Nonsmooth Optimization for Multidisk H_∞ Synthesis

Pierre Apkarian [*] Dominikus Noll [†]

Abstract

We discuss output feedback control design with multiple performance specifications, each measured in a weighted H_∞-norm. The multidisk design problem consists in finding a stabilizing output feedback controller which minimizes the different performance specifications simultaneously using a worst case strategy. This is less conservative than existing approaches, but difficult to solve algorithmically due to inherent nonsmoothness and nonconvexity. We present a nonsmooth optimization method suited for the multidisk problem and more generally, for programs where the maximum of an infinite family of nonconvex maximum eigenvalue functions is minimized. The method is shown to perform well on a control problem for a helicopter at hover.

Keywords: H_∞-synthesis, multi-channel design, multi-objective optimization, concurring performance specifications, static output feedback, reduced-order synthesis, decentralized control, PID, NP-hard problems, nonsmooth optimization, multidisk problems.

1 Introduction

Well designed feedback control systems are expected to respond favorably to an extended set of design goals including robustness, good regulation against disturbances, desirable responses to commands, and much else. Controller design therefore often involves a tradeoff between these objectives in order to achieve a suitable compromise. In this paper we discuss a class of multi-objective design problems, known as *multidisk problems* [11], where the performance channels are all measured in a weighted H_∞-norm, and where these performances are optimized *simultaneously* using a worst case strategy. Mathematically, the multidisk design problem may be regarded as minimizing the maximum of an infinite family of nonconvex maximum eigenvalue functions. The multidisk problem is of great practical importance, but difficult to solve due to its semi-infinite min-max structure. This explains why to date only a few heuristic approaches have been presented. Since our new approach is expected to reduce conservatism in existing multi-channel strategies, we comment on those subsequently.

[*]ONERA-CERT, Centre d'études et de recherche de Toulouse, Control System Department, 2 av. Edouard Belin, 31055 Toulouse, France - and - Mathématiques pour l'Industrie et la Physique, Université Paul Sabatier, Toulouse, France - Email : `apkarian@cert.fr` - Tel : +33 5.62.25.27.84 - Fax : +33 5.62.25.27.64.

[†]Université Paul Sabatier, Institut de Mathématiques de Toulouse (IMT), 118, route de Narbonne, 31062 Toulouse, France - Email : `noll@mip.ups-tlse.fr` - Tel : +33 5.61.55.86.22 - Fax : +33 5.61.55.83.85.

One clearly conservative approach to multi-channel problems alleviates the difficulty by setting up a single performance channel, using trade-offs between the conflicting performance specifications. This may then be solved via traditional methods (LMIs and AREs) suited for single channel design.

A second more sophisticated strategy uses the Q-parametrization of all stabilizing controllers of a system. For the two-disk problem, this is elaborated in [21], the related strong stabilization problem is considered in [26, 8]. Unfortunately, these approaches use the Youla parametrization [27], which leads to undesirably large feedback controllers, and moreover, makes it impossible to add structural constraints, also known as control law specifications, on the controller. We refer the reader to the work of Scherer [23] for an analysis of structured design problems for a specific class of plants, and for a Q-parametrization approach to multi-objective control problems. A thorough mathematical study of multidisk problems using tools from analytic function theory is presented by Helton in [11].

There exist yet another class of heuristic techniques for multidisk and multi-objective synthesis problems, which uses state-space LMI formulations. However, these techniques rely on sufficient conditions, and are in general extremely conservative [18]. The reader is referred to [24, 16, 10] to list just a few of these approaches, and to [4] for an extension to Linear Parameter-Varying systems.

Here we attack the multi-objective design problem directly using techniques from nonsmooth optimization. Such a strategy has already been proposed in [13], where the authors use the Q-parametrization, which allows them to treat the multi-disk problem via convex analysis. In order to avoid the inconveniences of the Q-parametrization, we follow a different and more flexible line which allows us in particular to add structural constraints on the controller. The price to be paid for this extension is that the optimization program is nonconvex and nonsmooth, so that computed solutions are only locally optimal. Experiments nonetheless show that the advantage of our local strategy is considerable. Note that a similar nonsmooth and nonconvex formulation of H_∞ synthesis is also investigated in [15]. There, authors propose a global search strategy based on a dynamical systems approach to determine solutions.

Our approach leads to optimization of composite functions of the form

$$f(K) = \max_{i=1,\dots,N} \|T_{w^i \to z^i}(K)\|_\infty, \tag{1}$$

where $T_{w^i \to z^i}(K)$ are performance specifications used to probe the closed-loop system. Each $T_{w^i \to z^i}$ is a smooth operator defined on the open domain $\mathcal{D} \subset \mathbb{R}^{(m_2+k)\times(p_2+k)}$ of stabilizing feedback controllers K, with values in the infinite dimensional space RH_∞ of rational stable transfer matrix functions. In consequence, the composite functions $\|\cdot\|_\infty \circ T_{w^i \to z^i}$ are neither smooth nor convex, but their structure can be exploited algorithmically. One central contribution of this work is a spectral bundle algorithm suited for local optimization of functions of the form f, and more generally, of semi-infinite nonconvex maximum eigenvalue functions with a related structure.

Notice that the max-H_∞-function $f(K)$ may be written as

$$f(K) = \max_{i=1,\dots,N} \|T_{w^i \to z^i}(K)\|_\infty = \max_{\omega\in\mathbb{R}} \overline{\sigma}\left(T(K,j\omega)\right) = \max_{\omega\in\mathbb{R}} \lambda_1\left(T(K,j\omega)T(K,j\omega)^H\right)^{1/2} \tag{2}$$

where $\overline{\sigma}(M)$ and $\lambda_1(MM^H)$ denote the maximum singular value respectively the maximum eigenvalue, and where $T(K,j\omega)$ has a block structure with N blocks regrouping the different performance channels. In particular, $T(K,j\omega)T(K,j\omega)^H$ is block diagonal with N blocks, so that f

may be interpreted as an infinite maximum of maximum eigenvalue functions. This structure will be exploited for the nonsmooth analysis of f.

The structure of the paper is as follows. In Section 4 we present a convergence result for our method and discuss practical aspects. A related approach is developed in [1] and [20]. In Section 2, we introduce the general setting of the multidisk H_∞ synthesis problem and discuss a few instances of practical interest. Tools and ingredients from nonsmooth analysis that are used in algorithmic constructions are introduced in Section 3. Nonsmooth descent techniques are developed in Section 4 and are illustrated in Section 5 for a helicopter control problem.

Notation

Let $\mathbb{R}^{n\times m}$ be the space of $n\times m$ matrices, equipped with the corresponding scalar product $\langle X, Y\rangle = \mathrm{Tr}(X^TY)$, where X^T is the transpose of the matrix X, $\mathrm{Tr}\, X$ its trace. For complex matrices, X^H denotes the transconjugate. For Hermitian or symmetric matrices, $X \succ Y$ means that $X - Y$ is positive definite, $X \succeq Y$ that $X - Y$ is positive semi-definite. We use the symbol λ_1 to denote the maximum eigenvalue of a symmetric or Hermitian matrix and $\overline{\sigma}$ to denote the maximum singular value of a general matrix. The Frobenius norm of a matrix M is denoted $\|M\|_F$ and is defined by $\|M\|_F = \sqrt{\mathrm{Tr}\,(M^HM)}$. The symbol $\otimes$ designates the usual Kronecker product for matrices. We shall use notion from nonsmooth analysis covered by [9]. In particular, for a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, $\partial f(x)$ denotes its Clarke subdifferential or generalized gradient at x, $f'(x; d)$ the Clarke directional derivative. The convex hull of vectors $v_1, \dots, v_q$ is denoted $\mathrm{co}\,\{v_1, \dots, v_q\}$.

2 Multidisk H_∞ synthesis

Consider a finite family of plants $P^i(s)$, $i = 1, \dots, N$, described in state-space form as

$$P^i(s): \quad \begin{bmatrix} \dot{x}^i \\ z^i \\ y^i \end{bmatrix} = \begin{bmatrix} A^i & B_1^i & B_2^i \\ C_1^i & D_{11}^i & D_{12}^i \\ C_2^i & D_{21}^i & D_{22}^i \end{bmatrix} \begin{bmatrix} x^i \\ w^i \\ u^i \end{bmatrix}, i = 1, \dots, N, \tag{3}$$

where $x^i \in \mathbb{R}^{n^i}$ is the state vector, $u^i \in \mathbb{R}^{m_2}$ the vector of control inputs, $w^i \in \mathbb{R}^{m_1^i}$ the vector of exogenous inputs, $y^i \in \mathbb{R}^{p_2}$ the vector of measurements and $z^i \in \mathbb{R}^{p_1^i}$ the controlled or performance vector of the ith plant. We require that all models in the family (3) have the *same* number of control and measurement signals, while their dynamics and performances need not be the same. This allows to embed a variety of problems within the same framework. For simplicity, it is assumed throughout that $D_{22}^i = 0$.

The general multidisk synthesis problem consists in designing a dynamic output feedback controller $u^i = K(s)y^i$ for the plant family in (3) with the following properties:

- **Internal stability:** For $w^i = 0$ the state vector x^i of the closed-loop systems (3) and the control $u^i = K(s)y^i$ tend to zero as time goes to infinity.

- **Performance:** Among all simultaneously stabilizing controllers, K minimizes the worst case performance function $f(K) = \max_{i=1,\dots,N} \|T_{w^i\to z^i}(K)\|_\infty$.

We assume that the controller K has the following frequency domain representation:

$$K(s) = C_K(sI - A_K)^{-1}B_K + D_K, \qquad A_K \in \mathbb{R}^{k\times k}, \tag{4}$$

where k is the order of the controller, and where the case $k = 0$ of a static controller $K(s) = D_K$ is included.

Often practical considerations impose additional structural constraints on the controller K. For instance it may be desired to design low-order controllers ($0 \leq k \ll n$) or controllers with prescribed-pattern, sparse controllers, decentralized controllers, observed-based controllers, PID control structures, synthesis on a finite set of transfer functions, and much else. Formally, the synthesis problem may then be represented as

$$\begin{array}{ll} \text{minimize} & f(K) = \max\limits_{i=1,\dots,N} \|T_{w^i\to z^i}(K)\|_\infty \\ \text{subject to} & K \text{ stabilizes all } P^i \text{ in (3)} \\ & K \in \mathcal{K} \end{array} \tag{5}$$

where $K \in \mathcal{K}$ represents a structural constraint on the controller (4). In most cases, this takes the more amenable form of an equality constraint $g(K) = 0$. The formulation in (3) and (5) is fairly general, so in order to motivate the present work, we conclude this section by considering some examples of particular interest.

Example 1. *Standard multichannel problems.* Here the (A^i, B_2^i, C_2^i) originate from the same system and only differ by the introduction of weighting functions that have been used channel-wise to translate design specifications. Other data are allowed to vary. The helicopter control problem discussed in the experimental section will be of this form. ■

Example 2. *Simultaneous design.* Here all data including (A^i, B_2^i, C_2^i) are arbitrary. This could for instance include cases where for safety reasons the nominal behavior of a plant as well as several failure scenarios have to be controlled simultaneously. ■

Example 3. *Design with stable controller.* At least one plant in the family, say $j = N$, is realized by a particular static system as follows:

$$P^N(s) := \begin{bmatrix} D_{11}^N & D_{12}^N \\ D_{21}^N & 0 \end{bmatrix} := \begin{bmatrix} 0 & I_{m_2} \\ I_{p_2} & 0 \end{bmatrix}.$$

This leads to $T_{w^N\to z^N}(K, s) = K(s)$. The search for an approximate local solution to design problems with stable controller can then be cast as

$$\underset{K(s)}{\text{minimize}} \quad \max\left(\max_{i=1,\dots,N-1} \|T_{w^i\to z^i}(K)\|_\infty,\ \rho\,\|T_{w^N\to z^N}(K)\|_\infty\right) \tag{6}$$

where ρ is a small parameter. The role of channel N is now to keep the controller away from becoming unstable during the iterations.

This also suggests a number of variants where explicit constraints on the controller are added. For example, practical considerations often prescribe controllers with small gains, a problem that naturally falls into the cast (6). In the same vein, approximate structural constraints can be assigned to the dynamic controller $K(s)$. Specifically, if a gain penalization on entry $K_{i,j}(s)$ is desired, it suffices to generate a plant in the family in the form:

$$P^j(s) := \begin{bmatrix} D_{11}^j & D_{12}^j \\ D_{21}^j & 0 \end{bmatrix} := \begin{bmatrix} 0 & e_i^T \\ e_j & 0 \end{bmatrix},$$

where e_i and e_j are suitable unit vectors. ■

Example 4. *Pure stabilization problems.* Our approach needs initial stabilizing controllers, $K \in \mathcal{D}$, since the different H_∞-norm terms must be well-defined. The initialization problem is therefore of particular interest. It can be handled by solving a special multidisk synthesis problem, by selecting appropriate data in (3). The reader is referred to [2] for more details.

The problem is then how to maintain the stability constraint $K \in \mathcal{D}$ in (5) during the optimization of K. This may be achieved indirectly as in (6), that is, by adding a term $\rho\|T_{w^N \to z^N}(K)\|_\infty$ with small $\rho > 0$, which in tandem with the performance specifications $i = 1, \ldots, N-1$ guarantees simultaneous closed-loop stability for all plants. Notice that $K \in \mathcal{D}$ is not a constraint in the usual sense of constrained optimization, because $\mathcal{D}$ is an open set, and because the objective function is not defined outside $\mathcal{D}$. This explains why an indirect approach to this constraint is taken. ■

3 Subdifferential of the max-H_∞ mapping

The success of our algorithmic constructions hinges on the following central fact.

Proposition 3.1 *Every closed-loop H_∞-mapping of the form $\|\cdot\|_\infty \circ T_{w\to z}$ defined on the set of asymptotically stabilizing controllers $K \in \mathbb{R}^{(m_2+k)\times(p_2+k)}$ is regular in the sense of Clarke [9]. Consequently, the same is true for the function $f(K) = \max_{i=1,\ldots,N} \|T_{w^i\to z^i}(K)\|_\infty$ defined on the open set $\mathcal{D}$ of simultaneously asymptotically stabilizing controllers K.*

Proof: Note that $\|\cdot\|_\infty$ is continuous and convex on the space $\mathcal{H}^\infty$ of stable transfer functions. Since $T_{w\to z}$ is differentiable on the open subset $\mathcal{D} \subset \mathbb{R}^{(m_2+k)\times(p_2+k)}$, it follows from [9] that the composite function $\|\cdot\|_\infty \circ T_{w\to z}$ is regular. Finally, since f is a finite maximum of such functions, the same result is true for f. ■

This result has important consequences for control problems involving H_∞ performances, because calculus rules for generalized gradients simplify. As we shall see, it allows us to compute the Clarke subdifferential of the max-H_∞ map $f = \max \circ (\|\cdot\|_\infty \circ T_{w^1\to z^1}, \ldots, \|\cdot\|_\infty \circ T_{w^N\to z^N})$.

In the sequel, the focus is on static feedback controllers, $k = 0$. Dynamic output feedback design is easily converted into static output feedback design through prior dynamic augmentation of the plant:

$$\begin{array}{llll} K \to \begin{bmatrix} A_K & B_K \\ C_K & D_K \end{bmatrix}, & A \to \begin{bmatrix} A & 0 \\ 0 & 0_k \end{bmatrix} & B_1 \to \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, & C_1 \to [\, C_1 \quad 0 \,] \\ B_2 \to \begin{bmatrix} 0 & B_2 \\ I_k & 0 \end{bmatrix}, & C_2 \to \begin{bmatrix} 0 & I_k \\ C_2 & 0 \end{bmatrix}, & D_{12} \to [\, 0 \quad D_{12} \,], & D_{21} \to \begin{bmatrix} 0 \\ D_{21} \end{bmatrix}. \end{array} \tag{7}$$

We refer the reader to [2] for further details. For static controllers, we will need the following notations:

$$
\begin{array}{c}
\mathcal{A}^i(K) := A^i + B_2^i K C_2^i,\ \mathcal{B}^i(K) := B_1^i + B_2^i K D_{21}^i,\ \mathcal{C}^i(K) := C_1^i + D_{12}^i K C_2^i, \\
\mathcal{D}^i(K) := D_{11}^i + D_{12}^i K D_{21}^i\,,
\end{array}
\tag{8}
$$

for closed-loop data. The computation of generalized subgradients greatly simplifies if we introduce the following definitions:

$$
\begin{bmatrix} T_{w^i\to z^i}(K,s) & G_{12}^i(K,s) \\ G_{21}^i(K,s) & \star \end{bmatrix} \quad := \quad \begin{bmatrix} \mathcal{C}^i(K) \\ C_2^i \end{bmatrix} (sI - \mathcal{A}^i(K))^{-1} \left[\, \mathcal{B}^i(K) \quad B_2^i \,\right] + \begin{bmatrix} \mathcal{D}^i(K) & D_{12}^i \\ D_{21}^i & \star \end{bmatrix}.
$$

We are almost ready to characterize the Clarke subdifferential of the composite function $f(K) = \max_{i=1,\dots,N} \|T_{w^i\to z^i}(K)\|_\infty$. Let us introduce two more notations. We let

$$
I(K) = \{i \in \{1,\dots,N\} : \|T_{w^i\to z^i}(K)\|_\infty = f(K)\},
\tag{9}
$$

the set of active indices at a given K. Moreover, for each $i \in I(K)$, we consider the set of active frequencies

$$
\Omega^i(K) = \{\omega \in \mathbb{R} : \overline{\sigma}\left(T_{w^i\to z^i}(K,j\omega)\right) = f(K)\}.
$$

We assume throughout that $\Omega^i(K)$ is a finite set, indexed as

$$
\Omega^i(K) = \{\omega_\nu^i : \nu = 1,\dots,p^i\},\ i \in I(K).
\tag{10}
$$

The set of all active frequencies is denotes as $\Omega(K)$.

Theorem 3.2 *Assume that the controller K is static, $k = 0$, and stabilizes the plant family in (3), that is, $K \in \mathcal{D}$. With the notations introduced in (9) and (10), let Q_ν^i be a matrix whose columns form an orthonormal basis of the eigenspace of $T_{w^i\to z^i}(K,j\omega_\nu^i)T_{w^i\to z^i}(K,j\omega_\nu^i)^H$ associated with the largest eigenvalue $\lambda_1\left(T_{w^i\to z^i}(K,j\omega_\nu^i)T_{w^i\to z^i}(K,j\omega_\nu^i)^H\right) = \overline{\sigma}(T_{w^i\to z^i}(K,j\omega_\nu^i))^2$. Then, the Clarke subdifferential of the mapping f at $K \in \mathcal{D}$ is the compact and convex set $\partial f(K) = \{\Phi_Y : Y \in \mathcal{S}(K)\}$, where*

$$
\Phi_Y = f(K)^{-1} \sum_{i\in I(K)} \sum_{\nu=1,\dots,p^i} \mathrm{Re}\left\{G_{21}^i(K,j\omega_\nu^i)\, T_{w^i\to z^i}(K,j\omega_\nu^i)^H Q_\nu^i Y_\nu^i (Q_\nu^i)^H G_{12}^i(K,j\omega_\nu^i)\right\}^T\,,
\tag{11}
$$

and

$$
\mathcal{S}(K) = \{Y = (Y_\nu^i)_{i\in I(K),\nu=1,\dots,p^i} : Y_\nu^i = (Y_\nu^i)^H \succeq 0,\ \sum_{i\in I(K)} \sum_{\nu=1,\dots,p^i} \mathrm{Tr}\, Y_\nu^i = 1\}.
\tag{12}
$$

Proof: Let $G \in \mathcal{H}^\infty$ be a nonzero stable transfer function. Suppose its H_∞-norm is attained at the finite set of frequencies $\omega_1,\dots,\omega_p$, possibly including ∞. Then the subgradients of $\|\cdot\|_\infty$ at G are linear functionals on $\mathcal{H}^\infty$ of the form

$$
\phi_Y(H) = \|G\|_\infty^{-1} \sum_{\nu=1}^{p} \mathrm{Re}\, G(j\omega_\nu)^H Q_\nu Y_\nu Q_\nu^H H(j\omega_\nu),
$$

where the columns of the matrix Q_ν are an orthonormal basis of the eigenspace of $G(j\omega_\nu)G(j\omega_\nu)^H$ associated with its largest eigenvalue $\|G\|_\infty^2$, and where $Y_\nu \succeq 0$, $\sum_{\nu=1}^{p} \mathrm{Tr}(Y_\nu) = 1$.

Next consider a composite function like $\|\cdot\|_\infty \circ T_{w\to z}$. By Proposition 3.1, this functions is regular, and the Clarke subdifferential is therefore obtained using the chain rule. In other words, $\partial\left(\|\cdot\|_\infty \circ T_{w\to z}\right)(K) = T'_{w\to z}(K)^* \partial\|\cdot\|_\infty \left(T_{w\to z}(K)\right)$. After computing the adjoint of the derivative of $T_{w\to z}$ at K, mapping the dual of $\mathcal{H}^\infty$ into $\mathbb{R}^{(m_2+k)\times(p_2+k)}$, we find that the subgradients of $\|\cdot\|_\infty \circ T_{w\to z}$ at K are precisely of the form

$$\Phi_Y = \|T_{w\to z}(K)\|_\infty^{-1} \sum_{\nu=1}^{p} \mathrm{Re}\left\{G_{21}(K, j\omega_\nu) T_{w\to z}(K, j\omega_\nu)^H Q_\nu Y_\nu Q_\nu^H G_{12}(K, j\omega_\nu)\right\}^T,$$

where $Y_\nu \succeq 0$, $\sum_{\nu=1}^{p} \mathrm{Tr}(Y_\nu) = 1$. Notice that $\Phi_Y \in \mathbb{R}^{(m_2+k)\times(p_2+k)}$ now acts on vectors K of that space via the standard scalar product $\langle K, \Phi_Y\rangle = \mathrm{Tr}(K^T \Phi_Y)$.

Next, according to Clarke [9, Proposition 2.3.12], the subdifferential of the finite maximum $f(K) = \max_{i\in I(K)} \|T_{w^i\to z^i}(K)\|_\infty$ is obtained through

$$\partial f(K) = \mathrm{co}\left\{\partial(\|.\|_\infty \circ T_{w^i\to z^i})(K) : i \in I(K)\right\}.$$

Combining with the above this leads to the set of subgradients

$$\Phi_Y = f(K)^{-1} \sum_{i\in I(K)} \tau_i \sum_{\nu=1,\dots,p^i} \mathrm{Re}\left\{G_{21}^i(K, j\omega_\nu^i)\, T_{w^i\to z^i}(K, j\omega_\nu^i)^H Q_\nu^i Y_\nu^i (Q_\nu^i)^H G_{12}^i(K, j\omega_\nu^i)\right\}^T, \quad (13)$$

where $\sum_{i\in I(K)} \tau_i = 1$, $\tau_i \geq 0$ and $\sum_{\nu=1,\dots,p^i} \mathrm{Tr}\, Y_\nu^i = 1$, $Y_\nu^i \succeq 0$.

Note finally that (13) is equivalent to

$$\Phi_Y = f(K)^{-1} \sum_{i\in I(K)} \sum_{\nu=1,\dots,p^i} \mathrm{Re}\left\{G_{21}^i(K, j\omega_\nu^i)\, T_{w^i\to z^i}(K, j\omega_\nu^i)^H Q_\nu^i\, \tau_i Y_\nu^i\, (Q_\nu^i)^H G_{12}^i(K, j\omega_\nu^i)\right\}^T. \quad (14)$$

From the definitions of the τ_i and the Y_ν^i, we have

$$\mathrm{Tr} \sum_{i\in I(K)} \sum_{\nu=1,\dots,p^i} \tau_i Y_\nu^i = \sum_{i\in I(K)} \tau_i \mathrm{Tr}\,\Big(\sum_{\nu=1,\dots,p^i} Y_\nu^i\Big) = \sum_{i\in I(K)} \tau_i = 1.$$

Therefore, redefining $\tau_i Y_\nu^i$ as Y_ν^i, immediately gives (11) where the Y are as in (12). ■

Remark. Under the assumption that the set $\Omega(K)$ of active frequencies is finite, a subset of the set $\partial\|\cdot\|_\infty(K)$ of all subgradients of the H_∞-norm has first been presented in [6]. The full characterization of $\partial\|\cdot\|_\infty(K)$, is given in [3, 2], and the formula for a composite function $\partial f(K)$ with $N = 1$ follows by computing the adjoint $T'_{w\to z}(K)^*$. The extension to general finite N above uses the max formula and is therefore a straightforward extension of the case $N = 1$.

Remark. The subdifferential in Theorem 3.2 is generated through the convex set of matrices (12). When singular values at peak points are simple, i.e., each $\overline{\sigma}\left(T_{w^i\to z^i}(K, j\omega_\nu)\right)$ with $i \in I(K)$, $\omega_\nu \in \Omega^i(K)$ has multiplicity 1, this description reduces to the set of subgradients

$$\Phi_Y = f(K)^{-1} \sum_{i\in I(K)} \sum_{\nu=1,\dots,p^i} \lambda_\nu^i\, \mathrm{Re}\left\{G_{21}(K, j\omega_\nu^i)\, T_{w^i\to z^i}(K, j\omega_\nu^i)^H q_\nu^i (q_\nu^i)^H G_{12}(K, j\omega_\nu^i)\right\}^T, \quad (15)$$

where $\lambda_\nu^i \geq 0$, $\sum_{i\in I(K)}\sum_{\nu=1,\dots,p^i}\lambda_\nu^i = 1$, and where q_ν^i is the normalized eigenvector associated with the maximum eigenvalue of the matrix $T_{w^i\to z^i}(K, j\omega_\nu^i)T_{w^i\to z^i}(K, j\omega_\nu^i)^H$. Whenever possible, this simplification is beneficial in our descent algorithms for solving multidisk H_∞ problems.

Example 5. *Strongly structured controllers.* We combine Theorem 3.2 with a chain rule to obtain the subdifferential of the max-H_∞ operator for specially structured controllers, including PID, lead-lag, fixed-pattern, decentralized, observer-based, companion form, and much else. This can be formalized as follows.

Assume, an affine parametrization of the state-space data of the controller is known:

$$K = K_0 + L\,\mathrm{diag}(\kappa)R$$

where κ is a vector of free parameters to be designed. Then, using chain rules for subdifferentials of regular functions [9], the subdifferential $\partial g(\kappa)$ of the max-H_∞ function $g(\kappa) := f\,(K_0 + L\,\mathrm{diag}(\kappa)R)$ is obtained as the set of subgradients

$$\Psi_Y = \mathrm{diag}(L^T\Phi_Y R^T)\,, \tag{16}$$

where Φ_Y is described in (11). In the notation above, we have used the convention that the diag operation applied to a vector κ generates a diagonal matrix with κ on the main diagonal, whereas the same operation applied to a matrix vectorizes the main diagonal. ■

Example 6. *Decentralized PID controllers.* We specialize to the case of decentralized PID controllers, a structure often used in industrial applications. This class of structured controllers is well-defined for $m\times m$ square plants and can be described in the form:

$$K(s) = \mathrm{diag}\left(K_P^1 + \frac{K_I^1}{s} + \frac{sK_D^1}{1+\tau^1 s}, \dots, K_P^m + \frac{K_I^m}{s} + \frac{sK_D^m}{1+\tau^m s}\right).$$

Introduce parameter vectors

$$\begin{aligned}
\kappa_\tau &:= (\tfrac{1}{\tau_1}, \dots, \tfrac{1}{\tau_m}),\\
\kappa_D &:= (-\tfrac{K_D^1}{\tau_1}, \dots, -\tfrac{K_D^m}{\tau_m}),\\
\kappa_I &:= (K_I^1, \dots, K_I^m),\\
\kappa_P &:= (K_P^1 + \tfrac{K_D^1}{\tau_1}, \dots, K_P^m + \tfrac{K_D^m}{\tau_m}),
\end{aligned}$$

then an affine state-space parametrization of decentralized PID controllers can be written as:

$$\begin{bmatrix} A_K & B_K \\ C_K & D_K \end{bmatrix} := \begin{bmatrix} I_m \otimes 0_{2,2} & I_m \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ I_m \otimes 0_{2,1} & I_m \otimes 0 \end{bmatrix} + L \quad \mathrm{diag}(\kappa_\tau, \kappa_D, \kappa_I, \kappa_P) \quad R\,,$$

where

$$L := \begin{bmatrix} I_m \otimes \begin{bmatrix} 0 \\ -1 \end{bmatrix} & I_m \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} & 0 & 0 \\ I_m & 0 & I_m & I_m \end{bmatrix}, \quad R := \begin{bmatrix} I_m \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} & 0 & I_m \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} & 0 \\ 0 & I_m & 0 & I_m \end{bmatrix}^T, \tag{17}$$

and where $\otimes$ denotes the standard Kronecker product of matrices. Then $\partial g(\kappa)$ is obtained by combining (16) and (17). ■

4 Nonsmooth descent techniques

With complete knowledge of the generalized subdifferential, several nonsmooth techniques can be derived for the determination of local solutions to multidisk H_∞ design problems. Different variants along with their convergence study are discussed in [20, 2]. In the sequel, we briefly discuss two strategies. The second is supported by a sound convergence theory and has been used in our numerical experiments of Section 5.

4.1 Steepest descent

A straightforward idea to compute locally optimal solutions to multidisk synthesis is the steepest descent algorithm. Following Clarke [9], and leaving apart structural constraints on the controller, K is critical for problem $(\mathcal{P})$ if and only $0 \in \partial f(K)$, where $\partial f(K)$ is described in (11). It therefore appears natural to consider the program

$$H_K := -\frac{\Phi_Y}{\|\Phi_Y\|_F}, \qquad \Phi_Y := \operatorname{argmin}\{\|\Phi_Y\|_F : Y \in \mathcal{S}(K)\},$$

which either shows that $0 \in \partial f(K)$ or produces the direction H_K of steepest descent at K when $0 \notin \partial f(K)$. Clearly, the nature of the set $\mathcal{S}(K)$ in (12) entails that direction H_K can be computed via standard SDP codes or even using classical convex QP when the subdifferential reduces to (15). In a steepest descent algorithm the descent direction computation is followed by a line search and the algorithm loops over these 2 steps until $0 \in \partial f(K)$ is approximately satisfied. A central weakness of this technique is that it may fail to converge to a critical point due to the nonsmoothness of $f(K)$. The reason of the failure is that the search direction H_K at K does not depend continuously on K thus iterates may undergo zigzagging. A more elaborate technique must therefore be formulated. This has been done in a general context in [2]. In the next section, we briefly discuss an attractive variant for which convergence to local solutions can be established. The reader is referred to [20, 2] for a thorough convergence study.

4.2 Convergent nonsmooth descent method

Given a static controller $K \in \mathcal{D}$, where $\mathcal{D} \subset \mathbb{R}^{m_2 \times p_2}$ is the set of simultaneously stabilizing controllers of (3), introduce the functions

$$f(K,\omega) := \max_{i=1,\dots,N} \overline{\sigma}\left(T_{w^i \to z^i}(K, j\omega)\right),$$

then $f(K) = \max_{\omega \in \mathbb{R}} f(K,\omega)$, and minimization of f may be interpreted as a semi-infinite minimization problem involving the infinite family $f(\cdot,\omega)$. At a given K, recall that $\Omega(K)$ is the set of active frequencies at K. Clearly, $f(K,\omega) \le f(K)$ for all $\omega \in \mathbb{R}$ and $f(K,\omega) = f(K)$ for $\omega \in \Omega(K)$. As a consequence of Theorem 3.2, the subdifferential of the function $f(K,\omega)$ at K is defined as the set of subgradients

$$\Phi_{Y,\omega} := f(K,\omega)^{-1} \sum_{i \in I_\omega(K)} \operatorname{Re}\left\{G_{21}^i(K,j\omega)\, T_{w^i \to z^i}(K,j\omega)^H Q_\omega^i Y_\omega^i (Q_\omega^i)^H G_{12}^i(K,j\omega)\right\}^T,$$

where

$$I_\omega(K) = \{i \in \{1,\dots,N\} : \overline{\sigma}(T_{w^i \to z^i}(K,j\omega)) = f(K,\omega)\}$$

is the index set of active models at frequency ω. The columns of the matrix Q_ω^i form an orthonormal basis of the eigenspace of $T_{w^i\to z^i}(K,j\omega)T_{w^i\to z^i}(K,j\omega)^H$ associated with its largest eigenvalue, and where

$$\sum_{i\in I_\omega(K)} \operatorname{Tr} Y_\omega^i = 1\,, \quad Y_\omega^i = (Y_\omega^i)^H \succeq 0\,.$$

During the following, we consider finite extensions $\Omega_e(K)$ of the set of active frequencies $\Omega(K)$. For any such set $\Omega_e(K)$, and for some fixed $\delta > 0$, we introduce the optimality function

$$\theta_e(K) := \inf_{H\in R^{m_2\times p_2}} \sup_{\omega\in\Omega_e(K)} \sup_{\sum\limits_{i\in I_\omega(K)} \operatorname{Tr} Y_\omega^i=1,\, Y_\omega^i\succeq 0} -f(K) + f(K,\omega) + \langle \Phi_{Y,\omega}, H\rangle + \frac{1}{2}\delta\|H\|_F^2\,. \tag{18}$$

When $\Omega_e(K) = \Omega(K)$, we use the notation $\theta(K)$. Since $\Omega(K) \subset \Omega_e(K)$ for any extension, we have $\theta(K) \le \theta_e(K)$.

As we shall see, the name optimality function for $\theta(K)$ and $\theta_e(K)$ is justified by the fact that $\theta_e(K) \le 0$ for all K, while $\theta_e(K) = 0$ implies that K is a critical point of f. Note that this type of optimality measure has first been introduced by E. Polak [22] for finite and infinite families of smooth functions.

Proposition 4.1 *The following dual formula is valid:*

$$\theta_e(K) = \sup_{\sum_{\omega\in\Omega_e(K)}\tau_\omega=1,\,\tau_\omega\ge 0} \sup_{\sum_{i\in I_\omega(K)}\operatorname{Tr} Y_\omega^i=1,\, Y_\omega^i\succeq 0} \sum_{\omega\in\Omega_e(K)} \tau_\omega(f(K,\omega) - f(K)) - \frac{1}{2\delta}\|\sum_{\omega\in\Omega_e(K)} \tau_\omega\Phi_{Y,\omega}\|_F^2\,. \tag{19}$$

Proof. To begin with, we convert the first supremum in (18) into a supremum over the simplex $\sum_{\omega\in\Omega_e(K)}\tau_\omega = 1$, $\tau_\omega \ge 0$. This operation leaves the function θ_e unchanged. We then have that

$$\begin{aligned}\theta_e(K) = \inf_{H\in R^{m_2\times p_2}} \sup_{\sum_{\omega\in\Omega_e(K)}\tau_\omega=1,\,\tau_\omega\ge 0} \sup_{\sum\limits_{i\in I_\omega(K)}\operatorname{Tr} Y_\omega^i=1,\, Y_\omega^i\succeq 0} \{-f(K)+ \\ \textstyle\sum_{\omega\in\Omega_e(K)}\tau_\omega f(K,\omega) \qquad + \sum_{\omega\in\Omega_e(K)}\tau_\omega\langle\Phi_{Y,\omega},H\rangle + \frac{\delta}{2}\|H\|_F^2\ \}\ .\end{aligned}$$

Next, invoking Fenchel duality, it is possible to interchange the inner double supremum with the outer infimum. The now inner infimum with respect to H becomes unconstrained and can be computed explicitly. We obtain the solution

$$H(K) := -\frac{1}{\delta}\sum_{\omega\in\Omega_e(K)}\tau_\omega\Phi_{Y,\omega}\,. \tag{20}$$

Substituting this expression back into (18) yields the desired dual expression for $\theta_e(K)$ given in (19). ∎

Since $f(K,\omega) \le f(K)$ for all ω, we infer that $\theta_e(K) \le 0$. Using the dual formula (19), one can see that equality $\theta_e(K) = 0$ can only occur when $\tau_\omega = 0$ for all $\omega \in \Omega_e(K)\setminus\Omega(K)$. Hence the condition $\theta_e(K) = 0$ is equivalent to

$$0 = \sup_{\sum\limits_{i\in I_\omega(K)}\operatorname{Tr} Y_\omega^i=1,\, Y_\omega^i\succeq 0} -\frac{1}{2\delta}\|\sum_{\omega\in\Omega(K)}\tau_\omega\Phi_{Y,\omega}\|_F^2 \tag{21}$$

independently of the extension $\Omega_e(K)$ of $\Omega(K)$ which was used. Note that in this expression $I_\omega(K) = I(K)$ since $f(K,\omega) = f(K)$ for $\omega \in \Omega(K)$. By the definition of $\Phi_{Y,\omega}$, equality (21) is equivalent to

$$\sum_{\omega\in\Omega(K)} \tau_\omega f(K)^{-1} \sum_{i\in I(K)} \mathrm{Re}\left\{G_{21}^i(K,j\omega)\, T_{w^i\to z^i}(K,j\omega)^H Q_\omega^i Y_\omega^i (Q_\omega^i)^H G_{12}^i(K,j\omega)\right\}^T = 0\,,$$

which by Theorem 3.2 is nothing else but the condition $0 \in \partial f(K)$. To sum up our reasoning, we have the following

Theorem 4.2 *A controller K is critical for the minimization of $f(K)$ if and only if $\theta_e(K) = 0$, where $\theta_e(K)$ may be computed with respect to any finite extension $\Omega_e(K)$ of the set of active frequencies $\Omega(K)$. Whenever $\theta_e(K) < 0$, direction (20), where τ_ω, Y_ω^i are solutions to the dual program (19), is a descent direction of $f(K)$ at K.* ■

Before we present our nonsmooth algorithm, we need to state another useful property of the descent direction. We have the following

Lemma 4.3 *$H(K)$ in (20) is a qualified descent direction in the sense that*

$$f'(K;H(K)) \leq \theta_e(K) - \frac{\delta}{2}\|H(K)\|^2 < 0, \tag{22}$$

where $f'(K;H)$ denotes the Clarke directional derivative of f at K in direction H.

Proof. By the definition of $H(K)$ we have

$$\theta_e(K) = \sup_{\omega\in\Omega_e(K)} \sup_{\mathrm{Tr}\sum_{i\in I_\omega(K)} Y_\omega^i=1, Y_\omega^i\succeq 0} -f(K) + f(K,\omega) + \langle \Phi_{Y,\omega}, H(K)\rangle + \frac{1}{2}\delta\|H(K)\|_F^2.$$

Since $\Omega(K) \subset \Omega_e(K)$, restricting the first supremum to $\omega \in \Omega(K)$ yields

$$\begin{aligned}
\theta_e(K) - \frac{1}{2}\delta\|H(K)\|_F^2 &\geq \sup_{\omega\in\Omega(K)} \sup_{\mathrm{Tr}\sum_{i\in I_\omega(K)} Y_\omega^i=1, Y_\omega^i\succeq 0} \langle \Phi_{Y,\omega}, H(K)\rangle \\
&= \sup\{\langle \Phi_Y, H(K)\rangle : \Phi_Y \in \partial f(K)\} \\
&= f'(K;H(K))\,.
\end{aligned}$$

Here we use the fact that for $\omega \in \Omega(K)$, the terms $-f(K) + f(K,\omega)$ vanish. ■

The above analysis suggests the following nonsmooth descent algorithm for the minimization of $f(K)$, where $0 < \alpha < 1$, $0 < \beta < 1$ and $\delta > 0$ are fixed parameters.

1. **Initialization**. Find a controller K which stabilizes the plants $P^1, \dots, P^N$ simultaneously.
2. **Generate frequencies**. Given the current K, compute $f(K)$ and obtain active frequencies $\Omega(K)$. Select a finite enriched set of frequencies $\Omega_e(K)$ containing $\Omega(K)$, as outlined in Section 4.4.
3. **Descent direction.** Compute $\theta_e(K)$ and the solution (τ, Y) of SDP (19). If $\theta_e(K) = 0$, stop, because $0 \in \partial f(K)$. Otherwise compute descent direction $H(K)$ given in (20).
4. **Line search.** Find largest $t = \beta^k$ such that $f(K + t\,H(K)) \le f(K) + t\alpha\theta(K)$ and such that $K + tH(K)$ remains stabilizing.
5. **Step.** Replace K by $K + tH(K)$, increase iteration counter by one, and go back to step 2.

Remark. Notice here that the line search in step 4 is successful since by Lemma 4.3,

$$\lim_{t\to 0} \frac{1}{t} \left(f\left(K + tH(K)\right) - f(K)\right) = f'\left(K; H(K)\right) \le \theta_e(K) - \frac{1}{2}\delta\|H(K)\|_F^2 < \theta_e(K) < 0.$$

Since $0 < \alpha < 1$, $\theta_e(K) < \alpha\theta_e(K)$, and the set of admissible t in the line search of step 4 contains a nonempty open interval $(0, \bar{t})$. Locating the largest $t = \beta^k \in (0, \bar{t})$ is therefore a finite procedure.

4.3 Convergence

The question which remains to be resolved is how to choose the frequency set $\Omega_e(K)$ in step 2 of the algorithm in order to achieve convergence of the method. A crucial observation is that the optimality function $\theta(K)$ may fail to behave continuously, as the sequence of iterates K_n approaches a limit point K^*. This is due to the fact that typically $\Omega(K^*)$ will contain frequencies which are not limit points of sequences of frequencies in $\Omega(K_n)$. This in turn is just another way to express the fact that the steepest descent direction behaves discontinuously.

In order to force continuity of the optimality function, one has to use extended sets $\Omega_e(K_n)$. The most general approach to assure convergence is to use a sequence of meshes $\Omega_h \subset \mathbb{R}$ with decreasing mesh size $h > 0$, in order to approximate the infinite maximum. Assume that in step 2 of the algorithm, $\Omega_e(K) = \Omega(K) \cup \Omega_{1/n}$, where n is the actual value of the iteration counter. This simply ensures that as $K_n \to K^*$, the set $\Omega(K^*)$ is contained in the set of accumulation points of the sequence $\Omega_e(K_n) \supset \Omega_{1/n}$.

Unfortunately, this is not a practically useful procedure, as we get subsets $\Omega_e(K)$ of increasingly large size. Our experiments show that, on the contrary, it is possible to limit the size of the sets $\Omega_e(K)$ throughout the process. One naturally wonders whether this may be justified theoretically. We show that this is indeed possible if some mild extra assumptions on the limit point K^* are made.

Let us assume that K^* is a limit point of the sequence K_n of iterates generated by the algorithm. Assume as before that Ω_ν^i is finite for every $i \in I(K^*)$, so that the set of active frequencies

$\Omega(K^*) = \{\omega_1^*, \dots, \omega_p^*\}$ is finite. Moreover, assume for the time being that the maximum eigenvalues $\lambda_1\left(T_{w^i\to z^i}(K^*, j\omega)T_{w^i\to z^i}(K^*, j\omega)^H\right)$ at $\omega = \omega_\nu^*$ have all multiplicity 1, so that the functions $f(K, \omega_\nu)$ are smooth in a neighborhood of (K^*, ω_ν^*). Now assume that

$$(\mathcal{H}) \qquad f_\omega(K^*, \omega_\nu^*) \neq 0, \text{ and } f''(K^*, \omega_\nu^*) \prec 0 \text{ for } \nu = 1, \dots, p.$$

Then by the implicit function theorem there exist p functions $\omega_\nu(K)$, $\nu = 1, \dots, p$, of class C^1 defined in a neighborhood of K^* such that $\omega_\nu(K^*) = \omega_\nu^*$ and $f'\left(K, \omega_\nu(K)\right) = 0$. Since $f''(K^*, \omega_\nu^*) \prec 0$ for $\nu = 1, \dots, p$, and since the $\omega_1^*, \dots, \omega_p^*$ are the only frequencies where the maximum is attained, it follows that in a neighborhood of K^*, the problem of minimizing f is equivalent to the program

$$\min_{K\in\mathcal{D}} \max_{\nu=1,\dots,p} f\left(K, \omega_\nu(K)\right).$$

In particular, $\Omega(K) \subset \{\omega_1(K), \dots, \omega_p(K)\}$ in a neighborhood of K^*, even though the inclusion may be strict at $K \neq K^*$. We refer to those local maxima $\omega_\nu(K)$ of $f(K, \cdot)$ for which $f\left(K, \omega_\nu(K)\right) < f(K)$ as secondary peaks, while those in $\Omega(K)$ are called peaks. Due to hypotheses $(\mathcal{H})$, in a neighborhood of K^*, there can be at most $p-1$ secondary peaks, because the only local maxima are the $\omega_\nu(K)$, and at least one of them must be peak.

Now for a finite maximum of smooth functions, $F_i(K) = f\left(K, \omega_i(K)\right)$, the optimality function (18) behaves continuously if defined as $\Omega_e(K) = \{\omega_1(K), \dots, \omega_p(K)\} \supseteq \Omega(K)$ in a neighborhood of K^*. In conclusion, we have the following

Theorem 4.4 *Let K^* be an accumulation point of the sequence K_n generated by the nonsmooth algorithm. Suppose that hypothesis $(\mathcal{H})$ is satisfied at K^*. Suppose that $\theta_e(K)$ is computed using the p primary and secondary peaks, that is, $\Omega_e(K) = \{\omega_1(K), \dots, \omega_p(K)\}$. Then K^* is a critical point of f.*

Proof. As θ_e now varies continuously with K, so does the descent direction $H(K)$ in (20). Suppose then we had $\theta(K^*) < 0$. Then a descent step $H(K^*)$ away from K^* is possible, and descent is at least $f\left(K^* + t(K^*)H(K^*)\right) \leq f(K^*) + \alpha t(K^*)\theta_e(K^*)$, as shown by the line search procedure and Lemma 4.3. Here $t(K)$ is the step size function defined by the line search in step 4 of the algorithm. We claim that the function $K \to t(K)\theta_e(K)$ is semicontinuous at K^* in the weak sense that there exists a neighborhood N^* of K^* such that $t(K)\theta_e(K) \leq \rho t(K^*)\theta(K^*)$ for all $K \in N^*$ and some fixed $0 < \rho < 1$. Accepting that this is the case, we argue as follows. Step 4 of the algorithm tells us that for $K_n \in N^*$, the value $f(K_{n+1})$ is below $f(K_n)$ by a gap of at least $\alpha t(K_n)\theta_e(K_n) \leq \rho\alpha t(K^*)\theta(K^*) < 0$. Since this gap occurs infinitely often, f must be unbounded below, a contradiction. But notice that semicontinuity of $t(K)\theta_e(K)$ in the sense indicated is clear from the continuity of θ_e and from the definition of the stepsize rule. ■

4.4 Extended sets of frequencies

Multidisk H_∞ synthesis requires repeated computations of H_∞ norms. This is done quite efficiently using the bisection algorithm [5, 7]. As a byproduct, this algorithm returns estimates of the primary and secondary peak frequencies $\omega \in \{\omega_1(K), \dots, \omega_p(K)\}$. In our numerical experiments, we have observed that it is generally beneficial to consider an extended set of frequencies including the primary and secondary peak set $\Omega_e(K) \supseteq \Omega(K)$. This renders the algorithm less dependent on the accuracy to which peak frequencies are computed, and the resulting descent direction (20) often behaves more smoothly. Moreover, using a larger $\Omega_e(K)$ allows us to capture more information on the frequency curve $\omega \mapsto \max_{i=1,\ldots,N} \overline{\sigma}(T_{w^i \to z^i}(K, j\omega))$ and thus better steps are performed. In our numerical testing, we have used the following simple scheme:

Selection of additional frequencies

1. Compute $f(K)$ via the bisection algorithm and detect $\Omega(K)$.
2. Guess the number p of active peaks in the limit K^* and add the $p - |\Omega(K)|$ largest secondary peaks.
3. Define a cut-off level $\gamma_c := \beta\, f(K)$ where $\beta \in (0,\, 1)$.
4. Determine nearly active models using γ_c. Model with index i is retained for frequency gridding whenever $\|T_{w^i \to z^i}(K)\|_\infty > \gamma_c$.
5. For each nearly active model i, grid those frequency intervals where $\overline{\sigma}(T_{w^i \to z^i}(K, j\omega)) > \gamma_c$. Keep track of the p largest (primary and secondary) peaks to assure that $\Omega_e(K)$ contains $\{\omega_1(K), \ldots, \omega_p(K)\}$.

Note that secondary peaks in step 2 as well as the intervals in step 5 can be located via the bisection algorithm in [5]. The accuracy of primary peaks is very high, while secondary peaks are usually obtained with slightly less precision. This does not present a serious problem in practice, as the precision increases as soon as the $\omega_i(K)$ tend to become active.

Both linearly or logarithmically spaced gridding may be used and we are not strict yet on what is a better choice in a given application. Typical values of β are $\beta = 0.8;\, 0.9$. In practice, if $\Omega_e(K)$ happens to be too large a set, we truncate to retain the first 300 frequencies with leading singular values. With this simple rule, the computational effort in generating descent directions is kept under control and hardly exceeds a second in most applications.

To conclude this section, let us argue why the proposed method of selecting $\Omega_e(K)$ does not put the convergence result Theorem 4.4 at stake, as long as the choice of the gridding in the $> \gamma_c$ zone is continuous.

Theorem 4.5 *Let the sequence K_n be generated by the nonsmooth algorithm, and suppose K^* is one of its accumulation points for which the hypotheses of Theorem 4.4 are satisfied. Suppose the criticality measure $\theta_e(K)$ is computed on the basis of an extended set of frequencies $\Omega_e(K)$ obtained via the above construction. Then K^* is a critical point of f.*

Proof. Consider a convergent subsequence $K_n \to K^*$, $n \in \mathcal{N}$. Since the number of models i is finite, we may select a subsequence $n \in \mathcal{N}'$ such that the same i are nearly active in $\mathcal{N}'$. Since γ_c depends continuously on K, the region of nearly active frequencies ω with $\overline{\sigma}\,(T_{w^i \to z^i}(K, j\omega)) > \gamma_c$ also depends continuously on K. Since the gridding operator in step 5 depends continuously on the set $\{\omega_1(K), \ldots, \omega_p(K)\}$, we see that $\Omega_e(K)$ also depends continuously on K. Then the stepsize $t(K)$ has the same properties as in Theorem 4.4, that is, $t(K)\theta_e(K)$ is semi continuous in the weak sense of Theorem 4.4, and the convergence argument remains essentially the same. ■

Guessing the correct number of peaks p at K^* may appear difficult in practice. However, the outlined procedure is stable in the sense that if we overestimate p, the result remains correct. An upper bound for p is also known, $p \leq \dim \operatorname{vec}(K) + 1$, which may be proved rigorously using Helly type theorems (see e.g. Hettich and Kortanek [12] and the references given there). A typical behavior of our selection strategy is depicted in Figure 2. These figures describe the evolution of the extended set $\Omega_e(K)$ in a run of our nonsmooth technique for example AC8 from [14]. Note that primary and secondary peaks remain in the extended set over the last 4 iterations.

To conclude this section, notice that one may go one step further in the analysis and dispense with the hypothesis that maximum eigenvalues have multiplicity one at peak frequencies if a suitable hypothesis replacing $(\mathcal{H})$ is made. We do not go into details here, as in our experiments, eigenvalues seem to have multiplicity one as a rule.

5 Applications

In this section, our nonsmooth method is used to design a variety of controllers for a helicopter at hover. The model is presented in [25], where a comparative study of classical design techniques is presented. Our tests use a slightly simplified model borrowed from [19]. The helicopter model is a 5th-order system. Control inputs are the main rotor collective pitch A_m (rad) and the tail rotor collective pitch A_t (rad). Performance outputs are the vertical position z (m) and the yaw angle ψ (rad). In this application, vertical and yaw axis dynamics are strongly coupled and unstable. One must therefore stabilize the helicopter dynamics and achieve good decoupling between the axes. Also, acceptable settling times must be obtained in response to step inputs. The synthesis interconnection is a standard $S/KS/T$ structure , see Figure 1.

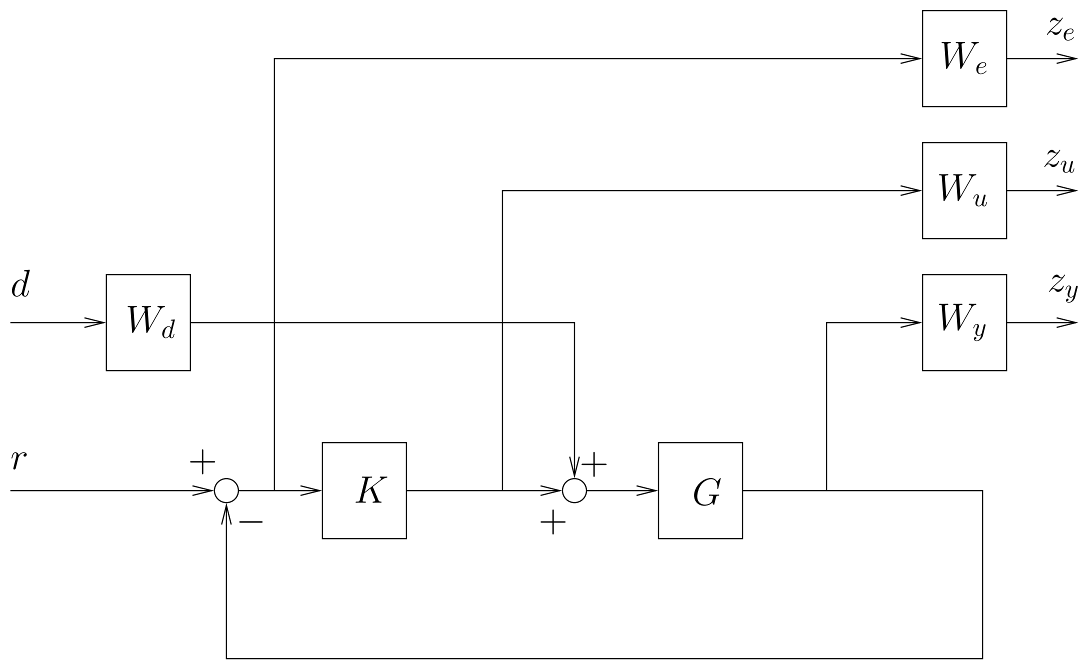

Figure 1: synthesis interconnection

The synthesis interconnection is complemented with an extra channel $T_{d\to z_y}$ to prevent inversion of the helicopter dynamics by the controller. Signal d is an input disturbance. r is a reference input defined as $r := [z_{\text{ref}}, \psi_{\text{ref}}]$. The outputs z_e, z_u and z_y suitably weighted correspond to reference tracking, attenuation of high frequency gains and robustness to neglected high frequency dynamics, respectively.

The authors in [19] use SQP and the Nelder-Mead moving polytope method to compute reduced-order and PID controllers. A difficulty with these techniques is that they do not offer any convergence certificate in the presence of nonsmoothness, a feature inherent to H_∞ performance. The Nelder-Mead method may even fail to converge in the smooth case. Illustrating examples of smooth and convex functions of two variables are discussed by McKinnon [17]. Not surprisingly, for a highly nonsmooth problem like the present one, the authors report failure of the SQP method, whereas the Nelder-Mead method is still capable of making progress as it only employs function values. However, solutions computed with such a strategy do not give any local convergence certificate.

Subsequently we attack the helicopter problem using a multidisk, more precisely multichannel, H_∞ strategy:

$$\underset{K(s)}{\text{minimize}}\; f(K) := \max\left\{\|T_{d\to z_y}(K)\|_\infty, \|T_{r\to z_e}(K)\|_\infty, \|T_{r\to z_u}(K)\|_\infty, \|T_{r\to z_y}(K)\|_\infty, 1e\text{–}7\,\|K\|_\infty\right\}. \tag{23}$$

Note that the last channel in the definition of $f(K)$ ensures stability of the controller. See our discussion in Section 2. The weighting filters in Figure 1 are the following [19]:

$$W_d := 1e\text{–}4\,I, \quad W_e := \frac{600}{100s+1}I, \quad W_u := 1e\text{–}4\,I, \quad W_y := \frac{0.1s}{1e\text{–}4s+1}\,.$$

Five different designs have been performed:

- Design of a full-order controller for the standard H_∞ problem through conventional DGKF techniques
$$\underset{K(s)}{\text{minimize}}\, \|T_{w\to z}\|_\infty$$
where $w := [\,d, r\,]^T$, $z := [\,z_e, z_u, z_y\,]^T$. This controller is of order 9.

- Design of a 4th-order controller through balanced truncation of the full-order controller. This is a stable and stabilizing controller and can therefore be used to initialize our nonsmooth technique.

- Direct design of a stable 4th-order controller for the multichannel H_∞ synthesis problem in (23) via our nonsmooth algorithm.

- Design of a decentralized PID controller for the multichannel H_∞ synthesis problem in (23) via our nonsmooth algorithm.

- Design of a decentralized PID controller for the multichannel H_∞ synthesis problem in (23), with prior pre-decoupling of the input matrix. Again, our nonsmooth algorithm is used.

Time-domain simulations for each design are displayed in Figures 3-7. Plots on the left hand column show the responses to a unit step in the vertical position z, while those on the right correspond to a unit step in the yaw angle ψ. The full-order H_∞ controller is of order 9 and achieves perfect decoupling of the vertical and yaw axes. Responses significantly deteriorate when the controller is reduced to 4th order via balanced truncation, Figure 4. A 20% coupling appears on a z-step demand while the responses in ψ shows unacceptable overshoot. In a next phase, we have used the 4th order controller to initialize our nonsmooth algorithm to compute a local solution to (23). The associated simulations are shown in Figure 5. Responses are now satisfactory both in terms of coupling and overshoot.

Finally, we have attempted to compute decentralized PID controllers for the same problem (23) using our nonsmooth algorithm. Since the controller now possesses pure integral action, we have dispensed with the stability constraint on K in $f(K)$. In view of the simulations in Figure 6, enforcing a pure PID structure seems a severe restriction in this application. Substantial coupling appears in response to z steps and responses show unacceptable overshoot. We have tried to improve these results by using a pre-decoupling H of the input matrix $B \leftarrow BH$ as is done classically in helicopter applications. The new PID controller, however, cannot be considered a valid solution as compared to the stable 4th order controller of Figure 5. If we insist on a decentralized PID structure, it appears difficult to satisfy all performance requirements by feedback alone. Open-loop compensation or feed-forward action might be necessary to achieve a desired input-output behavior.

For completeness, we provide values of the max-H_∞ objective in (23) associated with each controller in Table 1. Notice that the full-order H_∞ controller has a slightly worse max-H_∞ objective since channels are not separated in traditional H_∞ synthesis. Put differently, the 4th order controller computed by our method is even slightly better than the full-order controller obtained via a single channel with weights. This shows that the multi-channel setting, if combined with our nonsmooth method, allows to address the different specifications much more favorably. Irrelevant cross channels, such as $T_{d\to z_e}$ in this application, generally hinder proper minimization of meaningful specifications.

$K(s)$	full	truncated	order 4 nonsmooth	PID	PID pre-decoupling
$m(K)$	0.78	1.53	0.77	2.05	1.58

TABLE 1: max-H_∞ objectives

6 Conclusion

We have proposed a new algorithm to minimize the maximum H_∞ norm of a finite family of closed-loop channels, with the option to include structural constraints (control law specifications) on the controller dynamics. Our method is a first-order nonsmooth descent technique, which exploits the structure of the Clarke subdifferential of composite functions of the H_∞-norm. Its success hinges on the possibility to compute the H_∞-norm very efficiently using the bisection algorithm of [5]. The case of strongly structured controllers such as PID controllers, has been investigated, and formulas for the nonsmooth objective functions have been computed. A number of implementation details of the method have been discussed. Our approach is theoretically justified by a convergence result which shows convergence of the iterates towards a local minimum from an arbitrary starting point.

Our method is tested and shown to perform well on a multichannel helicopter control problem. This is in agreement with previous tests for single channel H_∞ synthesis in [2], where sizeable problems with up to 130 states have been solved.

References

[1] P. Apkarian and D. Noll, *Nonsmooth H_∞ Synthesis*, in 7th WSEAS Internatonal Conference on Automatic Control, Modelling and Simulation, Prague, 2005.

[2] ——, *Nonsmooth H_∞ Synthesis*, submitted, (2005).

[3] P. Apkarian, D. Noll, and D. Alazard, *Controller Design via Nonsmooth Multi-Directional Search*, in IFAC Conf. on System Structure and Control, Oaxaca, Mexico, Dec. 2004.

[4] P. Apkarian, P. Pellanda, and H. D. Tuan, *Mixed H_2/H_∞ Multi-Channel Linear Parameter-Varying Control in Discrete Time*, Syst. Control Letters, (2000), pp. 333–346.

[5] S. Boyd, V. Balakrishnan, and P. Kabamba, *A bisection method for computing the $\mathbf{H}_\infty$ norm of a transfer matrix and related problems*, Mathematics of Control, Signals, and Systems, 2 (1989), pp. 207–219.

[6] S. Boyd and C. Barratt, *Linear Controller Design: Limits of Performance*, Prentice-Hall, 1991.

[7] N. A. Bruinsma and M. Steinbuch, *A fast algorithm to compute the $\mathbf{H}_\infty$-norm of a transfer function matrix*, Syst. Control Letters, 14 (1990), pp. 287–293.

[8] U. Campos-Delgado and K. Zhou, *A Parametric Optimization Approach to H_∞ and H_2 Strong Stabilization*, Automatica, 39 (2003), pp. 1205–1211.

[9] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Canadian Math. Soc. Series, John Wiley & Sons, New York, 1983.

[10] M. C. de Oliveira, J. C. Geromel, and J. Bernussou, *An LMI Optimization Approach to Multiobjective Controller Design for Discrete-Time Systems*, in Proc. IEEE Conf. on Decision and Control, Phoenix,AZ, 1999, pp. 3611–3616.

[11] H. Dym, J. W. Helton, and O. Merino, *Multidisk problems in H_∞ optimization: A method for analyzing numerical algorithms*, Indiana Univ. Math. J., 51 (2002), pp. 1111–1160.

[12] R. Hettich and K. O. Kortanek, *Semi-infinite programming: theory, methods, and applications*, SIAM Review, 35 (1993), pp. 380–429.

[13] Z. HU, S. E. Salcudean, and P. D. Loewen, *Multiple Objective Control Problems via Nonsmooth Analysis*, in IFAC World Congress, San Fransisco, 1996, pp. 416–420.

[14] F. Leibfritz, *COMPL$_e$IB, COnstraint Matrix-optimization Problem LIbrary - a collection of test examples for nonlinear semidefinite programs, control system design and related problems*, tech. rep., Universität Trier, 2003.

[15] M. Mammadov and R. Orsi, *H_∞ Synthesis via a Nonsmooth, Nonconvex Optimization Approach*, 2005. submitted.

[16] I. Masubuchi, A. Ohara, and N. Suda, *LMI-based controller synthesis: a unified formulation and solution*, Int. J. Robust and Nonlinear Control, 8 (1998), pp. 669–686.

[17] K. I. M. McKinnon, *Convergence of the Nelder-Mead simplex method to a non-stationary point*, SIAM J. on Optimization, 9 (1998), pp. 148–158.

[18] O. Merino, H. Dym, and J. W. Helton, *Algorithms for solving multidisk problems in H_∞ optimization*, in Proc. IEEE Conf. on Decision and Control, 1999.

[19] H. E. Musch and M. Steiner, *Tuning advanced PID controllers via direct Hinf-norm minimization*, in European Control Conference, Brussels, Belgium, July 1997.

[20] D. Noll and P. Apkarian, *Spectral Bundle Methods for Non-Convex Maximum Eigenvalue Functions. Part 1: First-Order Methods*, to appear in Maths. Programming Series B, (2005).

[21] S. Norman and S. Boyd, *Numerical solution of a two-disk problem*, in Proc. American Control Conf., 1989, pp. 1745–1747. Reprinted in *Recent Advances in Robust Control*, pages 285-287, edited by P. Dorato and R. K. Yedavalli, IEEE Press.

[22] E. Polak, *Optimization : Algorithms and Consistent Approximations*, Applied Mathematical Sciences, 1997.

[23] C. Scherer, *Structured finite-dimensional controller design by convex optimization*, Linear Algebra and Appl., (2002), pp. 639–669.

[24] C. Scherer, P. Gahinet, and M. Chilali, *Multi-Objective Output-Feedback Control via LMI Optimization*, IEEE Trans. Aut. Control, 42 (1997), pp. 896–911.

[25] M. F. Weilenmann and H. P. Geering, *A Test Bench for Rotorcraft Hover Control*, in AIAA Guid., Nav. and Control Conf., Monterey, CA, August 1993, pp. 1371–1382.

[26] M. Zeren and H. Ozbay, *On the Strong Stabilization and Stable H_∞ Controller Design Problems for MIMO Systems*, Automatica, (2000), pp. 1674–1684.

[27] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*, Prentice Hall, 1996.

Acknowledgements

Thanks to Dr. Martin Weilenmann, EMPA IC-Engines lab, Switzerland and Dr. Urs Christen, Ford Forschungszentrum, Germany, for providing models and details on the helicopter example.

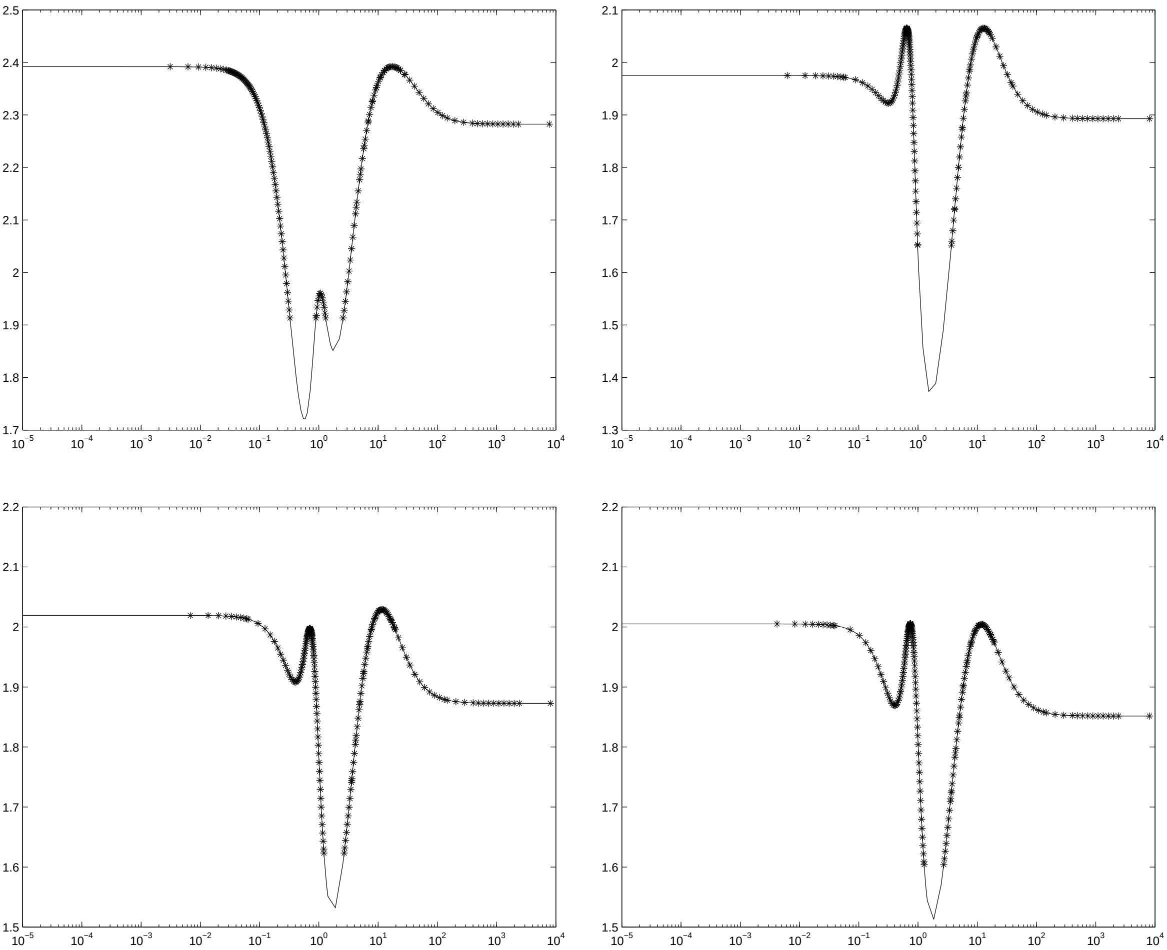

FIGURE 2: Extended set over last 4 iterations
solid line: max singular value
'*' frequencies in extended set

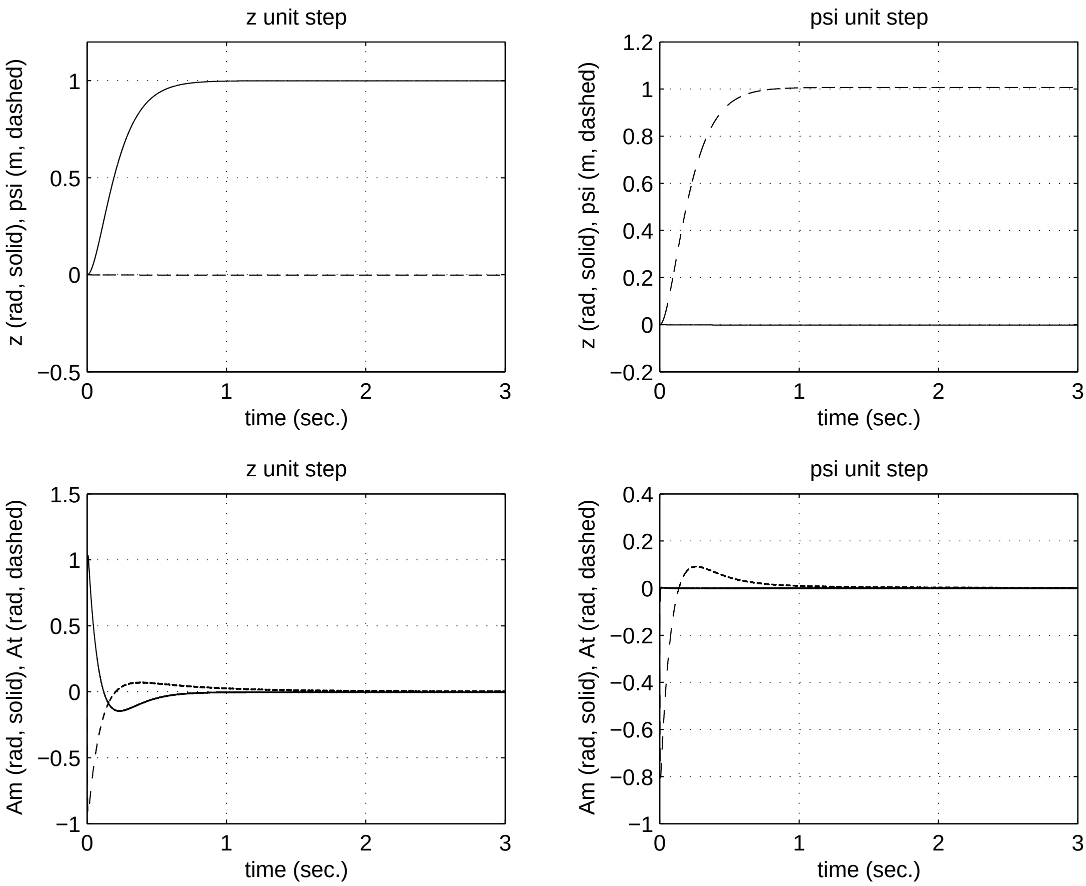

FIGURE 3: Full-order H_∞ controller
single channel

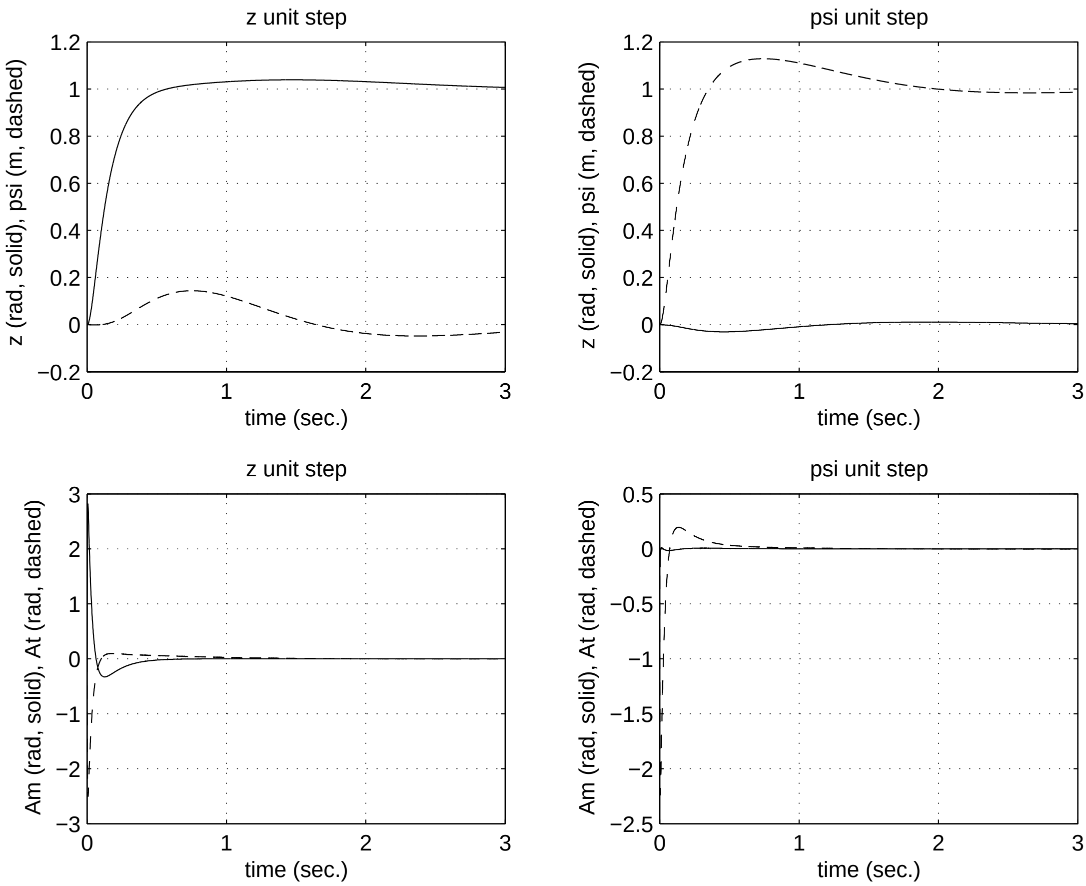

FIGURE 4: 4th reduced-order H_∞ controller
single channel

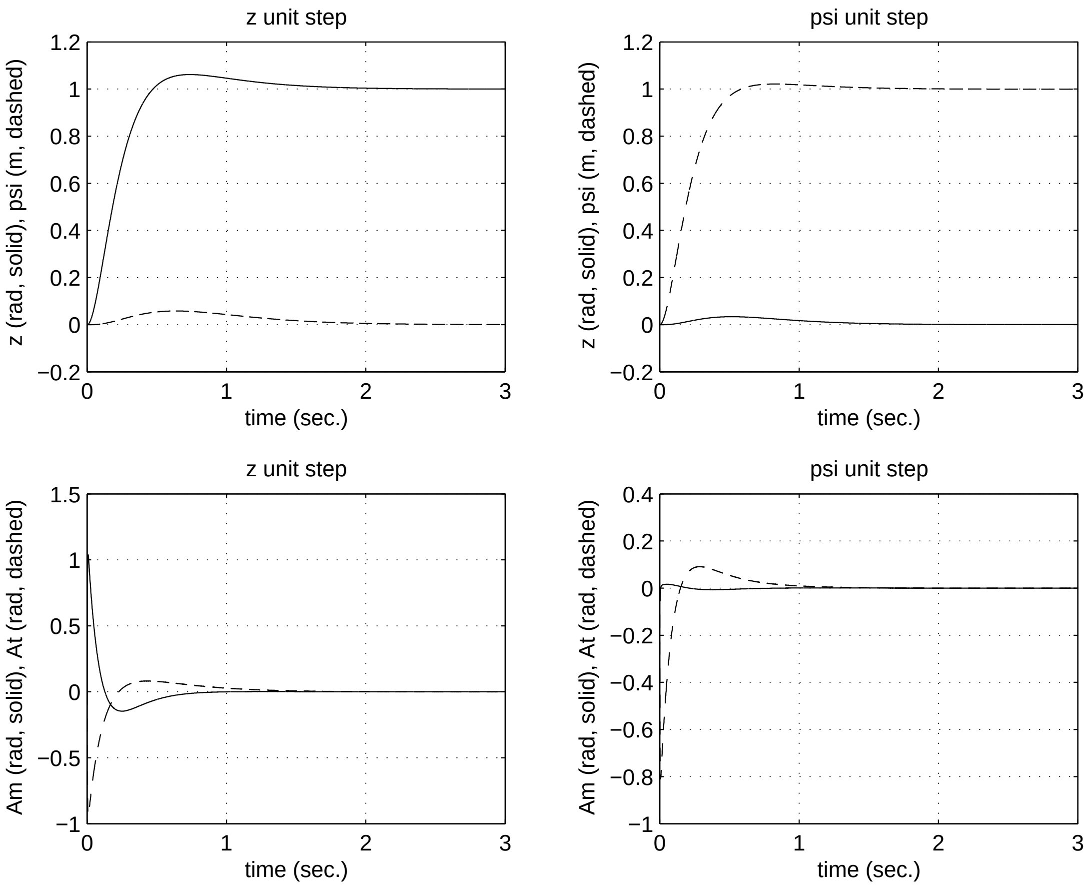

FIGURE 5: 4th-order controller from nonsmooth technique
multiple channels

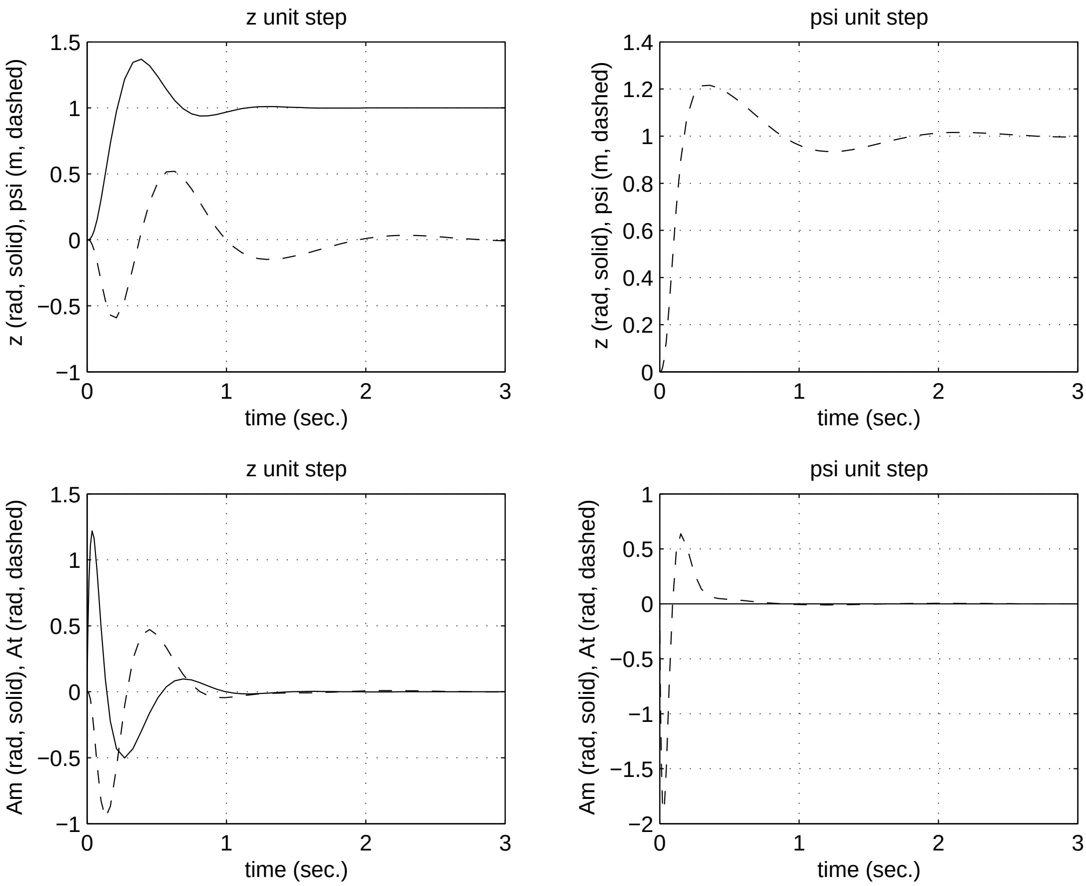

Figure 6: 4th-order PID controller from nonsmooth technique multiple channels

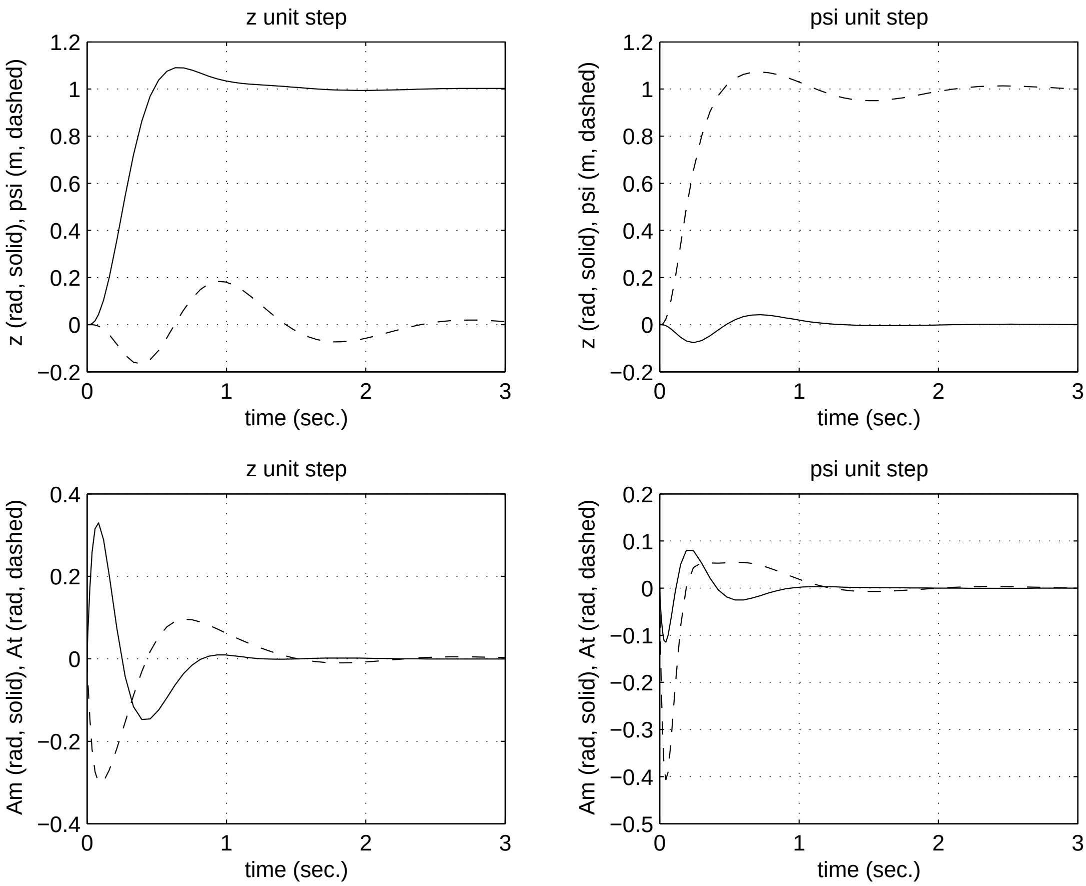

Figure 7: 4th-order PID controller from nonsmooth technique with pre-decoupling and multiple channels