Iterative Robust Experiment Design for MIMO System Identification via the S-Lemma

Dirkx, Nic; Tiels, Koen; Oomen, Tom

DOI
10.1109/CCTA54093.2023.10252362

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the 2023 IEEE Conference on Control Technology and Applications, CCTA 2023

2023 IEEE Conference on Control Technology and Applications (CCTA)
August 16-18, 2023. Bridgetown, Barbados

Iterative Robust Experiment Design for MIMO System Identification via the S-Lemma

Nic Dirkx[1], Koen Tiels[2], and Tom Oomen[3]

2023 IEEE Conference on Control Technology and Applications (CCTA) | 979-8-3503-3544-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/CCTA54093.2023.10252362

***Abstract*— Optimal input design plays an important role in system identification for complex and multivariable systems. A known paradox in input design is that the optimal inputs depend on the true but unknown system. The aim of this paper is to design inputs for multivariable systems that are robust to all system variations within a given continuous uncertainty set. In the presented approach, the robust design problem is cast as an infinite-dimensional min-max optimization problem, and tackled via the S-lemma in an iterative approximation scheme. Experimental results from a multivariable motion system show that the algorithm enables significant robustness improvements.**

I. INTRODUCTION

Good design of system identification experiments is an essential step towards high-quality identified models. This applies especially to complex multiple inputs multiple outputs (MIMO) systems such as wafer stages [1]. Optimal experiment design consists in a systematic approach for the design of inputs that maximize model accuracy within limited experimental resources. The topic of optimal experiment design has seen much research over the past decades, including [2], [3], and more recent surveys [4] and [5].

Paradoxically, an inherent issue in experiment design for system identification is that the optimal inputs depend on the true system itself, which is unknown prior to the identification experiment. In the majority of literature, this issue is circumvented by using a (point) estimate of the true system [6], e.g., in [7], [8]. In the presence of uncertainty in the estimated model, such nominal design approaches lack performance guarantees in view of the true system.

The aim in robust experiment design is to explicitly address the uncertainties in the prior model. The most widely considered approach is to design inputs that are worst-case optimal with respect to a model set that includes the true system, leading to a min-max design [9], [10], [11]. Typically, the model set is described as a compact continuous set, by which the design problem becomes a semi-infinite program (SIP), i.e., an optimization problem over a finite number of decision variables but with an infinite-dimensional constraint set. Such problems are generally difficult to solve and potentially NP-hard [12]. This makes robust experiment design a complicated task in general.

To handle the complexity in robust experiment design, several techniques have been developed that involve discretization of the continuous uncertainty set. In [11], the uncertainty set is gridded after which the problem is solved via standard convex optimization. In [13], a scenario approach is presented wherein the continuous uncertainty set is approximated by a finite number of constraints that are randomly selected. A drawback of these approaches is that the required size of the finite subset is typically large [13], which comes with a substantial computational burden [14]. In [15], an exchange algorithm is presented where in each iteration only the active constraints are maintained. This reduces the computational load, yet at the expense of accuracy.

Departing from discretization approaches, in [16] a separation of graph framework is used to substitute the infinite-dimensional constraint by a different one that implies the original constraint. While this enables providing hard robustness guarantees, the substitution comes with conservatism. Further, the approach is limited to a specific design criterion and single input single output (SISO) systems. In [10], the infinite-dimensional constraints are preserved by exploiting sum of squares formulations. The method involves introducing high-order auxiliary polynomials in the optimization program, which selection is non-straightforward, and gives rise to numerical complexities.

Although important developments have been made in the design of robust experiments, tractable design approaches for complex MIMO systems are scarce. The aim of this paper is to present a robust experiment design approach for complex MIMO systems, that does not require the inclusion of a large number of constraints or auxiliary parameters.

The contributions of this paper are:

1. An iterative robust experiment design algorithm for system identification of multivariable systems, based on the S-lemma,
2. An experimental validation on a MIMO motion system.

Throughout, proofs are omitted to conserve space.

Notations Operation $\Re(X)$ denotes the real part of X, and $\overline{X} = \text{conj}(X)$. Operator $\otimes$ denotes the Kronecker product. $\text{Tr}(X)$ is the trace. For $X = \begin{bmatrix} X_1, \ldots, X_n \end{bmatrix}$ with $X_i \in \mathbb{C}^m, i = 1, \ldots, n$, operation $\text{vec}(X) = \begin{bmatrix} X_1^T, \ldots, X_n^T \end{bmatrix}^T$. The operation $\text{diag}\left(\begin{bmatrix} X_1, \ldots, X_n \end{bmatrix}\right)$ results in a block-diagonal matrix with $X_i,\ i = 1, \ldots, n$ on its block-diagonal.

*This work was supported by the Research Programme VIDI under Project 15698, partly financed by the NWO.

[1]Nic Dirkx is with ASML, Veldhoven, The Netherlands. He is also with the Dept. of Mechanical Engineering, Control Systems Technology, Eindhoven University of Technology, Eindhoven, The Netherlands. nic.dirkx@asml.com

[2]Koen Tiels is with the Control Systems Technology Section, Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.

[3]Tom Oomen is with the Control Systems Technology Section, Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. He is also with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands.

II. PROBLEM FORMULATION

A. Complex mechatronic systems

Complex MIMO mechatronic systems are considered, such as the next-generation wafer stage system in Fig. 1. To perform fast motion tasks, the stage is designed to be lightweight. As a consequence, it exhibits complex dynamic behavior, e.g., due to mechanical flexibilities. To achieve extreme positioning performance, the stage is equipped with 8 actuators and 7 sensors, enabling explicit control of internal flexibilities [17]. Accurate system identification demands experiment design techniques that can handle such complex MIMO systems. In this paper, a robust experiment design approach for such systems is presented.

B. Identification framework

Consider the MIMO true system with inputs $u \in \mathbb{R}^{n_u}$ and outputs $y \in \mathbb{R}^{n_y}$,

$$\mathcal{S}: \quad y(t) = G_o(q)u(t) + H_o(q)e(t) \tag{1}$$

where $G_o(q)$ is a $n_y \times n_u$ stable transfer function matrix (TFM) and q is the forward shift operator $qu(t) = u(t+1)$. Furthermore, $e \in \mathbb{R}^{n_y}$ represents measurement noise, characterized as an independent and identically distributed (iid) standard normal random sequence with covariance Λ, and $H_o(q)$ is a $n_y \times n_y$ stable inversely monic TFM representing the noise dynamics.

The true system is identified within the model structure

$$\mathcal{M}: \quad y(t) = G(q,\theta)u(t) + H(q,\theta)e(t), \tag{2}$$

with $\theta = \begin{bmatrix}\rho^T & \mu^T\end{bmatrix}^T \in \mathbb{R}^{n_\theta}$, where $\rho \in \mathbb{R}^{n_\rho}$ and $\mu \in \mathbb{R}^{n_\mu}$ represent the parameters in G and H, respectively. It is assumed that the true system is within the model set, i.e., there exists a parameter θ_o such that $G(q,\theta_o) = G_o(q)$ and $H(q,\theta_o) = H_o(q)$.

The right matrix fraction description (RMFD) parametrization for $G(q,\rho)$ and $H(q,\mu)$ is considered,

$$\begin{aligned} G(q,\rho) &= N(q,\rho)D^{-1}(q,\rho), \\ H(q,\mu) &= N_H(q,\rho)D_H^{-1}(q,\mu). \end{aligned} \tag{3}$$

where $N(q,\rho) \in \mathbb{R}^{n_y \times n_u}, D(q,\rho) \in \mathbb{R}^{n_u \times n_u}, N_H(q,\rho) \in \mathbb{R}^{n_y \times n_y}, D_H(q,\rho) \in \mathbb{R}^{n_y \times n_y}$ are real polynomial matrices, affinely parametrized in ρ and μ, respectively. Furthermore, $D(q,\rho), N_H(q,\rho)$, and $D_H(q,\rho)$ are monic.

The covariance error in the estimates $\hat{\theta}$ of any parametric estimator is lower bounded by the Cramér-Rao lower bound, i.e., cov $\hat{\theta} \succeq M_{\theta,o}^{-1}$, where $M_{\theta,o}$ is the Fisher information matrix [3]. For the independent parametrization in (3), the information matrix is of the block-diagonal form $M_{\theta,o} = \text{diag}\big([M_{\rho,o},\ M_{\mu,o}]\big)$ [18]. Often, the main interest is in the part $M_{\rho,o}$, i.e., the part related to the plant model $G(q,\rho)$. For the multivariable system (1), the matrix $M_{\rho,o}$ is expressed as $M_{\rho,o} = M(\theta_o, \Phi_u)$, where [6]

$$M(\theta_o,\Phi_u) = \frac{N}{2\pi}\Re\int_{-\pi}^{\pi} \mathcal{F}^H(q,\theta_o)\left(\overline{\Phi_u} \otimes \Lambda^{-1}\right)\mathcal{F}(q,\theta_o)d\omega, \tag{4}$$

with N the sample size, Φ_u the input spectrum, and

$$\begin{aligned} \mathcal{F}(q,\theta) &= \left(I_{n_u} \otimes H^{-1}(q,\theta)\right)\Lambda_G(q,\theta), \\ \Lambda_G(q,\theta) &= \frac{\partial \text{vec}\left(G(q,\rho)\right)}{\partial \rho^T}. \end{aligned} \tag{5}$$

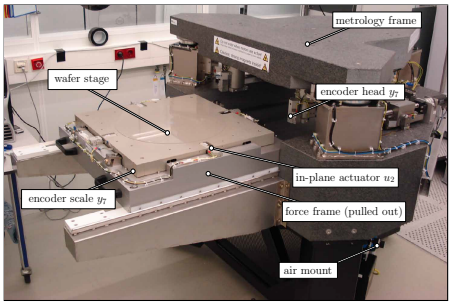

Fig. 1. Next-generation wafer stage setup.

Expression (4) shows that the lower bound on the achievable covariance in the parameter estimates is determined by the input spectrum Φ_u. Also, it appears from (4) that this bound depends on the true but unknown parameter θ_o. In this paper, optimal design of the input spectrum Φ_u is considered that is robust to the uncertainty in θ_o.

C. Robust experiment design problem

Optimal experiment design consists in computing an input spectrum Φ_u that achieves optimal model quality, in some sense, while respecting possible limitations in experimental resources, e.g., bounded inputs or outputs.

Two choices of quality measures are considered. The first measure is the parameter covariance matrix defined as $C_\theta(\theta_o,\Phi_u) = M^{-1}(\theta_o,\Phi_u)$. The second measure is the frequency-wise covariance of the identified model $C_G(q,\theta_o,\Phi_u) = \text{cov}\Big(\text{vec}(G(q,\theta_o))\Big)$, which is estimated using a first-order Taylor approximation as [7],

$$C_G(q,\theta_o,\Phi_u) = \Lambda_G(q,\theta_o)C_\theta(\theta_o,\Phi_u)\Lambda_G^H(q,\theta_o), \tag{6}$$

with Λ_G in (5). A typical property of input spectrum design is that the optimal spectrum depends on the true but unknown system parameter θ_o, as reflected by (6). In this paper, it is assumed that the true parameter θ_o is unknown, but is known to lie with α-probability within the ellipsoid

$$\Theta_{\text{init}} = \Big\{\theta : \varepsilon^T M_{\text{init}}\ \varepsilon \le \chi_\alpha^2(n_\theta)\Big\}, \tag{7}$$

with $\varepsilon := \theta - \theta_{\text{init}}$, and $\chi_\alpha^2(n_\theta)$ denoting the α-probability level for a χ^2 distribution with n_θ degrees of freedom.

The goal in this paper is to design an input spectrum Φ_u that maximizes the model accuracy in a worst-case sense with respect to the initial uncertainty region (7). Specifically, the following robust design problem is considered,

$$\begin{aligned} &\underset{\Phi_u,\gamma}{\text{minimize}}\ \gamma \\ &\text{subject to} \\ &C_X\left(e^{j\omega},\theta,\Phi_u(\omega)\right) \preceq \gamma, \quad \omega \in \Omega, \qquad \theta \in \Theta_{\text{init}} \\ &P_u(\Phi_u(\omega)) + wP_y(\theta,\Phi_u(\omega)) \le 1, \qquad \theta \in \Theta_{\text{init}}, \end{aligned} \tag{8}$$

$$\tag{9}$$

where $X \in \{\theta, G\}$ and $\Omega = (-\pi,\pi]$ and w is a user-selected weight. For the choice $X = G$, the objective is the minimization of the largest eigenvalue of the worst-case plant frequency-wise covariance, see (6). For the choice $X = \theta$, the objective becomes an E-optimality criterion [19].

The constraint (9) reflects limited experimental resources,

given by the weighted sum of the input power P_u and output power P_y, where

$$\begin{aligned} P_u(\Phi_u) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{Tr}\Big(\Phi_u(\omega)\Big)d\omega, \\ P_y(\theta,\Phi_u) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{Tr}\Big(G(q,\theta)\Phi_u(\omega)G^H(q,\theta)\Big)d\omega, \end{aligned} \tag{10}$$

and $w \geq 0$ a user-defined weight.

The key difference between the robust design problem (8) and nominal input design is that the constraints in (8) are evaluated over the uncertainty set Θ_{init}, while in nominal design these are evaluated around a point estimate of θ_o.

Problem (8) involves the optimization of the infinite-dimensional decision variable $\Phi_u(\omega)$ over an infinite-dimensional constraint set Θ_{init}. In the next section, the decision variable is discretized to obtain a tractable formulation.

D. Finite spectrum parametrization

To obtain a tractable formulation of (8), the infinite-dimensional spectrum Φ_u is parametrized as that of a multisine signal at frequencies $\omega_m,\ m = 1,\dots,L$, i.e., [16]

$$\Phi_u(\omega) = \pi \sum_{m=1}^{L} \tilde{\Phi}_m\Big(\delta(\omega-\omega_m)+\delta(\omega+\omega_m)\Big), \tag{11}$$

where $\tilde{\Phi}_m \in \mathbb{H}_+^{n_u}, \forall m$ and δ is the unit impulse function. Using (11), the information matrix $M(\theta,\Phi_u)$ in (4) becomes

$$M(\theta,\tilde{\Phi}) = N\Re\sum_{m=1}^{L} \mathcal{F}^H(q_m,\theta)\Big(\overline{\tilde{\Phi}_m}\otimes\Lambda^{-1}\Big)\mathcal{F}(q_m,\theta), \tag{12}$$

where $q_m = e^{j\omega_m}$, and the power $P = P_u + wP_y$ becomes

$$P(\theta,\tilde{\Phi}) = \sum_{m=1}^{L}\mathrm{Tr}\Big(\tilde{\Phi}_m + wG(q_m,\theta)\tilde{\Phi}_m G^H(q_m,\theta)\Big). \tag{13}$$

With these finite parametrizations, the problem (8) becomes a semi-infinite program (SIP) [12], i.e., an optimization problem over a finite number of decision variables but with an infinite-dimensional constraint set. SIPs are generally difficult to solve in exact sense [12]. In the next section, an iterative robust experiment design algorithm is presented that solves (8) approximately.

III. ROBUST EXPERIMENT DESIGN ALGORITHM

A. Main idea: Iterative robust design via the S-lemma

The presence of the infinite-dimensional constraint $\theta \in \Theta_{\text{init}}$ renders (8) a complex problem to solve. Commonly, the set constraint is (iteratively) approximated by a finite subset, e.g., [13], [15]. In this paper, the constraint $\varepsilon \in \Theta_{\text{init}}$ is not approximated but addressed in its infinite-dimensionality. Instead, the program (8) is approximated iteratively by a sequence of convex programs of the form

$$\begin{aligned} &\underset{\tilde{\Phi},\gamma}{\text{minimize}}\ \gamma \\ &\text{subject to } \tilde{f}(q_m,\varepsilon,\tilde{\Phi},\gamma) \geq 0 \quad \forall m, \quad \varepsilon\in\Theta_{\text{init}} \\ &\qquad\qquad \tilde{g}(\varepsilon,\tilde{\Phi},\gamma) \quad\ \geq 0, \qquad\quad \varepsilon\in\Theta_{\text{init}}, \end{aligned} \tag{14}$$

that are solved exactly at each iteration. Here, the functions $\tilde{f},\tilde{g}$ are quadratic approximations to the constraints in (8).

The key existing technical result that is employed to solve (14) is the S-lemma [20], which provides conditions under which the non-negativity of one quadratic function is a consequence of another one. Since, by virtue of (7), the constraint $\varepsilon \in \Theta_{\text{init}}$ in (14) is quadratic, the following result is due to the S-lemma.

Lemma 1: Given the quadratic function

$$\tilde{f}(\varepsilon) := \begin{bmatrix}\varepsilon\\1\end{bmatrix}^T \mathcal{A}_f \begin{bmatrix}\varepsilon\\1\end{bmatrix}. \tag{15}$$

Then, $\tilde{f}(\varepsilon) \geq 0$ for all $\varepsilon \in \Theta_{\text{init}}$, with Θ_{init} given by (7), if and only if there exists a scalar $\beta \geq 0$ such that

$$\Re(\mathcal{A}_f) - \beta\mathcal{E} \succeq 0, \tag{16}$$

where $\mathcal{E} = \mathrm{diag}\big(\ [-M_{\text{init}},\ \chi^2_\alpha(n_\theta)]\ \big)$.

When the functions $\tilde{f},\tilde{g}$ are constructed to be convex in $\tilde{\Phi}$ and γ, Lemma 1 enables formulating and solving (14) as a convex program. Hence, suitable selection of the functions $\tilde{f},\tilde{g}$ is crucial to facilitate solving (14) on the one hand, and ensuring that (14) reflects the original problem (8). To achieve this, program (8) is first recast as a convex SIP that is affine in the decision variables in the next section.

B. Reformulation to convex SIP

To enable approximating the input design problem by the form (14), the original program (8) is reformulated such that the constraints become affine in the decision variables. Exploiting the finite spectrum parametrization (8), the performance constraint in (8) is reformulated as an affine function in $\tilde{\Phi}$ by applying the Schur complement:

$$C_X(q,\theta,\tilde{\Phi}) \preceq \gamma \ \Leftrightarrow\ M(\theta,\tilde{\Phi}) - \frac{1}{\gamma}R_X(q,\theta) \succeq 0,$$

with $R_X(q,\theta) = \Lambda_X^H(q,\theta)\Lambda_X(q,\theta)$ and $\Lambda_\theta = I$. Additionally, introducing the coordinate transform $\tilde{\Psi} = \gamma\tilde{\Phi}$ enables expressing (8) exactly as the convex SIP,

$$\begin{aligned} &\underset{\tilde{\Psi},\gamma}{\text{minimize}}\ \gamma \\ &\text{subject to, for } \theta\in\Theta_{\text{init}}, \\ &f(q_m,\theta,\tilde{\Psi}) := \underline{\sigma}\Big(M(\theta,\tilde{\Psi}) - R_X(q_m,\theta)\Big) \geq 0 \quad \forall m, \\ &g(\theta,\tilde{\Psi},\gamma) \quad := -P_u(\tilde{\Psi}) - wP_y(\theta,\tilde{\Psi}) + \gamma \geq 0, \end{aligned} \tag{17}$$

with $\underline{\sigma}$ the smallest eigenvalue. In the next section, quadratic approximation functions to f,g in (17) are presented that enable solving (17) approximately using the S-lemma.

C. Quadratic approximations of f and g

In this section, the quadratic approximation functions $\tilde{f},\tilde{g}$ of f,g in (17) are presented.

To approximate the nonlinear function f by a quadratic function $\tilde{f}$ of form (15), first the information matrix $M(\theta,\tilde{\Psi})$ is approximated. To this end, the term $\mathcal{F}(q,\theta_*+\varepsilon_*)$ in (5) is approximated by a first-order Taylor expansion around the point θ_* as

$$\tilde{\mathcal{F}}(q,\theta_*,\varepsilon_*) = \mathcal{F}(q,\theta_*) + (I_n\otimes\varepsilon_*^T)\nabla_\theta\mathcal{F}(q,\theta_*), \tag{18}$$

where $\varepsilon_* = \theta - \theta_*$ and $n = n_u n_y$.
To express the deviations around θ_* to deviations around the center θ_{init} of Θ_{init}, the coordinate transformation $\varepsilon_* = \varepsilon + (\theta_{\text{init}} - \theta_*)$ is imposed onto (18). This yields

$$\tilde{\mathcal{F}}(q, \theta_*, \varepsilon) = \mathcal{F}_1(q, \theta_*) + \mathcal{F}_2(q, \theta_*, \varepsilon), \tag{19}$$

where

$$\mathcal{F}_1(q, \theta_*) = \mathcal{F}(q, \theta_*) + \Big(I_n \otimes (\theta_{\text{init}} - \theta_*)^T\Big)\nabla_\theta \mathcal{F}(q, \theta_*),$$
$$\mathcal{F}_2(q, \theta_*, \varepsilon) = (I_n \otimes \varepsilon^T)\nabla_\theta \mathcal{F}(q, \theta_*).$$

This gives rise to the approximated information matrix

$$\tilde{M}(\theta_*, \varepsilon, \tilde{\Psi}) = N\Re \sum_{m=1}^{L} \tilde{\mathcal{F}}^H(q_m, \theta_*, \varepsilon)\big(\overline{\tilde{\Psi}_m} \otimes \Lambda^{-1}\big)\tilde{\mathcal{F}}(q_m, \theta_*, \varepsilon).$$

Following the same procedure, the matrix R_X in (17) is approximated as

$$\tilde{R}_X(q, \theta_*, \varepsilon) = \tilde{\Lambda}_X^H(q, \theta_*, \varepsilon)\tilde{\Lambda}_X(q, \theta_*, \varepsilon). \tag{20}$$

Exploiting these approximations, the function $\tilde{f}(q, \theta_*, \varepsilon, \tilde{\Psi})$ represents the smallest eigenvalue of $\tilde{M} - \tilde{R}_X$ via

$$\tilde{f}(q, \theta_*, \varepsilon, \tilde{\Psi}) = \mathcal{U}^H\Big(\tilde{M}(\theta_*, \varepsilon, \tilde{\Psi}) - \tilde{R}_X(q, \theta_*, \varepsilon)\Big)\mathcal{U}, \tag{21}$$

where $\mathcal{U}(q, \theta_*, \tilde{\Psi}) \in \mathbb{C}^{n_\rho}$ is the eigenvector corresponding to the smallest eigenvalue of $\tilde{M}(\theta_*, \varepsilon, \tilde{\Psi}) - \tilde{R}_X(q, \theta_*, \varepsilon)$ at the point θ_*. The arguments of $\mathcal{U}$ in (21) have been omitted for notational brevity. Rearranging terms allows expressing $\tilde{f}$ in (21) as the quadratic function,

$$\tilde{f}(q, \theta_*, \varepsilon, \tilde{\Psi}) = \begin{bmatrix}\varepsilon \\ 1\end{bmatrix}^T \mathcal{A}_f(q, \theta_*, \tilde{\Psi}) \begin{bmatrix}\varepsilon \\ 1\end{bmatrix}, \tag{22}$$

where

$$\begin{aligned}\mathcal{A}_f = N \sum_{m=1}^{L} &\mathcal{Z}_f^H(q_m, \theta_*, \tilde{\Psi}_m)(\overline{\tilde{\Psi}_m} \otimes \Lambda^{-1})\mathcal{Z}_f(q_m, \theta_*, \tilde{\Psi}_m) \\ &- \mathcal{Z}_X^H(q, \theta_*, \tilde{\Psi})\mathcal{Z}_X(q, \theta_*, \tilde{\Psi})\end{aligned} \tag{23}$$

with

$$\mathcal{Z}_f = \big[(I_n \otimes \mathcal{U}^T(q, \theta, \tilde{\Psi}))\overline{\nabla_\theta \mathcal{F}(q, \theta)} \quad \mathcal{F}_1(q, \theta_*)\mathcal{U}(q, \theta, \tilde{\Psi})\big]$$
$$\mathcal{Z}_X = \big[(I_n \otimes \mathcal{U}^T(q, \theta, \tilde{\Psi}))\overline{\nabla_\theta \Lambda_X(q, \theta)} \quad \Lambda_{X_1}(q, \theta_*)\mathcal{U}(q, \theta, \tilde{\Psi})\big].$$

Similarly, the quadratic approximation function $\tilde{g}$ of g in (17) is constructed as

$$\tilde{g}(\theta_*, \varepsilon, \tilde{\Psi}, \gamma) = \begin{bmatrix}\varepsilon \\ 1\end{bmatrix}^T \mathcal{A}_g(\theta_*, \tilde{\Psi}, \gamma) \begin{bmatrix}\varepsilon \\ 1\end{bmatrix}. \tag{24}$$

The derivation of $\mathcal{A}_g$ is omitted due to space limitations. The function $\tilde{f}$ in (22) has the following properties:

Lemma 2: The approximation function $\tilde{f}$ in (22) has the following properties:

i. *The approximation is exact at the point θ_*, i.e., $\tilde{f}(\theta_*, \varepsilon, \tilde{\Psi}) = f(\theta_*, \tilde{\Psi})$ for $\varepsilon = \theta_* - \theta_o$.*
ii. *The Jacobian $\nabla_\varepsilon \tilde{f}(\theta_*, \varepsilon, \tilde{\Psi}) = \nabla_\theta f(\theta, \tilde{\Psi})$ at the point $\{\theta, \varepsilon\} = \{\theta_*, \theta_* - \theta_o\}$.*
iii. *Define $\theta_{f,\text{wc}} = \arg\min_{\theta \in \Theta_{\text{init}}} f(\theta, \tilde{\Psi})$. The approximation function $\tilde{f}$ attains the same minimum as f, i.e., $\inf_{\theta \in \Theta_{\text{init}}} f(\theta, \tilde{\Psi}) = \inf_{\varepsilon \in \Theta_{\text{init}}} \tilde{f}(\theta_{f,\text{wc}}, \varepsilon, \tilde{\Psi})$.*

A similar result applies to the function $\tilde{g}$ in (24).
The approximation functions $\tilde{f}, \tilde{g}$ are exploited in the robust experiment design algorithm in the next section.

D. Robust experiment design algorithm

In this section, the robust experiment design algorithm is presented to approximately solve (17).
The approximations $\tilde{f}, \tilde{g}$ in (22), (24) are quadratic in ε, which enables evaluation of the conditions $\tilde{f} \geq 0$, $\tilde{g} \geq 0$ over the ellipsoidal set Θ_{init}, by virtue of Lemma 1. However, the complexity that arises in view of input design (14) is that these functions are non-linear and non-convex functions of the inputs $\tilde{\Psi}$. The reasons are twofold: 1) the worst-case parameters $\theta_{f,\text{wc}}$ and $\theta_{g,\text{wc}}$ are non-convex functions of $\tilde{\Psi}$, and 2) the eigenvectors $\mathcal{U}(q, \theta_{f,\text{wc}}, \tilde{\Psi})$ in (21) are non-convex in $\theta_{f,\text{wc}}$ and $\tilde{\Psi}$. Differently stated, the points $\theta_{f,\text{wc}}, \theta_{g,\text{wc}}$ and the eigenvectors $\mathcal{U}$ depend on the solution $\tilde{\Psi}$, and cannot be determined before the problem (14) is solved.
This dependency is resolved by an iterative procedure, where the parameters $\{\theta_{f,\text{wc}}, \theta_{g,\text{wc}}, \mathcal{U}\}$ are based on a previous iteration. Let the function $\tilde{f}^{[x]}$ be defined as

$$\tilde{f}^{[x]}(q, \varepsilon, \tilde{\Psi}_u) = \begin{bmatrix}\varepsilon \\ 1\end{bmatrix}^T \mathcal{A}_f^{[x]}(q, \tilde{\Psi}) \begin{bmatrix}\varepsilon \\ 1\end{bmatrix}, \tag{25}$$

where $\mathcal{A}_f^{[x]}(q, \tilde{\Psi})$ is of the form (23) but with $\mathcal{Z}_f(q, \theta, \tilde{\Psi})$ and $\mathcal{Z}_X(q, \theta, \tilde{\Psi})$ substituted by the fix matrices

$$\begin{aligned}\mathcal{Z}_f^{[x]}(q) &= \mathcal{Z}_f(q, \theta_{f,\text{wc}}^{[x]}, \tilde{\Psi}^{[x]}), \\ \mathcal{Z}_X^{[x]}(q) &= \mathcal{Z}_X(q, \theta_{f,\text{wc}}^{[x]}, \tilde{\Psi}^{[x]}).\end{aligned} \tag{26}$$

As a result, (25) becomes convex in $\tilde{\Psi}$. Likewise, the function $\tilde{g}(\theta_*, \varepsilon, \tilde{\Psi}, \gamma)$ is convex in the variables $\varepsilon, \tilde{\Psi}$, and γ when evaluated at a fix point $\theta_{g,\text{wc}}^{[x]}$.
The obtained convex approximation functions are exploited in Algorithm 1. The algorithm alternates between updating $\{\theta_{f,\text{wc}}, \theta_{g,\text{wc}}, \mathcal{U}\}$ based on a previous (candidate) solution, and solving the following convex program,

$$\begin{aligned}&\underset{\tilde{\Psi}, \hat{\gamma}, \beta_f^{[x]}, \beta_g^{[x]}, \eta}{\text{minimize}} \quad \hat{\gamma} + \tau\eta^2 \\ &\text{subject to, for } x = a, b: \qquad (27) \\ &\Re\Big(\mathcal{A}_f^{[x]}(q_m, \tilde{\Psi})\Big) \quad - \beta_f^{[x]}\mathcal{E} + \eta\mathcal{I} \succeq 0, \quad m = 1, \ldots, L, \\ &\Re\Big(\mathcal{A}_g(\theta^{[x]}, \tilde{\Psi}, \hat{\gamma})\Big) - \beta_g^{[x]}\mathcal{E} \succeq 0, \\ &\Big\|\tilde{\Psi}_m - \tilde{\Psi}_m^{[a]}\Big\| \leq \epsilon, \quad m = 1, \ldots, L, \qquad (28) \\ &\hat{\gamma}, \beta_f^{[x]}, \beta_g^{[x]} \geq 0, \tilde{\Psi}_m \succeq 0,\end{aligned}$$

where $\mathcal{I} = \text{diag}([\delta_{1,n_\theta+1}, \ldots, \delta_{n_\theta+1,n_\theta+1}])$ with $\delta_{i,j}$ the Kronecker delta.
Program (27) computes, by virtue of Lemma 1, the spectrum $\tilde{\Psi}$ that satisfies $\tilde{f}^{[x]} + \eta \geq 0$ and $\tilde{g}^{[x]} \geq 0$ for $\varepsilon \in \Theta_{\text{init}}$ for the smallest value of $\hat{\gamma} + \tau\eta^2$. Herein, the second term is a penalty term, where $\tau > 0$ and η is a slack variable that softens the first constraint in (27) to ensure that the program is feasible in each iteration.
The approximations in (27) depend on the prior iterations ($[a]$) and prior candidate solutions ($[b]$). To achieve accurate

Algorithm 1 Robust Experiment Design

Input: Initial data: $\tilde{\Psi}^{[0]}, \theta^{[0]}_{f,\text{wc}}, \theta^{[0]}_{g,\text{wc}}, \gamma^{[0]}$. Algorithm parameters: $\kappa \in (0,1), \tau > 0, \epsilon^{[0]} > 0$. **Output**: $\tilde{\Psi}^{[i]}$.

1: **Initialize** Set $i = 1, v = 0$.
2: **Loop**
3: Set $[a] = [b] = [i-1]$.
4: Compute candidate solution:
5: Set step size: $\epsilon = \kappa^v \cdot \epsilon^{[0]}$.
6: Solve program (27) to obtain candidate $\tilde{\Psi}^{[*]}$.
7: Global search: determine points $\theta^{[*]}_{f,\text{wc}}$, $\theta^{[*]}_{g,\text{wc}}$ and $g_{\text{wc}} = g(\theta^{[*]}_{g,\text{wc}}, \tilde{\Psi}^{[*]})$, $\gamma^{[*]} = f(\theta^{[*]}_{f,\text{wc}}, \tilde{\Psi}^{[*]})$.
8: Make feasible: $\tilde{\Psi}^{[*]} \leftarrow \min(\frac{\gamma^{[*]}}{\gamma^{[*]} - g_{\text{wc}}}, 1)\tilde{\Psi}^{[*]}$.
9: Compute true performance: $\gamma^{[*]} = f(\theta^{[*]}_{f,\text{wc}}, \tilde{\Psi}^{[*]})$.
10: **if** $\gamma^{[*]} \leq \gamma^{[i-1]}$
11: Accept candidate: $\{\tilde{\Psi}^{[i]}, \gamma^{[i]}, \theta^{[i]}_{f,\text{wc}}, \theta^{[i]}_{g,\text{wc}}\} \leftarrow \{\tilde{\Psi}^{[*]}, \gamma^{[*]}, \theta^{[*]}_{f,\text{wc}}, \theta^{[*]}_{g,\text{wc}}\}$.
12: Update step size parameter: $v \leftarrow \max(v-1, 0)$.
13: Advance $i \leftarrow i+1$.
14: **else**
15: Update step size parameter: $v \leftarrow v+1$.
16: Set $[b] \leftarrow [*]$.
17: Return to line 4.
18: **end if**

approximations, the constraint (28) controls the stepsize, i.e., the distance between the new and the previous solution.

Algorithm 1 consists in, at each iteration i, the computation of a candidate spectrum $\tilde{\Psi}^{[*]}$ (lines 6 - 8) that satisfies the constraint on the true power, and improves the true performance $\gamma^{[*]}$ upon the previously achieved performance $\gamma^{[i-1]}$ (line 10). The true parameters $\theta^{[*]}_{f,\text{wc}}, \theta^{[*]}_{g,\text{wc}}$ are obtained by a global search strategy (line 7). Specifically, the set Θ_{init} is uniformly randomly sampled, where sufficiently dense sampling is assumed.

If a candidate solution does not improve performance (line 14), the step size parameter ϵ in (28) is reduced (lines 15 and 5). In addition, constraints based on the candidate solution are added to the program (27) via line 16.

The behavior of Algorithm 1 is characterized as follows.

Theorem 1: Algorithm 1 has the following properties:

i) *Program (27) is feasible in each iteration* $i \geq 1$.
ii) *The sequence* $\{\gamma^{[i]}_f\}$ *is monotonically non-decreasing.*

In the next section, the algorithm is confronted with an experimental motion system.

IV. EXPERIMENTAL CASE STUDY

A. Experiment description

1) Experimental setup: The presented approach is applied for input design of the steel beam in Fig. 2. The system exhibits predominant flexible dynamics, hereby mimicking the behavior of the next-generation wafer stage in Fig. 1. The beam has a translational and rotational motion degree of freedom, and is equipped with three voice-coil actuators and three contactless fiberoptic sensors.

2) Initial model: An initial 16th-order model G_{init} and uncertainty set Θ_{init} are estimated from an initial identification experiment. The frequency response of the (y_3, u_3)-entry of G_{init} is shown in (—) in Fig. 3. The system shows flexible behavior originating from six lightly damped flexible modes.

3) Goal: The goal in this case study is to compare input designs from three different design approaches:

a) Nominal design: Solution to program (8), where the set constraint $\theta \in \Theta_{\text{init}}$ is replaced by the point constraint $\theta = \theta_{\text{init}}$. The nominal spectrum is scaled to robustly satisfy the power constraint.
b) Robust design - SA: Solution to program (8), where the set constraint $\theta \in \Theta_{\text{init}}$ is replaced by a finite subset $\Theta_{\text{sample}} \subset \Theta_{\text{init}}$ that is augmented at each iteration with a random sample from Θ_{init}.
c) Robust design - Algorithm 1 as presented in this paper.

E-optimality is considered, i.e., $C_X = C_\theta$ in (8).

4) Settings: The frequency grid contains $L = 100$ frequencies. The settings used for approach c) are $\kappa = 0.8$, $\epsilon^{[0]} = 0.1L$, $w = 10^4$, and $\eta = 10$. The spectrum computed from the nominal approach a) is used as the initial spectrum in approaches b) and c).

B. Results

1) Convergence: The performance, in terms of the inverse of γ, i.e., $1/\gamma = \underline{\sigma}\left(M(\theta, \tilde{\Phi})\right)$, achieved by the different approaches is compared in Fig. 4. The values are based on the worst-case results from a Monte-Carlo analysis with 10.000 samples drawn randomly from Θ_{init}.

Algorithm 1 (—) shows monotonically non-decreasing performance. A factor 1.8 improvement compared to nominal design (- -) is achieved. Also, it outperforms the SA (—) for $N_{\text{iter}} \leq 40$. In the limit case $N_{\text{iter}} \to \infty$, the SA (—) achieves a factor 1.06 better performance (- -) than Algorithm 1 (- -). However, the computational burden of the SA is large, since it involves extending the constraint set at each iteration.

2) Monte-Carlo analysis: The distribution of the sampled performance indicators $\underline{\sigma}\left(M(\theta, \tilde{\Phi})\right)$ and the powers $P(\theta, \tilde{\Psi})$ obtained from the Monte-Carlo analysis is compared in Fig. 5. The nominal design (■) shows a long-tailed distribution in both the performance and the power. Algorithm 1 (■), for 40 iterations, renders a denser distribution, which enables an average power increase and in turn an improved robust performance. The distribution of the SA in the limit case is the most dense and yields the highest performance. This is expected behavior, since the SA converges to the global optimum in the limit $N_{\text{iter}} \to \infty$.

3) Spectrum: The input spectra are compared in Fig. 3. The nominal spectrum (—) has its energy highly concentrated around the resonance frequencies. The energy is more distributed in the spectrum obtained from Algorithm 1 (—) to achieve robustness to model uncertainty. Similar observations are made for the SA approach, but the results are not depicted to retain clarity in Fig. 3.

Overall, this case study demonstrates that Algorithm 1 provides a viable approach to robustify input spectrum design. The achieved performance in the limit $N_{\text{iter}} \to \infty$ is lower

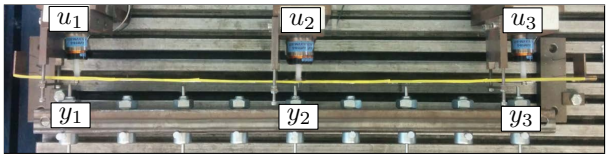

Fig. 2. Flexible beam (yellow) with actuators u_i and sensors y_i.

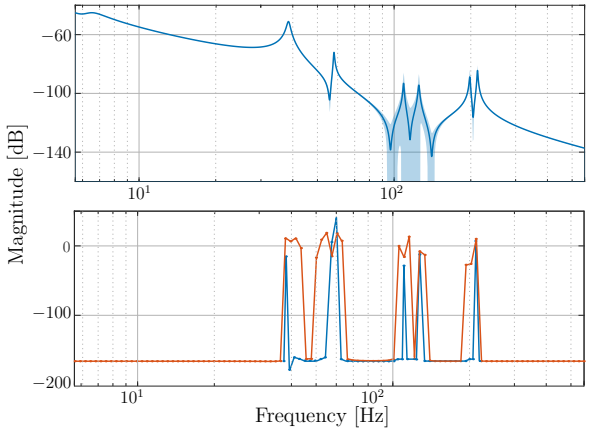

Fig. 3. *Top:* Entry of $G_{\text{init}}(3,3)$ including 95% confidence region. *Bottom:* Nominal spectrum (—) and robust spectrum using Algorithm 1 (—).

than using the SA, yet Algorithm 1 avoids the the inclusion of many constraints and the associated computational cost.

V. CONCLUSIONS

The method presented in this paper enables input spectrum design for high-quality identification for complex MIMO systems that is robust to ellipsoidal parametric uncertainty in the nominal plant model. The iterative algorithm alternates beween solving a convex approximate program and updating the approximation functions. The method avoids the discretization of the infinite-dimensional uncertainty set, as well as the inclusion of a large number of constraints in the optimization program, both of which are encountered in existing approaches such as the scenario approach. Experimental confrontation with a flexible beam setup shows that the method enables a significant robustification and achieves a performance close to that of the scenario approach.

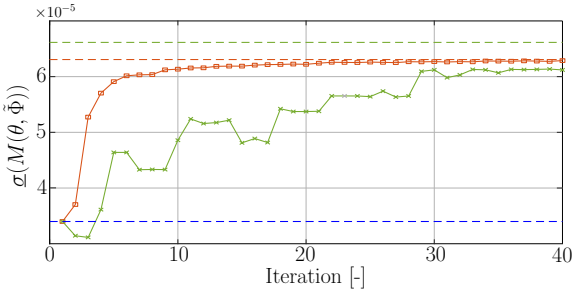

Fig. 4. The robust design (⊟) monotonically improves upon the nominal design (- -). Within $N_{\text{iter}} = 40$ iterations, it outperforms the SA (⋆) both in convergence speed and achieved performance. In the limit $N_{\text{iter}} \to \infty$, the SA achieves better performance (- -) than Algorithm 1 (- -).

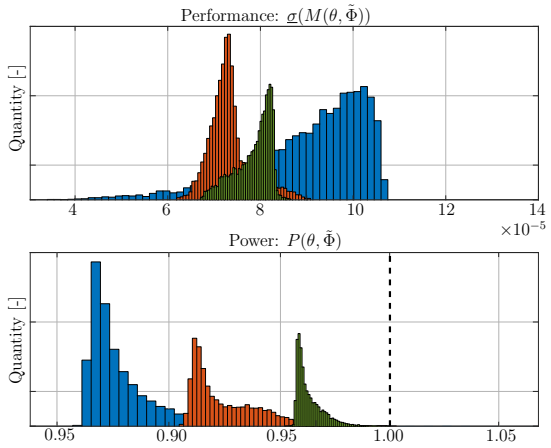

Fig. 5. Distribution of performance indicators (*top*) and powers (*bottom*). Algorithm 1 (■) and the SA (■) yield more compact distributions than nominal design (■), leading to better robust performance.

REFERENCES

[1] N. Dirkx, J. van de Wijdeven, and T. Oomen, "Frequency response function identification for multivariable motion control: Optimal experiment design with element-wise constraints," *Mechatronics*, vol. 71, p. 102440, 2020.

[2] V. V. Fedorov, *Theory of optimal experiments*. Elsevier, 2013.

[3] G. C. Goodwin, "Dynamic system identification: experiment design and data analysis," *Math. Sci. Eng.*, vol. 136, 1977.

[4] H. Hjalmarsson, "From experiment design to closed-loop control," *Automatica*, vol. 41, no. 3, pp. 393–438, 2005.

[5] M. Gevers, "Identification for control: From the early achievements to the revival of experiment design," *Eur. J. Control*, vol. 11, no. 4-5, pp. 335–352, 2005.

[6] M. Barenthin, X. Bombois, H. Hjalmarsson, and G. Scorletti, "Identification for control of multivariable systems: Controller validation and experiment design via LMIs," *Automatica*, vol. 44, no. 12, pp. 3070–3078, 2008.

[7] H. Jansson and H. Hjalmarsson, "Input design via LMIs admitting frequency-wise model specifications in confidence regions," *IEEE Trans. Automat. Contr.*, vol. 50, no. 10, pp. 1534–1549, 2005.

[8] X. Bombois, G. Scorletti, M. Gevers, P. M. Van den Hof, and R. Hildebrand, "Least costly identification experiment for control," *Automatica*, vol. 42, no. 10, pp. 1651–1662, 2006.

[9] L. Pronzato and É. Walter, "Robust experiment design via maximin optimization," *Math. Biosci.*, vol. 89, no. 2, pp. 161–176, 1988.

[10] J. Mårtensson and H. Hjalmarsson, "Robust input design using sum of squares constraints," *IFAC Proceedings Volumes*, vol. 39, no. 1, pp. 1352–1357, 2006.

[11] C. R. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer, "Robust optimal experiment design for system identification," *Automatica*, vol. 43, no. 6, pp. 993–1008, 2007.

[12] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Math. Oper. Res.*, vol. 23, no. 4, pp. 769–805, 1998.

[13] J. S. Welsh and C. R. Rojas, "A scenario based approach to robust experiment design," *IFAC Proceedings Volumes*, vol. 42, no. 10, pp. 186–191, 2009.

[14] R. Hettich and K. O. Kortanek, "Semi-infinite programming: theory, methods, and applications," *SIAM review*, vol. 35, no. 3, pp. 380–429, 1993.

[15] D. Katselis, C. R. Rojas, J. S. Welsh, and H. Hjalmarsson, "Robust experiment design for system identification via semi-infinite programming techniques," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 680–685, 2012.

[16] X. Bombois, F. Morelli, H. Hjalmarsson, L. Bako, and K. Colin, "Robust optimal identification experiment design for multisine excitation," *Automatica*, vol. 125, p. 109431, 2021.

[17] R. van Herpen, T. Oomen, E. Kikken, M. van de Wal, W. Aangenent, and M. Steinbuch, "Exploiting additional actuators and sensors for nano-positioning robust motion control," *Mechatronics*, vol. 24, no. 6, pp. 619–631, 2014.

[18] M. B. Zarrop, *Optimal experiment design for dynamic system identification*. Springer, 1979.

[19] J. Kiefer, "General equivalence theory for optimum designs (approximate theory)," *The annals of Statistics*, pp. 849–879, 1974.

[20] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.