A Fast Smoothing-Based Algorithm to Generate l_∞-Norm Constrained Signals for Multivariable Experiment Design

Dirkx, Nic; Bosselaar, Marcel; Oomen, Tom

DOI
10.1109/LCSYS.2021.3133655

Publication date
2022

Document Version
Accepted author manuscript

Published in
IEEE Control Systems Letters

A Fast Smoothing-based Algorithm to Generate l_∞-norm Constrained Signals for Multivariable Experiment Design

Nic Dirkx, Marcel Bosselaar, and Tom Oomen

***Abstract*— Handling peak amplitude constraints, or equivalently l_∞-norm constraints, is an important application demand in experiment design for system identification. The aim of this paper is to present a method for the design of excitation signals with prescribed power spectrum under l_∞-norm constraints for systems with many inputs and outputs. The method exploits an exponential smoothing function in an iterative algorithm. Fast convergence is achieved by a computationally efficient construction of the gradient and the Hessian matrix. Experimental results show excellent convergence behavior that overcomes local minima, while significantly reducing computation time compared to existing techniques.**

I. INTRODUCTION

The computation of optimal excitation signals plays a central role in the design of experiments for system identification. Typically, the aim of optimal experiment design (OED) is to maximize the signal-to-noise ratio in view of a selected model quality criterion, yet within the system its operating constraints [1], [2]. In many practical applications, including wafer scanners [3] and chemical processes [4], dealing with l_∞-norm constraints on the input and output signals is key to guarantee safe experiments.

The complexity of OED depends on the constraints, consequently the availability of algorithms heavily depends on the specific constraints that are imposed. In case of power constraints, OED problems can often be formulated as convex optimization programs, with the input spectrum as decision variable [5], [6]. Such problems can readily be solved by convex optimization techniques [7]. For l_∞-norm signal constraints, such approach cannot be applied, since the signal peak amplitude depends not only spectral power, but also on the phase [8]. The relationship between the peak amplitude and the phase is highly nonlinear and non-smooth [9], which complicates optimal design procedures, especially for multiple inputs multiple outputs (MIMO) systems.

To mitigate the full complexity in solving l_∞-norm constrained OED problems, various methods have been developed to solve the related problem of crest-factor minimization for signals with prescribed power spectrum. The crest-factor is the ratio between signal peak amplitude and signal power [8]. Exact solutions to crest-factor minimization are not available, hence methods tend to be of heuristic nature. In [10], a phase selection law is presented that often yields a low crest-factor for scalar-valued signals. In [12], [13], a time-frequency domain swapping method is presented for joint crest-factor reduction of input and output signals. Extension of these methods to MIMO systems is complicated due to the heuristic nature of the algorithms. An optimization-based approach to crest-factor minimization is presented in [9], wherein the l_∞-norm is iteratively approximated by the l_p-norm. The method can be applied to MIMO systems, but becomes computationally intensive for systems with many inputs and outputs.

An alternative approach to achieve a low crest-factor is by the generation of binary signals that approximate a prescribed spectrum, since binary signals achieve maximal power for a given amplitude constraint [11]. In [14], a simulated annealing approach is employed to globally solve this nonconvex problem. The method is computationally intensive. Recursive algorithms based on a receding horizon concept are presented in [15], [16]. The algorithms are of lower computational complexity but are not defined for MIMO systems. In [17], excitation signals are computed from a l_∞-norm constrained time-domain optimization problem, which typically leads to binary signals. The method involves solving a large scale semi-definite program, which becomes computationally intractable for large MIMO systems.

Although many approaches for OED have been developed for a wide range of system identification problems, tractable methods for current applications with increasing complexity, such as many inputs and outputs, are not available. The aim of this paper is to develop an efficient algorithm for l_∞-norm constrained design of excitation signals with prescribed power spectrum, that is particularly suitable for large scale and multivariable experiment design problems.

The main contributions of this paper are:

1. An algorithm for l_∞-norm minimization of multivariable signals that exploits an exponential smoothing function to overcome local minima (Section III),
2. an efficient and scalable computation of the gradient and Hessian matrix, including a quantification of computational complexity (Section IV),
3. an experimental validation on a multivariable Active Vibration Isolation System (Section V).

Smooth approximations provide a solution to deal with non-smoothness and non-convexity at the same time, and are

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

Marcel Bosselaar is with the Dept. of Mechanical Engineering, Control Systems Technology, Eindhoven University of Technology, Eindhoven, The Netherlands.

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.

applied in non-smooth convex [18], [19] and non-convex optimization [20], [22] to accelerate convergence and increase parameter space exploration.

Notations: Operator $\odot$ denotes the element-wise product. Operations $X = \mathcal{F}(x)$ and $x = \mathcal{F}^{-1}(X)$ denote the Discrete Fourier Transform (DFT) of x and the inverse DFT of X [8], respectively. $\overline{x}$ denotes the complex conjugate of x.

II. PROBLEM FORMULATION

A. Experiment design for system identification

In system identification the goal is to estimate a system model from measured data. The identification setup is given by the discrete-time input-output relation

$$y(n) = G(z) * u(n) + \nu(n), \tag{1}$$

where $G(z)$ represents a $n_y \times n_u$ open-loop or closed-loop LTI system to be identified, with z the z-domain operator. Signal $u(n) \in \mathbb{R}^{n_u}$ is a user-defined vector-valued excitation signal, where $n = [0, \dots, N-1]$ the discrete-time index, for sample size N. Signal $y \in \mathbb{R}^{n_y}$ represents the measurements, perturbed by an independent and identically distributed random sequence ν. Note that y may encompass both input and output signals, via suitable choice of G.

The quality of the identified model of G depends on the design of the excitations u. OED consists in the computation of optimal excitation signals u within constraints:

$$\begin{aligned} u^{\text{opt}} = \underset{u}{\text{minimize}}\ & \mathcal{I}(G, u) \\ \text{subject to}\ & l_\infty(y_p) \leq c_p, \quad p = 1, \dots, n_y, \end{aligned} \tag{2}$$

where criterion $\mathcal{I}$ expresses the quality of the to-be-identified model, y_p is the p^{th} signal in y and c_p denotes the corresponding constraint value. The l_∞-norm is defined as follows.

Definition 1: The l_∞-norm of a scalar-valued signal $x(n)$ is defined as its absolute peak value in the interval $[0, N-1]$,

$$l_\infty(x) = \max_{n \in [0, N-1]} |x(n)|. \tag{3}$$

B. Signal parametrization

Multisine excitation signals u with a band-limited spectrum [8] are considered, i.e., for the q^{th} input,

$$u_q(n, A, \Phi) = \sum_{k=1}^{N_k} a_{qk} \cos\left(\frac{2\pi k n}{N} + \phi_{qk}\right), \tag{4}$$

where $N_k \leq 1/2N - 1$ is the excited frequency band, and the amplitudes $a_{qk} > 0\ \forall q, k$ and phases $\phi_{qk} = [0, 2\pi),\ \forall q, k$ are collected in A and Φ, respectively.

The output signals y_p in (1) for multisine inputs u_q in (4), assuming $\nu = 0$, are given by

$$y_p(n) = \sum_{q=1}^{n_u} \sum_{k=1}^{N_k} b_{pqk} \cos\left(\frac{2\pi k n}{N} + \xi_{pqk}\right), \tag{5}$$

with $b_{pqk} = a_{qk}|G_{pqk}|$ and $\xi_{pqk} = \phi_{qk} + \angle G_{pqk}$, and where $|G_{pqk}|$ and $\angle G_{pqk}$ denote the magnitude and the phase of entry $[p, q]$ of G at the k^{th} spectral line, respectively.

C. Objective: l_∞-norm constrained excitation design

Solving the OED problem (2) consists in the design of both the spectral magnitudes A and the phases Φ in (4). In the present paper, it is assumed that computation of the magnitudes in view of some selected criterion $\mathcal{I}$ in (2) has been performed as prior step, e.g., by convex spectrum design approaches in [5], [6]. The design problem considered in this paper is the subsequent phase design step to generate l_∞-norm bounded signals, for such prescribed spectral magnitudes. By using the equivalence between $l_\infty(y_p) \leq c_p, \forall p$ and $l_\infty(\bar{y}) \leq 1$ with $\bar{y} = [y_1/c_1, \dots, y_{n_y}/c_{n_y}]$, this problem is formulated as the unconstrained l_∞-norm minimization problem over Φ, i.e.,

$$\underset{\Phi}{\text{minimize}}\ l_\infty\left(\bar{y}(\Phi)\right). \tag{6}$$

Prior knowledge of G is required to solve (6), which is assumed available, e.g., from preliminary identification [6]. Solving (6) is non-straightforward since 1) the function $l_\infty(\bar{y}(\Phi))$ is non-convex in the variables Φ, 2) the function $l_\infty(\bar{y}(\Phi))$ is non-smooth and not differentiable over the entire domain of Φ, and 3) identification problems tend to be large in the number of variables, especially for systems with many inputs and outputs. Consequently, a dedicated and efficient strategy is required to solve (6). The concept of the strategy is presented in the next section.

D. Gradual smoothing strategy

The key idea of the presented approach is to solve the non-smooth problem (6) via a sequence of approximate smooth problems, wherein the level of smoothness is gradually reduced. A similar strategy is employed in [9], using l_p-norm approximations of the l_∞-norm to solve (6). In the present paper, an exponential smoothing function is employed that achieves accurate approximations. Besides obtaining a differentiable problem, a gradual smoothing strategy reduces the susceptibility to local minima, as exemplified next.

Example 1: Consider a scalar-valued signal u as in (4) with $N_k = 52$, $a_{1k} = \sqrt{2/N_k}$, $\forall k$ and uniformly distributed random phases $\phi_{1k} \in [0, 2\pi)$, $k = 1, \dots, N_k - 2$. The contour maps of $l_\infty(u)$ and the approximations are shown in Fig. 1. The method is formalized in the next section.

III. A SMOOTHING-BASED ALGORITHM TO l_∞-NORM MINIMIZATION

A. Smooth approximation function

To handle the complexity introduced by both the non-smoothness and the non-convexity in (6), the function $l_\infty^2\left(\bar{y}(\Phi)\right)$ is approximated by the smooth function,

$$L(\bar{y}, \sigma) = \sigma \ln\left(\sum_{n=0}^{\bar{N}-1} \exp\left(\frac{\bar{y}^2(n)}{\sigma}\right)\right), \tag{7}$$

where $\bar{N} = n_y N$ and $\sigma > 0$. This function is based on the exponential penalty function presented in [21], [23] for solving convex programs with inequality constraints. It is also applied for smoothing of minmax problems in [22]. Furthermore, (7) is a specific form of the log-sum-exp function encountered

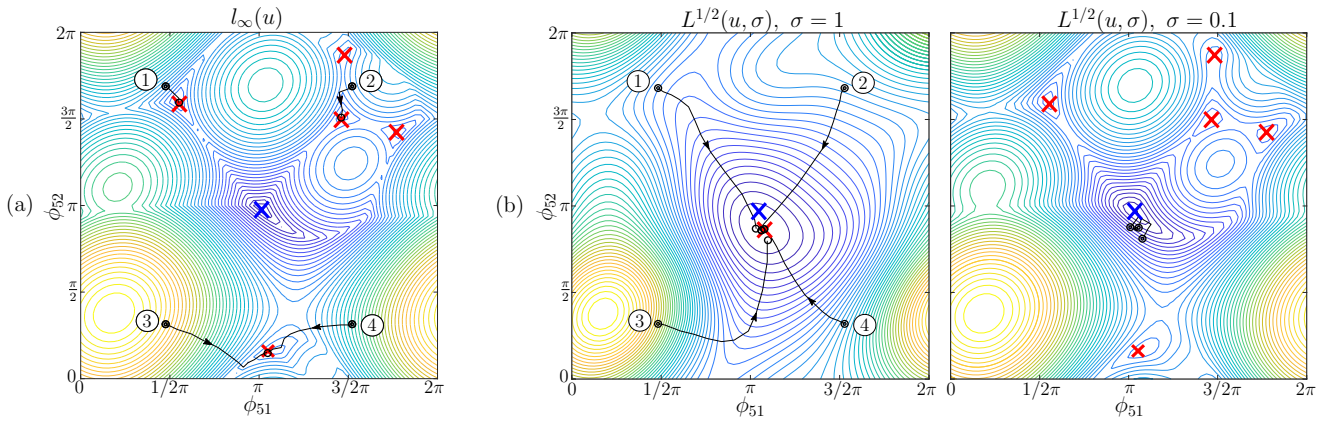

Fig. 1. Illustration of convergence behavior with and without smoothing strategy for Example 1. (a) Contour map of non-smooth $\boldsymbol{l_\infty(u)}$ as function of $\boldsymbol{\phi_{51}, \phi_{52}}$, with multiple local minima (×) and single global minimum (×). For this example, without a smoothing strategy, iteratively stepping in the steepest descent direction (black lines) from either of the numbered initial points leads to locally optimal solutions. (b) To improve performance, a gradual smoothing strategy is applied, depicted by the contour maps of the sequence of two smooth approximations of $\boldsymbol{l_\infty(u)}$, with local minima (×) and the global minimum of $\boldsymbol{l_\infty(u)}$ (×). *Left:* First, a high level of smoothing is applied. For this example, the approximant has a single local minimum (×) close to the global minimum (×) or the original problem. As a result, by iteratively stepping in the steepest descent direction from the numbered initial points, near global convergence is achieved. *Right:* Taking the previous solution as a new starting point, a next optimization is performed using a lower level of smoothing by which the global minimum is attained for all initial conditions.

in inference and classification in machine learning, e.g., [25]. Properties of the function $L(\bar{y},\sigma)$ are given in the following.

Lemma 1: Consider the function $L(\bar{y},\sigma)$ for $\sigma>0$. Then,

i) $l_\infty^2(\bar{y}) \le L(\bar{y},\sigma) \le l_\infty^2(\bar{y}) + \sigma\ln(N)$,
ii) $\lim_{\sigma\to 0}\ L(\bar{y},\sigma) = l_\infty^2(\bar{y})$,
iii) *$L(\bar{y},\sigma)$ is increasing with respect to σ.*
iv) *$L(\bar{y},\sigma)$ is twice continuously differentiable everywhere.*
v) *The gradient $\nabla_\Phi L(\bar{y},\sigma)$ is expressed as*

$$\nabla_\Phi L(\bar{y},\sigma) = \frac{\partial L}{\partial \phi} = \sum\nolimits_{n=0}^{\bar{N}-1} \lambda_n(\Phi,\sigma)\nabla_\Phi \bar{y}^2(\Phi,n), \quad (8)$$

where

$$\lambda_n(\Phi,\sigma) = \frac{\exp(\bar{y}^2(n)/\sigma)}{\sum_{n=0}^{\bar{N}-1}\exp(\bar{y}^2(n)/\sigma)}, \quad \sum_{n=0}^{\bar{N}-1}\lambda_n = 1. \quad (9)$$

See [25], [22] for a proof. Numerically stable implementations of L that avoid overflow are addressed in [22], [24]. The optimization problem based on L is formulated next.

B. Optimization problem formulation

The smoothing function $L(\bar{y},\sigma)$ is exploited to approximate the original non-smooth problem (6) by the smooth problem

$$\underset{\Phi}{\text{argmin}}\ L\left(\bar{y}(\Phi),\sigma\right) \quad (10)$$

for some selected $\sigma>0$. To obtain a quadratic form that is compatible with Gauss-Newton-type (GN) algorithms [26], the following equivalent problem to (10) is considered,

$$\underset{\Phi}{\text{argmin}}\ \tilde{L}\left(\bar{y}(\Phi),\sigma\right), \quad (11)$$

where

$$\tilde{L}(\bar{y},\sigma) = \frac{1}{2}\varepsilon^T(\bar{y},\sigma)\varepsilon(\bar{y},\sigma), \quad (12)$$

with $\varepsilon(\bar{y},\sigma) = \left[\exp\left(\frac{\bar{y}^2(0)}{2\sigma}\right),\ldots,\exp\left(\frac{\bar{y}^2(\bar{N}-1)}{2\sigma}\right)\right]^T$. The gradients of L and $\tilde{L}$ are equal up to a scale factor, since

$$\nabla_\Phi L(\bar{y},\sigma) = \sigma\tilde{L}^{-1}(\bar{y},\sigma)\nabla_\Phi\tilde{L}(\bar{y},\sigma), \quad (13)$$

and hence L and $\tilde{L}$ have identical minimizers.

Various optimization algorithms may be employed to solve (11). Typically, the parameters Φ are updated iteratively by an update law of the form

$$\Phi^{i+1} = \Phi^i - (\mathcal{H}^i)^{-1}(\mathcal{J}^i), \quad (14)$$

with i the iteration index, and where $\mathcal{J}$ and $\mathcal{H}$ contain curvature information of $\tilde{L}$. In the next section, an iterative algorithm is presented for solving (6).

C. A descent algorithm

This section presents Algorithm 1, that enables accurately solving (6) by a joint minimization of the smoothing function $L(\bar{y},\sigma)$ and a gradual decrease of the smoothing parameter. The algorithm bears resemblance to that in [22] for solving minmax problems, but differs in the selection of the search direction and the control of the smoothing parameter.

In step 2, a Steepest Descent (SD) direction [26] is selected. Step 3 involves a Wolfe condition [26] to determine the step size α^i. Step 4 sets the parameters $\mathcal{J},\mathcal{H}$ in (14) that give rise to a SD algorithm. The Φ-parameters are updated in step 5. In step 6, approximation parameter σ^i is controlled by the rate of descent of the function L. Properties ii) and iii) in Lemma 1 enable the following result.

Theorem 1: Let $\{L(\Phi^i,\sigma^i)\}$ be the sequence generated by Algorithm 1 for $i\to\infty$. Then,

(i) *Sequence $\{L(\Phi^i,\sigma^i)\}$ is monotonically non-increasing.*
(ii) *The point Φ^∞ is a stationary point of $l_\infty(\bar{y}(A,\Phi))$.*

Proof: (i) For fixed $\sigma^i>0$, $-d^i$ is a descent direction of $\tilde{L}$ and L, where the latter follows from (13). Step 3 guarantees that $L(\Phi^{i+1},\sigma^i)\le L(\Phi^i,\sigma^i)$. Step 6 guarantees that $\sigma^{i+1}\le\sigma^i$. Thus, by property iii) in Lemma 1, $L(\Phi^{i+1},\sigma^{i+1})\le L(\Phi^{i+1},\sigma^i)\le L(\Phi^i,\sigma^i)$. (ii) For fixed $\sigma^i>0$, steps 2-5 guarantee that $L(\Phi^i,\sigma^i)$ is reduced towards a stationary point for $i\to\infty$. Step 6 guarantees reduction of σ by $q\leftarrow 1$ when a

Algorithm 1 Smoothing-based l_∞-norm minimization

1: (Initialization). Given A and G, choose initial phases Φ^0 and $\sigma^0, \alpha^{\max}, c, \epsilon > 0, \tau \in (0,1)$ and set $i = 0$.
2: (Direction generation). $d^i = \nabla_\Phi \tilde{L}(\Phi^i, \sigma^i)$.
3: (Armijo line search). Find the largest α^i, $0 < \alpha^i < \alpha^{\max}$ that satisfies the sufficient descent condition,

$$L(\Phi^i - \alpha^i d^i, \sigma^i) \leq L(\Phi^i, \sigma^i) - \frac{\alpha^i c \sigma^i}{\tilde{L}(\Phi^i, \sigma^i)} \nabla_\Phi \tilde{L}^T(\Phi^i, \sigma^i) d^i.$$

4: (Select $\mathcal{H}$ and $\mathcal{J}$). $\mathcal{H}^i = 1, \mathcal{J}^i = \alpha^i d^i$.
5: (Φ-parameter update). $\Phi^{i+1} = \Phi^i - (\mathcal{H}^i)^{-1}(\mathcal{J}^i)$.
6: (σ-parameter update). $\sigma^{i+1} = \tau^q \sigma^i$, where $q \leftarrow 1$ if $|L(\Phi^i, \sigma^i) - L(\Phi^{i+1}, \sigma^i)| < \epsilon$ and $q \leftarrow 0$ otherwise.
7: (Termination). Stop if a stopping criterion is met, otherwise set $i \leftarrow i + 1$ and return to step 2.

stationary point is sufficiently approached. Hence, $\sigma^i \to 0$ for $i \to \infty$. Property ii) in Lemma 1 implies that a stationary point of $L(\Phi^i, \sigma^i)$ is a stationary point $l_\infty^2(\bar{y}(\Phi^i), \sigma^i)$ for $\sigma^i \to 0$, which is formalized in [22, Theorem 3.1]. ■

The convergence properties in Theorem 1 are preserved for any other parameter update in step 5 that reduces L in each iteration. This is exploited in Section IV for the selection of more advanced solvers.

A key aspect of Algorithm 1 is that a gradual refinement of the approximation reduces the sensitivity to local minima. Next result shows that certain local minima vanish when $\sigma \to \infty$.

Theorem 2: Consider the parameter set $\mathcal{M} = \{\Phi : \sum_{n=0}^{\bar{N}-1} \nabla_\Phi \bar{y}^2(\Phi, n) \neq 0\}$. Suppose that the set $\Phi^* = \left\{\Phi \in \mathcal{M} : \nabla_\Phi L(\bar{y}(\Phi), \sigma) = 0\right\}$ for $\sigma \to \infty$ is non-empty and includes local minimizers to $L(\bar{y}(\Phi), \sigma)$. Then, the set Φ^* is empty for $\sigma \to 0$.

Proof: From (9) it follows that $\lambda_n \to 1/\bar{N}, \forall n$ for $\sigma \to \infty$. Hence, by (8) the gradient becomes $\nabla_\Phi L(\bar{y}(\Phi), \sigma) = 1/\bar{N} \sum_{n=0}^{\bar{N}-1} \nabla_\Phi \bar{y}^2(\Phi, n) \neq 0$ for $\forall \Phi \in \mathcal{M}$. Thus, $L(\bar{y}, \sigma)$ has no local minima for $\Phi \in \mathcal{M}$ for $\sigma \to \infty$, since $\mathcal{M}$ is void. ■

Although global convergence cannot be guaranteed in general, experience has shown that gradual smoothing tends to lead to low signal peak amplitude values. The results of Algorithm 1 for the multisine defined in Example 1 are shown in Fig. 1.

IV. A FAST AND SCALABLE ALGORITHM

A. Efficient gradient and Hessian computation

The computation of the gradient and Hessian matrix are typically the most time-consuming operations in optimization. A direct computation of the gradient via differentiation of $\tilde{L}$ in (12) leads to the expression

$$\nabla_\Phi \tilde{L} = \left[\sum_{p=1}^{n_y} (J_{p,1}^T \varepsilon_p)^T, \ \ldots, \ \sum_{p=1}^{n_y} (J_{p,n_u}^T \varepsilon_p)^T\right]^T, \quad (15)$$

where $\varepsilon_p(n) = \exp\left(\frac{y_p^2(n)}{2c_p^2\sigma}\right)$ and $J_{p,q} \in \mathbb{R}^{N \times N_k}$ represent submatrices of the Jacobian J, with elements $[n, u]$ given by

$$[J_{p,q}]_{nu} = \frac{\partial \varepsilon_p(n)}{\partial \phi_{qu}} = -\frac{y_p(n)}{c_p \sigma} \varepsilon_p(n) b_{pqu} S_{pqu}(n), \quad (16)$$

and $S_{pqu}(n) = \sin(2\pi k_u n/N + \xi_{pqu})$.

An alternative and significantly more efficient method for computing the gradient is presented in the following result. The method exploits structural properties to formulate the gradient as a circular convolution, to benefit from the computational advantages of the DFT in circulant matrix multiplication [27].

Theorem 3: The vectors $J_{p,q}^T \varepsilon_p$ in (15) are given by

$$J_{p,q}^T \varepsilon_p = \frac{1}{c_p \sigma} \operatorname{Im}\left(\overline{\eta_{p,q}} \odot \begin{bmatrix} I_{N_k} \\ 0_{N_k} \end{bmatrix}^T \mathcal{F}^{-1}\left(\mathcal{D}_p^- + \mathcal{D}_p^+\right)\right), \quad (17)$$

where

$$\begin{aligned} \eta_{p,q} &= 1/\sqrt{2}\left[b_{pq1} e^{j\xi_{pq1}}, \ldots, b_{pqN_k} e^{j\xi_{pqN_k}}\right]^T, \\ \mathcal{D}_p^- &= \mathcal{F}(c_p^-) \odot \mathcal{F}\begin{bmatrix} \rho_p \\ 0_{N_k} \end{bmatrix}, \quad \mathcal{D}_p^+ = \mathcal{F}(c_p^+) \odot \mathcal{F}\begin{bmatrix} P\overline{\rho_p} \\ 0_{N_k} \end{bmatrix}, \end{aligned} \quad (18)$$

with P an $N_k \times N_k$ exchange matrix,

$$\begin{aligned} c_p^- &= \left[[Z_p]_0, \ \ldots, \ [Z_p]_{N_k-1}, \ 0, \ [\overline{Z_p}]_{N_k-1}, \ \ldots, \ [\overline{Z_p}]_1\right]^T, \\ c_p^+ &= \left[[Z_p]_{N_k+1}, \ \ldots, \ [Z_p]_{2N_k}, \ 0, \ [Z_p]_2, \ \ldots, \ [Z_p]_{N_k}\right]^T, \\ \rho_p &= 1/\sqrt{2} \sum_{w=1}^{n_u} \left[b_{pw1} e^{j\xi_{pw1}}, \ldots, b_{pwN_k} e^{j\xi_{pwN_k}}\right]^T, \end{aligned} \quad (19)$$

and where $[Z_p]_u$ is the DFT of $z_p = \exp(y_p^2/(c_p^2\sigma))$ at frequency line u, with $[Z_p]_{-u} = [\overline{Z_p}]_u$.

Proof: Substituting $y_p(n)$ in (16) by (5), performing the multiplication by ε_p, and using $z_p(n) = \varepsilon_p^2(n)$ gives,

$$[J_{p,q}^T \varepsilon_p]_u = -\frac{b_{pqu}}{c_p \sigma} \sum_{n=0}^{N-1} \sum_{w=1}^{n_u} \sum_{v=1}^{N_k} z_p(n) S_{pqu}(n) C_{pwv}(n) b_{pwv}$$

where $C_{pwv}(n) = \cos(2\pi k_v n/N + \xi_{pwv})$. Using that $S_{pqu} C_{pwv} = \frac{1}{2}(S_{pqu-pwv} + S_{pqu+pwv})$ with $S_{pqu\pm pwv} := \sin(2\pi(k_u \pm k_v)n/N + (\xi_{pqu} \pm \xi_{pwv}))$ and, by Euler's law, $\sin a \pm b = -\operatorname{Im}(e^{-j(a\pm b)})$, the elements of $J_{p,q}^T \varepsilon_p$ satisfy

$$\begin{aligned} [J_{p,q}^T \varepsilon_p]_u = \frac{b_{pqu}}{2c_p\sigma} \operatorname{Im} \sum_{w=1}^{n_u} \sum_{v=1}^{N_k} \Big(&[Z_p]_{u-v} e^{-j(\xi_{pqu} - \xi_{pwv})} + \\ &[Z_p]_{u+v} e^{-j(\xi_{pqu} + \xi_{pwv})} \Big) b_{pwv}. \end{aligned} \quad (20)$$

Herein, $[Z_p]_{u\pm v} = \sum_{n=0}^{N-1} z_p(n) e^{-j(2\pi(k_u \pm k_v)n/N}$ is equal to the DFT of z_p at frequency lines $(k_u \pm k_v)$. Due to the specific element-wise construction, (20) is expressed as the sum of two matrix-vector products involving a Toeplitz and a Hankel matrix, which allows writing

$$J_{p,q}^T \varepsilon_p = \frac{1}{c_p \sigma} \operatorname{Im}(\Psi_{p,q}^- + \Psi_{p,q}^+), \quad (21)$$

with

$$\Psi_{p,q}^- = \overline{\eta_{p,q}} \odot \mathcal{T}(Z_p^-) \rho_p, \ \Psi_{p,q}^+ = \overline{\eta_{p,q}} \odot \mathcal{T}(Z_p^+) P \overline{\rho_p}. \quad (22)$$

Here, $\mathcal{T}(a)$ is the Toeplitz matrix with basis vector a,

$$\mathcal{T}(a) = \begin{bmatrix} a_0 & a_{-1} & \ldots & a_{-(n-1)} \\ a_1 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{-1} \\ a_{n-1} & \ldots & a_1 & a_0 \end{bmatrix}, \ a = \begin{bmatrix} a_{-(n-1)} \\ a_{-(n-2)} \\ \vdots \\ a_{n-2} \\ a_{n-1} \end{bmatrix},$$

and the basis vectors Z_p^- and Z_p^+ are defined by

$$Z_p^- = \left[[Z_p]_{-(N_k-1)}, [Z_p]_{-(N_k-2)}, \dots, [Z_p]_{N_k-2}, [Z_p]_{N_k-1}\right]^T,$$
$$Z_p^+ = \left[[Z_p]_{(N_k-2)}, [Z_p]_{(N_k-1)}, \dots, [Z_p]_{2N_k-1}, [Z_p]_{2N_k}\right]^T.$$

By embedding the matrices $\mathcal{T}(Z_p^\pm)$ into circulant matrices $\mathcal{C}(c_p^\pm)$ with basis column vectors $c_p^\pm$ defined in (19), expression (22) is expressed as a circular convolution. Using the circular convolution theorem to expresss $\mathcal{C}(c)b = \mathcal{F}^{-1}(\mathcal{F}(c) \odot \mathcal{F}(b))$ [27], allows formulating $\Psi_{p,q}^\pm = \overline{\eta_{p,q}} \odot \left[I_{N_k} \quad 0_{N_k}\right] \mathcal{F}^{-1}(\mathcal{D}_{p,q}^\pm)$, with $\mathcal{D}_{p,q}^\pm$ and $\overline{\eta_{p,q}}$ defined in (18). Insertion into (21) and rearranging completes the proof. ■

In GN-type algorithms, the Hessian matrix $\nabla_\Phi^2 \tilde{L}$ is approximated based on the Jacobian matrix J, e.g., classical GN employs $\nabla_\Phi^2 \tilde{L} \approx J^T J$. A fast method for constructing $H = J^T J$ based on the DFT is presented next.

Theorem 4: Let J be the Jacobian matrix composed of the submatrices $J_{p,q}$ in (16) for $p = 1, \dots, n_y$ and $q = 1, \dots, n_u$. Let $H_{i,j} \in \mathbb{R}^{N_k \times N_k}$ for $i, j = 1, \dots n_u$ be the submatrices of $H = J^T J$. Then,

$$H_{i,j} = \sum_{w=1}^{n_y} \frac{1}{c_w^2 \sigma^2} \mathrm{Re}(\Gamma_{w,i,j}^- - \Gamma_{w,i,j}^+) \tag{23}$$

where $[\Gamma_{w,i,j}^\pm]_{uv} = \frac{b_{wiu} b_{wjv}}{2} e^{\sqrt{-1}(\phi_{wiu} \pm \phi_{wjv})} [Q_{w,i,j}]_{u \pm v}$. Herein, $[Q_{w,i,j}]_{u \pm v}$ is defined as the DFT of $q_{w,i,j} := \exp\left(y_w^2/(c_w^2\sigma)\right)x^2$ at frequency lines $(k_u \pm k_v)$, where $[Q_{w,i,j}]_{-u} = \overline{[Q_{w,i,j}]_u}$.

The proof follows the lines of [9, Appendix A].

B. Computational complexity

The complexity of the computation of the gradient and the Hessian is quantified in Table I for a SISO system. Herein, a comparison is performed between: direct computation from (16), the approach presented in [9] for the L_p-norm based method, and the fast approach by Theorems 3 and 4. The Fast Fourier Transform (FFT) implementation of the DFT is assumed. The complexity of the gradient and Hessian computation scales with $n_y n_u$ and $n_y n_u^2$, respectively.

Evidently, by exploiting circular convolution theory, the fast approach eliminates the quadratic complexity in N_k in the gradient computation for both the direct approach (since $NN_k > N_k^2$) and the L_p approach. For the Hessian, the cubic complexity in N_k in the direct method is eliminated in the L_p and the fast method by employing the FFT.

C. Solver selection

The performance of Algorithm 1 for three different solvers is compared: 1) GN with modified Levenberg-Marquardt regularization (GNLM) [28], 2) SD, and 3) the Polak-Ribière Conjugate Gradient (PRCG) method [29]. Steps 2-4 in Algorithm 1 are appropriately adjusted and the corresponding parameters $\mathcal{J}, \mathcal{H}$ in step 4 are given in Table II. Herein, λ is the LM parameter that ensures well-conditioned and positive definite $\mathcal{H}$ [28]. Parameter β is the PR parameter, and r is the previous step direction [29]. In the sequel, the tuning parameters $\{\sigma^0, \alpha^{\max}, c, \epsilon, \tau\} = \{1, 0.1, 10^{-4}, 10^{-4}, 0.7\}$ are used, which have been found to generally produce high-performance results.

TABLE I
COMPLEXITY OF GRADIENT AND HESSIAN COMPUTATIONS.

Method	Gradient $\nabla_\Phi \tilde{L} = J^T \varepsilon$	Hessian $J^T J$
Direct	$\mathcal{O}(NN_k)$	$\mathcal{O}(NN_k^2)$
L_p, [9]	$\mathcal{O}(N \log N + N_k^2)$	$\mathcal{O}(N \log N + N_k^2)$
Fast	$\mathcal{O}(N \log N + N_k \log N_k)$	$\mathcal{O}(N \log N + N_k^2)$

TABLE II
COMPUTATIONAL PERFORMANCE FOR DIFFERENT SOLVERS.

Solver	$\mathcal{J}$	$\mathcal{H}$	iters	time
GNLM	$\nabla_\Phi \tilde{L}$	$J^T J + \lambda \mathrm{diag}(J^T J)$	255	3.8
SD	$\alpha \nabla_\Phi \tilde{L}$	1	734	0.8
PRCG	$\alpha(\nabla_\Phi \tilde{L} + \beta r)$	1	322	0.3

Table II shows the results averaged over 100 simulations (performed on a standard laptop) with uniform spectral magnitudes, random initial phases Φ, and $N = 2 \cdot 10^5, N_k = 10^3, n_u, n_y = 1, b_{11k} = \sqrt{2/N_k}$ $\forall k$. All solvers achieve $l_\infty(\bar{y}) = 1.38$. GNLM requires the fewest iterations, yet requires the most time due to the computation of the Hessian. This motivates the use of the first-order solvers SD and PRCG. PRCG achieves the best performance by mitigating the typical chatter behavior of SD. In the next section, the performance of Alg. 1 is experimentally evaluated on a multivariable system.

V. EXPERIMENTAL VALIDATION

A. Experiment description

The presented algorithm is experimentally validated on a closed-loop controlled Active Vibration Isolation System (AVIS) with 6 inputs and 6 outputs [30]. A scalar-valued multisine signal u is applied to the first system input. The signal u has a uniform spectrum, a length of $N = 2^{15}$, and $N_k = 3000$. The constraints are imposed onto the 6 outputs, 6 input voltages, and 6 controller outputs, hence $n_y = 18$ and $\dim(G) = 18 \times 1$. The spectra of y are all non-uniform due to the closed-loop dynamics in G between u and y.

B. Results

1) Optimized signals: The output signals $y_p/c_p, p = 1, \dots, 6$ before and after optimization are shown in Fig. 2. Peak amplitude reduction up to a factor 3 is achieved and $l_\infty(y_p/c_p) \leq 1$ $\forall p$.

2) Performance comparison: The smoothing-based algorithm is compared to several existing crest-factor minimization techniques. The results are shown in Fig. 3. The random method [13] generates random realizations with $\phi_k \in [0, 2\pi)$, $\forall k$ and retains the realization with lowest l_∞-norm. This method hardly shows convergence. Schroeder phase selection [10] is fast but leads to a constraint excess of a factor > 2.5. The time-frequency domain swapping algorithm [12] is originally a SISO method, and is applied to the worst-case signal per iteration. This method fails to converge in the multivariable setting, since the optimization of a single signal y_p deteriorates the other signals. The L_p-method [9], using sequence $p = \{2^2, 2^3, \dots, 2^9\}$, converges in ∼40

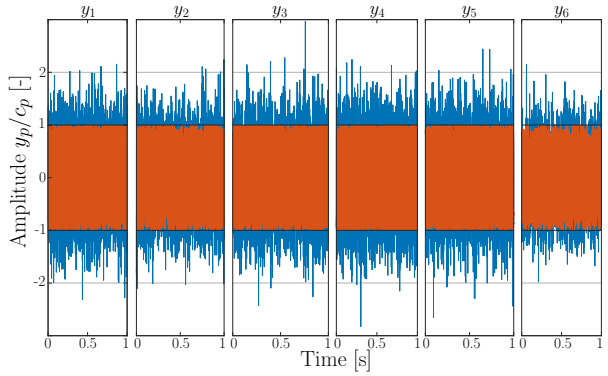

Fig. 2. Signals $\boldsymbol{y_1}, \ldots, \boldsymbol{y_6}$ before (—) and after (—) optimization using the smoothing-based approach.

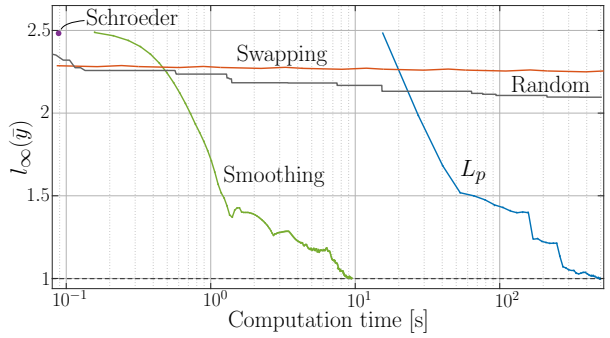

Fig. 3. Convergence of $\boldsymbol{l_\infty(\bar{y})}$ versus computation time for different methods. The smoothing-based algorithm outperforms existing methods significantly in view of computation time and achieved cost.

iterations, but the computation time of ∼500s is substantial. The smoothing method refers to Alg. 1 with PRCG solver, and gradient computation by Theorem 3. This method shows convergence in ∼150 iterations, and is ∼50 times faster than the L_p method. This demonstrates the capability of the smoothing-based algorithm for multivariable l_∞-norm constrained signal design, as well as its high computational efficiency.

VI. CONCLUSIONS

The presented method enables designing l_∞-norm constrained excitation signals to improve the quality of identified models of large MIMO systems. This is realized by an optimization-based approach that exploits a sequence of smooth approximations of the non-smooth objective function, to achieve a high level of robustness against local minima. Low computational complexity is achieved by exploiting structural properties in the gradient and Hessian matrix and results from circular convolution theory. Experimental results show a drastic improvement in both the achieved l_∞-norm and the computation time compared to existing techniques.

ACKNOWLEDGMENT

The authors gratefully thank Gijs Linskens and Rick van der Maas for their contributions.

REFERENCES

[1] L. Ljung, "System Identification: Theory for the User. 2nd ed", Prentice Hall PTR, Upper Saddle River, NJ, USA, 1999.
[2] G. Goodwin, R. Payne, "Dynamic system identification: experiment design and data analysis", Academic Press New York, 1977.
[3] M. Heertjes, H. Butler, N. Dirkx, S. van der Meulen, R. Ahlawat, K. O'Brien, J. Simonelli, K-T. Teng, Y. Zhao , "Control of Wafer Scanners: Methods and Developments", Proc. Am. Control Conf., pp. 3686-3703, 2020.
[4] D. Rivera, M. Braun, H. Mittelmann, "Constrained multisine input signals for plant-friendly identification of chemical process systems", J. of Process Control, vol.19, pp. 623-635, 2009.
[5] H. Jansson, 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.
[6] R. Hildebrand R, M. Gevers, "Identification for control: Optimal input design with respect to a worst-case ν-gap cost function", Proc IEEE Conf. Decis. Control., vol. 1, pp. 996 - 1001, 2004.
[7] S. Boyd, L. Vandenberghe, "Convex Optimization", Cambridge University Press, New York, NY, USA, 2004.
[8] R. Pintelon, J. Schoukens, "System identification: a frequency domain approach. 2nd ed", John Wiley & Sons, Inc, Hoboken, USA, 2012.
[9] P. Guillaume, J. Schoukens, R. Pintelon, I. Kollár, "Crest-factor minimization using nonlinear Chebyshev approximation methods", IEEE Trans. Instrum. Meas., vol. 40, no. 6, pp. 982 - 989, 1991.
[10] M. Schroeder, "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation, IEEE Trans. Inf. Theory., vol. 16, no. 1, pp. 85-89, 1970.
[11] A. Tan, K. Godfrey, "The generation of binary and near-binary pseudo-random signals: an overview", IEEE Trans. Instrum. Meas., vol. 51, no. 4, pp. 583-588, 2002.
[12] E. van der Ouderaa, J. Schoukens, J. Renneboog, "Peak Factor Minimization Using a Time-Frequency Domain Swapping Algorithm", IEEE Trans. Instrum. Meas., vol. IM-37, pp. 145-147, 1988.
[13] E. van der Ouderaa, J. Schoukens, J. Renneboog, "Peak factor minimization of input and output signals of linear systems", IEEE Trans. Instrum. Meas., vol. 37, no. 2, pp. 207-212, 1988.
[14] C. Yeong, S. Torquato, "Reconstructing random media", Physical Review E, vol. 57, no. 1, pp. 495-506, 1998.
[15] C. Rojas, J. Welsh, G. Goodwin, "A Receding Horizon Algorithm to Generate Binary Signals with a Prescribed Autocovariance", Proc. Am. Control Conf., pp. 122 - 127, 2007.
[16] C. Larsson, P. Hägg and H. Hjalmarsson, "Generation of excitation signals with prescribed autocorrelation for input and output constrained systems", Proc. Am. Control Conf., pp. 3918-3923 , 2013.
[17] I. Manchester, "Input design for system identification via convex relaxation", Proc. IEEE Conf. Decis. Control., pp. 2041-2046, 2010.
[18] Y. Nesterov, "Smooth minimization of non-smooth functions", Math. Programming., vol. 103, pp. 127-152, 2005.
[19] A. Beck, M. Teboulle, "Smoothing and first order methods: A unified framework" . SIAM J. Optim., vol. 22, no. 2, pp. 557–580, 2012.
[20] Y. Nesterov, V. Spokoiny, "Random gradient-free minimization of convex functions", Found. Comput. Math., vol. 17, no. 2, pp. 527–566, 2017.
[21] D.P. Bertsekas, "Approximation Procedures based on the method of multipliers", J. Optim. Theory Appl., vol. 23, pp. 487–510, 1977.
[22] S. Xu, "Smoothing method for minimax problem", Comput. Optim. Appl., vol. 20, no. 3, pp. 267-279, 2001.
[23] B. Kort, D. Bertsekas, "A new penalty function algorithm algorithm for constrained minimization", Proc. IEEE Conf. Decis. Control., New Orleans, Louisiana, 1972.
[24] P. Blanchard, D. Higham, N. Higham, "Accurately computing the log-sum-exp and softmax functions", IMA J. Numer. Anal. vol. 41. no. 4, pp. 2311-2330, 2021.
[25] G. Calafiore, S. Gaubert, C. Possieri, "Log-sum-exp neural networks and posynomial models for convex and log-log-convex data", IEEE Trans. Neural Netw. Learn. Syst., vol. 31, pp. 1–12, 2020.
[26] J. Nocedal, S.Wright, "Numerical Optimization", New York, Springer, 1999.
[27] G. Golub, C. Van Loan, "Matrix Computations", The Johns Hopkins University Press, Baltimore, third edition, 1996.
[28] R. Fletcher, "A modified Marquardt subroutine for non-linear least squares", AERE-R 6799, Harwell, 1971.
[29] L. Grippo, S. Lucidi, "A globally convergent version of the Polak-Ribière conjugate gradient method", Math. Program., vol. 78, pp. 375–391, 1997.
[30] R. Voorhoeve, A. van der Maas, T. Oomen, "Non-parametric identification of multivariable systems: A local rational modeling approach with application to a vibration isolation benchmark", Mech. Syst. Signal Process., vol. 105, pp. 129-152, 2018.