

A SUBSPACE METHOD FOR LARGE-SCALE EIGENVALUE OPTIMIZATION*

FATIH KANGAL[†], KARL MEERBERGEN[‡], EMRE MENGI[†], AND WIM MICHIELS[‡]

Abstract. We consider the minimization or maximization of the J th largest eigenvalue of an analytic and Hermitian matrix-valued function, and build on Mengi, Yildirim, and Kilic [*SIAM J. Matrix Anal. Appl.*, 35, pp. 699–724, 2014]. This work addresses the setting when the matrix-valued function involved is very large. We describe subspace procedures that convert the original problem into a small-scale one by means of orthogonal projections and restrictions to certain subspaces, and that gradually expand these subspaces based on the optimal solutions of small-scale problems. Global convergence and superlinear rate-of-convergence results with respect to the dimensions of the subspaces are presented in the infinite dimensional setting, where the matrix-valued function is replaced by a compact operator depending on parameters. In practice, it suffices to solve eigenvalue optimization problems involving matrices with sizes on the scale of tens, instead of the original problem involving matrices with sizes on the scale of thousands.

Key words. eigenvalue optimization, large scale, orthogonal projection, eigenvalue perturbation theory, parameter dependent compact operator, matrix-valued function

AMS subject classifications. 65F15, 90C26, 47B37, 47B07

DOI. 10.1137/16M1070025

1. Introduction. We are concerned with the global optimization problems

$$(MN) \text{ minimize } \{\lambda_J(\omega) \mid \omega \in \Omega\} \quad \text{and} \quad (MX) \text{ maximize } \{\lambda_J(\omega) \mid \omega \in \Omega\}.$$

The feasible region Ω of these optimization problems is a compact subset of \mathbb{R}^d . Furthermore, letting $\ell^2(\mathbb{N})$ denote the sequence space consisting of square summable infinite sequences of complex numbers equipped with the inner product $\langle w, v \rangle = \sum_{k=0}^{\infty} \overline{w_k} v_k$ as well as the norm $\|v\|_2 = \sqrt{\sum_{k=0}^{\infty} |v_k|^2}$, the objective function $\lambda_J(\omega)$ is the J th largest eigenvalue of a compact self-adjoint operator

$$(1.1) \quad \mathbf{A}(\omega) : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N}), \quad \mathbf{A}(\omega) := \sum_{\ell=1}^{\kappa} f_{\ell}(\omega) \mathbf{A}_{\ell}$$

for every $\omega \in \overline{\Omega}$, an open subset of \mathbb{R}^d containing the feasible region Ω . Above $\mathbf{A}_{\ell} : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$ and $f_{\ell} : \overline{\Omega} \rightarrow \mathbb{R}$ for $\ell = 1, \dots, \kappa$ represent given compact self-adjoint operators and real-analytic functions, respectively. Throughout the text, $\mathbf{A}(\omega)$ for each ω and $\mathbf{A}_1, \dots, \mathbf{A}_{\kappa}$ as in (1.1) could intuitively be considered as infinite dimensional Hermitian matrices.

*Received by the editors April 12, 2016; accepted for publication (in revised form) October 16, 2017; published electronically January 11, 2018.

<http://www.siam.org/journals/simax/39-1/M107002.html>

Funding: The work of the second and fourth authors was supported by OPTEC, the Optimization in Engineering Center of KU Leuven, the Research Foundation Flanders (project G.0712.11N), and by the project UCoCoS, funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement 675080. The work of the third author was supported in part by the European Commission grant PIRG-GA-268355, TUBITAK - FWO (Scientific and Technological Research Council of Turkey - Belgian Research Foundation, Flanders) joint grant 113T053, and by the BAGEP program of The Science Academy of Turkey.

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

[‡]Department of Computer Science, KU Leuven, University of Leuven, 3001 Heverlee, Belgium (Karl.Meerbergen@cs.kuleuven.be, Wim.Michiels@cs.kuleuven.be).

Our interest in the infinite dimensional eigenvalue optimization problems (MN) and (MX) rise from their finite dimensional counterparts, which for given Hermitian matrices $A_\ell \in \mathbb{C}^{n \times n}$ for $\ell = 1, \dots, \kappa$ involve the matrix-valued function

$$(1.2) \quad A(\omega) := \sum_{\ell=1}^{\kappa} f_\ell(\omega) A_\ell$$

instead of the parameter dependent operator $\mathbf{A}(\omega)$. These problems come with standard challenges due to their nonsmoothness and nonconvexity. But we would like to tackle a different challenge, when the matrices in them are very large, that is, n is very large. Thus, the primary purpose of this paper is to deal with large dimensionality, it does not address the inherent difficulties due to nonconvexity and nonsmoothness. We introduce the ideas in the idealized infinite dimensional setting only because this makes a rigorous convergence analysis possible.

To deal with large dimensionality, we propose restricting the domain and projecting the range of the map $v \mapsto \mathbf{A}(\omega)v$ to small subspaces. This gives rise to eigenvalue optimization problems involving small matrices, which we call reduced problems. Two greedy procedures are presented here to construct small subspaces so that the optimal solution of the reduced problem is close to the optimal solution of the original problem. For both procedures, we observe a superlinear rate of decay in the error with respect to the subspace dimension. The first procedure is more straightforward and constructs smaller subspaces, shown to converge at a superlinear rate when $d = 1$, but lacks a complete formal argument justifying its quick convergence when $d \geq 2$. The second constructs larger subspaces, but comes with a formal proof of superlinear convergence for all d .

While the proposed procedures operate on (MN) and (MX) similarly, there are remarkable differences in their convergence behaviors in these two contexts. The proposed subspace restrictions and projections on the map $v \mapsto \mathbf{A}(\omega)v$ lead to global lower envelopes for $\lambda_J(\omega)$. These lower envelopes in turn make the convergence to the globally smallest value of $\lambda_J(\omega)$ possible as the dimensions of the subspaces by the procedures grow to infinity, which we prove formally. Such a global convergence behavior does not hold for the maximization problem: if the subspace dimension is allowed to grow to infinity, the globally maximal values of the reduced problems converge to a locally maximal value of $\lambda_J(\omega)$ that is not necessarily globally maximal. But the maximization problem possesses a remarkable low-rank property: there exists a J dimensional subspace such that when the map $v \mapsto \mathbf{A}(\omega)v$ is restricted and projected to this subspace, the resulting reduced problem has the same globally largest value as $\lambda_J(\omega)$. The minimization problem does not enjoy an analogous low-rank property.

1.1. Motivation. Large eigenvalue optimization problems arise from various applications. For instance, the distance to instability from a large stable matrix with respect to the matrix 2-norm yields large singular value optimization problems [35] that can be converted into large eigenvalue optimization problems. The computation of the H-infinity norm of the transfer function of a linear time-invariant (LTI) control system can be considered as a generalization of the computation of the distance to instability. This is a norm for the operator that maps inputs of the LTI system into outputs, and plays a major role in robust control. The singular value optimization characterization for the H-infinity norm involves large matrices if the input, output, or, in particular, the intermediate state space have a large dimension. The state-of-the-art algorithms for H-infinity norm computation [4, 5] cannot cope with such

large-scale control systems. In engineering applications, the largest eigenvalue of a matrix-valued function is often sought to be minimized. A particular application is the numerical scheme for the design of the strongest column subject to volume constraints [6], where the sizes of the matrices depend on the fineness of a discretization imposed on a differential operator. These matrices can be very large if a fine grid is employed. The standard semidefinite program (SDP) formulations have received a lot of attention by the convex optimization community since the 1990s [36]. They concern the optimization over the cone of symmetric positive semidefinite matrices of a linear objective function subject to linear constraints. The dual problem of an SDP under mild assumptions can be cast as an eigenvalue optimization problem [15]. If the size of the matrix variable of an SDP is large, the associated eigenvalue optimization problem involves large matrices. The current SDP solvers are usually not suitable to deal with such large-scale problems.

1.2. Literature. Subspace projections and restrictions have been applied to particular eigenvalue optimization problems in the past. But general procedures such as the ones in this paper have not been proposed and studied thoroughly. A subspace restriction idea has been employed specifically for the computation of the pseudospectral abscissa of a matrix in [21]. This computational problem involves the optimization of a linear objective function subject to a constraint on the smallest singular value of an affine matrix-valued function. Fast convergence is observed in that work, and confirmed with a superlinear rate of convergence result. In the context of standard SDPs, subspace methods have been used for quite a while, in particular, the spectral bundle method [15] is based on subspace ideas. The small-scale optimization problems resulting from subspace restrictions and projections are solved by standard SDP solvers [36]. A thorough convergence analysis for them has not been performed; also their efficiency is not fully realized in practice. Extensions for convex quadratic SDPs [25] and linear matrix inequalities [27] have been considered. All these large-scale problems connected to SDPs are convex. We present unified procedures and their convergence analyses that are applicable regardless of whether the problem is convex or nonconvex.

1.3. Spectral properties and operator norm of $\mathbf{A}(\omega)$. The spectrum of $\mathbf{A}(\omega)$ is defined by

$$\sigma(\mathbf{A}(\omega)) := \{z \in \mathbb{R} \mid zI - \mathbf{A}(\omega) \text{ is not invertible}\}.$$

This set contains countably many real eigenvalues each with a finite multiplicity; also the only accumulation point of these eigenvalues is 0 (see [18, p. 185, Theorem 6.26]). We assume, throughout the text, that $\mathbf{A}(\omega)$ has at least J positive eigenvalues for all $\omega \in \Omega$.¹ This ensures that $\lambda_J(\omega)$ is well-defined over Ω .

The eigenspace associated with each eigenvalue of $\mathbf{A}(\omega)$ is also finite dimensional [22, Corollary 6.44]. Furthermore, the set of eigenvectors of $\mathbf{A}(\omega)$ can be chosen in a way so that they are orthonormal and complete in $\ell^2(\mathbb{N})$ [7, Theorem 4.2.23], i.e., there exist an orthonormal sequence $\{v^{(j)}\}$ in $\ell^2(\mathbb{N})$ and a sequence $\{\mu^{(j)}\}$ of real numbers such that $\mathbf{A}(\omega)v^{(j)} = \mu^{(j)}v^{(j)}$ for all $j = 1, 2, \dots$, $\mu^{(j)} \rightarrow 0$ as $j \rightarrow \infty$, and any vector $v \in \ell^2(\mathbb{N})$ can be expressed as $v = \sum_{j=1}^{\infty} \alpha_j v^{(j)}$ for some scalars α_j .

¹Recall that our setup is motivated by the finite dimensional case (large matrices). There the corresponding assumption can be made without losing generality: instead of the optimization of the J th largest eigenvalue of $A(\omega)$, one may equivalently consider the optimization of the J th largest eigenvalue of $A(\omega) + \tau I$ for $\tau \geq 0$ sufficiently large, since the added term only introduces a shift of τ of the eigenvalues, of both the original problems, and the problems obtained by orthogonal projection.

Since $\mathbf{A}_1, \dots, \mathbf{A}_\kappa$ and $\mathbf{A}(\omega)$ are compact, they are bounded. Hence, their operator norms

$$\begin{aligned}\|\mathbf{A}_\ell\|_2 &:= \sup \{ \|\mathbf{A}_\ell v\|_2 \mid v \in \ell^2(\mathbb{N}) \text{ such that } \|v\|_2 = 1 \} \quad \text{for } \ell = 1, \dots, \kappa \quad \text{and} \\ \|\mathbf{A}(\omega)\|_2 &:= \sup \{ \|\mathbf{A}(\omega)v\|_2 \mid v \in \ell^2(\mathbb{N}) \text{ such that } \|v\|_2 = 1 \}\end{aligned}$$

are also well-defined.

1.4. Outline. We formally define the reduced problems, then analyze the relation between the original and the reduced problems in the next section. Some of this analysis, in particular, the low-rank property discussed above, apply only to the maximization problem. Last two subsections of the next section are devoted to the introduction of the two subspace procedures for eigenvalue optimization. Interpolation properties between the original problem and the reduced problems formed by these two subspace procedures are also investigated there. Section 3 concerns the convergence analysis of the two subspace procedures. In particular, section 3.1 establishes the convergence of the subspace procedures globally to the smallest value of $\lambda_J(\omega)$ for the minimization problem as the subspace dimensions grow to infinity. In practice, we observe at least a superlinear rate of convergence with respect to the dimension of the subspaces for both of the procedures. We prove this formally in section 3.2 fully for one of the procedures and partly for the other. Section 4 focuses on variations and extensions of the two subspace procedures for eigenvalue optimization. Specifically, in section 4.1 we argue that a variant that disregards the subspaces formed in the past iterations works effectively for the maximization problem but not for the minimization problem, in sections 4.2 and 4.3 we extend the procedures to optimize a specified singular value of a compact operator depending on several parameters analytically, and in section 4.4 we provide a comparison of one of the subspace procedures for eigenvalue optimization with the cutting plane method [19], which also constructs global lower envelopes repeatedly. Section 5 describes the MATLAB software accompanying this work, and the efficiency of the proposed subspace procedures on particular applications, namely, the numerical radius, the distance to instability from a matrix, and the minimization of the largest eigenvalue of an affine matrix-valued function. The text concludes with a summary as well as research directions for the future in section 6.

2. The subspace procedures. Let \mathcal{V} be a finite dimensional subspace of $\ell^2(\mathbb{N})$, and $V := \{v_1, \dots, v_m\}$ be an orthonormal basis for \mathcal{V} . The linear operator

$$(2.1) \quad \mathbf{V} : \mathbb{C}^m \rightarrow \ell^2(\mathbb{N}), \quad \mathbf{V}\alpha := \sum_{j=1}^m \alpha_j v_j,$$

maps the coordinates of a vector $v \in \mathcal{V}$ relative to V to itself. On the other hand, its adjoint $\mathbf{V}^* : \ell^2(\mathbb{N}) \rightarrow \mathbb{C}^m$ projects a vector in $\ell^2(\mathbb{N})$ orthogonally onto \mathcal{V} and represents the orthogonal projection in coordinates relative to V . The procedures that will be introduced in this section are based on operators of the form

$$(2.2) \quad \mathbf{A}^\mathcal{V}(\omega) := \mathbf{V}^* \mathbf{A}(\omega) \mathbf{V},$$

which is the restriction of $\mathbf{A}(\omega)$ that acts only on \mathcal{V} , with its input and output represented in coordinates relative to V . This operator can be expressed in the form

$$\mathbf{A}^\mathcal{V}(\omega) = \sum_{\ell=1}^{\kappa} f_\ell(\omega) \mathbf{A}_\ell^\mathcal{V}, \quad \text{where } \mathbf{A}_\ell^\mathcal{V} := \mathbf{V}^* \mathbf{A}_\ell \mathbf{V},$$

which is beneficial from a computational point of view, because it is possible to form the $m \times m$ matrix representations of the operators $\mathbf{A}_\ell^\mathcal{V}$ in advance. The reduced eigenvalue optimization problems are defined in terms of the J th largest eigenvalue $\lambda_J^\mathcal{V}(\omega)$ of $\mathbf{A}^\mathcal{V}(\omega)$ as

$$(2.3) \quad \text{minimize } \{\lambda_J^\mathcal{V}(\omega) \mid \omega \in \Omega\} \quad \text{and} \quad \text{maximize } \{\lambda_J^\mathcal{V}(\omega) \mid \omega \in \Omega\}.$$

The discussions above generalize when \mathcal{V} is infinite dimensional. In the infinite dimensional setting, the operators $\mathbf{A}^\mathcal{V}(\omega)$ and $\mathbf{A}_\ell^\mathcal{V}$ are defined similarly in terms of an infinite countable orthonormal basis $V = \{v_j \mid j \in \mathbb{Z}^+\}$ for \mathcal{V} and the associated operator

$$(2.4) \quad \mathbf{V}\alpha := \sum_{j=1}^{\infty} \alpha_j v_j.$$

The eigenvalue function $\lambda_J^\mathcal{V}(\omega)$ is Lipschitz continuous, indeed there exists a uniform Lipschitz constant γ for all subspaces \mathcal{V} of $\ell^2(\mathbb{N})$ as formally stated and proven in the following lemma.

LEMMA 2.1 (Lipschitz continuity). *There exists a real scalar $\gamma > 0$ such that for all subspaces \mathcal{V} of $\ell^2(\mathbb{N})$, we have*

$$|\lambda_J^\mathcal{V}(\tilde{\omega}) - \lambda_J^\mathcal{V}(\omega)| \leq \gamma \cdot \|\tilde{\omega} - \omega\|_2 \quad \forall \tilde{\omega}, \omega \in \bar{\Omega}.$$

Proof. By Weyl's theorem (see [16, Theorem 4.3.1] for the finite dimensional case; its extension to the infinite dimensional setting is straightforward by exploiting the maximin characterization (2.5) of $\lambda_J^\mathcal{V}(\omega)$ given below)

$$\begin{aligned} |\lambda_J^\mathcal{V}(\tilde{\omega}) - \lambda_J^\mathcal{V}(\omega)| &\leq \|\mathbf{A}^\mathcal{V}(\tilde{\omega}) - \mathbf{A}^\mathcal{V}(\omega)\|_2 \\ &\leq \sum_{\ell=1}^{\kappa} |f_\ell(\tilde{\omega}) - f_\ell(\omega)| \|\mathbf{A}_\ell^\mathcal{V}\|_2 \quad \forall \tilde{\omega}, \omega \in \bar{\Omega}. \end{aligned}$$

In the last summation $\|\mathbf{A}_\ell^\mathcal{V}\|_2 \leq \|\mathbf{A}_\ell\|_2$ and the real analyticity of $f_\ell(\omega)$ implies its Lipschitz continuity, hence, the existence of a constant γ_ℓ satisfying

$$|f_\ell(\tilde{\omega}) - f_\ell(\omega)| \leq \gamma_\ell \|\tilde{\omega} - \omega\|_2 \quad \forall \tilde{\omega}, \omega \in \bar{\Omega}.$$

Combining these observations we obtain

$$|\lambda_J^\mathcal{V}(\tilde{\omega}) - \lambda_J^\mathcal{V}(\omega)| \leq \left(\sum_{\ell=1}^{\kappa} \gamma_\ell \|\mathbf{A}_\ell\|_2 \right) \cdot \|\tilde{\omega} - \omega\|_2 \quad \forall \tilde{\omega}, \omega \in \bar{\Omega}$$

as desired. \square

Throughout the rest of this section, we first investigate the relation between $\lambda_J(\omega)$ and $\lambda_J^\mathcal{V}(\omega)$ as well as their globally maximal values. Some of these theoretical results will be frequently used in the subsequent sections. The second and third parts of this section introduce two subspace procedures for the generation of small dimensional subspaces \mathcal{V} , leading to reduced eigenvalue optimization problems that approximate the original eigenvalue optimization problems accurately.

2.1. Relations between $\lambda_J^{\mathcal{V}}(\omega)$ and $\lambda_J(\omega)$. We start with a result about the monotonicity of $\lambda_J^{\mathcal{V}}(\omega)$ with respect to \mathcal{V} . This result is an immediate consequence of the following maximin characterization [8, pp. 1543–1544] (see also [16, Theorem 4.2.11] for the finite dimensional case) of $\lambda_J^{\mathcal{V}}(\omega)$ for all subspaces \mathcal{V} of $\ell^2(\mathbb{N})$:

$$(2.5) \quad \lambda_J^{\mathcal{V}}(\omega) = \sup_{S \subseteq \mathcal{V}, \dim S=J} \min_{v \in S, \|v\|_2=1} \langle \mathbf{A}(\omega)v, v \rangle,$$

where $\langle \cdot, \cdot \rangle$ stands for the standard inner product $\langle w, v \rangle := \sum_{k=0}^{\infty} \overline{w_k} v_k$ and the outer supremum is over all J dimensional subspaces of \mathcal{V} . Furthermore, if $\mathbf{A}^{\mathcal{V}}(\omega)$ has at least J positive eigenvalues, that is, if $\lambda_J^{\mathcal{V}}(\omega) > 0$, or if \mathcal{V} is finite dimensional, then the outer supremum in (2.5) is attained, hence, can be replaced by the maximum. The monotonicity result presented next will play a central role later when we analyze the convergence of the subspace procedures.

LEMMA 2.2 (monotonicity). *Let $\mathcal{V}_1, \mathcal{V}_2$ be subspaces of $\ell^2(\mathbb{N})$ of dimension larger than or equal to J such that $\mathcal{V}_1 \subseteq \mathcal{V}_2$. The following holds:*

$$(2.6) \quad \lambda_J^{\mathcal{V}_1}(\omega) \leq \lambda_J^{\mathcal{V}_2}(\omega) \leq \lambda_J(\omega).$$

A consequence of this monotonicity result is the following interpolatory property.

LEMMA 2.3 (interpolatory property). *Let \mathcal{V} be a subspace of $\ell^2(\mathbb{N})$ of dimension larger than or equal to J , and \mathbf{V} be the operator defined as in (2.1) or (2.4) in terms of an orthonormal basis V for \mathcal{V} . If $S := \text{span}\{s_1, \dots, s_J\} \subseteq \mathcal{V}$, where s_1, \dots, s_J are eigenvectors corresponding to the J largest eigenvalues of $\mathbf{A}(\omega)$, then the following hold:*

- (i) $\lambda_J(\omega) = \lambda_J^{\mathcal{V}}(\omega)$;
- (ii) $\mathbf{V}^* s_J$ is an eigenvector of $\mathbf{V}^* \mathbf{A}(\omega) \mathbf{V}$ corresponding to its eigenvalue $\lambda_J^{\mathcal{V}}(\omega)$.

Proof. (i) We assume each $s_\ell \in \mathcal{V}$ for $\ell = 1, \dots, J$ is of unit length without loss of generality. It follows that there exist $\alpha_1, \dots, \alpha_J$ of unit length such that $s_\ell = \mathbf{V}\alpha_\ell$ for $\ell = 1, \dots, J$. Now define $\mathcal{S}_\alpha := \text{span}\{\alpha_1, \dots, \alpha_J\} \subseteq \mathbb{C}^m$, and observe

$$\begin{aligned} \lambda_J(\omega) &= \langle \mathbf{A}(\omega) s_J, s_J \rangle = \langle \mathbf{A}(\omega) \mathbf{V}\alpha_J, \mathbf{V}\alpha_J \rangle \\ &= \langle \mathbf{V}^* \mathbf{A}(\omega) \mathbf{V}\alpha_J, \alpha_J \rangle \\ &= \min_{\alpha \in \mathcal{S}_\alpha, \|\alpha\|_2=1} \langle \mathbf{V}^* \mathbf{A}(\omega) \mathbf{V}\alpha, \alpha \rangle \leq \lambda_J^{\mathcal{V}}(\omega). \end{aligned}$$

The opposite inequality is immediate from Lemma 2.2, so $\lambda_J(\omega) = \lambda_J^{\mathcal{V}}(\omega)$ as claimed.

(ii) The equalities

$$\langle \mathbf{V}^* \mathbf{A}(\omega) \mathbf{V}\alpha_J, \alpha_J \rangle = \min_{\alpha \in \mathcal{S}_\alpha, \|\alpha\|_2=1} \langle \mathbf{V}^* \mathbf{A}(\omega) \mathbf{V}\alpha, \alpha \rangle = \lambda_J^{\mathcal{V}}(\omega)$$

imply that $\alpha_J = \mathbf{V}^* s_J$ is an eigenvector of $\mathbf{V}^* \mathbf{A}(\omega) \mathbf{V}$ corresponding to the eigenvalue $\lambda_J^{\mathcal{V}}(\omega)$. \square

Maximization of the J th largest eigenvalue over a low dimensional subspace is motivated by the next result. According to the result, it suffices to perform the optimization on a proper J dimensional subspace, which is hard to determine in advance. Here and elsewhere, $\arg \max_{\omega \in \Omega} \lambda_J(\omega)$ and $\arg \min_{\omega \in \Omega} \lambda_J(\omega)$ denote the set of global maximizers and global minimizers of $\lambda_J(\omega)$ over $\omega \in \Omega$, respectively.

LEMMA 2.4 (low-rank property of maximization problems). *For a given subspace $\mathcal{V} \subset \ell^2(\mathbb{N})$ with dimension larger than or equal to J , consider the following assertions:*

- (i) $\max_{\omega \in \Omega} \lambda_J^{\mathcal{V}}(\omega) = \max_{\omega \in \Omega} \lambda_J(\omega)$;
- (ii) $\mathcal{S}_* := \text{span}\{s_1, \dots, s_J\} \subseteq \mathcal{V}$, where s_1, \dots, s_J are eigenvectors corresponding to the J largest eigenvalues of $\mathbf{A}(\omega_*)$ at some $\omega_* \in \arg \max_{\omega \in \Omega} \lambda_J(\omega)$.

Assertion (ii) implies assertion (i). Furthermore, when $J = 1$, assertions (i) and (ii) are equivalent.

Proof. Suppose $s_1, \dots, s_J \in \mathcal{V}$ satisfies assertion (ii). By Lemma 2.3, we have

$$\max_{\omega \in \Omega} \lambda_J(\omega) = \lambda_J(\omega_*) = \lambda_J^{\mathcal{V}}(\omega_*) \leq \max_{\omega \in \Omega} \lambda_J^{\mathcal{V}}(\omega).$$

Lemma 2.2 implies the opposite inequality, proving assertion (i).

To prove that the assertions are equivalent when $J = 1$, assume that assertion (i) holds. Letting ω_* be any point in $\arg \max_{\omega \in \Omega} \lambda_1^{\mathcal{V}}(\omega)$, denoting by \mathbf{V} an operator defined as in (2.1) or (2.4) in terms of an orthonormal basis V for \mathcal{V} , and denoting by α a unit eigenvector corresponding to the largest eigenvalue of $\mathbf{V}^* \mathbf{A}(\omega_*) \mathbf{V}$ (note that $\lambda_1^{\mathcal{V}}(\omega_*) = \lambda_1(\omega_*) > 0$, so the unit eigenvector α is well-defined), we have

$$(2.7) \quad \max_{\omega \in \mathbb{R}^d} \lambda_1(\omega) = \lambda_1^{\mathcal{V}}(\omega_*) = \langle \mathbf{V}^* \mathbf{A}(\omega_*) \mathbf{V} \alpha, \alpha \rangle = \langle \mathbf{A}(\omega_*) \mathbf{V} \alpha, \mathbf{V} \alpha \rangle \leq \lambda_1(\omega_*).$$

Thus we deduce

$$\lambda_1(\omega_*) = \max_{\omega \in \Omega} \lambda_1(\omega) = \langle \mathbf{A}(\omega_*) \mathbf{V} \alpha, \mathbf{V} \alpha \rangle.$$

Consequently, $\mathbf{V} \alpha$ is a unit eigenvector of $\mathbf{A}(\omega_*)$ corresponding to its largest eigenvalue, where ω_* belongs to $\arg \max_{\omega \in \Omega} \lambda_1(\omega)$. Furthermore, $\text{span}\{\mathbf{V} \alpha\} \subseteq \mathcal{V}$. This proves assertion (ii). \square

When \mathcal{V} does not contain the optimal subspace \mathcal{S}_* (as in part (ii) of Lemma 2.4), the next result quantifies the gap between the eigenvalues of the original and the reduced operators in terms of the distance from \mathcal{S}_* to the J dimensional subspaces of \mathcal{V} . For this result, we define the distance between two finite dimensional subspaces $\tilde{\mathcal{S}}, \mathcal{S}$ of $\ell^2(\mathbb{N})$ of the same dimension by

$$d(\tilde{\mathcal{S}}, \mathcal{S}) := \max_{\tilde{v} \in \tilde{\mathcal{S}}, \|\tilde{v}\|_2=1} \min_{v \in \mathcal{S}} \|\tilde{v} - v\|_2.$$

This distance corresponds to the sine of the largest angle between the subspaces $\tilde{\mathcal{S}}$ and \mathcal{S} . Results of a similar nature can be found in the literature; see, for instance, [29, Theorem 11.7.1] and [32, Proposition 4.5] where the bounds are in terms of distances between one dimensional subspaces.

THEOREM 2.5 (accuracy of reduced problems). *Let \mathcal{V} be a subspace of $\ell^2(\mathbb{N})$ with dimension J or larger.*

- (i) *For each ω in $\overline{\Omega}$, we have*

$$(2.8) \quad \lambda_J(\omega) - \lambda_J^{\mathcal{V}}(\omega) = O(\varepsilon^2),$$

where

$$\varepsilon := \min \left\{ d(\tilde{\mathcal{S}}, \mathcal{S}) \mid \tilde{\mathcal{S}} \text{ is a } J \text{ dimensional subspace of } \mathcal{V} \right\}$$

and \mathcal{S} is the subspace spanned by the eigenvectors corresponding to the J largest eigenvalues of $\mathbf{A}(\omega)$.

(ii) *The equality*

$$(2.9) \quad \max_{\omega \in \Omega} \lambda_J(\omega) - \max_{\omega \in \Omega} \lambda_J^{\mathcal{V}}(\omega) = O(\varepsilon_*^2)$$

holds. Here, ε_* is given by

$$\varepsilon_* := \min \left\{ d(\tilde{\mathcal{S}}, \mathcal{S}_*) \mid \tilde{\mathcal{S}} \text{ is a } J \text{ dimensional subspace of } \mathcal{V} \right\}$$

for some subspace \mathcal{S}_* spanned by the eigenvectors corresponding to the J largest eigenvalues of $\mathbf{A}(\omega_*)$ at some $\omega_* \in \arg \max_{\omega \in \Omega} \lambda_J(\omega)$.

Proof. (i) Let $\hat{\mathcal{S}}$ be a J dimensional subspace of \mathcal{V} such that $\varepsilon = d(\hat{\mathcal{S}}, \mathcal{S})$. Furthermore, for a given unit vector $\hat{v} \in \hat{\mathcal{S}}$, let us use the notations

$$v(\hat{v}) := \arg \min \{ \|\hat{v} - v\|_2 \mid v \in \mathcal{S} \} \quad \text{and} \quad \delta(\hat{v}) := \hat{v} - v(\hat{v}),$$

where $\|\delta(\hat{v})\|_2 = O(\varepsilon)$. Observe that the inner minimization problem in the definition of $d(\hat{\mathcal{S}}, \mathcal{S})$ is a least-squares problem, so the minimizer $v(\hat{v})$ of $\|\hat{v} - v\|_2$ over $v \in \mathcal{S}$ defined above is unique and $\delta(\hat{v}) \perp \mathcal{S}$. Additionally, $\|v(\hat{v})\|_2^2 = 1 - O(\varepsilon^2)$ due to the properties $\|\hat{v}\|_2 = \|v(\hat{v}) + \delta(\hat{v})\|_2 = 1$ and $v(\hat{v}) \perp \delta(\hat{v})$.

Now the maximin characterization (2.5) for $\lambda_J^{\mathcal{V}}(\omega)$ implies

$$\begin{aligned} \lambda_J^{\mathcal{V}}(\omega) &\geq \min_{\hat{v} \in \hat{\mathcal{S}}, \|\hat{v}\|_2=1} \langle \mathbf{A}(\omega)\hat{v}, \hat{v} \rangle = \min_{\hat{v} \in \hat{\mathcal{S}}, \|\hat{v}\|_2=1} \langle \mathbf{A}(\omega) [v(\hat{v}) + \delta(\hat{v})], [v(\hat{v}) + \delta(\hat{v})] \rangle \\ &\geq \min_{\hat{v} \in \hat{\mathcal{S}}, \|\hat{v}\|_2=1} \langle \mathbf{A}(\omega)v(\hat{v}), v(\hat{v}) \rangle \\ &\quad + \min_{\hat{v} \in \hat{\mathcal{S}}, \|\hat{v}\|_2=1} 2\Re(\langle \mathbf{A}(\omega)v(\hat{v}), \delta(\hat{v}) \rangle) \\ &\quad + \min_{\hat{v} \in \hat{\mathcal{S}}, \|\hat{v}\|_2=1} \langle \mathbf{A}(\omega)\delta(\hat{v}), \delta(\hat{v}) \rangle \\ &\geq \lambda_J(\omega) \min_{\hat{v} \in \hat{\mathcal{S}}, \|\hat{v}\|_2=1} \|v(\hat{v})\|_2^2 - O(\varepsilon^2) \\ &= \lambda_J(\omega) - O(\varepsilon^2). \end{aligned}$$

Above, on the third line, we note that $\langle \mathbf{A}(\omega)v(\hat{v}), \delta(\hat{v}) \rangle = 0$ due to the fact $\mathbf{A}(\omega)v(\hat{v}) \in \mathcal{S}$ and $\delta(\hat{v}) \perp \mathcal{S}$, and on the second to the last line, we employ $\|v(\hat{v})\|_2^2 = 1 - O(\varepsilon^2)$. The desired equality (2.8) follows from $\lambda_J^{\mathcal{V}}(\omega) \leq \lambda_J(\omega)$ due to Lemma 2.2.

(ii) This is an immediate corollary of (i). In particular,

$$O(\varepsilon_*^2) = \lambda_J(\omega_*) - \lambda_J^{\mathcal{V}}(\omega_*) \geq \max_{\omega \in \Omega} \lambda_J(\omega) - \max_{\omega \in \Omega} \lambda_J^{\mathcal{V}}(\omega)$$

combined with $\max_{\omega \in \Omega} \lambda_J^{\mathcal{V}}(\omega) \leq \max_{\omega \in \Omega} \lambda_J(\omega)$ (due to Lemma 2.2) yields (2.9). \square

2.2. The greedy procedure. The basic greedy procedure solves the reduced eigenvalue optimization problem (2.3) for a given subspace \mathcal{V} . Denoting a global optimizer of the reduced problem with ω_* , the subspace is expanded with the addition of the eigenvectors corresponding to $\lambda_1(\omega_*), \dots, \lambda_J(\omega_*)$, then this is repeated with the expanded subspace. A formal description is given in Algorithm 1, where \mathcal{S}_k denotes the subspace at the k th step of the procedure, and $\lambda_J^{(k)}(\omega) := \lambda_J^{\mathcal{S}_k}(\omega)$. The reduced eigenvalue optimization problems on line 5 are nonsmooth and nonconvex. The description assumes that these problems can be solved globally. The algorithm in [26] works well in practice for this purpose when the number of parameters, d , is small.

These reduced problems are computationally cheap to solve, the main computational burden comes from line 6 which requires the computation of eigenvectors of the full problem. In the finite dimensional case, these large eigenvalue problems are typically solved by means of an iterative method, for instance by Lanczos' method.

Algorithm 1 The greedy subspace procedure.

Require: A parameter dependent compact self-adjoint operator $\mathbf{A}(\omega)$ of the form (1.1), and a compact subset $\Omega \subset \bar{\Omega}$ of \mathbb{R}^d .

- 1: $\omega^{(1)} \leftarrow$ a random point in Ω .
 - 2: $s_1^{(1)}, \dots, s_J^{(1)} \leftarrow$ eigenvectors corresponding to $\lambda_1(\omega^{(1)}), \dots, \lambda_J(\omega^{(1)})$.
 - 3: $\mathcal{S}_1 \leftarrow \text{span} \{s_1^{(1)}, \dots, s_J^{(1)}\}$.
 - 4: **for** $k = 2, 3, \dots$ **do**
 - 5: $\omega^{(k)} \leftarrow$ any $\omega_* \in \arg \min_{\omega \in \Omega} \lambda_J^{(k-1)}(\omega)$ for the minimization problem (MN), or
 $\omega^{(k)} \leftarrow$ any $\omega_* \in \arg \max_{\omega \in \Omega} \lambda_J^{(k-1)}(\omega)$ for the maximization problem (MX).
 - 6: $s_1^{(k)}, \dots, s_J^{(k)} \leftarrow$ eigenvectors corresponding to $\lambda_1(\omega^{(k)}), \dots, \lambda_J(\omega^{(k)})$.
 - 7: $\mathcal{S}_k \leftarrow \mathcal{S}_{k-1} \oplus \text{span} \{s_1^{(k)}, \dots, s_J^{(k)}\}$.
 - 8: **end for**
-

As numerical experiments demonstrate (see section 5), the power of this greedy subspace procedure is that high accuracy is often reached after a small number of steps for reduced problems of small size. This can mostly be attributed to the following interpolatory properties between $\lambda_J^{(k)}(\omega)$ and $\lambda_J(\omega)$.

LEMMA 2.6 (Hermite interpolation). *The following hold regarding Algorithm 1:*

- (i) $\lambda_J(\omega^{(\ell)}) = \lambda_J^{(k)}(\omega^{(\ell)})$ for $\ell = 1, \dots, k$;
- (ii) if $J > 1$, then $\lambda_{J-1}(\omega^{(\ell)}) = \lambda_{J-1}^{(k)}(\omega^{(\ell)})$ for $\ell = 1, \dots, k$;
- (iii) if $\lambda_J(\omega^{(\ell)})$ is simple, then $\lambda_J^{(k)}(\omega^{(\ell)})$ is also simple for $\ell = 1, 2, \dots, k$;
- (iv) if $\lambda_J(\omega^{(\ell)})$ is simple, then $\nabla \lambda_J(\omega^{(\ell)}) = \nabla \lambda_J^{(k)}(\omega^{(\ell)})$ for $\ell = 1, 2, \dots, k$.

Proof. (i)–(ii) Lines 2, 3, 6, and 7 of Algorithm 1 imply that $s_1^{(\ell)}, \dots, s_J^{(\ell)} \in \mathcal{S}_k$ for $\ell = 1, \dots, k$. An application of part (i) of Lemma 2.3 with $\mathcal{V} = \mathcal{S}_k$ yields $\lambda_j(\omega^{(\ell)}) = \lambda_j^{\mathcal{S}_k}(\omega^{(\ell)}) = \lambda_j^{(k)}(\omega^{(\ell)})$ for $\ell = 1, \dots, k$ and $j = J$, as well as $j = J - 1$ if $J > 1$, as desired.

(iii) Suppose $\lambda_J^{(k)}(\omega^{(\ell)})$ is not simple for some $\ell \in \{1, 2, \dots, k\}$. In this case there must exist two mutually orthogonal unit eigenvectors $\hat{\alpha}, \tilde{\alpha} \in \mathbb{C}^m$ corresponding to it. Let us denote by \mathbf{S}_k the operator as in (2.1) in terms of a basis \mathcal{S}_k for \mathcal{S}_k , and $\mathbf{A}^{\mathcal{S}_k}(\omega^{(\ell)}) = \mathbf{S}_k^* \mathbf{A}(\omega^{(\ell)}) \mathbf{S}_k$. It follows from part (i) that

$$\begin{aligned} \lambda_J(\omega^{(\ell)}) &= \lambda_J^{(k)}(\omega^{(\ell)}) = \langle \mathbf{A}(\omega^{(\ell)}) \mathbf{S}_k \hat{\alpha}, \mathbf{S}_k \hat{\alpha} \rangle = \min_{\alpha \in \hat{\mathcal{S}}, \|\alpha\|_2=1} \langle \mathbf{A}(\omega^{(\ell)}) \mathbf{S}_k \alpha, \mathbf{S}_k \alpha \rangle \\ &= \langle \mathbf{A}(\omega^{(\ell)}) \mathbf{S}_k \tilde{\alpha}, \mathbf{S}_k \tilde{\alpha} \rangle = \min_{\alpha \in \tilde{\mathcal{S}}, \|\alpha\|_2=1} \langle \mathbf{A}(\omega^{(\ell)}) \mathbf{S}_k \alpha, \mathbf{S}_k \alpha \rangle \end{aligned}$$

for some J dimensional subspaces $\hat{\mathcal{S}}, \tilde{\mathcal{S}}$ of $\ell^2(\mathbb{N})$ such that $\hat{\alpha} \in \hat{\mathcal{S}}, \tilde{\alpha} \in \tilde{\mathcal{S}}$. This shows that $\mathbf{S}_k \hat{\alpha}, \mathbf{S}_k \tilde{\alpha} \in \ell^2(\mathbb{N})$ are mutually orthogonal eigenvectors corresponding to $\lambda_J(\omega^{(\ell)})$, so $\lambda_J(\omega^{(\ell)})$ is not simple either.

(iv) It follows from part (iii) that $\lambda_J^{(k)}(\omega^{(\ell)})$ is also simple, so both $\lambda_J(\omega)$ and $\lambda_J^{(k)}(\omega)$ are differentiable at $\omega^{(\ell)}$; furthermore the associated unit eigenvectors can

be chosen in a way so that they are also differentiable at $\omega^{(\ell)}$ (see [31, pp. 57–58, Theorem 1] for the differentiability of $\lambda_J(\omega)$ and the associated unit eigenvector, and [31, pp. 33–34, Theorem 1] for the differentiability of $\lambda_J^{(k)}(\omega)$ and the associated unit eigenvector). By Lemma 2.3(ii), the eigenvector α_J of $\mathbf{A}^{\mathcal{S}_k}(\omega^{(\ell)}) = \mathbf{S}_k^* \mathbf{A}(\omega^{(\ell)}) \mathbf{S}_k$ corresponding to the eigenvalue $\lambda_J^{(k)}(\omega^{(\ell)}) = \lambda_J^{\mathcal{S}_k}(\omega^{(\ell)})$ satisfies $\alpha_J = \mathbf{S}_k^* s_J^{(\ell)}$. Equivalently, we have $s_J^{(\ell)} = \mathbf{S}_k \alpha_J$ (since $s_J^{(\ell)} \in \mathcal{S}_k$). By employing the analytical formulas for the derivatives of eigenvalue functions (see [23] for the finite dimensional matrix-valued case whose derivation exploits the Hermiticity of the matrix-valued function; the generalization to the infinite dimensional case, leading to the formula $\partial \lambda_J(\omega) / \partial \omega_q = \langle \partial \mathbf{A}(\omega) / \partial \omega_q s_J, s_J \rangle$, where s_J is a unit eigenvector corresponding to $\lambda_J(\omega)$, is straightforward by making use of the self-adjointness of $\mathbf{A}(\omega)$), for $q = 1, \dots, d$ we obtain

$$\begin{aligned} \frac{\partial \lambda_J^{(k)}(\omega^{(\ell)})}{\partial \omega_q} &= \left\langle \frac{\partial \mathbf{A}^{\mathcal{S}_k}(\omega^{(\ell)})}{\partial \omega_q} \alpha_J, \alpha_J \right\rangle \\ &= \left\langle \frac{\partial \mathbf{A}(\omega^{(\ell)})}{\partial \omega_q} \mathbf{S}_k \alpha_J, \mathbf{S}_k \alpha_J \right\rangle \\ &= \left\langle \frac{\partial \mathbf{A}(\omega^{(\ell)})}{\partial \omega_q} s_J^{(\ell)}, s_J^{(\ell)} \right\rangle = \frac{\partial \lambda_J(\omega^{(\ell)})}{\partial \omega_q}. \end{aligned}$$

This completes the proof. □

2.3. The extended greedy procedure. To better exploit the Hermite interpolation properties of Lemma 2.6, we extend the basic greedy procedure of the previous subsection with the inclusion of additional eigenvectors in the subspaces at points close to the optimizers of the reduced problems. The purpose here is to achieve

$$\lim_{k \rightarrow \infty} \left\| \nabla^2 \lambda_J(\omega^{(k)}) - \nabla^2 \lambda_J^{(k)}(\omega^{(k)}) \right\|_2 = 0$$

in addition to $\lambda_J(\omega^{(k)}) = \lambda_J^{(k)}(\omega^{(k)})$ and $\nabla \lambda_J(\omega^{(k)}) = \nabla \lambda_J^{(k)}(\omega^{(k)})$. These properties enable us to make an analogy with a quasi-Newton method for unconstrained smooth optimization, and come up with a theoretical superlinear rate-of-convergence result in the next section.

In the extended procedure also the reduced eigenvalue optimization problem (2.3) is solved for a subspace \mathcal{V} already constructed. Denoting the optimizer of the reduced problem with ω_* , in addition to the eigenvectors corresponding to $\lambda_1(\omega_*), \dots, \lambda_J(\omega_*)$, the eigenvectors corresponding to $\lambda_1(\omega_* + h e_{pq}), \dots, \lambda_J(\omega_* + h e_{pq})$ are added into the subspace \mathcal{V} for h decaying to zero if convergence occurs, for $p = 1, \dots, d$ and $q = p, \dots, d$. Here $e_{pq} := (1/\sqrt{2})(e_p + e_q)$ for $p \neq q$ as well as $e_{pp} := e_p$. We provide a formal description of this extended subspace procedure in Algorithm 2 below.

The reduced eigenvalue functions of the extended procedure possesses additional interpolatory properties stated in the lemmas below. Their proofs are similar to the proofs for Lemma 2.6, so we omit them.

LEMMA 2.7 (extended Hermite interpolation). *The assertions of Lemma 2.6 hold for Algorithm 2. Additionally, we have*

- (i) $\lambda_J(\omega^{(\ell)} + h^{(\ell)} e_{pq}) = \lambda_J^{(k)}(\omega^{(\ell)} + h^{(\ell)} e_{pq})$;
- (ii) if $\lambda_J(\omega^{(\ell)} + h^{(\ell)} e_{pq})$ is simple, then $\lambda_J^{(k)}(\omega^{(\ell)} + h^{(\ell)} e_{pq})$ is also simple;

Algorithm 2 The extended greedy subspace procedure.

Require: A parameter dependent compact self-adjoint operator $\mathbf{A}(\omega)$ of the form (1.1), and a compact subset $\Omega \subset \bar{\Omega}$ of \mathbb{R}^d .

- 1: $\omega^{(1)} \leftarrow$ a random point in Ω .
 - 2: $s_1^{(1)}, \dots, s_J^{(1)} \leftarrow$ eigenvectors corresponding to $\lambda_1(\omega^{(1)}), \dots, \lambda_J(\omega^{(1)})$.
 - 3: $\mathcal{S}_1 \leftarrow \text{span} \{s_1^{(1)}, \dots, s_J^{(1)}\}$.
 - 4: **for** $k = 2, 3, \dots$ **do**
 - 5: $\omega^{(k)} \leftarrow$ any $\omega_* \in \arg \min_{\omega \in \Omega} \lambda_J^{(k-1)}(\omega)$ for the minimization problem (MN), or
 $\omega^{(k)} \leftarrow$ any $\omega_* \in \arg \max_{\omega \in \Omega} \lambda_J^{(k-1)}(\omega)$ for the maximization problem (MX).
 - 6: $s_1^{(k)}, \dots, s_J^{(k)} \leftarrow$ eigenvectors corresponding to $\lambda_1(\omega^{(k)}), \dots, \lambda_J(\omega^{(k)})$.
 - 7: $h^{(k)} \leftarrow \|\omega^{(k)} - \omega^{(k-1)}\|_2$
 - 8: **for** $p = 1, \dots, d$, $q = p, \dots, d$ **do**
 - 9: $s_{1,pq}^{(k)}, \dots, s_{J,pq}^{(k)} \leftarrow$ eigenvectors corresponding to
 $\lambda_1(\omega^{(k)} + h^{(k)}e_{pq}), \dots, \lambda_J(\omega^{(k)} + h^{(k)}e_{pq})$.
 - 10: **end for**
 - 11: $\mathcal{S}_k \leftarrow \mathcal{S}_{k-1} \oplus \text{span} \{s_1^{(k)}, \dots, s_J^{(k)}\} \oplus \left\{ \bigoplus_{p=1, q=p}^d \text{span} \{s_{1,pq}^{(k)}, \dots, s_{J,pq}^{(k)}\} \right\}$.
 - 12: **end for**
-

(iii) if $\lambda_J(\omega^{(\ell)} + h^{(\ell)}e_{pq})$ is simple, $\nabla \lambda_J(\omega^{(\ell)} + h^{(\ell)}e_{pq}) = \nabla \lambda_J^{(k)}(\omega^{(\ell)} + h^{(\ell)}e_{pq})$ for every $\ell = 1, \dots, k$, $p = 1, \dots, d$, and $q = p, \dots, d$.

The main motivation for the inclusion of additional eigenvectors in the subspaces is the deduction of theoretical bounds on the proximity of the second derivatives, which we present next. This result is initially established under the assumption that the third derivatives of the reduced eigenvalue functions $\omega \mapsto \lambda_J^{(k)}(\omega)$ are bounded uniformly with respect to k provided k is large enough. Subsequently, we show in Proposition 2.9 that this assumption is always satisfied, hence, can be dropped.

LEMMA 2.8. *Suppose that the sequence $\{\omega^{(k)}\}$ by Algorithm 2 (or Algorithm 1 when $d = 1$ by defining $h^{(k)} := |\omega^{(k)} - \omega^{(k-1)}|$) is convergent, that its limit $\omega_* := \lim_{k \rightarrow \infty} \omega^{(k)}$ lies strictly in the interior of Ω , and that $\lambda_J(\omega_*)$ is simple. Assume, furthermore, that in an open ball containing ω_* , all third derivatives of functions $\omega \mapsto \lambda_J^{(k)}(\omega)$ are bounded uniformly with respect to $k \geq k_0$ with k_0 sufficiently large. Then the following assertions hold:*

- (i) *There exists a constant $C > 0$ such that*

$$(2.10) \quad \left| \frac{\partial^2 \lambda_J(\omega^{(k)})}{\partial \omega_p \partial \omega_q} - \frac{\partial^2 \lambda_J^{(k)}(\omega^{(k)})}{\partial \omega_p \partial \omega_q} \right|_2 \leq Ch^{(k)} \quad \text{for } p, q = 1, \dots, d,$$

in particular,

$$\left\| \nabla^2 \lambda_J(\omega^{(k)}) - \nabla^2 \lambda_J^{(k)}(\omega^{(k)}) \right\|_2 \leq dCh^{(k)}$$

for all k large enough.

- (ii) *Additionally, if $\nabla^2 \lambda_J(\omega_*)$ is invertible, then*

$$(2.11) \quad \left\| \left[\nabla^2 \lambda_J(\omega^{(k)}) \right]^{-1} - \left[\nabla^2 \lambda_J^{(k)}(\omega^{(k)}) \right]^{-1} \right\|_2 = O(h^{(k)})$$

for all k large enough.

Proof. (i) First we specify a ball centered at ω_* in which $\lambda_J(\omega)$ and $\lambda_J^{(k)}(\omega)$ for all large k are simple. The argument initially assumes $J > 1$. Letting $\varepsilon := \min\{\lambda_{J-1}(\omega_*) - \lambda_J(\omega_*), \lambda_J(\omega_*) - \lambda_{J+1}(\omega_*)\}$, consider the ball $\mathcal{B}(\omega_*, \varepsilon/(8\gamma))$, where γ is the Lipschitz constant as in Lemma 2.1. Without loss of generality, let us assume $\mathcal{B}(\omega_*, \varepsilon/(8\gamma)) \subseteq \Omega$ (i.e., otherwise choose ε even smaller so that $\mathcal{B}(\omega_*, \varepsilon/(8\gamma)) \subseteq \Omega$). Now, due to $\lambda_{J-1}(\omega_*) - \lambda_J(\omega_*) \geq \varepsilon$ as well as $\lambda_J(\omega_*) - \lambda_{J+1}(\omega_*) \geq \varepsilon$, and by Lemma 2.1, we have

$$(2.12) \quad \lambda_{J-1}(\omega) - \lambda_J(\omega) \geq 3\varepsilon/4 \quad \text{and} \quad \lambda_J(\omega) - \lambda_{J+1}(\omega) \geq 3\varepsilon/4 \quad \forall \omega \in \mathcal{B}(\omega_*, \varepsilon/(8\gamma)).$$

Next choose k large enough so that $\mathcal{B}(\omega^{(k)}, h^{(k)}) \subset \mathcal{B}(\omega_*, \varepsilon/(8\gamma))$. We will show that $\lambda_J^{(k)}(\omega)$ is also simple in $\mathcal{B}(\omega_*, \varepsilon/(8\gamma))$. In this respect we note that $\lambda_j(\omega^{(k)}) = \lambda_j^{(k)}(\omega^{(k)})$ for $j = J - 1, J$ due to parts (i) and (ii) of Lemma 2.6, so from (2.12) we have

$$\begin{aligned} \lambda_{J-1}^{(k)}(\omega^{(k)}) - \lambda_J^{(k)}(\omega^{(k)}) &\geq 3\varepsilon/4 \quad \text{and} \\ \lambda_J^{(k)}(\omega^{(k)}) - \lambda_{J+1}^{(k)}(\omega^{(k)}) &\geq \lambda_J^{(k)}(\omega^{(k)}) - \lambda_{J+1}(\omega^{(k)}) \geq 3\varepsilon/4, \end{aligned}$$

where in the last line we also used $\lambda_{J+1}^{(k)}(\omega^{(k)}) \leq \lambda_{J+1}(\omega^{(k)})$ which holds due to monotonicity (Lemma 2.2). Another application of Lemma 2.1 yields

$$\lambda_{J-1}^{(k)}(\omega) - \lambda_J^{(k)}(\omega) \geq \varepsilon/4 \quad \text{and} \quad \lambda_J^{(k)}(\omega) - \lambda_{J+1}^{(k)}(\omega) \geq \varepsilon/4 \quad \forall \omega \in \mathcal{B}(\omega_*, \varepsilon/(8\gamma)).$$

We remark that the argument above applies to the case $J = 1$ trivially by considering only the gaps $\lambda_J(\omega) - \lambda_{J+1}(\omega)$ as well as $\lambda_J^{(k)}(\omega) - \lambda_{J+1}^{(k)}(\omega)$ and showing they remain bounded away from zero for all ω inside $\mathcal{B}(\omega_*, \varepsilon/(8\gamma))$.

We prove the desired bounds (2.10) first for the sequence $\{\omega^{(k)}\}$ generated by Algorithm 2. The simplicity of the eigenvalue functions $\lambda_J(\omega)$ and $\lambda_J^{(k)}(\omega)$ on $\mathcal{B}(\omega_*, \varepsilon/(8\gamma))$ imply that, for each p, q , the functions

$$(2.13) \quad \ell_{pq}(t) := \lambda_J(\omega^{(k)} + th^{(k)}e_{pq}) \quad \text{and} \quad \ell_{pq,k}(t) := \lambda_J^{(k)}(\omega^{(k)} + th^{(k)}e_{pq})$$

are analytic on $(0, 1)$. By applying Taylor's theorem to $\ell_{pq}(t), \ell_{pq,k}(t)$ on the interval $(0, 1)$, we obtain

$$\begin{aligned} \ell_{pq}(1) &= \ell_{pq}(0) + \ell'_{pq}(0) + \ell''_{pq}(0)/2 + \ell'''_{pq}(\eta)/6, \\ \ell_{pq,k}(1) &= \ell_{pq,k}(0) + \ell'_{pq,k}(0) + \ell''_{pq,k}(0)/2 + \ell'''_{pq,k}(\eta^{(k)})/6 \end{aligned}$$

for some $\eta, \eta^{(k)} \in (0, 1)$. Now by employing $\ell_{pq}(0) = \ell_{pq,k}(0)$, $\ell'_{pq}(0) = \ell'_{pq,k}(0)$ due to parts (i), (iv) of Lemma 2.6, and $\ell_{pq}(1) = \ell_{pq,k}(1)$ due to part (i) of Lemma 2.7, we deduce

$$\frac{\ell''_{pq}(0) - \ell''_{pq,k}(0)}{2} = \frac{\ell'''_{pq,k}(\eta^{(k)}) - \ell'''_{pq}(\eta)}{6}.$$

In the last expression,

$$\ell''_{pq}(0) = [h^{(k)}]^2 e_{pq}^T \nabla^2 \lambda_J(\omega^{(k)}) e_{pq} \quad \text{and} \quad \ell''_{pq,k}(0) = [h^{(k)}]^2 e_{pq}^T \nabla^2 \lambda_J^{(k)}(\omega^{(k)}) e_{pq}$$

as well as $\ell''_{pq,k}(\eta^{(k)}) = O([h^{(k)}]^3)$ and $\ell'''_{pq}(\eta) = O([h^{(k)}]^3)$, so we have

$$(2.14) \quad \frac{[h^{(k)}]^2 \left| e_{pq}^T \left[\frac{\nabla^2 \lambda_J(\omega^{(k)}) - \nabla^2 \lambda_J^{(k)}(\omega^{(k)})}{2} \right] e_{pq} \right|}{2} \leq \frac{D}{6} [h^{(k)}]^3$$

for some constant D independent of k .

Choosing $q = p$ in (2.14) yields

$$\left| \frac{\partial^2 \lambda_J(\omega^{(k)})}{\partial \omega_p^2} - \frac{\partial^2 \lambda_J^{(k)}(\omega^{(k)})}{\partial \omega_p^2} \right| \leq \frac{D}{3} h^{(k)}.$$

Additionally, for $q \neq p$, rewriting (2.14) as

$$\left| \frac{1}{2} \left(\sum_{\ell=p,q} \frac{\partial^2 \lambda_J(\omega^{(k)})}{\partial \omega_\ell^2} - \frac{\partial^2 \lambda_J^{(k)}(\omega^{(k)})}{\partial \omega_\ell^2} \right) + \left(\frac{\partial^2 \lambda_J(\omega^{(k)})}{\partial \omega_p \partial \omega_q} - \frac{\partial^2 \lambda_J^{(k)}(\omega^{(k)})}{\partial \omega_p \partial \omega_q} \right) \right| \leq \frac{D}{3} h^{(k)},$$

we obtain

$$\left| \frac{\partial^2 \lambda_J(\omega^{(k)})}{\partial \omega_p \partial \omega_q} - \frac{\partial^2 \lambda_J^{(k)}(\omega^{(k)})}{\partial \omega_p \partial \omega_q} \right| \leq \frac{2D}{3} h^{(k)}$$

leading us to (2.10). The proof above establishing the bounds (2.10) also applies to the sequence $\{\omega^{(k)}\}$ generated by Algorithm 1 when $d = 1$ by defining $\tilde{h}^{(k)} := \omega^{(k-1)} - \omega^{(k)}$ (so that $h^{(k)} = |\tilde{h}^{(k)}|$) and letting

$$\ell_{pq}(t) := \lambda_J(\omega^{(k)} + t\tilde{h}^{(k)}e_{pq}) \quad \text{and} \quad \ell_{pq,k}(t) := \lambda_J^{(k)}(\omega^{(k)} + t\tilde{h}^{(k)}e_{pq})$$

instead of (2.13).

(ii) From part (i), as well as by the existence of $\nabla^2 \lambda_J(\omega^{(k)})$, $\nabla^2 \lambda_J^{(k)}(\omega^{(k)})$ for all k large enough so that $\mathcal{B}(\omega^{(k)}, h^{(k)}) \subset \mathcal{B}(\omega_*, \varepsilon/(8\gamma))$ and by the continuity of $\nabla^2 \lambda_J(\omega)$ at $\omega = \omega_*$, we have

$$\lim_{k \rightarrow \infty} \nabla^2 \lambda_J(\omega^{(k)}) = \lim_{k \rightarrow \infty} \nabla^2 \lambda_J^{(k)}(\omega^{(k)}) = \nabla^2 \lambda_J(\omega_*).$$

Exploiting this and the invertibility of $\nabla^2 \lambda_J(\omega_*)$, we deduce that $\nabla^2 \lambda_J(\omega^{(k)})$ and $\nabla^2 \lambda_J^{(k)}(\omega^{(k)})$ are invertible for all k large enough.

For such a k , letting $A = \nabla^2 \lambda_J(\omega^{(k)})$, $\tilde{A} = \nabla^2 \lambda_J^{(k)}(\omega^{(k)})$ as well as $X = [\nabla^2 \lambda_J(\omega^{(k)})]^{-1}$, $\tilde{X} = [\nabla^2 \lambda_J^{(k)}(\omega^{(k)})]^{-1}$, the classical adjugate formula (i.e., $[B^{-1}]_{ij} = \frac{(-1)^{i+j} \det(B_{ji})}{\det(B)}$ for an invertible B , where B_{ji} denotes the matrix obtained from B by removing its j th row and the i th column) yields

$$\tilde{x}_{ij} = \frac{(-1)^{i+j} \det(\tilde{A}_{ji})}{\det(\tilde{A})} \quad \text{for } i, j = 1, \dots, d.$$

By part (i), in particular (2.10), we have $\det(\tilde{A}_{ji}) = \det(A_{ji}) + O(h^{(k)})$ and $\det(\tilde{A}) = \det(A) + O(h^{(k)})$, so we deduce

$$\tilde{x}_{ij} = \frac{(-1)^{i+j} \det(A_{ji}) + O(h^{(k)})}{\det(A) + O(h^{(k)})} = \frac{(-1)^{i+j} \det(A_{ji})}{\det(A)} + O(h^{(k)}) = x_{ij} + O(h^{(k)})$$

leading to (2.11). \square

PROPOSITION 2.9. *Suppose that the sequence $\{\omega^{(k)}\}$ by Algorithm 1 or Algorithm 2 is convergent, its limit $\omega_* := \lim_{k \rightarrow \infty} \omega^{(k)}$ lies strictly in the interior of Ω , and that $\lambda_J(\omega_*)$ is simple. Then for k sufficiently large all third derivatives of functions $\omega \mapsto \lambda_J^{(k)}(\omega)$ exist in an open ball containing ω_* , and they can be bounded uniformly with respect to k .*

Proof. For the sake of clarity we first consider the case when $d = 1$. Following from the assumptions of the proposition and from the elements spelled out in the proof of Lemma 2.8, there exists an open interval $\mathcal{I} := (\omega_* - r_1, \omega_* + r_1)$, and numbers $k_1 \in \mathbb{N}$ and $\ell \in \mathbb{R}, \ell > 0$, such that for every $k \geq k_1$ and all $\omega \in \mathcal{I}$, the eigenvalue $\lambda_J^{(k)}(\omega)$ is simple, as well as

$$(2.15) \quad \min \left\{ \left| \lambda_J^{(k)}(\omega) - \lambda_{J-1}^{(k)}(\omega) \right|, \left| \lambda_J^{(k)}(\omega) - \lambda_{J+1}^{(k)}(\omega) \right| \right\} \geq \ell.$$

Since functions f_ℓ are assumed real analytic, there exist analytic extensions on the complex plane at ω_* , which we refer to by functions \hat{f}_ℓ , $\ell = 1, \dots, \kappa$. We denote by $\hat{\mathbf{A}}^{S_k}(\omega)$ the corresponding extension of $\mathbf{A}^{S_k}(\omega)$. Following the ideas spelled out in the proof of [20, Lemma 3], we can express

$$(2.16) \quad \begin{aligned} \left\| \hat{\mathbf{A}}^{S_k}(\omega) - \hat{\mathbf{A}}^{S_k}(\omega_*) \right\|_2 &\leq \sum_{\ell=1}^{\kappa} \left\| \mathbf{S}_k^* \mathbf{A}_\ell \mathbf{S}_k \right\|_2 \left| \hat{f}_\ell(\omega) - \hat{f}_\ell(\omega_*) \right| \\ &\leq \sum_{\ell=1}^{\kappa} \left\| \mathbf{A}_\ell \right\|_2 \left| \hat{f}_\ell(\omega) - \hat{f}_\ell(\omega_*) \right|. \end{aligned}$$

Note that if ω is nonreal, the difference $\hat{\mathbf{A}}^{S_k}(\omega) - \hat{\mathbf{A}}^{S_k}(\omega_*)$ might be non-Hermitian. However, an important conclusion from the above inequalities is that $\left\| \hat{\mathbf{A}}^{S_k}(\omega) - \hat{\mathbf{A}}^{S_k}(\omega_*) \right\|_2$ can be bounded from above uniformly independent of k , that is independent of the dimension of the subspace S_k , for all ω inside a disk centered at ω_* . Furthermore, this upper bound can be chosen arbitrarily close to 0 by reducing the radius of the disk.

From (2.15), (2.16), and [33, Theorem 5.1], we conclude that there exists a number $r_2 \in (0, r_1)$ independent of $k \geq k_1$ (i.e., the dimension of the subspace), such that the eigenvalue function $\hat{\lambda}_J^{(k)}$, obtained by replacing \mathbf{A}^{S_k} with its analytic extension $\hat{\mathbf{A}}^{S_k}$, remains simple on and inside the disk $\{\omega \in \mathbb{C} : |\omega - \omega_*| \leq r_2\}$ for every $k \geq k_1$. Hence $\hat{\lambda}_J^{(k)}$ is well-defined and analytic inside this disk for every $k \geq k_1$. For $\tilde{\omega} \in \mathbb{R}$ satisfying $|\tilde{\omega} - \omega_*| < r_2/2$, we have

$$\frac{d^3 \hat{\lambda}_J^{(k)}}{d\omega^3}(\tilde{\omega}) = \frac{3!}{2\pi i} \oint_{|\omega - \tilde{\omega}| = r_2/2} \frac{\hat{\lambda}_J^{(k)}(\omega)}{(\omega - \tilde{\omega})^4} d\omega$$

with i the imaginary unit. We can bound

$$(2.17) \quad \begin{aligned} \left| \frac{d^3 \hat{\lambda}_J^{(k)}}{d\omega^3}(\tilde{\omega}) \right| &\leq \frac{96}{r_2^3} \max_{\omega \in \mathbb{C}, |\omega - \tilde{\omega}| = r_2/2} \left| \hat{\lambda}_J^{(k)}(\omega) \right| \\ &\leq \frac{96}{r_2^3} \sum_{\ell=1}^{\kappa} \left\| \mathbf{A}_\ell^{S_k} \right\|_2 \left(\max_{\omega \in \mathbb{C}, |\omega - \tilde{\omega}| = r_2/2} \left| \hat{f}_\ell(\omega) \right| \right) \\ &\leq \frac{96}{r_2^3} \sum_{\ell=1}^{\kappa} \left\| \mathbf{A}_\ell \right\|_2 \left(\max_{\omega \in \mathbb{C}, |\omega - \tilde{\omega}| = r_2/2} \left| \hat{f}_\ell(\omega) \right| \right), \end{aligned}$$

hence, an upper bound is established that does not depend on $k \geq k_1$.

For the case $d > 1$ only a slight adaptation is needed for the mixed derivatives. We sketch the main principles by means of the case when $d = 2$. From the ideas behind (2.17) a bound on $\hat{\lambda}_J^{(k)}$ induces a bound, uniform in k , on its partial derivative $\frac{\partial^2 \hat{\lambda}_J^{(k)}}{\partial \omega_1^2}$ in an open set containing ω_* . Likewise the bound on the latter induces a bound on $\frac{\partial^3 \hat{\lambda}_J^{(k)}}{\partial \omega_1^2 \partial \omega_2}$. \square

3. Convergence analysis. In this section we analyze the convergence properties of the two subspace procedures introduced. The first part concerns global convergence: it elaborates on whether the iterates of Algorithm 1 and Algorithm 2 necessarily converge as the subspace dimensions grow to infinity, and if they do converge, where they converge to. The second part establishes a superlinear rate-of-convergence result for the iterates of Algorithm 2 (and Algorithm 1 when $d = 1$).

3.1. Global convergence. The subspace procedures, when applied to the minimization problem (MN), converge globally. A formal statement of this global convergence property together with its proof are given in what follows. Note that the sequence $\{\omega^{(k)}\}$ generated by Algorithm 1 or Algorithm 2 belongs to the bounded set Ω , so it must have convergent subsequences.

THEOREM 3.1. *The following hold for both Algorithm 1 and Algorithm 2.*

- (i) *The limit of every convergent subsequence of the sequence $\{\omega^{(k)}\}$ for the minimization problem (MN) is a global minimizer of $\lambda_J(\omega)$ over Ω .*
- (ii) $\lim_{k \rightarrow \infty} \lambda_J^{(k)}(\omega^{(k+1)}) = \lim_{k \rightarrow \infty} \min_{\omega \in \Omega} \lambda_J^{(k)}(\omega) = \min_{\omega \in \Omega} \lambda_J(\omega)$.

Proof. (i) Let $\{\omega^{(\ell_k)}\}$ be a convergent subsequence of $\{\omega^{(k)}\}$. By the monotonicity property, that is by Lemma 2.2, we have

$$(3.1) \quad \begin{aligned} \min_{\omega \in \Omega} \lambda_J(\omega) &\geq \min_{\omega \in \Omega} \lambda_J^{(\ell_{k+1}-1)}(\omega) \\ &= \lambda_J^{(\ell_{k+1}-1)}(\omega^{(\ell_{k+1})}) \geq \lambda_J^{(\ell_k)}(\omega^{(\ell_{k+1})}), \end{aligned}$$

and, by the interpolatory property, that is part (i) of Lemma 2.6,

$$(3.2) \quad \min_{\omega \in \Omega} \lambda_J(\omega) \leq \lambda_J(\omega^{(\ell_k)}) = \lambda_J^{(\ell_k)}(\omega^{(\ell_k)}).$$

But observe that the Lipschitz continuity of $\lambda_J^{(\ell_k)}(\omega)$ (see Lemma 2.1) implies

$$\lim_{k \rightarrow \infty} \left| \lambda_J^{(\ell_k)}(\omega^{(\ell_{k+1})}) - \lambda_J^{(\ell_k)}(\omega^{(\ell_k)}) \right| = \gamma \lim_{k \rightarrow \infty} \left\| \omega^{(\ell_{k+1})} - \omega^{(\ell_k)} \right\|_2 = 0.$$

Consequently, taking the limits in (3.1) and (3.2) as $k \rightarrow \infty$ and employing the interpolatory property (part (i) of Lemma 2.6), we deduce

$$(3.3) \quad \lim_{k \rightarrow \infty} \lambda_J(\omega^{(\ell_k)}) = \lim_{k \rightarrow \infty} \lambda_J^{(\ell_k)}(\omega^{(\ell_k)}) = \lim_{k \rightarrow \infty} \lambda_J^{(\ell_k)}(\omega^{(\ell_{k+1})}) = \min_{\omega \in \Omega} \lambda_J(\omega).$$

Finally, by the continuity of $\lambda_J(\omega)$, the sequence $\{\omega^{(\ell_k)}\}$ must converge to a global minimizer of $\lambda_J(\omega)$.

(ii) Let $\lambda_* := \min_{\omega \in \Omega} \lambda_J(\omega)$. We first show that the sequence $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$ is convergent. In this respect observe that for every pair of positive integers p, k such

that $p > k$, due to monotonicity (Lemma 2.2) we have

$$\begin{aligned} \lambda_* &\geq \min_{\omega \in \Omega} \lambda_J^{(p)}(\omega) = \lambda_J^{(p)}\left(\omega^{(p+1)}\right) \\ &\geq \lambda_J^{(k)}\left(\omega^{(p+1)}\right) \geq \min_{\omega \in \Omega} \lambda_J^{(k)}(\omega) = \lambda_J^{(k)}\left(\omega^{(k+1)}\right). \end{aligned}$$

Hence the sequence $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$ is monotonically increasing bounded above by λ_* proving its convergence.

Now let $\{\omega^{(\ell_k)}\}$ be a convergent subsequence of $\{\omega^{(k)}\}$ as in part (i), and let us consider the sequence $\{\lambda_J^{(\ell_{k+1}-1)}(\omega^{(\ell_{k+1})})\}$, which is a subsequence of $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$. We complete the proof by establishing the convergence of this subsequence to λ_* . The monotonicity and the boundedness of $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$ from above by λ_* imply

$$(3.4) \quad \lambda_J^{(\ell_k)}\left(\omega^{(\ell_{k+1})}\right) \leq \lambda_J^{(\ell_{k+1}-1)}(\omega^{(\ell_{k+1})}) \leq \lambda_*.$$

Above, as shown in part (i) (in particular, see (3.3)), $\lim_{k \rightarrow \infty} \lambda_J^{(\ell_k)}(\omega^{(\ell_{k+1})}) = \lambda_*$, so we must also have $\lim_{k \rightarrow \infty} \lambda_J^{(\ell_{k+1}-1)}(\omega^{(\ell_{k+1})}) = \lambda_*$ as desired. \square

As for the maximization problem (MX), it does not seem possible to conclude with such a convergence result to a global maximizer. This is because only a lower bound (but not an upper bound) is available in terms of the reduced eigenvalue functions for the maximum value of $\lambda_J(\omega)$ over $\omega \in \Omega$. However, the sequence $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$ is still convergent as shown next.

THEOREM 3.2. *The sequence $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$ generated by Algorithms 1 and 2 for the maximization problem (MX) converges.*

Proof. Letting $\lambda^* := \max_{\omega \in \Omega} \lambda_J(\omega)$, the sequence by Algorithms 1 and 2 satisfies

$$(3.5) \quad \begin{aligned} \lambda_J^{(k-1)}\left(\omega^{(k)}\right) &\leq \lambda_J\left(\omega^{(k)}\right) = \lambda_J^{(k)}\left(\omega^{(k)}\right) \\ &\leq \max_{\omega \in \Omega} \lambda_J^{(k)}(\omega) = \lambda_J^{(k)}\left(\omega^{(k+1)}\right) \leq \lambda^*, \end{aligned}$$

where the first and the last inequality follow from the monotonicity, and the equality in the first line is a consequence of the interpolatory property. These inequalities lead to the conclusion that the sequence $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$ is monotone increasing and bounded above by λ^* , hence convergent. \square

We observe in practice that the sequence $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$ converges to a $\tilde{\lambda}$ such that $\lambda_J(\tilde{\omega}) = \tilde{\lambda}$ for some $\tilde{\omega}$ that is a local maximizer of $\lambda_J(\omega)$, that is not necessarily a global maximizer.

We illustrate these convergence results for the minimization as well as for the maximization of the largest eigenvalue $\lambda_1(\omega)$ of

$$(3.6) \quad A(\omega) = \frac{Ae^{i\omega} + A^*e^{-i\omega}}{2} = \cos(\omega) \left(\frac{A + A^*}{2}\right) + \sin(\omega) \left(\frac{iA - iA^*}{2}\right)$$

over $\omega \in [0, 2\pi]$ and for a particular matrix A . The maximum of the largest eigenvalue of $A(\omega)$ over $\omega \in [0, 2\pi]$ corresponds to the numerical radius of A [17, 13]; see section 5.2 for more on the numerical radius. Here we particularly choose A as the

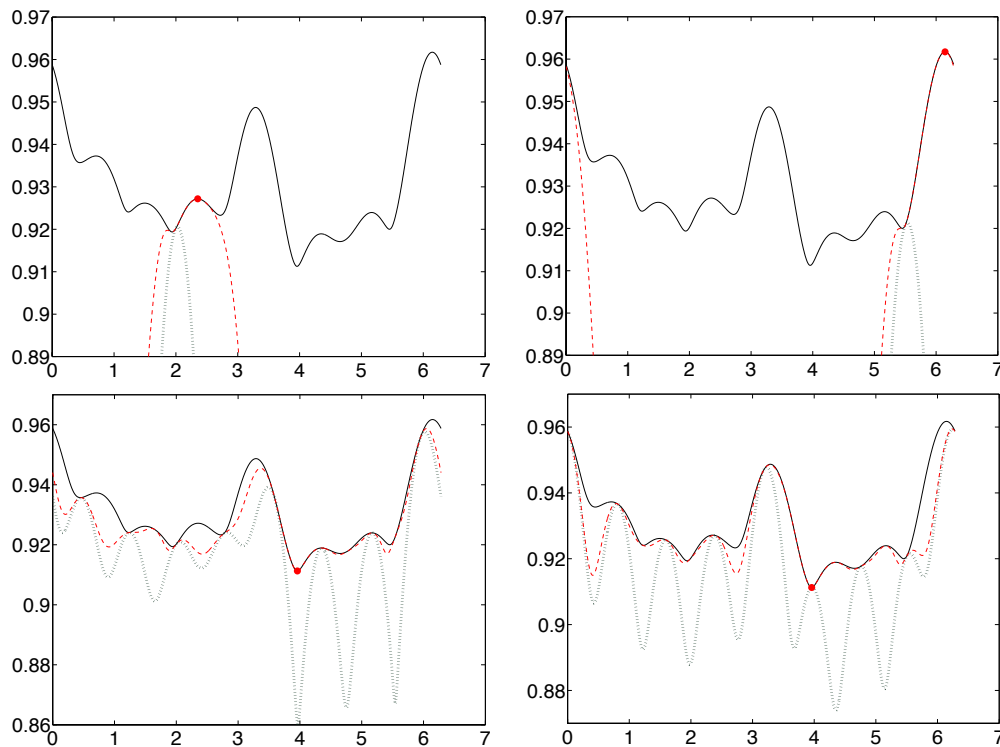


FIG. 1. Top and bottom rows illustrate Algorithm 1 to maximize and minimize, respectively, the largest eigenvalue $\lambda_1(\omega)$ (solid curves) of $A(\omega)$ as in (3.6) over $\omega \in [0, 2\pi]$ (the horizontal axis). For both the maximization and the minimization, Algorithm 1 is initiated with two different initial points $\omega^{(1)} = 2$ on the left column and $\omega^{(1)} = 5.5$ on the right column. The dotted curves on the top and at the bottom represent the reduced eigenvalue functions with one and eight dimensional subspaces, respectively. The dashed curves correspond to the reduced eigenvalue functions right before convergence, with five dimensional subspaces on the top and fifteen dimensional subspaces at the bottom. The dots represent the converged points, that is, denoting the converged ω value with $\tilde{\omega}$ and letting $\tilde{\lambda} := \lambda_1(\tilde{\omega})$ it marks $(\tilde{\omega}, \tilde{\lambda})$.

400×400 matrix whose real part comes from a five-point finite difference discretization of a Poisson equation, and whose complex part has random entries selected independently from a normal distribution with zero mean and standard deviation equal to 20. An application of the basic subspace procedure (Algorithm 1) for the maximization of $\lambda_1(\omega)$ starting with $\omega^{(1)} = 2$ results in convergence to a local maximizer $\tilde{\omega} \approx 2.353$, whereas initiating the procedure with $\omega^{(1)} = 5.5$ leads to convergence to $\omega^* \approx 6.145$, the unique global maximizer of $\lambda_1(\omega)$ over $[0, 2\pi]$. This is depicted on the top row in Figure 1 on the left and on the right, respectively. The situation is quite different when the subspace procedure is applied to minimize $\lambda_1(\omega)$. It converges to the global minimizer $\omega_* \approx 3.959$ of $\lambda_1(\omega)$ regardless of whether the procedure is initiated with $\omega^{(1)} = 2$ or $\omega^{(1)} = 5.5$ as depicted on the bottom row of Figure 1. Notice that the subspace procedure for the maximization problem constructs reduced eigenvalue functions that capture $\lambda_1(\omega)$ well only locally around the maximizers. In contrast, for the minimization problem the reduced eigenvalue functions capture $\lambda_1(\omega)$ globally, but their accuracy is higher around the minimizers.

3.2. The rate of convergence. Next we are concerned with how quickly the iterates of the subspace procedures converge when they do converge to a smooth stationary point of $\lambda_J(\omega)$. Note that by Theorem 3.1, for the minimization problem, if the sequence $\{\omega^{(k)}\}$ by Algorithm 2 (or Algorithm 1) converges to a point ω_* where $\lambda_J(\omega)$ is simple, then ω_* must be a smooth stationary point of $\lambda_J(\omega)$. The analysis below applies to Algorithm 2 (and Algorithm 1 when $d = 1$) in a unified way, both for the minimization problem and for the maximization problem.

THEOREM 3.3 (superlinear convergence). *Suppose that the sequence $\{\omega^{(k)}\}$ by Algorithm 2 (or Algorithm 1 when $d = 1$) converges to a point ω_* in the interior of Ω such that*

- (i) $\lambda_J(\omega_*)$ is simple, (ii) $\nabla\lambda_J(\omega_*) = 0$, and (iii) $\nabla^2\lambda_J(\omega_*)$ is invertible.

Then there exists a constant $\alpha > 0$ such that

$$(3.7) \quad \frac{\|\omega^{(k+1)} - \omega_*\|_2}{\|\omega^{(k)} - \omega_*\|_2 \max\{\|\omega^{(k)} - \omega_*\|_2, \|\omega^{(k-1)} - \omega_*\|_2\}} \leq \alpha \quad \forall k.$$

Proof. The arguments in the first two paragraphs of the proof of Lemma 2.10 establish the existence of a ball $\mathcal{B}(\omega_*, \eta)$ containing $\mathcal{B}(\omega^{(k)}, h^{(k)})$ for all k large enough, say for $k \geq k'$, such that $\lambda_J(\omega)$ as well as $\lambda_J^{(k)}(\omega)$ for $k \geq k'$ are simple for all $\omega \in \mathcal{B}(\omega_*, \eta)$. The Hessians $\nabla^2\lambda_J(\omega)$ and $\nabla^2\lambda_J^{(k)}(\omega)$ for $k \geq k'$ are Lipschitz continuous inside this ball.

Additionally, following the arguments in the proof of part (ii) of Lemma 2.10, the invertibility of $\nabla^2\lambda_J(\omega_*)$ ensures that $\nabla^2\lambda_J(\omega^{(k)})$ and $\nabla^2\lambda_J^{(k)}(\omega^{(k)})$ are invertible (indeed $\|[\nabla^2\lambda_J(\omega^{(k)})]^{-1}\|_2$ and $\|[\nabla^2\lambda_J^{(k)}(\omega^{(k)})]^{-1}\|_2$ are bounded from above by some constant) for all k large enough, say for $k \geq k'' \geq k'$.

Now for $k \geq k''$ by Taylor's theorem with integral remainder,

$$0 = \nabla\lambda_J(\omega_*) = \nabla\lambda_J(\omega^{(k)}) + \int_0^1 \nabla^2\lambda_J(\omega^{(k)} + \alpha(\omega_* - \omega^{(k)})) (\omega_* - \omega^{(k)}) \, d\alpha.$$

Employing $\nabla\lambda_J(\omega^{(k)}) = \nabla\lambda_J^{(k)}(\omega^{(k)})$ (Lemma 2.6(iii)) in this equation and left multiplying both sides of the equation by the inverse of $\nabla^2\lambda_J(\omega^{(k)})$ yield

$$(3.8) \quad \begin{aligned} 0 = & \left[\nabla^2\lambda_J(\omega^{(k)}) \right]^{-1} \cdot \nabla\lambda_J^{(k)}(\omega^{(k)}) + (\omega_* - \omega^{(k)}) \\ & + \left[\nabla^2\lambda_J(\omega^{(k)}) \right]^{-1} \int_0^1 \left[\nabla^2\lambda_J(\omega^{(k)} + \alpha(\omega_* - \omega^{(k)})) \right. \\ & \quad \left. - \nabla^2\lambda_J(\omega^{(k)}) \right] (\omega_* - \omega^{(k)}) \, d\alpha. \end{aligned}$$

Taylor expansion of $\nabla\lambda_J^{(k)}(\omega)$ about $\omega^{(k)}$, noting $\nabla\lambda_J^{(k)}(\omega^{(k+1)}) = 0$, implies

$$\left[\nabla^2\lambda_J^{(k)}(\omega^{(k)}) \right]^{-1} \nabla\lambda_J^{(k)}(\omega^{(k)}) = -(\omega^{(k+1)} - \omega^{(k)}) + O\left(\left[h^{(k+1)}\right]^2\right).$$

Using this equality in (3.8) leads us to

$$\begin{aligned}
 (3.9) \quad 0 &= \omega_* - \omega^{(k+1)} + \left\{ \left[\nabla^2 \lambda_J(\omega^{(k)}) \right]^{-1} - \left[\nabla^2 \lambda_J^{(k)}(\omega^{(k)}) \right]^{-1} \right\} \nabla \lambda_J(\omega^{(k)}) \\
 &+ O\left(\left[h^{(k+1)} \right]^2 \right) + \left[\nabla^2 \lambda_J(\omega^{(k)}) \right]^{-1} \\
 &\times \int_0^1 \left[\nabla^2 \lambda_J(\omega^{(k)} + \alpha(\omega_* - \omega^{(k)})) - \nabla^2 \lambda_J(\omega^{(k)}) \right] (\omega_* - \omega^{(k)}) \, d\alpha,
 \end{aligned}$$

which, by taking the 2-norms and employing the triangle inequality, yields

$$\begin{aligned}
 (3.10) \quad &\left\| \omega^{(k+1)} - \omega_* \right\|_2 \\
 &\leq \left\| \left[\nabla^2 \lambda_J(\omega^{(k)}) \right]^{-1} - \left[\nabla^2 \lambda_J^{(k)}(\omega^{(k)}) \right]^{-1} \right\|_2 \left\| \nabla \lambda_J(\omega^{(k)}) \right\|_2 \\
 &+ O\left(\left[h^{(k+1)} \right]^2 \right) + \left\| \left[\nabla^2 \lambda_J(\omega^{(k)}) \right]^{-1} \right\|_2 \\
 &\times \int_0^1 \left\| \nabla^2 \lambda_J(\omega^{(k)} + \alpha(\omega_* - \omega^{(k)})) - \nabla^2 \lambda_J(\omega^{(k)}) \right\|_2 \left\| \omega^{(k)} - \omega_* \right\|_2 \, d\alpha.
 \end{aligned}$$

To conclude with the desired superlinear convergence result, we bound the terms on the right-hand side of the inequality in (3.10) from above in terms of $\|\omega^{(k+1)} - \omega_*\|_2$, $\|\omega^{(k)} - \omega_*\|_2$, and $\|\omega^{(k-1)} - \omega_*\|_2$. To this end, first note that the terms in the third line of (3.10) are $O(\|\omega^{(k)} - \omega_*\|_2^2)$; this is because of the boundedness of $\|[\nabla^2 \lambda_J(\omega^{(k)})]^{-1}\|_2$ and the Lipschitz continuity of $\nabla^2 \lambda_J(\omega)$ inside $\mathcal{B}(\omega_*, \eta)$. Second, $\|\nabla \lambda_J(\omega^{(k)})\|_2 = O(\|\omega^{(k)} - \omega_*\|)$, as can be seen from a Taylor expansion of $\nabla \lambda_J(\omega)$ about $\omega^{(k)}$ and by exploiting $\nabla \lambda_J(\omega_*) = 0$. Finally, due to part (ii) of Lemma 2.8, we have $\|[\nabla^2 \lambda_J(\omega^{(k)})]^{-1} - [\nabla^2 \lambda_J^{(k)}(\omega^{(k)})]^{-1}\|_2 = O(h^{(k)})$. Applying all these bounds to (3.10) gives rise to

$$\left\| \omega^{(k+1)} - \omega_* \right\|_2 \leq O\left(h^{(k)} \left\| \omega^{(k)} - \omega_* \right\|_2 \right) + O\left(\left[h^{(k+1)} \right]^2 \right) + O\left(\left\| \omega^{(k)} - \omega_* \right\|_2^2 \right).$$

The desired result (3.7) follows noting $h^{(k)} \leq 2 \max\{\|\omega^{(k)} - \omega_*\|_2, \|\omega^{(k-1)} - \omega_*\|_2\}$ and $[h^{(k+1)}]^2 \leq 2(\|\omega^{(k)} - \omega_*\|_2^2 + \|\omega^{(k+1)} - \omega_*\|_2^2)$. \square

Regarding, specifically, the minimization of the largest eigenvalue when $d = 1$, the order of the superlinear rate of convergence for Algorithm 1 is shown to be at least $1 + \sqrt{2}$ in [20, Theorem 1].

It does not seem straightforward to extend the rate of convergence result above to Algorithm 1 when $d \geq 2$, because relations such as the ones given by Lemma 2.8 between the second derivatives of $\lambda_J(\omega)$ and $\lambda_J^{(k)}(\omega)$ are not evident. We observe a superlinear rate of convergence for Algorithm 1 in practice as well. Algorithm 2 requires a slightly fewer iterations to reach a prescribed accuracy, but Algorithm 1 attains the prescribed accuracy with subspaces of smaller dimension. These observations are illustrated in Table 1 on the problem of minimizing the largest eigenvalue $\lambda_1(\omega)$ of

$$(3.11) \quad A(\omega) = A_0 + \omega_1 A_1 + \cdots + \omega_d A_d$$

TABLE 1

The minimization of the largest eigenvalues of two particular matrix-valued functions of the form (3.11) with two and three parameters using Algorithms 1 and 2. The minimal values $\lambda_1^{(k)}(\omega^{(k+1)})$ of the reduced eigenvalue functions, as well as the errors $\|\omega^{(k+1)} - \omega_*\|$ of the minimizers are listed with respect to k and $p := \dim \mathcal{S}_k$. (Top row) $d = 2$, A_2, A_1, A_0 are 400×400 such that A_2 is random symmetric, A_1 is random Hermitian, whereas A_0 is obtained by adding $2A_2 - A_1$ to another random symmetric matrix; (bottom row) $d = 3$, A_3, A_2, A_1, A_0 are 400×400 such that A_3, A_2, A_1 are random symmetric, while A_0 is obtained from a random symmetric matrix by adding $A_3 + 2A_2 - 3A_1$.

$d = 2$							
Algorithm 1				Algorithm 2			
k	p	$\lambda_1^{(k)}(\omega^{(k+1)})$	$\ \omega^{(k+1)} - \omega_*\ _2$	k	p	$\lambda_1^{(k)}(\omega^{(k+1)})$	$\ \omega^{(k+1)} - \omega_*\ _2$
6	6	27.8402784689	0.3245035160	3	12	7.0566934870	0.3112458616
7	7	27.9933586806	0.1134918678	4	16	24.9336629638	0.5250223403
8	8	28.4522934270	0.0247284535	5	20	27.6531948787	0.1830729388
9	9	28.5008552358	0.0007732659	6	24	28.4440816653	0.0142076774
10	10	28.5010522075	0.0000243843	7	28	28.5010327361	0.0001556439
11	11	28.5010523924	0.0000000694	8	32	28.5010523924	0.0000000344
$d = 3$							
Algorithm 1				Algorithm 2			
k	p	$\lambda_1^{(k)}(\omega^{(k+1)})$	$\ \omega^{(k+1)} - \omega_*\ _2$	k	p	$\lambda_1^{(k)}(\omega^{(k+1)})$	$\ \omega^{(k+1)} - \omega_*\ _2$
10	10	28.1244577720	0.1137180915	3	21	22.9802867532	0.8996337159
11	11	28.1897386962	0.0482696055	4	28	26.1225009081	0.6602340340
12	12	28.2359716412	0.0051064698	5	35	27.0151485822	0.1737449902
13	13	28.2388852363	0.0005705897	6	42	28.0850898434	0.0374522435
14	14	28.2389043164	0.0000124871	7	49	28.2387293186	0.0002142688
15	15	28.2389043663	0.0000001102	8	56	28.2389043663	0.0000001112

for given Hermitian matrices $A_0, A_1, \dots, A_d \in \mathbb{C}^{n \times n}$ for $d = 2, 3$, where we restrict $\omega_1, \dots, \omega_d$ to the interval $[-60, 60]$. Such eigenvalue optimization problems are convex [28], and arise from a classical structural design problem [6, 24]. Additionally, as discussed in section 5.4, they are closely related to SDPs. In Table 1 the minimal values $\lambda_1^{(k)}(\omega^{(k+1)})$ of the reduced eigenvalue functions $\lambda_1^{(k)}(\omega)$, as well as the error $\|\omega^{(k+1)} - \omega_*\|_2$, converge superlinearly with respect to k both for $d = 2$ and for $d = 3$. In both cases the extended subspace procedure on the right column achieves 10 decimal digits accuracy (in the sense that $\lambda_1(\omega_*) - \lambda_1^{(k)}(\omega^{(k+1)}) \leq 10^{-10}$) after 8 iterations but with subspaces of dimension 32 and 56 for $d = 2$ and $d = 3$, respectively. On the other hand, the basic subspace procedure on the left column requires 11 and 15 iterations, which are also the dimensions of the subspaces constructed, for $d = 2$ and $d = 3$, respectively, to achieve the same accuracy. Observe also that the minimal values $\lambda_1^{(k)}(\omega^{(k+1)})$ seem to converge at a rate even faster than the rate of decay of the errors $\|\omega^{(k+1)} - \omega_*\|_2$.

In the one parameter case (i.e., $d = 1$) Theorem 3.3 applies to Algorithm 1 as well to establish its superlinear convergence. The convergence of Algorithm 1 on the example of Figure 1 (concerning the minimization or maximization of the largest eigenvalue of the matrix-valued function in (3.6)) starting with $\omega^{(1)} = 5.5$ is depicted in Table 2. For the maximization and minimization the iterates $\{\omega^{(k)}\}$ converge to the global maximizer (≈ 6.145) and global minimizer (≈ 3.959) of $\lambda_1(\omega)$ at a superlinear rate, which is realized earlier for the maximization problem. The optimal values of the reduced eigenvalue function $\lambda_1^{(k)}(\omega^{(k+1)})$ converge to the globally maximal value and minimal value of $\lambda_1(\omega)$ at an even faster rate.

TABLE 2

Iterates of Algorithm 1 on the example depicted on the right-hand column of Figure 1 involving the maximization and minimization of the largest eigenvalue of a matrix-valued function of the form (3.6) starting with $\omega^{(1)} = 5.5$. The optimal values $\lambda_1^{(k)}(\omega^{(k+1)})$ of the reduced eigenvalue functions, the errors $|\omega^{(k+1)} - \omega_*|$ of the optimizers, and the optimizers $\omega^{(k+1)}$ are listed with respect to k .

Maximization			
k	$\lambda_1^{(k)}(\omega^{(k+1)})$	$ \omega^{(k+1)} - \omega_* $	$\omega^{(k+1)}$
1	0.9213417656	0.6098999764	5.5354507094
2	0.9272678333	0.4809562556	5.6643944303
3	0.9452307484	0.2569647709	5.8883859150
4	0.9596256053	0.0659648186	6.0793858672
5	0.9617115893	0.0019013105	6.1434493753
6	0.9617265293	0.0000002797	6.1453504061
Minimization			
k	$\lambda_1^{(k)}(\omega^{(k+1)})$	$ \omega^{(k+1)} - \omega_* $	$\omega^{(k+1)}$
10	0.8905833092	1.2006093051	5.1592214939
11	0.8953986993	2.7480373651	1.2105748237
12	0.9039755162	0.2734960246	3.6851161642
13	0.9112584133	0.0063764350	3.9522357538
14	0.9112669429	0.0000006079	3.9586115809
15	0.9112669442	0.0000000000	3.9586121888

4. Variations and extensions.

4.1. A greedy subspace procedure without past. The rate-of-convergence analysis of the previous section and, in particular, the proof of Theorem 3.3, for Algorithm 2 (for Algorithm 1 when $d = 1$) makes use of the eigenvectors added into the subspace in the last iteration (in the last two iterations) only. Hence Theorem 3.3 and its superlinear convergence assertion still hold for Algorithm 2 even if its line 11 is changed as

$$\mathcal{S}_k \leftarrow \text{span} \left\{ s_1^{(k)}, \dots, s_J^{(k)} \right\} \oplus \left\{ \bigoplus_{p=1, q=p}^d \text{span} \left\{ s_{1,pq}^{(k)}, \dots, s_{J,pq}^{(k)} \right\} \right\},$$

that is even if the previous subspace is completely discarded. We refer to this variant of Algorithm 2 as *Algorithm 2 without past*. Similarly for Algorithm 1, when $d = 1$, the superlinear convergence assertion of Theorem 3.3 still holds if only the eigenvectors from the last two iterations are kept inside the subspace, that is, if line 7 of Algorithm 1 is replaced by

$$\mathcal{S}_k \leftarrow \text{span} \left\{ s_1^{(k-1)}, \dots, s_J^{(k-1)} \right\} \oplus \text{span} \left\{ s_1^{(k)}, \dots, s_J^{(k)} \right\}.$$

We refer to this variant of Algorithm 1 for the case $d = 1$ as *Algorithm 1 without past*.

The main issue with these subspace procedures without past is the convergence. In section 3.1 the convergence of the sequence $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$ is established both for the minimization problem and for the maximization problem (by Theorems 3.1 and 3.2, respectively). When the eigenvectors from the past iterations are discarded, the convergence of $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$ is not guaranteed anymore for the minimization problem, because the monotonicity of $\lambda_J^{(k)}(\omega)$ with respect to k is no longer true. On the other hand, all the equalities and inequalities in (3.5) concerning the monotonicity and boundedness of $\{\lambda_J^{(k)}(\omega^{(k+1)})\}$ for the maximization problem can be verified to hold, so this sequence is still guaranteed to converge.

TABLE 3

The optimal values $\lambda_1^{(k)}(\omega^{(k+1)})$ of the reduced eigenvalue functions by the variant of Algorithm 1 that keeps only the subspaces from the last two iterations are listed with respect to k for the example of the right column of Figure 1 (which involves the maximization and minimization of the largest eigenvalue of a matrix-valued function of the form (3.6) starting with $\omega^{(1)} = 5.5$).

Maximization		Minimization	
k	$\lambda_1^{(k)}(\omega^{(k+1)})$	k	$\lambda_1^{(k)}(\omega^{(k+1)})$
1	0.9213417656	45	-0.3266668608
2	0.9322945953	46	-0.5147292664
3	0.9595070069	47	-0.3348522694
4	0.9615567360	48	-0.4859510415
5	0.9617263748	49	-0.3259949226
6	0.9617265294	50	-0.5140279537

The remarks of the previous paragraph are illustrated in Table 3 which concerns the application of Algorithm 1 without past to the example of Table 2. For the maximization problem, the maximizers $\lambda_1^{(k)}(\omega^{(k+1)})$ of the reduced eigenvalue functions $\lambda_1^{(k)}(\omega)$ converge to the globally largest value of $\lambda_1(\omega)$ at a superlinear rate with respect to k , even though for every k the subspace dimension is two. For the minimization problem, the values in the table depict that the sequence $\{\lambda_1^{(k)}(\omega^{(k+1)})\}$ does not converge.

The subspace procedures without past work effectively in practice for the maximization problem. On the other hand, for the minimization problem, it may fail to converge. In [20] a convergent subspace framework making use of three dimensional subspaces is devised for the Crawford number computation, which involves the maximization of the smallest eigenvalue (equivalently, the minimization of the largest eigenvalue) of a matrix-valued function depending on one variable. The three dimensional subspaces are formed of eigenvectors from the past iterations, but not necessarily the eigenvectors from the last three iterations. The approach is built on the concavity of the smallest eigenvalue function involved and the knowledge of an interval containing the global maximizer. It does not seem easy to extend the ideas in [20] to the general nonconvex setting that involves the minimization of $\lambda_J(\omega)$.

4.2. Subspace procedures for singular value optimization. The ideas of the previous sections can be extended to maximize or minimize the J th largest singular value $\sigma_J(\omega)$ of a compact operator

$$(4.1) \quad \mathbf{B}(\omega) := \sum_{\ell=1}^{\kappa} f_{\ell}(\omega) \mathbf{B}_{\ell}$$

over a compact subset Ω of \mathbf{R}^d . As before $\mathbf{B}(\omega)$ is defined over ω that belongs to an open subset $\bar{\Omega}$ containing the feasible region Ω . In representation (4.1) of $\mathbf{B}(\omega)$ above, $f_{\ell} : \bar{\Omega} \rightarrow \mathbb{R}$ is a real-analytic function and $\mathbf{B}_{\ell} : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$ is a compact operator for $\ell = 1, \dots, \kappa$. Once again, the infinite dimensional problem is motivated by the finite dimensional case where the matrix-valued function

$$(4.2) \quad B(\omega) := \sum_{\ell=1}^{\kappa} f_{\ell}(\omega) B_{\ell}$$

for given $B_{\ell} \in \mathbb{C}^{p \times q}$ $\ell = 1, \dots, \kappa$ with large dimensions, takes the role of $\mathbf{B}(\omega)$.

Let \mathcal{U}, \mathcal{V} be two subspaces of $\ell^2(\mathbb{N})$ with equal dimension, and let U, V be orthonormal bases for these subspaces. The subspace procedures to optimize $\sigma_J(\omega)$ are based on the optimization of the J th largest singular value $\sigma_J^{\mathcal{U}, \mathcal{V}}(\omega)$ of an operator of the form

$$(4.3) \quad \mathbf{B}^{\mathcal{U}, \mathcal{V}}(\omega) := \mathbf{U}^* \mathbf{B}(\omega) \mathbf{V},$$

where \mathbf{V} (\mathbf{U}) represents the operator as in (2.1) or (2.4) that maps the coordinates of a vector in \mathcal{V} (\mathcal{U}) relative to the basis V (U) to itself. This operator can be written as

$$\mathbf{B}^{\mathcal{U}, \mathcal{V}}(\omega) = \sum_{\ell=1}^{\kappa} f_{\ell}(\omega) \mathbf{B}_{\ell}^{\mathcal{U}, \mathcal{V}}, \quad \text{where } \mathbf{B}_{\ell}^{\mathcal{U}, \mathcal{V}} := \mathbf{U}^* \mathbf{B}_{\ell} \mathbf{V},$$

which helps to reduce computational costs.

The subspace projections and restrictions here for singular value optimization are closely related to the ones employed for eigenvalue optimization. Indeed $\sigma_J(\omega)$ and $\sigma_J^{\mathcal{U}, \mathcal{V}}(\omega)$ correspond to the J th largest eigenvalues of

$$\begin{aligned} \begin{bmatrix} 0 & \mathbf{B}(\omega) \\ \mathbf{B}^*(\omega) & 0 \end{bmatrix} & \quad \text{and} \quad \begin{bmatrix} 0 & \mathbf{U}^* \mathbf{B}(\omega) \mathbf{V} \\ \mathbf{V}^* \mathbf{B}^*(\omega) \mathbf{U} & 0 \end{bmatrix} \\ & = \begin{bmatrix} \mathbf{U}^* & \mathbf{V}^* \end{bmatrix} \begin{bmatrix} 0 & \mathbf{B}(\omega) \\ \mathbf{B}^*(\omega) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}, \end{aligned}$$

respectively. Hence the results in section 2, specifically Lemmas 2.1–2.4 and Theorem 2.5, extend to relate the singular values $\sigma_J(\omega)$ and $\sigma_J^{\mathcal{U}, \mathcal{V}}(\omega)$. For instance, monotonicity amounts to the following: for four subspaces $\mathcal{U}_1, \mathcal{U}_2, \mathcal{V}_1, \mathcal{V}_2$ of $\ell^2(\mathbb{N})$ such that $\mathcal{U}_1 \subseteq \mathcal{U}_2$ and $\mathcal{V}_1 \subseteq \mathcal{V}_2$, we have

$$\sigma_J^{\mathcal{U}_1, \mathcal{V}_1}(\omega) \leq \sigma_J^{\mathcal{U}_2, \mathcal{V}_2}(\omega) \leq \sigma_J(\omega).$$

The extended greedy subspace procedure for singular value optimization forms \mathcal{U} and \mathcal{V} from the left singular vectors and right singular vectors of $\mathbf{B}(\omega)$ at the optimizers of the reduced problems and at nearby points. A precise description is given in Algorithm 3 below, where \mathcal{U}_k and \mathcal{V}_k denote the left and the right subspace at step k and $\sigma_J^{(k)}(\omega) := \sigma_J^{\mathcal{U}_k, \mathcal{V}_k}(\omega)$. The basic greedy procedure is defined similarly by modifying lines 11 and 12 as

$$(4.4) \quad \mathcal{U}_k \leftarrow \mathcal{U}_{k-1} \oplus \text{span} \{u_1^{(k)}, \dots, u_J^{(k)}\} \quad \text{and} \quad \mathcal{V}_k \leftarrow \mathcal{V}_{k-1} \oplus \text{span} \{v_1^{(k)}, \dots, v_J^{(k)}\}.$$

In the description, by a consistent pair of left and right singular vectors u and v corresponding to a singular value σ of $\mathbf{B}(\omega)$, we mean the vectors satisfying $\mathbf{B}(\omega)v = \sigma u$ and $\mathbf{B}^*(\omega)u = \sigma v$ simultaneously.

Observe that the reduced singular value function $\sigma_J^{(k)}(\omega)$ is the same as $\lambda_J^{(k)}(\omega)$ formed by Algorithm 2 when it is applied to

$$(4.5) \quad \mathbf{A}(\omega) := \begin{bmatrix} 0 & \mathbf{B}(\omega) \\ \mathbf{B}^*(\omega) & 0 \end{bmatrix},$$

so Algorithm 3 applied to $\mathbf{B}(\omega)$ and Algorithm 2 applied to $\mathbf{A}(\omega)$ lead the same sequence $\{\omega^{(k)}\}$. Similarly, the basic greedy subspace procedure for singular value

Algorithm 3 The extended greedy subspace procedure for singular value optimization.

Require: A parameter dependent compact operator $\mathbf{B}(\omega)$ of the form (4.2), and a compact subset $\Omega \subset \bar{\Omega}$ of \mathbb{R}^d .

- 1: $\omega^{(1)} \leftarrow$ a random point in Ω .
 - 2: $u_1^{(1)}, \dots, u_J^{(1)}$ and $v_1^{(1)}, \dots, v_J^{(1)} \leftarrow$ consistent left and right singular vectors corresponding to $\sigma_1(\omega^{(1)}), \dots, \sigma_J(\omega^{(1)})$.
 - 3: $\mathcal{U}_1 \leftarrow \text{span} \{u_1^{(1)}, \dots, u_J^{(1)}\}$ and $\mathcal{V}_1 \leftarrow \text{span} \{v_1^{(1)}, \dots, v_J^{(1)}\}$.
 - 4: **for** $k = 2, 3, \dots$ **do**
 - 5: $\omega^{(k)} \leftarrow$ any $\omega_* \in \arg \min_{\omega \in \Omega} \sigma_J^{(k-1)}(\omega)$ for the minimization problem, or
 $\omega^{(k)} \leftarrow$ any $\omega_* \in \arg \max_{\omega \in \Omega} \sigma_J^{(k-1)}(\omega)$ for the maximization problem.
 - 6: $u_1^{(k)}, \dots, u_J^{(k)}$ and $v_1^{(k)}, \dots, v_J^{(k)} \leftarrow$ consistent left and right singular vectors corresponding to $\sigma_1(\omega^{(k)}), \dots, \sigma_J(\omega^{(k)})$.
 - 7: $h^{(k)} \leftarrow \|\omega^{(k)} - \omega^{(k-1)}\|_2$
 - 8: **for** $p = 1, \dots, d,$ $q = p, \dots, d$ **do**
 - 9: $u_{1,pq}^{(k)}, \dots, u_{J,pq}^{(k)}$ and $v_{1,pq}^{(k)}, \dots, v_{J,pq}^{(k)} \leftarrow$ consistent left and right singular vectors corresponding to $\sigma_1(\omega^{(k)} + h^{(k)}e_{pq}), \dots, \sigma_J(\omega^{(k)} + h^{(k)}e_{pq})$.
 - 10: **end for**
 - 11: $\mathcal{U}_k \leftarrow \mathcal{U}_{k-1} \oplus \text{span} \{u_1^{(k)}, \dots, u_J^{(k)}\} \oplus \left\{ \bigoplus_{p=1, q=p}^d \text{span} \{u_{1,pq}^{(k)}, \dots, u_{J,pq}^{(k)}\} \right\}$.
 - 12: $\mathcal{V}_k \leftarrow \mathcal{V}_{k-1} \oplus \text{span} \{v_1^{(k)}, \dots, v_J^{(k)}\} \oplus \left\{ \bigoplus_{p=1, q=p}^d \text{span} \{v_{1,pq}^{(k)}, \dots, v_{J,pq}^{(k)}\} \right\}$.
 - 13: **end for**
-

optimization is equivalent to Algorithm 1 operating on $\mathbf{A}(\omega)$. All of the convergence results deduced in section 3 for eigenvalue optimization, in particular,

1. global convergence for the minimization problem (Theorem 3.1),
2. convergence of the sequence of maximal values of the reduced problems for the maximization problem (Theorem 3.2),
3. superlinear rate of convergence for smooth optimizers (Theorem 3.3),

carry over to this singular value optimization setting. Some of these results in section 3 are proven assuming the simplicity of $\lambda_J(\omega_*)$ at a particular ω_* , which translates into a simplicity and a positivity assumption on $\sigma_J(\omega_*)$ in the singular value setting. Regarding issue 2 above, we observe in practice the convergence of $\{\sigma_J^{(k)}(\omega^{(k+1)})\}$ for the maximization problem to σ_* such that $\sigma_J(\tilde{\omega}) = \sigma_*$ for some local maximizer $\tilde{\omega}$, that is not necessarily a global maximizer, similar to what we observe in the eigenvalue optimization setting.

4.3. Optimization of the J th smallest singular value. The minimum (maximum) of the J th smallest eigenvalue $\lambda_{-J}(\omega)$ of $\mathbf{A}(\omega)$ is equal to the negative of the maximum (minimum) of the J th largest eigenvalue of $-\mathbf{A}(\omega)$. Hence Algorithms 1 and 2 can be adapted to optimize $\lambda_{-J}(\omega)$.

The optimization of the J th smallest singular value has a different nature. In particular it cannot be converted into an optimization problem involving the J th largest singular value. This is partly seen by the observation that the J th smallest singular value $\sigma_{-J}(\omega)$ of an operator $\mathbf{B}(\omega)$ of the form (4.1) corresponds to an eigenvalue of $\mathbf{A}(\omega)$ as in (4.5), right in the middle of its spectrum. For the restricted operator $\mathbf{B}^{\mathcal{U}, \mathcal{V}}(\omega)$ defined as in (4.3) and its J th smallest singular value $\sigma_{-J}^{\mathcal{U}, \mathcal{V}}(\omega)$, monotonicity is lost as the subspaces \mathcal{U}, \mathcal{V} expand. In particular $\sigma_{-J}^{\mathcal{U}, \mathcal{V}}(\omega) \geq \sigma_{-J}(\omega)$ does not neces-

sarily hold; the restriction of the domain of $\mathbf{B}(\omega)$ to \mathcal{V} causes an increase in the J th smallest singular value, whereas the projection of the range onto \mathcal{U} causes a decrease in the singular value. As a consequence of the loss of monotonicity, the interpolation properties do not hold anymore either.

A neater and theoretically more sound approach is to employ only the restrictions from the right-hand side, that is a subspace procedure that operates on $\mathbf{B}^{\mathcal{V}}(\omega) := \mathbf{B}(\omega)\mathbf{V}$ and its J th smallest singular value. The resulting subspace procedures are equivalent to those for eigenvalue optimization applied to $\mathbf{B}^*(\omega)\mathbf{B}(\omega)$, so monotonicity and interpolation properties as well as all the theoretical convergence properties are regained.

4.4. A comparison with the cutting plane methods. The cutting-plane method was introduced by Kelley to solve convex minimization problems [19]. For unconstrained minimization problems, the approach underestimates convex functions with piecewise linear functions globally, solves the resulting linear program, and refines the underestimator with the addition of the new linear approximation around the optimizer of the linear program [3, section 9.3.2].

Let us consider again the minimization of the largest eigenvalue $\lambda_1(\omega)$ of an affine and Hermitian matrix-valued function

$$A(\omega) = A_0 + \omega_1 A_j + \cdots + \omega_d A_d$$

for given Hermitian matrices $A_0, \dots, A_d \in \mathbb{C}^{n \times n}$. Recall that the largest eigenvalue $\lambda_1(\omega)$ is convex [28], so the cutting-plane method is applicable for its minimization as outlined in Algorithm 4. In this description $\widehat{\lambda}_1^{(k)}(\omega)$ represents the piecewise linear underestimator for $\lambda_1(\omega)$ and is defined by

$$\begin{aligned} \widehat{\lambda}_1^{(k)}(\omega) &:= \max_{\ell=1, \dots, k} \lambda_1(\widehat{\omega}^{(\ell)}) + \nabla \lambda_1(\widehat{\omega}^{(\ell)})^T (\omega - \widehat{\omega}^{(\ell)}) \\ &= \max_{\ell=1, \dots, k} (\widehat{v}^{(\ell)})^* A(\omega) \widehat{v}^{(\ell)}, \end{aligned}$$

where $\widehat{v}^{(\ell)}$ denotes a unit eigenvector corresponding to the largest eigenvalue of $A(\widehat{\omega}^{(\ell)})$. Furthermore, in line 4 of the outline, $\widehat{\omega}^{(k)}$ denotes the unique minimizer of the convex function $\widehat{\lambda}_1^{(k-1)}(\omega)$.

Algorithm 4 The cutting-plane method to minimize the largest eigenvalue.

Require: A matrix-valued function $A : \mathbb{R}^d \rightarrow \mathbb{C}^{n \times n}$ that is affine and Hermitian

- 1: $\widehat{\omega}^{(1)} \leftarrow$ a random vector in \mathbb{R}^d .
 - 2: $\widehat{v}^{(1)} \leftarrow$ a unit eigenvector corresponding to largest eigenvalue of $A(\widehat{\omega}^{(1)})$.
 - 3: **for** $k = 2, 3, \dots$ **do**
 - 4: $\widehat{\omega}^{(k)} \leftarrow \arg \min_{\omega \in \Omega} \widehat{\lambda}_1^{(k-1)}(\omega)$.
 - 5: $\widehat{v}^{(k)} \leftarrow$ a unit eigenvector corresponding to largest eigenvalue of $A(\widehat{\omega}^{(k)})$.
 - 6: **end for**
-

The basic greedy subspace procedure Algorithm 1 for the minimization of the largest eigenvalue $\lambda_1(\omega)$ in this finite dimensional matrix-valued setting minimizes $\lambda_1^{(k)}(\omega)$, which, by the monotonicity property, underestimates $\lambda_1(\omega)$. The sequences $\{\omega^{(k)}\}$ and $\{\widehat{\omega}^{(k)}\}$ generated by Algorithm 1 and Algorithm 4 are not the same in general, but, for simplicity, let us suppose $\omega^{(\ell)} = \widehat{\omega}^{(\ell)}$ for $\ell = 1, \dots, k$ and for some k .

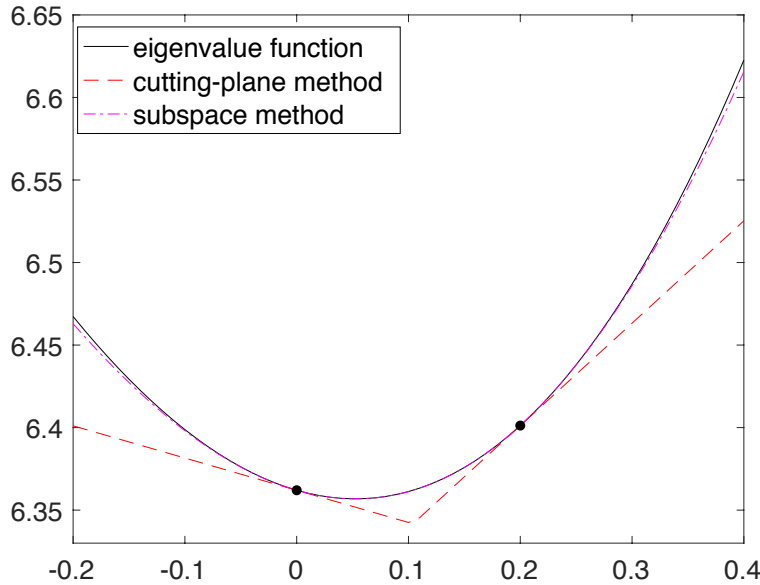


FIG. 2. A comparison of the approximations by the cutting-plane method and the subspace procedure: (solid curve) the largest eigenvalue of $A(\omega) = A_0 + \omega A_1$, for randomly selected 10×10 symmetric A_0, A_1 with respect to $\omega \in [-0.2, 0.4]$; (dashed curve) the function $\widehat{\lambda}_1^{(2)}(\omega)$ used by the cutting-plane method with $\widehat{\omega}^{(1)} = 0$ and $\widehat{\omega}^{(2)} = 0.2$; (dashed-dotted curve) the function $\lambda_1^{(2)}(\omega)$ used by the subspace procedure with $\omega^{(1)} = 0$ and $\omega^{(2)} = 0.2$; additionally, the black dots represent $(0, \lambda_1(0))$ and $(0.2, \lambda_1(0.2))$.

In this case, for each $\omega \in \Omega$, we have

$$\begin{aligned} \lambda_1(\omega) &\geq \lambda_1^{(k)}(\omega) = \max_{\alpha \in \mathbb{C}^m, \|\alpha\|_2=1} \alpha^* V_k^* A(\omega) V_k \alpha \\ &\geq \max_{\ell=1, \dots, k} \left(\widehat{v}^{(\ell)} \right)^* A(\omega) \widehat{v}^{(\ell)} = \widehat{\lambda}_1^{(k)}(\omega), \end{aligned}$$

where the columns of the matrix V_k form an orthonormal basis for the subspace $\text{span}\{\widehat{v}^{(1)}, \dots, \widehat{v}^{(k)}\}$. This illustrates that, for this special case concerning the minimization of the largest eigenvalue of an affine matrix-valued function, the underestimators used by the basic greedy subspace procedure are more accurate than those used by the cutting-plane method.

The better accuracy of the subspace procedure is apparent in Figure 2 for the matrix-valued function $A(\omega) = A_0 + \omega A_1$, where A_0, A_1 are 10×10 random symmetric matrices. In this figure, the function $\widehat{\lambda}_1^{(2)}(\omega)$ used by the cutting-plane method is the maximum of two linear approximations about $\widehat{\omega}^{(1)} = 0$ and $\widehat{\omega}^{(2)} = 0.2$, while a two dimensional subspace is used by the subspace procedure with $\omega^{(1)} = 0$ and $\omega^{(2)} = 0.2$.

5. Numerical experiments. The next section specifies the MATLAB software accompanying this work, and some important implementation details of this software. The rest of the section is devoted to three applications of large-scale eigenvalue and singular value optimization, namely, the numerical radius, the distance to instability from a matrix or a pencil, and the minimization of the largest eigenvalue of an affine matrix-valued function, which is closely related to semi-definite programming. On these examples we illustrate the power of the subspace procedures, introduced and analyzed in the previous sections, in practice.

5.1. Software and implementation details. MATLAB implementations accompanying this work are made publicly available on the internet.² The current version includes generic routines for the following purposes (for a prescribed positive integer J):

- minimization of the J th largest eigenvalue (based on Algorithm 1);
- maximization of the J th largest eigenvalue (based on Algorithm 1 without past if $d = 1$, or Algorithm 2 without past if $d > 1$);
- minimization of the J th smallest singular value (based on an adaptation of Algorithm 1 without past if $d = 1$, or Algorithm 2 without past if $d > 1$, for singular value optimization).

The MATLAB routines above terminate if one of the conditions

$$(1) \left| \lambda_J^{(k)}(\omega^{(k+1)}) - \lambda_J^{(k-1)}(\omega^{(k)}) \right| \leq \text{tol} \quad \left(\text{or} \quad \left| \sigma_J^{(k)}(\omega^{(k+1)}) - \sigma_J^{(k-1)}(\omega^{(k)}) \right| \leq \text{tol} \right),$$

$$(2) \# \text{ subspace iterations} > \sqrt{n},$$

hold for a prescribed tolerance `tol`, where n is the size of the matrices A_ℓ in (1.2) for eigenvalue optimization, or the maximum of the dimensions p and q of the matrices B_ℓ in (4.2) for singular value optimization. For all of the experiments in this section `tol` = 10^{-12} is used unless otherwise specified. The second condition is never used in practice for the examples in this section, because the other condition is fulfilled after a few subspace iterations.

The reduced eigenvalue optimization and singular value optimization problems are solved by means of the the MATLAB package `eigopt` [26, section 10]. These routines keep a lower bound and an upper bound for the optimal value of the reduced problem, and terminate when they differ by less than a prescribed tolerance. We set this tolerance equal to 0.1 tol . Additionally, `eigopt` requires a global lower (upper) bound γ for the minimization (maximization) problem on the second derivatives of the eigenvalue functions or the singular value functions; particular choices for the three applications in this section are specified below. Finally, `eigopt` performs the optimization on a box, which must be supplied by the user. Once again, particular box choices for the applications in this section are specified below.

Large-scale eigenvalue and singular value problems are solved iteratively by means of `eigs` and `svds` in MATLAB. If the optimal values of the reduced eigenvalue or singular value functions at two consecutive iterations are close enough (i.e., if they differ by an amount less than 10^{-2}), then we provide the shift $\lambda_{\text{cur}} + 5|\lambda_{\text{cur}} - \lambda_{\text{pre}}|$ (or $\sigma_{\text{cur}} + 5|\sigma_{\text{cur}} - \sigma_{\text{pre}}|$, in the case of singular values)—with $\lambda_{\text{cur}}, \lambda_{\text{pre}}$ ($\sigma_{\text{cur}}, \sigma_{\text{pre}}$) denoting the optimal values of the current, previous reduced J th largest eigenvalue functions (J th smallest singular value functions)—to `eigs` (`svds`) and require it to compute the J eigenvalues (singular values) closest to this shift. Otherwise, `eigs` and `svds` are called without shifts.

5.2. Numerical radius. The numerical radius $r(A)$ of a matrix $A \in \mathbb{C}^{n \times n}$, as also indicated at the end of section 3.1, is the modulus of the outermost point in the field of values [17, 13] formally defined by

$$r(A) := \{ |z^* A z| \mid z \in \mathbb{C}^n, \|z\|_2 = 1 \}.$$

²<http://home.ku.edu.tr/~emengi/software/leigopt>.

TABLE 4

The computed value of the numerical radius by Algorithm 1 without past (i.e., the subspaces S_k are spanned by the eigenvectors at the last two iterations, hence, two dimensional), the number of subspace iterations, and the runtimes in seconds are listed for the Grcar matrix of size n for various n . Specifically, the number of subspace iterations (2nd column) to reach the prescribed accuracy, the numerical radius (3rd column), the total runtime (4th column), the time spent for large-scale eigenvalue computations (5th column), and the time spent for the solution of the reduced eigenvalue optimization problems (6th column) are reported w.r.t. the sizes of the matrices.

n	# iter	$r(A)$	Total time	Eigval comp	Reduced prob
320	11	3.240793870067	8.3	0.8	7.4
640	12	3.241243679341	9.1	1.4	7.7
1280	13	3.241357030535	9.7	2.5	7.2
2560	15	3.241385481170	10.9	4.6	6.3
5120	16	3.241392607964	14.4	8.8	5.4
10240	18	3.241394391431	30.8	26.6	3.9
20480	19	3.241394837519	87.4	82.8	4.0

This quantity is used, for instance, to analyze the convergence of the classical iterative schemes for linear systems [1, 9]. Recall from section 3.1 that it has an eigenvalue optimization characterization, specifically

$$r(A) = \max_{\omega \in [0, 2\pi]} \lambda_1(\omega), \quad \text{where } A(\omega) = \frac{Ae^{i\omega} + A^*e^{-i\omega}}{2}.$$

We apply Algorithm 1 without past as discussed in section 4.1 for the computation of the numerical radius of two families of matrices. The box and γ supplied to `eigopt` are $[0, 2\pi]$ and $2\|A\|_2$, respectively. The latter is not guaranteed to be an upper bound on the second derivatives of the eigenvalue function, but it works well in practice in our experience.

Example 1. The Grcar matrix is a Toeplitz matrix with 1s on the main, first, second, third superdiagonals, -1 s on the first subdiagonal, which exhibits ill-conditioned eigenvalues. It is used as a test matrix in the previous works [13, 34] concerning the estimation of the numerical radius. Table 4 lists the computed values of the numerical radius by the subspace procedure and runtimes in seconds for the Grcar matrices of sizes varying between 320 – 20480. As the matrix sizes increase, the solution of large-scale eigenvalue problems take nearly all of the computation time. In contrast to this, the time spent to solve reduced eigenvalue optimization problems is very small and it is more or less constant as the sizes of the matrices increase. The reported values of the numerical radius for the Grcar matrices of sizes 320 and 640 match with the results reported in [34] up to 12 decimal digits.

Example 2. This example concerns the gear matrix which is another Toeplitz matrix with 1s on the superdiagonal, subdiagonal, 1 at position $(1, n)$, and -1 at $(n, 1)$. This matrix also has ill-conditioned, real eigenvalues lying in the interval $(-2, 2)$. The computed values of the numerical radius by the subspace procedure for the gear matrix of size n with n varying in 320 – 20480, as well as the runtimes, are given in Table 5. Once again, the computation time for increasing n is dominated by large-scale eigenvalue computations. The results reported for the gear matrices of sizes 320 and 640 match again with those reported in [34] up to prescribed accuracy.

5.3. Distance to instability. The distance to instability from a square matrix $A \in \mathbb{C}^{n \times n}$ with all eigenvalues on the open left half of the complex plane, defined by

$$\mathcal{D}(A) := \min\{\|\Delta\|_2 \mid \exists \omega \in \mathbb{R}, \det(A + \Delta - \omega I) = 0\},$$

TABLE 5

This table lists the computed value of the numerical radius by Algorithm 1 without past (i.e., the subspaces \mathcal{S}_k are spanned by the eigenvectors at the last two iterations, hence two dimensional) for the gear matrix of size n for various n , as well as the number of subspace iterations, and runtimes in seconds. For the meaning of each column see the caption of Table 4.

n	# iter	$r(A)$	Total time	Eigval comp	Reduced prob
320	5	1.999904217490	5.2	0.4	4.8
640	5	1.999975979457	5.5	0.8	4.6
1280	6	1.999993985476	6.2	1.2	5.0
2560	5	1.999998495194	5.8	1.9	3.9
5120	5	1.999999623651	6.3	3.0	3.3
10240	5	1.999999905895	7.7	5.6	2.1
20480	5	1.999999976471	16.4	14.6	1.6

is suggested in [35] as a measure of robust stability for the autonomous system $x'(t) = Ax(t)$. An application of the Eckart–Young theorem [12, Theorem 2.5.3] yields the characterization

$$\mathcal{D}(A) = \min_{\omega \in \mathbb{R}} \sigma_{-1}(\omega),$$

where $\sigma_{-1}(\omega)$ denotes the smallest singular value of $B(\omega) = A - \omega I$.

Here also, we adopt Algorithm 1 without past (see section 4.1), noting that at step $k+1$ the two dimensional right subspace is the span of the right singular vectors corresponding to $\sigma_{-1}(\omega^{(k)})$ and $\sigma_{-1}(\omega^{(k-1)})$. We illustrate numerical results on two families of sparse matrices. We set γ , the global lower bound for the second derivatives of the singular value function for `eigopt`, equal to $-2\|A\|_2$, which is a heuristic that works well in practice. The boxes supplied to `eigopt` are $[-60, 60]$ and $[150, 160]$ for the first and second families, respectively. These boxes indeed contain the global minimizers.

Example 3. Tolosa matrices arise from the stability analysis of an airplane. They are used as test examples in previous works [14, 11] concerning the computation of the distance to instability. These are stable matrices with all eigenvalues lying in the left half of the complex plane, but they are nearly unstable (see [11, Figure 4] for the spectrum of the 340×340 Tolosa matrix), indeed their perturbations at a distance 0.002 have eigenvalues on the right half-plane. We run the subspace procedure on the Tolosa matrices of size 340, 1090, 2000, 4000, which are all available through the matrix market [2]. According to Table 6 only four iterations suffice to reach prescribed accuracy. The time required for large-scale singular value computations increases with respect to the size of the matrices. However the majority of the time is consumed for the solution of the reduced problems, but this is only because the matrices are relatively small. The computed value of $\mathcal{D}(A)$ is the same in each case and match with the results reported in [11] up to prescribed accuracy.

Example 4. This example is taken from [14]. A finite difference discretization of an Orr–Sommerfeld operator with step size h for planar Poiseuille flow leads to an $n \times n$ generalized eigenvalue problem $B_n v = \lambda L_n v$ or an $n \times n$ standard eigenvalue problem $L_n^{-1} B_n v = \lambda v$, where $n = 2/h - 1$,

$$L_n = \frac{1}{h^2} \text{tridiag}(1, -(2+h^2), 1),$$

$$B_n = \frac{1}{100} L_n^2 - i(U_n L_n + 2I),$$

TABLE 6

The runtimes in seconds, number of iterations (2nd column) and computed values of $\mathcal{D}(A)$ (3rd column) are listed for Algorithm 1 without past applied to compute the distance to instability from the Tolosa matrices of sizes 340, 1090, 2000, 4000. The subspaces \mathcal{S}_k are two dimensional and spanned by the right singular vectors computed at the last two iterations. The total runtimes, the time for large-scale singular value computations, and the time for the reduced optimization problems in seconds are provided in the 4th, 5th, and 6th columns, respectively.

n	# iter	$\mathcal{D}(A)$	Total time	Singval comp	Reduced prob
340	4	0.001999796888	9.1	1.0	8.0
1090	4	0.001999796888	9.7	1.4	8.2
2000	4	0.001999796888	9.9	1.8	8.1
4000	4	0.001999796888	10.3	2.7	7.5

TABLE 7

The number of iterations, computed values of the distance to instability, and runtimes are given for Algorithm 1 without past applied to estimate the distance to instability from the Orr–Sommerfeld matrices. Once again the subspaces are always two dimensional and spanned by the right singular vectors computed at the last two iterations. For the meaning of each column we refer to the caption of Table 6.

n	# iter	$\mathcal{D}(A)$	Total time	Singval comp	Reduced prob
400	9	0.001978	5.0	0.9	4.1
1000	8	0.001978	5.3	1.7	3.4
2000	7	0.001978	5.9	3.0	2.8
4000	8	0.001978	7.9	5.5	2.3
8000	8	0.001979	15.2	11.7	3.4
16000	7	0.001938	30.8	27.7	2.9

and $U_n = \text{diag}(1 - x_1^2, \dots, 1 - x_n^2)$ with $x_k = 1 + k \cdot h$ for $k = 1, \dots, n$. Matrix $A_n = L_n^{-1}B_n$ is stable with eigenvalues on the left half-plane, yet nearly unstable.

We apply the subspace procedure for the computation of the distance to instability for the Orr–Sommerfeld matrix A_n of sizes $n = 400, 1000, 2000, 4000, 8000, 16000$. Note that A_n is not sparse, yet applications of Arnoldi's method for large-scale smallest singular value computations on $A_n - \omega i I$ require the solutions of the linear systems of the form $(A_n - \omega i I)v_{k+1} = v_k$ for v_{k+1} for a given v_k . We equivalently solve the sparse linear system

$$(5.1) \quad (B_n - \omega i L_n)v_{k+1} = L_n v_k$$

in practice. In Table 7 we again observe that the total computation time is dominated by large-scale singular value computations as n increases, whereas the contribution of the time for the solution of the reduced problems to the total running time is very little for large n . The computed values of $\mathcal{D}(A)$ in the table are listed only to six decimal digits, because for large n `eigs` could not compute singular values beyond 7–8 decimal digits in a reliable fashion. We attribute this to the fact that the norm of B_n and L_n increase considerably as n increases, so it is not possible to solve the linear system (5.1) with high accuracy, for instance, $\|B_{16000}\|_2 \approx 6.467 \cdot 10^{13}$. For smaller n the computed solutions are accurate up to 12 decimal digits, for instance, the computed value of the distance to instability for A_{400} by the subspace procedure is 0.001978172281 which differs from the result reported in [11] by an amount less than 10^{-12} .

5.4. Minimization of the largest eigenvalue. A problem that drew substantial interest in the late 1980s and early 1990s [28, 10] concerns the minimization of

the largest eigenvalue of $\lambda_1(\omega)$ of

$$(5.2) \quad A(\omega) := A_0 + \omega_1 A_1 + \cdots + \omega_d A_d$$

for given symmetric matrices $A_0, \dots, A_d \in \mathbb{R}^{n \times n}$. This problem is already discussed in section 3.2 in the more general case, when A_0, \dots, A_d are complex and Hermitian. Numerical results over there on random matrices indicate that the sequences generated by both the basic subspace procedure and the extended one converge at least superlinearly. An important application is in the context of semidefinite programming: under mild assumptions the dual of an SDP can be expressed as an unconstrained minimization problem with the objective function $\lambda_1(\omega) + b^T \omega$ for some $b \in \mathbb{R}^d$ and with $\lambda_1(\omega)$ denoting the largest eigenvalue of a matrix-valued function $A(\omega)$ of the form (5.2).

We apply the subspace procedures Algorithms 1 and 2 for a particular notoriously difficult family of matrix-valued functions depending on two parameters. In this example the subspaces from the previous iterations are kept fully. Note that the MATLAB software is based on Algorithm 1; additionally, we apply Algorithm 2 for comparison purposes. As for the parameters for `eigopt`, since the largest eigenvalue function is convex, γ (the global lower bound on the second derivatives of the eigenvalue function) in theory can be chosen zero, instead we set $\gamma = -10^{-6}$ for numerical reliability. We have specified the box containing the minimizer as $[-10, 10] \times [-10, 10]$.

Example 5. In [28], for $A_0 \in \mathbb{R}^{n \times n}$ with its (k, j) entry equal to

$$\begin{cases} \min\{k, j\} & \text{if } |k - j| > 1, \\ \min\{k, j\} + 0.1 & \text{if } |k - j| = 1, \\ 0, & k = j, \end{cases}$$

the spectral radius (i.e., the absolute value of the eigenvalue furthest away from the origin) of $C_n(\omega) = A_0 - \sum_{j=1}^d \omega_j e_j e_j^T$ is minimized. It is observed in that paper on the $n = 10$ case that at the optimal ω , the eigenvalue with the largest modulus has multiplicity three. This problem would correspond to an SDP relaxation of a max-cut problem, if A_0 had been a Laplace matrix of a graph [15, section 7].

Here we minimize the spectral radius of

$$(5.3) \quad \tilde{C}_n(\omega) = \frac{1}{100n} A_0 - \omega_1 I_u - \omega_2 I_l$$

for $n = 250, 500, 1000, 2000$, where $I_u = \text{diag}(I_{n/2}, 0_{n/2})$ and $I_l = \text{diag}(0_{n/2}, I_{n/2})$. The scaling in front of A_0 is to make sure that the unique minimizer of the problem is inside $[-10, 10] \times [-10, 10]$. This problem can equivalently be posed as the minimization of the largest eigenvalue of $A(\omega) := \text{diag}(\tilde{C}_n(\omega), -\tilde{C}_n(\omega))$, so fits within the problem class described by (5.2). Table 8 lists the minimal spectral radius values computed by the basic and extended subspace procedures along with number of subspace iterations and computation time. The total computation times are again dominated by the solutions of large eigenvalue problems. Furthermore, even though the basic subspace procedure usually requires more iterations to reach the prescribed accuracy, overall it solves fewer large eigenvalue problems and takes less computation time as compared to the extended subspace procedure.

In all cases the computed minimizer ω_* is such that the largest eigenvalue of $A(\omega_*)$ has algebraic multiplicity three, so $\lambda_1(\omega)$ is not differentiable at ω_* . For instance, for

TABLE 8

The table lists the number of subspace iterations, the dimension of the subspace \mathcal{S}_k at termination (p), computational times in seconds, and the computed minimal value (ρ_*) of the spectral radius when Algorithm 1 (on the top) and Algorithm 2 (at the bottom) are applied to minimize the largest eigenvalue of $A(\omega) = \text{diag}(\tilde{C}_n(\omega), -\tilde{C}_n(\omega))$ for the matrix-valued function $\tilde{C}_n(\omega)$ defined as in (5.3). Note that $A(\omega)$, whose largest eigenvalue is minimized, is of size $2n$. Also note that the subspaces from all of the previous iterations are kept in these examples.

n	# iter	p	ρ_*	Total time	Eigval comp	Reduced prob
250	7	7	0.509646245274	3.8	1.5	2.2
500	7	7	1.016261471669	4.8	2.8	1.6
1000	8	8	3.584040976076	13.9	11.8	1.2
2000	7	7	4.055903987776	68.7	65.8	0.7
n	# iter	p	ρ_*	Total time	Eigval comp	Reduced prob
250	6	24	0.509646245274	4.7	2.8	1.6
500	6	24	1.016261471669	7.6	5.4	1.1
1000	7	28	3.584040976076	22.1	17.8	1.0
2000	7	28	4.055903987776	115.9	106.7	0.8

TABLE 9

(Left) The five largest eigenvalues of $A(\omega) = \text{diag}(\tilde{C}_{500}(\omega), -\tilde{C}_{500}(\omega))$ at ω_* , where $\tilde{C}_n(\omega)$ is defined as in (5.3) and ω_* is the minimizer of the largest eigenvalue of $A(\omega)$ for $n = 500$. (Right) The last four iterates $\{\lambda_1^{(k)}(\omega^{(k+1)})\}$ of Algorithm 1 (which keeps all of the subspaces from the previous iterations) when the large eigenvalue computations are performed directly by calling `eig` in MATLAB.

Eigenvalues of $A(\omega_*)$
1.016261471669
1.016261471669
1.016261471669
1.016234975093
1.016234803766

k	$\lambda_1^{(k)}(\omega^{(k+1)})$
3	1.016260414001
4	1.016261417096
5	1.016261471669
6	1.016261471669

$n = 500$, that is, when the matrix-valued function $A(\omega)$ is of size 1000, the five largest eigenvalues of $A(\omega_*)$ are listed in Table 9 on the left. Even the gaps between the remaining 497 positive eigenvalues are very small, as indeed among 500 positive eigenvalues 495 of them lie in an interval of length 0.017. The fact that most of the eigenvalues belong to a small interval causes poor convergence properties for `eigs`. On the other hand, it appears that the nonsmoothness does not affect the superlinear convergence of the iterates $\{\lambda_1^{(k)}(\omega^{(k+1)})\}$, as depicted in Table 9 on the right. This quick convergence is also apparent from the number of subspace iterations in Table 8. Theorem 3.3 does not apply to this nonsmooth case. It is an open problem to come up with a formal argument explaining the quick convergence in this nonsmooth setting.

6. Concluding remarks. We have proposed subspace procedures to cope with large-scale eigenvalue and singular value optimization problems. To optimize the J th largest eigenvalue of a Hermitian and analytic matrix-valued function $A(\omega)$ over ω for a prescribed integer J , the subspace procedures operate on a small matrix-valued function that acts like $A(\omega)$ in a small subspace. The subspace is expanded with the addition of the eigenvectors of $A(\omega)$ at the optimizer of the eigenvalue function of the small matrix-valued function and, possibly, at nearby points. A similar strategy is adopted to optimize the J th largest singular value of an analytic matrix-valued function $B(\omega)$. In that context, it is advantageous to use two different subspace restrictions on the input and the output to $B(\omega)$ so that the resulting small matrix-

TABLE 10

Problem	Preferred subspace procedure
(1) $\min_{\omega \in \Omega} \lambda_J(\omega)$	Algorithm 1
(2) $\max_{\omega \in \Omega} \lambda_J(\omega)$	Algorithm 2 (Algorithm 1) without past if $d > 1$ (if $d = 1$)
(3) $\min_{\omega \in \Omega} \sigma_J(\omega)$	Algorithm 3(b)
(4) $\max_{\omega \in \Omega} \sigma_J(\omega)$	Algorithm 3 (Algorithm 3(b)) without past if $d > 1$ (if $d = 1$)
(5) $\min_{\omega \in \Omega} \sigma_{-J}(\omega)$	Algorithm 2(s) (Algorithm 1(s)) without past if $d > 1$ (if $d = 1$)
(6) $\max_{\omega \in \Omega} \sigma_{-J}(\omega)$	Algorithm 1(s)

valued function acts like the original one only in these small input and output spaces. The subspaces are expanded with the inclusion of the left and right singular vectors of $B(\omega)$ at the optimizers of the small problems and at nearby points. The optimization of the J th smallest singular value involves some subtlety; here, it seems suitable to apply restrictions only on the input to $B(\omega)$, which is equivalent to the frameworks for eigenvalue optimization applied to $B(\omega)^*B(\omega)$. The preferred subspace procedures for particular cases are summarized in Table 10.

In the table, Algorithm 3(b) refers to the basic greedy procedure for singular value optimization that uses two-sided projections, i.e., Algorithm 3 but with lines 11 and 12 replaced by (4.4). Additionally, Algorithms 1(s) and 2(s) refer to the adaptations of Algorithms 1 and 2 for singular value optimization, which form the subspace from the right singular vectors rather than the eigenvectors. Note that the minimization and maximization of a J th smallest eigenvalue are not listed in the table, since they can be posed as the maximization and minimization of a J th largest eigenvalue, respectively.

We have performed convergence and rate-of-convergence analyses for these subspace procedures by extending $A(\omega)$ and $B(\omega)$ to infinite dimension, by replacing them with compact operators $\mathbf{A}(\omega) : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$ and $\mathbf{B}(\omega) : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$, the former of which is also self-adjoint. Most remarkably, Theorem 3.1 establishes global convergence for Algorithms 1–2 when the J th largest eigenvalue is minimized, and Theorem 3.3 establishes a superlinear rate of convergence for Algorithm 1 when $d = 1$ and Algorithm 2 for minimizing and maximizing the J th largest eigenvalue. The superlinear convergence result is established under the simplicity assumption on the J th largest eigenvalue at the optimizer, even though we observe superlinear convergence in numerical experiments (e.g., see Example 5 in section 5.4) where this simplicity assumption is violated. The convergence results also extend to Algorithm 3 for the optimization of the J th largest singular value. The convergence properties of the proposed subspace frameworks for the six problems in the table above are as follows.

(1)&(3)&(6): Proven global convergence at a proven (observed) superlinear rate if $d = 1$ ($d > 1$);

(2)&(4)&(5): Proven convergence but not necessarily to a global solution, observed local convergence at a proven superlinear rate.

Two additional problems where the subspace procedures and their convergence analyses developed here may be applicable are large-scale sparse estimation problems and SDPs. For instance, sparse estimation problems that can be cast as nuclear norm minimization problems [30] seem worth exploring because of their connection with singular value optimization. As noted in the text, the dual of a standard SDP can often be cast as an eigenvalue optimization problem involving the minimization of the largest eigenvalue. A systematic integration of the subspace frameworks proposed here for eigenvalue optimization into large-scale SDPs, motivated by their theoretical convergence properties, is a direction that is worth investigating.

Acknowledgments. We are grateful to the two anonymous referees, Daniel Szyld, and Daniel Kressner for valuable comments on this manuscript.

REFERENCES

- [1] O. AXELSSON, H. LU, AND B. POLMAN, *On the numerical radius of matrices and its application to iterative solution methods*, Linear Multilinear Algebra, 37 (1994), pp. 225–238.
- [2] B. BOISVERT, R. POZO, K. REMINGTON, B. MILLER, AND R. LIPMAN, *Matrix Market*, <https://math.nist.gov/MatrixMarket/>.
- [3] J. F. BONNANS, J. C. GILBERT, C. LEMARECHAL, AND C. A. SAGASTIZABAL, *Numerical Optimization*, Springer, Berlin, 1998.
- [4] 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.
- [5] 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.
- [6] S. J. COX AND M. L. OVERTON, *On the optimal design of columns against buckling*, SIAM J. Math. Anal., 23 (1992), pp. 287–325.
- [7] E. B. DAVIES, *Linear Operators and their Spectra*, Cambridge University Press, Cambridge, 2007.
- [8] N. DUNFORD AND J. T. SCHWARTZ, *Linear Operators. Part II: Spectral Theory. Self Adjoint Operators in Hilbert Spaces*, Wiley, New York, 1998.
- [9] M. EIERMANN, *Field of values and iterative methods*, Linear Algebra Appl., 180 (1993), pp. 167–197.
- [10] M. K. H. FAN AND B. NEKOOIE, *On minimizing the largest eigenvalue of a symmetric matrix*, Linear Algebra Appl., 214 (1995), pp. 225–246.
- [11] M. A. FREITAG AND A. SPENCE, *A Newton-based method for the calculation of the distance to instability*, Linear Algebra Appl., 435 (2011), pp. 3189–3205.
- [12] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1996.
- [13] C. HE AND G. A. WATSON, *An algorithm for computing the numerical radius*, IMA J. Numer. Anal., 17 (1997), pp. 329–342.
- [14] C. HE AND G. A. WATSON, *An algorithm for computing the distance to instability*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 101–116.
- [15] C. HELMBERG AND F. RENDL, *A spectral bundle method for semidefinite programming*, SIAM J. Optim., 10 (2000), pp. 673–696.
- [16] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
- [17] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
- [18] T. KATO, *Perturbation Theory for Linear Operators*, Springer, Berlin, 1995.
- [19] J. E. KELLEY, *The cutting-plane method for solving convex programs*, J. Soc. Ind. Appl. Math., 8 (1960), pp. 703–712, <https://doi.org/10.1137/0108053>.
- [20] D. KRESSNER, D. LU, AND B. VANDEREYCKEN, *Subspace Acceleration for the Crawford Number and Related Eigenvalue Optimization Problems*, Technical report, Université de Genève, Geneva, 2017.
- [21] 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.

- [22] C. S. KUBRUSLY, *Elements of Operator Theory*, Birkhäuser, Boston, 2001.
- [23] P. LANCASTER, *On eigenvalues of matrices dependent on a parameter*, Numer. Math., 6 (1964), pp. 377–387.
- [24] A. S. LEWIS AND M. L. OVERTON, *Eigenvalue optimization*, Acta Numer., 5 (1996), pp. 149–190, <https://doi.org/10.1017/S0962492900002646>.
- [25] H. LIN, *An inexact spectral bundle method for convex quadratic semidefinite programming*, Comput. Optim. Appl., 53 (2012), pp. 45–89, <https://doi.org/10.1007/s10589-011-9443-x>.
- [26] 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.
- [27] S. A. MILLER AND R. S. SMITH, *A bundle method for efficiently solving large structured linear matrix inequalities*, in Proceedings of the 2000 American Control Conference, Vol. 2, IEEE, Piscataway, NJ, 2000, pp. 1405–1409.
- [28] M. L. OVERTON, *On minimizing the maximum eigenvalue of a symmetric matrix*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 256–268, <https://doi.org/10.1137/0609021>.
- [29] B. N. PARLETT, *The Symmetric Eigenvalue Problem*, Class. Appl. Math. 20, SIAM, Philadelphia, 1998.
- [30] B. RECHT, M. FAZEL, AND P. A. PARRILO, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.
- [31] F. RELICH, *Perturbation Theory of Eigenvalue Problems*, Gordon and Breach, New York, 1969.
- [32] Y. SAAD, *Numerical Methods for Large Eigenvalue Problems*, Class. Appl. Math. 66, SIAM, Philadelphia, 2011.
- [33] G. W. STEWART AND J. SUN, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
- [34] F. UHLIG, *Geometric computation of the numerical radius of a matrix*, IMA J. Numer. Anal., 52 (2009), pp. 335–353.
- [35] C. F. VAN LOAN, *How near is a stable matrix to an unstable matrix?*, Contemp. Math., 47 (1985), pp. 465–477.
- [36] L. VANDENBERGHE AND S. BOYD, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95, <https://doi.org/10.1137/1038003>.