LARGE-SCALE AND GLOBAL MAXIMIZATION OF THE DISTANCE TO INSTABILITY[*]

EMRE MENGI[†]

Abstract. The larger the distance to instability from a matrix is, the more robustly stable the associated autonomous dynamical system is in the presence of uncertainties and typically the less severe transient behavior its solution exhibits. Motivated by these issues, we consider the maximization of the distance to instability of a matrix dependent on several parameters, a nonconvex optimization problem that is likely to be nonsmooth. In the first part we propose a globally convergent algorithm when the matrix is of small size and depends on a few parameters. In the second part we deal with the problems involving large matrices. We tailor a subspace framework that reduces the size of the matrix drastically. The strength of the tailored subspace framework is proven with a global convergence result as the subspaces grow and a superlinear rate-of-convergence result with respect to the subspace dimension.

Key words. eigenvalue optimization, maximin optimization, distance to instability, robust stability, subspace framework, large-scale optimization, global optimization, eigenvalue perturbation theory

AMS subject classifications. 65F15, 90C26, 93D09, 93D15, 49K35

DOI. 10.1137/18M1177019

1. Introduction. Our concern in this work is the maximization of the distance to instability of a matrix dependent on parameters over a set of admissible parameter values. We assume throughout that the matrix dependent on parameters can be represented in the form

$$A(x) := f_1(x)A_1 + f_2(x)A_2 + \cdots + f_\kappa(x)A_\kappa$$

for given $A_1, \ldots, A_\kappa \in \mathbb{C}^{n \times n}$ and $f_1, \ldots, f_\kappa : \Omega \to \mathbb{R}$ that are real analytic on their domain Ω, which is an open subset of $\mathbb{R}^d$. Moreover the distance to instability of such an $A(x)$ is defined by

$$
\begin{aligned}
\mathcal{D}(A(x)) \quad &:= \quad \min\{\|\Delta\|_2 \mid \exists z \in \mathbb{C}^+ \;\; \det(A(x) + \Delta - zI) = 0\} \\
&= \quad \min_{z \in \mathbb{C}^+} \sigma_{\min}(A(x) - zI) \\
&= \quad \begin{cases} \min_{\omega \in \mathbb{R}} \; \sigma_{\min}(A(x) - \omega iI) & \text{if } \Lambda(A(x)) \cap \mathbb{C}^+ = \emptyset, \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}
$$

where $\mathbb{C}^+ := \{z \in \mathbb{C} \mid \mathrm{Re}(z) \geq 0\}$ denotes the closed right-half of the complex plane, $\sigma_{\min}(\cdot)$ and $\Lambda(\cdot)$ denote the smallest singular value and spectrum, respectively, of their matrix arguments. The second equality above is a simple consequence of the Eckart–Young theorem [18, Theorem 2.5.3], whereas the third equality follows from the maximum modulus principle. The condition $\Lambda(A(x)) \cap \mathbb{C}^+ = \emptyset$ is an algebraic way of stating that the autonomous system $y' = A(x)y$ is asymptotically stable. If this condition is violated, the distance to instability of $A(x)$ is defined to be zero.

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

The problem at our hands, maximization of the distance to instability, can formally be expressed as

$$
(1) \quad
\begin{aligned}
\max_{x \in \widetilde{\Omega}} \mathcal{D}(A(x)) &= \max_{x \in \widetilde{\Omega}} \min_{z \in \mathbb{C}^+} \sigma_{\min}(A(x) - zI) \\
&= \max \left\{ \min_{\omega \in \mathbb{R}} \sigma_{\min}(A(x) - \omega iI) \ \middle| \ x \in \widetilde{\Omega} \ \text{s.t.} \ \Lambda(A(x)) \cap \mathbb{C}^+ = \emptyset \right\}
\end{aligned}
$$

with $\widetilde{\Omega}$ denoting a compact subset of Ω. One classical problem in control theory is stabilization by output feedback control, which, for given $A \in \mathbb{C}^{n \times n}, B \in \mathbb{C}^{n \times m}, C \in \mathbb{C}^{p \times n}$, involves finding a $K \in \mathbb{C}^{m \times p}$ such that all eigenvalues of $A + BKC$ are contained in the open left-half of the complex plane. A more robust stabilization procedure would attempt to minimize the spectral abscissa (i.e., the real part of the rightmost eigenvalue) of $A + BKC$ over all K. Even more robust approaches would aim to minimize the ε-pseudospectral abscissa (i.e., the real part of the rightmost point among the eigenvalues of all matrices within an ε neighborhood, for instance, with respect to the matrix 2-norm) of $A + BKC$ for a prescribed $\varepsilon > 0$, or maximize the distance to instability of $A + BKC$ [14]. The latter can be formally stated as

$$
(2) \quad \max_{K \in \mathbb{C}^{m \times p}} \mathcal{D}(A + BKC).
$$

If the entries of K are constrained to lie in prescribed boxes in the complex plane, then (2) falls into the scope of (1) that we consider here.

1.1. Literature. Optimization of spectral abscissa has been an active field of research since the beginning of the 2000s. It is observed in [12, 11] that the nonsmoothness at the optimal point is a common phenomenon. In particular, it is shown in these works that, for a particular affine family of matrices, the optimal matrix with the smallest spectral abscissa is a Jordan block, and this property remains to be true for slightly perturbed problems. These observations led to a concentrated effort put into the development of numerical algorithms for nonconvex and nonsmooth optimization. For this purpose, a gradient sampling algorithm is introduced in [13], its convergence is analyzed in [15]. An implementation of a hybrid algorithm based on this gradient sampling algorithm as well as BFGS, called HANSO, is made publicly available [36]. A special adaptation of this software HIFOO [10, 22, 4] for fixed-order $\mathcal{H}_\infty$ controller design has found various applications. HIFOO can be used to optimize spectral abscissa, pseudospectral abscissa, and distance to instability, but it would converge only to a locally optimal solution. The problem of stabilization by output feedback is referred to as one of the important open problems in control theory in [5]; indeed, it is well known that the problem of finding a K with box constraints on the entries of K such that $A + BKC$ has all of its eigenvalues on the open left-half of the complex plane is NP-hard [35]. HIFOO and the related ideas can be applied to find K minimizing the spectral abscissa, pseudospectral abscissa, or maximizing the distance to instability of $A + BKC$ [14], but again it would lead to a solution that is locally optimal. In a different direction, bundling techniques [2, 1] and spectral bundle methods [3] have been proposed for $\mathcal{H}_\infty$-synthesis and $\mathcal{H}_\infty$-norm minimization. More recently, sequential quadratic and linear programming techniques that take all of the eigenvalues (instead of only the rightmost eigenvalue) into account have been employed for spectral abscissa minimization [27]. All of these techniques in the literature are meant to find a locally optimal solution even if there are only a few optimization parameters. Furthermore, none of them is specifically designed for

large-scale problems; for instance, none of these approaches is meant for problems on the order of a few thousands.

The current state-of-the-art regarding the computation of the distance to instability, which for this work is the objective function to be maximized, is at a mature stage. There are very reliable numerical techniques that converge very quickly and that are meant for small- to medium-scale problems. All of these techniques [16, 6, 8] to compute the distance to instability of an $n \times n$ matrix A are based on repeatedly finding the level sets of the singular value function $f(\omega) := \sigma_{\min}(A - \omega \mathrm{i} I)$ by extracting the imaginary eigenvalues of a $2n \times 2n$ Hamiltonian matrix. For larger problems, a fixed-point iteration is proposed in [21], and a technique that operates on the roots of an implicitly defined determinant function is discussed in [17]. These numerical techniques are meant for larger problems and work directly on $n \times n$ problems, but they can get stagnated at local minimizers of the singular value function. A subspace framework is proposed in [26] to cope with large-scale distance to instability computations, and observed to converge quickly with respect to the subspace dimension.

1.2. Outline and contributions. We first deal with the maximization of the distance to instability when the matrices $A_1, \ldots, A_\kappa$ involved are of small size in section 2. In particular, we discuss how the algorithm in [34] can be adapted for this purpose. This results in Algorithm 1, which is globally convergent, unlike the methods employed in the literature, but aims at problems depending on only a few parameters.

The rest of the paper is devoted to large-scale problems when $A_1, \ldots, A_\kappa$ are of large size. Subspace frameworks based on one-sided restrictions of the matrix-valued function are proposed. At every step of the subspace frameworks the distance to instability is maximized for such a restricted problem; then the subspace is expanded with the addition of a singular vector at the maximizing parameter value so that Hermite interpolation properties hold between the full and the restricted distance to instability functions at this parameter value. We first present a basic framework along these lines, namely, Algorithm 2 in section 3, and later an extended version Algorithm 3 in section 4.2.3, which Hermite interpolates not only at the optimal parameter values but also at nearby points. A detailed convergence analysis for the subspace frameworks is carried out in section 4; remarkably, the global convergence of the subspace frameworks is established, and a superlinear rate-of-convergence with respect to the subspace dimension is deduced for the basic framework when $d = 1$ and for the extended framework for every d. The practical implication of these convergence results is that the frameworks are capable of computing global maximizers of $\mathcal{D}(A(x))$ over $x \in \widetilde{\Omega}$ with high accuracy by replacing $A(x)$ with matrix-valued functions of size much smaller. Efficient solutions of the restricted problems are addressed in section 5. The proposed subspace frameworks in sections 3 and 4 perform the inner minimization for the restricted problems on the right-hand side of the complex plane. Section 6 argues that these inner minimization problems can rather be performed on the imaginary axis if $A(x)$ is asymptotically stable for all $x \in \widetilde{\Omega}$. Finally, the performance of the proposed subspace frameworks in practice are illustrated on examples in section 7.

The literature lacks studies that address maximin or minimax optimization problems involving a prescribed eigenvalue of a large-scale matrix-valued function. To our knowledge, this is the first work that proposes subspace frameworks to deal with the large-scale nature of such problems; what is striking is the strong convergence results that we deduce for the subspace frameworks. The proposed subspace frameworks

and their convergence analyses set examples for various other minimax or maximin eigenvalue optimization problems, including the minimization of the $\mathcal{H}_\infty$-norm and the ε-pseudospectral abscissa for a prescribed ε. The approach for the small-scale problem is based on [34] and global piecewise quadratic estimators, whereas the subspace frameworks for the large-scale problem are inspired from [25]. However, all those previous works concern the minimization of the Jth largest eigenvalue, while the problem we deal with here has the maximin structure which introduces additional challenges.

The report [31] is a slightly longer version of this article. In particular, it includes a few additional numerical results omitted here for brevity.

2. Small-scale problems. The algorithm we discuss in this section is meant to compute a globally optimal solution when the problem has a few optimization parameters, e.g., $d = 1$ or $d = 2$. If there are more than a few parameters, one can for instance resort to HIFOO [10] and be content with a locally optimal solution. Let us first assume that $A(x)$ is asymptotically stable, i.e., $\Lambda(A(x)) \cap \mathbb{C}^+ = \emptyset$, for all $x \in \widetilde{\Omega}$. In this case, for the solution of (1), we could equivalently deal with

$$\max_{x \in \widetilde{\Omega}} \ [\mathcal{D}(A(x))]^2 \ = \ \max_{x \in \widetilde{\Omega}} \min_{\omega \in \mathbb{R}} \ \lambda(x, \omega),$$

where $\lambda(x, \omega)$ denotes the smallest eigenvalue of $M(x, \omega) := (A(x) - \mathrm{i}\omega I)^*(A(x) - \mathrm{i}\omega I)$. Let us define $\omega(x)$ implicitly by

$$\omega(x) := \arg\min_{\omega \in \mathbb{R}} \ \lambda(x, \omega).$$

In the case the global minimizer of the problem on the right is not unique, we define $\omega(x)$ arbitrarily as any global minimizer. Hence, the objective that we would like to maximize is

$$[\mathcal{D}(A(x))]^2 \ = \ \lambda(x, \omega(x)).$$

Let us consider an $\widetilde{x}$ where the global minimizer $\omega(\widetilde{x})$ of $\lambda(\widetilde{x}, \omega)$ over all ω is unique and the eigenvalue $\lambda(\widetilde{x}, \omega(\widetilde{x}))$ is simple. The eigenvalue function $\lambda(x, \omega)$ is twice continuously differentiable at $(x, \omega) = (\widetilde{x}, \omega(\widetilde{x}))$. Throughout this section, we assume that for every $\widetilde{x}$ such that the global minimizer $\omega(\widetilde{x})$ is unique and $\lambda(\widetilde{x}, \omega(\widetilde{x}))$ is simple, the property

$$(3) \qquad\qquad \partial^2 \lambda(x, \omega)/\partial\omega^2 \,|_{(x,\omega)=(\widetilde{x},\omega(\widetilde{x}))} \neq 0$$

holds. Observe that the second derivative above cannot be negative, since $\omega(\widetilde{x})$ is a minimizer of $\lambda(\widetilde{x}, \omega)$ over $\omega \in \mathbb{R}$. Hence, under assumption (3), we have

$$\partial^2 \lambda(x, \omega)/\partial\omega^2 \,|_{(x,\omega)=(\widetilde{x},\omega(\widetilde{x}))} > 0.$$

Now the implicit function theorem ensures that the function $\omega(x)$ is defined uniquely in a neighborhood of $\widetilde{x}$ and twice continuously differentiable in this neighborhood. This in turn implies that $[\mathcal{D}(A(x))]^2$ is twice continuously differentiable in this neighborhood of $\widetilde{x}$. Its derivatives satisfy the properties stated by the next theorem. For part (i) of the theorem, we refer to [34, Lemma 5.1], whereas part (ii) is immediate from the fact that $\partial\lambda(x, \omega)/\partial\omega \,|_{(x,\omega)=(\widetilde{x},\omega(\widetilde{x}))} = 0$ and the chain rule. Note that, in what follows, the derivatives $\lambda_x(\cdot), \omega'(x), \lambda_{x\omega}(\cdot)$ are row vectors, whereas $\nabla^2_{xx}\lambda(\cdot)$ denotes the $d \times d$ Hessian of $\lambda(\cdot)$ with respect to x only. The notations $\Phi'_-(\widetilde{v}), \Phi'_+(\widetilde{v})$ stand for the left-hand, right-hand derivatives of the univariate function $\Phi(v)$ at $\widetilde{v}$.

THEOREM 1. *The following hold for every $\widetilde{x} \in \mathbb{R}^d$:*

(i) *For every $p \in \mathbb{R}^d$, letting $\Phi : \mathbb{R} \to \mathbb{R}$, $\Phi(v) := [\mathcal{D}(A(\widetilde{x} + vp))]^2$, we have*
$\Phi'_-(0) \geq \Phi'_+(0)$.

(ii) *If $\widetilde{x}$ is such that the global minimizer $\omega(\widetilde{x})$ of $\lambda(\widetilde{x}, \omega)$ over all ω is unique and the eigenvalue $\lambda(\widetilde{x}, \omega(\widetilde{x}))$ is simple, then*
- $\nabla[\mathcal{D}(A(\widetilde{x}))]^2 = \lambda_x(x, \omega)^T |_{(x,\omega)=(\widetilde{x}, \omega(\widetilde{x}))}$, *and*
- $\nabla^2[\mathcal{D}(A(\widetilde{x}))]^2 = \left\{ \nabla^2_{xx}\lambda(x, \omega) + \lambda_{x\omega}(x, \omega)^T \omega'(x) \right\} |_{(x,\omega)=(\widetilde{x}, \omega(\widetilde{x}))}$.

The next result follows from part (i) of Theorem 1. Its proof is similar to that of Theorem 5.2 in [34]. Only here the result is in terms of upper envelopes for a smallest eigenvalue function, whereas the result in [34] introduces lower envelopes for a largest eigenvalue function. Here and elsewhere, $\lambda_{\max}(\cdot)$ refers to the largest eigenvalue of a Hermitian matrix argument.

THEOREM 2 (upper support function). *Let $x^{(k)}$ be a point such that the global minimizer $\omega(x^{(k)})$ of the eigenvalue function $\lambda(x^{(k)}, \omega)$ over all ω is unique and $\lambda(x^{(k)}, \omega(x^{(k)}))$ is simple. Furthermore, let γ satisfy $\lambda_{\max}(\nabla^2[\mathcal{D}(A(x))]^2) \leq \gamma$ for all x such that the global minimizer $\omega(x)$ is unique and $\lambda(x, \omega(x))$ is simple. For every $x \in \mathbb{R}^d$, we have*

$$(4) \quad \begin{aligned} [\mathcal{D}(A(x))]^2 &\leq \\ q(x; x^{(k)}) &:= [\mathcal{D}(A(x^{(k)}))]^2 + \left\{ \nabla[\mathcal{D}(A(x^{(k)}))]^2 \right\}^T (x - x^{(k)}) + \frac{\gamma}{2}\|x - x^{(k)}\|_2^2. \end{aligned}$$

We refer the function $q(x; x^{(k)})$ in (4) as the upper support function for $[\mathcal{D}(A(x))]^2$ about $x^{(k)}$. In [34], based on such support functions, a globally convergent optimization algorithm due to Breiman and Cutler [7] has been adapted for eigenvalue optimization. Here, we adapt that algorithm to maximize $[\mathcal{D}(A(x))]^2$ globally.

2.1. Analytical deduction of γ. Before spelling out the algorithm formally, let us elaborate on how one can obtain γ as in Theorem (2) analytically.

THEOREM 3. *The function $\mathcal{D}(A(x))$ satisfies*

$$\lambda_{\max}(\nabla^2[\mathcal{D}(A(\widetilde{x}))]^2) \leq \lambda_{\max}\left(\nabla^2_{xx}M(\widetilde{x}, \omega(\widetilde{x}))\right)$$

for all $\widetilde{x}$ such that the global minimizer $\omega(\widetilde{x})$ is unique and $\lambda(\widetilde{x}, \omega(\widetilde{x}))$ is simple, where

$$\nabla^2_{xx}M(x, \omega) := \begin{bmatrix} \frac{\partial^2 M(x,\omega)}{\partial x_1^2} & \frac{\partial^2 M(x,\omega)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 M(x,\omega)}{\partial x_1 \partial x_d} \\ \frac{\partial^2 M(x,\omega)}{\partial x_2 \partial x_1} & \frac{\partial^2 M(x,\omega)}{\partial x_2^2} & \cdots & \frac{\partial^2 M(x,\omega)}{\partial x_2 \partial x_d} \\ & & \ddots & \\ \frac{\partial^2 M(x,\omega)}{\partial x_d \partial x_1} & \frac{\partial^2 M(x,\omega)}{\partial x_d \partial x_2} & \cdots & \frac{\partial^2 M(x,\omega)}{\partial x_d^2} \end{bmatrix}.$$

Proof. The arguments are similar to those in the proof of Theorem 6.1 in [30]. Let $\widetilde{x}$ be a point such that the minimum $\omega(\widetilde{x})$ is unique. By part (ii) of Theorem 1,

$$(5) \quad \nabla^2[\mathcal{D}(A(\widetilde{x}))]^2 = \left\{ \nabla^2_{xx}\lambda(x, \omega) + \lambda_{x\omega}(x, \omega)^T \omega'(x) \right\} \Bigg|_{(x,\omega)=(\widetilde{x}, \omega(\widetilde{x}))}.$$

The implicitly defined function $\omega(x)$ satisfies

$$\partial \lambda(x, \omega(x))/\partial \omega = 0$$

for all x in a neighborhood of $\widetilde{x}$. Differentiating this equation with respect to x at $(x, \omega(x)) = (\widetilde{x}, \omega(\widetilde{x}))$ yields

$$\omega'(\widetilde{x}) = -\left\{ \frac{\lambda_{x\omega}(x,\omega)}{\lambda_{\omega\omega}(x,\omega)} \right\}\bigg|_{(x,\omega)=(\widetilde{x},\omega(\widetilde{x}))},$$

where the derivative in the denominator is positive due to (3). Now plug this expression for $\omega(\widetilde{x})$ in (5) to obtain

$$\nabla^2[\mathcal{D}(A(\widetilde{x}))]^2 = \left\{ \nabla_{xx}^2\lambda(x,\omega) - \frac{\lambda_{x\omega}(x,\omega)^T\lambda_{x\omega}(x,\omega)}{\lambda_{\omega\omega}(x,\omega)} \right\}\bigg|_{(x,\omega)=(\widetilde{x},\omega(\widetilde{x}))}.$$

The last expression and the formulas [28] for the second derivatives of $\lambda(x,\omega)$ imply

$$\lambda_{\max}\left(\nabla^2[\mathcal{D}(A(\widetilde{x}))]^2\right) \leq \lambda_{\max}\left(\nabla_{xx}^2\lambda(\widetilde{x},\omega(\widetilde{x}))\right)$$
$$= \lambda_{\max}\left\{ H(\widetilde{x},\omega(\widetilde{x})) + 2\sum_{j=1}^{n-1}\frac{1}{\lambda(\widetilde{x},\omega(\widetilde{x})) - \lambda_j(\widetilde{x},\omega(\widetilde{x}))}\Re(H_j(\widetilde{x},\omega(\widetilde{x}))) \right\},$$

where

$$[H(\widetilde{x},\omega(\widetilde{x}))]_{k,\ell} = v^*\left\{ \frac{\partial^2 M(\widetilde{x},\omega(\widetilde{x}))}{\partial x_k \partial x_\ell} \right\}v,$$
$$[H_j(\widetilde{x},\omega(\widetilde{x}))]_{k,\ell} = \left[v^*\left\{ \frac{\partial M(\widetilde{x},\omega(\widetilde{x}))}{\partial x_k} \right\}v_j \right]\left[v_j^*\left\{ \frac{\partial M(\widetilde{x},\omega(\widetilde{x}))}{\partial x_\ell} \right\}v \right],$$

and $\lambda_j(\widetilde{x},\omega(\widetilde{x}))$ denotes the jth largest eigenvalue of $M(\widetilde{x},\omega(\widetilde{x}))$, whereas v_j, v represent unit eigenvectors corresponding to $\lambda_j(\widetilde{x},\omega(\widetilde{x}))$, $\lambda(\widetilde{x},\omega(\widetilde{x}))$. As shown in the proof of [30, Theorem 6.1], the term $\Re(H_j(\widetilde{x},\omega(\widetilde{x})))$ is positive semidefinite implying

$$\lambda_{\max}\left(\nabla^2[\mathcal{D}(A(\widetilde{x}))]^2\right) \leq \lambda_{\max}(H(\widetilde{x},\omega(\widetilde{x}))) \leq \lambda_{\max}\left(\nabla_{xx}^2 M(\widetilde{x},\omega(\widetilde{x}))\right),$$

where for the last inequality we again refer to the proof of [30, Theorem 6.1]. $\square$

It follows from

$$\nabla_{xx}^2 M(\widetilde{x},\omega(\widetilde{x})) = \nabla^2\left[A(\widetilde{x})^*A(\widetilde{x})\right] + \omega(\widetilde{x})\cdot\nabla^2\left[\mathrm{i}A(\widetilde{x}) - \mathrm{i}A^*(\widetilde{x})\right],$$

and the inequality $|\omega(\widetilde{x})| \leq 2\|A(\widetilde{x})\|_2$ (see [37, Lemma 2.1]) that

$$\lambda_{\max}\left(\nabla_{xx}^2 M(\widetilde{x},\omega(\widetilde{x}))\right) \leq \left\| \nabla^2\left\{ \sum_{j=1}^{\kappa} f_j(\widetilde{x})A_j \right\}^*\left\{ \sum_{j=1}^{\kappa} f_j(\widetilde{x})A_j \right\} \right\|_2$$
$$+ 4\left\{ \sum_{j=1}^{\kappa} |f_j(\widetilde{x})|\|A_j\|_2 \right\}\left\{ \sum_{j=1}^{\kappa} \|\nabla^2 f_j(\widetilde{x})\|_2\|A_j\|_2 \right\}$$
$$\leq 2\left\{ \sum_{j=1}^{\kappa} \|\nabla f_j(\widetilde{x})\|_2\|A_j\|_2 \right\}^2$$
$$+ 6\left\{ \sum_{j=1}^{\kappa} |f_j(\widetilde{x})|\|A_j\|_2 \right\}\left\{ \sum_{j=1}^{\kappa} \|\nabla^2 f_j(\widetilde{x})\|_2\|A_j\|_2 \right\}.$$

Hence, any upper bound on

$$\max_{x \in \Omega} \ 2g_1(x)^2 + 6g_0(x)g_2(x) \quad \text{with} \ g_2(x) := \sum_{j=1}^{\kappa} \|\nabla^2 f_j(x)\|_2 \|A_j\|_2,$$

$$g_1(x) := \sum_{j=1}^{\kappa} \|\nabla f_j(x)\|_2 \|A_j\|_2, \ \ g_0(x) := \sum_{j=1}^{\kappa} |f_j(x)| \|A_j\|_2$$

is a theoretically sound choice for γ in Theorem 2.

The expression above may look complicated at first look; however, for instance, in the affine case when $A(x) = B_0 + \sum_{j=1}^{d} x_j B_j$, that is considered widely in the literature, it leads to the conclusion that

$$\gamma \ = \ 2 \left\{ \sum_{j=1}^{d} \|B_j\| \right\}^2$$

is a sound choice that can be used in Theorem 2. A slightly tighter bound in this affine case can be obtained by observing

$$(6) \quad \nabla_{xx}^2 M(x, \omega) \ = \ \begin{bmatrix} 2B_1^* B_1 & B_1^* B_2 + B_2^* B_1 & \dots & B_1^* B_d + B_d^* B_1 \\ B_2^* B_1 + B_1^* B_2 & 2B_2^* B_2 & & B_2^* B_d + B_d^* B_2 \\ \vdots & & \ddots & \\ B_d^* B_1 + B_1^* B_d & B_d^* B_2 + B_2^* B_d & & 2B_d^* B_d \end{bmatrix},$$

so γ can also be chosen as the largest eigenvalue of the matrix on the right-hand side above.

Remark 4. It is often assumed so far that $\lambda(\widetilde{x}, \omega)$ has $\omega(\widetilde{x})$ as its sole global minimizer, which is needed to ensure the twice continuous differentiability of $\mathcal{D}(A(x))$ at $\widetilde{x}$. This assumption is always violated if the matrices $A_1, \dots, A_\kappa$, hence the matrix-valued function $A(x)$ for all $x \in \Omega$, are real; in this case the singular value function has the symmetry $\sigma_{\min}(A(x) - \omega \mathrm{i} I) = \sigma_{\min}(A(x) + \omega \mathrm{i} I)$ for all $\omega \in \mathbb{R}$, implying if $\omega(x)$ is a global minimizer, so is $-\omega(x)$. This symmetry in the global minimizers in the real case does not cause nondifferentiability, as long as there is a unique nonnegative (and a unique nonpositive) global minimizer, as the symmetry is preserved for all x. Hence, all of the discussions and results in this section carry over to the real matrix-valued setting provided $\lambda(\widetilde{x}, \omega)$ has a unique nonnegative global minimizer over ω.

2.2. The algorithm. The algorithm that we employ for small-scale problems is borrowed from [7, 34]. It is presented in Algorithm 1 below for completeness. At the kth iteration, the piecewise quadratic function $\min\{q(x; x^{(j)}) \mid j = 1, \dots, k-1\}$ that lies above $[\mathcal{D}(A(x))]^2$ globally is maximized. Then the piecewise quadratic function is refined with the addition of a new quadratic piece, namely, the upper support function $q(x; x^{(k)})$ about the computed maximizer $x^{(k)}$. Every convergent subsequence of the sequence $\{x^{(k)}\}$ is guaranteed to converge to a global maximizer of $[\mathcal{D}(A(x))]^2$ [34, Theorem 8.1]. We observe in practice that this convergence occurs typically at a linear rate, but a formal proof of this observation is open.

Solution of the subproblem in line 4, that is, the maximization of the smallest of $k-1$ quadratic functions with constant curvature, turns out to be challenging. It is possible to pose this as a bunch of nonconvex quadratic programming problems. For a few parameters, these quadratic programming problems are tractable and can

be solved efficiently [7, 34]. In practice, we use `eigopt` [33], which is a MATLAB implementation of this algorithm.

Algorithm 1 Solution of small-scale problems.

Input: The matrix-valued function $A(x)$ of the form (1) and the feasible region $\widetilde{\Omega}$.
Output: The sequence $\{x^{(k)}\}$.
1: $x^{(1)} \leftarrow$ a random point in $\widetilde{\Omega}$.
2: Calculate $[\mathcal{D}(A(x^{(1)}))]^2$ and $\nabla[\mathcal{D}(A(x^{(1)}))]^2$.
3: **for** $k = 2, 3, \ldots$ **do**
4:    $x^{(k)} \leftarrow \arg\max_{x \in \widetilde{\Omega}} \min\{q(x; x^{(j)}) \mid j = 1, \ldots, k-1\}$.
5:    Calculate $[\mathcal{D}(A(x^{(k)}))]^2$ and $\nabla[\mathcal{D}(A(x^{(k)}))]^2$.
6: **end for**

2.3. General case without uniform stability. Consider the problem at our hands with the full generality, that is, consider

$$\max_{x \in \widetilde{\Omega}} [\mathcal{D}(A(x))]^2 = \max_{x \in \widetilde{\Omega}} \min_{\omega \in \mathbb{R}} \begin{cases} \lambda(x, \omega) & \text{if } \Lambda(A(x)) \cap \mathbb{C}^+ = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

The global property that the left-hand derivatives are greater than or equal to the right-hand derivatives indicated in part (i) of Theorem 1 is lost at x values where $A(x)$ has none of the eigenvalues on the open right-half of the complex plane, but one or more eigenvalues on the imaginary axis (for an illustrative example see the top left of Figure 1; see, in particular, the distance function over there depending on one parameter near 0, where the distance becomes zero). A consequence is that the quadratic function $q(x; \widetilde{x})$ constructed about $\widetilde{x}$ such that $A(\widetilde{x})$ is not asymptotically stable is not necessarily an upper support function, that is, $[\mathcal{D}(A(x))]^2 \leq q(x; \widetilde{x})$ does not necessarily hold for all $x \in \widetilde{\Omega}$. However, such points where asymptotic stability is lost are far away from global maximizers that we are seeking. Indeed, if γ is chosen large enough, the quadratic functions about these points still bound the distance to instability function from above locally in a neighborhood of each global maximizer. This correct representation around the global maximizers is sufficient for the convergence of the algorithm to the globally maximal value. Formally, it can be shown that there exists γ such that every convergent subsequence of the sequence $\{x^{(k)}\}$ generated by Algorithm 1 converges to a global maximizer of $[\mathcal{D}(A(x))]^2$ over $x \in \widetilde{\Omega}$. In the affine case, when $A(x) = B_0 + \sum_{j=1}^{\kappa} x_j B_j$, the choice of γ set equal to the largest eigenvalue of the matrix in (6) works well in practice in our experience.

2.4. Numerical results. We present numerical results on two sets of examples[1] both of which concern the stabilization by output feedback control problem. These results are obtained by applications of a MATLAB implementation of Algorithm 1 that is publicly available [32].

The first set concerns the turbo-generator example in [24, Appendix E], used as a test example also in [14]. This example involves the robust stabilization of $A + BKC$ over $K \in \mathbb{R}^{2 \times 2}$, where $A \in \mathbb{R}^{10 \times 10}, B \in \mathbb{R}^{10 \times 2}, C \in \mathbb{R}^{2 \times 10}$. Denoting the jth columns of B, C^T with b_j, c_j, we maximize
  (i) $\mathcal{D}(A + kb_2 c_2^T)$ over $k \in [-0.5, 0.1]$, and
  (ii) $\mathcal{D}(A + k_1 b_2 c_2^T + k_2 b_2 c_1^T)$ over $k_1, k_2 \in [-0.5, 0.1]$.

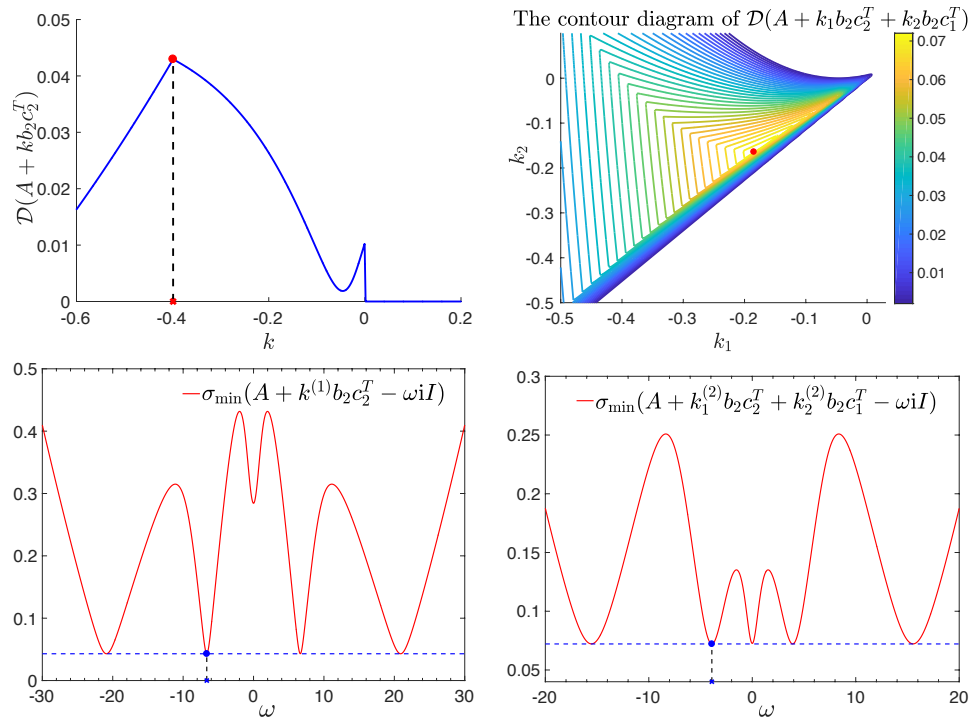[1] Available at http://home.ku.edu.tr/~emengi/software/max_di/Data_&_Updates.html.

FIG. 1. *The plots concern the turbo-generator example* [24, Appendix E]. *(Top Left) The plot of* $\mathcal{D}(A + kb_2c_2^T)$ *as a function of* $k \in [-0.5, 0.1]$. *The cross at the horizontal axis marks the computed global maximizer* $k^{(1)} = -0.3990$ *by Algorithm 1, whereas the dot marks* $(k^{(1)}, \mathcal{D}(A + k^{(1)}b_2c_2^T))$. *(Bottom Left) The plot of the singular value* $\sigma_{\min}(A + k^{(1)}b_2c_2^T - \omega \mathrm{i}I)$ *as a function of* ω. *The cross at the horizontal axis marks* $\omega_* = -6.6423$, *one of the* ω *values minimizing* $\sigma_{\min}(A + k^{(1)}b_2c_2^T - \omega \mathrm{i}I)$ *globally, while the dot marks* $(\omega_*, \sigma_{\min}(A + k^{(1)}b_2c_2^T - \omega_* \mathrm{i}I))$. *The dashed horizontal line is the highest horizontal line that bounds the graph of the singular value function from below. (Top Right) The contour diagram of* $\mathcal{D}(A + k_1 b_2 c_2^T + k_2 b_2 c_1^T)$ *with respect to* $k_1 \in [-0.5, 0.1]$ *in the horizontal axis and* $k_2 \in [-0.5, 0.1]$ *in the vertical axis. The global maximizer* $(k_1^{(2)}, k_2^{(2)}) = (-0.1847, -0.1644)$ *is marked with a dot. (Bottom Right) The plot of the singular value* $\sigma_{\min}(A + k_1^{(2)}b_2c_2^T + k_2^{(2)}b_2c_1^T - \omega \mathrm{i}I)$ *as a function of* ω. *The global minimizer marked with a cross in the horizontal axis is* $\omega_* = -3.9057$. *The dot and the dashed line stand for the same quantities as in bottom left.*

The original matrix A is stable with $\mathcal{D}(A) = 0.0077$, whereas $\mathcal{D}(A + k^{(1)}b_2c_2^T) = 0.0430$ at the global maximizer $k^{(1)} = -0.3990$ of (i), and $\mathcal{D}(A + k_1^{(2)}b_2c_2^T + k_2^{(2)}b_2c_1^T) = 0.0722$ at the global maximizer $k^{(2)} = (k_1^{(2)}, k_2^{(2)}) = (-0.1847, -0.1644)$ of (ii). The rightmost eigenvalues of A, $A + k^{(1)}b_2c_2^T$, $A + k_1^{(2)}b_2c_2^T + k_2^{(2)}b_2c_1^T$ are located at -0.2345, -0.3583 ± 6.6403, -0.5019, respectively. The results are displayed in Figure 1. In the one parameter case, the figure at top left indicates the existence of two maximizers, both of which are nonsmooth, and only one of which is a global maximizer. The algorithm correctly converges to the global maximizer. The two parameter case is highly nonsmooth, indeed, $\sigma_{\min}(A + k_1^{(2)}b_2c_2^T + k_2^{(2)}b_2c_1^T - \omega \mathrm{i}I)$ is minimized at three distinct nonpositive ω values, as shown at bottom right in Figure 1. This is reflected into the contour diagram at top right as steep changes close to the global maximizer.

The nonsmoothness at the maximizer does not always occur. In Figure 2 on the left, for random $A \in \mathbb{R}^{200 \times 200}$ and $b, c \in \mathbb{R}^{200}$, the plot of $\mathcal{D}(A + kbc^T)$ with respect
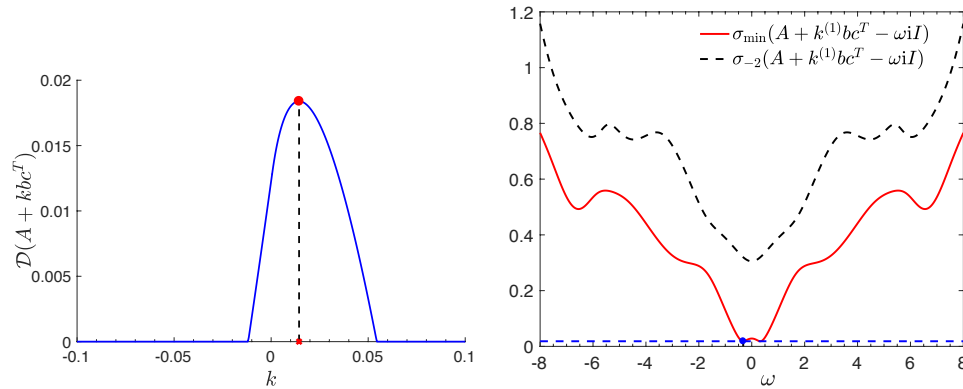
FIG. 2. (Left) *The plot of the distance to instability* $\mathcal{D}(A + kbc^T)$ *with respect to* $k \in [-0.1, 0.1]$ *for a random* $A \in \mathbb{R}^{200 \times 200}$ *and* $b, c \in \mathbb{R}^{200}$. *The distance function is smooth at the maximizer* $k^{(1)} = 0.0144$ *(marked with a cross on the horizontal axis), and attains the maximal value* $\mathcal{D}(A + k^{(1)}bc^T) = 0.0184$. *(Right) The plots of the smallest and second smallest singular value functions* $\sigma_{\min}(A + k^{(1)}bc^T - \omega iI)$ *(solid curve) and* $\sigma_{-2}(A + k^{(1)}bc^T - \omega iI)$ *(dashed curve) with respect to* ω. *The point* $\omega_* = -0.3231$ *is the unique negative global minimizer of the smallest singular value function, where the smallest singular value takes the value* 0.0184.

TABLE 1
γ *values for the the turbo-generator example, and the random* 200×200 *example.*

	$d = 1$	$d = 2$
Turbo-generator	137	291
Random 200×200	72836	-

to $k \in [-0.1, 0.1]$ is illustrated. The distance function is smooth at the computed maximizer $k^{(1)} = 0.0144$. Indeed, as depicted on the right in Figure 2, the singular value function $\sigma_{\min}(A + k^{(1)}bc^T - \omega iI)$ attains its minimum at a unique negative ω value. Our numerical experiments indicate both the smooth and the nonsmooth maximizers are possible. Based on our numerical experiments, it is not possible to call one of these cases generic and the other nongeneric.

In each of these examples, γ is set equal to the largest eigenvalue of the matrix in (6). The precise values are listed in Table 1.

3. A subspace framework for large-scale problems. The ideas here are inspired by [25]. But the eigenvalue optimization problems considered in that work concern minimization or maximization of the Jth largest eigenvalue, whereas the problem here, namely, (1), is a maximin problem. At first glance, this looks challenging, but it turns out that the ideas over there can be extended to our setting.

Our projected reduced problems will be of the form

$$(7) \qquad \max_{x \in \Omega} \min_{z \in \mathbb{C}^+} \sigma_{\min}(A^V(x) - zV),$$

where

$$(8) \qquad A^V(x) \; := \; f_1(x)A_1 V + f_2(x)A_2 V + \cdots + f_\kappa(x)A_\kappa V$$

and $V \in \mathbb{C}^{n \times \ell}$ with $n \gg \ell$ has orthonormal columns. Hence we resort to one-sided projections. Since the problem at hand is non-Hermitian, it may be tempting

to use two-sided projections. Unfortunately, it is difficult to conclude with strong convergence properties for such a two-sided framework and its practical reliability is also in doubt. This is basically due to the loss of a monotonicity property discussed in the next section with the two-sided projections.

The subspace framework is presented formally in Algorithm 2 below, where we use the notation

$$(9) \qquad \mathcal{D}^{\mathcal{V}}(A(x)) \quad := \quad \min_{z \in \mathbb{C}^+} \ \sigma_{\min}(A^V(x) - zV),$$

$\mathcal{V} := \mathrm{Col}(V)$ (i.e., $\mathcal{V}$ is the column space of V), and $\mathrm{orth}(C)$ stands for a matrix whose columns form an orthonormal basis for the column space of the matrix C. Note that when $\mathcal{V} = \mathbb{C}^n$ and $A(x)$ is asymptotically stable (i.e., when $\Lambda(A(x)) \cap \mathbb{C}^+ = \emptyset$), the minimization problem in (9) is attained on the imaginary axis due to the maximum modulus principle and we have $\mathcal{D}^{\mathcal{V}}(A(x)) = \mathcal{D}(A(x))$. But this attainment property on the imaginary axis is not necessarily true in the rectangular case when $\mathcal{V}$ is a strict subspace of $\mathbb{C}^n$.

At every iteration, the subspace framework first solves a projected small-scale problem. For this purpose, we benefit from the ideas in the previous section if there are only a few parameters; some details are discussed in section 5. If the problem depends on more than a few parameters, then a viable choice for the solution of the small-scale problem is HIFOO [10]. Throughout the rest, we assume that the small-scale problem is solved globally. At a maximizer of this small-scale problem, if the full problem is asymptotically stable, then we compute the distance to instability of the full problem, in particular, we retrieve the ω value where this distance is attained. We expand the subspace with the addition of a right singular vector corresponding to the smallest singular value at the optimal values of x and ω. On the other hand, if the full problem is not asymptotically stable at the maximizer of the small-scale problem, then the full problem at the maximizer has an eigenvalue on the closed right-half of the complex plane. The subspace is expanded with the inclusion of an eigenvector corresponding to such an eigenvalue.

Remark 5. It may sound plausible to define the reduced problem (7) so that the inner minimization problem is over the imaginary axis, that is, $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is defined by $\min_{\omega \in \mathbb{R}} \sigma_{\min}(A^{V_\ell}(x) - \mathrm{i}\omega V_\ell)$. Some difficulty arises with this approach if the full problem is unstable at the maximizer of the reduced problem. As we shall see, Algorithm 2 is designed in a way so that interpolation properties hold between the full and reduced problems at the maximizers of the reduced problems. Without these interpolation properties, the essential features of the framework such as its global and quick convergence are lost. It does not seem possible to fulfill these interpolation properties at a maximizer of the reduced problem where the full problem is unstable if the inner minimization problem is performed over the imaginary axis. However, if the full problem $A(x)$ is uniformly stable at all $x \in \widetilde{\Omega}$, then this idea of restricting the inner minimization to the imaginary axis works well; this is discussed in more detail in section 6.

3.1. Theoretical properties of the subspace framework. We first present two results concerning the subspace framework that will play prominent roles in the convergence analysis. The first one is a monotonicity result regarding $\mathcal{D}^{\mathcal{V}}(A(x))$ with respect to the subspace $\mathcal{V}$. We note once again that when $\mathcal{V} = \mathbb{C}^n$, we have $\mathcal{D}^{\mathcal{V}}(A(x)) = \mathcal{D}(A(x))$.

Algorithm 2 The subspace framework.

Input: The matrix-valued function $A(x)$ of the form (1) with the feasible region $\widetilde{\Omega}$.
Output: The sequences $\{x^{(\ell)}\}$, $\{z^{(\ell)}\}$.
1: $x^{(1)} \leftarrow$ a random point in $\widetilde{\Omega}$.
2: **if** $\Lambda(A(x^{(1)})) \cap \mathbb{C}^+ = \emptyset$ **then**
3:     $z^{(1)} \leftarrow \mathrm{i} \cdot \arg\min_{\omega \in \mathbb{R}} \ \sigma_{\min}(A(x^{(1)}) - \omega \mathrm{i} I)$.
4:     $V_1 \leftarrow$ a unit right singular vector corresponding to $\sigma_{\min}(A(x^{(1)}) - z^{(1)} I)$.
5: **else**
6:     $z^{(1)} \leftarrow$ an eigenvalue in $\Lambda(A(x^{(1)})) \cap \mathbb{C}^+$.
7:     $V_1 \leftarrow$ a unit eigenvector corresponding to the eigenvalue $z^{(1)}$ of $A(x^{(1)})$.
8: **end if**
9: $\mathcal{V}_1 \leftarrow \mathrm{span}\{V_1\}$.
10: **for** $\ell = 1, 2, \ldots$ **do**
11:     $x^{(\ell+1)} \leftarrow \arg\max_{x \in \widetilde{\Omega}} \mathcal{D}^{\mathcal{V}_\ell}(A(x))$.
12:     **if** $\Lambda(A(x^{(\ell+1)})) \cap \mathbb{C}^+ = \emptyset$ **then**
13:        $z^{(\ell+1)} \leftarrow \mathrm{i} \cdot \arg\min_{\omega \in \mathbb{R}} \ \sigma_{\min}(A(x^{(\ell+1)}) - \omega \mathrm{i} I)$.
14:        $v_{\ell+1} \leftarrow$ a right singular vector corresponding to $\sigma_{\min}(A(x^{(\ell+1)}) - z^{(\ell+1)} I)$.
15:     **else**
16:        $z^{(\ell+1)} \leftarrow$ an eigenvalue in $\Lambda(A(x^{(\ell+1)})) \cap \mathbb{C}^+$.
17:        $v_{\ell+1} \leftarrow$ an eigenvector corresponding to the eigenvalue $z^{(\ell+1)}$ of $A(x^{(\ell+1)})$.
18:     **end if**
19:     $V_{\ell+1} \leftarrow \mathrm{orth}\left(\begin{bmatrix} V_\ell & v_{\ell+1} \end{bmatrix} \right)$ and $\mathcal{V}_{\ell+1} \leftarrow \mathrm{Col}(V_{\ell+1})$.
20: **end for**

THEOREM 6 (monotonicity). *For two subspaces $\mathcal{V}, \mathcal{W}$ of $\mathbb{C}^n$ such that $\mathcal{V} \subseteq \mathcal{W}$, we have*

$$\sigma_{\min}(A(x) - zI) \ \leq \ \sigma_{\min}(A^W(x) - zW) \ \leq \ \sigma_{\min}(A^V(x) - zV) \ \ \forall z \in \mathbb{C}^+ \ \forall x \in \Omega$$

and

$$\mathcal{D}(A(x)) \ \leq \ \mathcal{D}^{\mathcal{W}}(A(x)) \ \leq \ \mathcal{D}^{\mathcal{V}}(A(x)) \ \ \forall x \in \Omega,$$

where V, W are matrices with columns forming orthonormal bases for $\mathcal{V}, \mathcal{W}$.

Proof. Recall that $\mathcal{D}^{\mathcal{V}}(A(x)) = \min_{z \in \mathbb{C}^+} \sigma_{\min}(A^V(x) - zV)$ and $\mathcal{D}^{\mathcal{W}}(A(x)) = \min_{z \in \mathbb{C}^+} \sigma_{\min}(A^W(x) - zW)$. Additionally, since $\mathcal{W} \supseteq \mathcal{V}$, we have

$$\begin{aligned} \sigma_{\min}(A^W(x) - zW) &= \min\{\|(A(x) - zI)w\|_2 \mid w \in \mathcal{W}, \ \|w\|_2 = 1\} \\ &\leq \min\{\|(A(x) - zI)w\|_2 \mid w \in \mathcal{V}, \ \|w\|_2 = 1\} \\ &= \sigma_{\min}(A^V(x) - zV) \end{aligned}$$

for all $z \in \mathbb{C}^+$. These assertions in turn imply $\mathcal{D}^{\mathcal{W}}(A(x)) \leq \mathcal{D}^{\mathcal{V}}(A(x))$. In a similar way, since $\mathbb{C}^n \supseteq \mathcal{W}$, we have $\sigma_{\min}(A(x) - zI) \leq \sigma_{\min}(A^W(x) - zW)$ for all $z \in \mathbb{C}^+$ and $\mathcal{D}(A(x)) \leq \mathcal{D}^{\mathcal{W}}(A(x))$. $\quad\square$

The next result concerns Algorithm 2. It establishes interpolation properties between $\mathcal{D}(A(x))$ and $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$. These interpolation properties also extend to first derivatives of these distance functions, but we elaborate on that issue later when we analyze the rate-of-convergence.

THEOREM 7 (interpolation).   *The following hold regarding Algorithm* 2 *for a given* $\ell \in \mathbb{Z}^+$ *and* $j = 1, \ldots, \ell$:

$$
\begin{aligned}
\mathcal{D}(A(x^{(j)})) &= \sigma_{\min}(A(x^{(j)}) - z^{(j)}I) \\
&= \sigma_{\min}(A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell) = \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(j)})).
\end{aligned}
\tag{10}
$$

In particular, if $A(x^{(j)})$ *is asymptotically stable, then the globally smallest value of* $\sigma_{\min}(A^{V_\ell}(x^{(j)}) - zV_\ell)$ *over all* $z \in \mathbb{C}^+$ *is attained on the imaginary axis.*

Proof. By Theorem 6, we have $\sigma_{\min}(A(x^{(j)}) - z^{(j)}I) \leq \sigma_{\min}(A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell)$ and $\mathcal{D}(A(x^{(j)})) \leq \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(j)}))$. Additionally, since $v_j \in \mathcal{V}_\ell$, there exists α_j of unit length such that $v_j = V_\ell \alpha_j$, which implies

$$
\begin{aligned}
\mathcal{D}(A(x^{(j)})) &= \sigma_{\min}(A(x^{(j)}) - z^{(j)}I) \\
&= \|(A(x^{(j)}) - z^{(j)}I)v_j\|_2 = \|(A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell)\alpha_j\|_2 \\
&\geq \sigma_{\min}(A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell) \geq \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(j)})),
\end{aligned}
\tag{11}
$$

regardless of whether $A(x^{(j)})$ is asymptotically stable or not. Specifically, when $A(x^{(j)})$ is not asymptotically stable, then $\mathcal{D}(A(x^{(j)})) = 0$ by definition, and all other terms in (11) also vanish. This proves (10).

If $A(x^{(j)})$ is asymptotically stable, then $z^{(j)}$ as defined in Algorithm 2 is purely imaginary and (11) indicates that

$$
\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(j)})) = \min_{z \in \mathbb{C}^+} \sigma_{\min}(A^{V_\ell}(x^{(j)}) - zV_\ell) = \sigma_{\min}(A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell). \qquad \square
$$

4. Convergence analysis of the subspace framework. We carry out the convergence analysis in the infinite dimensional setting. Formally, this means that $A_1, \ldots, A_\kappa$ now become linear bounded operators acting on $\ell_2(\mathbb{N})$, the Hilbert space of square summable complex infinite sequences with the inner product $\langle v, w \rangle := \sum_{j=0}^{\infty} \overline{v}_j w_j$ and the norm $\|v\| := \sqrt{\sum_{j=0}^{\infty} |v_j|^2}$. Moreover some of the arguments below refer to the operator norms defined by

$$
\begin{aligned}
\|A_j\| &:= \sup\{\|A_j v\| \mid v \in \ell_2(\mathbb{N}) \text{ such that } \|v\| = 1\} \text{ for } j = 1, \ldots, \kappa, \\
\|A(x)\| &:= \sup\{\|A(x)v\| \mid v \in \ell_2(\mathbb{N}) \text{ such that } \|v\| = 1\}.
\end{aligned}
$$

We assume $A(x)$ for all x has countably many singular values each with finite multiplicity, and 0 is not an accumulation point of these singular values. Compact operators have countably many singular values, but their singular values accumulate at 0, indeed this is the sole accumulation point of the singular values. Hence, for instance, the singular values of $A(x) = \widetilde{A}(x) + \tau I$ with $\widetilde{A}(x)$ denoting any compact operator and τ any nonzero scalar satisfy the assumptions; the singular values are countably many, do not accumulate at 0, additionally each singular value has finite multiplicity.

Throughout this section, we use the notations

$$
\sigma^{\mathcal{V}}(x, z) := \sigma_{\min}(A^V(x) - zV) \quad \text{and} \quad \sigma(x, z) := \sigma_{\min}(A(x) - zI)
$$

for V whose columns form an orthonormal basis for the subspace $\mathcal{V}$, and for $z \in \mathbb{C}^+$, $x \in \Omega$, so that $\mathcal{D}^{\mathcal{V}}(A(x)) = \min_{z \in \mathbb{C}^+} \sigma^{\mathcal{V}}(x, z)$.

4.1. Global convergence. The main result of this subsection, that is, Theorem 9, establishes the global convergence of Algorithm 2, which turns out to be a consequence of the monotonicity and the interpolation properties of the previous section,[2] as well as a Lipschitz continuity property. The latter is formally stated and proven below, where

$$\delta_{\mathcal{Z}}^{\mathcal{V}}(A(x)) := \inf\{\sigma^{\mathcal{V}}(x,z) \mid z \in \mathcal{Z}\}$$

for a given set $\mathcal{Z} \subseteq \mathbb{C}^+$.

LEMMA 8 (uniform Lipschitz continuity). *There exists a constant ζ such that for all subspaces $\mathcal{V}$ the following hold:*

(i) *For all $z \in \mathbb{C}^+$, $|\sigma^{\mathcal{V}}(\widetilde{x},z) - \sigma^{\mathcal{V}}(\widehat{x},z)| \leq \zeta\|\widetilde{x} - \widehat{x}\|_2 \quad \forall \widetilde{x}, \widehat{x} \in \Omega$.*

(ii) *For any given set $\mathcal{Z} \subseteq \mathbb{C}^+$, $|\delta_{\mathcal{Z}}^{\mathcal{V}}(A(\widetilde{x})) - \delta_{\mathcal{Z}}^{\mathcal{V}}(A(\widehat{x}))| \leq \zeta\|\widetilde{x} - \widehat{x}\|_2 \quad \forall \widetilde{x}, \widehat{x} \in \Omega$.*

(iii) *$|\mathcal{D}^{\mathcal{V}}(A(\widetilde{x})) - \mathcal{D}^{\mathcal{V}}(A(\widehat{x}))| \leq \zeta\|\widetilde{x} - \widehat{x}\|_2 \quad \forall \widetilde{x}, \widehat{x} \in \Omega$.*

Proof. (i) Let V be such that its columns form an othonormal basis for $\mathcal{V}$. Weyl's theorem [23, Theorem 4.3.1] implies

$$|\sigma^{\mathcal{V}}(\widetilde{x},z) - \sigma^{\mathcal{V}}(\widehat{x},z)| \leq \|A^V(\widetilde{x}) - A^V(\widehat{x})\|$$

$$\leq \sum_{j=1}^{\kappa} |f_j(\widetilde{x}) - f_j(\widehat{x})|\|A_j\|$$

for all $\widetilde{x}, \widehat{x} \in \Omega$. But since each f_j is real analytic, it is also Lipschitz continuous amounting to the existence of constants γ_j such that $|f_j(\widetilde{x}) - f_j(\widehat{x})| \leq \gamma_j\|\widetilde{x} - \widehat{x}\|_2$. It follows that

$$|\sigma^{\mathcal{V}}(\widetilde{x},z) - \sigma^{\mathcal{V}}(\widehat{x},z)| \leq \left(\sum_{j=1}^{\kappa} \gamma_j\|A_j\|\right)\|\widetilde{x} - \widehat{x}\|_2 \quad \forall \widetilde{x}, \widehat{x} \in \Omega$$

as desired.

(ii) Without loss of generality, we assume that the set $\mathcal{Z}$ is closed, as otherwise the argument below applies by replacing $\mathcal{Z}$ with $\mathrm{Cl}(\mathcal{Z})$, that is, the closure of $\mathcal{Z}$, since $\delta_{\mathcal{Z}}^{\mathcal{V}}(A(x)) = \delta_{\mathrm{Cl}(\mathcal{Z})}^{\mathcal{V}}(A(x))$. It turns out that, due to the assumption that $\mathcal{Z}$ is closed and the fact $\sigma^{\mathcal{V}}(x,z) \to \infty$ as $|z| \to \infty$, the quantity $\delta_{\mathcal{Z}}^{\mathcal{V}}(A(x))$ must be attained at some $z \in \mathcal{Z}$.

Now consider any two points $\widetilde{x}, \widehat{x} \in \Omega$ for which we must have

$$\delta_{\mathcal{Z}}^{\mathcal{V}}(A(\widetilde{x})) = \min_{z \in \mathcal{Z}} \sigma^{\mathcal{V}}(\widetilde{x},z) \quad \text{and} \quad \delta_{\mathcal{Z}}^{\mathcal{V}}(A(\widehat{x})) = \min_{z \in \mathcal{Z}} \sigma^{\mathcal{V}}(\widehat{x},z).$$

Without loss of generality, assume $\delta_{\mathcal{Z}}^{\mathcal{V}}(A(\widetilde{x})) \geq \delta_{\mathcal{Z}}^{\mathcal{V}}(A(\widehat{x}))$. Also, let $\widetilde{z}, \widehat{z}$ be such that

$$\delta_{\mathcal{Z}}^{\mathcal{V}}(A(\widetilde{x})) = \sigma^{\mathcal{V}}(\widetilde{x},\widetilde{z}) \quad \text{and} \quad \delta_{\mathcal{Z}}^{\mathcal{V}}(A(\widehat{x})) = \sigma^{\mathcal{V}}(\widehat{x},\widehat{z}).$$

Finally, observe

$$|\delta_{\mathcal{Z}}^{\mathcal{V}}(A(\widetilde{x})) - \delta_{\mathcal{Z}}^{\mathcal{V}}(A(\widehat{x}))| = |\sigma^{\mathcal{V}}(\widetilde{x},\widetilde{z}) - \sigma^{\mathcal{V}}(\widehat{x},\widehat{z})|$$

$$\leq |\sigma^{\mathcal{V}}(\widetilde{x},\widehat{z}) - \sigma^{\mathcal{V}}(\widehat{x},\widehat{z})|$$

$$\leq \zeta\|\widetilde{x} - \widehat{x}\|_2,$$

where the last inequality is due to part (i).

---

[2]The monotonicity and interpolation results, Theorems 6 and 7, hold in the infinite dimensional setting under consideration over the Hilbert space $\ell_2(\mathbb{N})$. The arguments in the convergence analyses make use of infinite dimensional versions of these theorems.

(iii) Observe that $\mathcal{D}^{\mathcal{V}}(A(x)) = \delta_{\mathcal{Z}}^{\mathcal{V}}(A(x))$ for $\mathcal{Z} = \mathbb{C}^+$, so this immediately follows from part (ii). $\qquad\square$

Now we are ready for the global convergence result. We believe that an argument in support of global convergence in the finite dimensional setting based on the supposition that the subspace becomes the whole space eventually is not a proper argument. In [25], it is observed that a subspace framework to maximize the largest eigenvalue stagnates at a local maximizer, that is, not necessarily a global maximizer, after a few iterations.

THEOREM 9 (global convergence). *Every convergent subsequence of the sequence $\{x^{(\ell)}\}$ generated by Algorithm 2 in the infinite dimensional setting converges to a global maximizer of $\mathcal{D}(A(x))$ over all $x \in \widetilde{\Omega}$. Furthermore,*

$$(12) \qquad \lim_{\ell \to \infty} \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell+1)})) = \lim_{\ell \to \infty} \max_{x \in \widetilde{\Omega}} \mathcal{D}^{\mathcal{V}_\ell}(A(x)) = \max_{x \in \widetilde{\Omega}} \mathcal{D}(A(x)).$$

Proof. To prove the claim that every convergent subsequence converges to a global maximizer of $\mathcal{D}(A(x))$, let us consider a convergent subsequence $\{x^{(j_\ell)}\}$ of $\{x^{(\ell)}\}$. First observe that

$$\max_{x \in \widetilde{\Omega}} \mathcal{D}(A(x)) \geq \mathcal{D}(A(x^{(j_\ell)})) = \mathcal{D}^{\mathcal{V}_{j_\ell}}(A(x^{(j_\ell)})),$$

where the equality is due to the interpolation property (part (i) of Theorem 7), as well as

$$\max_{x \in \widetilde{\Omega}} \mathcal{D}(A(x)) \leq \max_{x \in \widetilde{\Omega}} \mathcal{D}^{\mathcal{V}_{j_{\ell+1}-1}}(A(x)) = \mathcal{D}^{\mathcal{V}_{j_{\ell+1}-1}}(A(x^{(j_{\ell+1})})) \leq \mathcal{D}^{\mathcal{V}_{j_\ell}}(A(x^{(j_{\ell+1})})),$$

where both of the inequalities are due to the monotonicity property (Theorem 6). But by part (iii) of Lemma 8, for every subspace $\mathcal{V}$, we have

$$|\mathcal{D}^{\mathcal{V}}(A(\widetilde{x})) - \mathcal{D}^{\mathcal{V}}(A(\widehat{x}))| \leq \zeta \|\widetilde{x} - \widehat{x}\|_2 \quad \forall \widetilde{x}, \widehat{x} \in \Omega,$$

which implies

$$\lim_{\ell \to \infty} \left| \mathcal{D}^{\mathcal{V}_{j_\ell}}(A(x^{(j_{\ell+1})})) - \mathcal{D}^{\mathcal{V}_{j_\ell}}(A(x^{(j_\ell)})) \right| \leq \lim_{\ell \to \infty} \zeta \left\| x^{(j_{\ell+1})} - x^{(j_\ell)} \right\|_2 = 0.$$

Hence, employing the interpolation property once again,

$$\lim_{\ell \to \infty} \mathcal{D}(A(x^{(j_\ell)})) = \lim_{\ell \to \infty} \mathcal{D}^{\mathcal{V}_{j_\ell}}(A(x^{(j_\ell)})) = \max_{x \in \widetilde{\Omega}} \mathcal{D}(A(x)).$$

It follows from the continuity of $\mathcal{D}(A(x))$ that the subsequence $\{x^{(j_\ell)}\}$ converges to a point in $\arg\max_{x \in \widetilde{\Omega}} \mathcal{D}(A(x))$.

Finally, to deduce (12), we proceed as in part (ii) of the proof of [25, Theorem 3.1]. Following similar arguments, the sequence $\{\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell+1)}))\}$ can be shown to be monotonically decreasing and bounded below by $\mathcal{D}_* := \max_{x \in \widetilde{\Omega}} \mathcal{D}(A(x))$, so it is convergent. The proof is completed by constructing a subsequence of $\{\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell+1)}))\}$ that converges to $\mathcal{D}_*$. In particular, for any convergent subsequence $\{x^{(j_\ell)}\}$ of $\{x^{(\ell)}\}$, the sequence $\{\mathcal{D}^{\mathcal{V}_{j_{\ell+1}-1}}(A(x^{(j_{\ell+1})}))\}$ is a subsequence of $\{\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell+1)}))\}$ and satisfies

$$\mathcal{D}^{\mathcal{V}_{j_\ell}}(A(x^{(j_{\ell+1})})) \geq \mathcal{D}^{\mathcal{V}_{j_{\ell+1}-1}}(A(x^{(j_{\ell+1})})) \geq \mathcal{D}_*.$$

Since we have

$$\lim_{\ell \to \infty} \mathcal{D}^{\mathcal{V}_{j_\ell}}(A(x^{(j_{\ell+1})})) = \mathcal{D}_*$$

from the previous paragraph, $\{\mathcal{D}^{\mathcal{V}_{j_{\ell+1}-1}}(A(x^{(j_{\ell+1})}))\}$ also converges to $\mathcal{D}_*$ as desired. $\qquad\square$

4.2. Local rate-of-convergence. Now we assume that the sequence $\{x^{(\ell)}\}$ itself is convergent. Theorem 9 in this case ensures the convergence of the sequence $\{x^{(\ell)}\}$ to a global maximizer x_* of $\mathcal{D}(A(x))$ over $\widetilde{\Omega}$. For instance, if $\mathcal{D}(A(x))$ has a unique global maximizer, due to assertion (12), the sequence $\{x^{(\ell)}\}$ must converge to this unique global maximizer. In this subsection we quantify the speed of this convergence. The main result establishes a local superlinear rate for the convergence of Algorithm 2 in the one parameter case (i.e., $d = 1$) and of an extended version, namely, Algorithm 3 below, in the multiparameter case under the assumption that $\mathcal{D}(A(x))$ is smooth and its Hessian is invertible at the converged maximizer.

Throughout this section $A(x)$ is assumed to be asymptotically stable at some $x \in \widetilde{\Omega}$. A consequence is that $A(x^{(\ell)})$ is asymptotically stable for all large ℓ.

4.2.1. Derivatives of singular value functions. We first present results that relate the first and second derivatives of $\sigma(x, z)$ with those of $\sigma^{\mathcal{V}_\ell}(x, z)$. From here on, the complex variable z is written as $z = \alpha + \mathrm{i}\omega$ for $\alpha, \omega \in \mathbb{R}$. Hence, the notations $\sigma_\alpha^{\mathcal{V}}(x, z)$, $\sigma_\omega^{\mathcal{V}}(x, z)$ stand for differentiation of $\sigma^{\mathcal{V}}(\cdot)$ with respect to the real, imaginary parts of z.

THEOREM 10 (Hermite interpolation of singular value functions). *The following are satisfied by Algorithm 2 for all $\ell \in \mathbb{Z}^+$ and $j = 1, \ldots, \ell$ such that $A(x^{(j)})$ is asymptotically stable:*
  (i) *If $u^{(j)}, v^{(j)}$ consist of a consistent pair of unit left, right singular vectors corresponding to $\sigma(x^{(j)}, z^{(j)})$, then $u^{(j)}, \vartheta^{(j)} := V_\ell^* v^{(j)}$ consist of a consistent pair of unit left, right singular vectors corresponding to $\sigma^{\mathcal{V}_\ell}(x^{(j)}, z^{(j)})$.*
  (ii) *$\sigma_\omega(x^{(j)}, z^{(j)}) = \sigma_\omega^{\mathcal{V}_\ell}(x^{(j)}, z^{(j)})$.*
  (iii) *If the singular value $\sigma(x^{(j)}, z^{(j)})$ is simple, then $\sigma_\alpha(x^{(j)}, z^{(j)}) = \sigma_\alpha^{\mathcal{V}_\ell}(x^{(j)}, z^{(j)})$.*

Proof. (i) Letting $\sigma := \sigma_{\min}(A(x^{(j)}) - z^{(j)}I)$ and $u^{(j)}, v^{(j)}$ be a corresponding pair of consistent left, right singular vectors, this follows from the following line of reasoning:

$$\left(A(x^{(j)}) - z^{(j)}I\right)v^{(j)} = \sigma u^{(j)} \quad \text{and} \quad (u^{(j)})^*\left(A(x^{(j)}) - z^{(j)}I\right) = \sigma(v^{(j)})^* \Longrightarrow$$

$$\left(A(x^{(j)}) - z^{(j)}I\right)V_\ell\vartheta^{(j)} = \sigma u^{(j)} \quad \text{and} \quad (u^{(j)})^*\left(A(x^{(j)}) - z^{(j)}I\right) = \sigma(\vartheta^{(j)})^*V_\ell^* \Longrightarrow$$

$$\left(A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell\right)\vartheta^{(j)} = \sigma u^{(j)} \quad \text{and} \quad (u^{(j)})^*\left(A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell\right) = \sigma(\vartheta^{(j)})^*.$$

Hence, $\vartheta^{(j)}, u^{(j)}$ form a pair of unit right, left singular vectors of $A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell$ corresponding to σ.

(ii) We first remark that the singular value functions $\sigma(x, z)$ and $\sigma^{\mathcal{V}_\ell}(x, z)$ are differentiable at $(x, z) = (x^{(j)}, z^{(j)})$ with respect to the imaginary part of z. This is due to the fact that both of the functions $\sigma(x^{(j)}, \mathrm{i}\omega)$ and $\sigma^{\mathcal{V}_\ell}(x^{(j)}, \mathrm{i}\omega)$ over $\omega \in \mathbb{R}$ have a local minimizer at $\omega^{(j)} \in \mathbb{R}$ such that $z^{(j)} = \mathrm{i}\omega^{(j)}$, which is implied by Theorem 7, in particular, (10). Using the analytical formulas for the derivatives of singular value functions [9, 28], we obtain

$$\sigma_\omega(x^{(j)}, z^{(j)}) = \Re\left(1275(u^{(j)})^*\frac{\partial\left\{A(x^{(j)}) - z^{(j)}I\right\}}{\partial\omega}v^{(j)}\right)$$

$$= \Im\left((u^{(j)})^*v^{(j)}\right) = \Im\left((u^{(j)})^*V_\ell\vartheta^{(j)}\right)$$

$$= \Re\left((u^{(j)})^*\frac{\partial\left\{A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell\right\}}{\partial\omega}\vartheta^{(j)}\right) = \sigma_\omega^{\mathcal{V}_\ell}(x^{(j)}, z^{(j)}).$$

(iii) It is an easy exercise to see that the simplicity of $\sigma(x^{(j)}, z^{(j)})$ implies the simplicity of $\sigma^{\mathcal{V}_\ell}(x^{(j)}, z^{(j)})$, so both $\sigma(x, z)$ and $\sigma^{\mathcal{V}_\ell}(x, z)$ are differentiable with respect to the real part of z at $(x, z) = (x^{(j)}, z^{(j)})$. Once again an application of the analytical formulas for singular value functions yields

$$
\begin{aligned}
\sigma_\alpha(x^{(j)}, z^{(j)}) &= \Re\left((u^{(j)})^* \frac{\partial\{A(x^{(j)}) - z^{(j)}I\}}{\partial\alpha} v^{(j)}\right) \\
&= -\Re\left((u^{(j)})^* v^{(j)}\right) = -\Re\left((u^{(j)})^* V_\ell \vartheta^{(j)}\right) \\
&= \Re\left((u^{(j)})^* \frac{\partial\{A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell\}}{\partial\alpha} \vartheta^{(j)}\right) = \sigma_\alpha^{\mathcal{V}_\ell}(x^{(j)}, z^{(j)}).
\end{aligned}
$$

∎

LEMMA 11. *For Algorithm 2 for all $\ell \in \mathbb{Z}^+$ such that $A(x^{(\ell)})$ is asymptotically stable, we have*

$$
\sigma_{\omega\omega}(x^{(\ell)}, z^{(\ell)}) \ \leq \ \sigma_{\omega\omega}^{\mathcal{V}_\ell}(x^{(\ell)}, z^{(\ell)}).
$$

Proof. The functions $\sigma(x, z)$, $\sigma^{\mathcal{V}_\ell}(x, z)$ are twice continuously differentiable with respect to the imaginary part of z at $(x, z) = (x^{(\ell)}, z^{(\ell)})$, since $\omega^{(\ell)} \in \mathbb{R}$ such that $z^{(\ell)} = i\omega^{(\ell)}$ is a minimizer of $\sigma(x^{(\ell)}, i\omega)$ and $\sigma^{\mathcal{V}_\ell}(x^{(\ell)}, i\omega)$ over $\omega \in \mathbb{R}$ as a consequence of (10). Furthermore, the second derivatives are related by

$$
\begin{aligned}
\sigma_{\omega\omega}(x^{(\ell)}, z^{(\ell)}) &= \lim_{h\to 0} \frac{\sigma(x^{(\ell)}, i(\omega^{(\ell)} + h)) - 2\sigma(x^{(\ell)}, i\omega^{(\ell)}) + \sigma(x^{(\ell)}, i(\omega^{(\ell)} - h))}{h^2} \\
&\leq \lim_{h\to 0} \frac{\sigma^{\mathcal{V}_\ell}(x^{(\ell)}, i(\omega^{(\ell)} + h)) - 2\sigma^{\mathcal{V}_\ell}(x^{(\ell)}, i\omega^{(\ell)}) + \sigma^{\mathcal{V}_\ell}(x^{(\ell)}, i(\omega^{(\ell)} - h))}{h^2} \\
&= \sigma_{\omega\omega}^{\mathcal{V}_\ell}(x^{(\ell)}, z^{(\ell)}),
\end{aligned}
$$

where the inequality follows from $\sigma(x^{(\ell)}, i(\omega^{(\ell)} \pm h)) \leq \sigma^{\mathcal{V}_\ell}(x^{(\ell)}, i(\omega^{(\ell)} \pm h))$ due to monotonicity (Theorem 6) and $\sigma(x^{(\ell)}, i\omega^{(\ell)}) = \sigma^{\mathcal{V}_\ell}(x^{(\ell)}, i\omega^{(\ell)})$ due to the interpolation property (Theorem 7). ∎

4.2.2. Derivatives of distance functions. The key to our rate-of-convergence analysis is the interpolation properties between derivatives of $\mathcal{D}(A(x))$ and $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$. As a starting point for this analysis, we extend the interpolation result of Theorem 7 to first derivatives. In what follows $\mathcal{B}(y, \nu) := \{\widetilde{y} \mid \|\widetilde{y} - y\|_2 \leq \nu\}$ refers to the closed ball (closed interval if y is a scalar) of radius ν centered at y either in a real Euclidean space or in a complex Euclidean space depending on whether y is real or complex.

THEOREM 12 (Hermite interpolation of distance functions). *The following hold regarding Algorithm 2 for every $\ell \in \mathbb{Z}^+$ and $j = 1, \ldots, \ell$ such that $A(x^{(j)})$ is asymptotically stable: If $\mathcal{D}(A(x))$ and $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ are differentiable at $x = x^{(j)}$, then*

$$
\nabla \mathcal{D}(A(x^{(j)})) \ = \ \nabla \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(j)})).
$$

Proof. Suppose $v^{(j)}, u^{(j)}$ consist of a consistent pair of unit right, left singular vectors of $A(x^{(j)}) - z^{(j)}I$ corresponding to $\sigma_{\min}(A(x^{(j)}) - z^{(j)}I)$. By part (i) of Theorem 10 the vectors $\vartheta^{(j)} := V_\ell^* v^{(j)}, u^{(j)}$ form a consistent pair of unit right, left singular vectors of $A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell$ corresponding to $\sigma_{\min}(A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell)$. By employing the analytical formulas for the derivatives of singular value functions, we

obtain

$$\frac{\partial \mathcal{D}(A(x^{(j)}))}{\partial x_s} = \frac{\partial \sigma_{\min}(A(x^{(j)}) - z^{(j)}I)}{\partial x_s} = \Re \left((u^{(j)})^* \frac{\partial \left\{ A(x^{(j)}) - z^{(j)}I \right\}}{\partial x_s} v^{(j)} \right)$$

$$= \Re \left((u^{(j)})^* \frac{\partial \left\{ A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell \right\}}{\partial x_s} \vartheta^{(j)} \right) = \frac{\partial \sigma_{\min}(A^{V_\ell}(x^{(j)}) - z^{(j)}V_\ell)}{\partial x_s} = \frac{\partial \mathcal{D}^{V_\ell}(A(x^{(j)}))}{\partial x_s}$$

for $s = 1, \ldots, d$. This completes the proof. $\qquad\square$

4.2.3. Extended subspace framework. Before going into a formal rate-of-convergence analysis, we shall comment briefly on the extended subspace framework, formally defined in Algorithm 3. In the description of the algorithm, letting e_p be the pth column of the $d \times d$ identity matrix, we employ the notations $e_{p,q} := 1/\sqrt{2}(e_p + e_q)$ if $p \neq q$ and $e_{p,p} := e_p$. The extended version adds the singular vectors not only at $x^{(\ell)}$, but also at the nearby points $x^{(\ell)} + h^{(\ell)}e_{p,q}$ for $p = 1, \ldots, d$, $q = p, \ldots, d$. This obviously brings additional expenses, mainly, the computation of the distance to instability at these nearby points, as well as the computation of the corresponding right singular vectors. For instance, the cost of every iteration for $d = 2$ is about four times that of the basic framework. Hence, this extended framework aims to address the large-scale problems depending on a few parameters.

The interpolation and Hermite interpolation results of Theorems 7 and 12 do hold beyond $x^{(\ell)}$ also at the nearby points $x^{(\ell)} + h^{(\ell)}e_{p,q}$ for Algorithm 3. These are formally stated in the next theorem. We omit its proof as the arguments are similar to those in the proofs of Theorems 7 and 12.

THEOREM 13. *The assertions of Theorems* 7 *and* 12 *are also satisfied by the sequences generated by Algorithm* 3. *Additionally, for Algorithm* 3, *the following hold for all $\ell \in \mathbb{Z}^+$, $j = 1, \ldots, \ell$, $p = 1, \ldots, d$, and $q = p, \ldots, d$:*
   (i) $\mathcal{D}(A(x_{p,q}^{(j)})) = \sigma_{\min}(A(x_{p,q}^{(j)}) - z_{p,q}^{(j)}I) = \sigma_{\min}(A^{V_\ell}(x_{p,q}^{(j)}) - z_{p,q}^{(j)}V_\ell) = \mathcal{D}^{V_\ell}(A(x_{p,q}^{(j)}))$.
   (ii) *If $\mathcal{D}(A(x))$ and $\mathcal{D}^{V_\ell}(A(x))$ are differentiable at $x = x_{p,q}^{(j)}$, then*

$$\nabla \mathcal{D}(A(x_{p,q}^{(j)})) = \nabla \mathcal{D}^{V_\ell}(A(x_{p,q}^{(j)})).$$

Additionally, we remark that the global convergence result of Theorem 9 also applies to the sequence $\{x^{(\ell)}\}$ generated by Algorithm 3.

4.2.4. Rate-of-convergence analysis. The rate-of-convergence analysis here is inspired by [25]. The Hermite interpolation property (Theorems 12 and 13) especially facilitates this. However, the fact that $\mathcal{D}(A(x)), \mathcal{D}^{V_\ell}(A(x))$ are defined in terms of the global minimizers of $\sigma(x,z), \sigma^{V_\ell}(x,z)$ over $z \in \mathbb{C}^+$ brings subtleties. Our first task is to show that $\mathcal{D}(A(x)) = \sigma(x, \mathrm{i}\omega(x))$, $\mathcal{D}^{V_\ell}(x) = \sigma^{V_\ell}(x, \mathrm{i}\omega^{V_\ell}(x))$ for all x close to $x_* := \lim_{\ell \to \infty} x^{(\ell)}$ (recall that $\{x^{(\ell)}\}$ itself is assumed to be convergent, and Theorem 9 ensures that its limit x_* is a global maximizer of $\mathcal{D}(A(x))$ over $x \in \widetilde{\Omega}$), where $\omega(x)$, $\omega^{V_\ell}(x)$ are the real-valued functions defined implicitly by $\sigma_\omega(x, \mathrm{i}\omega(x)) = 0$, $\sigma_\omega^{V_\ell}(x, \mathrm{i}\omega^{V_\ell}(x)) = 0$. Throughout this section, for $x \in \mathbb{R}^d, z \in \mathbb{C}$, and $\epsilon_1, \epsilon_2 > 0$, we employ the notations

$$\mathcal{N}(x, z; \epsilon_1, \epsilon_2) := \{(\widetilde{x}, \widetilde{z}) \in \mathbb{R}^d \times \mathbb{C} \mid \|\widetilde{x} - x\|_2 \leq \epsilon_1, \; |\widetilde{z} - z| \leq \epsilon_2\}, \; \text{and}$$

$$\mathcal{N}(x, \Im z; \epsilon_1, \epsilon_2) := \{(\widetilde{x}, \widetilde{\omega}) \in \mathbb{R}^d \times \mathbb{R} \mid \|\widetilde{x} - x\|_2 \leq \epsilon_1, \; |\widetilde{\omega} - \Im z| \leq \epsilon_2\}$$

for neighborhoods of (x, z) and $(x, \Im z)$. Moreover, $\mathrm{i}\mathbb{R}$ refers to the set of purely imaginary numbers.

Algorithm 3 The extended subspace framework.

Input: The matrix-valued function $A(x)$ of the form (1) with the feasible region $\widetilde{\Omega}$.
Output: The sequences $\{x^{(\ell)}\}$, $\{z^{(\ell)}\}$.

1: $x^{(1)} \leftarrow$ a random point in $\widetilde{\Omega}$.
2: **if** $\Lambda(A(x^{(1)})) \cap \mathbb{C}^+ = \emptyset$ **then**
3:    $z^{(1)} \leftarrow \mathrm{i} \cdot \arg\min_{\omega \in \mathbb{R}} \; \sigma_{\min}(A(x^{(1)}) - \omega \mathrm{i} I)$.
4:    $V_1 \leftarrow$ a unit right singular vector corresponding to $\sigma_{\min}(A(x^{(1)}) - z^{(1)}I)$.
5: **else**
6:    $z^{(1)} \leftarrow$ an eigenvalue in $\Lambda(A(x^{(1)})) \cap \mathbb{C}^+$.
7:    $V_1 \leftarrow$ a unit eigenvector corresponding to the eigenvalue $z^{(1)}$ of $A(x^{(1)})$.
8: **end if**
9: $\mathcal{V}_1 \leftarrow \mathrm{span}\{V_1\}$.
10: **for** $\ell = 1, 2, \ldots$ **do**
11:    $x^{(\ell+1)} \leftarrow \arg\max_{x \in \widetilde{\Omega}} \mathcal{D}^{\mathcal{V}_\ell}(A(x))$.
12:    **if** $\Lambda(A(x^{(\ell+1)})) \cap \mathbb{C}^+ = \emptyset$ **then**
13:      $z^{(\ell+1)} \leftarrow \mathrm{i} \cdot \arg\min_{\omega \in \mathbb{R}} \; \sigma_{\min}(A(x^{(\ell+1)}) - \omega \mathrm{i} I)$.
14:      $v_{\ell+1} \leftarrow$ a right singular vector corresponding to $\sigma_{\min}(A(x^{(\ell+1)}) - z^{(\ell+1)}I)$.
15:    **else**
16:      $z^{(\ell+1)} \leftarrow$ an eigenvalue in $\Lambda(A(x^{(\ell+1)})) \cap \mathbb{C}^+$.
17:      $v_{\ell+1} \leftarrow$ an eigenvector corresponding to the eigenvalue $z^{(\ell+1)}$ of $A(x^{(\ell+1)})$.
18:    **end if**
19:    $h^{(\ell+1)} \leftarrow \|x^{(\ell+1)} - x^{(\ell)}\|_2$
20:    **for** $p = 1, \ldots, d$ **do**
21:      **for** $q = p, \ldots, d$ **do**
22:        $x_{p,q}^{(\ell+1)} \leftarrow x^{(\ell+1)} + h^{(\ell+1)}e_{p,q}$.
23:        **if** $\Lambda(A(x_{p,q}^{(\ell+1)})) \cap \mathbb{C}^+ = \emptyset$ **then**
24:          $z_{p,q}^{(\ell+1)} \leftarrow \mathrm{i} \cdot \arg\min_{\omega \in \mathbb{R}} \; \sigma_{\min}(A(x_{p,q}^{(\ell+1)}) - \omega \mathrm{i} I)$.
25:          $v_{p,q}^{(\ell+1)} \leftarrow$ a right singular vector
                              corresponding to $\sigma_{\min}(A(x_{p,q}^{(\ell+1)}) - z_{p,q}^{(\ell+1)}I)$.
26:        **else**
27:          $z_{p,q}^{(\ell+1)} \leftarrow$ an eigenvalue in $\Lambda(A(x_{p,q}^{(\ell+1)})) \cap \mathbb{C}^+$.
28:          $v_{p,q}^{(\ell+1)} \leftarrow$ an eigenvector
                              corresponding to the eigenvalue $z_{p,q}^{(\ell+1)}$ of $A(x_{p,q}^{(\ell+1)})$.
29:        **end if**
30:      **end for**
31:    **end for**
32:    $V_{\ell+1} \leftarrow \mathrm{orth}\left(\begin{bmatrix} V_\ell & v_{\ell+1} & v_{1,1}^{(\ell+1)} & \ldots & v_{1,d}^{(\ell+1)} & v_{2,2}^{(\ell+1)} & \ldots & v_{d,d}^{(\ell+1)} \end{bmatrix}\right)$.
33:    $\mathcal{V}_{\ell+1} \leftarrow \mathrm{Col}(V_{\ell+1})$.
34: **end for**

LEMMA 14. *Suppose that $\sigma(x_*, z_*)$ is simple, where $z_* \in \mathrm{i}\mathbb{R}$ is the global min-imizer of $\sigma(x_*, z)$ over $z \in \mathbb{C}^+$, which we assume is unique. Additionally, assume $\sigma_{\omega\omega}(x_*, z_*) = \delta > 0$. There exist neighborhoods $\mathcal{N}(x_*, z_*; \epsilon_1, \epsilon_2)$ and $\mathcal{N}(x_*, \Im z_*; \widetilde{\epsilon}_1, \widetilde{\epsilon}_2)$ of (x_*, z_*) and $(x_*, \Im z_*)$ with $\widetilde{\epsilon}_1 \leq \epsilon_1$, $\widetilde{\epsilon}_2 \leq \epsilon_2$ such that*
   (i) *the singular values $\sigma(x, z)$ and $\sigma^{\mathcal{V}_\ell}(x, z)$ for all ℓ large enough remain simple and their first three derivatives are bounded above by constants uniformly for all $(x, z) \in \mathcal{N}(x_*, z_*; \epsilon_1, \epsilon_2)$, where the constants are independent of ℓ,*

(ii) $\sigma_{\omega\omega}(\widetilde{x}, \mathrm{i}\widetilde{\omega}) \geq 3\delta/4 \ \ \forall(\widetilde{x}, \widetilde{\omega}) \in \mathcal{N}(x_*, \Im z_*; \widetilde{\epsilon}_1, \widetilde{\epsilon}_2)$, and
(iii) $\forall \ell$ large enough, $\sigma_{\omega\omega}^{\mathcal{V}_\ell}(\widetilde{x}, \mathrm{i}\widetilde{\omega}) \geq \delta/2 \ \ \forall(\widetilde{x}, \widetilde{\omega}) \in \mathcal{N}(x_*, \Im z_*; \widetilde{\epsilon}_1, \widetilde{\epsilon}_2)$.

Proof. For assertion (i) we refer to [25, Proposition 2.9]. By the boundedness of $\sigma_{\omega\omega\omega}(x, \mathrm{i}\omega)$ in $\mathcal{N}(x_*, z_*; \epsilon_1, \epsilon_2)$ and $\sigma_{\omega\omega}(x_*, z_*) = \delta$, we infer a neighborhood $\widehat{\mathcal{N}} := \mathcal{N}(x_*, \Im z_*; \widehat{\epsilon}_1, \widehat{\epsilon}_2) \subseteq \mathcal{N}(x_*, z_*, \epsilon_1, \epsilon_2)$ of $(x_*, \Im z_*)$ such that $\sigma_{\omega\omega}(\widetilde{x}, \mathrm{i}\widetilde{\omega}) \geq 3\delta/4$ for all $(\widetilde{x}, \widetilde{\omega}) \in \widehat{\mathcal{N}}$.

Now, since $\lim_{\ell \to \infty} \sigma(x^{(\ell)}, z^{(\ell)}) = \lim_{\ell \to \infty} \mathcal{D}(A(x^{(\ell)})) = \sigma(x_*, z_*)$, by the uniqueness of the global minimizer z_* and the continuity of $\sigma(x, z)$, we must have $z^{(\ell)} \to z_*$. Choose ℓ large enough so that $(x^{(\ell)}, \Im z^{(\ell)})$ is in $\widehat{\mathcal{N}}$. By Lemma 11

$$\sigma_{\omega\omega}^{\mathcal{V}_\ell}(x^{(\ell)}, z^{(\ell)}) \geq \sigma_{\omega\omega}(x^{(\ell)}, z^{(\ell)}) \geq 3\delta/4$$

for all such large ℓ. Finally, since $\sigma_{\omega\omega\omega}^{\mathcal{V}_\ell}(x, \mathrm{i}\omega)$ are also uniformly bounded in $\widehat{\mathcal{N}}$ by a constant independent of ℓ, there must exist a neighborhood $\widetilde{\mathcal{N}} \subseteq \widehat{\mathcal{N}}$ of $(x_*, \Im z_*)$ such that $\sigma_{\omega\omega}(\widetilde{x}, \mathrm{i}\widetilde{\omega}) \geq \delta/2$ for all $(\widetilde{x}, \widetilde{\omega}) \in \widetilde{\mathcal{N}}$ and all ℓ sufficiently large. $\square$

The significance of the last lemma is that it implies the existence of the functions $\widetilde{\omega}(x)$ and $\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ for x near x_* such that $\mathrm{i}\widetilde{\omega}(x)$ and $\mathrm{i}\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ satisfy the first order optimality conditions to be a minimizer of $\sigma(x, z)$ and $\sigma^{\mathcal{V}_\ell}(x, z)$, respectively, over $z \in \mathbb{C}^+$. This is formally stated by the next result.

LEMMA 15 (local representations of the minimizers). *Suppose (x_*, z_*) is as in Lemma 14 and satisfies the assumptions of that lemma. Additionally, suppose $\sigma_\alpha(x_*, z_*) > 0$. For some $\eta_1, \eta_2 > 0$, the following hold:*
  (i) *There exists a unique three times differentiable function $\widetilde{\omega}(x) : \mathcal{B}(x_*, \eta_1) \to \mathcal{B}(\Im z_*, \eta_2)$ such that $\widetilde{\omega}(x_*) = \Im z_*$, as well as*

  (13) $\qquad \sigma_\omega(x, \mathrm{i}\widetilde{\omega}(x)) = 0 \quad \text{and} \quad \sigma_{\omega\omega}(x, \mathrm{i}\widetilde{\omega}(x)) \geq \delta/2 \quad \forall x \in \mathcal{B}(x_*, \eta_1).$

  Furthermore, $\mathrm{i}\widetilde{\omega}(x)$ is the unique point in the ball $\mathcal{B}(z_*, \eta_2)$ that satisfies the first order optimality conditions to be a minimizer of $\sigma(x, z)$ over $z \in \mathbb{C}^+$ for all $x \in \mathcal{B}(x_*, \eta_1)$;
  (ii) *For all ℓ large enough, in particular satisfying $(x^{(\ell)}, z^{(\ell)}) \in \mathcal{N}(x_*, z_*; \eta_1, \eta_2)$, there exists a unique three times differentiable function $\widetilde{\omega}^{\mathcal{V}_\ell}(x) : \mathcal{B}(x_*, \eta_1) \to \mathcal{B}(\Im z_*, \eta_2)$ such that $\widetilde{\omega}^{\mathcal{V}_\ell}(x^{(\ell)}) = \Im z^{(\ell)}$, as well as*

  (14) $\quad \sigma_\omega^{\mathcal{V}_\ell}(x, \mathrm{i}\widetilde{\omega}^{\mathcal{V}_\ell}(x)) = 0 \quad \text{and} \quad \sigma_{\omega\omega}^{\mathcal{V}_\ell}(x, \mathrm{i}\widetilde{\omega}^{\mathcal{V}_\ell}(x)) \geq \delta/2 \quad \forall x \in \mathcal{B}(x_*, \eta_1).$

  Furthermore, for such ℓ and for all $x \in \mathcal{B}(x_*, \eta_1)$, $\mathrm{i}\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ is the unique point in the ball $\mathcal{B}(z_*, \eta_2)$ that satisfies the first order optimality conditions to be a minimizer of $\sigma^{\mathcal{V}_\ell}(x, z)$ over $z \in \mathbb{C}^+$.

Proof. Consider the neighborhood $\mathcal{N}(x_*, \Im z_*; \widetilde{\epsilon}_1, \widetilde{\epsilon}_2)$ as in Lemma 14. For all (x, ω) in this neighborhood, the singular values $\sigma(x, \mathrm{i}\omega)$ and $\sigma^{\mathcal{V}_\ell}(x, \mathrm{i}\omega)$ for large ℓ are simple and their second derivatives with respect to ω are bounded below by $\delta/2$ uniformly. By the implicit function theorem, for some $\eta_1 < \widetilde{\epsilon}_1$, $\eta_2 < \widetilde{\epsilon}_2$, there exist a unique function $\widetilde{\omega}(x)$ satisfying (13), and a unique function $\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ satisfying (14) for all ℓ large enough, in particular, $(x^{(\ell)}, z^{(\ell)}) \in \mathcal{N}(x_*, z_*; \eta_1, \eta_2)$. Note that the uniformity of the radii η_1, η_2 over all such ℓ is due to the uniform lower bound $\delta/2$ on the second derivatives $\sigma_{\omega\omega}(x, \mathrm{i}\omega)$ and $\sigma_{\omega\omega}^{\mathcal{V}_\ell}(x, \mathrm{i}\omega)$ in $\mathcal{N}(x_*, \Im z_*; \widetilde{\epsilon}_1, \widetilde{\epsilon}_2)$. We also remark that $\widetilde{\omega}(x)$ and $\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ are three times differentiable, because $\sigma(x, \mathrm{i}\omega)$ and $\sigma^{\mathcal{V}_\ell}(x, \mathrm{i}\omega)$ are simple, hence, real analytic, in $\mathcal{N}(x_*, \Im z_*; \widetilde{\epsilon}_1, \widetilde{\epsilon}_2)$.

We complete the proof by arguing that $i\widetilde{\omega}(x)$ and $i\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ are the unique first order optimal points in a ball in the complex plane around z_*. To this end, $\sigma_\alpha(x_*, z_*) > 0$ by assumption, and $\sigma_\alpha(x, z) > 0$ in a neighborhood $\widetilde{\mathcal{N}}$ of (x_*, z_*) due to continuous differentiability of $\sigma(x, z)$ around (x_*, z_*) as implied by part (i) of Lemma 14. Choose ℓ even larger if necessary so that $(x^{(\ell)}, z^{(\ell)})$ is in this neighborhood and $\sigma_\alpha^{\mathcal{V}_\ell}(x^{(\ell)}, z^{(\ell)}) = \sigma_\alpha(x^{(\ell)}, z^{(\ell)}) > 0$, where we employ part (iii) of Theorem 10 for the equality. By Lemma 14 once again, in particular due to analyticity of $\sigma^{\mathcal{V}_\ell}(x, z)$ and a uniform upper bound on $\sigma_{\alpha\alpha}^{\mathcal{V}_\ell}(x, z)$ independent of ℓ, there exists a neighborhood $\widehat{\mathcal{N}} \subseteq \widetilde{\mathcal{N}}$ of (x_*, z_*) such that

$$\sigma_\alpha(x, z) > 0 \quad \text{and} \quad \sigma_\alpha^{\mathcal{V}_\ell}(x, z) > 0 \quad \forall (x, z) \in \widehat{\mathcal{N}}$$

for all ℓ large enough. Now η_1, η_2 as in the previous paragraph can be chosen small enough if necessary so that $\mathcal{B}(x_*, \eta_1) \times \mathcal{B}(z_*, \eta_2) \subseteq \widehat{\mathcal{N}}$. Consequently, for each $x \in \mathcal{B}(x_*, \eta_1)$, since $i\widetilde{\omega}(x), i\widetilde{\omega}^{\mathcal{V}_\ell}(x) \in \mathcal{B}(z_*, \eta_2)$, we have

$$\sigma_\omega(x, i\widetilde{\omega}(x)) = 0, \ \ \sigma_\alpha(x, i\widetilde{\omega}(x)) > 0 \quad \text{and} \quad \sigma_\omega^{\mathcal{V}_\ell}(x, i\widetilde{\omega}^{\mathcal{V}_\ell}(x)) = 0, \ \ \sigma_\alpha^{\mathcal{V}_\ell}(x, i\widetilde{\omega}^{\mathcal{V}_\ell}(x)) > 0,$$

that is, $i\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ and $i\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ satisfy the first order optimality conditions to be a minimizer of $\sigma(x, z)$ and $\sigma^{\mathcal{V}_\ell}(x, z)$ over $z \in \mathbb{C}^+$. The uniqueness of $i\widetilde{\omega}(x)$, $i\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ as first order optimal points inside $\mathcal{B}(z_*, \eta_2)$ follows from the uniqueness of $\widetilde{\omega}(x), \widetilde{\omega}^{\mathcal{V}_\ell}(x) \in \mathcal{B}(\Im z_*, \eta_2)$ satisfying $\sigma_\omega(x, i\widetilde{\omega}(x)) = 0$, $\sigma_\omega^{\mathcal{V}_\ell}(x, i\widetilde{\omega}^{\mathcal{V}_\ell}(x)) = 0$, as well as the fact that $\sigma_\alpha(x, z) \neq 0$, $\sigma_\alpha^{\mathcal{V}_\ell}(x, z) \neq 0$ for $z \in \mathcal{B}(z_*, \eta_2)$ such that $\Re z > 0$. □

The next result has two important conclusions. First it shows that the unique first order optimal points of the previous lemma are indeed global minimizers where $\mathcal{D}(A(x))$ and $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ are attained. Second, it establishes that the smoothness assumption on $\mathcal{D}(A(x))$ at x_* implies the existence of a ball centered at x_* in which $\mathcal{D}(A(x))$, as well as $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ for all large ℓ, are smooth. In this result and elsewhere, $\sigma_{-2}(\cdot)$ refers to the second smallest singular value of its matrix argument.

LEMMA 16 (uniform smoothness). *Suppose that the sequence $\{x^{(\ell)}\}$ by Algorithms 2 or 3 converges to a point x_* that is strictly in the interior of $\widetilde{\Omega}$ and that $\mathcal{D}(A(x_*))$ is attained at a unique z_*, the singular value $\sigma(x_*, z_*) > 0$ is simple, $\sigma_{\omega\omega}(x_*, z_*) > 0$, and $\sigma_\alpha(x_*, z_*) > 0$. Let η_1, $\widetilde{\omega}(x)$, $\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ for all ℓ large enough, say $\ell \geq \ell_1$, be as in Lemma 15.*

Then there exist an $\eta \leq \eta_1$ and an $\varepsilon > 0$ satisfying the following:
(i) *We have $\min\{\sigma(x, i\widetilde{\omega}(x)), \ \sigma_{-2}(x, i\widetilde{\omega}(x)) - \sigma(x, i\widetilde{\omega}(x))\} \geq \varepsilon$, and $\mathcal{D}(A(x))$ is uniquely attained at $i\widetilde{\omega}(x)$ for all $x \in \mathcal{B}(x_*, \eta)$.*
(ii) *Additionally, for all ℓ large enough, say $\ell \geq \ell_2 \geq \ell_1$ and for all $x \in \mathcal{B}(x_*, \eta)$, $\min\{\sigma^{\mathcal{V}_\ell}(x, i\widetilde{\omega}^{\mathcal{V}_\ell}(x)), \ \sigma_{-2}^{\mathcal{V}_\ell}(x, i\widetilde{\omega}^{\mathcal{V}_\ell}(x)) - \sigma^{\mathcal{V}_\ell}(x, i\widetilde{\omega}^{\mathcal{V}_\ell}(x))\} \geq \varepsilon$, and $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is attained at $i\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ uniquely.*

Proof. Following the arguments in [25, Lemma 2.8], due to the simplicity and positivity of $\sigma(x_*, z_*)$ and the interpolation properties $\sigma(x^{(\ell)}, z^{(\ell)}) = \sigma^{\mathcal{V}_\ell}(x^{(\ell)}, z^{(\ell)})$, there exists an ε and a neighborhood $\mathcal{N}(x_*, \Im z_*; \widetilde{\eta}_1, \widetilde{\eta}_2)$ of $(x_*, \Im z_*)$ such that

$$(15) \qquad \min\{\sigma(x, i\omega), \ \sigma_{-2}(x, i\omega) - \sigma(x, i\omega)\} \geq \varepsilon \quad \forall (x, \omega) \in \mathcal{N}(x_*, \Im z_*; \widetilde{\eta}_1, \widetilde{\eta}_2)$$

and, for all ℓ large enough,

$$(16) \qquad \min\{\sigma^{\mathcal{V}_\ell}(x, i\omega), \ \sigma_{-2}^{\mathcal{V}_\ell}(x, i\omega) - \sigma^{\mathcal{V}_\ell}(x, i\omega)\} \geq \varepsilon \quad \forall (x, \omega) \in \mathcal{N}(x_*, \Im z_*; \widetilde{\eta}_1, \widetilde{\eta}_2).$$

Furthermore, let $\widetilde{\omega}(x)$ and $\widetilde{\omega}^{\mathcal{V}_\ell}(x)$ for ℓ large enough, say $\ell \geq \ell_1$, be as in Lemma 15 satisfying its assertions (i) and (ii) for some $\eta_1 \leq \widetilde{\eta}_1$, $\eta_2 \leq \widetilde{\eta}_2$.

First we prove that $\mathcal{D}(A(x))$ is minimized uniquely at $\mathrm{i}\widetilde{\omega}(x)$ for x in a ball centered at x_*. To this end, for each $x \in \mathcal{B}(x_*, \eta_1)$, we define the functions

$$\delta_2(x) := \inf\left\{\sigma(x,z) \mid z \in \mathbb{C}^+ \backslash \mathcal{B}(z_*, \eta_2)\right\}, \quad \delta_1(x) := \min\left\{\sigma(x,z) \mid z \in \mathcal{B}(z_*, \eta_2)\right\},$$
$$\delta(x) := \delta_2(x) - \delta_1(x).$$

In particular, let $\delta_* := \delta(x_*) := \delta_2(x_*) - \delta_1(x_*) = \delta_2(x_*) - \mathcal{D}(A(x_*)) > 0$. Consider the ball $\mathcal{B}(x_*, \delta_*/8\zeta)$, where the constant ζ is as in Lemma 8. For each $x \in \mathcal{B}(x_*, \eta)$ with $\eta := \min\{\delta_*/(8\zeta), \eta_1\}$, by part (ii) of Lemma 8,

$$\delta_2(x_*) - \delta_2(x) \leq \delta_*/8 \quad \text{and} \quad \delta_1(x) - \delta_1(x_*) \leq \delta_*/8,$$

implying

$$(17) \qquad \delta(x) = \delta_2(x) - \delta_1(x) \geq \delta_2(x_*) - \delta_1(x_*) - \delta_*/4 = 3\delta_*/4.$$

This means that $\sigma(x,z)$ is minimized over $z \in \mathbb{C}^+$ globally by some point in the interior of $\mathcal{B}(z_*, \eta_2)$. This has to be $\mathrm{i}\widetilde{\omega}(x)$, because Lemma 15 asserts that all other points in $\mathcal{B}(z_*, \eta_2)$ violate the first order optimality conditions to be a minimizer of $\sigma(x,z)$ over $z \in \mathbb{C}^+$. Note also that, since $\eta \leq \eta_1 \leq \widetilde{\eta}_1$ as well as $\widetilde{\omega}(x) \in \mathcal{B}(\Im z_*, \eta_2)$ with $\eta_2 \leq \widetilde{\eta}_2$, it follows from (15) that

$$\min\{\sigma(x, \mathrm{i}\widetilde{\omega}(x)), \ \sigma_{-2}(x, \mathrm{i}\widetilde{\omega}(x)) - \sigma(x, \mathrm{i}\widetilde{\omega}(x))\} \geq \varepsilon \quad \forall x \in \mathcal{B}(x_*, \eta).$$

This completes the proof of assertion (i).

Now there exists $\ell_2 \geq \ell_1$ such that $x^{(\ell)} \in \mathcal{B}(x_*, \eta)$ and $z^{(\ell)} \in \mathcal{B}(z_*, \eta_2)$ for all $\ell \geq \ell_2$. We show the satisfaction of assertion (ii) for all $\ell \geq \ell_2$. For such an ℓ and for an $x \in \mathcal{B}(x_*, \eta)$, define

$$\delta_2^{\mathcal{V}_\ell}(x) := \inf\left\{\sigma^{\mathcal{V}_\ell}(x,z) \mid z \in \mathbb{C}^+ \backslash \mathcal{B}(z_*, \eta_2)\right\},$$
$$\delta_1^{\mathcal{V}_\ell}(x) := \min\left\{\sigma^{\mathcal{V}_\ell}(x,z) \mid z \in \mathcal{B}(z_*, \eta_2)\right\}, \quad \delta^{\mathcal{V}_\ell}(x) := \delta_2^{\mathcal{V}_\ell}(x) - \delta_1^{\mathcal{V}_\ell}(x).$$

The monotonicity property $\sigma^{\mathcal{V}_\ell}(x,z) \geq \sigma(x,z)$ for all z implies $\delta_j^{\mathcal{V}_\ell}(x) \geq \delta_j(x)$ for $j = 1, 2$. Additionally, $\delta_1^{\mathcal{V}_\ell}(x^{(\ell)}) \leq \sigma^{\mathcal{V}_\ell}(x^{(\ell)}, z^{(\ell)}) = \sigma(x^{(\ell)}, z^{(\ell)}) = \delta_1(x^{(\ell)})$ implies $\delta_1^{\mathcal{V}_\ell}(x^{(\ell)}) = \delta_1(x^{(\ell)})$. These observations lead us to

$$(18) \qquad \delta^{\mathcal{V}_\ell}(x^{(\ell)}) = \delta_2^{\mathcal{V}_\ell}(x^{(\ell)}) - \delta_1^{\mathcal{V}_\ell}(x^{(\ell)}) \geq \delta_2(x^{(\ell)}) - \delta_1(x^{(\ell)}) \geq 3\delta_*/4,$$

where the last bound is due to (17). Now for each $x \in \mathcal{B}(x_*, \eta)$, by part (ii) of Lemma 8, we have

$$\delta_2^{\mathcal{V}_\ell}(x^{(\ell)}) - \delta_2^{\mathcal{V}_\ell}(x) \leq \zeta\|x^{(\ell)} - x\|_2 \leq \zeta(\|x^{(\ell)} - x_*\|_2 + \|x_* - x\|_2) \leq \delta_*/4 \quad \text{and}$$
$$\delta_1^{\mathcal{V}_\ell}(x) - \delta_1^{\mathcal{V}_\ell}(x^{(\ell)}) \leq \zeta\|x - x^{(\ell)}\|_2 \leq \zeta(\|x - x_*\|_2 + \|x_* - x^{(\ell)}\|_2) \leq \delta_*/4,$$

that gives rise to

$$\delta^{\mathcal{V}_\ell}(x) = \delta_2^{\mathcal{V}_\ell}(x) - \delta_1^{\mathcal{V}_\ell}(x) \geq \delta_2^{\mathcal{V}_\ell}(x^{(\ell)}) - \delta_1^{\mathcal{V}_\ell}(x^{(\ell)}) - \delta_*/2 \geq \delta_*/4,$$

where the last lower bound follows from (18). Hence, the global minimizer of $\sigma^{\mathcal{V}_\ell}(x,z)$ over $z \in \mathbb{C}^+$ also lies strictly in the interior of $\mathcal{B}(z_*, \eta_2)$. Once again this global

minimizer has to be $i\widetilde{\omega}^{\mathcal{V}_\ell}(x)$, as the other points in $\mathcal{B}(z_*, \eta_2)$ violate the first order conditions to be a minimizer of $\sigma^{\mathcal{V}_\ell}(x, z)$ over $z \in \mathbb{C}^+$ due to Lemma 15. Finally, since $\eta \leq \widetilde{\eta}_1$ and $\widetilde{\omega}^{\mathcal{V}_\ell}(x) \in \mathcal{B}(\Im z_*, \eta_2)$ with $\eta_2 \leq \widetilde{\eta}_2$, we deduce the following uniform gap from (16):

$$\min\{\sigma^{\mathcal{V}_\ell}(x, i\widetilde{\omega}^{\mathcal{V}_\ell}(x)), \, \sigma^{\mathcal{V}_\ell}_{-2}(x, i\widetilde{\omega}^{\mathcal{V}_\ell}(x)) - \sigma^{\mathcal{V}_\ell}(x, i\widetilde{\omega}^{\mathcal{V}_\ell}(x))\} \geq \varepsilon \quad \forall x \in \mathcal{B}(x_*, \eta),$$

completing the proof of assertion (ii). $\qquad\square$

Now we state two results regarding the second and third derivatives of $\mathcal{D}(A(x))$ and $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ for x inside the ball $\mathcal{B}(x_*, \eta)$ specified by Lemma 16. Note that there are no interpolation properties between the second and higher order derivatives of these two distance to instability functions.

LEMMA 17 (uniform boundedness of the third derivatives). *Suppose that $\{x^{(\ell)}\}$ by Algorithms 2 or 3 converges to a point x_* that satisfies the assumptions in Lemma 16. There exist $\eta \in \mathbb{R}^+$ and $\xi \in \mathbb{R}^+$ such that, for all ℓ large enough, we have*

$$\left| \frac{\partial^3 \left[\mathcal{D}^{\mathcal{V}_\ell}(A(\widetilde{x})) \right]}{\partial x_p \partial x_q \partial x_s} \right| \leq \xi \quad \forall \widetilde{x} \in \mathcal{B}(x_*, \eta)$$

for $p, q, s = 1, \ldots, d$.

Proof. By Lemma 16 there exists an η such that, for all ℓ large enough and for all $x \in \mathcal{B}(x_*, \eta)$, $\mathcal{D}^{\mathcal{V}_\ell}(A(x)) = \sigma(x, i\omega^{\mathcal{V}_\ell}(x))$ for some real-valued three times differentiable function $\omega^{\mathcal{V}_\ell}(x)$, and the singular value $\sigma(x, i\omega^{\mathcal{V}_\ell}(x))$ is simple, bounded away from zero. Consequently, $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is three times differentiable on $\mathcal{B}(x_*, \eta)$. Each third derivative of $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ at a given $\widetilde{x} \in \mathcal{B}(x_*, \eta)$ can be expressed in terms of the first three derivatives of $\sigma^{\mathcal{V}_\ell}(x, z)$ at $(x, z) = (\widetilde{x}, i\omega^{\mathcal{V}_\ell}(\widetilde{x}))$ and the first two derivatives of $\omega^{\mathcal{V}_\ell}(x)$ at $x = \widetilde{x}$, all of which can be bounded solely in terms of $\|A_j\|$ and $\max_{x \in \mathcal{B}(x_*, \eta)} |f_j(x)|$ for $j = 1, \ldots, \kappa$, as well as the reciprocal of the uniform gap $\sigma^{\mathcal{V}_\ell}_{-2}(x, i\omega^{\mathcal{V}_\ell}(x)) - \sigma^{\mathcal{V}_\ell}(x, i\omega^{\mathcal{V}_\ell}(x)) \geq \varepsilon > 0$ established in Lemma 16 for all $x \in \mathcal{B}(x_*, \eta)$. $\qquad\square$

LEMMA 18 (accuracy of the second derivatives). *Suppose that $\{x^{(\ell)}\}$ by Algorithm 2 when $d = 1$ or Algorithm 3 converges to a point x_* that satisfies the assumptions in Lemma 16. Then, for all ℓ large enough, we have*

$$\left\| \nabla^2 \mathcal{D}(A(x^{(\ell)})) - \nabla^2 \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell)})) \right\|_2 = O(h^{(\ell)}),$$

where we define $h^{(\ell)} := |x^{(\ell)} - x^{(\ell-1)}|$ for Algorithm 2 when $d = 1$. Furthermore, if $\nabla^2 \mathcal{D}(A(x_*))$ is invertible, then for all ℓ large enough

$$\left\| \left[\nabla^2 \mathcal{D}(A(x^{(\ell)})) \right]^{-1} - \left[\nabla^2 \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell)})) \right]^{-1} \right\|_2 = O(h^{(\ell)}).$$

Proof. By Lemma 16 there exists an η such that both $\mathcal{D}(A(x))$ as well as $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ for ℓ large enough are three times continuously differentiable for all $x \in \mathcal{B}(x_*, \eta)$. Furthermore, we can choose ℓ even larger if necessary so that $\mathcal{B}(x^{(\ell)}, h^{(\ell)}) \subseteq \mathcal{B}(x_*, \eta)$.

Now, for Algorithm 3 exploiting the properties

$$\mathcal{D}(A(x^{(\ell)})) = \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell)})), \quad \nabla \mathcal{D}(A(x^{(\ell)})) = \nabla \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell)})), \quad \text{and}$$
$$\mathcal{D}(A(x^{(\ell)} + h^{(\ell)} e_{p,q})) = \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell)} + h^{(\ell)} e_{p,q}))$$

for $p = 1, \ldots, d$, $q = p, \ldots, d$, and for Algorithm 2 when $d = 1$ exploiting

$$\mathcal{D}(A(x^{(\ell)})) = \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell)})), \quad \mathcal{D}(A(x^{(\ell-1)})) = \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell-1)})), \quad \text{and}$$
$$\mathcal{D}'(A(x^{(\ell)})) = \left[\mathcal{D}^{\mathcal{V}_\ell}\right]'(A(x^{(\ell)})),$$

as well as the uniform boundedness of the third derivatives of $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ independent of ℓ (see Lemma 17 above), the results follow from an application of Taylor's theorem with a third order remainder. For details we refer to the proof of Lemma 2.8 in [25]; only $\mathcal{D}(A(x)), \mathcal{D}^{\mathcal{V}_\ell}(A(x))$ here takes the role of $\lambda_J(\omega), \lambda_J^{(\ell)}(\omega)$ in that proof. □

The superlinear rate-of-convergence result below is a consequence of the Hermite interpolation properties, as well as Lemma 18 that relates the second derivatives of the distance functions. Intuitively, an analogy can be made between the reduced function $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ and the model function used by a quasi-Newton method about $x^{(\ell)}$; the reduced function satisfies

$$\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell)})) = \mathcal{D}(A(x^{(\ell)})), \ \nabla\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell)})) = \nabla\mathcal{D}(A(x^{(\ell)})),$$
$$\nabla^2\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell)})) \approx \nabla^2\mathcal{D}(A(x^{(\ell)})),$$

where the gap between the Hessians decay to zero, moreover, $x^{(\ell+1)}$ is defined as the maximizer of $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$. Formally, the proof of Theorem 3.3 in [25] applies identically with $\mathcal{D}(A(x)), \mathcal{D}^{\mathcal{V}_\ell}(A(x))$ taking the role of $\lambda_J(\omega), \lambda_J^{(\ell)}(\omega)$ in that work to deduce this superlinear rate-of-convergence result .

THEOREM 19 (local superlinear convergence). *Suppose that the sequence $\{x^{(\ell)}\}$ generated by Algorithm 2 when $d = 1$ or by Algorithm 3 converges to a point x_* that is strictly in the interior of $\widetilde{\Omega}$ such that (i) $\mathcal{D}(A(x_*))$ is attained at $z_* \in i\mathbb{R}$ uniquely, $\sigma(x_*, z_*)$ is simple, $\sigma_{\omega\omega}(x_*, z_*) > 0$, $\sigma_\alpha(x_*, z_*) > 0$, and (ii) $\nabla^2\mathcal{D}(A(x_*))$ is invertible. Then, there exists a constant $\mu \in \mathbb{R}^+$ such that*

$$\frac{\|x^{(\ell+1)} - x_*\|_2}{\|x^{(\ell)} - x_*\|_2 \max\{\|x^{(\ell)} - x_*\|_2, \|x^{(\ell-1)} - x_*\|_2\}} \ \leq \ \mu \quad \forall \ell \geq 2.$$

Remark 20. As indicated in Remark 4, if $A(x)$ is real valued, then the global minimizers of $\sigma(x_*, i\omega)$ over $\omega \in \mathbb{R}$ are in plus, minus pairs. This means that the uniqueness assumption on $z_* \in i\mathbb{R}$, the point where $\mathcal{D}(A(x_*))$ is attained, in the main rate-of-convergence result (Theorem 19) and the auxiliary results leading to this main result, is no longer true. In this case, this superlinear rate-of-convergence result still holds, but under the assumption that $\mathcal{D}(A(x_*))$ is attained uniquely over all purely imaginary numbers with nonnegative (or with nonpositive) imaginary parts. It is straightforward to modify the analysis above to this setting by restricting z to the first quadrant in the complex plane, requiring $z^{(\ell)}$ to have a nonnegative imaginary part.

5. Solutions of the reduced problems.

5.1. Computation of $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$. Algorithms 2 and 3 require the maximization of the distance $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ over $x \in \widetilde{\Omega}$ at step ℓ. For this maximization the objective

$$(19) \qquad\qquad \mathcal{D}^{\mathcal{V}_\ell}(A(x)) \ = \ \min_{z \in \mathbb{C}^+} \sigma_{\min}(A(x)V_\ell - zV_\ell)$$

needs to be computed at several x in $\mathbb{R}^d$. Assuming the dimension of $\mathcal{V}_\ell$ is small, which is usually the case in practice as the framework converges at a superlinear rate, it turns out this can be performed efficiently and reliably by means of simple extensions of the techniques to compute the distance to uncontrollability [19, 20] characterized by

$$(20) \qquad \min_{z \in \mathbb{C}} \sigma_{\min} \left(\begin{bmatrix} F - zI & G \end{bmatrix} \right)$$

for a given pair $F \in \mathbb{C}^{q \times q}, G \in \mathbb{C}^{q \times m}$ with $q \geq m$. Essential steps are outlined next.

We first compute a reduced QR factorization of the form

$$\begin{bmatrix} V_\ell & A(x)V_\ell \end{bmatrix} = \underbrace{\begin{bmatrix} V_\ell & \widetilde{V}_\ell \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} I_\ell & R_A \\ 0 & R_B \end{bmatrix}}_{R},$$

where $Q \in \mathbb{C}^{n \times 2\ell}$ has orthonormal columns, $R \in \mathbb{R}^{2\ell \times 2\ell}$ is upper triangular, and I_ℓ denotes the identity matrix of size $\ell \times \ell$. Then the singular values of $A(x)V_\ell - zV_\ell$ and

$$Q^T \left\{ A(x)V_\ell - zV_\ell \right\} = \begin{bmatrix} R_A \\ R_B \end{bmatrix} - z \begin{bmatrix} I_\ell \\ 0 \end{bmatrix} = \begin{bmatrix} R_A - zI_\ell \\ R_B \end{bmatrix}$$

are the same for every z. Hence (19) reduces to[3]

$$(21) \qquad \mathcal{D}^{\mathcal{V}_\ell}(A(x)) = \min_{z \in \mathbb{C}^+} \sigma_{\min} \left(\begin{bmatrix} R_A^T - zI_\ell & R_B^T \end{bmatrix} \right)$$

which is of the form (20) except that the minimization has to be performed over $\mathbb{C}^+$ rather than $\mathbb{C}$. Modifications of the techniques in [19, 20] to solve (21), in particular, to perform optimization over $\mathbb{C}^+$, are straightforward.

In practice we adopt the approach in [20] combined with BFGS (that employs line searches ensuring the satisfaction of the weak Wolfe conditions by the iterates). First BFGS converges to a local minimizer of (21), which is then subjected to the trisection test from [20] to determine whether the local minimizer is a global minimizer or not. If it is a global minimizer, $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is indeed computed. Otherwise, the trisection test provides a point in $\mathbb{C}^+$ where the singular value function takes a smaller value compared with the converged local minimizer, and BFGS is restarted with this point. Each trisection test requires the extraction of the real eigenvalues of either a $2\ell^2 \times 2\ell^2$ matrix or a $4\ell^2 \times 4\ell^2$ matrix pencil. In practice we employ the latter; even though these are larger generalized eigenvalue problems, they turn out to be better conditioned and it is usually possible to solve them reliably in the presence of rounding errors.

5.2. Maximization of $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$. Suppose first that $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is attained on the imaginary axis for all $x \in \widetilde{\Omega}$. Then the regularity result in part (i) of Theorem 1 remains to be true but with $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ taking the role of $\mathcal{D}(A(x))$. A consequence is that upper support functions of Theorem 2 extend to this rectangular setting, that is, for a given $\widetilde{x}$ where $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is differentiable, we have

$$[\mathcal{D}^{\mathcal{V}_\ell}(A(x))]^2 \leq q(x; \widetilde{x}) := [\mathcal{D}^{\mathcal{V}_\ell}(A(\widetilde{x}))]^2 + \nabla[\mathcal{D}^{\mathcal{V}_\ell}(A(\widetilde{x}))]^2(x - \widetilde{x}) + \frac{\gamma}{2}\|x - \widetilde{x}\|_2^2 \quad \forall x,$$

---

[3]An alternative approach to compute $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ based on a reduced QR factorization of $\begin{bmatrix} V_\ell & A_1V_\ell & \dots & A_\kappa V_\ell \end{bmatrix}$ is discussed in a longer version of this text [31]. The advantage of that approach is that it requires the computation of only one QR factorization for the minimization of $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ over x rather than one QR factorization for each evaluation of $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$. But the difference it makes appears to be insignificant in practice, because V_ℓ has typically a few columns and the computational cost of these QR factorizations is linear, quite small compared to other ingredients of the algorithm. We thank Daniel Kressner for pointing out the alternative approach.

where γ now satisfies $\lambda_{\max}(\nabla^2[\mathcal{D}^{\mathcal{V}_\ell}(A(x))]^2) \leq \gamma$ for all x such that $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is differentiable. Furthermore, estimation of such a γ is facilitated by simple adaptations of Theorem 3, that is, defining $M^{\mathcal{V}_\ell}(x,z) := (A^{V_\ell}(x) - zV_\ell)^*(A^{V_\ell}(x) - zV_\ell)$ and denoting with $z(x)$ the point where $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is attained, the bound

$$\lambda_{\max}(\nabla^2[\mathcal{D}^{\mathcal{V}_\ell}(A(x))]^2) \leq \lambda_{\max}\left(\nabla^2_{xx}M^{\mathcal{V}_\ell}(x,z(x))\right)$$

holds for all x where $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is differentiable. Specifically, for the affine case, we have $A^{V_\ell}(x) = B_0 V_\ell + \sum_{j=1}^{\kappa} x_j B_j V_\ell$ and $\nabla^2_{xx}M^{\mathcal{V}_\ell}(x,z)$ is given by the right-hand side of (6) by replacing B_j with $B_j V_\ell$ for $j = 1, \ldots, \kappa$.

The remarks of the previous paragraph also apply if $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is attained strictly on the right-hand side of the complex plane for all $x \in \widetilde{\Omega}$. In this case the arguments deal with both the complex and the imaginary parts of the point $z(x)$ where $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is attained, which, at points of differentiability, are defined implicitly by $\sigma_\alpha^{\mathcal{V}_\ell}(x, z(x)) = 0$, $\sigma_\omega^{\mathcal{V}_\ell}(x, z(x)) = 0$.

Hence, we employ Algorithm 1 for the maximization of $\mathcal{D}^{\mathcal{V}_\ell}(x)$. But naturally $q(x; x^{(j)})$ is now defined in terms of $\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(j)}))$, $\nabla \mathcal{D}^{\mathcal{V}_\ell}(A(x^{(j)}))$ and γ is a global upper bound on $\lambda_{\max}(\nabla^2[\mathcal{D}^{\mathcal{V}_\ell}(A(x))]^2)$ for all $x \in \widetilde{\Omega}$. The algorithm is guaranteed to converge under the assumptions of the previous paragraphs.

A likely situation is that $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ is attained on the imaginary axis for some x, yet strictly on the right-hand side of the complex plane for some other values of x. Even in this case, the global convergence of the algorithm can be asserted provided γ is large enough. Some difficulty arises due to the possibility of points $\widetilde{x}$ such that $\mathcal{D}^{\mathcal{V}_\ell}(A(\widetilde{x}))$ is attained at $\widetilde{z} \in i\mathbb{R}$ with $\sigma_\alpha^{\mathcal{V}_\ell}(\widetilde{x}, \widetilde{z}) = 0$. The left-hand derivative of $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ in a particular direction at such a point can be smaller than the right-hand derivative, which means $q(x; \widetilde{x})$ is not an upper support function anymore. However, such a point $\widetilde{x}$ is away from maximizers of $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$, and $q(x; \widetilde{x})$ is still an upper support function for $\mathcal{D}^{\mathcal{V}_\ell}(A(x))$ near a global maximizer locally provided γ is large enough, from which the global convergence of the algorithm can be deduced.

In practice we choose γ based on the largest eigenvalue of $\nabla^2_{xx}M^{\mathcal{V}_\ell}(x, z(x))$, which seems to be working well.

6. A subspace framework for uniformly stable problems. If it is known in advance that $A(x)$ has all of its eigenvalues in the closed left-half of the complex plane at all $x \in \widetilde{\Omega}$ (such is the case, for instance, for dissipative Hamiltonian systems, which naturally arise from multibody problems, circuit simulation, finite element modeling of a disk brake [29]), then the reduced problems can be simplified. In particular, there is theoretical ground to perform the inner minimization problems over the imaginary axis. Formally, we operate on the reduced problems

$$\max_{x \in \widetilde{\Omega}} \widetilde{\mathcal{D}}^{\mathcal{V}}(A(x)) \qquad \text{with} \quad \widetilde{\mathcal{D}}^{\mathcal{V}}(A(x)) := \min_{\omega \in \mathbb{R}} \sigma_{\min}(A^V(x) - \omega i V),$$

where V is a matrix whose columns form an orthonormal basis for $\mathcal{V}$ and $A^V(x)$ is defined as in (8).

The corresponding greedy framework is presented in Algorithm 4. As before at every iteration, we solve a reduced problem. We compute the distance to instability of the full problem at the maximizing x, in particular, retrieve $\omega \in \mathbb{R}$ where the full distance is attained, and expand the subspace with the inclusion of a right singular vector corresponding to $\sigma_{\min}(A(x) - i\omega I)$ at the optimal x and ω.

Following the footsteps of the arguments in sections 3 and 4, it is possible to deduce (i) the monotonicity, that is $\mathcal{V} \supseteq \mathcal{W} \implies \widetilde{\mathcal{D}}^{\mathcal{V}}(A(x)) \leq \widetilde{\mathcal{D}}^{\mathcal{W}}(A(x))$, and (ii) Hermite interpolation, that is

(1) $\widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x^{(j)})) = \mathcal{D}(A(x^{(j)}))$, as well as
(2) if $\mathcal{D}(A(x))$ and $\widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x))$ are differentiable at $x = x^{(j)}$, then $\nabla \widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x^{(j)})) = \nabla \mathcal{D}(A(x^{(j)}))$

for $j = 1, \ldots, \ell$. These properties, as before, pave the way for global convergence and superlinear rate-of-convergence results presented formally below.

THEOREM 21. *Suppose that the eigenvalues of $A(x)$ are contained in the closed left-half of the complex plane at all $x \in \widetilde{\Omega}$. Every convergent subsequence of the sequence $\{x^{(\ell)}\}$ generated by Algorithm 4 in the infinite dimensional setting converges to a global maximizer of $\mathcal{D}(A(x))$ over all $x \in \widetilde{\Omega}$. Furthermore,*

$$\lim_{\ell \to \infty} \widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x^{(\ell+1)})) = \lim_{\ell \to \infty} \max_{x \in \widetilde{\Omega}} \widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x)) = \max_{x \in \widetilde{\Omega}} \mathcal{D}(A(x)).$$

THEOREM 22. *Suppose that the eigenvalues of $A(x)$ are contained in the closed left-half of the complex plane at all $x \in \widetilde{\Omega}$. Suppose also that the sequence $\{x^{(\ell)}\}$ generated by Algorithm 4 when $d = 1$ converges to a point x_* that is strictly in the interior of $\widetilde{\Omega}$ such that (i) $\mathcal{D}(A(x_*))$ is attained at ω_* uniquely, $\sigma_{\min}(A(x_*) - \omega_* i I)$ is positive and simple, $\partial^2 [\sigma_{\min}(A(x_*) - \omega_* i I)] / \partial \omega^2 > 0$, as well as (ii) $\mathcal{D}''(A(x_*)) \neq 0$. Then, there exists a constant $\mu \in \mathbb{R}^+$ such that*

$$\frac{|x^{(\ell+1)} - x_*|}{|x^{(\ell)} - x_*| \max\{|x^{(\ell)} - x_*|, |x^{(\ell-1)} - x_*|\}} \leq \mu \quad \forall \ell \geq 2.$$

It is also possible to define an extended variant of Algorithm 4, reminiscent of Algorithm 3, to which the global convergence result (Theorem 21) and the superlinear convergence result (Theorem 22) apply for every d.

Algorithm 4 The subspace framework for uniformly stable problems.

Input: The matrix-valued function $A(x)$ of the form (1) with the feasible region $\widetilde{\Omega}$.
Output: The sequences $\{x^{(\ell)}\}$, $\{\omega^{(\ell)}\}$.
 1: $x^{(1)} \leftarrow$ a random point in $\widetilde{\Omega}$.
 2: $\omega^{(1)} \leftarrow \arg\min_{\omega \in \mathbb{R}} \sigma_{\min}(A(x^{(1)}) - \omega i I)$.
 3: $V_1 \leftarrow$ a unit right singular vector corresponding to $\sigma_{\min}(A(x^{(1)}) - \omega^{(1)} i I)$.
 4: $\mathcal{V}_1 \leftarrow \text{span}\{V_1\}$.
 5: **for** $\ell = 1, 2, \ldots$ **do**
 6: $\quad x^{(\ell+1)} \leftarrow \arg\max_{x \in \widetilde{\Omega}} \widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x))$.
 7: $\quad \omega^{(\ell+1)} \leftarrow \arg\min_{\omega \in \mathbb{R}} \sigma_{\min}(A(x^{(\ell+1)}) - \omega i I)$.
 8: $\quad v_{\ell+1} \leftarrow$ a right singular vector corresponding to $\sigma_{\min}(A(x^{(\ell+1)}) - \omega^{(\ell+1)} i I)$.
 9: $\quad V_{\ell+1} \leftarrow \text{orth}\left(\begin{bmatrix} V_\ell & v_{\ell+1} \end{bmatrix}\right)$ and $\mathcal{V}_{\ell+1} \leftarrow \text{Col}(V_{\ell+1})$.
10: **end for**

When solving the reduced problem at step ℓ, the objective $\widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x))$ needs to be computed at several points. To perform the calculation of this objective efficiently at a given $x \in \mathbb{R}^d$, inspired by the approach in [26] in the context of a subspace framework for the pseudospectral abscissa, we first compute a reduced QR factorization

$$\begin{bmatrix} V_\ell & A(x)V_\ell \end{bmatrix} = Q \begin{bmatrix} R_1 & R_2 \end{bmatrix},$$

where $Q \in \mathbb{C}^{n \times 2\ell}$, $R_1, R_2 \in \mathbb{C}^{2\ell \times \ell}$. It follows that the singular values of $A(x)V_\ell - \omega i V_\ell$ and $R_2 - \omega i R_1$ are the same for all $\omega \in \mathbb{R}$, so

$$\widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x)) = \min_{\omega \in \mathbb{R}} \sigma_{\min}(R_2 - \omega i R_1).$$

We solve the singular value minimization problem on the right above by employing the globally convergent level-set algorithms for the distance to instability [6, 8]. These minimization problems are cheap to solve, because they involve $2\ell \times \ell$ matrix-valued functions.

Finally for the maximization of $\widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x))$ we apply Algorithm 1 in section 2 with $\widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x))$ taking the role of $\mathcal{D}(A(x))$. Every convergent subsequence of the resulting sequence, under the assumption that $A(x)$ is asymptotically stable for all $x \in \widetilde{\Omega}$, is guaranteed to converge to a global maximizer of $\widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x))$ provided γ is chosen such that $\lambda_{\max}(\nabla^2[\widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x))]^2) \leq \gamma$ for all $x \in \widetilde{\Omega}$. In the affine case with $A^{\mathcal{V}_\ell}(x) = B_0 V_\ell + \sum_{j=1}^{\kappa} x_j B_j V_\ell$, a theoretically sound choice for γ is given by the largest eigenvalue of the matrix on the right-hand side of (6) but with blocks defined in terms of $B_j V_\ell$ rather than B_j.

7. Numerical results for the subspace framework. In this section we report the results obtained from several numerical experiments that are carried out with a publicly available MATLAB implementation [32] of Algorithm 2 to maximize $\mathcal{D}(A(x))$ when $A(x)$ is affine and x is subject to box constraints. We do not make an implementation of Algorithm 4 available at this point, as it typically does not converge to a desired maximizer unless the problem is uniformly stable. But some numerical results illustrating the convergence of Algorithm 4 are also reported at the end of section 7.1

7.1. Illustration of the algorithm on a 200 × 200 example. We first demonstrate Algorithm 2 and its convergence on the example of Figure 2, which concerns the maximization of $\mathcal{D}(A + kbc^T)$ for a random $A \in \mathbb{R}^{200 \times 200}$ and random $b, c \in \mathbb{R}^{200}$. Recall that the distance function is smooth at the global maximizer for this example.

To make the example more challenging we replace A with $\widetilde{A} = A - 0.08bc^T$, which translates the graph of the distance function horizontally by 0.08. We maximize $\mathcal{D}(\widetilde{A} + kbc^T)$ over $k \in [-0.2, 0.2]$. This distance function is asymptotically stable only in a small subinterval of $[-0.2, 0.2]$, and $\widetilde{A}$ itself is unstable. The locally convergent methods in the literature would fail and stagnate at an unstable point unless they are initiated with an asymptotically stable point on this small interval, which is hard to know in advance. In contrast, Algorithm 2 locates an asymptotically stable point that maximizes the distance to instability regardless of the initial random point for Hermite interpolation. (As a convention we always choose the initial point as the midpoint of the box or the interval in our implementation, so for this particular example we start with 0. But any other point in the interval should work equally well.)

Figure 3 displays the full distance function (the solid curve) and the reduced distance functions (the dashed, dashed-dotted, dotted curves) at the third, fourth, fifth steps of the algorithm. The dashed curve representing the reduced function at the third step with a three dimensional subspace interpolates the full distance function at $k = -0.2, 0, 0.2$ (marked with circles). This reduced function is maximized at $k = 0.10612$, so at the next iteration the full distance function is interpolated at this maximizer as well. This leads to the dashed-dotted curve in the figure, which represents the reduced function at the fourth step of the algorithm with a four dimensional subspace. Next the full distance function is interpolated at the maximizer $k = 0.08019$ of the dashed-dotted curve, in addition to the existing interpolation points, leading to the dotted curve representing the reduced function at step five with a five dimensional subspace. Observe that, as the subspace dimension increases, the reduced functions tend to the full distance function. Observe also that, due to monotonicity, each re-
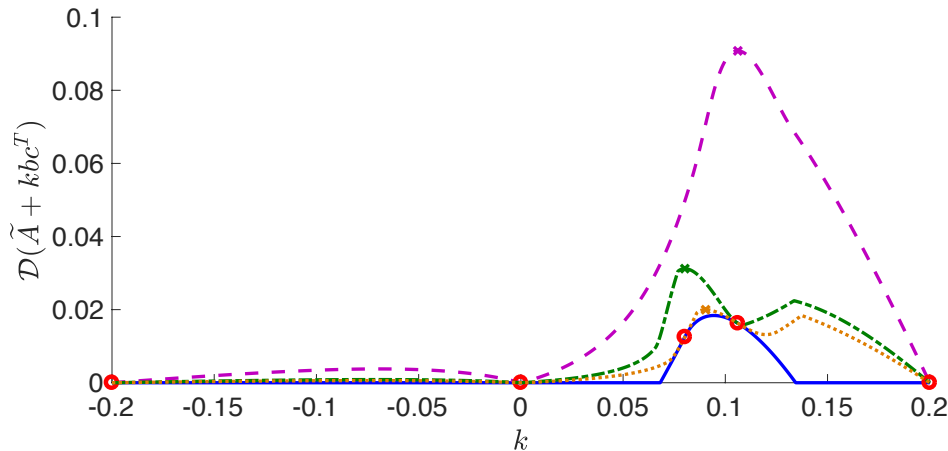
FIG. 3. *This figure displays the progress of Algorithm 2 on a variation of the example of Figure 2, which concerns the robust stability of a random matrix $A \in \mathbb{R}^{200 \times 200}$ subject to rank one updates kbc^T for given random $b, c \in \mathbb{R}^{200}$. Here we maximize $\mathcal{D}(\widetilde{A} + kbc^T)$ over $k \in [-0.2, 0.2]$, where $\widetilde{A} := A - 0.08bc^T$. The solid curve is a plot of the full distance function $\mathcal{D}(\widetilde{A} + kbc^T)$ with respect to k, whereas the dashed, dashed-dotted, dotted curves are reduced functions operated on at step k of Algorithm 2 with a k dimensional subspace for $k = 3, 4, 5$, respectively. The crosses mark the global maxima of the reduced functions, whereas the circles mark the points where the full distance function is Hermite interpolated.*

duced function is squeezed in between the full distance function that lies below, and the reduced functions with smaller subspaces that lie above.

The iterates of Algorithm 2 seem to converge at least at a superlinear rate on this example; this is depicted in Table 2; in particular, see the decay of the errors of the maximal values of the reduced distance functions on the third column of the bottom row, as well as the errors of the maximizers of the reduced distance functions on the third column of the top row.

The minimum of $\sigma^{\mathcal{V}_\ell}(x^{(\ell+1)}, z)$ over $z \in \mathbb{C}^+$ must be attained on the imaginary axis at the later steps of Algorithm 2 in theory. This is a consequence of the facts that $\|x^{(\ell+1)} - x^{(\ell)}\| \to 0$, as well as $\sigma^{\mathcal{V}_\ell}(x^{(\ell)}, z)$ must be minimized at $z^{(\ell)}$ on the imaginary axis due to Theorem 7. This property holds to be true for all the examples that we have experimented with. For the particular example, the global minimizer of $\sigma^{\mathcal{V}_\ell}(x^{(\ell+1)}, z)$ over $z \in \mathbb{C}^+$ is listed with respect to ℓ on the fourth column in the top row of Table 2; notice that the minimizer is strictly on the right-hand side in the first step, but in all other steps the minimizer is always on the imaginary axis.

The γ values employed for the solution of the reduced problems increase monotonically as the subspace dimension increases. Typically in the first steps with small dimensional subspaces we deal with subproblems that require much smaller values of γ compared to the full problem. This is a feature that contributes to the efficient solution of the reduced problems. The γ values employed for the reduced problems for the particular example are listed on the rightmost column of the top row in Table 2; much smaller values of γ are used compared with the full problem for which the bound that follows from (6) is $\gamma = 2\|bc^T\|_2^2 = 2\|b\|_2^2\|c\|_2^2 = 72836$.

Algorithm 4 which performs the inner minimization over the imaginary axis rather than the right-half of the complex plane fails to converge, indeed stagnates at an unstable point, on this example on the interval $[-0.2, 0.2]$. A similar phenomenon occurs even on smaller intervals, such as $[0.05, 0.15]$, as long as the interval contains

TABLE 2

The table concerns an application of Algorithm 2 to the example of Figure 3 on the interval $[-0.2, 0.2]$. It lists the iterates generated by Algorithm 2, as well as their errors, the points in $\mathbb{C}^+$ where $\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell+1)}))$ are attained, and the γ values used for the reduced problems. The notations $\mathcal{D}_*$ and x_* represent the maximal value of $\mathcal{D}(\widetilde{A} + kbc^T)$ over $k \in [-0.2, 0.2]$ and the corresponding global maximizer, respectively.

ℓ	$x^{(\ell+1)}$	$|x^{(\ell+1)} - x_*|$	$\arg\min_{z \in \mathbb{C}^+} \sigma^{\mathcal{V}_\ell}(x^{(\ell+1)}, z)$	γ
1	-0.2000000000	0.2943923006	0.348268	164
2	0.2000000000	0.1056076994	0	948
3	0.1061223556	0.0117300550	0.367974i	983
4	0.0801946598	0.0141976408	-0.249139i	1346
5	0.0903625728	0.0040297278	-0.294728i	1465
6	0.0944455439	0.0000532433	-0.323530i	2201
7	0.0943923006	0.0000000143	-0.323085i	2374

ℓ	$\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell+1)}))$	$\mathcal{D}^{\mathcal{V}_\ell}(A(x^{(\ell+1)})) - \mathcal{D}_*$
1	1.8127144436	1.7943202283
2	0.7324009260	0.7140067108
3	0.0907849141	0.0723906988
4	0.0312089149	0.0128146997
5	0.0199709057	0.0015766904
6	0.0184001783	0.0000059630
7	0.0183942153	0.0000000000

TABLE 3

This table concerns an application of Algorithm 4 to the example of Figure 3 on the interval $[0.07, 0.13]$ where uniform stability holds. It lists the maximal values of the reduced distance functions and their errors with respect to the subspace dimension.

ℓ	$\widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x^{(\ell+1)}))$	$\widetilde{\mathcal{D}}^{\mathcal{V}_\ell}(A(x^{(\ell+1)})) - \mathcal{D}_*$
1	0.4386346102	0.4202403949
2	0.2183039380	0.1999097228
3	0.0188802168	0.0004860015
4	0.0183970429	0.0000028276
5	0.0183942153	0.0000000000

unstable points. This algorithm however converges to the correct global maximizer at a superlinear rate on the interval $[0.07, 0.13]$ where uniform stability holds; this is consistent with what is expected in theory, in particular, with Theorems 21 and 22. The maximal values of the reduced distance functions by Algorithm 4 on this example are listed in Table 3.

7.2. Results on large random examples. We test the performance of Algorithm 2 once again to maximize $\mathcal{D}(A + kbc^T)$ over $k \in [-3, 3]$ with random $A \in \mathbb{R}^{n \times n}$, $b, c \in \mathbb{R}^n$ for $n = 400, 800, 1200, 1600, 2000$. In each of these examples, A is shifted by a multiple of the identity and normalized so that $\|A\|_2 = 10$ and its spectral abscissa lies in $[-0.31, -0.23]$, and the vectors b, c are normalized so that $\|b\|_2 = \|c\|_2 = \sqrt{50}$. The exact data are available on the web.[4]

Convergence to a global maximizer at a superlinear rate is realized in practice for each one of these examples. Figure 4 confirms the correctness of the computed maximizers for $n = 400$ (dotted curve), $n = 800$ (dashed curve), $n = 1200$ (solid curve); the distance to instability functions are displayed in these figures along with the computed global maxima marked with circles.

---

[4]http://home.ku.edu.tr/~emengi/software/max_di/Data_&_Updates.html.
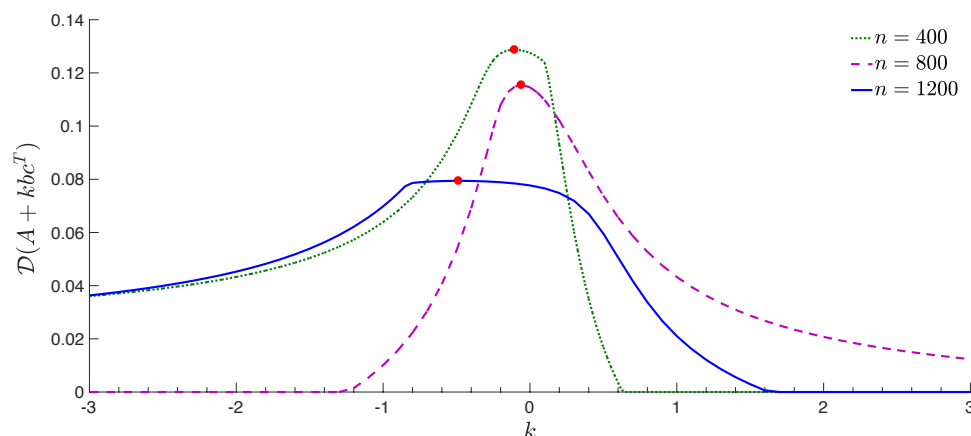
FIG. 4. *The plots of $\mathcal{D}(A + kbc^T)$ as a function of $k \in [-3, 3]$ for the examples of section 7.2 for $n = 400$ (dotted curve), $n = 800$ (dashed curve), $n = 1200$ (solid curve). The circle marks the computed global maximum for each one of the three cases.*

TABLE 4

The number of subspace iterations (2nd column), run-times (3rd–5th columns), computed globally maximal value of the distance $\mathcal{D}(A+kbc^T)$ (7th column), and the corresponding global maximizer (6th column) are listed for each one of the examples in section 7.2 with respect to n. The 3rd, 4th, 5th columns provide the total runtime, total time for the reduced problems, total time for large-scale distance to instability computations in seconds.

n	# iter	time	reduced	dist. instab.	x_*	$\mathcal{D}_*$
400	7	154	123	30	-0.10565	0.12870882
800	6	200	72	124	-0.05943	0.11545563
1200	6	632	245	372	-0.48694	0.07941192
1600	9	1646	366	1228	-0.05009	0.07829585
2000	10	4288	279	3896	0.04762	0.08436380

The main purpose of these experiments is to illustrate the factors determining the efficiency of Algorithm 2 as the sizes of the matrices increase. According to Table 4 the main factor that dominates the runtimes as n increases is the computation of the distance to instability for the full problem a few times, i.e., the solution of the large-scale singular value minimization problems in lines 3 and 13. The number of large-scale distance to instability computations is at most one more than the number of subspace iterations, which varies between 6–10 for these examples. The solution of the reduced problems take a significant portion of the runtime for small n (e.g., $n = 400$), but its effect on the total runtime diminishes as n increases. Indeed the total time consumed for the solution of the reduced problems appear more or less independent of n.

For all of these examples we would be using $\gamma = 2\|bc^T\|_2^2 = 5000$, had we relied on formula (6) and had we been performing optimization directly on the full problem. But as in the previous subsection, we benefit from projections and use $\gamma = 2\|V_\ell^T b(V_\ell^T c)^T\|_2^2$ for the reduced problem at step ℓ, which turns out to be much smaller than the one for the full problem. Indeed the γ values chosen this way for the reduced problems never exceed 400 for the particular examples.

8. Concluding remarks. We have focused on the maximization of the distance to instability of a matrix dependent on several parameters. In the systems setting this

is motivated by maximizing the robust stability and minimizing the transient behavior of the associated autonomous control system. Existing approaches including BFGS, gradient sampling, and bundle methods converge to a locally optimal solution. This local convergence attribute is problematic especially if the starting point is unstable; then all of these methods would stagnate at the unstable point. Unlike these existing approaches in the literature, we have described approaches that converge to a globally optimal point. In particular, even if our approaches are initiated with unstable points, they are capable of locating stable points, and beyond robustly stable points where distance to instability is maximized.

The problems where the parameter-dependent matrix is small are dealt with by means of a support-function-based approach, an adaptation of the approach in [34] that approximates the distance function with a piecewise quadratic model which also lies above the distance function globally. Arguments are presented in support of its global convergence.

The problems where the parameter-dependent matrix is large are dealt with by means of a subspace framework. Here the matrices are restricted to subspaces from the right-hand side leading to rectangular problems. At every iteration such a restricted problem is solved efficiently by employing the support function based approach for small problems. Then the subspace is expanded carefully in a way so that Hermite interpolation properties hold between the full problem and the restricted problem at the optimal point of the restricted problem. We have established formally that the Hermite interpolation property along with a monotonicity property give rise to global convergence at a superlinear rate with respect to the subspace dimension. The proposed approaches for both the small-scale and the large-scale problems are implemented in MATLAB and made available on the internet [32].

The present work extend the approaches in [34] and [25] to a more challenging setting; whereas the approaches in those previous works concern the minimization of the Jth largest eigenvalue, the approaches here are tailored for a maximin optimization problem with a smallest eigenvalue as the objective. There are several other eigenvalue optimization problems of a similar spirit with a maximin or a minimax structure, e.g., minimization of the numerical radius, minimization of the $\mathcal{H}_\infty$-norm of a parameter dependent control system, minimization of the pseudospectral abscissa to name a few. We believe that the ideas and approaches developed in this paper set an example, and are likely to be applicable to those contexts as well.

REFERENCES

[1] P. APKARIAN AND D. NOLL, *Controller design via nonsmooth multidirectional search*, SIAM J. Control Optim., 44 (2006), pp. 1923–1949.
[2] P. APKARIAN AND D. NOLL, *Nonsmooth H_∞ synthesis*, IEEE Trans. Automat. Control, 51 (2006), pp. 71–86.
[3] P. APKARIAN, D. NOLL, AND O. PROT, *A trust region spectral bundle method for nonconvex eigenvalue optimization*, SIAM J. Optim., 19 (2008), pp. 281–306.
[4] D. ARZELIER, D. GEORGIA, S. GUMUSSOY, AND D. HENRION, *H_2 for HIFOO*, preprint, https://arXiv.org/abs/1010.1442v1 (2010).

[5] V. BLONDEL AND J. N. TSITSIKLIS, *NP-hardness of some linear control design problems*, SIAM J. Control Optim., 35 (1997), pp. 2118–2127.

[6] 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.

[7] L. BREIMAN AND A. CUTLER, *A deterministic algorithm for global optimization*, Math. Program., 58 (1993), pp. 179–199, https://doi.org/10.1007/BF01581266.

[8] 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.

[9] 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.

[10] J. V. BURKE, D. HENRION, A. S. LEWIS, AND M. L. OVERTON, *HIFOO – A MATLAB package for fixed-order controller design and H_∞ optimization*, IFAC Proc., 39 (2006), pp. 339–344.

[11] J. V. BURKE, A. S. LEWIS, AND M. L. OVERTON, *Optimal stability and eigenvalue multiplicity*, Found. Comput. Math., 1 (2001), pp. 205–225.

[12] J. V. BURKE, A. S. LEWIS, AND M. L. OVERTON, *Optimizing matrix stability*, Proc. Amer. Math. Soc., 129 (2001), pp. 1635–1642.

[13] J. V. BURKE, A. S. LEWIS, AND M. L. OVERTON, *Two numerical methods for optimizing matrix stability*, Linear Algebra Appl., 351–352 (2002), pp. 117–145.

[14] J. V. BURKE, A. S. LEWIS, AND M. L. OVERTON, *A nonsmooth, nonconvex optimization approach to robust stabilization by static output feedback and low-order controllers*, IFAC Proc., 36 (2003), pp. 175–181.

[15] J. V. BURKE, A. S. LEWIS, AND M. L. OVERTON, *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization*, SIAM J. Optim., 15 (2005), pp. 751–779.

[16] 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.

[17] 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.

[18] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.

[19] M. GU, *New methods for estimating the distance to uncontrollability*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 989–1003.

[20] M. GU, E. MENGI, M. L. OVERTON, J. XIA, AND J. ZHU, *Fast methods for estimating the distance to uncontrollability*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 477–502.

[21] 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.

[22] S. GUMUSSOY, D. HENRION, M. MILLSTONE, AND M. L. OVERTON, *Multiobjective robust control with HIFOO 2.0*, IFAC Proc., 42 (2009), pp. 144–149.

[23] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, 1985.

[24] Y. S. HUNG AND A. G. J. MACFARLANE, *Multivariable Feedback: A Quasi-Classical Approach*, Springer, Berlin, 1982.

[25] F. KANGAL, K. MEERBERGEN, E. MENGI, AND W. MICHIELS, *A subspace method for large scale eigenvalue optimization*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 48–82.

[26] D. KRESSNER AND B. VANDEREYCKEN, *Subspace methods for computing the pseudospectral abscissa and the stability radius*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 292–313.

[27] V. KUNGURTSEV, W. MICHIELS, AND M. DIEHL, *An Inequality-Constrained SL/QP Method for Minimizing the Spectral Abscissa*, preprint, https://arXiv.org/abs/1411.2362v1 (2014).

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

[29] C. MEHL, V. MEHRMANN, AND P. SHARMA, *Stability radii for linear Hamiltonian systems with dissipation under structure-preserving perturbations*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 1625–1654, https://doi.org/10.1137/16M1067330.

[30] E. MENGI, *A support function based algorithm for optimization with eigenvalue constraints*, SIAM J. Optim., 27 (2017), pp. 246–268.

[31] E. MENGI, *Large-Scale and Global Maximization of the Distance to Instability*, preprint, https://arXiv.org/abs/1803.08712v3 (2018).

[32] E. MENGI, *Maximization of Distance to Instability: MATLAB Routines*, http://home.ku.edu.tr/~emengi/software/max_di (2018).

[33] E. MENGI, E. A. YILDIRIM, AND M. KILIC, *Eigopt: Software for Eigenvalue Optimization*, http://home.ku.edu.tr/~emengi/software/eigopt.html (2014).

[34] 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.

[35] A. NEMIROVSKII, *Several NP-hard problems arising in robust stability analysis*, Math. Control Signals Systems, 6 (1993), pp. 99–105.

[36] M. L. OVERTON, *HANSO: A Hybrid Algorithm for Nonsmooth Optimization*, http://cs.nyu.edu/overton/software/hanso (2009).

[37] C. F. VAN LOAN, *How near is a stable matrix to an unstable matrix?*, Contemp. Math., 47 (1985), pp. 465–478.