SIAM J. MATRIX ANAL. APPL.
Vol. 38, No. 4, pp. 1496–1516

LARGE-SCALE COMPUTATION OF $\mathcal{L}_\infty$-NORMS BY A GREEDY SUBSPACE METHOD[*]

NICAT ALIYEV[†], PETER BENNER[‡], EMRE MENGI[†], PAUL SCHWERDTNER[§], AND MATTHIAS VOIGT[†]

Abstract. We are concerned with the computation of the $\mathcal{L}_\infty$-norm for an $\mathcal{L}_\infty$-function of the form $H(s) = C(s)D(s)^{-1}B(s)$, where the middle factor is the inverse of a meromorphic matrix-valued function, and $C(s)$, $B(s)$ are meromorphic functions mapping to short-and-fat and tall-and-skinny matrices, respectively. For instance, transfer functions of descriptor systems and delay systems fall into this family. We focus on the case where the middle factor is large scale. We propose a subspace projection method to obtain approximations of the function H where the middle factor is of much smaller dimension. The $\mathcal{L}_\infty$-norms are computed for the resulting reduced functions, then the subspaces are refined by means of the optimal points on the imaginary axis where the $\mathcal{L}_\infty$-norm of the reduced function is attained. The subspace method is designed so that certain Hermite interpolation properties hold between the largest singular values of the original and reduced functions. This leads to a locally superlinearly convergent algorithm with respect to the subspace dimension, which we prove and illustrate on various numerical examples.

Key words. $\mathcal{L}_\infty$-norm, large scale, projection, singular values, Hermite interpolation, descriptor systems, delay systems, model order reduction, greedy search, reduced basis

AMS subject classifications. 34K17, 65D05, 65F15, 90C06, 90C26, 93D03

DOI. 10.1137/16M1086200

1. Introduction.

We consider the computation of the $\mathcal{L}_\infty$-norm of a matrix-valued function of the form

$$H : \Omega \to \mathbb{C}^{p\times m}, \quad H(s) := C(s)D(s)^{-1}B(s), \tag{1}$$

and specifically address the case when the middle square factor D is of large size. In what follows, a subspace method is derived to reduce the size of D making efficient computation of the $\mathcal{L}_\infty$-norm of H possible. The domain Ω is an open subset of the complex plane and assumed to enclose the imaginary axis $\mathrm{i}\mathbb{R}$.

Furthermore, it is assumed that the functions $B : \Omega \to \mathbb{C}^{n\times m}$, $C : \Omega \to \mathbb{C}^{p\times n}$, and $D : \Omega \to \mathbb{C}^{n\times n}$ are defined by

$$\begin{aligned} B(s) &:= f_1(s)B_1 + \cdots + f_{\kappa_B}(s)B_{\kappa_B}, \\ C(s) &:= g_1(s)C_1 + \cdots + g_{\kappa_C}(s)C_{\kappa_C}, \\ D(s) &:= h_1(s)D_1 + \cdots + h_{\kappa_D}(s)D_{\kappa_D} \end{aligned} \tag{2}$$

[*] Received by the editors July 22, 2016; accepted for publication (in revised form) by D. Szyld August 30, 2017; published electronically December 12, 2017.
http://www.siam.org/journals/simax/38-4/M108620.html

Funding: The work of the third author was supported in part by the BAGEP program of The Science Academy of Turkey. The work of the fourth author was supported by the DFG priority program 1897: “Calm, Smooth and Smart – Novel Approaches for Influencing Vibrations by Means of Deliberately Introduced Dissipation.” The work of the fifth author was supported by the Einstein Foundation Berlin within the framework of the Einstein Center for Mathematics (ECMath).

[†] Department of Mathematics, Koç University, Rumeli Feneri Yolu 34450, Sarıyer, Istanbul, Turkey (naliyev@ku.edu.tr, emengi@ku.edu.tr).

[‡] Max Planck Institute for Dynamics of Complex Technical Systems, 39106 Magdeburg, Germany (benner@mpi-magdeburg.mpg.de).

[§] Technische Universität Berlin, Institut für Mathematik, 10623 Berlin, Germany (p.schwerdtner@campus.tu-berlin.de, mvoigt@math.tu-berlin.de).

for given matrices $B_1, \ldots, B_{\kappa_B} \in \mathbb{C}^{n\times m}$, $C_1, \ldots, C_{\kappa_C} \in \mathbb{C}^{p\times n}$, $D_1, \ldots, D_{\kappa_D} \in \mathbb{C}^{n\times n}$ and given functions $f_1, \ldots, f_{\kappa_B}, g_1, \ldots, g_{\kappa_C}, h_1, \ldots, h_{\kappa_D} : \Omega \to \mathbb{C}$ that are assumed to be meromorphic in Ω.

For example, if $sE - A$ is a regular pencil, then the transfer function

$$H(s) = C(sE - A)^{-1}B$$

of the descriptor system

$$Ex'(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t), \tag{3}$$

and, more generally, the transfer function

$$H(s) = C\left(sE - A_0 - \sum_{j=1}^{m} \mathrm{e}^{-s\tau_j} A_j\right)^{-1} B$$

of the delay differential-algebraic system

$$Ex'(t) = A_0 x(t) + \sum_{j=1}^{m} A_j x(t - \tau_j) + Bu(t), \quad y(t) = Cx(t), \tag{4}$$

are encompassed by framework (1)–(2). Some other examples are transfer functions of higher order systems and systems containing input and output delays, as well as transfer functions of the form

$$H(s) = sB^*\left(s^2 I_n - \sqrt{s}D_2 + D_3\right)^{-1} B$$

resulting from the spatial discretization of electromagnetic field equations, i.e., the Maxwell equations, describing the electrodynamical behavior of microwave devices with surface losses (see [13] and references therein).

We are concerned with the computation of the $\mathcal{L}_\infty$-norm of H, particularly for the case where n is very large and further $n \gg m$, p. We define the spaces

$$\mathcal{L}_\infty^{p\times m} := \left\{ H|_{\mathrm{i}\mathbb{R}} \;\middle|\; H : \Omega \to \mathbb{C}^{p\times m} \text{ is analytic for an open domain } \Omega \subseteq \mathbb{C} \text{ with } \mathrm{i}\mathbb{R} \subset \Omega \right.$$
$$\left. \text{and } \sup_{\omega\in\mathbb{R}} \|H(\mathrm{i}\omega)\|_2 < \infty \right\},$$
$$\mathcal{H}_\infty^{p\times m} := \left\{ H : \mathbb{C}^+ \to \mathbb{C}^{p\times m} \;\middle|\; H \text{ is analytic and } \sup_{s\in\mathbb{C}^+} \|H(s)\|_2 < \infty \right\},$$

where $\mathbb{C}^+ := \{s \in \mathbb{C} \mid \mathrm{Re}(s) > 0\}$ denotes the open right complex half-plane. In this paper, the function H (more precisely, its restriction to the imaginary axis) is assumed to be in $\mathcal{L}_\infty^{p\times m}$. For ease of notation, we write $H \in \mathcal{L}_\infty^{p\times m}$ instead of $H|_{\mathrm{i}\mathbb{R}} \in \mathcal{L}_\infty^{p\times m}$. For such, the $\mathcal{L}_\infty$-norm is defined by

$$\|H\|_{\mathcal{L}_\infty} := \sup_{\omega\in\mathbb{R}} \|H(\mathrm{i}\omega)\|_2 = \sup_{\omega\in\mathbb{R}} \sigma(H(\mathrm{i}\omega)),$$

where $\sigma(\cdot)$ denotes the largest singular value of its matrix argument. Throughout the text we refer to each function in $\mathcal{L}_\infty^{p\times m}$ as an $\mathcal{L}_\infty$-*function*.

In most applications one is often rather interested in functions which are in $\mathcal{H}_\infty^{p\times m}$. For such, using the maximum principle for analytic functions, one can show that the $\mathcal{H}_\infty$-norm is equivalent to the $\mathcal{L}_\infty$-norm, i.e.,

$$\|H\|_{\mathcal{H}_\infty} := \sup_{s\in\mathbb{C}^+} \|H(s)\|_2 = \sup_{s\in\partial\mathbb{C}^+} \|H(s)\|_2 = \sup_{\omega\in\mathbb{R}} \sigma(H(\mathrm{i}\omega)).$$

1.1. Motivation. The $\mathcal{H}_\infty$-norm plays an indispensable role in the assessment of robust stability as well as in robust control. For instance, assume that we are given an exponentially stable delay differential-algebraic equation

$$Ex'(t) = A_0x(t) + A_1x(t-\tau),$$

and consider the perturbed delay differential-algebraic equation [12]

$$(E + B_1\Delta_1 C)x'(t) = (A_0 + B_2\Delta_2 C)x(t) + (A_1 + B_3\Delta_3 C)x(t-\tau), \tag{5}$$

where $\Delta_i \in \mathbb{C}^{m_i\times p}$, $i = 1, 2, 3$, are the perturbations and $B_i \in \mathbb{C}^{n\times m_i}$ and $C \in \mathbb{C}^{p\times n}$ are matrices that define the perturbation structure. Define the function

$$H(s) := C(sE - A_0 - \mathrm{e}^{-s\tau}A_1)^{-1}\begin{bmatrix} -sB_1 & B_2 & \mathrm{e}^{-s\tau}B_3 \end{bmatrix}.$$

In [12] it is shown that under certain conditions on the matrices E, A_0, A_1 (ensuring a "strangeness-free" system) and some further restrictions on the perturbation matrices B_1, B_2, B_3, the $\mathcal{H}_\infty$-norm is the reciprocal of the structured stability radius, similar to the standard state-space case [21, 22]. In other words, with $\Delta := \begin{bmatrix} \Delta_1^* & \Delta_2^* & \Delta_3^* \end{bmatrix}^*$ we have

$$\inf\left\{ \|\Delta\|_2 \;\middle|\; \text{system (5) is not exponentially stable} \right\} = \|H\|_{\mathcal{H}_\infty}^{-1}.$$

This connection also motivates the importance of the $\mathcal{H}_\infty$-norm in robust control, and the minimization of the $\mathcal{H}_\infty$-norm over system parameters. Consider, for example, the system (see, e.g., [36])

$$\begin{aligned} Ex'(t) &= A_0x(t) + B_1u(t) + B_2w(t), \\ y(t) &= C_1x(t-\tau), \\ z(t) &= C_2x(t), \end{aligned}$$

where u is the control input, y is the (delayed) measured output, w is an input representing noise or unmodeled dynamics, and z is the performance output, respectively. By imposing the feedback law $u(t) = Fy(t)$, we obtain the closed-loop system

$$Ex'(t) = A_0x(t) + B_1FC_1x(t-\tau) + B_2w(t), \quad z(t) = C_2x(t).$$

With $A_{1,F} := B_1FC_1$ its transfer function from w to z is given by

$$H_F(s) = C_2(sE - A_0 - \mathrm{e}^{-s\tau}A_{1,F})^{-1}B_2.$$

The goal of robust control is to determine a stabilizing feedback F such that the closed-loop $\mathcal{H}_\infty$-norm, i.e., $\|H_F\|_{\mathcal{H}_\infty}$, is minimized in order to achieve a maximum robustness of stability of the performance output z with respect to disturbances and noise that enter the system via the input w. For standard state-space systems this $\mathcal{H}_\infty$

optimization problem is addressed by the MATLAB package HIFOO [8]. In the past ten years this software has found manifold applications in industry, some of which are outlined in [29]. Since HIFOO performs a couple of $\mathcal{H}_\infty$-norm evaluations, an efficient $\mathcal{H}_\infty$-norm computation will be beneficial for the performance of the optimization procedure.

1.2. Literature.

Studies concerning the computation of the $\mathcal{L}_\infty$- or $\mathcal{H}_\infty$-norm have been conducted since the late 1980s. Byers' work [9] focuses on the computation of the distance to instability for a matrix, which can be viewed as a special $\mathcal{L}_\infty$-norm computation problem for the transfer function of a standard state-space system (3) with B, C, and E being identities. This idea has been independently adapted for the computation of the $\mathcal{L}_\infty$-norm of transfer functions of standard state-space systems by Boyd and Balakrishnan [5], as well as Bruinsma and Steinbuch [6]. An extension of these methods to transfer functions of descriptor systems is discussed in [3]. These are level-set based optimization approaches, and require the repeated solution of Hamiltonian eigenvalue problems of size twice the order of the system. Consequently, they are not suitable for systems beyond medium scale.

For larger problems, several approaches have been proposed in recent years. For instance, the characterization of the $\mathcal{L}_\infty$-norm via a Hamiltonian eigenvalue problem has been used to formulate an associated root-finding problem which can be solved using Newton's method [14]. This approach requires solutions of linear systems of size equal to the order of the system. Some other approaches [17, 34, 30] are restricted to the case of the $\mathcal{H}_\infty$-norm only. They are based on the relation of the $\mathcal{H}_\infty$-norm to the structured stability radius and structured ε-pseudospectra [33, 25]; these approaches compute the rightmost point of the structured ε-pseudospectrum repeatedly for various values of ε. However, all of these methods for larger problems converge only locally and there is no guarantee that the global maximum of $\sigma(H(\mathrm{i}\cdot))$ is found.

The delay-system setting is addressed by a few works [18, 19] only. These are extensions of the level-set based approach of Byers, but involve infinite dimensional operators. None of these works benefits from a subspace projection idea and their use is typically limited to systems of the order of a thousand at most.

1.3. Contributions and outline.

Our approach is based on a reduction of the middle factor $D(s)$ in (1) to a much smaller dimension using two-sided projections. The $\mathcal{L}_\infty$-norm is computed for the resulting reduced matrix-valued function, then the subspaces are expanded using the singular vectors of $H(\mathrm{i}\omega_r)$, where $\mathrm{i}\omega_r$ is the point on the imaginary axis (including infinity) at which the reduced function attains its $\mathcal{L}_\infty$-norm. Our expansion strategy leads to superlinear convergence with respect to the subspace dimension which we observe in practice and prove in theory. This work is inspired by a recent work [24] on a subspace method in the context of eigenvalue optimization. However, unlike [24], the matrix-valued function $H(\mathrm{i}\cdot)$ (whose largest singular value is to be maximized) is $p \times m$, where p, m are typically small, the large-scale nature of the problem in this paper is due to the size of $D(s)$. Dealing with the large dimensionality of $D(s)$ requires a different approach compared to the one proposed for eigenvalue optimization in [24].

We expose our work in the following order. In the next section, we formally introduce the reduced matrix-valued functions and present a result (Theorem 2.1) that points out how Hermite interpolation of the original $\mathcal{L}_\infty$-function can be achieved by a reduced matrix-valued function. This interpolation result gives rise to the formal definition of the subspace method as in Algorithm 1. The method is devised in order to lift the Hermite interpolation properties to the largest singular value functions associated with the original $\mathcal{L}_\infty$-function and the reduced matrix-valued function. The local superlinear convergence of the subspace method can be attributed to these interpolation properties. This convergence is proven rigorously in section 3. Important implementation details of the proposed method and the results of our numerical experiments are discussed in section 4.

2. Our approach.

Two-sided subspace projections are widely used in model order reduction [11, 35, 1, 16]. In the context of a descriptor system of the form (3), this amounts to restricting the state space to a subspace $\mathcal{V}$ of dimension much smaller than the original state space, and imposing a Petrov–Galerkin condition with respect to another subspace $\mathcal{W}$. Formally, introducing matrices V, W whose columns span $\mathcal{V}$, $\mathcal{W}$, respectively, and denoting the reduced state at time t with $\tilde{x}(t)$, the reduced system is defined by

$$W^* \left(EV\tilde{x}'(t) - AV\tilde{x}(t) - Bu(t)\right) = 0 \quad \text{and} \quad y(t) = CV\tilde{x}(t).$$

The transfer functions associated with the original descriptor system and the reduced one above are

$$H(s) = C(sE-A)^{-1}B \quad \text{and} \quad \widetilde{H}(s) = CV(sW^*EV - W^*AV)^{-1}W^*B.$$

The representation of the reduced transfer function above is under the assumption that $\mathcal{V}$ and $\mathcal{W}$ are of equal dimension.

More generally, let us consider general $\mathcal{L}_\infty$-functions in the framework of (1). We define the reduced function by

$$\widetilde{H} : \Omega \to \mathbb{C}^{p\times m}, \quad \widetilde{H}(s) := \widetilde{C}(s)\widetilde{D}(s)^{-1}\widetilde{B}(s), \tag{6}$$

where

$$\widetilde{B}(s) := f_1(s)\widetilde{B}_1 + \cdots + f_{\kappa_B}(s)\widetilde{B}_{\kappa_B}, \qquad \widetilde{B}_j := W^*B_j, \qquad j = 1,\dots,\kappa_B, \tag{7a}$$

$$\widetilde{C}(s) := g_1(s)\widetilde{C}_1 + \cdots + g_{\kappa_C}(s)\widetilde{C}_{\kappa_C}, \qquad \widetilde{C}_j := C_jV, \qquad j = 1,\dots,\kappa_C, \tag{7b}$$

$$\widetilde{D}(s) := h_1(s)\widetilde{D}_1 + \cdots + h_{\kappa_D}(s)\widetilde{D}_{\kappa_D}, \qquad \widetilde{D}_j := W^*D_jV, \qquad j = 1,\dots,\kappa_D. \tag{7c}$$

Throughout the rest of this work, we focus on matrices $V, W \in \mathbb{C}^{n\times\tilde{n}}$ with $\tilde{n} \ll n$ whose columns span the subspaces $\mathcal{V}$, $\mathcal{W}$, respectively. Furthermore, in what follows, we always assume that the subspaces $\mathcal{V}$, $\mathcal{W}$ are such that $\widetilde{H}$ is well-defined and bounded on the imaginary axis and that $\widetilde{D}$ is invertible almost everywhere on the imaginary axis. The following result is fundamental to our approach. It is a special case of [2, Theorem 1].

Theorem 2.1. *Let $\mu \in \mathbb{C}$ be such that $C(\mu)$, $D(\mu)$, and $B(\mu)$ are analytic and both $D(\mu)$and $\widetilde{D}(\mu)$ are invertible. Suppose also that $b \in \mathbb{C}^m$ and $c \in \mathbb{C}^p$ are given nonzero vectors. Then the following statements hold:*

(i) *If $D(\mu)^{-1}B(\mu)b \in \operatorname{Col}(V)$, then $H(\mu)b = \widetilde{H}(\mu)b$.*

(ii) *If $(c^*C(\mu)D(\mu)^{-1})^* \in \operatorname{Col}(W)$, then $c^*H(\mu) = c^*\widetilde{H}(\mu)$.*

(iii) *If $D(\mu)^{-1}B(\mu)b \in \operatorname{Col}(V)$ and $(c^*C(\mu)D(\mu)^{-1})^* \in \operatorname{Col}(W)$, then $c^*H'(\mu)b = c^*\widetilde{H}'(\mu)\,b$.*

For the computation of the $\mathcal{L}_\infty$-norm, we form subspaces $\mathcal{W}$, $\mathcal{V}$ that give rise to the Hermite interpolation of $\sigma(H(s))$ by $\sigma(\widetilde{H}(s))$ at some nodes $\mu_1,\dots,\mu_\ell$, that is,

$$\sigma(H(\mu_j)) = \sigma\big(\widetilde{H}(\mu_j)\big) \quad \text{and} \quad \sigma'(H(\mu_j)) = \sigma'\big(\widetilde{H}(\mu_j)\big) \quad \text{for} \quad j = 1,\dots,\ell.$$

Theorem 2.1 above is helpful in this direction. It is immediate from part (i) of the theorem that if $D(\mu)^{-1}B(\mu)v \in \mathcal{V}$ for a right singular vector v associated with $\sigma(H(\mu))$, then we have $\sigma(H(\mu)) \leq \sigma(\widetilde{H}(\mu))$. The same conclusion can be drawn from part (ii) if $(w^*C(\mu)D(\mu)^{-1})^* \in \mathcal{W}$ for a left singular vector w associated with $\sigma(H(\mu))$. Furthermore, it can be shown that if $D(\mu)^{-1}B(\mu)v_j \in \mathcal{V}$ for each right singular vector v_j of $H(\mu)$ and $(w_j^*C(\mu)D(\mu)^{-1})^* \in \mathcal{W}$ for each left singular vector w_j of $H(\mu)$, then the equality $H(\mu) = \widetilde{H}(\mu)$ is attained (see Lemma 3.1, part (i) below), implying $\sigma(H(\mu)) = \sigma(\widetilde{H}(\mu))$. Additionally, $H(\mu) = \widetilde{H}(\mu)$ have the same right and left singular vectors v, w corresponding to $\sigma(H(\mu)) = \sigma(\widetilde{H}(\mu))$ and $D(\mu)^{-1}B(\mu)v \in \mathcal{V}, (w^*C(\mu)D(\mu)^{-1})^* \in \mathcal{W}$. Consequently, part (iii) of Theorem 2.1 leads to the desired Hermite interpolation property

$$\sigma'(H(\mu)) = \operatorname{Re}\left(w^* H'(\mu)\, v\right) = \operatorname{Re}\left(w^* \widetilde{H}'(\mu)\, v\right) = \sigma'\left(\widetilde{H}(\mu)\right),$$

where the first and the third equality follow from the analytical formulas for the derivatives of singular value functions [26, 7].

An observation that enhances efficiency is that the singular vectors v_j, w_j do not need to be calculated explicitly. It is sufficient that we have

$$\left\{D(\mu)^{-1}B(\mu)v_j \;\middle|\; v_j \text{ is a right singular vector of } H(\mu)\right\} \subseteq \mathcal{V}, \quad \text{and} \tag{8}$$

$$\left\{\left(w_j^*C(\mu)D(\mu)^{-1}\right)^* \;\middle|\; w_j \text{ is a left singular vector of } H(\mu)\right\} \subseteq \mathcal{W} \tag{9}$$

in order to obtain the Hermite interpolation property. Note that $\mathcal{V}$ and $\mathcal{W}$ must have the same dimension, otherwise the middle factor $\widetilde{D}(s)$ of $\widetilde{H}(s)$ defined by (7) is not square and Theorem 2.1 fails. Clearly, the choices $\mathcal{V} = \operatorname{Col}(D(\mu)^{-1}B(\mu))$ and $\mathcal{W} = \operatorname{Col}((C(\mu)D(\mu)^{-1})^*)$ yield the desired inclusions (8) and (9), but have different dimensions unless $m = p$. When $m < p$, we have

$$\begin{aligned}&\left\{\left(w_j^*C(\mu)D(\mu)^{-1}\right)^* \;\middle|\; w_j \text{ is a left singular vector of } H(\mu)\right\}\\ &\qquad = \left\{\left((H(\mu)v_j)^*C(\mu)D(\mu)^{-1}\right)^* \;\middle|\; v_j \text{ is a right singular vector of } H(\mu)\right\}\\ &\qquad\qquad \subseteq \operatorname{Col}\left(\left(C(\mu)D(\mu)^{-1}\right)^* H(\mu)\right),\end{aligned}$$

so the subspaces $\mathcal{V} = \operatorname{Col}(D(\mu)^{-1}B(\mu))$ and $\mathcal{W} = \operatorname{Col}((C(\mu)D(\mu)^{-1})^*H(\mu))$ have equal dimension and satisfy (8) and (9), respectively. Similarly, when $m > p$, it can be deduced that the subspaces $\mathcal{V} = \operatorname{Col}(D(\mu)^{-1}B(\mu)H(\mu)^*)$ and $\mathcal{W} = \operatorname{Col}((C(\mu)D(\mu)^{-1})^*)$ are of equal dimension, and satisfy (8) and (9).

The subspace method is described below in Algorithm 1. It generates matrices V_r, W_r and acts on the subspaces $\operatorname{Col}(V_r)$, $\operatorname{Col}(W_r)$ of growing dimension as r increases for $r = 1, 2, \ldots$. In the description, the notation $\widetilde{H}_r(s)$ refers to the reduced function $\widetilde{H}(s)$ defined as in (6) and (7), but with the particular choices $V = V_r$ and $W = W_r$. Thus, at iteration r on line 10, the algorithm maximizes $\sigma\left(\widetilde{H}_{r-1}(s)\right)$ over the imaginary axis and retrieves the global maximizer $\mathrm{i}\omega_r$. Then, it expands the subspaces $\operatorname{Col}(V_{r-1}), \operatorname{Col}(W_{r-1})$ and thus forms $\widetilde{H}_r(s)$ such that the Hermite interpolation properties $\sigma(H(\mathrm{i}\omega_r)) = \sigma\left(\widetilde{H}_r(\mathrm{i}\omega_r)\right), \sigma'(H(\mathrm{i}\omega_r)) = \sigma'\left(\widetilde{H}_r(\mathrm{i}\omega_r)\right)$ hold. In practice we observe that Algorithm 1 converges to a local maximizer of $\sigma(\omega) := \sigma(H(\mathrm{i}\omega))$ (that is not necessarily a global maximizer) at a superlinear rate of convergence. The next section is devoted to a formal proof of this superlinear rate of convergence. Numerical experiments showing this convergence are reported in section 4.

Algorithm 1 Subspace method for the computation of the $\mathcal{L}_\infty$-norm.

```
Input: matrices $B_1, \dots, B_{\kappa_B} \in \mathbb{C}^{n\times m}$, $C_1, \dots, C_{\kappa_C} \in \mathbb{C}^{p\times n}$, $D_1, \dots, D_{\kappa_D} \in \mathbb{C}^{n\times n}$
    and functions $f_1, \dots, f_{\kappa_B}, g_1, \dots, g_{\kappa_C}, h_1, \dots, h_{\kappa_D}$ as in (2).
Output: the $\mathcal{L}_\infty$-norm of $H \in \mathcal{L}_\infty^{p\times m}$ with $H$ as in (1) and (2).
$\omega_1 \leftarrow$ a random number in $\mathbb{R}$.
if $m = p$ then
    $V_1 \leftarrow D(\mathrm{i}\omega_1)^{-1}B(\mathrm{i}\omega_1)$ and $W_1 \leftarrow \left(C(\mathrm{i}\omega_1)D(\mathrm{i}\omega_1)^{-1}\right)^*$.
else if $m < p$ then
    $V_1 \leftarrow D(\mathrm{i}\omega_1)^{-1}B(\mathrm{i}\omega_1)$ and $W_1 \leftarrow \left(C(\mathrm{i}\omega_1)D(\mathrm{i}\omega_1)^{-1}\right)^* H(\mathrm{i}\omega_1)$.
else
    $V_1 \leftarrow D(\mathrm{i}\omega_1)^{-1}B(\mathrm{i}\omega_1)H(\mathrm{i}\omega_1)^*$ and $W_1 \leftarrow \left(C(\mathrm{i}\omega_1)D(\mathrm{i}\omega_1)^{-1}\right)^*$.
end if
for $r = 2, 3, \dots$ do
    Form $\widetilde{H}_{r-1}$ as in (6) and (7) and set $\omega_r \leftarrow \arg\max_{\omega\in\mathbb{R}\cup\{\infty\}} \sigma(\widetilde{H}_{r-1}(\mathrm{i}\omega))$.
    if $m = p$ then
        $\widetilde{V}_r \leftarrow D(\mathrm{i}\omega_r)^{-1}B(\mathrm{i}\omega_r)$ and $\widetilde{W}_r \leftarrow \left(C(\mathrm{i}\omega_r)D(\mathrm{i}\omega_r)^{-1}\right)^*$.
    else if $m < p$ then
        $\widetilde{V}_r \leftarrow D(\mathrm{i}\omega_r)^{-1}B(\mathrm{i}\omega_r)$ and $\widetilde{W}_r \leftarrow \left(C(\mathrm{i}\omega_r)D(\mathrm{i}\omega_r)^{-1}\right)^* H(\mathrm{i}\omega_r)$.
    else
        $\widetilde{V}_r \leftarrow D(\mathrm{i}\omega_r)^{-1}B(\mathrm{i}\omega_r)H(\mathrm{i}\omega_r)^*$ and $\widetilde{W}_r \leftarrow \left(C(\mathrm{i}\omega_r)D(\mathrm{i}\omega_r)^{-1}\right)^*$.
    end if
    $V_r \leftarrow \operatorname{orth}\left(\begin{bmatrix} V_{r-1} & \widetilde{V}_r \end{bmatrix}\right)$ and $W_r \leftarrow \operatorname{orth}\left(\begin{bmatrix} W_{r-1} & \widetilde{W}_r \end{bmatrix}\right)$.
end for
```

Remark 2.2. The procedure described in Algorithm 1 resembles the reduced basis approach for model order reduction of parametrized systems; see, e.g., [20]. The key ingredients are projection onto a subspace, solving the resulting low-dimensional problem, a subprocedure to maximize (minimize) a desired quantity for the reduced parametrized system, and expanding the subspace by a snapshot of the full-order problem at the argmax/min returned by the subprocedure. As all these ingredients are used in Algorithm 1, it can be considered as a reduced basis method.

Remark 2.3. In Algorithm 1 the subspaces from all of the previous iterations are kept. An alternative would be to keep the subspaces only from the last two iterations. The rate of convergence analysis in the next section also applies to this variant, since that analysis (specifically Theorem 3.3) relies on the interpolation properties only at the last two iterates. Thus, the variant with only subspaces from the last two iterations is also guaranteed to converge at a superlinear rate, which we observe in practice. However, for the numerical experiments discussed in section 4, all the previous subspaces are kept. This results in better global convergence properties, and usually avoidance of stagnation at a local maximizer of $\sigma(\cdot)$ that is not a global maximizer. The cost of keeping additional subspaces is usually small, because the algorithm often needs fewer iterations than the variant that uses only the subspaces from the last two iterations. Typically it converges quickly up to prescribed tolerances in less than 10 iterations (see the numerical results in sections 4.2 and 4.3).

3. Rate of convergence analysis.

In this section, we prove that the aforementioned Hermite interpolation properties of the subspace method lead to a superlinear convergence with respect to the subspace dimension, under the assumption that the method converges locally. The argument revolves around the singular value functions

$\sigma(\omega) := \sigma(H(\mathrm{i}\omega))$ and $\sigma_r(\omega) := \sigma(\widetilde{H}_r(\mathrm{i}\omega))$. Occasionally, the jth largest singular values of $H(\mathrm{i}\omega)$ and $\widetilde{H}_r(\mathrm{i}\omega)$ are also needed, which we denote by $\sigma^{(j)}(\omega)$ and $\sigma_r^{(j)}(\omega)$, respectively. We also reserve the notations $\underline{\sigma}(\omega)$ and $\underline{\sigma}_r(\omega)$ specifically for the second largest singular values of these matrix-valued functions. When $\min\{m,p\} < j$, then we define $\sigma^{(j)}(\omega) = \sigma_r^{(j)}(\omega) = 0$ for all $\omega \in \mathbb{R}$. We first formally state and prove Hermite interpolation properties of the singular value functions.

LEMMA 3.1. *The following statements hold regarding Algorithm* 1 *for $k = 1, \ldots, r$:*

(i) $H(\mathrm{i}\omega_k) = \widetilde{H}_r(\mathrm{i}\omega_k)$.

(ii) $\sigma(\omega_k) = \sigma_r(\omega_k)$ *and* $\underline{\sigma}(\omega_k) = \underline{\sigma}_r(\omega_k)$.

(iii) *If $\sigma(\omega_k)$ is simple, then $\sigma(\omega)$, $\sigma_r(\omega)$ are differentiable at ω_k and $\sigma'(\omega_k) = \sigma_r'(\omega_k)$.*

Proof.

(i) When $m \leq p$, for each $k \in \{1, \ldots, r\}$, we have

$$\mathrm{Col}(D(\mathrm{i}\omega_k)^{-1}B(\mathrm{i}\omega_k)) \subseteq \mathrm{Col}(V_r),$$

due to lines 3, 5, 12, 14, and 18 of Algorithm 1. Thus, $D(\mathrm{i}\omega_k)^{-1}B(\mathrm{i}\omega_k)e_j \in \mathrm{Col}(V_r)$ for $j = 1, \ldots, m$. It follows that $H(\mathrm{i}\omega_k)e_j = \widetilde{H}_r(\mathrm{i}\omega_k)e_j$ from part (i) of Theorem 2.1 for $j = 1, \ldots, m$, that is, $H(\mathrm{i}\omega_k) = \widetilde{H}_r(\mathrm{i}\omega_k)$. On the other hand, when $m > p$, for each $k \in \{1, \ldots, r\}$, the inclusion $\mathrm{Col}((C(\mathrm{i}\omega_k)D(\mathrm{i}\omega_k)^{-1})^*) \subseteq \mathrm{Col}(W_r)$ follows from lines 7, 16, and 18 of Algorithm 1. Consequently, $(e_j^* C(\mathrm{i}\omega)D(\mathrm{i}\omega_k)^{-1})^* \in \mathrm{Col}(W_r)$, so $e_j^* H(\mathrm{i}\omega_k) = e_j^* \widetilde{H}_r(\mathrm{i}\omega_k)$ by part (ii) of Theorem 2.1 for each $j = 1, \ldots, p$, that is, $H(\mathrm{i}\omega_k) = \widetilde{H}_r(\mathrm{i}\omega_k)$.

(ii) This is immediate from part (i).

(iii) Suppose that $\sigma(\omega_k)$ is simple for a particular $k \in \{1, \ldots, r\}$. This implies that $\sigma(\omega)$, $\sigma_r(\omega)$ are differentiable at ω_k [31]. The left and right singular vectors corresponding to $\sigma(\omega_k)$ and $\sigma_r(\omega_k)$ are the same, since $H(\mathrm{i}\omega_k) = \widetilde{H}_r(\mathrm{i}\omega_k)$ due to part (i). Let us denote them by $w \in \mathbb{C}^p$ and $v \in \mathbb{C}^m$, respectively, and without loss of generality (w.l.o.g.), assume these are unit vectors. Suppose $m \leq p$. In this case, $\mathrm{Col}(D(\mathrm{i}\omega_k)^{-1}B(\mathrm{i}\omega_k)) \subseteq \mathrm{Col}(V_r)$ and $\mathrm{Col}((C(\mathrm{i}\omega_k)D(\mathrm{i}\omega_k)^{-1})^* H(\mathrm{i}\omega_k)) \subseteq \mathrm{Col}(W_r)$, so we have $D(\mathrm{i}\omega_k)^{-1}B(\mathrm{i}\omega_k)v \in \mathrm{Col}(V_r)$ and

$$\left(C(\mathrm{i}\omega_k)D(\mathrm{i}\omega_k)^{-1}\right)^* H(\mathrm{i}\omega_k)v = \sigma(\omega_k)\left(w^* C(\mathrm{i}\omega_k)D(\mathrm{i}\omega_k)^{-1}\right)^* \in \mathrm{Col}(W_r).$$

When we have $m > p$, the inclusions $\mathrm{Col}(D(\mathrm{i}\omega_k)^{-1}B(\mathrm{i}\omega_k)H(\mathrm{i}\omega_k)^*) \subseteq \mathrm{Col}(V_r)$ and $\mathrm{Col}((C(\mathrm{i}\omega_k)D(\mathrm{i}\omega_k)^{-1})^*) \subseteq \mathrm{Col}(W_r)$ hold. This implies that

$$(w^* C(\mathrm{i}\omega_k)D(\mathrm{i}\omega_k)^{-1})^* \in \mathrm{Col}(W_r)$$

and

$$D(\mathrm{i}\omega_k)^{-1}B(\mathrm{i}\omega_k)H(\mathrm{i}\omega_k)^* w = \sigma(\omega_k)D(\mathrm{i}\omega_k)^{-1}B(\mathrm{i}\omega_k)v \in \mathrm{Col}(V_r).$$

In both cases, part (iii) of Theorem 2.1 yields $w^* H'(\mathrm{i}\omega_k)v = w^* \widetilde{H}_r'(\mathrm{i}\omega_k)v$. Finally, by exploiting the analytical formulas for the derivatives of singular value functions [26, 7], we deduce

$$\sigma'(\omega_k) = \mathrm{Re}(w^* H'(\mathrm{i}\omega_k)v) = \mathrm{Re}\left(w^* \widetilde{H}_r'(\mathrm{i}\omega_k)v\right) = \sigma_r'(\omega_k). \qquad \square$$

The next result concerns how accurately $\sigma_r''(\cdot)$ approximates $\sigma''(\cdot)$ at ω_r. We view ω_k for every $k > 1$ as a function of the initial point ω_1 for the next result and the subsequent rate of convergence result. A consequence is that the function $\sigma_k(\cdot)$ also depends on ω_1. Furthermore, in what follows, for a given bounded interval $\mathcal{I} \subseteq \mathbb{R}$, we consider ω_1 such that $\omega_k \in \mathcal{I}$ for each $k \geq 1$. Due to the analyticity of the function H on the imaginary axis (recall that $H \in \mathcal{L}_\infty^{p\times m}$), there exists a Lipschitz constant $\eta_{\mathcal{I}} > 0$ such that

$$\|H(\mathrm{i}\omega) - H(\mathrm{i}\widetilde{\omega})\|_2 \leq \eta_{\mathcal{I}}|\omega - \widetilde{\omega}| \qquad \forall \omega,\, \widetilde{\omega} \in \mathcal{I}. \tag{10}$$

Additionally, for a given $\eta \geq \eta_{\mathcal{I}}$ and $\varphi > 0$, we consider ω_1 such that

$$\|\widetilde{H}_k^{(j)}(\mathrm{i}\omega) - \widetilde{H}_k^{(j)}(\mathrm{i}\widetilde{\omega})\|_2 \leq \eta|\omega - \widetilde{\omega}| \qquad \forall \omega,\, \widetilde{\omega} \in \mathcal{I} \tag{11}$$

for $k \geq 1$ and $j = 0, 1, 2$ as well as

$$\|\widetilde{H}_k^{(j)}(\mathrm{i}\omega)\|_2 \leq \varphi \qquad \forall \omega \in \mathcal{I} \tag{12}$$

for $k \geq 1$ and $j = 1, 2$. Condition (10), in particular the existence of the constant $\eta_{\mathcal{I}}$, is a simple consequence of the analyticity of $H(\mathrm{i}\cdot)$ and the boundedness of $\mathcal{I}$, whereas conditions (11) and (12) are assumptions, which are typically satisfied in practice because of the interpolation properties between $H(\mathrm{i}\cdot)$ and $\widetilde{H}_k(\mathrm{i}\cdot)$. These conditions imply the Lipschitz continuity of $\sigma''(\cdot)$ and $\sigma_k''(\cdot)$ on $\mathcal{I}$ with Lipschitz constants independent of ω_1, which is established and exploited by the proof of the next lemma.

Lemma 3.2. *For a given $\zeta \in \mathbb{R}^+$, an integer $r \geq 2$, a bounded interval $\mathcal{I} \subseteq \mathbb{R}$, $\varphi \in \mathbb{R}^+$, and $\eta \geq \eta_{\mathcal{I}}$, where $\eta_{\mathcal{I}}$ is as in* (10)*, suppose ω_1 is chosen in a way so that $\omega_k \in \mathcal{I}$ for $k \geq 1$, conditions* (11) *and* (12) *hold, as well as*

$$\sigma^{(j)}(\omega_r) - \sigma^{(j+1)}(\omega_r) \geq \zeta \geq c\eta|\omega_r - \omega_{r-1}| \quad \text{for } \ j = 1, \dots, \min\{m, p\} - 1 \tag{13}$$

for some constant $c > 2$. Then we have

$$|\sigma''(\omega_r) - \sigma_r''(\omega_r)| \leq \mu|\omega_r - \omega_{r-1}|$$

for some constant μ independent of ω_1.

Proof. We start by establishing the simplicity of $\sigma(\omega)$ and $\sigma_r(\omega)$ on the closed interval with endpoints ω_{r-1}, ω_r, which we denote with $\mathcal{I}_r$. To this end, for each $\omega \in \mathcal{I}_r$, we have

$$\begin{aligned} &|\sigma(\omega_r) - \sigma(\omega)| \leq \|H(\mathrm{i}\omega_r) - H(\mathrm{i}\omega)\|_2 \leq \eta|\omega_r - \omega| \leq \eta|\omega_r - \omega_{r-1}| \quad \text{and} \\ &|\underline{\sigma}(\omega_r) - \underline{\sigma}(\omega)| \leq \|H(\mathrm{i}\omega_r) - H(\mathrm{i}\omega)\|_2 \leq \eta|\omega_r - \omega| \leq \eta|\omega_r - \omega_{r-1}|, \end{aligned}$$

due to Weyl's theorem [23, Theorem 4.3.1] and inequality (10) regarding the Lipschitz continuity of H. Hence we have

$$\sigma(\omega) \geq \sigma(\omega_r) - \eta|\omega_r - \omega_{r-1}| \quad \text{and} \quad \underline{\sigma}(\omega) \leq \underline{\sigma}(\omega_r) + \eta|\omega_r - \omega_{r-1}|,$$

that is,

$$\sigma(\omega) - \underline{\sigma}(\omega) \geq \{\sigma(\omega_r) - \underline{\sigma}(\omega_r)\} - 2\eta|\omega_r - \omega_{r-1}| \geq (c-2)\eta|\omega_r - \omega_{r-1}| > 0.$$

Above, the second inequality follows from (13). This shows that $\sigma(\omega)$ is simple for each $\omega \in \mathcal{I}_r$. Furthermore, by part (ii) of Lemma 3.1, we have $\sigma_r(\omega_r) = \sigma(\omega_r)$

and $\underline{\sigma}_r(\omega_r) = \underline{\sigma}(\omega_r)$. An analogous argument with $\sigma_r(\cdot)$ taking the role of $\sigma(\cdot)$ also shows the simplicity of $\sigma_r(\omega)$ for each $\omega \in \mathcal{I}_r$. It follows that both $\sigma(\cdot)$ and $\sigma_r(\cdot)$ are analytic on $\mathcal{I}_r$. Similarly, $\sigma^{(j)}(\cdot)$ and $\sigma_r^{(j)}(\cdot)$ can also be shown to be simple on $\mathcal{I}_r$ for $j = 2, \dots, \min\{m, p\} - 1$.

To relate the second derivatives, we exploit part (iii) of Lemma 3.1, in particular, $\sigma'(\omega_k) = \sigma_r'(\omega_k)$ for $k = r-1,\, r$. These interpolation properties imply

$$\sigma''(\xi)\,(\omega_r - \omega_{r-1}) = \sigma'(\omega_r) - \sigma'(\omega_{r-1}) = \sigma_r'(\omega_r) - \sigma_r'(\omega_{r-1}) = \sigma_r''\big(\widehat{\xi}\big)\,(\omega_r - \omega_{r-1})$$

for some $\xi,\, \widehat{\xi} \in \mathcal{I}_r$ leading to

$$\sigma''(\xi) - \sigma_r''(\widehat{\xi}) = 0. \tag{14}$$

Moreover, the second derivatives of $\sigma(\cdot)$, $\sigma_r(\cdot)$ are Lipschitz continuous in $\mathcal{I}_r$, so there exist positive constants γ_1, γ_2 such that

$$\begin{aligned} \big|\sigma_r''(\omega_r) - \sigma_r''\big(\widehat{\xi}\big)\big| &\le \gamma_1 \big|\widehat{\xi} - \omega_r\big| \le \gamma_1 |\omega_r - \omega_{r-1}| \quad \text{and} \\ |\sigma''(\omega_r) - \sigma''(\xi)| &\le \gamma_2 |\xi - \omega_r| \le \gamma_2 |\omega_r - \omega_{r-1}|. \end{aligned} \tag{15}$$

We claim that the Lipschitz constant γ_1 can be expressed solely in terms of η, φ, ζ (satisfying (11), (12), (13), respectively).

To see this, let us denote a unit eigenvector corresponding to the largest eigenvalue of

$$\begin{bmatrix} 0 & \widetilde{H}_r(\mathrm{i}\omega) \\ \big[\widetilde{H}_r(\mathrm{i}\omega)\big]^* & 0 \end{bmatrix} \tag{16}$$

by $v_r(\omega)$, and a unit eigenvector corresponding to the jth largest eigenvalue $\lambda_{r,j}(\omega)$ of this matrix by $v_{r,j}(\omega)$. Then the claim is evident from the analytical expression [26]

$$\begin{aligned} \sigma_r''(\omega) = {} & v_r(\omega)^* \begin{bmatrix} 0 & \widetilde{H}_r''(\mathrm{i}\omega) \\ \big[\widetilde{H}_r''(\mathrm{i}\omega)\big]^* & 0 \end{bmatrix} v_r(\omega) \\ & + 2 \sum_{j=2}^{2r} \frac{1}{\sigma_r(\omega) - \lambda_{r,j}(\omega)} \left| v_{r,j}(\omega)^* \begin{bmatrix} 0 & \widetilde{H}_r'(\mathrm{i}\omega) \\ \big[\widetilde{H}_r'(\mathrm{i}\omega)\big]^* & 0 \end{bmatrix} v_r(\omega) \right|^2, \end{aligned}$$

where $\sigma_r(\cdot)$, $\lambda_{r,j}(\cdot)$, $v_r(\cdot)$, $v_{r,j}(\cdot)$, $\widetilde{H}_r'(\mathrm{i}\cdot)$, and $\widetilde{H}_r''(\mathrm{i}\cdot)$ are Lipschitz continuous on $\mathcal{I}_r$ with Lipschitz constants depending on η and ζ only. Here we remark that the terms $\sigma_r(\omega) - \lambda_{r,j}(\omega)$ can be bounded from below by a quantity solely depending on η and ζ, because of the interpolation properties $\sigma(\omega_r) = \sigma_r(\omega_r)$, $\underline{\sigma}(\omega_r) = \underline{\sigma}_r(\omega_r)$, and assumption (13). Furthermore, we exploit the fact that if f, g are Lipschitz continuous functions with Lipschitz constants β_1, β_2 on a closed interval, then fg is also Lipschitz continuous with Lipschitz constant $\beta_1 g_* + \beta_2 f_*$, where f_*, g_* are the maximum values of f, g attained on the interval. The second derivative $\sigma''(\omega)$ can be expressed in a way analogous to the expression for $\sigma_r''(\omega)$ above, in terms of the matrix obtained from the one in (16) by replacing $\widetilde{H}_r(\mathrm{i}\omega)$ by $H(\mathrm{i}\omega)$, as well as the eigenvalues and eigenvectors of this matrix. Hence, the Lipschitz constant γ_2 in (15) can be expressed in terms of η, ζ, and an upper bound on $\|H^{(j)}(\mathrm{i}\omega)\|_2$ for $j = 1, 2$ and for all $\omega \in \mathcal{I}$. Finally, (14) and inequalities (15) yield

$$\begin{aligned}|\sigma''(\omega_r) - \sigma_r''(\omega_r)| &= |\sigma''(\omega_r) - \sigma_r''(\omega_r) + \sigma_r''(\widehat{\xi}) - \sigma''(\xi)| \\ &\leq |\sigma''(\omega_r) - \sigma''(\xi)| + \left|\sigma_r''(\omega_r) - \sigma_r''(\widehat{\xi})\right| \leq (\gamma_1 + \gamma_2)|\omega_r - \omega_{r-1}|,\end{aligned}$$

hence the result follows. ∎

The main result presented next assumes ω_{r-1}, ω_r, ω_{r+1} are sufficiently close to a local maximizer of $\sigma(\cdot)$ for certain values of ω_1 and a given $r \geq 2$. This is a convergence assumption which we observe in practice.

Theorem 3.3 (local superlinear convergence). *Let ω_* be a local maximizer of $\sigma(\omega)$ such that $\sigma(\omega_*)$ is simple, and $\sigma''(\omega_*) \neq 0$. Furthermore, let*

$$\zeta := \min\{\sigma^{(j)}(\omega_*) - \sigma^{(j+1)}(\omega_*) \mid j = 1, \ldots, \min\{m, p\} - 1\}.$$

For a given integer $r \geq 2$, a bounded interval $\mathcal{I} \subseteq \mathbb{R}$ containing ω_* in its interior, $\varphi \in \mathbb{R}^+$, and $\eta \geq \eta_{\mathcal{I}}$, where $\eta_{\mathcal{I}}$ is as in (10)*, suppose ω_1 is chosen in a way so that $\omega_k \in \mathcal{I}$ for $k \geq 1$, conditions* (11), (12) *hold, and $\delta := \max\{|\omega_{r+1} - \omega_*|, |\omega_r - \omega_*|, |\omega_{r-1} - \omega_*|\}$ is sufficiently small, in particular*

$$\zeta \geq 8\eta\delta. \tag{17}$$

Then we have

$$\frac{|\omega_{r+1} - \omega_*|}{|\omega_r - \omega_*| \cdot \max\{|\omega_{r-1} - \omega_*|, |\omega_r - \omega_*|\}} \leq \nu$$

for some constant ν independent of ω_1.

Proof. The proof is split into two parts. In the first part, we deduce the analyticity of the singular value functions $\sigma(\cdot)$, $\sigma_r(\cdot)$ on $\mathcal{I}(\omega_*, \delta) := [\omega_* - \delta, \omega_* + \delta]$, bound their second derivatives from below, and the third derivative of $\sigma_r(\cdot)$ from above uniformly on $\mathcal{I}(\omega_*, \delta)$ by quantities that do not depend on ω_1. Then the second part makes use of these uniform bounds to relate $|\omega_{r+1} - \omega_*|$ with $|\omega_r - \omega_*|$, $|\omega_{r-1} - \omega_*|$ and conclude a superlinear rate of convergence.

Part 1: We first show the analyticity of $\sigma(\cdot)$ and $\sigma_r(\cdot)$ on $\mathcal{I}(\omega_*, \delta)$. Condition (17) together with Weyl's theorem [23, Theorem 4.3.1] ensures that

$$\sigma(\omega) - \underline{\sigma}(\omega) \geq \{\sigma(\omega_*) - \underline{\sigma}(\omega_*)\} - 2\eta\delta \geq 3\zeta/4$$

for each $\omega \in \mathcal{I}(\omega_*, \delta)$, meaning $\sigma(\omega)$ is simple on this interval. Moreover,

$$\sigma_r(\omega_r) - \underline{\sigma}_r(\omega_r) = \sigma(\omega_r) - \underline{\sigma}(\omega_r) \geq 3\zeta/4.$$

But $|\mathcal{I}(\omega_*, \delta)| = 2\delta$, so, by Weyl's theorem, we also have

$$\sigma_r(\omega) - \underline{\sigma}_r(\omega) \geq \{\sigma_r(\omega_r) - \underline{\sigma}_r(\omega_r)\} - 4\eta\delta \geq \zeta/4$$

for all $\omega \in \mathcal{I}(\omega_*, \delta)$. Consequently, $\sigma(\cdot)$ and $\sigma_r(\cdot)$ are analytic on $\mathcal{I}(\omega_*, \delta)$. Indeed, similar arguments show the more general inequalities $\sigma^{(j)}(\omega) - \sigma^{(j+1)}(\omega) \geq 3\zeta/4$ and $\sigma_r^{(j)}(\omega) - \sigma_r^{(j+1)}(\omega) \geq \zeta/4$ for each $\omega \in \mathcal{I}(\omega_*, \delta)$ and for $j = 2, \ldots, \min\{m, p\} - 1$.

Second, we show that the second derivatives of $\sigma(\cdot)$ and $\sigma_r(\cdot)$ are bounded away from zero on $\mathcal{I}(\omega_*, \delta)$. For the former, w.l.o.g. due to $\sigma''(\omega_*) \neq 0$, we simply consider δ small enough (much smaller than $|\sigma''(\omega_*)|$) so that

$$|\sigma''(\omega)| \geq \ell_1 \quad \forall \omega \in \mathcal{I}(\omega_*, \delta)$$

for some constant $\ell_1 \gg \delta > 0$. For the latter, for $j = 2, \dots, \min\{m, p\} - 1$, we obtain

$$3\eta|\omega_r - \omega_{r-1}| \leq 3\eta(|\omega_r - \omega_*| + |\omega_* - \omega_{r-1}|) \leq 6\eta\delta \leq \frac{3\zeta}{4} \leq \sigma^{(j)}(\omega_r) - \sigma^{(j+1)}(\omega_r),$$

so Lemma 3.2 implies

$$|\sigma_r''(\omega_r) - \sigma''(\omega_r)| \leq \mu|\omega_r - \omega_{r-1}| \leq 2\mu \max\{|\omega_r - \omega_*|, |\omega_{r-1} - \omega_*|\} \leq 2\delta\mu. \tag{18}$$

That is, $|\sigma_r''(\omega_r)| \geq |\sigma''(\omega_r)| - 2\delta\mu \geq \ell_1 - 2\delta\mu$. It follows that

$$|\sigma_r''(\omega)| \geq \ell_2 \quad \forall \omega \in \mathcal{I}(\omega_*, \delta)$$

for another constant $\ell_2 > 0$.

Third, we show that the absolute value of the third derivative of $\sigma_r(\cdot)$ is bounded from above by a quantity that does not depend on ω_1 on $\mathcal{I}(\omega_*, \delta)$. Repeating the arguments in the proof of Lemma 3.2, the second derivative $\sigma_r''(\cdot)$ is Lipschitz continuous on $\mathcal{I}(\omega_*, \delta)$ with a Lipschitz constant γ_1 that depends on η, φ, and ζ only. This in turn, together with the analyticity of $\sigma_r(\cdot)$ on $\mathcal{I}(\omega_*, \delta)$, implies

$$|\sigma_r'''(\omega)| \leq \gamma_1 \quad \forall \omega \in \mathcal{I}(\omega_*, \delta).$$

Part 2: We express $|\omega_{r+1} - \omega_*|$ in terms of $|\omega_r - \omega_*|$ and $|\omega_{r-1} - \omega_*|$, and conclude with the superlinear convergence result as desired. Analyticity of $\sigma(\omega)$ implies

$$0 = \sigma'(\omega_*) = \sigma'(\omega_r) + \int_0^1 \sigma''\left(\omega_r + t(\omega_* - \omega_r)\right)(\omega_* - \omega_r)\,\mathrm{d}t.$$

In the last equation, we employ $\sigma'(\omega_r) = \sigma_r'(\omega_r)$ (part (iii) of Lemma 3.1), divide both sides by $\sigma''(\omega_r)$, and reorganize to obtain

$$0 = \frac{\sigma_r'(\omega_r)}{\sigma''(\omega_r)} + (\omega_* - \omega_r) + \frac{1}{\sigma''(\omega_r)} \int_0^1 \left[\sigma''\left(\omega_r + t(\omega_* - \omega_r)\right) - \sigma''(\omega_r)\right](\omega_* - \omega_r)\,\mathrm{d}t. \tag{19}$$

In what follows, we exploit $\sigma''(\omega_r) \approx \sigma_r''(\omega_r)$ as a consequence of Lemma 3.2, and $\sigma_r'(\omega_r)/\sigma_r''(\omega_r) \approx -(\omega_{r+1} - \omega_r)$ as a consequence of $\sigma_r'(\omega_{r+1}) = 0$. These observations lead us to $\sigma_r'(\omega_r)/\sigma''(\omega_r) + (\omega_* - \omega_r) \approx (\omega_* - \omega_{r+1})$ in (19). Formally, an application of Taylor's theorem with Lagrange remainder to $\sigma_r'(\cdot)$ and optimality of ω_{r+1} with respect to $\sigma_r(\cdot)$ give rise to

$$0 = \sigma_r'(\omega_{r+1}) = \sigma_r'(\omega_r) + \sigma_r''(\omega_r)(\omega_{r+1} - \omega_r) + \frac{\sigma_r'''(\xi)}{2}(\omega_{r+1} - \omega_r)^2,$$

which can be rearranged as

$$\frac{\sigma_r'(\omega_r)}{\sigma_r''(\omega_r)} = -(\omega_{r+1} - \omega_r) - \frac{\sigma_r'''(\xi)}{2\sigma_r''(\omega_r)}(\omega_{r+1} - \omega_r)^2 \tag{20}$$

for some $\xi \in \mathcal{I}(\omega_*, \delta)$. By combining (19) and (20), we deduce

$$\begin{aligned} 0 = (\omega_* - \omega_{r+1}) + \left(\frac{1}{\sigma''(\omega_r)} - \frac{1}{\sigma_r''(\omega_r)}\right)\sigma'(\omega_r) - \frac{\sigma_r'''(\xi)}{2\sigma_r''(\omega_r)}(\omega_{r+1} - \omega_r)^2 \\ + \frac{1}{\sigma''(\omega_r)} \int_0^1 \left[\sigma''\left(\omega_r + t(\omega_* - \omega_r)\right) - \sigma''(\omega_r)\right](\omega_* - \omega_r)\mathrm{d}t, \end{aligned}$$

implying

$$
(21) \quad |\omega_* - \omega_{r+1}| \leq \left| \frac{\sigma_r''(\omega_r) - \sigma''(\omega_r)}{\sigma''(\omega_r)\sigma_r''(\omega_r)} \right| |\sigma'(\omega_r)| + \left| \frac{\sigma_r'''(\xi)}{2\sigma_r''(\omega_r)} \right| |\omega_{r+1} - \omega_r|^2 + \left| \frac{\gamma}{2\sigma''(\omega_r)} \right| |\omega_* - \omega_r|^2,
$$

where we used the Lipschitz continuity of $\sigma''(\cdot)$ on the interval $\mathcal{I}(\omega_*, \delta)$, in particular, we used the existence of a Lipschitz constant $\gamma > 0$ such that

$$
|\sigma''(\omega_r + t(\omega_* - \omega_r)) - \sigma''(\omega_r)| \leq \gamma t |\omega_* - \omega_r| \quad \forall t \in [0, 1].
$$

Finally, by Young's inequality, we have $|\omega_{r+1} - \omega_r|^2 \leq 2|\omega_{r+1} - \omega_*|^2 + 2|\omega_r - \omega_*|^2$. Thus, the expression (21) can be rewritten as

$$
(22) \qquad (1 - c_1\delta)|\omega_{r+1} - \omega_*| \leq c_2|\sigma_r''(\omega_r) - \sigma''(\omega_r)||\sigma'(\omega_r)| + c_3|\omega_r - \omega_*|^2,
$$

where $c_1 = \gamma_1/\ell_2$, $c_2 = 1/(\ell_1 \cdot \ell_2)$, and $c_3 = \gamma/(2\ell_1) + \gamma_1/\ell_2$. The term $|\sigma_r''(\omega_r) - \sigma''(\omega_r)|$ on the right in (22) is bounded from above by $\max\{|\omega_r - \omega_*|, |\omega_{r-1} - \omega_*|\}$ up to a constant by (18), whereas the term $|\sigma'(\omega_r)|$ on the right is bounded from above by $|\omega_r - \omega_*|$ up to a constant by the mean value theorem. If δ is sufficiently small, the term on the left-hand side of (22) can be bounded from below by $c_4|\omega_{r+1} - \omega_*|$ for some constant $c_4 > 0$. Hence, the result follows. □

4. Numerical experiments.

In this section, we report on the numerical results obtained by our MATLAB implementation of Algorithm 1.[1] We first describe a few important implementation details and the test setup. After that we report on the respective numerical results.

4.1. Implementation details and test setup.

At each iteration of Algorithm 1, the $\mathcal{L}_\infty$-norm of a reduced function needs to be computed in line 10. This global nonconvex optimization problem is solved by means of the approach due to Boyd and Balakrishnan for transfer functions of linear state-space systems [5] (and [3] for the case of descriptor systems), and by means of the algorithm in [28] for general $\mathcal{L}_\infty$-functions. The Boyd–Balakrishnan algorithm requires the solution of an eigenvalue problem of size twice the order of the original system, but these are fairly small eigenvalue problems which can be solved efficiently and robustly using well-established factorization approaches. A structure-preserving algorithm for this task has been implemented as a FORTRAN subroutine for which we have used a MEX file to call it from MATLAB.

Algorithm 1 is terminated in practice when the relative distance between ω_r and ω_{r-1} is less than a prescribed tolerance for some $r > 1$, or the number of iterations exceeds a specified integer. Formally, we terminate when

$$
|\omega_r - \omega_{r-1}| < \varepsilon \cdot \frac{1}{2} |\omega_r + \omega_{r-1}| \quad \text{or} \quad r > r_{\max}.
$$

For our numerical experiments, we set $\varepsilon = 10^{-6}$ and $r_{\max} = 30$.

Algorithm 1 converges locally. To reduce the possibility of stagnating at a local maximizer that is not a global maximizer, we initialize the algorithm with r_0 interpolation points $\omega_1, \ldots, \omega_{r_0}$, instead of only one. In our numerical experiments we have set $r_0 = 10$ as a default value, but there are more complicated examples that need a larger amount of initial interpolation points. For instance, the peec example (see

[1] This is available from http://www.math.tu-berlin.de/index.php?id=186267&L=1.

below for further details) needs 80 initial points. We distribute the initial interpolation points equidistantly on the imaginary axis with the imaginary parts located in the interval $[0, \omega_{\max}]$, where $\omega_{\max}$ is a problem-dependent parameter that is highly influenced by the location of the poles of H.

Another problem arises when the number of inputs and outputs is large. In this case, the dimension $\tilde{n}$ of the middle factor $\widetilde{D}(s)$ of $\widetilde{H}(s)$ will grow by $\min\{m, p\}$ in each interpolation step. To avoid a too fast growth of $\tilde{n}$, we have implemented an option in our implementation that allows us to update the projection spaces by using only the singular vectors corresponding the the largest singular value of $H(\mathrm{i}\omega_r)$. This means that in Algorithm 1, lines 11 and 13–17 disappear and line 12 is replaced by the following.

```
Compute the left and right singular vectors $w$ and $v$ of $\tilde{H}_{r-1}(\mathrm{i}\omega_r)$
corresponding to the largest singular value.
$\widetilde{V}_r \leftarrow D(\mathrm{i}\omega_r)^{-1}B(\mathrm{i}\omega_r)v$   and   $\widetilde{W}_r \leftarrow \left(w^* C(\mathrm{i}\omega_r)D(\mathrm{i}\omega_r)^{-1}\right)^*$.
```

Similar changes are also made in lines 2–8. Note that in this way we may lose the Hermite interpolation property of the maximum singular values, since in general we only have $\sigma(\omega_k) \le \sigma_r(\omega_k)$, $k = 1, \ldots, r$. Therefore, we also do not necessarily have local superlinear convergence. We have tested this option on the `mimo8x8_system`, `mimo28x28_system`, and `mimo46x46_system` examples, which have 8, 28, and 46 inputs and outputs, respectively. The approach works well on these examples. However, a more rigorous analysis of this remains an open problem.

In the next two subsections we report on the outcome of our numerical experiments. These have been performed on a machine with a 4 Intel Core 3.30 GHz i5-4590 CPUs and 16 GB RAM in MATLAB 9.0.0.341360 (R2016a) running on Linux version 3.12.67-64-default. First we test our algorithm on 33 linear systems taken from [32, 27, 15, 10] in section 4.2. The data of these examples is freely available on the websites of Rommes[2] and the SLICOT benchmark collection[3]. The first 13 of these examples are standard state-space models ($E = I_n$), the other ones are descriptor systems with singular E. All these examples have transfer functions in $\mathcal{H}_\infty^{p\times m}$, so in fact we compute the $\mathcal{H}_\infty$-norm. Furthermore, in section 4.3, we consider an example of a time-delay system provided in [2].

4.2. Results for descriptor systems. In this subsection, we compare the results with the ones generated by the approach in [4], which is based on structured pseudospectra and locating their rightmost points in the complex plane repeatedly. In this approach, perturbed transfer functions of the form

$$H_\Delta(s) = C(sE - (A + B\Delta C))^{-1}B$$

with $\Delta \in \mathbb{C}^{m\times p}$ are considered. There, a perturbation Δ of minimal spectral norm such that the perturbed transfer function H_Δ is not in $\mathcal{H}_\infty^{p\times m}$ is determined by a sequence of structured rank-1 perturbations.

Table 1 summarizes the results of the 33 numerical experiments. For all examples, the correct norm value has been found up to the termination tolerance. In this table, the number of additional iterations after the construction of the initial reduced function needed to retrieve the $\mathcal{L}_\infty$-norm by Algorithm 1 up to the prescribed relative

[2] See http://sites.google.com/site/rommes/software.

[3] See http://slicot.org/20-site/126-benchmark-examples-for-model-reduction.

TABLE 1
Numerical results for 33 test examples and comparison with the pseudospectral approach from [4].

						Computed $\mathcal{L}_\infty$-norm		Optimal frequency ω_{opt}		Time in s		
#	Example	n	m	p	n_{iter}	[4]	Algor. 1	[4]	Algor. 1	[4]	Algor. 1	Ratio
1	build	48	1	1	6	5.27633e−03	5.27633e−03	5.20608e+00	5.20608e+00	1.06	0.08	14.0
2	pde	84	1	1	1	1.08358e+01	1.08358e+01	0.00000e+00	0.00000e+00	0.84	0.03	27.4
3	CDplayer	120	2	2	1	2.31982e+06	2.31982e+06	2.25682e+01	2.25682e+01	0.90	0.02	41.6
4	iss	270	3	3	7	1.15887e−01	1.15887e−01	7.75093e−01	7.75093e−01	0.85	0.24	3.5
5	beam	348	1	1	1	4.55487e+03	4.55487e+03	1.04575e−01	1.04575e−01	11.06	0.08	135.8
6	S10PI_n1	528	1	1	7	3.97454e+00	3.97454e+00	7.53151e+03	7.53151e+03	0.79	0.08	10.3
7	S20PI_n1	1028	1	1	5	3.44317e+00	3.44317e+00	7.61831e+03	7.61831e+03	1.79	0.07	24.0
8	S40PI_n1	2028	1	1	7	3.34732e+00	3.34732e+00	6.95875e+03	6.95875e+03	1.95	0.14	13.6
9	S80PI_n1	4028	1	1	5	3.37016e+00	3.37016e+00	6.96149e+03	6.96149e+03	3.84	0.19	20.3
10	M10PI_n1	528	3	3	7	4.05662e+00	4.05662e+00	7.53181e+03	7.53181e+03	1.21	0.35	3.5
11	M20PI_n1	1028	3	3	12	3.87260e+00	3.87260e+00	5.06412e+03	5.06412e+03	1.42	0.85	1.7
12	M40PI_n1	2028	3	3	8	3.81767e+00	3.81767e+00	5.07107e+03	5.07107e+03	2.24	0.51	4.4
13	M80PI_n1	4028	3	3	9	3.80375e+00	3.80375e+00	5.07279e+03	5.07279e+03	3.82	0.83	4.6
14	peec	480	1	1	1	3.52624e−01	3.52610e−01	5.46349e+00	5.46349e+00	9.26	2.13	4.3
15	S10PI_n	682	1	1	7	3.97454e+00	3.97454e+00	7.53151e+03	7.53151e+03	1.03	0.09	11.3
16	S20PI_n	1182	1	1	5	3.44317e+00	3.44317e+00	7.61831e+03	7.61831e+03	1.90	0.08	24.7
17	S40PI_n	2182	1	1	7	3.34732e+00	3.34732e+00	6.95875e+03	6.95875e+03	2.12	0.15	14.6
18	S80PI_n	4182	1	1	5	3.37016e+00	3.37016e+00	6.96149e+03	6.96149e+03	3.96	0.20	20.1
19	M10PI_n	682	3	3	7	4.05662e+00	4.05662e+00	7.53181e+03	7.53181e+03	1.40	0.35	4.0
20	M20PI_n	1182	3	3	10	3.87260e+00	3.87260e+00	5.06412e+03	5.06412e+03	1.44	0.60	2.4
21	M40PI_n	2182	3	3	8	3.81767e+00	3.81767e+00	5.07107e+03	5.07107e+03	2.12	0.51	4.1
22	M80PI_n	4182	3	3	9	3.80375e+00	3.80375e+00	5.07279e+03	5.07279e+03	3.96	0.85	4.7
23	bips98_606	7135	4	4	1	2.01956e+02	2.01956e+02	3.81763e+00	3.81762e+00	14.18	0.66	21.4
24	bips98_1142	9735	4	4	1	1.60427e+02	1.60427e+02	4.93005e+00	4.93006e+00	29.25	0.83	35.4
25	bips98_1450	11305	4	4	1	1.97389e+02	1.97389e+02	5.64575e+00	5.64571e+00	26.16	0.94	27.9
26	bips07_1693	13275	4	4	1	2.04168e+02	2.04168e+02	5.53766e+00	5.53765e+00	66.59	1.07	62.3
27	bips07_1998	15066	4	4	2	1.97064e+02	1.97064e+02	6.39968e+00	6.39960e+00	40.37	1.63	24.8
28	bips07_2476	16861	4	4	2	1.89579e+02	1.89579e+02	5.88971e+00	5.88973e+00	64.88	1.96	33.1
29	bips07_3078	21128	4	4	1	2.09445e+02	2.09445e+02	5.55792e+00	5.55793e+00	35.18	2.08	16.9
30	xingo_afonso_itaipu	13250	1	1	2	4.05605e+00	4.05605e+00	1.09165e+00	1.09165e+00	14.38	0.56	24.6
31	mimo8x8_system	13309	8	8	2	5.34292e−02	5.34292e−02	1.03313e+00	1.03312e+00	26.74	1.27	21.0
32	mimo28x28_system	13251	28	28	3	1.18618e−01	1.18618e−01	1.07935e+00	1.07935e+00	24.78	2.62	9.5
33	mimo46x46_system	13250	46	46	3	2.05631e+02	2.05631e+02	1.07908e+00	1.07908e+00	36.84	3.76	9.8

Table 2

The errors of the iterates of Algorithm 1*, the ratios of the errors of the iterates, and the errors of the largest singular values at these iterates are listed for the* S80PI_n *example. Here, the shorthands $\sigma_r := \sigma_r(\omega_{r+1})$ and $\sigma_* := \sigma(\omega_*)$ are used. As the "exact" solution we have taken the one we obtain after iteration* 6*.*

Iteration # (r)	$\lvert\omega_{r+1} - \omega_*\rvert$	$\lvert\omega_{r+1} - \omega_*\rvert / \lvert\omega_r - \omega_*\rvert$	$\lvert\sigma_r - \sigma_*\rvert$
1 (Initial model)	6.718e+02	—	1.207e+02
2	2.595e+03	3.863e+00	4.388e+00
3	1.156e+01	4.455e−03	2.905e+00
4	6.571e−01	5.684e−02	8.662e−03
5	0	0	8.878e−09

tolerance $\varepsilon = 10^{-6}$ is denoted by n_{iter}. The order of the system, the input dimension, and the output dimension are denoted by n, m, p, respectively. It is evident from Table 1 that the correct value of the $\mathcal{H}_\infty$-norm is found by Algorithm 1 for each of the problems. In terms of the runtime, Algorithm 1 outperforms the pseudospectral approach. The ratios between the time required by the pseudospectral approach and that required by Algorithm 1 are listed in the last column of Table 1.

Finally, local superlinear convergence consistent with Theorem 3.3 is observed in all cases. Specifically, for the `S80PI_n` example, the errors of the iterates are reported in Table 2. Five additional iterations after the construction of the initial reduced function suffice to compute the $\mathcal{L}_\infty$-norm with a desired relative tolerance of $\varepsilon = 10^{-6}$. In fact, we see that once the algorithm started converging to a local maximizer, it only needs one or two more iterations until convergence.

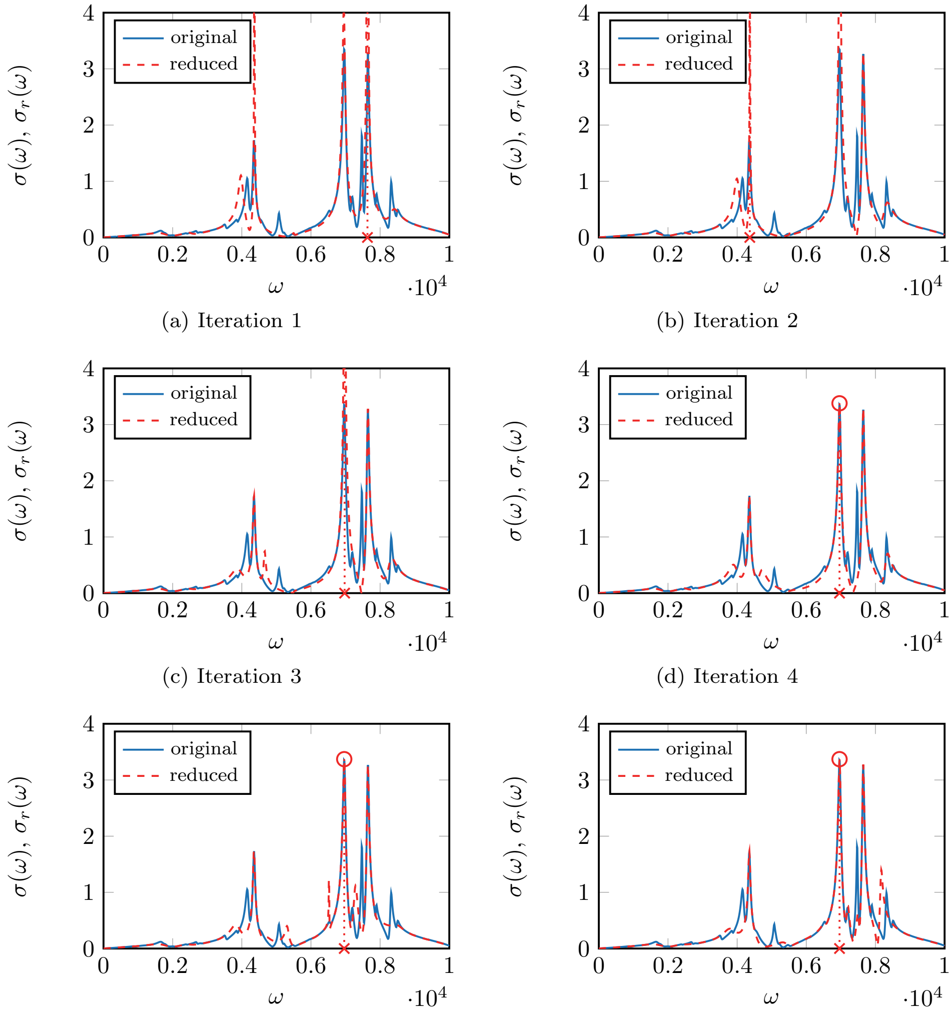
However, in some numerical examples we also observe that many iterations may be needed until convergence to a local maximizer takes place. If there exist many local maximizers of $\sigma(\cdot)$, then the algorithm often collects more global information of H in the beginning and only starts converging to a local maximizer after a certain number of iterations. In particular, this is the case in examples # 6–13 and # 15–22. An illustration of this fact is given in Figure 1, where the intermediate reduced functions for the `S80PI_n` example are depicted.

A further strong influence of the performance of Algorithm 1 is the number and location of the initial interpolation points. In the examples above, we have usually taken 10 initial points distributed equidistantly in an interval $[0, \omega_{\max}]$. Often, the algorithm also converges to the correct global maximizer with fewer initial points but then it may happen that more iterations are needed, since less global information is known. To illustrate this behavior we consider the `M80PI_n` example. For this example we generate interpolation points by the MATLAB command

```
options.initialPoints = linspace( 0.1, 10000, ninit );
```

and we let `ninit` grow from 1 to 30. For all 30 configurations, the correct norm value has been computed. The results are depicted in Figure 2. It can be seen that there is a certain trade-off between the number of initial interpolation points; a larger value of `ninit` may drastically reduce the number of additional iterations, but it also increases the subspace dimensions, which results in more effort for solving the small intermediate problems. For our example, values of about 20 initial interpolation points result in the best behavior (except for some smaller values, where the global optimizer has already been almost hit).

4.3. Results for time-delay systems. Next we test our approach on transfer functions of time-delay systems. Our experiments are performed on the following example taken from [2].



(a) Iteration 1

(b) Iteration 2

(c) Iteration 3

(d) Iteration 4

(e) Iteration 5

(f) Iteration 6

FIG. 1. *Intermediate reduced functions obtained by Algorithm* 1 *for the* S80PI_n *example. The original function is depicted in blue, while the reduced functions are represented by the dashed red lines. The red crosses and circles indicate the locations of the maximizers and the $\mathcal{L}_\infty$-norms of the reduced functions, respectively. In the first iteration (on the top left), the reduced model represents the initial model obtained from* 10 *equally spaced interpolation points.*

Example 4.1. Consider the delay system

$$Ex'(t) = A_0x(t) + A_1x(t-\tau) + Bu(t), \quad y(t) = Cx(t) \tag{23}$$

with $E = \theta I_n + T$, $A_0 = \frac{1}{\tau}(\frac{1}{\beta}+1)(T-\theta I_n)$, $A_1 = \frac{1}{\tau}(\frac{1}{\beta}-1)(T-\theta I_n)$, where T is the $n \times n$ matrix with ones on the subdiagonal and superdiagonal, as well as in the entries at position $(1,1)$ and (n,n), and zeros elsewhere. The scalars β and θ are parameters, and τ is the delay parameter.

We choose $\tau = 1$, $\beta = 0.01$, and $\theta = 5$. Additionally, we set $B = e_1 + e_2$, the sum of the first two columns of the $n \times n$ identity matrix, $C = B^*$, and experiment with various values of n.

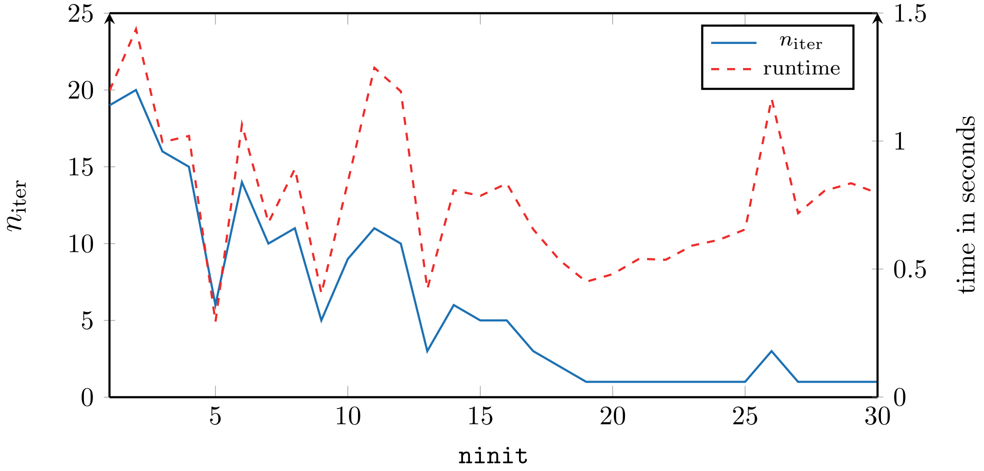


FIG. 2. *Behavior of Algorithm* 1 *on the* `M80PI_n` *example with respect to the number of initial interpolation points. The value* n_{iter} *refers to the number of additional iterations after the construction of the initial reduced function.*

Since Example 4.1 has the nonrational transfer function

$$H(s) = C\left(sE - A_0 - \mathrm{e}^{-s\tau}A_1\right)^{-1}B,$$

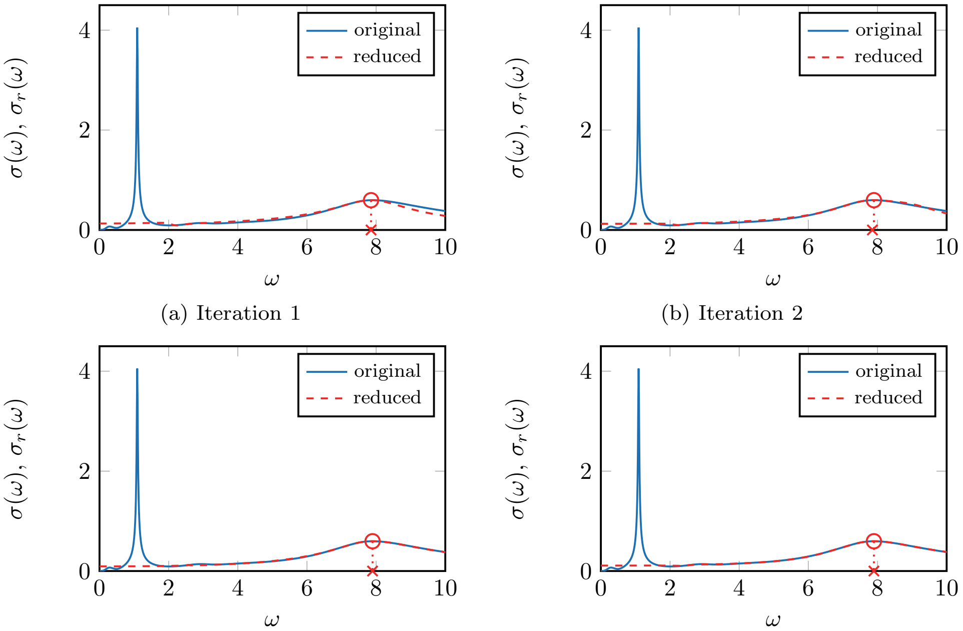
the Boyd–Balakrishnan algorithm cannot be applied here. Instead we use `eigopt` [28] to solve the small subproblems in Algorithm 1. The MATLAB package `eigopt` requires additional inputs. Specifically, a frequency interval in which the $\mathcal{L}_\infty$-norm is attained has to be supplied. We choose the interval $[0, 50]$, in which ω_* is located. In this interval, 8 local maxima of $\sigma(\cdot)$ can be found. Outside this interval, there exist infinitely many more such local maxima, but they result in much smaller maximum singular values. A second parameter the user has to supply is a global lower bound γ on the second derivative of $-\sigma(\cdot)$. In our example, the minimum of this second derivative is always about -93.08, so we choose $\gamma = -100$. The value of γ has a strong influence on the runtime; the lower the γ, the more piecewise quadratic support functions are constructed by `eigopt` which increases the computational complexity.

The runtimes and the runtime ratios between `eigopt` and Algorithm 1 are given in Table 3. For all values of n, `eigopt` and Algorithm 1 return the same (correct) value of the $\mathcal{L}_\infty$-norm, namely, $\|H\|_{\mathcal{L}_\infty} = 0.23766$. This value is attained for $\omega_* = 3.07547$. After the construction of the initial reduced transfer function, Algorithm 1 only needs one more iteration until convergence. The table also shows that Algorithm 1 is only more efficient for larger values of n. This is because the computation of H and its singular values becomes a dominant factor for larger n. For smaller values of n, Algorithm 1 carries out two calls of `eigopt`; for this reason, it needs almost twice the time to solve the original problem by a single run of `eigopt`.

4.4. Limitations of the method. As mentioned above, our algorithm converges only locally. (The same property holds for all other methods for large-scale $\mathcal{L}_\infty$-norm calculations to this date.) It is important to interpolate H at the parts of the imaginary axis that are close to the poles of H. If not enough interpolation points are taken, then the global maximizer of $\sigma(\cdot)$ may be missed. To illustrate this,

TABLE 3
Comparison of `eigopt` *and Algorithm* 1 *on Example* 4.1.

	Time in seconds		
n	`eigopt`	Algor. 1	ratio
100	0.92	1.83	0.51
300	0.98	1.86	0.53
1000	1.25	1.85	0.67
3000	1.76	1.86	0.95
10000	3.92	1.89	2.07
30000	10.47	2.02	5.19
100000	36.36	2.52	14.38
300000	113.10	3.80	29.75
1000000	403.13	9.09	44.34



(a) Iteration 1

(b) Iteration 2

(c) Iteration 1

(d) Iteration 2

FIG. 3. *Intermediate reduced functions obtained by Algorithm* 1 *for the* `xingo_afonso_itaipu` *example. The original function is depicted in blue, while the reduced functions are represented by the dashed red lines. The red crosses and circles indicate the locations of the maximizers and the* $\mathcal{L}_\infty$*-norms of the reduced functions, respectively. The initial reduced model (on the top left) is obtained from* 2 *interpolation points at* $\omega = 2.5$ *and* $\omega = 7.5$.

consider the `xingo_afonso_itaipu` example but with only two initial interpolation points 2.5i and 7.5i. With these points only, the global maximizer at 1.092 is not detected, instead the algorithm converges to the local maximizer at 7.897 which is not a global maximizer. The intermediate iterates are depicted in Figure 3. A remedy to this problem is to use more initial interpolation points for the initial iteration.

5. Concluding remarks. We have introduced an approach for the computation of the $\mathcal{L}_\infty$-norm of an $\mathcal{L}_\infty$-function of the form $H(s) = C(s)D(s)^{-1}B(s)$ in the large-scale setting, i.e., the middle factor is the inverse of a large-scale meromorphic matrix-valued function, and $C(s)$, $B(s)$ are meromorphic functions mapping to short-and-fat and tall-and-skinny matrices, respectively. Our approach is based on a subspace projection idea that is frequently used in model order reduction. More precisely, we approximate the given $\mathcal{L}_\infty$-function by a reduced function obtained by employing two-sided projections on the factors of the original $\mathcal{L}_\infty$-function. The middle factor of the resulting reduced function is of much smaller dimension. We compute the $\mathcal{L}_\infty$-norm of the reduced function by established methods. Then we expand the projection spaces by using the singular vectors of the original function at the point on the imaginary axis, where the $\mathcal{L}_\infty$-norm of the reduced function is attained. We have proven that our selection strategy for the subspaces leads to Hermite interpolation properties of the largest singular values of the original and reduced functions. These Hermite interpolation properties in turn give rise to a superlinear convergence with respect to the subspace dimension.

We have demonstrated on various numerical examples that our method can lead to substantial speedups compared to known methods. Moreover, it can be applied to a much larger class of functions such as transfer functions of delay systems. Thus, our method may lead to significant computational benefits in the field of $\mathcal{H}_\infty$-optimization.

Acknowledgments. The authors are grateful to two anonymous reviewers and Daniel Szyld for their invaluable feedback.

REFERENCES

[1] U. Baur, P. Benner, and L. Feng, *Model order reduction for linear and nonlinear systems: A system-theoretic perspective*, Arch. Comput. Methods Eng., 21 (2014), pp. 331–358, https://doi.org/10.1007/s11831-014-9111-2.

[2] C. Beattie and S. Gugercin, *Interpolatory projection methods for structure-preserving model reduction*, Systems Control Lett., 58 (2009), pp. 225–232, https://doi.org/10.1016/j.sysconle.2008.10.016.

[3] P. Benner, V. Sima, and M. Voigt, *$\mathcal{L}_\infty$-norm computation for continuous-time descriptor systems using structured matrix pencils*, IEEE Trans. Automat. Control, 57 (2012), pp. 233–238.

[4] P. Benner and M. Voigt, *A structured pseudospectral method for $\mathcal{H}_\infty$-norm computation of large-scale descriptor systems*, Math. Control Signals Systems, 26 (2014), pp. 303–338.

[5] S. Boyd and V. Balakrishnan, *A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its L_∞-norm*, Systems Control Lett., 15 (1990), pp. 1–7.

[6] N. A. Bruinsma and M. Steinbuch, *A fast algorithm to compute the H_∞-norm of a transfer function matrix*, Systems Control Lett., 14 (1990), pp. 287–293.

[7] A. Bunse-Gerstner, R. Byers, V. Mehrmann, and N. K. Nichols, *Numerical computation of an analytic singular value decomposition of a matrix valued function*, Numer. Math., 60 (1991), pp. 1–39.

[8] J. V. Burke, D. Henrion, A. S. Lewis, and M. L. Overton, *HIFOO – A MATLAB package for fixed-order controller design and H_∞ optimization*, in Proceedings of the 5th IFAC Symposium on Robust Control Design, Toulouse, France, Curran, Red Hook, NY, 2006, pp. 339–344.

[9] R. Byers, *A bisection method for measuring the distance of a stable matrix to the unstable matrices*, SIAM J. Sci. Stat. Comput., 9 (1988), pp. 875–881.

[10] Y. Chahlaoui and P. Van Dooren, *A Collection of Benchmark Examples for Model Reduction of Linear Time Invariant Dynamical Systems*, Technical report, SLICOT Working Note 2002–2, 2002.

[11] C. De Villemagne and R. E. Skelton, *Model reduction using a projection formulation*, Internat. J. Control., 40 (1987), pp. 2141–2169.

[12] N. H. Du, V. H. Linh, V. Mehrmann, and D. D. Thuan, *Stability and robust stability of linear time-invariant delay differential-algebraic equations*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 1631–1654.

[13] L. Feng and P. Benner, *Model order reduction for systems with non-rational transfer function arising in computational electromagnetics*, in Scientific Computing in Electrical Engineer-

ing SCEE 2008, J. Roos and L. R. J. Costa, eds., Math. Ind. 14, Springer, Berlin, 2010, pp. 512–522.

[14] M. A. Freitag, A. Spence, and P. Van Dooren, *Calculating the H_∞-norm using the implicit determinant method*, Linear Algebra Appl., 35 (2014), pp. 619–635.

[15] F. Freitas, J. Rommes, and N. Martins, *Gramian-based reduction method applied to large sparse power system descriptor models*, IEEE Trans. Power Syst., 23 (2008), pp. 1258–1270.

[16] S. Gugercin, T. Stykel, and S. Wyatt, *Model reduction of descriptor systems by interpolatory projection methods*, SIAM J. Sci. Comput., 35 (2013), pp. B1010–B1033.

[17] N. Guglielmi, M. Gürbüzbalaban, and M. L. Overton, *Fast approximation of the H_∞-norm via optimization over spectral value sets*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 709–737.

[18] S. Gumussoy and W. Michiels, *Computing $\mathcal{H}_\infty$ norms of time-delay systems*, in Proceedings of the Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Shanghai, China, IEEE, Piscataway, NJ, 2009, pp. 263–268.

[19] S. Gumussoy and W. Michiels, *Computation of extremum singular values and the strong H-infinity norm of SISO time-delay systems*, Automatica J. IFAC, 54 (2015), pp. 266–271.

[20] B. Haasdonk, *Reduced Basis Methods for Parametrized PDEs—A Tutorial Introduction for Stationary and Instationary Problems*, in Model Reduction and Approximation: Theory and Algorithms, Comput. Sci. Eng., P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, eds., SIAM, Philadelphia, 2017, pp. 65–136.

[21] D. Hinrichsen and A. J. Pritchard, *Stability radii of linear systems*, Systems Control Lett., 7 (1986), pp. 1–10.

[22] D. Hinrichsen and A. J. Pritchard, *Stability radius for structured perturbations and the algebraic Riccati equation*, Systems Control Lett., 8 (1986), pp. 105–113.

[23] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[24] F. Kangal, K. Meerbergen, E. Mengi, and W. Michiels, *A subspace method for large scale eigenvalue optimization*, SIAM J. Matrix Anal. Appl., to appear.

[25] M. Karow, *Geometry of Spectral Value Sets*, dissertation, Universität Bremen, Bremen, Germany, 2003.

[26] P. Lancaster, *On eigenvalues of matrices dependent on a parameter*, Numer. Math., 6 (1964), pp. 377–387.

[27] N. Martins, P. C. Pellanda, and J. Rommes, *Computation of transfer function dominant zeros with applications to oscillation damping control of large power systems*, IEEE Trans. Power Syst., 22 (2007), pp. 1657–1664.

[28] E. Mengi, E. A. Yildirim, and M. Kiliç, *Numerical optimization of eigenvalues of Hermitian matrix functions*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 699–724, https://doi.org/10.1137/130933472.

[29] T. Mitchell and M. L. Overton, *Fixed low-order controller design and H_∞ optimization for large-scale dynamical systems*, in Proceedings of the 8th IFAC Symposium on Robust Control Design, Bratislava, Slovakia, Elsevier, 2015, pp. 25–30.

[30] T. Mitchell and M. L. Overton, *Hybrid expansion-contraction: A robust scaleable method for approximating the H_∞ norm*, IMA J. Numer. Anal., 36 (2016), pp. 985–1014.

[31] F. Rellich, *Perturbation Theory of Eigenvalue Problems*, Notes Math. Appl., Gordon and Breach, New York, 1969.

[32] J. Rommes and N. Martins, *Efficient computation of multivariate transfer function dominant poles using subspace acceleration*, IEEE Trans. Power Syst., 21 (2006), pp. 1471–1483.

[33] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, Princeton, NJ, 2005.

[34] M. Voigt, *On Linear-Quadratic Optimal Control and Robustness of Differential-Algebraic Systems*, Logos, Berlin, 2015.

[35] A. Yousuff, D. A. Wagie, and R. E. Skelton, *Linear system approximation via covariance equivalent realizations*, J. Math. Anal. Appl., 106 (1985), pp. 91–115.

[36] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*, Prentice-Hall, Englewood Cliffs, NJ, 1996.