IDENTIFICATION FOR CONTROL: OPTIMAL INPUT DESIGN WITH RESPECT TO A WORST-CASE ν-GAP COST FUNCTION

ROLAND HILDEBRAND, AND MICHEL GEVERS*

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. The subject of the present contribution is to 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.

Key words. identification for control, worst-case ν-gap, parametric uncertainty region

AMS subject classifications. Primary 93E12; Secondary 93D21, 49M05

1. Introduction.

This paper continues the line of research that aims at connecting prediction error identification methods with robust control theory ([2],[3],[4],[10]). 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, it is desirable 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 allows one to connect the "size" of an uncertainty set with a set of robustly stabilizing controllers is the worst-case ν-gap $\delta_{WC}(\hat{G},\mathcal{D})$ introduced in [10]. It is the supremum of the Vinnicombe ν-gap (see e.g. [28]) between the identified model $\hat{G}$ and all plants in the uncertainty set $\mathcal{D}$. Specifically, if $\delta_{WC}(\hat{G},\mathcal{D}) = \beta$, then all controllers C that stabilize the model $\hat{G}$ with a stability margin $b_{\hat{G},C} > \beta$ stabilize all plants in $\mathcal{D}$.

In previous papers ([3],[4],[10]) a special type of uncertainty sets $\mathcal{D}$ 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. The latter is defined to be the probability with which the true plant is lying inside the considered uncertainty set.

The goal of this paper is to minimize the worst-case ν-gap of such uncertainty regions $\mathcal{D}$ 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 throughout the last century. Important results were obtained by Kiefer

*CESAME, Université Catholique de Louvain, Bâtiment EULER, 4 av. Georges Lemaitre, 1348 Louvain-la-Neuve, Belgium ({Hildebrand, Gevers}@auto.ucl.ac.be). The European Commission is herewith acknowledged for its financial support in part to the research reported on in this paper. 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).

and Wolfowitz (see e.g. [15],[16]), Fedorov (e.g. [9]), Mehra (e.g. [20],[21]), Goodwin, Payne and Zarrop (e.g. [12],[30]) and others.

We shall 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. [30]). For typical number of data this leads to a sufficiently good approximation of the optimal input. The latter can be obtained only by computationally expensive time domain optimization (see e.g. [11],[23],[26],[6]). 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, i.e. the one that minimizes the considered cost function, 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 common cost functions are $\ln(\det \bar{M}^{-1})$ (D-optimality), $\mathrm{tr}\bar{M}^{-1}$ (A-optimality), $\mathrm{tr}W\bar{M}^{-1}$, where $W \geq 0$ (L-optimality), $\lambda_{\max}(\bar{M}^{-1})$ (E-optimality), $\Phi_s = (p^{-1}\mathrm{tr}\bar{M}^{-s})^{1/s}$, where p is the dimension of the parameter vector and $s = 0, 1, \dots, \infty$ (Φ-optimality). All mentioned cost functions except $\Phi_\infty = \lambda_{\max}(\bar{M}^{-1})$ depend analytically on the entries of $\bar{M}$ and Kiefer-Wolfowitz theory can effectively be applied to them (see [15]). All above-mentioned criteria are convex and monotonic with respect to $\bar{M}$ (see [30, p.39]).

In this paper, we optimize the input power spectrum with respect to the worst-case ν-gap of the uncertainty region $\mathcal{D}$. This is a nonstandard cost function, which is nonsmooth and thus more difficult to treat than the common above-mentioned 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 (see [15, section 4G]) and application of Kiefer-Wolfowitz theory does not make them more tractable. However, the proposed criteria satisfy the natural condition of monotonicity with respect to $\bar{M}$, as well as the condition of quasiconvexity, which is slightly weaker than convexity.

It follows from a classical result on trigonometric moment spaces (see [14, chapter VI, Theorem 4.1]) 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. [5]. Since the worst-case ν-gap and the other proposed criteria are quasiconvex with respect to the input power spectrum, the apparatus of convex analysis and the theory of LMI's can be applied to solve this optimization problem. For recent results in convex optimization see e.g. [22].

In the last years several authors successfully treated input design problems arising in Identification for Control with convex optimization methods. In [18], 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 the 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 ([5],[29]).

In this paper we optimize over the whole set of nonnegative input power spectra. It can be shown [30] that under the assumptions made above 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. In the last century important contributions were made by Krein (see e.g. [17]), Karlin, Shapley [13] and others. For a comprehensive treatment, see textbook [14] by Karlin and Studden. It follows from a well-known fact of Tchebycheff system theory (see e.g. [14]) 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. A restatement of this assertion is provided in Theorem 3.6 in this paper. In view of this, we propose an algorithm that yields optimal input power spectra which are discrete. Given the result just quoted, this is in no way a restriction. 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. [30, p.29]) or other functions.

Another approach, which leads to a suboptimal discrete input power spectrum, was proposed by van den Eijnde and Schoukens ([25],[27]). Here a finite subset of frequencies is prespecified and the optimal input power spectrum is sought within this subset. Advantages of this suboptimal method are less computational effort and an easier way to generate an input signal with the desired spectrum.

Let us mention also the paper [7], where identification in the ν-gap metric was treated outside the context of input design. The identification of a model was performed from a set of frequency response measurements in a way that aimed at minimizing the ν-gap between the true plant and the model.

We stress that 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 paper easily carry over to other model structures and to input power or output power/energy constraints.

The remainder of the paper 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 will show that the set over which the optimization takes place is amenable to an LMI formulation. 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. Sections 3 to 5 are the key part of the paper. The results obtained therein allow the problem to be treated with standard convex analysis methods. In section 6 we provide some results that are useful for designing stopping criteria for iterative search algorithms and quality assessment of the solution. 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 task is accomplished in section 7. In section 8 we present a simulation example, which demonstrates the superiority of the proposed cost functions over the classical design criteria D- and E-optimality. Finally, in section 9 we draw some conclusions.

2. Problem setting. Let us consider an ARX model structure

$$y(t) + a_1 y(t-1) + \ldots + a_{n_a} y(t-n_a) = b_1 u(t-n_k) + \ldots + b_{n_b} u(t-n_k-n_b+1) + e(t),$$

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

$$y = z^{-n_k+1}\frac{b_1 z^{-1} + \ldots + b_{n_b} z^{-n_b}}{1 + a_1 z^{-1} + \ldots + a_{n_a} z^{-n_a}} u + \frac{1}{1 + a_1 z^{-1} + \ldots + a_{n_a} z^{-n_a}} e =$$
$$= z^{-n_k+1}\frac{B(\theta)}{A(\theta)} u + \frac{1}{A(\theta)} e = G(\theta) u + \frac{1}{A(\theta)} e,$$

where A, B are obviously defined polynomials in the delay operator with coefficients depending on the parameter vector. Note that by our stability assumption A has no zeros on the unit circle and hence $|A|^2$ is strictly positive there.

Suppose 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 (see [19]) that the estimate $\hat{\theta}$ is asymptotically unbiased as $N \to \infty$ and its covariance for large N is given by $E(\theta_0 - \hat{\theta})(\theta_0 - \hat{\theta})^T \approx \frac{\lambda_0}{N}(\bar{E}\psi\psi^T)^{-1}$, where $\psi^T = (-z^{-1}y, \ldots, -z^{-n_a}y, z^{-n_k}u, \ldots, z^{-n_k-n_b+1}u)$ is the gradient of the predictor with respect to θ at $\theta = \theta_0$. The power spectrum of ψ is given by

$$\Phi_\psi = \begin{pmatrix} -z^{-n_k}\frac{B}{A} \\ \vdots \\ -z^{-n_a-n_k+1}\frac{B}{A} \\ z^{-n_k} \\ \vdots \\ z^{-n_k-n_b+1} \end{pmatrix} \Phi_u(-z^{n_k}\frac{\bar{B}}{\bar{A}} \cdots z^{n_k+n_b-1}) + \begin{pmatrix} -\frac{z^{-1}}{A} \\ \vdots \\ -\frac{z^{-n_a}}{A} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \lambda_0(-\frac{z}{\bar{A}} \cdots 0).$$

This yields the following asymptotic expression for the parameter covariance.

$$E(\theta_0 - \hat{\theta})(\theta_0 - \hat{\theta})^T \approx \left(\frac{N}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|A|^2} \left(\frac{\Phi_u}{\lambda_0} \begin{pmatrix} -z^{-1}B \\ \vdots \\ -z^{-n_a}B \\ z^{-1}A \\ \vdots \\ z^{-n_b}A \end{pmatrix} \begin{pmatrix} -z^{-1}B \\ \vdots \\ -z^{-n_a}B \\ z^{-1}A \\ \vdots \\ z^{-n_b}A \end{pmatrix}^* + \begin{pmatrix} -z^{-1} \\ \vdots \\ -z^{-n_a} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} -z^{-1} \\ \vdots \\ -z^{-n_a} \\ 0 \\ \vdots \\ 0 \end{pmatrix}^* \right) d\omega \right)^{-1} = M^{-1},$$

where $z = e^{j\omega}$. The inverse of the parameter covariance matrix is the Fisher information matrix. Let us denote the asymptotic expression for the information matrix by M and the average information matrix per data sample (see e.g. [30, p.24]) by $\bar{M}$,

$\bar{M} = \frac{1}{N}M$. We obtain

$$\bar{M} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Phi_u}{\lambda_0|A|^2} \begin{pmatrix} |B|^2 \begin{pmatrix} 1 & \dots & z^{n_a-1} \\ \vdots & \ddots & \vdots \\ z^{-n_a+1} & \dots & 1 \end{pmatrix} & -B\bar{A} \begin{pmatrix} 1 & \dots & z^{n_b-1} \\ \vdots & \ddots & \vdots \\ z^{-n_a+1} & \dots & z^{n_b-n_a} \end{pmatrix} \\ -\bar{B}A \begin{pmatrix} 1 & \dots & z^{n_a-1} \\ \vdots & \ddots & \vdots \\ z^{-n_b+1} & \dots & z^{n_a-n_b} \end{pmatrix} & |A|^2 \begin{pmatrix} 1 & \dots & z^{n_b-1} \\ \vdots & \ddots & \vdots \\ z^{-n_b+1} & \dots & 1 \end{pmatrix} \end{pmatrix}$$

$$+\frac{1}{|A|^2} \begin{pmatrix} \begin{matrix} 1 & \dots & z^{n_a-1} \\ \vdots & \ddots & \vdots \\ z^{-n_a+1} & \dots & 1 \end{matrix} & 0 \\ 0 & 0 \end{pmatrix} d\omega. \tag{2.1}$$

Note that in the expansion of $\bar{M}$, we have $A = A(\theta_0)$, $B = B(\theta_0)$. Since the parameter estimate $\hat{\theta}$ is asymptotically normally distributed (see [19]), we can assume, following [10], that the true parameter vector θ_0 lies with a prespecified probability $\alpha \in (0,1)$ in the uncertainty ellipsoid

$$U = \left\{ \theta | \frac{N}{\chi^2_{n_a+n_b}(\alpha)} (\theta - \hat{\theta})^T \bar{M}(\theta - \hat{\theta}) < 1 \right\}, \tag{2.2}$$

where χ^2_l is the χ^2 probability distribution with l degrees of freedom.

The uncertainty ellipsoid U corresponds to an uncertainty set

$$\mathcal{D} = \left\{ G(z,\theta) = z^{-n_k+1}\frac{B(\theta)}{A(\theta)} | \theta \in U \right\} = \left\{ G(z,\theta) = \frac{Z_N(z)\theta}{1 + Z_D(z)\theta} | \theta \in U \right\}$$

in the space of transfer functions. Here

$$Z_N = z^{-n_k+1}(0 \ \cdots \ 0 \ z^{-1} \ \cdots \ z^{-n_b}), \qquad Z_D = (z^{-1} \ \cdots \ z^{-n_a} \ 0 \ \cdots \ 0) \tag{2.3}$$

are row vectors of dimension $n_a + n_b$. The set $\mathcal{D}$ belongs to the class of generic prediction error model uncertainty sets as defined in [10].

The worst-case ν-gap between the identified model $G(\hat{\theta})$ and the uncertainty region $\mathcal{D}$ is defined by

$$\delta_{WC}(G(\hat{\theta}), \mathcal{D}) = \sup_{\theta \in U} \delta_\nu(G(\hat{\theta}), G(\theta)), \tag{2.4}$$

where δ_ν denotes the Vinnicombe ν-gap between two plants (see [28]). Since $G(\hat{\theta})$ belongs to $\mathcal{D}$, the worst-case ν-gap can be expressed in the following way (see [10, Lemma 5.1]).

$$\delta_{WC}(G(\hat{\theta}), \mathcal{D}) = \sup_{\omega \in [0,\pi]} \kappa_{WC}(G(e^{j\omega}, \hat{\theta}), \mathcal{D}), \tag{2.5}$$

where $\kappa_{WC}(G(e^{j\omega}, \hat{\theta}), \mathcal{D})$ is called the worst-case chordal distance between $G(\hat{\theta})$ and $\mathcal{D}$ at frequency ω and is defined by

$$\kappa_{WC}(G(e^{j\omega}, \hat{\theta}), \mathcal{D}) = \sup_{\theta \in U} \frac{|G(e^{j\omega}, \hat{\theta}) - G(e^{j\omega}, \theta)|}{\sqrt{(1 + |G(e^{j\omega}, \hat{\theta})|^2)(1 + |G(e^{j\omega}, \theta)|^2)}}. \tag{2.6}$$

The worst-case ν-gap is directly related to the robustness properties of the uncertainty region $\mathcal{D}$. The smaller it is, the larger is the set of controllers stabilizing simultaneously all plants in $\mathcal{D}$. Therefore our primary goal shall be to minimize the quantity $\delta_{WC}(G(\hat{\theta}), \mathcal{D}) = \max_{\omega \in [0,\pi]} \kappa_{WC}(G(e^{j\omega}, \hat{\theta}), \mathcal{D})$ by choosing an input with an appropriate power spectrum.

To be more precise, by input spectrum we mean a nonnegative measure on $[-\pi, \pi]$ such that the equality $\int_{-\pi}^{\pi} \Phi_u \varphi(\omega)\, d\omega = \int_{-\pi}^{\pi} \Phi_u \varphi(-\omega)\, d\omega$ holds for all functions $\varphi(\omega) \in C^\infty([-\pi, \pi])$. To any such measure Φ_u on $[-\pi, \pi]$ corresponds a unique nonnegative measure $\bar{\Phi}_u$ on $[0, \pi]$ such that $\int_{-\pi}^{\pi} \Phi_u \varphi(\omega)\, d\omega = \int_0^{\pi} \bar{\Phi}_u \frac{\varphi(\omega)+\varphi(-\omega)}{2}\, d\omega$ for all $\varphi \in C^\infty([-\pi, \pi])$. For details of constructing $\bar{\Phi}_u$ from Φ_u see e.g. [30, p.23]. In the sequel we will denote the single-sided measure $\bar{\Phi}_u$ also by Φ_u. Since the measures are defined on different intervals, confusion is excluded.

To restrict the class of admissible power spectra we impose an input energy constraint

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega \leq c, \tag{2.7}$$

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

The worst-case ν-gap depends on Φ_u via the average per data sample information matrix $\bar{M}$, which enters in the expression for the set U. Furthermore, it depends via $\bar{M}$ 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 naturally not available before the identification experiment is performed. 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. Denote this value by $\bar{\theta}$, and denote the approximation of λ_0 by $\bar{\lambda}$.

We can now formulate our main

Problem 1 Find Φ_u satisfying (2.7) such that $\bar{M}(\Phi_u)$ defined by (2.1) minimizes the cost function $\mathcal{J}_1 = \delta_{WC}(G(\hat{\theta}), \mathcal{D})$ defined by equations (2.5),(2.6).

Along with the worst-case ν-gap of the uncertainty region $\mathcal{D}$, we will consider another cost function, which is easier to compute and is an approximation of δ_{WC}.

Let us approximate cost function $\mathcal{J}_1 = \mathcal{J}_1(\bar{M})$ by its asymptotic expression for large information matrices. For a fixed positive definite matrix $\bar{M}_0$ the size of the parameter ellipsoid U defined by any multiple $\bar{M} = \beta \bar{M}_0$ of $\bar{M}_0$, where $\beta > 0$, is proportional to $\beta^{-1/2}$. Since for small ellipsoids the worst-case ν-gap is asymptotically proportional to the size of the former, it follows that for large β the value of $\mathcal{J}_1(\bar{M})$ diminishes asymptotically proportionately to $\beta^{-1/2}$. Thus we can approximate $\mathcal{J}_1$ by

$$\mathcal{J}_2 = \lim_{\varepsilon \to 0} \frac{\mathcal{J}_1(\varepsilon^{-2} \bar{M})}{\varepsilon}. \tag{2.8}$$

Problem 2 Find Φ_u satisfying (2.7) such that $\bar{M}(\Phi_u)$ defined by (2.1) minimizes cost function $\mathcal{J}_2$ defined by equation (2.8).

The goal of the present paper is the development of numerical algorithms for solving both Problems 1 and 2. There is a two-fold reason for introducing cost function $\mathcal{J}_2$. Beside its much lower computational complexity, it turns out that identification with an input power spectrum minimizing $\mathcal{J}_2$ in many cases gives better results than one with an input power spectrum minimizing $\mathcal{J}_1$. This apparently counter-intuitive observation has the following reason. Both cost functions depend on the identified parameter value $\hat{\theta}$, the true parameter value θ_0 and the noise covariance λ_0. As mentioned above, these quantities are unknown and must be replaced by estimates obtained e.g. from a preliminary identification experiment. This approximation introduces an error to the argument of the minimum of the cost functions $\mathcal{J}_1$ and $\mathcal{J}_2$, i.e. to the solutions of Problems 1 and 2. Now simulations show that the impact of this effect on $\arg\min \mathcal{J}_2$ is lower than that on $\arg\min \mathcal{J}_1$ and that this difference as a rule overweighs the error introduced by approximating cost function $\mathcal{J}_1$ by $\mathcal{J}_2$. We will address this issue again 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 due to Payne and Goodwin [24].

PROPOSITION 3.1. *The average information 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.*

We find it convenient to give a proof here in order to provide explicit expressions that clarify the structure of $\bar{M}$.

Proof. Define $a_0 = 1$ and $n = n_a + n_b - 1$. Then we have

$$
\begin{aligned}
|B|^2 &= \sum_{k=-(n_b-1)}^{n_b-1} \left(\sum_{j=\max(1,1-k)}^{\min(n_b,n_b-k)} b_{j+k} b_j \right) z^k, \\
-B\bar{A} &= \sum_{k=-n_b}^{n_a-1} \left(\sum_{j=\max(1,-k)}^{\min(n_b,n_a-k)} -a_{j+k} b_j \right) z^k, \\
|A|^2 &= \sum_{k=-n_a}^{n_a} \left(\sum_{j=\max(0,-k)}^{\min(n_a,n_a-k)} a_{j+k} a_j \right) z^k. \qquad (3.1)
\end{aligned}
$$

Using (3.1) in (2.1) and ordering by powers of z, we can rewrite (2.1) as $\bar{M} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Phi_u}{\lambda_0|A|^2}\left(\sum_{i=-n}^{n} \tilde{M}_i z^i\right) d\omega + \tilde{M}$. The matrices $\tilde{M}_i$ are constant and depend only on the coefficients of A and B. By $\tilde{M}$ the integral over the second term in (2.1) is denoted. It is a constant matrix and independent of Φ_u. $\tilde{M}$ is most easily computed using the method proposed in [19, p.50]. Note that $\tilde{M}_i = \tilde{M}_{-i}^T$. Hence we obtain

$$
\begin{aligned}
\bar{M} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Phi_u}{\lambda_0|A|^2}\, d\omega \tilde{M}_0 + \sum_{i=1}^{n}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Phi_u}{\lambda_0|A|^2} z^i \, d\omega (\tilde{M}_i + \tilde{M}_i^T) \right) + \tilde{M} \\
&= \frac{1}{\pi}\int_{0}^{\pi} \frac{\Phi_u}{\lambda_0|A|^2}\, d\omega \frac{\tilde{M}_0}{2} + \sum_{i=1}^{n}\left(\frac{1}{\pi}\int_{0}^{\pi} \frac{\Phi_u}{\lambda_0|A|^2} \cos(i\omega)\, d\omega \frac{\tilde{M}_i + \tilde{M}_i^T}{2} \right) + \tilde{M}. \qquad (3.2)
\end{aligned}
$$

Thus $\bar{M}$ is contained in the $(n+1)$-dimensional affine subspace that is spanned by $\tilde{M}_0, \tilde{M}_i + \tilde{M}_i^T$, $i = 1,\dots,n$, and shifted by $\tilde{M}$. This completes the proof. $\square$

Let us compose a vector $\tilde{x} \in \mathbf{R}^{n+1}$ of real numbers $\tilde{x}_i$, $i = 0,\dots,n$, defined by $\tilde{x}_i = \frac{1}{\pi}\int_0^{\pi} \frac{\Phi_u}{\lambda_0|A|^2}\cos(i\omega)\, d\omega$.

DEFINITION 3.2. *The quantities $\tilde{x}_k = \frac{1}{\pi}\int_0^{\pi} \frac{\Phi_u}{\lambda_0|A|^2}\cos(k\omega)\, d\omega$, $k \in \mathbf{N}$, are called* trigonometric moments *of the measure $\frac{\Phi_u}{\pi\lambda_0|A|^2}$.*

Since $\frac{1}{\pi\lambda_0|A|^2}$ is strictly positive on $\omega \in [0,\pi]$, we have the following result [30].

PROPOSITION 3.3. *The set* $\{\tilde{x}(\Phi_u) \,|\, \Phi_u$ *is a nonnegative measure on* $[0,\pi]\}$ *equals the moment space* $\mathcal{M}^{(n+1)}$ *of the Tchebycheff system* $\{1, \cos\omega, \dots, \cos n\omega\}$ *on* $[0,\pi]$.

For definition and properties of moment spaces see e.g. [14].

The characterization of the space $\mathcal{M}^{(n+1)}$ is a special case of the extensively studied classical trigonometric moment problem. The following theorem is a consequence of the general result [14, Chapter VI, Theorem 4.1]. It asserts that $\mathcal{M}^{(n+1)}$ can be characterized as the feasible set of an LMI.

THEOREM 3.4. *A point* $\tilde{x} \in \mathbf{R}^{n+1}$ *belongs to the space* $\mathcal{M}^{(n+1)}$ *if and only if the Töplitz matrix composed of the elements of* $\tilde{x}$ *is positive semidefinite, i.e.*

$$T(\tilde{x}) = \begin{pmatrix} \tilde{x}_0 & \tilde{x}_1 & \ddots & \tilde{x}_n \\ \tilde{x}_1 & \tilde{x}_0 & \ddots & \tilde{x}_{n-1} \\ \ddots & \ddots & \ddots & \ddots \\ \tilde{x}_n & \tilde{x}_{n-1} & \ddots & \tilde{x}_0 \end{pmatrix} \geq 0. \qquad \square \tag{3.3}$$

Since $n+1 \geq 2$, the strict LMI $T(\tilde{x}) > 0$ is feasible. Hence the feasible set of the strict version is the interior of $\mathcal{M}^{(n+1)}$ (see [5, section 2.5]). By $\mathcal{M}$ denote the set of average information matrices corresponding to the interior of $\mathcal{M}^{(n+1)}$. From (3.2) we have

$$\mathcal{M} = \left\{ \bar{M}(\tilde{x}) = \tilde{x}_0 \frac{\tilde{M}_0}{2} + \sum_{i=1}^{n} \tilde{x}_i \frac{\tilde{M}_i + \tilde{M}_i^T}{2} + \tilde{M} \,|\, T(\tilde{x}) > 0 \right\}.$$

DEFINITION 3.5. *Let* Φ_u *be a discrete double-sided power spectrum with support* $supp\,\Phi_u \subset [-\pi,\pi]$. *The number of points in the intersection* $supp\,\Phi_u \cap [-\pi,\pi)$, *divided by two, is called the index of* Φ_u: $index(\Phi_u) = \frac{1}{2}\#(supp\,\Phi_u \cap [-\pi,\pi))$. *The index of a single-sided nonnegative discrete measure on* $[0,\pi]$ *is defined as the index of the corresponding double-sided power spectrum.*

Remark. This definition of the index is consistent with its definition for nonnegative discrete measures on the interval $[0,\pi]$ (see e.g. [14]).

The notion of the index also allows us to characterize the interior of the moment space $\mathcal{M}^{(n+1)}$. The following theorem is a standard result on moment spaces.

THEOREM 3.6. *(see e.g. [14]) Let* $\tilde{x}$ *be a point in* $\mathcal{M}^{(n+1)}$. *Then the following conditions hold.*

i) $\tilde{x} \in Bd(\mathcal{M}^{(n+1)})$ *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.*

ii) $\tilde{x} \in Int(\mathcal{M}^{(n+1)})$ *if and only if there exists a discrete nonnegative measure on* $[0,\pi]$ *with index* $\frac{n+1}{2}$ *that induces* $\tilde{x}$. *There are exactly two such measures. Exactly one of them contains the frequency* π.

iii) Let $\tilde{x} \in Int(\mathcal{M}^{(n+1)})$ *and* $\omega \in [0,\pi]$. *Then there exists a unique discrete nonnegative measure on* $[0,\pi]$ *which induces* $\tilde{x}$, *has index not exceeding* $\frac{n+2}{2}$, *and contains the frequency* ω. $\square$

Remark. Measures with index $\frac{n+1}{2}$ which induce $\tilde{x}$ are called *principal realizations* of $\tilde{x}$. The one containing π is called *upper principal*, the other *lower principal* realization. Measures with index not exceeding $\frac{n+2}{2}$ are called *canonical.*

We see that the interior of $\mathcal{M}^{(n+1)}$ is characterized by those points $\tilde{x}$ which can be represented by a discrete measure with index not less than $\frac{n+1}{2}$.

Now we shall characterize the set of input power spectra Φ_u that lead to nonsingular average information matrices $\bar{M}$.

PROPOSITION 3.7. *Let Φ_u be a power spectrum and $\bar{M}$ the corresponding average information matrix. Then $\bar{M}$ is singular if and only if Φ_u is discrete and its index is less than $\frac{n_b}{2}$.*

Proof. "$\Rightarrow$": Suppose $\bar{M}(\Phi_u)$ is singular. Then there exists a nonzero vector $v = (v_1, \dots, v_{n_a+n_b})^T \in \mathbf{R}^{n_a+n_b}$ such that $v^T\bar{M}v = 0$. Expanding $\bar{M}$, we obtain

$$v^T\bar{M}v = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\Phi_u}{\lambda_0|A|^2}|-v_1z^{-1}B-\ldots-v_{n_a}z^{-n_a}B+v_{n_a+1}z^{-1}A+\ldots+v_{n_a+n_b}z^{-n_b}A|^2 \right.$$
$$\left. +\frac{1}{|A|^2}|-v_1z^{-1}-\ldots-v_{n_a}z^{-n_a}|^2\right)d\omega = 0$$

with $z = e^{j\omega}$. This yields $-v_1z^{-1} - \ldots - v_{n_a}z^{-n_a} = 0$ for all z on the unit circle and $-v_1z^{-1}B - \ldots - v_{n_a}z^{-n_a}B + v_{n_a+1}z^{-1}A + \ldots + v_{n_a+n_b}z^{-n_b}A = 0$ for all $z = e^{j\omega}$ such that $\omega \in \operatorname{supp}\Phi_u$. From the first identity we obtain $v_1 = \ldots = v_{n_a} = 0$. Inserting this in the second equality, we get $v_{n_a+1} + \ldots + v_{n_a+n_b}z^{-n_b+1} = 0$. Since $v \neq 0$, this equation can have at most $n_b - 1$ different roots. Since Φ_u has to be concentrated at these roots, it is discrete and its index cannot exceed $\frac{n_b-1}{2}$.

"$\Leftarrow$": Suppose Φ_u is discrete with index less than $\frac{n_b}{2}$. Denote the frequencies of Φ_u by $\omega_1, \dots, \omega_{k'}$. They correspond to k different points $z_1, \dots, z_k$ on the unit circle, where $k < n_b$. We have

$$\bar{M} = \frac{1}{2\pi}\sum_{i=1}^{k}\frac{\alpha_i}{\lambda_0|A(z_i)|^2}\begin{pmatrix}-z_i^{-1}B(z_i)\\ \vdots\\ -z_i^{-n_a}B(z_i)\\ z_i^{-1}A(z_i)\\ \vdots\\ z_i^{-n_b}A(z_i)\end{pmatrix}\begin{pmatrix}-z_i^{-1}B(z_i)\\ \vdots\\ -z_i^{-n_a}B(z_i)\\ z_i^{-1}A(z_i)\\ \vdots\\ z_i^{-n_b}A(z_i)\end{pmatrix}^* + \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{|A|^2}\begin{pmatrix}-z^{-1}\\ \vdots\\ -z^{-n_a}\\ 0\\ \vdots\\ 0\end{pmatrix}\begin{pmatrix}-z^{-1}\\ \vdots\\ -z^{-n_a}\\ 0\\ \vdots\\ 0\end{pmatrix}^* d\omega.$$

Here $\alpha_i > 0$ are the weightings of the different frequencies. It is easily seen that the matrices under the sign of the sum are of (complex) rank one, while the integral is a matrix which has a rank of at most n_a. Thus the rank of $\bar{M}$ does not exceed $n_a + n_b - 1$ and $\bar{M}$ is singular. This concludes the proof. $\Box$

COROLLARY 3.8. *Any $\bar{M} \in \mathcal{M}$ is strictly positive definite.* $\Box$

This corollary ensures the existence of the inverse $\bar{M}^{-1}$ in the interior of the search space.

By inspecting (2.2),(2.4) and (2.8), the reader will have no difficulty to prove the following monotonicity property.

PROPOSITION 3.9. *Let $\bar{M}_1, \bar{M}_2$ be two positive semidefinite average information matrices, and suppose $\bar{M}_1 \leq \bar{M}_2$. Then the values of the cost functions $\mathcal{J}_1, \mathcal{J}_2$ at $\bar{M}_2$ do not exceed the respective values at $\bar{M}_1$.* $\Box$

Now we shall include the input energy constraint (2.7) into our framework. By Proposition 3.9, for any constant $\beta > 1$ the value of each of the considered cost functions at a particular input power spectrum Φ_u will be not less than its value at the input power spectrum $\beta\Phi_u$. Thus we can replace constraint (2.7) by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_u(\omega)d\omega = c. \qquad (3.4)$$

In [30] it was shown that relations like (3.4) determine affine hyperplanes in the space of feasible average information matrices. Indeed, we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_u(\omega)d\omega = \frac{\lambda_0}{2\pi}\int_{-\pi}^{\pi}\frac{\Phi_u}{\lambda_0|A|^2}\left(\sum_{i=0}^{n_a}a_i^2+\sum_{i=1}^{n_a}\left(2\sum_{k=0}^{n_a-i}a_{k+i}a_k\right)\cos i\omega\right)d\omega =$$
$$= \lambda_0\left(\tilde{x}_0\sum_{i=0}^{n_a}a_i^2+\sum_{i=1}^{n_a}\tilde{x}_i\left(2\sum_{k=0}^{n_a-i}a_{k+i}a_k\right)\right) = c.$$

Thus we get

$$\tilde{x}_0 = \frac{1}{\sum_{i=0}^{n_a}a_i^2}\left(\frac{c}{\lambda_0}-\sum_{i=1}^{n_a}\tilde{x}_i\left(2\sum_{k=0}^{n_a-i}a_{k+i}a_k\right)\right). \tag{3.5}$$

Inserting (3.5) into (3.3), we obtain an LMI on the variables $\tilde{x}_1,\dots,\tilde{x}_n$, i.e. in an n-dimensional space instead of the initial $n+1$-dimensional one. The feasible set of LMI (3.5),(3.3) is a subset of $\mathbf{R}^n$, parametrized by new variables $x_1,\dots,x_n$, which we define by $x_1=\tilde{x}_1,\dots,x_n=\tilde{x}_n$. Denote by $\mathcal{X}_c$ the interior of this set and by $\mathcal{M}_c$ the set of average information matrices corresponding to points in $\mathcal{X}_c$. Thus the optimization takes place over the closure of $\mathcal{X}_c$. Let us stack the variables $x_1,\dots,x_n$ into a vector $x\in\mathbf{R}^n$, to be distinguished from $\tilde{x}\in\mathbf{R}^{n+1}$. While the latter parametrizes the set $\mathcal{M}$, the former parametrizes the set $\mathcal{M}_c$ or $\mathcal{X}_c$.

Using (3.5), we can represent average information matrices in the closure of $\mathcal{M}_c$ as affine functions of the variables $x_1,\dots,x_n$. We have

$$\bar{M} = \bar{M}_0+\sum_{i=1}^{n}x_i\bar{M}_i, \tag{3.6}$$

where

$$\begin{aligned}
\bar{M}_0 &= \frac{c}{2\lambda_0\sum_{i=0}^{n_a}a_i^2}\tilde{M}_0+\tilde{M},\\
\bar{M}_i &= \frac{\tilde{M}_i+\tilde{M}_i^T}{2}-\frac{\sum_{k=0}^{n_a-i}a_{k+i}a_k}{\sum_{i=0}^{n_a}a_i^2}\tilde{M}_0, \qquad i=1,\dots,n_a,\\
\bar{M}_i &= \frac{\tilde{M}_i+\tilde{M}_i^T}{2}, \qquad i=n_a+1,\dots,n.
\end{aligned}$$

Thus the closure of $\mathcal{M}_c$ is contained in an n-dimensional affine subspace of the space of symmetric $(n_a+n_b)\times(n_a+n_b)$-matrices.

PROPOSITION 3.10. *The search space of Problems 1 and 2 can be represented as a section of the trigonometric moment cone $\mathcal{M}^{(n+1)}$ and is thus a bounded closed n-dimensional convex set. It is parametrized by the variables $x_1,\dots,x_n$.*

Proof. What is left to prove is that relation (3.5) defines a section of the moment cone $\mathcal{M}^{(n+1)}$. Let $\tilde{x}$ be an arbitrary non-zero moment point and $\Phi_u(\omega)$ a measure generating this moment point. Then the ray $\beta\tilde{x}$, $\beta>0$, will be generated by the ray $\beta\Phi_u(\omega)$ of measures. On the latter, exactly one measure satisfies relation (3.4). Therefore exactly one point on the ray $\beta\tilde{x}$ satisfies relation (3.5). □

In this section 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 finite-dimensional problem of searching the minimum over a section of a moment cone. Moreover, we described the search space as an LMI, namely (3.3), (3.5), and showed that it is bounded.

4. Quasiconvexity. In the previous section we proved the search space to be a bounded convex set. In this section we prove quasiconvexity of cost functions $\mathcal{J}_1, \mathcal{J}_2$ and thus of Problems 1 and 2.

PROPOSITION 4.1. *On $\mathcal{M}$ cost function $\mathcal{J}_1$ is quasiconvex with respect to $\bar{M}$.*

Proof. The worst-case chordal distance can be expressed as a solution to a generalized eigenvalue problem (GEVP) [10, Theorem 5.1]. We have $\kappa_{WC}(G(e^{j\omega}, \hat{\theta}), \mathcal{D}) = \sqrt{\gamma_{opt}}$, where γ_{opt} is the solution of the GEVP

$$\text{minimize } \gamma \text{ subject to } \quad F_0 + \gamma F_1 + \tau R \geq 0,\ \tau \geq 0. \tag{4.1}$$

Here F_0, F_1, R are symmetric matrices given by

$$F_0 = V \begin{pmatrix} -1 & 0 & -ImG(e^{j\omega}, \hat{\theta}) & ReG(e^{j\omega}, \hat{\theta}) \\ 0 & -1 & ReG(e^{j\omega}, \hat{\theta}) & ImG(e^{j\omega}, \hat{\theta}) \\ -ImG(e^{j\omega}, \hat{\theta}) & ReG(e^{j\omega}, \hat{\theta}) & -|G(e^{j\omega}, \hat{\theta})|^2 & 0 \\ ReG(e^{j\omega}, \hat{\theta}) & ImG(e^{j\omega}, \hat{\theta}) & 0 & -|G(e^{j\omega}, \hat{\theta})|^2 \end{pmatrix} V^T,$$

$$F_1 = (1 + |G(e^{j\omega}, \hat{\theta})|^2) V V^T,$$

$$R = \begin{pmatrix} I_{n_a+n_b} \\ -\hat{\theta}^T \end{pmatrix} \bar{M} \begin{pmatrix} I_{n_a+n_b} & -\hat{\theta} \end{pmatrix} - \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\chi^2_{n_a+n_b}(\alpha)}{N} \end{pmatrix}, \tag{4.2}$$

where 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_D given by (2.3).

We will now show that γ_{opt} is quasiconvex with respect to R. Choose $\lambda \in (0, 1)$ and let R_1, R_2 be symmetric matrices of appropriate dimension. Suppose γ, τ_1, τ_2 are nonnegative numbers such that $F_0 + \gamma F_1 + \tau_1 R_1 \geq 0, F_0 + \gamma F_1 + \tau_2 R_2 \geq 0$. We have to show that there exists $\tau \geq 0$ such that $F_0 + \gamma F_1 + \tau(\lambda R_1 + (1 - \lambda)R_2) \geq 0$. If $\tau_1 = 0$ or $\tau_2 = 0$, then we can choose $\tau = 0$. Let $\tau_1 \tau_2 > 0$. Define

$$\lambda' = \frac{\lambda \tau_2}{\lambda \tau_2 + (1 - \lambda)\tau_1}, \quad \tau = \frac{\tau_1 \tau_2}{\lambda \tau_2 + (1 - \lambda)\tau_1}.$$

Obviously $\lambda' \in (0, 1)$ and $\tau > 0$. It is easily verified that $\lambda\tau = \lambda'\tau_1, (1 - \lambda)\tau = (1 - \lambda')\tau_2$. Hence we have

$$F_0 + \gamma F_1 + \tau(\lambda R_1 + (1 - \lambda)R_2) = \lambda'(F_0 + \gamma F_1 + \tau_1 R_1) + (1 - \lambda')(F_0 + \gamma F_1 + \tau_2 R_2) \geq 0.$$

Thus if γ is feasible for $R = R_1$ and for $R = R_2$, then it is also feasible for any linear convex combination of R_1, R_2. It follows that $\gamma_{opt}|_{R=\lambda R_1+(1-\lambda)R_2} \leq \max\{\gamma_{opt}|_{R=R_1}, \gamma_{opt}|_{R=R_2}\}$, i.e. quasiconvexity of γ_{opt} with respect to R.

Suppose $\omega \in [0, \pi]$ is fixed. Note that R affinely depends on $\bar{M}$, while F_0 and F_1 are constant for given ω. Therefore γ_{opt} is quasiconvex with respect to $\bar{M}$ for fixed ω. But quasiconvexity is preserved under the operation of taking the maximum over a family of functions and under rescaling by a strictly monotonic function (in this case the square root). This completes the proof. □

PROPOSITION 4.2. *On $\mathcal{M}$ cost function $\mathcal{J}_2$ is quasiconvex with respect to $\bar{M}$.*

Proof. Let us compute cost function $\mathcal{J}_2$.

$$\mathcal{J}_2 = \lim_{\varepsilon\to 0}\frac{\mathcal{J}_1(\varepsilon^{-2}\bar{M})}{\varepsilon} = \sup_{\omega\in[0,\pi]}\lim_{\varepsilon\to 0}\left(\varepsilon^{-1}\sup_{z\in U_\varepsilon(\omega)}\frac{|G(e^{j\omega},\hat{\theta})-z|}{\sqrt{1+|G(e^{j\omega},\hat{\theta})|^2}\sqrt{1+|z|^2}}\right) =$$

$$= \sup_{\omega\in[0,\pi]}\frac{d}{d\varepsilon}|_{\varepsilon=0}\sup_{z\in U_\varepsilon(\omega)}\frac{|G(e^{j\omega},\hat{\theta})-z|}{\sqrt{1+|G(e^{j\omega},\hat{\theta})|^2}\sqrt{1+|z|^2}}.$$

Here $U_\varepsilon(\omega)$ denotes the set $\left\{z = G(e^{j\omega},\theta)|\frac{N}{\chi^2_{n_a+n_b}(\alpha)}(\theta-\hat{\theta})^T\bar{M}(\theta-\hat{\theta}) < \varepsilon^2\right\}$. The expression $\sqrt{1+|z|^2}$ tends to $\sqrt{1+|G(e^{j\omega},\hat{\theta})|^2}$ as $\varepsilon\to 0$, therefore

$$\mathcal{J}_2 = \sup_{\omega\in[0,\pi]}\frac{\frac{d}{d\varepsilon}|_{\varepsilon=0}\sup_{z\in U_\varepsilon(\omega)}|z-G(e^{j\omega},\hat{\theta})|}{1+|G(e^{j\omega},\hat{\theta})|^2} = \sup_{\omega\in[0,\pi]}\frac{\varepsilon^{-1}\sup_{z\in U_\varepsilon(\omega)}|T(\theta-\hat{\theta})|}{1+|G(e^{j\omega},\hat{\theta})|^2},$$

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

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

If T has full rank, then, following [2], we can write the term $\varepsilon^{-1}\sup_{z\in U_\varepsilon(\omega)}|T(\theta-\hat{\theta})|$ as $(\lambda_{\min}((T(\frac{N}{\chi^2_{n_a+n_b}(\alpha)}\bar{M})^{-1}T^T)^{-1}))^{-1/2} = \sqrt{\frac{\chi^2_{n_a+n_b}(\alpha)}{N}}(\lambda_{\max}(T\bar{M}^{-1}T^T))^{1/2}$. By $\lambda_{\min}$ and $\lambda_{\max}$ the minimal and maximal eigenvalue, respectively, are denoted.

If T is rank deficient, we can find vectors $w\in\mathbf{R}^{n_a+n_b}$ and $w_1\in\mathbf{R}^2$ such that $|w_1| = 1$ and $T = w_1w^T$. We exclude the trivial case $T = 0$ from consideration and assume $w\neq 0$. Then $\varepsilon^{-1}\sup_{z\in U_\varepsilon(\omega)}|T(\theta-\hat{\theta})| = ((w^T(\frac{N}{\chi^2_{n_a+n_b}(\alpha)}\bar{M})^{-1}w)^{-1})^{-1/2} = \sqrt{\frac{\chi^2_{n_a+n_b}(\alpha)}{N}}(w^T\bar{M}^{-1}w)^{1/2}$. But we have anyway $w^T\bar{M}^{-1}w = \lambda_{\max}(T\bar{M}^{-1}T^T)$.

Hence in either case we obtain

$$\mathcal{J}_2 = \sqrt{\frac{\chi^2_{n_a+n_b}(\alpha)}{N}}\sup_{\omega\in[0,\pi]}\frac{(\lambda_{\max}(T(\omega)\bar{M}^{-1}T(\omega)^T))^{1/2}}{1+|G(e^{j\omega},\hat{\theta})|^2}. \tag{4.3}$$

It is well-known that the inverse P^{-1} of a symmetric positive definite matrix P and the maximal eigenvalue $\lambda_{\max}(Q)$ of a symmetric positive semidefinite matrix Q are convex functions with respect to P or Q respectively (see e.g. [8]). Hence $\lambda_{\max}(T\bar{M}^{-1}T^T)$ is convex with respect to $\bar{M}$ for fixed ω. Since the operation of taking the maximum over a family of functions preserves convexity, we have that $\mathcal{J}_2^2$ is a convex function with respect to $\bar{M}$. By strict monotonicity of the square root this yields quasiconvexity of $\mathcal{J}_2$. □

In the preceding two sections we have shown that Problems 1 and 2 stated in section 2 are quasiconvex. In the next section we will provide the necessary tools that allow the user to apply standard convex algorithms to solve these problems numerically.

5. Cutting planes. Most methods in convex analysis are based on the notion of a cutting plane (see e.g. [5]). Suppose $S\subset\mathbf{R}^m$ is a convex set and $f : S\to\mathbf{R}$ is a quasiconvex function defined on S.

DEFINITION 5.1. *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 $g^T(x - x^{(0)}) \geq 0$.*

Thus the global minimum of f on S lies in the halfspace $\{x \,|\, g^T(x - x^{(0)}) \leq 0\}$. By definition of quasiconvexity a cutting plane always exists.

In this section we will compute cutting planes for cost functions $\mathcal{J}_1, \mathcal{J}_2$ at an arbitrary point $x^{(0)} \in \mathcal{X}_c$.

Let $\bar{M}^{(0)}$ be the average information matrix corresponding to $x^{(0)}$. By Corollary 3.8 the matrix $\bar{M}^{(0)}$ is positive definite.

We shall now compute a cutting plane for $\mathcal{J}_1 = \max_{\omega\in[0,\pi]} \kappa_{WC}(G(e^{j\omega}, \hat{\theta}), \mathcal{D})$. Denote by $\omega^{(0)}$ the frequency where the worst-case chordal distance κ_{WC} attains its maximum. The value of $\omega^{(0)}$ can be foung e.g. by a grid search with subsequent refinement using a denser grid in the vicinity of the maximum. It is easily seen that a cutting plane to the function $\kappa_{WC}(G(e^{j\omega^{(0)}}, \hat{\theta}), \mathcal{D})$ or its square will also be a cutting plane to $\mathcal{J}_1$. In the sequel we will assume that ω is equal to $\omega^{(0)}$ and omit it as argument.

Thus our goal is to find a cutting plane for the optimum value γ_{opt} of GEVP (4.1),(4.2), considered as a function of x. Note that F_0, F_1 are independent of x, while R depends on x via $\bar{M}$. By (3.6), we can represent R as $R(x) = R_0 + \sum_{i=1}^n x_i R_i$ with

$$R_0 = \begin{pmatrix} I_{n_a+n_b} \\ -\hat{\theta}^T \end{pmatrix} \bar{M}_0 \begin{pmatrix} I_{n_a+n_b} & -\hat{\theta} \end{pmatrix} - \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\chi^2_{n_a+n_b}(\alpha)}{N} \end{pmatrix},$$

$$R_i = \begin{pmatrix} I_{n_a+n_b} \\ -\hat{\theta}^T \end{pmatrix} \bar{M}_i \begin{pmatrix} I_{n_a+n_b} & -\hat{\theta} \end{pmatrix}.$$

Let $\gamma_{opt}^{(0)}, \tau_{opt}^{(0)}$ be the optimal values for γ, τ in GEVP (4.1),(4.2) at $x = x^{(0)}$. Then the matrix $F_0 + \gamma_{opt}^{(0)} F_1 + \tau_{opt}^{(0)} R$ is both singular and positive semidefinite. Let V^0 be the nullspace of this matrix.

PROPOSITION 5.2. *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 $\mathcal{J}_1$. If g is zero, $\mathcal{J}_1$ achieves a minimum at $x^{(0)}$.*

The proof of this proposition can be found in the Appendix.

Let us now compute a cutting plane for cost function $\mathcal{J}_2$, which is given by (4.3). 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 5.3. *Let $g \in \mathbf{R}^n$ be defined componentwise by $g_i = -v^T T(\omega^{(0)})\bar{M}^{-1}\bar{M}_i\bar{M}^{-1}T(\omega^{(0)})^T v$. Then g defines a cutting plane for the cost function $\mathcal{J}_2$ at $x^{(0)}$, if $g \neq 0$, and $\mathcal{J}_2$ attains a minimum at $x^{(0)}$, if $g = 0$.*

Proof. Consider $f(x) = \sqrt{\frac{\chi^2_{n_a+n_b}(\alpha)}{N}} \frac{(v^T T(\omega^{(0)})(\bar{M}_0 + \sum_{i=1}^n x_i \bar{M}_i)^{-1} T(\omega^{(0)})^T v)^{1/2}}{1+|G(e^{j\omega^{(0)}},\hat{\theta})|^2}$. By definition we have $f(x^{(0)}) = \mathcal{J}_2(x^{(0)})$, but $f(x) \leq \mathcal{J}_2(x)$ for any $x \in \mathcal{X}_c$.

Let $x \in \mathcal{X}_c$ be a point such that $g^T(x - x^{(0)}) \geq 0$. We shall show that $f(x^{(0)}) \leq f(x)$, which would imply $\mathcal{J}_2(x^{(0)}) \leq \mathcal{J}_2(x)$. This is equivalent to $\tilde{f}(x^{(0)}) \leq \tilde{f}(x)$, where $\tilde{f}$ is defined by

$$\begin{aligned}\tilde{f}(x) &= \frac{N}{\chi^2_{n_a+n_b}(\alpha)}(1+|G(e^{j\omega^{(0)}},\hat{\theta})|^2)^2 f^2(x) = \\ &= v^T T(\omega^{(0)})(\bar{M}_0 + \sum_{i=1}^n x_i \bar{M}_i)^{-1} T(\omega^{(0)})^T v = \\ &= \mathrm{tr}(T(\omega^{(0)})^T v v^T T(\omega^{(0)})(\bar{M}_0 + \sum_{i=1}^n x_i \bar{M}_i)^{-1}).\end{aligned}$$

In other words, we have to show that g defines a cutting plane for $\tilde{f}$. It is well-known (see e.g. [30, p.39]) that $\tilde{f}$, being of the form $\mathrm{tr} W\bar{M}^{-1}$ with $W \geq 0$, is a smooth convex function on $\mathcal{X}_c$. Hence a cutting plane to $\tilde{f}$ is defined by its gradient, which is identical to g.

If $g = 0$, then $\tilde{f}$ attains a minimum at $x^{(0)}$. Hence f attains a minimum at $x^{(0)}$, which yields $\mathcal{J}_2(x^{(0)}) \leq \mathcal{J}_2(x)$ for any $x \in \mathcal{X}_c$. This concludes the proof. □

The results of sections 3 to 5, i.e. the LMI description of the feasible set and the knowledge of cutting planes, allow the user to apply a whole range of convex optimization methods for solving Problems 1 and 2. For a description of different methods see e.g. [5],[22].

6. Error assessment of the solution. Suppose we seek the minimum of a quasiconvex cost function $\mathcal{J}(x)$ on the closure of $\mathcal{X}_c$. Let us assume that with some method an approximation $x^{(0)} \in \mathcal{X}_c$ of the optimal value x^* was obtained together with an upper bound on the scalar product $g^T(x^{(0)} - x^*)$ (which is usually delivered by standard convex analysis methods), where g is a vector defining a cutting plane to $\mathcal{J}$ at $x^{(0)}$.

In this section we assess the quality of the approximation $x^{(0)}$, i.e. we derive a bound on the error $\mathcal{J}(x^{(0)}) - \mathcal{J}(x^*)$. The results presented can be used for designing termination criteria for iterative optimization algorithms, guaranteeing a prespecified level of accuracy.

PROPOSITION 6.1. *Let $x^{(0)} \in \mathcal{X}_c$ be a feasible point and $\omega^{(0)}$ a frequency at which the worst-case chordal distance $\kappa_{WC}(G(e^{j\omega},\hat{\theta}),\mathcal{D}(x^{(0)}))$ attains its maximum. Suppose cost function $\mathcal{J}_1$ attains its minimum at x^*. Let vectors v and g be defined as in Proposition 5.2. If $v^T F_1 v > 0$, then the following bound on the error $\mathcal{J}_1(x^{(0)}) - \mathcal{J}_1(x^*)$ holds.*

$$\mathcal{J}_1^2(x^{(0)}) - \mathcal{J}_1^2(x^*) \leq N\frac{\mathcal{J}_1^2(x^{(0)})}{\chi^2_{n_a+n_b}(\alpha) v^T F_1 v}(1+|G(e^{j\omega^{(0)}},\hat{\theta})|^2)^2 |1 + Z_D\hat{\theta}|^2 g^T(x^{(0)} - x^*),$$

where Z_D is defined in (2.3).

Note that the condition $v^T F_1 v > 0$ is satisfied whenever $\mathcal{J}_1(x^{(0)}) < 1$. This inequality holds at least in the vicinity of x^* if $\mathcal{J}_1$ is not identically 1 on $\mathcal{X}_c$.

Proof. of Proposition 6.1. Denote by γ^*_{opt} the square of the worst-case chordal distance $\kappa_{WC}(G(e^{j\omega^{(0)}},\hat{\theta}),\mathcal{D}(x^*))$ at frequency $\omega^{(0)}$ and at the point x^*. Let τ^*_{opt} be the corresponding optimal value of τ. Then we have $\mathcal{J}_1^2(x^*) \geq \gamma^*_{opt}$.

By definition we have at frequency $\omega^{(0)}$ the relations $v^T(F_0 + \gamma^{(0)}_{opt} F_1 + \tau^{(0)}_{opt} R)v = 0$, $(\tau - \tau^{(0)}_{opt}) v^T R(x^{(0)}) v \leq 0$ for any $\tau \geq 0$ and $v^T(R(x) - R(x^{(0)}))v \leq -g^T(x - x^{(0)})$ for any x. Hence

$$v^T(F_0 + \gamma^*_{opt} F_1 + \tau^*_{opt} R(x^*))v \leq (\gamma^*_{opt} - \gamma^{(0)}_{opt}) v^T F_1 v - \tau^*_{opt} g^T(x^* - x^{(0)}).$$

Since the left-hand side of this inequality is nonnegative, we obtain

$$\mathcal{J}_1^2(x^{(0)}) - \mathcal{J}_1^2(x^*) \le \gamma_{opt}^{(0)} - \gamma_{opt}^* \le \frac{\tau_{opt}^*}{v^T F_1 v} g^T(x^{(0)} - x^*).$$

Let us now derive a bound on τ_{opt}^*. We have $(v')^T(F_0 + \gamma_{opt}^* F_1 + \tau_{opt}^* R)v' \ge 0$ for any vector $v' \in \mathbf{R}^{n_a+n_b+1}$. Choose $v' = (\hat\theta^T \ 1)^T$. By direct calculation one can show that $(v')^T F_0 v' = 0, (v')^T F_1 v' = (1+|G|^2)^2|1+Z_D\hat\theta|^2, (v')^T R v' = -\frac{\chi^2_{n_a+n_b}(\alpha)}{N}$. Thus we have

$$\tau_{opt}^* \le \frac{N}{\chi^2_{n_a+n_b}(\alpha)} \gamma_{opt}^* (1+|G|^2)^2 |1+Z_D\hat\theta|^2 \le \frac{N}{\chi^2_{n_a+n_b}(\alpha)} \gamma_{opt}^{(0)} (1+|G|^2)^2 |1+Z_D\hat\theta|^2.$$

Combining the obtained inequalities, we complete the proof. □

PROPOSITION 6.2. *Let $x^{(0)} \in \mathcal{X}_c$ be a feasible point. Let $\omega^{(0)}$ be a frequency at which the quantity $\frac{\lambda_{\max}(T(\omega)\bar M^{-1}(x^{(0)})T(\omega)^T)}{(1+|G(e^{j\omega},\hat\theta)|^2)^2}$ attains its maximum. Suppose cost function $\mathcal{J}_2$ attains its minimum at x^*. Let g be defined as in Proposition 5.3. Then the following bound on the error $\mathcal{J}_2(x^{(0)}) - \mathcal{J}_2(x^*)$ holds.*

$$\mathcal{J}_2^2(x^{(0)}) - \mathcal{J}_2^2(x^*) \le \frac{\chi^2_{n_a+n_b}(\alpha)}{N(1+|G(e^{j\omega^{(0)}},\hat\theta)|^2)^2} g^T(x^{(0)} - x^*).$$

Proof. Recall that we defined two functions $f(x), \tilde f(x)$ in the proof of Proposition 5.3 and identified g as the gradient of $\tilde f$. Since $\tilde f$ is convex, we can bound it by its first order Taylor polynomial, i.e. $\tilde f(x^{(0)}) - \tilde f(x^*) \le g^T(x^{(0)} - x^*)$. Therefore we have

$$\mathcal{J}_2^2(x^{(0)}) - \mathcal{J}_2^2(x^*) \le f^2(x^{(0)}) - f^2(x^*) = \frac{\chi^2_{n_a+n_b}(\alpha)}{N(1+|G(e^{j\omega^{(0)}},\hat\theta)|^2)^2}(\tilde f(x^{(0)}) - \tilde f(x^*)) \le$$
$$\le \frac{\chi^2_{n_a+n_b}(\alpha)}{N(1+|G(e^{j\omega^{(0)}},\hat\theta)|^2)^2} g^T(x^{(0)} - x^*). \quad \square$$

The propositions proven in this section enable the user to tell whether a given solution $x^{(0)}$, delivered e.g. by the current iteration step, satisfies the prespecified accuracy requirements. This information can be used e.g. to decide whether further iterations are necessary.

7. Design of input signals. Let us now turn to the question of how to design an input signal from $x^{(0)}$. By Theorem 3.6, there exist moment points which can be realized only by discrete spectra. On the other hand, any moment point can be realized by a discrete spectrum. Therefore we propose here the following 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. This procedure in no way restricts the optimality of the solution.

We weaken the condition $x^{(0)} \in \mathcal{X}_c$ and suppose that $x^{(0)}$ is in the closure of $\mathcal{X}_c$. The point $x^{(0)}$ corresponds to a point $\tilde x = (\tilde x_0, \tilde x_1, \dots, \tilde x_n)$ in moment space $\mathcal{M}^{(n+1)}$. Here $\tilde x_i$ equals the i-th component of $x^{(0)}$ for $i = 1, \dots, n$, and $\tilde x_0$ is given by (3.5).

Our goal will be to construct a realization of $\tilde x$. By Theorem 3.6, there exists a discrete realization with index not greater than $\frac{n+1}{2}$.

Denote by $\tilde x^s(\omega) \in \mathcal{M}^{(n+1)}$ the moment point induced by the design measure that satisfies constraint (3.4) and concentrates all power at the single frequency ω.

Then the i-th entry of $\tilde{x}^s(\omega)$ is given by $\frac{c}{\lambda_0|A(\omega)|^2}\cos(i\omega)$. Since (3.4) defines an affine section of the convex cone $\mathcal{M}^{(n+1)}$ and $\tilde{x}$ satisfies (3.4), $\tilde{x}$ is a convex combination of points on the curve $\{\tilde{x}^s(\omega)\,|\,\omega\in[0,\pi]\}$.

If $\tilde{x}$ equals $\tilde{x}^s(\pi)$, then we have already found a realization of index $\frac{1}{2}$. In this case this is the only possible realization.

Suppose now that $\tilde{x}$ is not equal to $\tilde{x}^s(\pi)$. To construct a realization of $\tilde{x}$, we will exploit an idea that is used to prove Theorem 3.6 (see e.g. [14]).

Consider the line going through the points $\tilde{x}$ and $\tilde{x}^s(\pi)$. This line has an interval in common with the convex set $\mathcal{M}^{(n+1)}$. This interval is finite, because it lies on the section defined by (3.4), and nondegenerated, because it contains two different points $\tilde{x}$ and $\tilde{x}^s(\pi)$. By Theorem 3.6 part i), $\tilde{x}^s(\pi)$ is one of the endpoints of this interval. Denote the other endpoint by $\tilde{x}^{bd}$. The computation of $\tilde{x}^{bd}$ from $\tilde{x}$ and $\tilde{x}^s(\pi)$ can be reduced to a standard GEVP using LMI description (3.3) of the set $\mathcal{M}^{(n+1)}$. For treatment of this type of problems see e.g. [5].

Thus $\tilde{x}$ is a linear convex combination of the points $\tilde{x}^s(\pi)$ and $\tilde{x}^{bd}$. Any realization of $\tilde{x}^{bd}$ will deliver us a realization of $\tilde{x}$. Note that $\tilde{x}^{bd}$ lies on the boundary of $\mathcal{M}^{(n+1)}$. By Theorem 3.6, part i) it has only one realization, which is of index less than $\frac{n+1}{2}$. Hence the realization of $\tilde{x}$ that we obtain with the described procedure, will have an index not exceeding $\frac{n+1}{2}$. If $\tilde{x}$ lies in the interior of $\mathcal{M}^{(n+1)}$, then this realization contains the frequency π and is therefore the upper principal realization of $\tilde{x}$.

We shall now construct the realization of $\tilde{x}^{bd}$. Denote the frequencies which are involved in this realization by ω_i, $i=1,\dots,k$. Then the point $\tilde{x}^{bd}$ is a nondegenerated linear convex combination of $\tilde{x}^s(\omega_1),\dots,\tilde{x}^s(\omega_k)$. We can write $\tilde{x}^{bd}=\sum_{i=1}^k\lambda_i\tilde{x}^s(\omega_i)$, where $\lambda_i>0$ and $\sum_{i=1}^k\lambda_i=1$.

Since $\tilde{x}^{bd}$ lies on the boundary of $\mathcal{M}^{(n+1)}$, there exists a supporting hyperplane E at $\tilde{x}^{bd}$. Note that E is a linear subspace, because $\mathcal{M}^{(n+1)}$ is a convex cone. The construction of a supporting hyperplane proceeding from LMI description (3.3) of $\mathcal{M}^{(n+1)}$ is a standard procedure and is described e.g. in [5].

Lemma 7.1. *The points $\tilde{x}^s(\omega_1),\dots,\tilde{x}^s(\omega_k)$ lie in E.*

Proof. Denote by n_E the normal vector to E that points toward $\mathcal{M}^{(n+1)}$ and by L_E the linear functional $x\mapsto\langle n_E,x\rangle$ defined by n_E. For any $\omega\in[0,\pi]$ we have $L_E(\tilde{x}^s(\omega))\geq 0$. On the other hand, $L_E(\tilde{x}^{bd})=\sum_{i=1}^k\lambda_iL_E(\tilde{x}^s(\omega_i))=0$, because $\tilde{x}^{bd}$ lies in E. Hence for all i we have $L_E(\tilde{x}^s(\omega_i))=0$, i.e. $\tilde{x}^s(\omega_i)\in E$. □

Lemma 7.2. *There exist maximally $\frac{n}{2}+1$ frequencies such that the corresponding points $\tilde{x}^s(\omega)$ lie in E.*

Proof. Consider $p_E:[0,\pi]\to\mathbf{R}$ defined by $p_E(\omega)=L_E(\tilde{x}^s(\omega))\frac{\lambda_0|A(\omega)|^2}{c}$. By definition, p_E is a trigonometric polynomial. Since $L_E(\tilde{x}^s(\omega))$ is nonnegative for all ω, p_E is too. Now we can apply a classical result from Tchebycheff system theory which states that p_E can have at most $\frac{n}{2}+1$ zeros. But the zeros of p_E lie exactly at those frequencies whose corresponding points $\tilde{x}^s(\omega)$ lie in E. □

Now we are able to obtain a finite set of frequencies that is guaranteed to contain $\omega_1,\dots,\omega_k$. Namely, we have to find the zeros, which are at the same time local minima, of the trigonometric polynomial $p_E(\omega)$.

Once we have found a set of frequencies $\omega_1,\dots,\omega_k,\omega_{k+1},\dots,\omega_{k+k'}$ such that the convex hull of the points $\tilde{x}^s(\omega_1),\dots,\tilde{x}^s(\omega_{k+k'})$ contains $\tilde{x}^{bd}$, it is a standard LQ programming problem to find the weights associated with the different frequencies. Namely, the weights λ_j minimize the squared distance $|\tilde{x}^{bd}-\sum_{j=1}^{k+k'}\lambda_j\tilde{x}^s(\omega_j)|^2$, which is a quadratic polynomial in the λ_j. Note that the number of frequencies involved is not greater than $\frac{n}{2}+1$ and hence not greater than n. Therefore the points $\tilde{x}^s(\omega_1),\ldots,\tilde{x}^s(\omega_{k+k'})$ are linearly independent and the minimized polynomial has a positive definite quadratic part. An efficient algorithm for solving this type of problems is e.g. the Beale algorithm (see [1]).

Suppose that a discrete realization of $\tilde{x}$ with frequencies $\omega_1,\ldots,\omega_m$ and associated weights $\lambda_1,\ldots,\lambda_m$ is available. Then 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 $\omega_i\in\{0,\pi\}$, has the desired input power spectrum (see e.g. [30]).

Often it is also possible to obtain the desired power spectrum by using binary signals (see [30, p.29] and references cited therein).

8. Simulation results. Consider the true system $y=G_0u+H_0e=\frac{B(z)}{A(z)}u+\frac{1}{A(z)}e$ 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 data points to be collected is $N=1000$. The aim is to minimize the worst-case ν-gap of the uncertainty region around the identified model corresponding to a confidence level of $\alpha=0.95$.

In a Monte-Carlo simulation, 500 runs were performed. Each run consisted of five identification experiments: one preliminary experiment and four mutually independent second experiments based on this preliminary experiment, corresponding to the four different cost functions $\mathcal{J}_1$, $\mathcal{J}_2$, D-optimality and E-optimality.

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. In two of the second experiments, the input power spectrum minimized the cost functions $\mathcal{J}_1,\mathcal{J}_2$, respectively. The actual input sequence was a multisine having the evaluated optimal power spectrum. For comparison, two other second experiments with D-optimal and E-optimal input power spectra were performed. After each identification experiment the worst-case ν-gap of the identified uncertainty region was recorded.

The noise realizations for the five experiments within one run and for different runs were different, as well as the input realizations for the preliminary experiments of the different runs.

In figure 8.1 the worst-case ν-gap obtained from the preliminary experiment with white noise input, as well as from the experiments with inputs optimized with respect to $\mathcal{J}_1$ and $\mathcal{J}_2$ respectively, are shown for the first 50 simulation runs. 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 criteria $\mathcal{J}_1,\mathcal{J}_2$ 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 $\mathcal{J}_1,\mathcal{J}_2$ 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 $\mathcal{J}_2$, which is a first order approximation of the exact cost function $\mathcal{J}_1$, give better results than $\mathcal{J}_1$, despite the fact that the plotted quantity is in fact $\mathcal{J}_1$. As

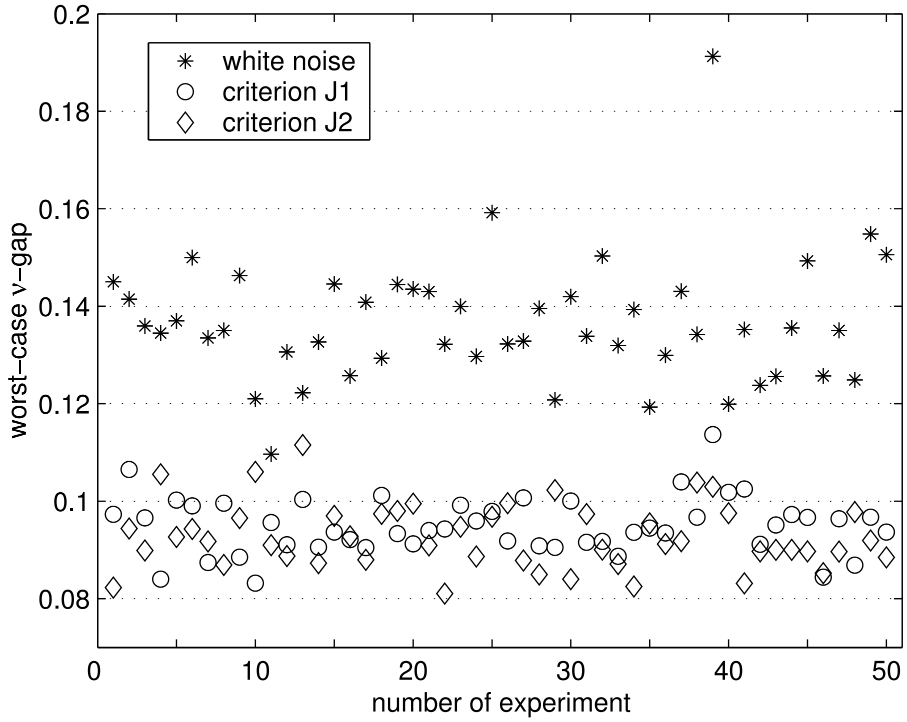

FIG. 8.1. *Identification with white and subsequently estimated optimal input*

mentioned already in section 2, this tendency was observed also in simulations with other systems. The reason is that the optimum of the input power spectrum with respect to $\mathcal{J}_2$ is less dependent on the preliminary estimate $\bar{\theta}$ of the true parameter vector than the optimum with respect to $\mathcal{J}_1$. Given the lower complexity of $\mathcal{J}_2$ and hence the lower computational effort in comparison with $\mathcal{J}_1$, it is recommendable to use primarily the former.

9. Conclusions. Let us summarize the results obtained in the present paper. We have to design an input sequence for an identification experiment that makes the worst-case ν-gap between the identified model and the uncertainty region around it as small as possible. The design takes place via power spectrum optimization. Two nonstandard cost criteria $\mathcal{J}_1$ and $\mathcal{J}_2$ are defined, which reflect the optimization task with different accuracy. $\mathcal{J}_1$ is the exact worst-case ν-gap one would want to minimize, while $\mathcal{J}_2$ is an approximation of $\mathcal{J}_1$. 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 reduced to a standard convex optimization problem involving LMI constraints. In Propositions 5.2 and 5.3 we demonstrate how to construct cutting planes to the cost functions $\mathcal{J}_1, \mathcal{J}_2$, which is essential for applying standard numerical methods such as the ellipsoid algorithm. In Propositions 6.1 and 6.2 we derive bounds on the difference between the actually achieved and the optimal value of the cost functions, which allows to estimate the quality of the optimization result and to design stopping criteria for iterative search algorithms. We have also briefly touched the problem of designing an input sequence with a prespecified power spectrum.

Simulations show clearly the superiority of the proposed cost functions over classical design criteria. They also suggest to use cost function $\mathcal{J}_2$ rather than $\mathcal{J}_1$, due to both lower computational effort and higher performance.

Acknowledgements. The authors would like to thank Yurii Nesterov for the very useful help he provided during the work on this paper.

REFERENCES

[1] J. Abadie, ed., *Nonlinear programming*, North Holland Co., Amsterdam, 1967, ch. 7.

[2] X. Bombois, B. Anderson, and M. Gevers, *Frequency domain image of a set of linearly parametrized transfer functions*, in Proceedings of the 6th ECC, 2001.

[3] X. Bombois, M. Gevers, and G. Scorletti, *A measure of robust stability for a set of parametrized transfer functions*, IEEE Trans. Automat. Control, AC-45 (2000).

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

[5] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*, vol. 15 of SIAM Stud. Appl. Math., SIAM, 1994.

[6] B. L. Cooley, J. H. Lee, and S. P. Boyd, *Control-relevant experiment design: A plant-friendly, LMI-based approach*, in Proceedings of the ACC, Philadelphia, 1998.

[7] P. Date and G. Vinnicombe, *An algorithm for identification in the ν-gap metric*, in Proceedings of the 38th Conference of Decision and Control, Phoenix, Arizona, December 1999, pp. 3230–3235.

[8] K. Fan, *On a theorem of Weyl concerning eigenvalues of linear transformations*, Proc. Natl. Acad. Sci. USA, 35 (1959).

[9] V. Fedorov, *Theory of Optimal Experiments*, vol. 12 of Probab. Math. Statist., Academic Press, New York – London, 1972.

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

[11] G. Goodwin and R. Payne, *Design and characterization of optimal test signals for linear SISO parameter estimation*, in Preprints of 3rd IFAC Symposium, The Hague, 1973. Paper TT-1.

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

[13] S. Karlin and L. Shapley, *Geometry of Moment Spaces*, vol. 12 of Mem. Amer. Math. Soc., Amer. Math. Soc., 1953.

[14] S. Karlin and W. Studden, *Tchebycheff systems: with applications in analysis and statistics*, vol. XV of Pure Appl. Math., Interscience Publishers, 1966.

[15] J. Kiefer, *General equivalence theory for optimum designs (approximate theory)*, Ann. Statist., 2 (1974), pp. 849–879.

[16] J. Kiefer and J. Wolfowitz, *Optimum designs in regression problems*, The Annals of Mathematical Statistics, 30 (1959), pp. 271–294.

[17] M. Krein, *The ideas of P.L. Čebyšev and A.A. Markov in the theory of limiting values of integrals and their further developments*, Amer. Math. Soc. Transl. Ser. 2, 12 (1951), pp. 1–122.

[18] K. Lindqvist and H.Hjalmarsson, *Identification for control: Adaptive input design using convex optimization*, in Proceedings of the 40th Conference on Decision and Control, Orlando, Florida, USA, 12 2001.

[19] L. Ljung, *System identification: theory for the user*, Prentice-Hall Information and System Sciences Series, Prentice Hall, second ed., 1999.

[20] R. Mehra, *Optimal input signals for parameter estimation in dynamic systems — survey and new results*, IEEE Trans. Automat. Control, AC-19 (1974), pp. 753–768.

[21] ———, *Optimal inputs for linear system identification*, IEEE Trans. Automat. Control, AC-19 (1974), pp. 192–200.

[22] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*, vol. 13 of SIAM Stud. Appl. Math., SIAM, Philadelphia, 1994.

[23] R. Payne, *Optimal Experiment Design for Dynamic System Identification*, PhD thesis, Imperial College, London, 1974.

[24] R. Payne and G. Goodwin, *Simplification of frequency domain experiment design for SISO systems*, Publication 74/3, Dept. of Computing and Control, Imperial College, London, 1974.
[25] J. Schoukens, P. Guillaume, and R. Pintelon, *Perturbation signals for system identification*, Prentice Hall International Series in Acoustics, Speech and Signal Processing, Prentice Hall, 1993, ch. 3, pp. 126–160.
[26] A. van den Bos, *Selection of periodic test signals for estimation of linear system dynamics*, in Preprints of 3rd IFAC Symposium, The Hague, 1973. Paper TT-3.
[27] E. van den Eijnde and J. Schoukens, *On the design of optimal excitation signals*, in 9th IFAC/IFORS Symposium on Identification and System Parameter Estimation, Budapest, Hungary, 1991, pp. 827–832.
[28] G. Vinnicombe, *Frequency domain uncertainty and the graph topology*, IEEE Trans. Automat. Control, AC-38 (1993), pp. 1371–1383.
[29] S. Wu, S. P. Boyd, and L. Vandenberghe, *FIR filter design via semidefinite programming and spectral factorization*, in Proceedings of the 35th CDC, Kobe, Japan, 1996.
[30] M. Zarrop, *Optimal experiment design for dynamic system identification*, vol. 21 of Lecture Notes in Control and Inform. Sci., Springer, Berlin, New York, 1979.

Appendix A. Proof of Proposition 5.2.

Lemma A.1. *Let γ_{opt}, τ_{opt} be the optimal values of γ, τ in GEVP (4.1),(4.2). Then the following conditions hold.*

i) The matrix F_0 is negative semidefinite.
ii) The nullspace of F_1 is a subset of the nullspace of F_0.
iii) The matrix $F_0 + F_1$ is positive semidefinite.
iv) The nullspace of F_1 is a strict subset of the nullspace of $F_0 + F_1$.
v) $\tau_{opt} > 0$ if and only if the restriction of R on the nullspace of $F_0 + F_1$ is strictly positive definite.
vi) $\gamma_{opt} = 1$ if and only if $\tau_{opt} = 0$.

Proof. i) follows from the representation $F_0 = -VWW^TV^T$, where W is a 4×2-matrix given by

$$W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ ImG(\hat{\theta}) & -ReG(\hat{\theta}) \\ -ReG(\hat{\theta}) & -ImG(\hat{\theta}) \end{pmatrix}.$$

The nullspace of F_1 is given by the kernel of V^T. The latter is contained in the kernel of F_0, which yields ii).

iii) follows from the representation

$$F_0 + F_1 = V \begin{pmatrix} |G(\hat{\theta})|^2 & 0 & -ImG(\hat{\theta}) & ReG(\hat{\theta}) \\ 0 & |G(\hat{\theta})|^2 & ReG(\hat{\theta}) & ImG(\hat{\theta}) \\ -ImG(\hat{\theta}) & ReG(\hat{\theta}) & 1 & 0 \\ ReG(\hat{\theta}) & ImG(\hat{\theta}) & 0 & 1 \end{pmatrix} V^T = VW_\perp W_\perp^T V^T,$$

where $W_\perp$ is a 4×2-matrix given by

$$W_\perp = \begin{pmatrix} |G(\hat{\theta})| & 0 \\ 0 & |G(\hat{\theta})| \\ -\sin argG(\hat{\theta}) & \cos argG(\hat{\theta}) \\ \cos argG(\hat{\theta}) & \sin argG(\hat{\theta}) \end{pmatrix}.$$

Here $argG(\hat{\theta})$ is an arbitrary number if $G(\hat{\theta}) = 0$. Note that $W^TW = (1 + |G(\hat{\theta})|^2)I_2$, $W^TW_\perp = 0$ and $W_\perp^TW_\perp = (1 + |G(\hat{\theta})|^2)I_2$.

By ii) the nullspace of F_1 is a subset of the nullspace of $F_0 + F_1$. We shall now show that the vector $v = (\zeta, 0, \dots, 0, 1, 0, \dots, 0, -\cos n_k\omega\, ReG(\hat\theta) + \sin n_k\omega\, ReG(\hat\theta) - \zeta\cos\omega)^T \in \mathbf{R}^{n_a+n_b+1}$, where $\zeta = -\cot n_k\omega\, ImG(\hat\theta) - ReG(\hat\theta)$ if $\omega \in (0,\pi)$ and ζ arbitrary otherwise, is not contained in the kernel of V^T but is contained in the kernel of $W_\perp^T V^T$. The "1" in v is situated at position $n_a + 1$. Indeed, by (2.3) we obtain

$$V^T v = \begin{pmatrix} \cos n_k\omega \\ -\sin n_k\omega \\ -\zeta\sin\omega \\ -\cos n_k\omega\, ReG(\hat\theta) + \sin n_k\omega\, ImG(\hat\theta) \end{pmatrix} =$$

$$= \cos n_k\omega \begin{pmatrix} 1 \\ 0 \\ ImG(\hat\theta) \\ -ReG(\hat\theta) \end{pmatrix} - \sin n_k\omega \begin{pmatrix} 0 \\ 1 \\ -ReG(\hat\theta) \\ ImG(\hat\theta) \end{pmatrix} \neq 0,$$

because $\cos n_k\omega$, $-\sin n_k\omega$ cannot both vanish. In case $\omega \in \{0,\pi\}$ the equality holds by $ImG(\hat\theta) = 0$. On the other hand, we have $W_\perp^T V^T v = W_\perp^T W(\cos n_k\omega, -\sin n_k\omega)^T = 0$. This concludes the proof of iv).

Let us prove v) and vi). Denote the nullspace of $F_0 + F_1$ by $\bar V^0$ and its orthogonal complement by $\bar V^\perp$. By definition there exists a positive number β_1 such that for any $v^\perp \in \bar V^\perp$ we have $(v^\perp)^T(F_0 + F_1)v^\perp \geq \beta_1|v^\perp|^2$.

Suppose the restriction of R on $\bar V^0$ is strictly positive definite. Then there exists a positive number β_2 such that for any $v^0 \in \bar V^0$ we have $(v^0)^T R v^0 \geq \beta_2|v^0|^2$. Let $v = v^0 + v^\perp$ be an arbitrary vector with v^0, $v^\perp$ being its orthogonal projections on $\bar V^0$, $\bar V^\perp$ respectively. Let $\tau > 0$ be a positive number. Then we have

$$v^T(F_0 + F_1 + \tau R)v = (v^\perp)^T(F_0 + F_1)v^\perp + \tau((v^\perp)^T R v^\perp + 2(v^0)^T R v^\perp + (v^0)^T R v^0)$$

$$\geq \beta_1|v^\perp|^2 + \tau(\lambda_{\min}(R)|v^\perp|^2 - 2\min\{\lambda_{\min}(R), -\lambda_{\max}(R))|v^\perp||v^0| + \beta_2|v^0|^2\} =$$

$$\begin{pmatrix} |v^\perp| \\ |v^0| \end{pmatrix}^T \begin{pmatrix} \beta_1 + \tau\lambda_{\min}(R) & -\tau\min\{\lambda_{\min}(R), -\lambda_{\max}(R)\} \\ -\tau\min\{\lambda_{\min}(R), -\lambda_{\max}(R)\} & \tau\beta_2 \end{pmatrix} \begin{pmatrix} |v^\perp| \\ |v^0| \end{pmatrix}.$$

It is easily seen that the 2×2-matrix in the middle is positive definite if τ is small enough. Therefore there exists $\tau > 0$ such that the matrix $F_0 + F_1 + \tau R$ is strictly positive definite, while the matrix $F_0 + F_1$ is not. Thus in this case we have $\tau_{opt} \neq 0$ and $\gamma_{opt} < 1$.

Now suppose the restriction of R on $\bar V^0$ is not strictly positive definite. Since $\bar M$ is strictly positive definite, it follows from expression (4.2) that R has $n_a + n_b$ positive eigenvalues and one negative eigenvalue. Thus it can be represented as a difference $R = R_+ - R_-$, where R_+, R_- are positive semidefinite matrices of rank $n_a + n_b, 1$ respectively and the linear hulls $\bar V_+, \bar V_-$ of their columns are orthogonal to each other. The whole space $\mathbf{R}^{n_a+n_b+1}$ splits into a direct sum $\bar V_+ \oplus \bar V_-$.

Let $v^0 \in \bar V^0$ be a nonzero vector such that $(v^0)^T R v^0 \leq 0$. The vector v^0 can be represented as a sum $v^0 = v_+ + v_-$, where $v_+ \in \bar V_+$, $v_- \in \bar V_-$. Since $(v^0)^T R v^0 = (v_+)^T R_+ v_+ - (v_-)^T R_- v_- \leq 0$, the assumption $v_- = 0$ would imply $v_+ = 0$, which contradicts $v^0 \neq 0$. Hence $v_- \neq 0$. We can represent v_- as a sum $v_- = v_-^0 + v_-^\perp$, where $v_-^0 \in \bar V^0, v_-^\perp \in \bar V^\perp$. Let $\varepsilon > 0$ be a positive number and consider the vector $v = v^0 +$

εv_-. We have $v = v_+ + (1+\varepsilon)v_-$. Hence $v^T R v = (v_+)^T R_+ v_+ - (1+\varepsilon)^2 (v_-)^T R_- v_- = (v^0)^T R v^0 - (2\varepsilon + \varepsilon^2)(v_-)^T R_- v_-$. On the other hand, $v = (v^0 + \varepsilon v_-^0) + \varepsilon v_-^\perp$. Since $v^0 + \varepsilon v_-^0 \in \bar{V}^0$, this yields $v^T(F_0 + F_1)v = \varepsilon^2 (v_-^\perp)^T (F_0 + F_1) v_-^\perp$. We obtain

$$v^T(F_0 + F_1 + \tau R)v \leq \varepsilon^2 (v_-^\perp)^T (F_0 + F_1) v_-^\perp - \tau(2\varepsilon + \varepsilon^2)(v_-)^T R_- v_-.$$

Note that $(v_-)^T R_- v_-$ is strictly positive. Hence for any prespecified $\tau > 0$ we can choose a small $\varepsilon > 0$ such that $v^T(F_0 + F_1 + \tau R)v < 0$. Thus for any positive τ the matrix $F_0 + F_1 + \tau R$ is not positive semidefinite, while $F_0 + F_1$ is. This implies $\tau_{opt} = 0$. $\gamma_{opt} = 1$ now follows from iv).

The proof of the lemma is complete. □

Proof. of Proposition 5.2. Denote the orthogonal complement of V^0 by $V^\perp$. Then the restriction on $V^\perp$ of the quadratic form defined by the matrix $F_0 + \gamma_{opt}^{(0)} F_1 + \tau_{opt}^{(0)} R$ is strictly positive definite. Hence there exists a positive number β_1 such that for any vector $v^\perp \in V^\perp$ we have $(v^\perp)^T(F_0 + \gamma_{opt}^{(0)} F_1 + \tau_{opt}^{(0)} R)v^\perp \geq \beta_1 |v^\perp|^2$.

Suppose the restriction on V^0 of the quadratic form defined by the matrix R is strictly positive definite. Then there exists a positive number β_2 such that for any vector $v^0 \in V^0$ we have $(v^0)^T R v^0 \geq \beta_2 |v^0|^2$.

Let $v = v^0 + v^\perp$ be an arbitrary vector, where $v^0 \in V^0$ and $v^\perp \in V^\perp$ are its orthogonal projections on the subspaces V^0 and $V^\perp$ respectively. Let $\varepsilon > 0$ be a positive number. Then we have

$$\begin{aligned}
&v^T(F_0 + \gamma_{opt}^{(0)} F_1 + (\tau_{opt}^{(0)} + \varepsilon)R)v = \\
&= (v^\perp)^T(F_0 + \gamma_{opt}^{(0)} F_1 + \tau_{opt}^{(0)} R)v^\perp + \varepsilon((v^\perp)^T R v^\perp + 2(v^0)^T R v^\perp + (v^0)^T R v^0) \\
&\geq \beta_1 |v^\perp|^2 + \varepsilon(\lambda_{\min}(R)|v^\perp|^2 + 2\min\{\lambda_{\min}(R), -\lambda_{\max}(R))|v^\perp||v^0| + \beta_2 |v^0|^2\} =
\end{aligned}$$

$$\begin{pmatrix} |v^\perp| \\ |v^0| \end{pmatrix}^T \begin{pmatrix} \beta_1 + \varepsilon\lambda_{\min}(R) & \varepsilon\min\{\lambda_{\min}(R), -\lambda_{\max}(R)\} \\ \varepsilon\min\{\lambda_{\min}(R), -\lambda_{\max}(R)\} & \varepsilon\beta_2 \end{pmatrix} \begin{pmatrix} |v^\perp| \\ |v^0| \end{pmatrix}.$$

It is easily seen that the 2×2-matrix in the middle is positive definite if ε is small enough. This implies that there exists a number $\tau > \tau_{opt}^{(0)}$ such that the matrix $F_0 + \gamma_{opt}^{(0)} F_1 + \tau R$ is strictly positive definite. This contradicts the optimality of $\gamma_{opt}^{(0)}$.

In a similar way it is shown that if the restriction on V^0 of the quadratic form R is strictly negative definite, then there exists a number $\varepsilon > 0$ such that for any $\tau \in [\tau_{opt}^{(0)} - \varepsilon, \tau_{opt}^{(0)})$ the matrix $F_0 + \gamma_{opt}^{(0)} F_1 + \tau R$ is strictly positive definite.

Thus the restriction on V^0 of the quadratic form R is neither strictly positive nor strictly negative definite if $\tau_{opt}^{(0)} > 0$ and it is negative semidefinite if $\tau_{opt}^{(0)} = 0$. This proves the first part of the proposition.

Now let $v \in V^0$ be a unit length vector satisfying the conditions of Proposition 5.2. Let $g \in \mathbf{R}^n$ be given componentwise by $g_i = -v^T R_i v$. Let $x \in \mathcal{X}_c$ be a vector satisfying the inequality $g^T(x - x^{(0)}) \geq 0$. Let τ be a nonnegative number and let γ be strictly less than $\gamma_{opt}^{(0)}$.

By assumption we have $v^T(F_0 + \gamma_{opt}^{(0)} F_1 + \tau_{opt}^{(0)} R(x^{(0)}))v = 0$. We obtain

$$v^T(F_0 + \gamma F_1 + \tau R(x))v = v^T((\gamma - \gamma_{opt}^{(0)})F_1 + \tau(R(x) - R(x^{(0)})) + (\tau - \tau_{opt}^{(0)})R(x^{(0)}))v$$

$$= (\gamma - \gamma_{opt}^{(0)})v^T F_1 v + (-\tau g^T(x - x^{(0)})) + (\tau - \tau_{opt}^{(0)})v^T R(x^{(0)})v \leq 0.$$

The last inequality follows from the fact that none of the three terms on the left-hand side exceeds zero. The first term is nonpositive because F_1 is positive semidefinite. The second term does not exceed zero by assumption on x. The third term is not greater than zero because by assumption on v we have $v^T R(x^{(0)})v \leq 0$ and the condition $\tau - \tau_{opt}^{(0)} < 0$ yields $\tau_{opt}^{(0)} > 0$ and hence $v^T R(x^{(0)})v = 0$. If the inequality is strict, then $F_0 + \gamma F_1 + \tau R(x)$ is not positive semidefinite.

Now assume that $v^T(F_0 + \gamma F_1 + \tau R(x))v = 0$. Then we have $v^T F_1 v = 0$ and v is an element of the nullspace of F_1. By Lemma A.1, part iv), it is also an element of the nullspace of $F_0 + F_1$. Note that $v^T R(x^{(0)})v \leq 0$. By Lemma A.1, part v) we then have $\tau_{opt}^{(0)} = 0$ and by part vi) $\gamma_{opt}^{(0)} = 1$. Further we have either $\tau = \tau_{opt}^{(0)}$ or $v^T R(x^{(0)})v = 0$.

If $\tau = \tau_{opt}^{(0)} = 0$, then by Lemma A.1, parts iii) and iv), the matrix $F_0 + \gamma F_1 = F_0 + \gamma F_1 + \tau R(x)$ is not positive semidefinite.

If $\tau > 0$, then $v^T R(x^{(0)})v = 0$ and $v^T R(x)v = v^T(R(x) - R(x^{(0)}))v = -g^T(x - x^{(0)}) \leq 0$. Since v belongs to the nullspace of $F_0 + F_1$, by Lemma A.1, part v) we have $\tau_{opt}(x) = 0$ and by part vi) $\gamma_{opt}(x) = 1 > \gamma$. Hence the pair (γ, τ) is again not feasible for GEVP (4.1),(4.2) at x.

Thus in any case $\gamma_{opt}(x)$ is not less than $\gamma_{opt}^{(0)}$ and the vector g, if nonzero, defines a cutting plane for γ_{opt} and hence for $\mathcal{J}_1$.

If $g = 0$, however, then any x satisfies the relation $g^T(x - x^{(0)}) \geq 0$ and $\gamma_{opt}^{(0)}$ does not exceed γ_{opt} at any other point $x \in \mathcal{X}_c$. Hence we have $\mathcal{J}_1(x) \geq \kappa_{WC}(G(e^{j\omega^{(0)}}, \hat{\theta}), \mathcal{D}) = \sqrt{\gamma_{opt}(x)} \geq \sqrt{\gamma_{opt}^{(0)}} = \mathcal{J}_1(x^{(0)})$ and $\mathcal{J}_1$ attains a minimum at $x^{(0)}$.

This concludes the proof of the second part of Proposition 5.2. □