SIAM J. OPTIM.
Vol. 27, No. 1, pp. 246–268

A SUPPORT FUNCTION BASED ALGORITHM FOR OPTIMIZATION WITH EIGENVALUE CONSTRAINTS*

EMRE MENGI†

Abstract. Optimization of convex functions subject to eigenvalue constraints is intriguing because of peculiar analytical properties of eigenvalue functions and is of practical interest because of a wide range of applications in fields such as structural design and control theory. Here we focus on the optimization of a linear objective subject to a constraint on the smallest eigenvalue of an analytic and Hermitian matrix-valued function. We propose a numerical approach based on quadratic support functions that overestimate the smallest eigenvalue function globally. The quadratic support functions are derived by employing variational properties of the smallest eigenvalue function over a set of Hermitian matrices. We establish the local convergence of the algorithm under mild assumptions and deduce a precise rate of convergence result by viewing the algorithm as a fixed point iteration. The convergence analysis reveals that the algorithm is immune to the nonsmooth nature of the smallest eigenvalue. We illustrate the practical applicability of the algorithm on the pseudospectral functions.

Key words. nonsmooth optimization, analytical properties of eigenvalues, support functions, Karush–Kuhn–Tucker conditions, fixed point theory, pseudospectra

AMS subject classifications. 65F15, 90C26, 49J52

DOI. 10.1137/140966551

1. Introduction.

Given an analytic and Hermitian matrix-valued function $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times n}$, this work concerns optimization problems of the form

$$\text{maximize} \quad c^T\omega \qquad \text{subject to} \quad \lambda_{\min}(\mathcal{A}(\omega)) \le 0, \tag{1.1}$$

where $c \in \mathbb{R}^d$ is fixed and $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue, and where we assume that the feasible set $\{\omega \in \mathbb{R}^d \mid \lambda_{\min}(\mathcal{A}(\omega)) \le 0\}$ is bounded in order to ensure the well-posedness of the problem. We do not presume convexity or linearity on $\mathcal{A}(\omega)$. Thus, some or all entries of $\mathcal{A}(\omega)$ can be nonlinear functions of ω. Furthermore, the feasible set for (1.1), that is, the set of ω such that $\lambda_{\min}(\mathcal{A}(\omega)) \le 0$, can be nonconvex. Two quantities that can be cast as optimization problems of the form (1.1) are the ϵ-pseudospectral abscissa and radius [23] of a given square matrix A. These quantities correspond to the real part of the rightmost and the modulus of the outermost points in the set composed of the eigenvalues of all matrices within an ϵ-neighborhood of A with respect to the matrix 2-norm. They have received considerable attention because of their association with the transient behavior of the autonomous systems $x'(t) = Ax(t)$ and $x_k = Ax_{k-1}$.

Various other applications in engineering give rise to problems of the form (1.1) either by replacing the smallest eigenvalue with the largest eigenvalue or by reversing the direction of the inequality. In structural design, a classical problem is the

*Received by the editors April 25, 2014; accepted for publication (in revised form) November 10, 2016; published electronically February 22, 2017.
http://www.siam.org/journals/siopt/27-1/96655.html
Funding: The work of the author was supported in part by European Commission grant PIRG-GA-2010-268355, TUBITAK-FWO (Scientific and Technological Research Council of Turkey–Belgian Research Foundation, Flanders) joint grant 113T053, and the BAGEP program of Turkish Academy of Science.

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

minimization of the volume or weight subject to a lower bound on the smallest eigenvalue [1]. In robust control theory, it is desirable to design a system subject to the largest eigenvalue not exceeding a prescribed tolerance. Even though we focus on framework (1.1) and derive an algorithm for that setting, it appears that the algorithm that we come up with works well in practice for such variations. An important family that is related to (1.1), with the direction of the inequality reversed, is the family of nonconvex semidefinite programs. The algorithm may work well for the solution of certain nonconvex semidefinite programs locally. We elaborate on such variations later in this paper.

In a recent work [19], we pursued support function based ideas to optimize a prescribed eigenvalue of $\mathcal{A}(\omega)$ globally on a box $\mathcal{B} \subset \mathbb{R}^d$. Remarkably, such ideas yield a linearly convergent algorithm that overcomes nonconvexity and nonsmoothness intrinsic to eigenvalue functions. Here we explore the use of support functions for the numerical solution of (1.1). We convexify and smooth the problem (1.1) by approximating the eigenvalue function $\lambda_{\min}(\mathcal{A}(\cdot))$ with a quadratic support function $q_k(\cdot)$ about a given $\omega^{(k)} \in \mathbb{R}^d$ satisfying

$$q_k(\omega^{(k)}) = \lambda_{\min}(\mathcal{A}(\omega^{(k)})) \quad \text{and} \quad q_k(\omega) \geq \lambda_{\min}(\mathcal{A}(\omega))$$

for all $\omega \in \mathbb{R}^d$. Here $\omega^{(k)}$ is assumed to be feasible so that $\lambda_{\min}(\mathcal{A}(\omega^{(k)})) \leq 0$. Thus, the original nonconvex and nonsmooth problem (1.1) is replaced by

$$\text{maximize} \quad c^T\omega \qquad \text{subject to} \quad q_k(\omega) \leq 0, \tag{1.2}$$

which can be solved analytically, and has a maximizer ω_* that is suboptimal yet feasible with respect to the original problem. We build a new quadratic support function $q_{k+1}(\omega)$ about the maximizer $\omega^{(k+1)} = \omega_*$, replace the constraint in (1.2) with $q_{k+1}(\omega) \leq 0$, and solve the updated convex smooth optimization problem. The practicality of the algorithm rests on a global upper bound on the second derivatives of the smallest eigenvalue function, as the quadratic support functions are built on the existence of such upper bounds. In various cases it is feasible to deduce such bounds analytically.

Optimization based on support functions dates back to the cutting plane method introduced by Kelley [13] and Cheney and Goldstein [6]. These original cutting plane methods are developed to minimize a linear objective subject to convex constraints by employing support functions. With focused research on combinatorial and integer optimization problems, the cutting plane methods are also adopted to solve relaxations of these computationally hard problems [20]. The original cutting plane method is suitable for convex constraints only, and extensions in the presence of nonconvex constraints to locate a locally optimal solution do not appear straightforward. More recently, inspired by Kelley's cutting plane method, bundle methods became popular for nonsmooth optimization built around linear support functions defined in terms of subgradients [14, 17, 16]. Bundle methods are especially effective for the unconstrained optimization of a convex nonsmooth function. In the nonconvex setting, construction of global support functions in terms of subgradients becomes a challenge; see, for instance, [11] and references therein for the latest developments concerning nonconvex bundle methods.

Outline. In the next section we derive support functions for $\lambda_{\min}(\mathcal{A}(\omega))$ and specify the algorithm based on the solution of the convex and smooth problem (1.2). In section 3, we show that the sequence generated by the algorithm converges to a point where the first order optimality conditions for a local maximizer of (1.1) hold

under mild assumptions. Remarkably the argument in section 3 establishes that the convergence to a first order optimal point occurs also in the nonsmooth case, that is, regardless of the multiplicity of the smallest eigenvalue at the converged point. Section 4 is on a practical variant of the algorithm that converges to a local maximizer of (1.1). Section 5 is devoted to a rate of convergence analysis of the algorithm, by viewing the algorithm as a fixed point iteration. The practicality of the algorithm relies on the deduction of an upper bound on the second derivatives of the smallest eigenvalue function, either analytically or numerically. We present a result in section 6 that facilitates deducing such an upper bound analytically. The practical usefulness of the algorithm is illustrated on the pseudospectral functions in section 7. Finally, section 8 concerns the applications of the algorithm to the variations of framework (1.1) either with a largest eigenvalue constraint instead of the smallest eigenvalue constraint or with a lower bound constraint on the smallest eigenvalue.

2. Derivation of the algorithm. We begin with the derivation of the support functions, which depends on the analytical properties of $\lambda_{\min}(\mathcal{A}(\omega))$. We summarize the relevant classical results below [21, 15].

LEMMA 2.1. *Let $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times n}$ be Hermitian and analytic, and let $\Phi : \mathbb{R} \to \mathbb{C}^{n\times n}$ be defined by $\Phi(\alpha) := \mathcal{A}(\hat{\omega} + \alpha p)$ for given $\hat{\omega}, p \in \mathbb{R}^d$. Then the following hold:*

(i) *There exist functions $\phi_1, \ldots, \phi_n : \mathbb{R} \to \mathbb{R}$ that are real analytic on $\mathbb{R}$ such that $\{\phi_1(\alpha), \ldots, \phi_n(\alpha)\}$ correspond to the set of eigenvalues of $\Phi(\alpha)$ for all $\alpha \in \mathbb{R}$.*

(ii) *Suppose that $\phi(\alpha) := \lambda_{\min}(\Phi(\alpha))$ is simple for all α on an open interval $\mathcal{I}$ in $\mathbb{R}$. Then $\phi(\alpha)$ is analytic on $\mathcal{I}$.*

(iii) *The left-hand $\phi'_-(\alpha)$ and the right-hand $\phi'_+(\alpha)$ derivatives of $\phi(\alpha) := \lambda_{\min}(\Phi(\alpha))$ exist everywhere. Furthermore, $\phi'_-(\alpha) \geq \phi'_+(\alpha)$ at all $\alpha \in \mathbb{R}$.*

(iv) *The eigenvalue function $\lambda_{\min}(\mathcal{A}(\omega))$ is twice continuously differentiable at all $\omega \in \mathbb{R}^d$ where it is simple. Furthermore, at all such ω we have*

$$\frac{\partial \lambda_{\min}(\mathcal{A}(\omega))}{\partial \omega_j} = v_n(\omega)^* \frac{\partial \mathcal{A}(\omega)}{\partial \omega_j} v_n(\omega)$$

and

$$\frac{\partial^2 \lambda_{\min}(\mathcal{A}(\omega))}{\partial \omega_k \, \partial \omega_\ell} = v_n^*(\omega) \frac{\partial^2 \mathcal{A}(\omega)}{\partial \omega_k \, \partial \omega_l} v_n(\omega) + 2\cdot\Re\left[\sum_{m=1}^{n-1} \frac{1}{\lambda_{\min}(\mathcal{A}(\omega)) - \lambda_m(\omega)} \left(v_n(\omega)^* \frac{\partial \mathcal{A}(\omega)}{\partial \omega_k} v_m(\omega)\right)\left(v_m(\omega)^* \frac{\partial \mathcal{A}(\omega)}{\partial \omega_\ell} v_n(\omega)\right)\right],$$

where $\lambda_m(\omega)$ denotes the mth largest eigenvalue of $\mathcal{A}(\omega)$ and $v_m(\omega)$ denotes an associated eigenvector that is analytic along every line in $\mathbb{R}^d$ such that $\{v_1(\omega), \ldots, v_n(\omega)\}$ is orthonormal.

In the theorem below and elsewhere, $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$, the scalar γ is positive and denotes a global upper bound satisfying

$$\lambda_{\max}\left[\nabla^2 \lambda_{\min}(\mathcal{A}(\omega))\right] \leq \gamma \;\; \forall \omega \in \mathbb{R}^d \text{ such that } \lambda_{\min}(\mathcal{A}(\omega)) \text{ is simple}, \tag{2.1}$$

and $\lambda_{\max}[\cdot]$ represents the largest eigenvalue of its matrix argument. This result may appear to follow from a straightforward application of Taylor's theorem. But the nonsmooth nature of $\lambda_{\min}(\mathcal{A}(\omega))$ complicates matters.

THEOREM 2.2 (support functions). *Suppose $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times n}$ is Hermitian and analytic, $\gamma \in \mathbb{R}$ satisfies (2.1), and $\omega^{(k)} \in \mathbb{R}^d$ is such that $\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ is simple. Then*

$$\lambda_{\min}(\mathcal{A}(\omega)) \;\le\; q_k(\omega) := \lambda_k + \nabla\lambda_k^T(\omega-\omega^{(k)}) + \frac{\gamma}{2}\|\omega-\omega^{(k)}\|^2 \quad \forall \omega \in \mathbb{R}^d, \tag{2.2}$$

where $\lambda_k := \lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ and $\nabla\lambda_k := \nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$.

Proof. Let $p = (\omega-\omega^{(k)})/\|\omega-\omega^{(k)}\|$, and define $\phi(\alpha) := \lambda_{\min}\left(\mathcal{A}(\omega^{(k)}+\alpha p)\right)$. Denote the points on $(0, \|\omega-\omega^{(k)}\|)$ where $\lambda_{\min}\left(\mathcal{A}(\omega^{(k)}+\alpha p)\right)$ is not simple with $\alpha_1, \ldots, \alpha_m$. There are finitely many such points, because $\lambda_{\min}\left(\mathcal{A}(\omega^{(k)}+\alpha p)\right)$ is the minimum of n analytic functions from part (i) of Lemma 2.1. Indeed, two analytic functions are identical or can intersect each other at finitely many points on a finite interval.

Partition the open interval $(0, \|\omega-\omega^{(k)}\|)$ into open subintervals $\mathcal{I}_j := (\alpha_j, \alpha_{j+1})$ for $j = 0, \ldots, m$ where $\alpha_0 = 0, \alpha_{m+1} = \|\omega-\omega^{(k)}\|$. On each open subinterval $\mathcal{I}_j$ the function $\phi(\alpha)$ is analytic from part (ii) of Lemma 2.1. Indeed $\phi(\alpha) = \phi_{i_j}(\alpha)$ on the closure of $\mathcal{I}_j$ for some analytic $\phi_{i_j}(\alpha)$ stated in part (i) of Lemma 2.1, and moreover $\phi'_+(\alpha_j) = \phi'_{i_j}(\alpha_j)$. Thus applying Taylor's theorem, for each α on the closure of $\mathcal{I}_j$, we deduce

$$\begin{aligned}
\phi(\alpha) &= \phi(\alpha_j) + \phi'_+(\alpha_j)(\alpha-\alpha_j) + \frac{\phi''(\eta)}{2}(\alpha-\alpha_j)^2 \\
&= \phi(\alpha_j) + \phi'_+(\alpha_j)(\alpha-\alpha_j) + \frac{p^T\nabla^2\lambda_{\min}(\mathcal{A}(\omega^{(k)}+\eta p))p}{2}(\alpha-\alpha_j)^2 \\
&\le \phi(\alpha_j) + \phi'_+(\alpha_j)(\alpha-\alpha_j) + \frac{\gamma}{2}(\alpha-\alpha_j)^2
\end{aligned} \tag{2.3}$$

for some $\eta \in \mathcal{I}_j$, where the second equality is due to the twice differentiability of $\lambda_{\min}(\mathcal{A}(\omega))$ at $\omega^{(k)}+\eta p$ (part (iv) of Lemma 2.1), and the last inequality is due to $p^T\nabla^2\lambda_{\min}(\mathcal{A}(\omega^{(k)}+\eta p))p \le \lambda_{\max}\left[\nabla^2\lambda_{\min}(\mathcal{A}(\omega^{(k)}+\eta p))\right]$ and (2.1). Furthermore, for $\alpha \in \mathcal{I}_j$, an application of the mean value theorem yields

$$\frac{\phi'(\alpha)-\phi'_+(\alpha_j)}{\alpha-\alpha_j} = \phi''(\varepsilon) = p^T\nabla^2\lambda_{\min}(\mathcal{A}(\omega^{(k)}+\varepsilon p))p \le \gamma$$

for some $\varepsilon \in (\alpha_j, \alpha)$. Thus

$$\phi'(\alpha) \le \phi'_+(\alpha_j) + \gamma(\alpha-\alpha_j). \tag{2.4}$$

Next we claim

$$\phi(\alpha_{m+1}) \le \phi(\alpha_j) + \phi'_+(\alpha_j)(\alpha_{m+1}-\alpha_j) + \frac{\gamma}{2}(\alpha_{m+1}-\alpha_j)^2 \tag{2.5}$$

for $j = 0, \ldots, m$. This is certainly true for $j = m$ from (2.3) on the interval $\mathcal{I}_m$ and with $\alpha = \alpha_{m+1}$. Suppose that inequality (2.5) holds for $k \ge 1$. Exploiting $\phi'_+(\alpha_k) \le \phi'_-(\alpha_k)$ (see part (iii) of Lemma 2.1) and then applying inequalities (2.3) and (2.4) with $\alpha = \alpha_k$ on the interval $\mathcal{I}_{k-1}$ lead us to

$$\begin{aligned}
\phi(\alpha_{m+1}) &\le \phi(\alpha_k) + \phi'_-(\alpha_k)(\alpha_{m+1}-\alpha_k) + \frac{\gamma}{2}(\alpha_{m+1}-\alpha_k)^2 \\
&\le \phi(\alpha_{k-1}) + \phi'_+(\alpha_{k-1})(\alpha_k-\alpha_{k-1}) + \frac{\gamma}{2}(\alpha_k-\alpha_{k-1})^2 \\
&\quad + [\phi'_+(\alpha_{k-1}) + \gamma(\alpha_k-\alpha_{k-1})](\alpha_{m+1}-\alpha_k) + \frac{\gamma}{2}(\alpha_{m+1}-\alpha_k)^2 \\
&= \phi(\alpha_{k-1}) + \phi'_+(\alpha_{k-1})(\alpha_{m+1}-\alpha_{k-1}) + \frac{\gamma}{2}(\alpha_{m+1}-\alpha_{k-1})^2.
\end{aligned}$$

Thus, by induction, we conclude

$$\phi(\alpha_{m+1}) \leq \phi(\alpha_0) + \phi'_+(\alpha_0)(\alpha_{m+1} - \alpha_0) + \frac{\gamma}{2}(\alpha_{m+1} - \alpha_0)^2.$$

Recalling $\phi(\alpha_0) = \phi(0) = \lambda_k$, $\phi'_+(\alpha_0) = \phi'(0) = \nabla\lambda_k^T p$, and $(\alpha_{m+1} - \alpha_0) = \|\omega - \omega^{(k)}\|$, we obtain the desired inequality

$$\lambda_{\min}(\mathcal{A}(\omega)) \leq \lambda_k + \nabla\lambda_k^T(\omega - \omega^{(k)}) + \frac{\gamma}{2}\|\omega - \omega^{(k)}\|^2. \qquad \square$$

Note. The following more general form of Theorem 2.2 is also apparent by replacing $\mathbb{R}^d$ with a convex subset $\mathcal{C} \subseteq \mathbb{R}^d$: If the scalar γ is an upper bound as in (2.1) on $\mathcal{C}$ rather than on $\mathbb{R}^d$, for each $\omega^{(k)} \in \mathcal{C}$ where $\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ is simple, the inequality in (2.2) holds for all $\omega \in \mathcal{C}$, that is, $q_k(\omega)$ is a support function on $\mathcal{C}$.

Given a feasible point $\omega^{(0)} \in \mathbb{R}^d$ satisfying $\lambda_{\min}(\mathcal{A}(\omega^{(0)})) \leq 0$, the algorithm generates a sequence $\{\omega^{(k)}\}$ of feasible points in $\mathbb{R}^d$. The update of $\omega^{(k)}$ is based on the solution of the following convex and smooth optimization problem:

(2.6)
$$\text{maximize} \quad c^T\omega \qquad \text{subject to} \quad q_k(\omega) := \lambda_k + \nabla\lambda_k^T(\omega - \omega^{(k)}) + \frac{\gamma}{2}\|\omega - \omega^{(k)}\|^2 \leq 0.$$

The next point $\omega^{(k+1)}$ is defined to be the unique maximizer of the problem above. Notice that the feasible set of the convex problem (2.6) is contained inside the feasible set of the original problem (1.1), i.e.,

$$\mathcal{F}_k := \{\omega \in \mathbb{R}^d \mid q_k(\omega) \leq 0\} \qquad \subseteq \qquad \mathcal{F} := \{\omega \in \mathbb{R}^d \mid \lambda_{\min}(\mathcal{A}(\omega)) \leq 0\}.$$

Thus $\omega^{(k+1)} \in \mathcal{F}_k \subseteq \mathcal{F}$ remains feasible. It is assumed by the algorithm that $\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ is simple for each $\omega^{(k)}$ in the sequence, which introduces no difficulties in practice, since the set of points ω such that $\lambda_{\min}(\mathcal{A}(\omega))$ is not simple is a subset of $\mathbb{R}^d$ of measure zero. On the other hand, being close to a multiple eigenvalue does not cause harm.

There is a degenerate case for the convergence of the algorithm as suggested by the following observation.

THEOREM 2.3. *Suppose the maximizer ω_* of* (2.6) *is such that $\nabla q_k(\omega_*) = 0$. Then* (i) *$\omega_* = \omega^{(k)}$ and* (ii) *$\nabla\lambda_k = 0$.*

Proof. The maximizer ω_* of (2.6) must be attained on the boundary of $\mathcal{F}_k$, that is, $q_k(\omega_*) = 0$. Assuming

$$\nabla q_k(\omega_*) = \nabla\lambda_k + \gamma(\omega_* - \omega^{(k)}) = 0 \tag{2.7}$$

yields $q_k(\omega_*) = \lambda_k - \frac{\gamma}{2}\|\omega_* - \omega^{(k)}\|^2 = 0$. Furthermore, since $\omega^{(k)}$ is feasible, $\lambda_k = \lambda_{\min}(\mathcal{A}(\omega^{(k)})) \leq 0$. This would imply $\lambda_k = 0$ and $\omega_* = \omega^{(k)}$. Now the second assertion follows from (2.7). $\square$

The first assertion of the theorem above means that $\omega^{(s)} = \omega^{(k)}$ and $q_s(\omega) \equiv q_k(\omega)$ for each $s > k$. Thus, convergence to a point ω_* such that $\nabla\lambda_{\min}(\mathcal{A}(\omega_*)) = 0$ seems possible. We rule this out by assuming $\nabla\lambda_k \neq 0$ for each k.

Now we apply the Karush–Kuhn–Tucker conditions to the constrained problem (2.6). The maximizer ω_* must satisfy

$$c = \mu\nabla q_k(\omega_*) \quad \text{and} \quad q_k(\omega_*) = 0 \tag{2.8}$$

for some positive $\mu \in \mathbb{R}$. Solving the equations above for ω_* and the positive scalar μ, and setting $\omega^{(k+1)} = \omega_*$, leads us to the update rule

$$\omega^{(k+1)} = \omega^{(k)} + \frac{1}{\gamma}\left[\frac{1}{\mu_+} \cdot c - \nabla\lambda_k\right], \quad \text{where} \quad \mu_+ = \frac{\|c\|}{\sqrt{\|\nabla\lambda_k\|^2 - 2\gamma\lambda_k}}. \tag{2.9}$$

Given a feasible point $\omega^{(0)} \in \mathcal{F}$ to start with, the basic algorithm generates the sequence $\{\omega^{(k)}\}$ defined by the update rule (2.9).

3. Convergence. We establish that the sequence $\{\omega^{(k)}\}$ converges to a point ω_* that satisfies the first order necessary condition

$$\exists \mu > 0 \quad \text{s.t.} \quad c = \mu\nabla\lambda_{\min}(\mathcal{A}(\omega_*)) \qquad \text{and} \qquad \lambda_{\min}(\mathcal{A}(\omega_*)) = 0 \tag{3.1}$$

for optimality for the optimization problem (1.1) under mild assumptions and when $\lambda_{\min}(\mathcal{A}(\omega_*))$ is simple. We also prove that the first order necessary condition holds in terms of generalized gradients [7, 8], in the nonsmooth case, when $\lambda_{\min}(\mathcal{A}(\omega_*))$ is not simple. Violation of the first order condition (3.1) is equivalent to the existence of a feasible ascent direction as stated below. This follows from an application of Farkas' lemma [4, p. 263] to our setting.

LEMMA 3.1. *Suppose that $\lambda_{\min}(\mathcal{A}(\omega_*))$ is simple and $\nabla\lambda_{\min}(\mathcal{A}(\omega_*)) \neq 0$. Furthermore, suppose $\omega_* \in \mathcal{F}$ is a point where the optimality condition* (3.1) *does not hold. Then there exists $p \in \mathbb{R}^d$ such that*

$$c^T p > 0 \quad \textit{and} \quad \nabla\lambda_{\min}(\mathcal{A}(\omega_*))^T p < 0.$$

The next theorem relates the local maximizers of problems (1.1) and (2.6).

THEOREM 3.2. *The following hold for each $\omega^{(k)} \in \mathbb{R}^d$ such that $\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ is simple and $\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)})) \neq 0$:*

(i) *If the point $\omega^{(k)}$ is a local maximizer of* (1.1)*, then it is a local maximizer of* (2.6).

(ii) *On the other hand, if $\omega^{(k)}$ is such that $c \neq \mu\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ for all $\mu > 0$ or $\lambda_{\min}(\mathcal{A}(\omega^{(k)})) \neq 0$, that is, the first order optimality condition for* (1.1) *is violated at $\omega^{(k)}$, then $\omega^{(k)}$ is not a local maximizer of* (2.6).

Proof. For assertion (i), if $\omega^{(k)} \in \mathcal{F}$ is a local maximizer of (1.1), then there exists a $\delta > 0$ such that

$$c^T\omega^{(k)} \geq c^T\omega \qquad \forall \omega \in \mathcal{B}(\omega^{(k)}, \delta) \cap \mathcal{F},$$

where $\mathcal{B}(\omega^{(k)}, \delta) := \{\omega \in \mathbb{R}^d \mid \|\omega - \omega^{(k)}\| \leq \delta\}$. But notice that $\omega^{(k)} \in \mathcal{F}_k$ (i.e., $q_k(\omega^{(k)}) = 0$), and due to the property $\mathcal{F}_k \subseteq \mathcal{F}$ we have $c^T\omega^{(k)} \geq c^T\omega$ for all $\omega \in \mathcal{B}(\omega^{(k)}, \delta) \cap \mathcal{F}_k$, meaning $\omega^{(k)}$ is a local maximizer of (2.6). This proves (i).

For assertion (ii), if $\omega^{(k)} \in \mathcal{F}$ does not satisfy the first order optimality condition for problem (1.1), then there exists a direction $p \in \mathbb{R}^d$ such that $c^T p > 0$ and $\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))^T p < 0$ due to Lemma 3.1. But then, for all small $\alpha > 0$, we have

$$q_k(\omega^{(k)} + \alpha p) = \lambda_k + \nabla\lambda_k^T(\alpha p) + O(\alpha^2) < \lambda_k \leq 0 \quad \text{and} \quad c^T(\omega^{(k)} + \alpha p) > c^T\omega^{(k)},$$

where $\lambda_k := \lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ and $\nabla\lambda_k := \nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$. Thus $\omega^{(k)}$ is not a local maximizer of (2.6). ∎

An implication of part (i) of the theorem above is that if $\omega^{(k)} \in \mathcal{F}$ is a local maximizer of (1.1), then $\omega^{(s)} = \omega^{(k)}$ for each $s > k$. This type of finite convergence to a local maximizer is unlikely. For the main convergence results, we first observe that the values of the objective function must converge.

LEMMA 3.3. *The sequence* $\{c^T\omega^{(k)}\}$ *is monotone increasing and convergent.*

Proof. Recall that

$$\omega^{(k+1)} = \arg \max_{q_k(\omega)\leq 0} c^T\omega.$$

Since $\omega^{(k)}$ is feasible with respect to the problem above, we must have $c^T\omega^{(k+1)} \geq c^T\omega^{(k)}$. Thus the sequence $\{c^T\omega^{(k)}\}$ is monotone increasing. Moreover, denoting a global maximizer for the original problem (1.1) with ω_*, due to $\mathcal{F}_k \subseteq \mathcal{F}$, we have $c^T\omega_* \geq c^T\omega^{(k+1)}$. Thus the sequence $\{c^T\omega^{(k)}\}$ is also bounded above, meaning the sequence must converge. ∎

Since each $\omega^{(k)}$ belongs to the bounded set $\mathcal{F}$, the sequence $\{\omega^{(k)}\}$ must have convergent subsequences by the Bolzano–Weierstrass theorem. First we establish the convergence of each of these subsequences to a point satisfying the optimality condition (3.1).

THEOREM 3.4 (convergence of subsequences). *Suppose the sequence* $\{\omega^{(k)}\}$ *is such that* $\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ *is simple, and* $\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)})) \neq 0$ *for each* $k \in \mathbb{N}$. *Consider a convergent subsequence of* $\{\omega^{(k)}\}$ *with limit* ω_* *such that* $\lambda_{\min}(\mathcal{A}(\omega_*))$ *is simple. The point* ω_* *satisfies the first order necessary condition* (3.1).

Proof. Let us denote the convergent subsequence of $\{\omega^{(k)}\}$ with $\{\omega^{(k_j)}\}$. Furthermore, let $\lambda_{k_j} := \lambda_{\min}(\mathcal{A}(\omega^{(k_j)}))$, $\lambda_* := \lambda_{\min}(\mathcal{A}(\omega_*))$, and $\nabla\lambda_{k_j} := \nabla\lambda_{\min}(\mathcal{A}(\omega^{(k_j)}))$, $\nabla\lambda_* := \nabla\lambda_{\min}(\mathcal{A}(\omega_*))$. Note that $\lim_{j\to\infty}\omega^{(k_j)} = \omega_*$ must be feasible, since all $\omega^{(k)}$ are feasible and $\lambda_{\min}(\mathcal{A}(\omega))$ varies continuously with respect to ω.

For the sake of contradiction, suppose ω_* violates the optimality condition (3.1). We infer from Lemma 3.1 the existence of a direction $p \in \mathbb{R}^d$ such that $c^Tp = \eta > 0$ and $\nabla\lambda_*^Tp = -\beta < 0$. Without loss of generality, we can assume $\|p\| = 1$. There exists a ball $\mathcal{B}(\omega_*, \delta)$ such that $\lambda_{\min}(\mathcal{A}(\omega))$ is simple for all $\omega \in \mathcal{B}(\omega_*, \delta)$, due to the continuity of the eigenvalues of $\mathcal{A}(\omega)$. Furthermore, there exists an integer j' such that each $\omega^{(k_j)}$ for $j \geq j'$ lies in $\mathcal{B}(\omega_*, \delta)$. Also, part (iv) of Lemma 2.1 (indicating the continuity of the partial derivatives of $\lambda_{\min}(\mathcal{A}(\omega))$ on $\mathcal{B}(\omega_*, \delta)$) implies the existence of an integer $j'' \geq j'$ such that

$$\|\nabla\lambda_* - \nabla\lambda_{k_j}\| \leq \beta/2 \qquad \forall j \geq j''. \tag{3.2}$$

We benefit from the convergence of $\{c^T\omega^{(k)}\}$ (Lemma 3.3); specifically, below we show that the existence of the feasible ascent direction p conflicts with the monotonicity of this sequence. In this respect we note that $\lim_{k\to\infty} c^T\omega^{(k)} = c^T\omega_*$. For some k' we must have

$$0 \leq (c^T\omega_* - c^T\omega^{(k)}) \leq (\eta\beta)/(2\gamma) \qquad \forall k \geq k'. \tag{3.3}$$

Consider any $\omega^{(k_j)}$ such that $j \geq j''$ and $k_j \geq k'$. Recalling $\lambda_{k_j} \leq 0$, since $\omega^{(k_j)}$ is feasible, the corresponding support function satisfies

$$q_{k_j}(\omega^{(k_j)} + \alpha p) = \lambda_{k_j} + \nabla\lambda_{k_j}^T(\alpha p) + \frac{\gamma}{2}\alpha^2 \leq 0$$

for all $\alpha \in [0, (2/\gamma)(-\nabla\lambda_{k_j}^Tp)]$. Furthermore,

$$\|\nabla\lambda_* - \nabla\lambda_{k_j}\| \geq (\nabla\lambda_{k_j} - \nabla\lambda_*)^Tp \quad \Longrightarrow \quad -\nabla\lambda_{k_j}^Tp \geq -\nabla\lambda_*^Tp - \|\nabla\lambda_* - \nabla\lambda_{k_j}\| \geq \beta/2,$$

where the last inequality is due to (3.2). Thus for $\tilde{\alpha} = (2/\gamma)(-\nabla\lambda_{k_j}^T p)$, and employing (3.3), we deduce

$$\begin{aligned} c^T(\omega^{(k_j)} + \tilde{\alpha}p) &= c^T\omega^{(k_j)} + (\tilde{\alpha})c^T p \\ &\geq c^T\omega_* + (\eta\beta)/(2\gamma) > c^T\omega_*. \end{aligned}$$

Now $\omega^{(k_j)} + \tilde{\alpha}p$ is a feasible point of the optimization problem $\max\left\{c^T\omega \mid q_{k_j}(\omega) \leq 0\right\}$, yielding $c^T(\omega^{(k_j+1)}) \geq c^T(\omega^{(k_j)} + \tilde{\alpha}p) > c^T\omega_*$. This last inequality contradicts (3.3). Thus, ω_* must satisfy the optimality condition (3.1). ∎

The convergence of the sequence $\{\omega^{(k)}\}$ itself is established below, with the additional mild assumption that $\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\|$ is uniformly bounded away from zero.

LEMMA 3.5. *Suppose* $\{\omega^{(k)}\}$ *is such that* $\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ *is simple, and* $\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)})) \neq 0$ *for each* $k \in \mathbb{N}$. *Then*

$$\lambda_{\min}(\mathcal{A}(\omega^{(k)})) \to 0 \quad \text{as } k \to \infty. \tag{3.4}$$

Proof. We adopt the notation used in the proof of Theorem 3.4, in particular, $\lambda_k := \lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ and $\nabla\lambda_k := \nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$. Furthermore, $M := \sup\{\|\nabla\lambda_k\| \mid k \in \mathbb{N}\}$. Below we show that the violation of (3.4) also contradicts the monotonicity of $\{c^T\omega^{(k)}\}$.

Let us suppose the contrary of (3.4), that is, there exists an $\epsilon > 0$ such that

$$\forall N\ \exists k > N, \quad \lambda_k \leq -\epsilon.$$

For all k large enough, it is also true that $c^T\omega_* - c^T\omega^{(k)} \leq m \cdot \|c\|^2$ (by Lemma 3.3), where

$$m := \frac{1}{2}\min\left(\frac{\epsilon}{2M\|c\|}, \frac{\sqrt{\epsilon}}{\|c\|\sqrt{\gamma}}\right).$$

Thus, choose k' large enough so that

$$\lambda_{k'} \leq -\epsilon \qquad \text{and} \qquad c^T\omega_* - c^T\omega^{(k')} \leq m \cdot \|c\|^2.$$

Let us consider the support function $q_{k'}(\omega^{(k')} + \alpha p)$ restricted to the direction $p := c$, and for the step-lengths $\alpha \in [0, 2m]$. The condition $q_{k'}(\omega^{(k')} + \alpha p) \leq 0$ holds for all $\alpha \in [0, 2m]$, since

$$\begin{aligned} q_{k'}(\omega^{(k')} + \alpha p) &= \lambda_{k'} + \nabla\lambda_{k'}^T(\alpha p) + \frac{\gamma}{2}\alpha^2\|p\|^2 \\ &\leq -\epsilon + \alpha\|\nabla\lambda_{k'}\|\|p\| + \frac{\gamma}{2}\alpha^2\|p\|^2 \\ &\leq -\epsilon + \epsilon/2 + \frac{\gamma}{2}\alpha^2\|c\|^2 \leq 0. \end{aligned}$$

In particular, for $\tilde{\alpha} := 2m$, we have

$$\begin{aligned} c^T(\omega^{(k')} + \tilde{\alpha}p) &= c^T\omega^{(k')} + \tilde{\alpha} \cdot \|c\|^2 \\ &\geq c^T\omega_* + m \cdot \|c\|^2 > c^T\omega_*, \end{aligned}$$

implying $c^T\omega^{(k'+1)} > c^T\omega_*$, which contradicts the monotonicity of $\{c^T\omega^{(k)}\}$. ∎

THEOREM 3.6 (convergence). *Suppose* $\{\omega^{(k)}\}$ *is such that* $\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ *is simple, and there exists a real scalar* $m > 0$ *satisfying* $\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\| \geq m$ *for each* $k \in \mathbb{N}$. *Then* $\lim_{k\to\infty} \theta_k = 0$, *where*

$$\theta_k := \arccos\left(\frac{c^T\ \nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))}{\|c\|\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\|}\right).$$

Proof. The update rule (2.9) can be arranged as

$$\omega^{(k+1)} - \omega^{(k)} = \frac{1}{\gamma}\left[\frac{\sqrt{\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\|^2 - 2\gamma\lambda_{\min}(\mathcal{A}(\omega^{(k)}))}}{\|c\|}\cdot c - \nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\right].$$

Multiplying both sides by c^T and taking the limits of both sides yield

$$\lim_{k\to\infty}\left[\|c\|\cdot\sqrt{\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\|^2 - 2\gamma\lambda_{\min}(\mathcal{A}(\omega^{(k)}))} - c^T\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\right] = 0.$$

Furthermore, by Lemma 3.5, the sequence $\{\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\}$ approaches zero, and the equation above simplifies to

$$\begin{aligned}\lim_{k\to\infty}&\left[\|c\|\cdot\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\| - c^T\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\right]\\ &= \lim_{k\to\infty}\|c\|\cdot\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\|\cdot[1-\cos(\theta_k)] = 0.\end{aligned}$$

Since $\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\|$ is assumed to be uniformly bounded away from zero, $\lim_{k\to\infty}$ $\theta_k = 0$ as desired. ∎

A corollary of the last two results is that whenever $\omega^{(k)} \to \omega_*$, the first order necessary conditions hold at ω_* in the more general nonsmooth setting, that is, regardless of the multiplicity of $\lambda_{\min}(\mathcal{A}(\omega_*))$. To see this, first note that the eigenvalue function $\lambda_{\min}(\mathcal{A}(\omega))$ is differentiable everywhere except on a subset Ω of $\mathbb{R}^d$ of measure zero. The generalized gradient of $\lambda_{\min}(\mathcal{A}(\cdot))$ at ω_* is given by [8, p. 11]

$$\partial\lambda_{\min}(\mathcal{A}(\omega_*)) := \mathrm{co}\Big\{\lim_{k\to\infty}\nabla\lambda_{\min}(\mathcal{A}(\tilde{\omega}^{(k)}))\ \mid\ \tilde{\omega}^{(k)}\to\omega_*,\ \ \tilde{\omega}^{(k)}\notin\Omega\ \forall k,\ \Big\{\nabla\lambda_{\min}(\mathcal{A}(\widetilde{\omega}^{(k)}))\Big\}\ \text{is convergent}\Big\},$$

where $\mathrm{co}(H)$ denotes the convex hull of the set H. The first order necessary condition in terms of this generalized gradient, regardless of the multiplicity of $\lambda_{\min}(\mathcal{A}(\omega_*))$, takes the form [3, Theorem 6.1.8]

$$\exists\mu > 0\ \ \text{s.t.}\ \ c \in \mu\cdot\partial\lambda_{\min}(\mathcal{A}(\omega_*)) \qquad \text{and} \qquad \lambda_{\min}(\mathcal{A}(\omega_*)) = 0. \tag{3.5}$$

The latter condition is evident from Lemma 3.5. Additionally, there exists a subsequence $\{\omega^{(k_j)}\}$ of $\{\omega^{(k)}\}$ such that $\{\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k_j)}))\}$ is convergent, because we have assumed $\{\omega^{(k)}\}$ is convergent and $\lambda_{\min}(\mathcal{A}(\omega))$ is the minimum of n continuously differentiable functions. We have $c/\|c\| = \lim_{j\to\infty}\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k_j)}))/\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k_j)}))\|$ due to Theorem 3.6. Consequently,

$$\tilde{\mu}c = \lim_{j\to\infty}\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k_j)})) \in \partial\lambda_{\min}(\mathcal{A}(\omega_*)),$$

where $\tilde{\mu} = (\lim_{j\to\infty}\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k_j)}))\|)/\|c\|$. Thus, the former condition in (3.5) also holds.

4. A practical algorithm. In finite precision arithmetic, we terminate when $(\cos\theta_k - 1)$ and $\lambda_{\min}(\mathcal{A}(\omega^{(k)}))$ are less than specified tolerances in absolute value. Termination occurs at smooth points due to rounding errors, and they satisfy the optimality conditions (3.1) up to rounding errors, thanks to Lemma 3.5 and Theorem 3.6. Rarely it happens that this point does not correspond to a local maximizer. This hurdle can be effectively overcome, and the algorithm can be made to converge to a local maximizer by exploiting the following second order necessary condition for the problem at hand.

THEOREM 4.1. *Suppose that ω_* is a local maximizer of (1.1) such that $\lambda_{\min}(\mathcal{A}(\omega_*))$ is simple, and $\nabla\lambda_{\min}(\mathcal{A}(\omega_*)) \neq 0$. Then (1) $\lambda_{\min}(\mathcal{A}(\omega_*)) = 0$, (2) $c = \mu\nabla\lambda_{\min}(\mathcal{A}(\omega_*))$ for some $\mu > 0$, and (3) $V^T\nabla^2\lambda_{\min}(\mathcal{A}(\omega_*))V \succeq 0$, where $V \in \mathbb{C}^{d\times(d-1)}$ is a matrix whose columns form an orthonormal basis for the subspace orthogonal to c.*

Once termination at a point $\omega^{(k)}$ satisfying (3.1) occurs up to rounding errors, we check whether the projected Hessian $\mathcal{H}_{V,k} := V^T\nabla^2\lambda_{\min}(\mathcal{A}(\omega^{(k)}))V$ is positive semidefinite. Here we disregard the highly unlikely possibility of a singular $\mathcal{H}_{V,k}$. If $\mathcal{H}_{V,k}$ is positive definite, we terminate with $\omega^{(k)}$ as a local maximizer due to the second order sufficient conditions, i.e., conditions (1)–(2) of Theorem 4.1 together with $V^T\nabla^2\lambda_{\min}(\mathcal{A}(\omega_*))V \succ 0$ guarantee that ω_* is a local maximizer. Otherwise, if $\mathcal{H}_{V,k}$ has a negative eigenvalue, then $\omega^{(k)}$ does not correspond to a local maximizer by Theorem 4.1; indeed the direction $p = Vv_-$ is a descent direction for $\lambda_{\min}(\mathcal{A}(\omega))$ at $\omega^{(k)}$, where v_- is an eigenvector corresponding to any negative eigenvalue of $\mathcal{H}_{V,k}$. For all $\alpha \in \mathbb{R}$ small in absolute value, we have $\lambda_{\min}(\mathcal{A}(\omega^{(k)} + \alpha p)) < 0$. Thus, for such an α, the point $\omega^{(k)} + \alpha p$ is feasible with respect to (1.1), yet first order optimality conditions (3.1) are violated at $\omega^{(k)} + \alpha p$. We restart the algorithm with the point $\omega^{(k)} + \alpha p$. Due to part (ii) of Theorem 3.2, this point is not a local maximizer of the new subproblem

$$\text{maximize } c^T\omega \qquad \text{subject to}$$
$$\lambda_{\min}(\mathcal{A}(\omega^{(k)} + \alpha p)) + \nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)} + \alpha p))^T(\omega - (\omega^{(k)} + \alpha p)) + \frac{\gamma}{2}\|\omega - (\omega^{(k)} + \alpha p)\|^2 \leq 0.$$

Thus, the algorithm terminates at another point $\hat{\omega}^{(\hat{k})}$, where again the first order optimality condition (3.1) holds up to rounding errors, and such that $c^T\hat{\omega}^{(\hat{k})} > c^T(\omega^{(k)} + \alpha p) = c^T\omega^{(k)}$. (The last equality is due to the orthogonality of c to the column space of V where p lies.) These remarks are summarized in Algorithm 1 below.

The progress of this practical algorithm is illustrated in Figure 1 on an example where a restart is applied. In the figure, the curves represent the boundary of the feasible set $\mathcal{F} \subset \mathbb{R}^2$. Here the rightmost point in $\mathcal{F}$ is sought, so $c = (1, 0)$. When the algorithm is started with $x_0 = (0, 0)$, it initially converges to a point ω_* where the first order optimality condition (3.1) holds. This is not a local maximizer; indeed the projected Hessian is the scalar -1.0368. Thus the algorithm is restarted by moving from ω_* vertically in the direction indicated by the dashed line segment in the figure. After this restart, the algorithm converges to the rightmost point in $\mathcal{F}$, which is marked by the magenta disk.

5. A fixed point view: rate of convergence. In this section, under the assumption that $\{\omega^{(k)}\}$ itself converges to a smooth local maximizer ω_*, we deduce a linear rate of convergence revealing also the factors affecting the speed of convergence. Throughout the section we use the shorthand $\lambda_* := \lambda_{\min}(\mathcal{A}(\omega_*))$,

Algorithm 1. Support function based constrained eigenvalue optimization.

Require: An analytic and Hermitian matrix-valued function $\mathcal{A} : \mathbb{R}^d \to \mathbb{C}^{n\times n}$, a vector $c \in \mathbb{R}^d$, and tolerances $\epsilon_1, \epsilon_2 \in \mathbb{R}^+$ for termination

$V \in \mathbb{R}^{d\times(d-1)} \leftarrow$ a matrix whose set of columns forms an orthonormal basis for the subspace of $\mathbb{R}^d$ orthogonal to c
Pick an arbitrary $\omega^{(0)} \in \mathbb{R}^d$ feasible with respect to (1.1); $k \leftarrow 0$
Repeat
loop
  While $\left|\frac{c^T\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))}{\|c\|\|\nabla\lambda_{\min}(\mathcal{A}(\omega^{(k)}))\|} - 1\right| > \epsilon_1$ **or** $|\lambda_{\min}(\mathcal{A}(\omega^{(k)}))| > \epsilon_2$ **do**
  loop
    $\omega^{(k+1)}$ is given by the update rule (2.9); $k \leftarrow k+1$
  end loop
  $\mathcal{H}_V \leftarrow V^T\nabla^2\lambda_{\min}(\mathcal{A}(\omega^{(k)}))V$
  if $\lambda_{\min}(\mathcal{H}_V) < 0$ **then**
    $v_n \leftarrow$ a unit eigenvector corresponding to $\lambda_{\min}(\mathcal{H}_V)$; $p \leftarrow Vv_n$
    $\alpha \leftarrow 1$
    While $\lambda_{\min}(\mathcal{A}(\omega^{(k)} + \alpha p)) > 0$ **do**
    loop
      $\alpha \leftarrow \alpha/2$
    end loop
    $\omega^{(k)} \leftarrow \omega^{(k)} + \alpha p$
  end if
end loop
Until $\lambda_{\min}(\mathcal{H}_V) > 0$
Output: $\omega^{(k)}$.

$\nabla\lambda_* := \nabla\lambda_{\min}(\mathcal{A}(\omega_*))$, and $\nabla^2\lambda_* := \nabla^2\lambda_{\min}(\mathcal{A}(\omega_*))$, respectively. As we shall see, the rate of convergence depends on the eigenvalue distribution of the projected Hessian

$$\mathcal{H}_V \;:=\; V^T\,\nabla^2\lambda_*\,V,$$

where $V \in \mathbb{R}^{d\times d-1}$ is an isometry with columns formed by an orthonormal basis for the subspace orthogonal to $\nabla\lambda_*$. In particular, the convergence is faster when the eigenvalues of $\mathcal{H}_V$ are closer to γ. In the extreme case, when $\mathcal{H}_V = \gamma I$, the rate of convergence becomes superlinear.

We will put a fixed point theory in use: it follows from (2.9) that the sequence $\{\omega^{(k)}\}$ is a fixed point sequence $\omega^{(k+1)} = f(\omega^{(k)})$ where

$$f(\omega) := \omega + \frac{1}{\gamma}\left[\frac{\sqrt{\|\nabla\lambda_{\min}(\mathcal{A}(\omega))\|^2 - 2\gamma\lambda_{\min}(\mathcal{A}(\omega))}}{\|c\|}\cdot c - \nabla\lambda_{\min}(\mathcal{A}(\omega))\right]. \tag{5.1}$$

It is straightforward to verify that any fixed point of $f(\omega)$ satisfies the first order optimality conditions (parts (1) and (2) of Theorem 4.1), or otherwise $\nabla\lambda_{\min}(\mathcal{A}(\omega))$ vanishes at the fixed point. Furthermore, the Jacobian of $f(\omega)$ given by

$$J(\omega) = I + \frac{1}{\gamma}\left[\frac{c\cdot\nabla\lambda_{\min}(\mathcal{A}(\omega))^T(\nabla^2\lambda_{\min}(\mathcal{A}(\omega)) - \gamma I)}{\|c\|\sqrt{\|\nabla\lambda_{\min}(\mathcal{A}(\omega))\|^2 - 2\gamma\lambda_{\min}(\mathcal{A}(\omega))}} - \nabla^2\lambda_{\min}(\mathcal{A}(\omega))\right]$$

for ω close to ω_* determines the rate of convergence.

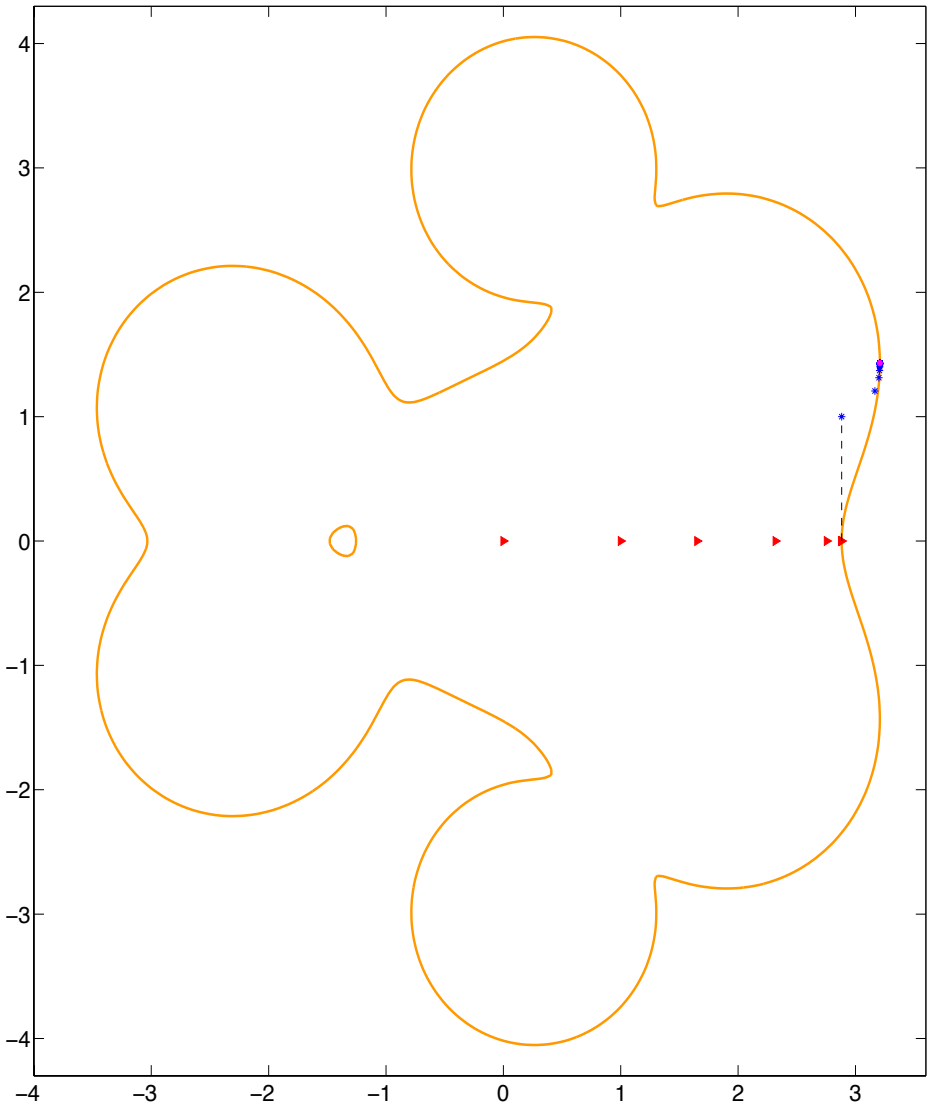

FIG. 1. *Progress of Algorithm* 1 *on an example. The initial iterates of the algorithm (red triangles) converge to a point* ω_* *satisfying the first order optimality conditions that is not a local maximizer. Then step* p *indicated by the dashed line segment is taken. The algorithm restarted with* $\omega_* + p$ *generates a sequence (blue asterisks) converging to a local maximizer marked with the magenta disk.*

At the local maximizer ω_*, due to parts (1) and (2) of Theorem 4.1, we must have $\lambda_* = 0$ and $c/\|c\| = \nabla\lambda_*/\|\nabla\lambda_*\|$. Thus, simple calculations yield

$$J(\omega_*) = \left[I - \frac{\nabla\lambda_* \cdot \nabla\lambda_*^T}{\|\nabla\lambda_*\|^2}\right]\left[I - \frac{1}{\gamma}\nabla^2\lambda_*\right] = (VV^T)\left[I - \frac{1}{\gamma}\nabla^2\lambda_*\right]. \tag{5.2}$$

We investigate the norm and the eigenvalue distribution of the projected Jacobian

$$V^T J(\omega_*)V = I - \frac{1}{\gamma}V^T\nabla^2\lambda_* V = \frac{1}{\gamma}(\gamma I - \mathcal{H}_V) \tag{5.3}$$

acting on $\mathrm{Col}(V)$ (the column space of V) in order to draw conclusions about the rate of convergence of the algorithm. In what follows, $\|A\|_2 := \max_{x\in\mathbb{F}^n,\|x\|=1} \|Ax\|_2$ denotes the 2-norm of a matrix $A \in \mathbb{F}^{n\times n}$ with real entries ($\mathbb{F} = \mathbb{R}$) or complex entries ($\mathbb{F} = \mathbb{C}$).

LEMMA 5.1. *Suppose $S \in \mathbb{R}^{k\times k}$ is a symmetric positive semidefinite matrix such that $\|S\|_2 \le 1$. Then $I - S$ is positive semidefinite and $\|I - S\|_2 \le 1$. Furthermore, if S is positive definite, then $\|I - S\|_2 < 1$.*

Proof. Since S is symmetric positive semidefinite and $\|S\|_2 \le 1$, each eigenvalue of S lies in $[0, 1]$. This in turn implies that each eigenvalue of $I - S$ belongs to $[0, 1]$. Consequently, $I - S$ is positive semidefinite and $\|I - S\|_2 \le 1$.

Additionally, if S is positive definite, each eigenvalue of S lies in $(0, 1]$, meaning that the eigenvalues of $I - S$ belong to $[0, 1)$. Consequently, $\|I - S\|_2 < 1$. ∎

THEOREM 5.2. *Let ω_* be a local maximizer of (1.1) such that λ_* is simple, and $\nabla\lambda_* \neq 0$. Then $V^T J(\omega_*)V$ is positive semidefinite and $\|V^T J(\omega_*)V\|_2 \le 1$. Additionally, if $\mathcal{H}_V \succ 0$, then $\|V^T J(\omega_*)V\|_2 < 1$.*

Proof. Observe that $\mathcal{H}_V \succeq 0$ due to part (3) of Theorem 4.1 and $\|\mathcal{H}_V\|_2 \le \gamma$ due to (2.1). Now the assertion immediately follows from (5.3) and Lemma 5.1 by letting $S = (1/\gamma)V^T\nabla^2\lambda_* V = (1/\gamma)\mathcal{H}_V$. ∎

We note that most often $\mathcal{H}_V$ is not only positive semidefinite but also positive definite amounting to the satisfaction of the sufficient conditions for the optimality of ω_*. Thus typically $\|V^T J(\omega_*)V\|_2 < 1$.

Our interest in the Jacobian and its projected variant at ω_* stem from $(\omega^{(k+1)} - \omega_*) \approx J(\omega_*)(\omega^{(k)} - \omega_*)$ for large k, which can be deduced from the mean value theorem by exploiting $\omega^{(k+1)} - \omega_* = f(\omega^{(k)}) - f(\omega_*)$. Remarkably, the projected Jacobian $V^T J(\omega_*)V$ assumes the role of $J(\omega_*)$ for large k in this identity, as it turns out $\omega^{(k)} - \omega_*$ lies almost on $\mathrm{Col}(V)$ for such k and has very little component in the direction of $u := \nabla\lambda_*/\|\nabla\lambda_*\|$. This in turn leads to the conclusion that the rate of convergence is typically linear and depends on the singular values, equivalently the eigenvalues, of $V^T J(\omega_*)V = (1/\gamma)(\gamma I - \mathcal{H}_V)$. Next, these assertions are presented formally and proven.

THEOREM 5.3 (rate of convergence). *Suppose that $\{\omega^{(k)}\}$ converges to a local maximizer ω_*, λ_* is simple, and $\nabla\lambda_* \neq 0$. Either*

$$\liminf_{k\to\infty} \frac{\|\omega^{(k+1)} - \omega_*\|}{\|\omega^{(k)} - \omega_*\|} = 0 \tag{5.4}$$

or otherwise both of the following hold:

(i) $u^T(\omega^{(k)} - \omega_*)/\|\omega^{(k)} - \omega_*\| \to 0$, $\|V^T(\omega^{(k)} - \omega_*)\|/\|\omega^{(k)} - \omega_*\| \to 1$, *as* $k \to \infty$.

(ii) *For each $\mu_1, \mu_2 \in \mathbb{R}^+$ satisfying*

$$\mu_1 < (1/\gamma)\sigma_{\min}(\gamma I - \mathcal{H}_V) \text{ and } \mu_2 > (1/\gamma)\|\gamma I - \mathcal{H}_V\|_2,$$

there exists an integer K such that for all $k > K$ we have

$$\mu_1 \le \frac{\|\omega^{(k+1)} - \omega_*\|}{\|\omega^{(k)} - \omega_*\|} \le \mu_2 \ , \tag{5.5}$$

where $\sigma_{\min}(\cdot)$ denotes the smallest singular value of its matrix argument.

Proof. Letting $p_k = \omega^{(k)} - \omega_*$, we have

(5.6)
$$p_{k+1} = \omega^{(k+1)} - \omega_* = f(\omega^{(k)}) - f(\omega_*) = J_k \cdot p_k, \quad \text{where } J_k = \begin{bmatrix} \nabla f_1(\omega_* + \eta_1 p_k)^T \\ \nabla f_2(\omega_* + \eta_2 p_k)^T \\ \vdots \\ \nabla f_d(\omega_* + \eta_d p_k)^T \end{bmatrix}$$

for some $\eta_1, \ldots, \eta_d \in (0,1)$ by the mean value theorem. Above $f_j(\omega)$ denotes the jth component of $f(\omega)$. This recurrence can be rearranged as

(5.7)
$$\begin{aligned} p_{k+1} &= J(\omega_*)p_k + [J_k - J(\omega_*)]p_k \\ &= (VV^T)\left[I - \frac{1}{\gamma}\nabla^2\lambda_*\right]p_k + [J_k - J(\omega_*)]p_k. \end{aligned}$$

Let us assume that

(5.8)
$$\liminf_{k\to\infty} \frac{\|\omega^{(k+1)} - \omega_*\|}{\|\omega^{(k)} - \omega_*\|} \neq 0.$$

Multiplying both sides of (5.7) by u^T from the left and taking the absolute value lead us to

$$|u^T p_{k+1}| = |u^T[J_k - J(\omega_*)]p_k| \implies \lim_{k\to\infty} \frac{|u^T p_{k+1}|}{\|p_k\|} = \lim_{k\to\infty} \left| u^T[J_k - J(\omega_*)]\frac{p_k}{\|p_k\|} \right| = 0.$$

Furthermore, due to assumption (5.8), there exists a real scalar $m' > 0$ and an integer K' such that

$$m' < \frac{\|p_{k+1}\|}{\|p_k\|} = \frac{\sqrt{\|V^T p_{k+1}\|^2 + |u^T p_{k+1}|^2}}{\|p_k\|} \quad \forall k > K'.$$

Since $u^T p_{k+1}/\|p_k\| \to 0$ as $k \to \infty$, this in turn implies the existence of a real scalar $m'' > 0$ and an integer K'' such that

$$m'' < \frac{\|V^T p_{k+1}\|}{\|p_k\|} \quad \forall k > K''.$$

Consequently, we obtain

(5.9)
$$\lim_{k\to\infty} \left(\frac{|u^T p_{k+1}|}{\|p_k\|}\right) / \left(\frac{\|V^T p_{k+1}\|}{\|p_k\|}\right) = \lim_{k\to\infty} \frac{|u^T p_{k+1}|}{\|V^T p_{k+1}\|} = 0.$$

Moreover, by taking the limit of

$$1 = \frac{\sqrt{(u^T p_k)^2 + \|V^T p_k\|^2}}{\|p_k\|} = \frac{\|V^T p_k\|}{\|p_k\|}\sqrt{1 + \frac{(u^T p_k)^2}{\|V^T p_k\|^2}}$$

and exploiting (5.9), we deduce

$$\lim_{k\to\infty} \frac{\|V^T p_k\|}{\|p_k\|} = 1 \text{ and } \lim_{k\to\infty} \frac{u^T p_k}{\|p_k\|} = 0$$

as desired.

To prove (5.5), we rewrite recurrence (5.7) in the following form:

(5.10)
$$\frac{p_{k+1}}{\|p_k\|} = VV^T\left[I-\frac{1}{\gamma}\nabla^2\lambda_*\right]VV^T\frac{p_k}{\|p_k\|} + VV^T\left[I-\frac{1}{\gamma}\nabla^2\lambda_*\right]uu^T\frac{p_k}{\|p_k\|} + [J_k - J(\omega_*)]\frac{p_k}{\|p_k\|}.$$

The last two terms on the right vanish as $k \to \infty$, so they can be made arbitrarily close to 0 by choosing k large enough. Furthermore, since the norm of $V^T p_k/\|p_k\|$ approaches 1 in the limit, by choosing k large enough, the first term on the right can be made to lie in any interval

$$\begin{aligned}[\mu_1', \mu_2'] &\supset \left[\sigma_{\min}\left(VV^T\left[I-\frac{1}{\gamma}\nabla^2\lambda_*\right]V\right), \left\|VV^T\left[I-\frac{1}{\gamma}\nabla^2\lambda_*\right]V\right\|_2\right] \\ &= \left[(1/\gamma)\sigma_{\min}(\gamma I - \mathcal{H}_V), (1/\gamma)\|\gamma I - \mathcal{H}_V\|_2\right].\end{aligned}$$

Note that the singular values of $V^T[I-(1/\gamma)\nabla^2\lambda_*]V = (1/\gamma)(\gamma I - \mathcal{H}_V)$ are not affected by left-multiplication with V. This is because this left-multiplication corresponds to merely a change of basis. Consequently, assertion (ii) follows. ∎

Remark 5.4. It is apparent from recurrence (5.7) that the condition $\liminf_{k\to\infty} \|\omega^{(k+1)} - \omega_*\|/\|\omega^{(k)} - \omega_*\| = 0$ holds if and only if there exists a subsequence $\{\omega^{(k_j)}\}$ of $\{\omega^{(k)}\}$ such that either

(i) $\lim_{j\to\infty} \operatorname{span}\left\{(\omega^{(k_j)} - \omega_*)/\|\omega^{(k_j)} - \omega_*\|\right\} \subseteq \operatorname{Null}\left(I - (1/\gamma)\nabla^2\lambda_*\right)$, or

(ii) $\lim_{j\to\infty} \operatorname{span}\left\{\left[I - (1/\gamma)\nabla^2\lambda_*\right](\omega^{(k_j)} - \omega_*)/\|\omega^{(k_j)} - \omega_*\|\right\} = \operatorname{span}\{\nabla\lambda_*\}$.

6. Estimation of an upper bound on second derivatives. The practicality of the algorithm presented and analyzed relies on the availability of an upper bound γ satisfying (2.1). The next result is helpful in determining such a γ analytically. An analogous result was proven in [19, Theorem 6.1] for the Hessian of a weighted sum of the j largest eigenvalues.

THEOREM 6.1. *Let $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times n}$ be a Hermitian and analytic matrix-valued function. Then*

$$\lambda_{\max}\left[\nabla^2\lambda_{\min}(\mathcal{A}(\omega))\right] \le \lambda_{\max}\left(\nabla^2\mathcal{A}(\omega)\right)$$

for all $\omega \in \mathbb{R}^d$ such that $\lambda_{\min}(\mathcal{A}(\omega))$ is simple, where

$$\nabla^2\mathcal{A}(\omega) := \begin{bmatrix} \frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_1^2} & \frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_1\partial\omega_2} & \cdots & \frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_1\partial\omega_d} \\ \frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_2\partial\omega_1} & \frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_2^2} & \cdots & \frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_2\partial\omega_d} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_d\partial\omega_1} & \frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_d\partial\omega_2} & \cdots & \frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_d^2} \end{bmatrix}.$$

Proof. By Theorem 2.1 part (iv), we have

$$\nabla^2\lambda_{\min}(\mathcal{A}(\omega)) = H_n(\omega) + 2\sum_{m=1}^{n-1}\frac{1}{\lambda_{\min}(\mathcal{A}(\omega)) - \lambda_m(\omega)}\Re\left(H_{n,m}(\omega)\right),$$

where the entries of $H_n(\omega)$ and $H_{n,m}(\omega)$ at position (k,ℓ) are given by

$$v_n^*(\omega)\frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_k\,\partial\omega_\ell}v_n(\omega) \quad\text{and}\quad \left(v_n(\omega)^*\frac{\partial\mathcal{A}(\omega)}{\partial\omega_k}v_m(\omega)\right)\left(v_m(\omega)^*\frac{\partial\mathcal{A}(\omega)}{\partial\omega_\ell}v_n(\omega)\right),$$

respectively. It is straightforward to verify that $H_{n,m}(\omega)$ is positive semidefinite, since for each $u \in \mathbb{C}^d$ we have

$$u^* H_{n,m}(\omega) u = \left| \sum_{\ell=1}^{d} h_\ell^{(n,m)} u_\ell \right|^2 \geq 0, \qquad \text{where } \; h_\ell^{(n,m)} = v_m(\omega)^* \frac{\partial \mathcal{A}(\omega)}{\partial \omega_\ell} v_n(\omega).$$

This implies that $\Re\left(H_{n,m}(\omega)\right)$ is also positive semidefinite due to $u^T \Re\left(H_{n,m}(\omega)\right) u = u^T H_{n,m}(\omega) u \geq 0$ for each $u \in \mathbb{R}^d$. Thus we deduce

$$\lambda_{\max}\left[\nabla^2 \lambda_{\min}(\mathcal{A}(\omega))\right] \leq \lambda_{\max}(H_n(\omega)) \leq \lambda_{\max}(\nabla^2 \mathcal{A}(\omega)).$$

Denoting the Kronecker product with $\otimes$, the last inequality above follows from $H_n(\omega) = [I_d \otimes v_n^*(\omega)] \nabla^2 \mathcal{A}(\omega) [I_d \otimes v_n(\omega)]$ and the observation that there exists a unit vector $v \in \mathbb{C}^d$ satisfying

$$\lambda_{\max}(H_n(\omega)) = v^* H_n(\omega) v = \left[v^* \otimes v_n^*(\omega)\right] \nabla^2 \mathcal{A}(\omega) \left[v \otimes v_n(\omega)\right] \leq \lambda_{\max}(\nabla^2 \mathcal{A}(\omega)). \quad \square$$

7. Case study: pseudospectral functions.

Pseudospectral abscissa. The ϵ-pseudospectrum of a matrix $A \in \mathbb{C}^{n \times n}$ is the subset of the complex plane consisting of the eigenvalues of all matrices within an ϵ-neighborhood of A, formally defined by[1]

$$\Lambda_\epsilon(A) := \bigcup_{\|\Delta\|_2 \leq \epsilon} \Lambda(A + \Delta),$$

with the singular value characterization [23]

$$\Lambda_\epsilon(A) = \{z \in \mathbb{C} \mid \sigma_{\min}(A - zI) \leq \epsilon\}. \tag{7.1}$$

The rightmost point in this set $\alpha_\epsilon(A)$ is called the ϵ-pseudospectral abscissa and is an indicator of the transient behavior of the dynamical system $x'(t) = Ax(t)$. Specifically, a large $\alpha_\epsilon(A)$ relative to ϵ indicates that $\|x(t)\|$ exhibits growth for some initial condition, even if the system is asymptotically stable, as implied by the Kreiss matrix theorem for matrix exponentials [23, Theorem 18.5].

Globally and locally convergent algorithms for $\alpha_\epsilon(A)$ have been suggested in [5] and [10], respectively. Algorithm 1 can also be employed for the computation of $\alpha_\epsilon(A)$, since $\alpha_\epsilon(A)$ can be cast as the optimization problem

$$\begin{aligned} &\text{maximize}_{\omega \in \mathbb{R}^2} \;\; \omega_1 \quad \text{subject to} \;\; \lambda_{\min}(\mathcal{A}(\omega)) \leq 0, \\ &\mathcal{A}(\omega) := [A - (\omega_1 + i\omega_2) I]^* \left[A - (\omega_1 + i\omega_2) I\right] - \epsilon^2 I \end{aligned}$$

that fits into framework (1.1). It follows from the expressions in part (iv) of Lemma 2.1 that

$$\nabla \lambda_{\min}(\mathcal{A}(\omega)) = \left(\; v_n(\omega)^* (-A - A^* + 2\omega_1 I) v_n(\omega), \;\; v_n(\omega)^* (iA - iA^* + 2\omega_2 I) v_n(\omega) \; \right).$$

Furthermore, the matrix $\nabla^2 \mathcal{A}(\omega)$ in Theorem 6.1 is given by

$$\nabla^2 \mathcal{A}(\omega) = \begin{bmatrix} \frac{\partial \mathcal{A}^2(\omega)}{\partial \omega_1^2} & \frac{\partial \mathcal{A}^2(\omega)}{\partial \omega_1 \partial \omega_2} \\ \frac{\partial \mathcal{A}^2(\omega)}{\partial \omega_2 \partial \omega_1} & \frac{\partial \mathcal{A}^2(\omega)}{\partial \omega_2^2} \end{bmatrix} = 2I.$$

Consequently, $\lambda_{\max}\left[\nabla^2 \lambda_{\min}(\mathcal{A}(\omega))\right] \leq \gamma := 2$ for all ω such that $\lambda_{\min}(\mathcal{A}(\omega))$ is simple.

[1] This definition is in terms of the matrix 2-norm. It could be defined in terms of other matrix norms. But the singular value characterization would not be valid anymore. The ϵ-pseudospectrum we consider here in terms of the matrix 2-norm is the one most widely used.

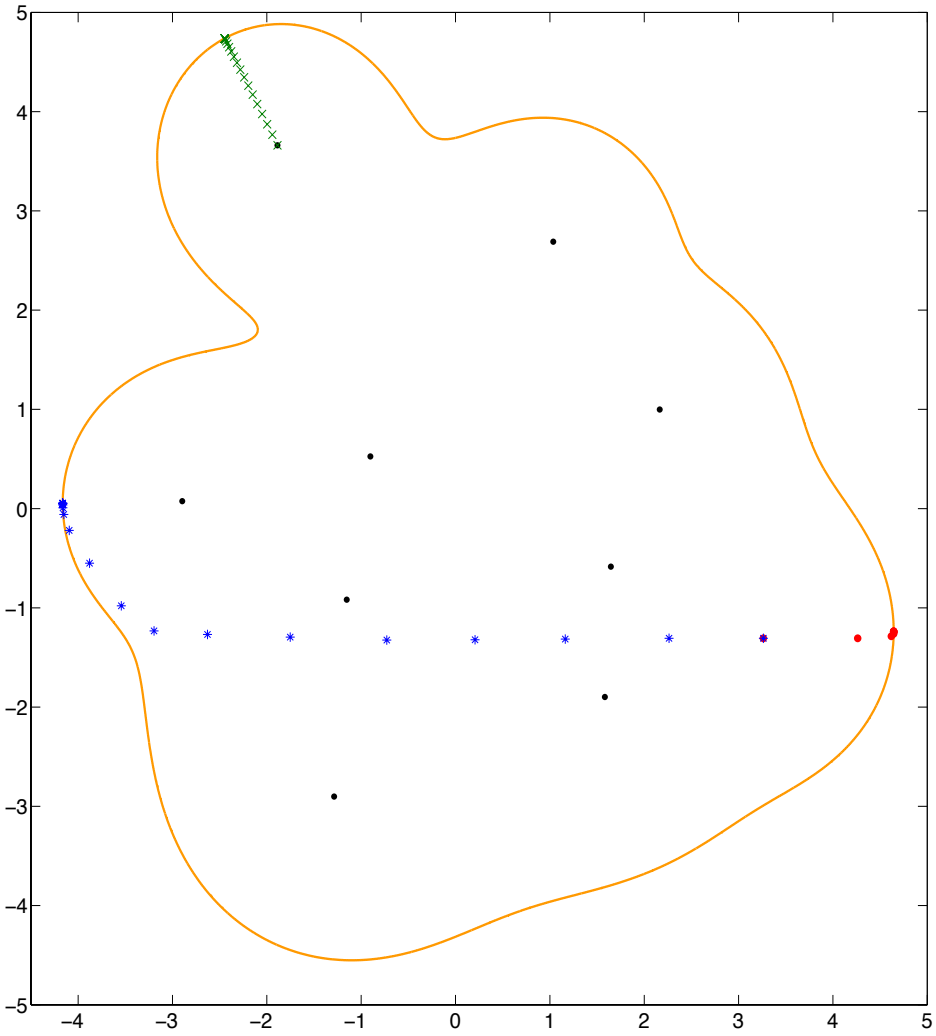

FIG. 2. *The progress of the algorithm to compute the rightmost, leftmost, and outermost points in* $\Lambda_\epsilon(A)$ *is illustrated for a random* 10×10 *matrix and* $\epsilon = 1$. *The orange curve represents the boundary of* $\Lambda_\epsilon(A)$, *while the red disks, blue asterisks, and green crosses mark the iterates of the algorithm to compute the rightmost, leftmost, and outermost points in* $\Lambda_\epsilon(A)$. *The real part of the rightmost and the modulus of the outermost points correspond to* $\alpha_\epsilon(A)$ *and* $\rho_\epsilon(A)$, *respectively.*

We illustrate Algorithm 1 on a 10×10 random matrix with entries selected from a normal distribution with zero mean and unit variance. We compute the real parts of both the rightmost point and the leftmost point in $\Lambda_\epsilon(A)$ for $\epsilon = 1$. Initially, $\omega^{(0)}$ is chosen as the rightmost eigenvalue. The iterates for the rightmost and leftmost points are depicted in Figure 2 with red disks and blue asterisks, respectively. The algorithm requires 39 iterations to compute these extreme points accurate up to 12 decimal digits. In Figure 3, on the left the later iterates of the algorithm for the rightmost point are shown. At the later iterations, the directions $(\omega^{(k)} - \omega_*)$ become more or less orthogonal to $\nabla\lambda_{\min}(\mathcal{A}(\omega_*))$, equivalently, tangent to the boundary of

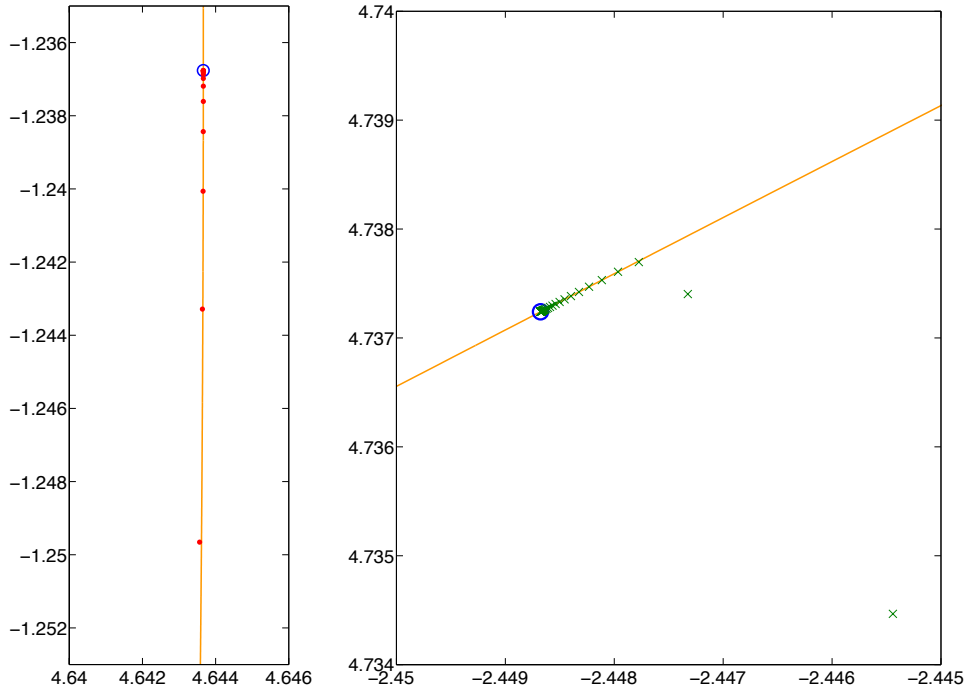

FIG. 3. *Left: A closer look at the iterates (red disks) for the rightmost point in $\Lambda_\epsilon(A)$; the blue circle marks the rightmost point computed by the highly accurate algorithm in* [5]. *Right: Later iterates (green crosses) for the outermost point in $\Lambda_\epsilon(A)$; the blue circle marks the outermost point computed by the algorithm in* [18].

TABLE 1

The errors $\|\omega^{(k)} - \omega_*\|$ for various k to compute the rightmost point in $\Lambda_\epsilon(A)$, rounded to 12 *decimal digits, are listed for the random matrix example and $\epsilon = 1$.*

k	29	30	31	32	33
$\|\omega^{(k)} - \omega_*\|$	$1.041 \cdot 10^{-9}$	$5.27 \cdot 10^{-10}$	$2.67 \cdot 10^{-10}$	$1.35 \cdot 10^{-10}$	$6.9 \cdot 10^{-11}$

$\Lambda_\epsilon(A)$, as suggested by assertion (i) of Theorem 5.3. As for the rate of convergence, the errors $\|\omega^{(k)} - \omega_*\|$ involved in computing the rightmost point at later iterations listed in Table 1 indicate a linear convergence. The projected Hessian reduces to the scalar $\mathcal{H}_V = \frac{\partial^2 \lambda_{\min}(\mathcal{A}(\omega_*))}{\partial \omega_2^2}$, so assertion (ii) of Theorem 5.3 implies

$$\lim_{k\to\infty} \frac{\|\omega^{(k+1)} - \omega_*\|}{\|\omega^{(k)} - \omega_*\|} = \left|1 - \frac{1}{\gamma}\mathcal{H}_V\right| = \left|1 - \frac{1}{2}\frac{\partial^2 \lambda_{\min}(\mathcal{A}(\omega_*))}{\partial \omega_2^2}\right|.$$

For the particular 10×10 example, the ratio above is 0.506 (rounded to three decimal digits). This ratio of decay in the error is confirmed in practice by Table 1.

This application of the algorithm here to the pseudospectral abscissa and the algorithm in [10] both generate sequences converging locally at a linear rate. However, the algorithm here is based on the computation of smallest singular values repeatedly. On the other hand, the algorithm in [10] requires the computation of rightmost eigenvalues, which is typically less reliable than the computation of smallest singular

values. The algorithm in [5] is globally convergent. But it is considerably less efficient compared to the algorithms here and in [10], as it requires the computation of all eigenvalues of Hamiltonian matrices of twice the size of the original problem.

Pseudospectral radius. The modulus of the outermost point $\rho_\epsilon(A)$ in the ϵ-pseudospectrum $\Lambda_\epsilon(A)$ is called the ϵ-pseudospectral radius [23]. This quantity is associated with the transient behavior of the discrete dynamical system $x_{k+1} = Ax_k$, as implied by the Kreiss matrix theorem for matrix powers [23, Theorem 18.1].

When $\Lambda_\epsilon(A)$ is defined in terms of the matrix 2-norm and using the singular value characterization (7.1), we deduce

$$\begin{aligned} \rho_\epsilon(A) = \text{maximize}_{\omega\in\mathbb{R}^2} \ \ \omega_1 \quad \text{subject to} \ \ \lambda_{\min}(\mathcal{A}(\omega)) \leq 0, \\ \mathcal{A}(\omega) := \left(A - \omega_1 e^{i\omega_2} I\right)^* \left(A - \omega_1 e^{i\omega_2} I\right) - \epsilon^2 I. \end{aligned} \tag{7.2}$$

Now the expressions for the first derivatives take the form

$$\nabla \lambda_{\min}(\mathcal{A}(\omega)) = \big(v_n(\omega)^*(-2\Re(e^{-i\omega_2}A)+2\omega_1 I)v_n(\omega),\ v_n(\omega)^*(-2\Im(\omega_1 e^{-i\omega_2}A))v_n(\omega)\big),$$

whereas

$$\nabla^2 \mathcal{A}(\omega) = \begin{bmatrix} 2I & -2\Im(e^{-i\omega_2}A) \\ -2\Im(e^{-i\omega_2}A) & 2\Re(\omega_1 e^{-i\omega_2}A) \end{bmatrix}.$$

Since $\rho_\epsilon(A) \leq \|A\|_2 + \epsilon$, for all feasible ω we have $\omega_1 \leq \|A\|_2 + \epsilon$. Thus, Gershgorin's theorem [12, Theorem 6.1.1] applied to $\nabla^2\mathcal{A}(\omega)$, combined with Theorem 6.1, yields

$$\lambda_{\max}\left[\nabla^2 \lambda_{\min}(\mathcal{A}(\omega))\right] \leq \gamma := \max\left(2 + 2\|A\|_2, 2\epsilon\|A\|_2 + 2\|A\|_2^2 + 2\|A\|_2\right)$$

for all $\omega \in \mathcal{B}(0, \|A\|_2 + \epsilon)$ such that $\lambda_{\min}(\mathcal{A}(\omega))$ is simple. Here, we benefit from the fact that $q_k(\omega)$ for each k is a support function on the ball $\mathcal{B}(0, \|A\|_2 + \epsilon)$, which contains the feasible set $\mathcal{F}$ (see the note after Theorem 2.2).

We apply Algorithm 1 to compute $\rho_\epsilon(A)$ starting with $\omega^{(0)}$ set equal to the eigenvalue with the largest modulus. In Figure 2, the green crosses mark the iterates of the algorithm on the 10×10 random matrix of the previous part and for $\epsilon = 1$. The directions $(\omega^{(k)} - \omega_*)$ for the later iterates $\omega^{(k)}$, as illustrated on the right in Figure 3, again become tangent to the boundary of $\Lambda_\epsilon(A)$, which is in harmony with assertion (i) of Theorem 5.3. For the rate of convergence, by assertion (ii) of Theorem 5.3, we have

$$\lim_{k\to\infty} \frac{\|\omega^{(k+1)} - \omega_*\|}{\|\omega^{(k)} - \omega_*\|} = \left|1 - \frac{1}{\gamma}\frac{\partial^2 \lambda_{\min}(\mathcal{A}(\omega_*))}{\partial \omega_2^2}\right|$$

for the matrix-valued function $\mathcal{A}(\omega)$ in (7.2). This quantity is equal to 0.791 (rounded to three decimal digits) for the particular example. This is confirmed by Table 2, as it indicates a linear convergence with the ratio of two consecutive errors roughly equal to 0.791.

TABLE 2

The errors $\|\omega^{(k)} - \omega_*\|$ for various k to compute the outermost point in $\Lambda_\epsilon(A)$ are listed for the random matrix example and $\epsilon = 1$.

k	41	42	43	44	45
$\|\omega^{(k)} - \omega_*\|$	1.105×10^{-7}	$8.742 \cdot 10^{-8}$	$6.918 \cdot 10^{-8}$	$5.474 \cdot 10^{-8}$	$4.332 \cdot 10^{-8}$

8. Variations. We consider the following variations of the main framework:

$$\text{maximize} \quad c^T\omega \qquad \text{subject to} \quad \lambda_{\max}(\mathcal{A}(\omega)) \le 0, \quad \text{and} \tag{8.1}$$

$$\text{maximize} \quad c^T\omega \qquad \text{subject to} \quad \lambda_{\min}(\mathcal{A}(\omega)) \ge 0. \tag{8.2}$$

The family of convex as well as nonconvex semidefinite programs are closely related to these frameworks. More specific applications include certain robust stability problems in control theory which can be addressed within the first framework, and structural design problems (where a smallest eigenvalue corresponding to the buckling load of the structure [9] is required to be bounded below) that are relevant to the second framework. Indeed, each one of these frameworks can be turned into the other, as the conditions $\lambda_{\min}(\mathcal{A}(\omega)) \ge 0$ and $\lambda_{\max}(-\mathcal{A}(\omega)) \le 0$ are equivalent. Algorithm 1 defined in terms of $\lambda_{\max}(\cdot)$ rather than $\lambda_{\min}(\cdot)$ usually works well in practice in our experience. Letting $\hat{\lambda}_k := \lambda_{\max}(\mathcal{A}(\omega^{(k)}))$ and $\nabla\hat{\lambda}_k := \nabla\lambda_{\max}(\mathcal{A}(\omega^{(k)}))$, this is explained by the property

$$\lambda_{\max}(\mathcal{A}(\omega)) \;\le\; \hat{q}_k(\omega) := \hat{\lambda}_k + \nabla\hat{\lambda}_k^T(\omega - \omega^{(k)}) + \frac{\gamma}{2}\|\omega - \omega^{(k)}\|^2 \quad \forall\omega \in \mathbb{R}^d \tag{8.3}$$

that seems to hold for large γ.

We illustrate the algorithm for (8.1) and (8.2) on two examples. In both of these examples, we set $\gamma = 100$ in (8.3) and $\epsilon_1 = \epsilon_2 = 10^{-6}$ in Algorithm 1. The first one concerns the stabilization of a linear control system by output feedback, that is, for a given system

$$x'(t) = Ax(t) + Bu(t), \qquad y(t) = Cx(t),$$

where $A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{n\times m}, C \in \mathbb{C}^{p\times n}$ determination of a matrix $K \in \mathbb{C}^{m\times p}$, consequently an output feedback $u(t) = Ky(t)$, such that the system is asymptotically stable. This could equivalently be expressed as the determination of a matrix K such that $A + BKC$ has all of its eigenvalues on the left half of the complex plane. This problem has drawn substantial interest since the 1970s; for instance, derivation of algorithms and simple mathematical characterizations to check the existence of stabilizing output feedbacks have been referred to as major open problems in the survey paper [2]. We refer to [22] and references therein for recent developments regarding output feedback stabilization. Here we focus on the case when A is Hermitian, $C = B^*$, and $K = K(\omega) = \sum_{j=1}^{\kappa}\omega_j K_j$, where K_j is a given Hermitian matrix and $\omega_j \in \mathbb{R}$ is a parameter for $j = 1, \ldots, \kappa$. We assume that the system is stabilizable by output feedback for certain values of parameters $\omega_1, \ldots, \omega_\kappa$ and aim for a robustly stable system that is as close to the original system as possible in a certain sense. Formally, we would like to solve

$$\text{minimize} \quad \sum_{j=1}^{\kappa}\omega_j\|K_j\|_2 \qquad \text{subject to} \quad \lambda_{\max}(A + BK(\omega)B^*) \le -\delta \tag{8.4}$$

for a given $\delta > 0$ over the parameters $\omega \in \mathbb{R}^\kappa$. This falls into the scope of framework (8.1). In Figure 4 on the top, the progress of the algorithm is shown for such an example with two parameters (i.e., $\kappa = 2$), $\delta = 0.1$, $A \in \mathbb{C}^{20\times 20}$, $B \in \mathbb{C}^{20\times 10}$, and $K_1, K_2 \in \mathbb{C}^{10\times 10}$ are random matrices.[2] The curves represent the level sets of $\lambda_{\max}(A + BK(\omega)B^*)$, in particular the rightmost curve represents the -0.1 level set. The algorithm started from $\omega^{(0)} = (11, 11)$ requires 546 iterations to converge

[2]The matrices are available at http://home.ku.edu.tr/~emengi/software/robustcontrol.mat.

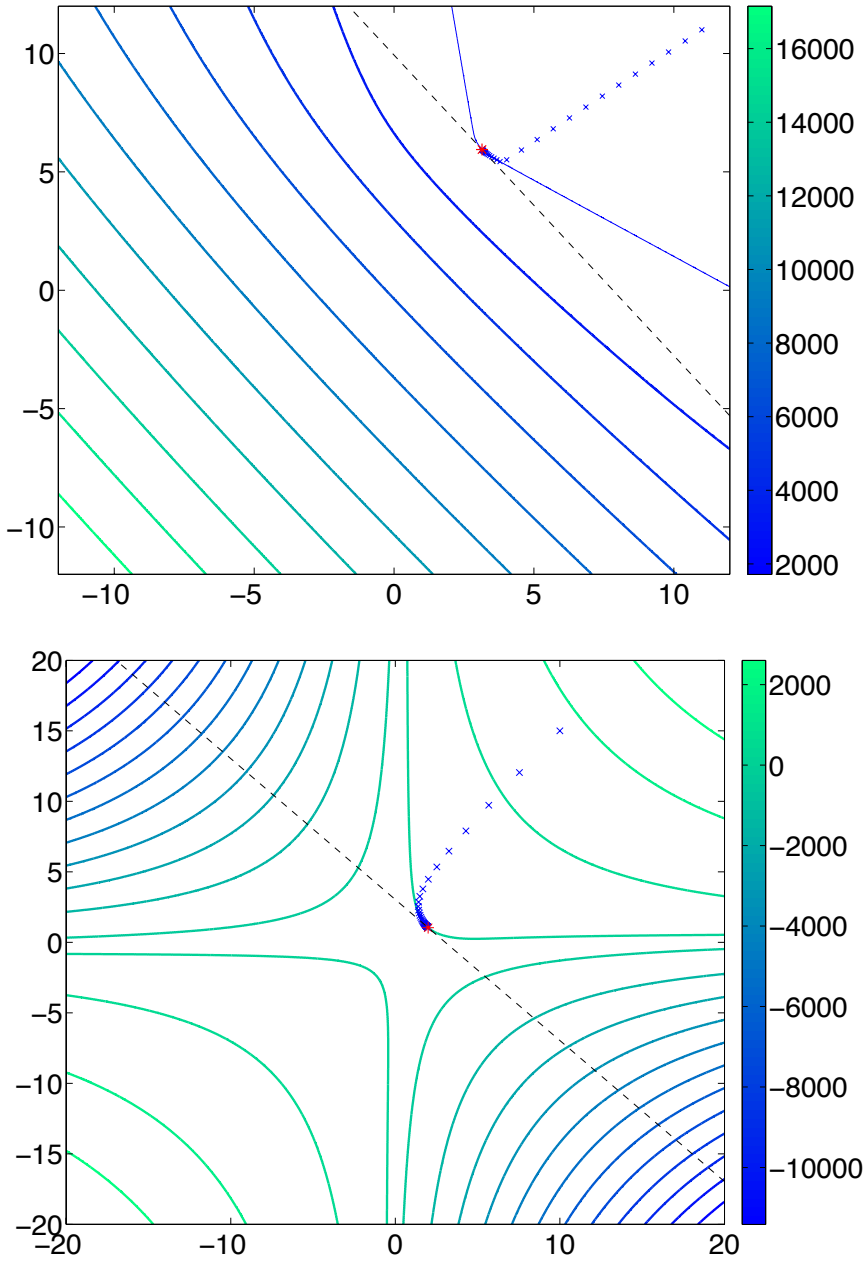

FIG. 4. *Illustration of the algorithm for the solutions of* (8.4), (8.5). *In each figure, the curves represent the level sets of the associated eigenvalue function. On the top, the rightmost curve corresponds to the* -0.1 *level set. At the bottom, the curve to which the dashed line is tangent corresponds to the* 0.1 *level set. The blue crosses are the iterates generated by the algorithm, whereas the red asterisks are the converged minimizer* ω_* *up to tolerances. The dashed line represents* $\{\omega \mid c^T\omega = c^T\omega_*\}$ *for* $c = [\|K_1\|_2 \ \|K_2\|_2]^T$ *on the top and for* $c = [1 \ 1]^T$ *at the bottom. In each case, this line is tangent to the boundary of a level set at* ω_*, *specifically,* -0.1 *level set of the eigenvalue function on the top and* 0.1 *level set of the eigenvalue function at the bottom.*

to $\omega_* = (3.1431, 5.9440)$ (marked by the red asterisk in the figure). This converged point ω_* is a local minimizer up to tolerances, as indeed $\lambda_{\max}(A + BK(\omega_*)B^*) = -0.10000002$ and $\left|c^T \nabla \lambda_{\max}(\mathcal{A}(\omega_*))/(\|c\|\|\nabla \lambda_{\max}(\mathcal{A}(\omega_*))\|) - 1\right| = 9.96 \cdot 10^{-7}$ for $c = \begin{bmatrix} \|K_1\|_2 & \|K_2\|_2 \end{bmatrix}^T$. Convergence behavior similar to the pseudospectral functions is observed here. For instance, it is clear from the figure that $(\omega^{(k)} - \omega_*)/\|\omega^{(k)} - \omega_*\|$ becomes tangent to the boundary of the feasible set in the limit.

The second example concerns

$$\text{minimize} \quad \sum_{j=1}^{\kappa} \omega_j \qquad \text{subject to} \quad \lambda_{\min}(\mathcal{A}(\omega)) \geq \beta \tag{8.5}$$

for a given $\beta > 0$. This problem can be interpreted as the minimization of the volume of a column subject to a constraint on its buckling load. It may be desirable to include other (typically linear and bilinear) constraints in addition to the eigenvalue constraint and impose the smallest eigenvalue constraint in terms of a generalized eigenvalue problem as in [1]. Here, we employ this simplified formulation, but in terms of a bilinear matrix-valued function $\mathcal{A}(\omega) := \sum_{j=1}^{\kappa} \sum_{k=1}^{\kappa} \omega_j \omega_k A_{jk} + \sum_{j=1}^{\kappa} \omega_j A_j + A_0$, which makes the problem nonconvex. We consider the problem with respect to two parameters (i.e., $\kappa = 2$), $\beta = 0.1$ and $A_0, A_1, A_2, A_{12} \in \mathbb{C}^{20\times 20}$, that are random Hermitian.[3] An illustration of the algorithm is given at the bottom in Figure 4. The curves represent the level sets of $\lambda_{\min}(\mathcal{A}(\omega))$, in particular the innermost red curve represents the 0.1 level set. The algorithm initiated with $\omega^{(0)} = (10, 15)$ terminates at $\omega_* = (1.9963, 1.0400)$ after 60 iterations. It can be verified that this point is a local minimizer up to tolerances, indeed $\lambda_{\min}(\mathcal{A}(\omega_*)) = 0.1000008$ and $\left|(c^T \nabla \lambda_{\min}(\mathcal{A}(\omega_*)))/(\|c\|\|\nabla \lambda_{\min}(\mathcal{A}(\omega_*))\|) - 1\right| = 8.91 \cdot 10^{-8}$, where $c = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$.

9. Conclusion. Various applications give rise to problems that involve optimization of a linear objective function subject to minimum eigenvalue constraints, or optimization problems of similar spirit, such as the calculation of the pseudospectral functions, shape optimization problems in structural design, and robust stability problems in control theory. We explored support function based approaches that construct global overestimators for the eigenvalue function appearing in the constraint. The global overestimators are quadratic functions and utilize derivatives. They lead to a linearly convergent algorithm robust against the nonsmooth nature of smallest eigenvalue functions. We established a local convergence result regardless of the multiplicity of the eigenvalue at the optimal points. Thus, the algorithm is immune to nonsmoothness at the optimal points and at nearby points. The rate of the convergence of the algorithm is also analyzed, leading us to a fine understanding of the factor affecting the rate of convergence, basically the eigenvalue distribution of a projected Hessian matrix. A MATLAB implementation of this algorithm is available on the web.[4]

Efficient extensions of the algorithm when there are multiple smallest eigenvalue constraints, possibly mixed with quadratic constraints, may be worth exploring. When the linear objective function is replaced by a convex objective function, the algorithm seems applicable, but the convex and smooth subproblems need to be solved numerically, for instance, via an interior point method. Convergence and rate of convergence analyses in these more general settings may be intricate and worth studying.

[3] The matrices are available at http://home.ku.edu.tr/~emengi/software/structuredesign_bilinear.mat.

[4] http://home.ku.edu.tr/~emengi/software/eigoptc.html.

Acknowledgments. The author is grateful to Karl Meerbergen, Wim Michiels, Roel Van Beeumen, and Emre A. Yıldırım for fruitful discussions, as well as two anonymous reviewers for their comments. Part of this manuscript was written during the author's visit to Courant Institute of Mathematical Sciences, which was hosted by Michael L. Overton.

REFERENCES

[1] W. Achtziger and M. Kocvara, *Structural topology optimization with eigenvalues*, SIAM J. Optim., 18 (2007), pp. 1129–1164.
[2] V. Blondel, M. Gevers, and A. Lindquist, *Survey on the state of systems and control*, Eur. J. Control, 1 (1995), pp. 5–23.
[3] J. M. Borwein and A. S. Lewis, *Convex and Nonlinear Optimization*, Springer-Verlag, Berlin, 2000.
[4] S. Boyd and L. Vandenberghe, *Convex Optimization*, 7th ed., Cambridge University Press, New York, 2009.
[5] J. V. Burke, A. S. Lewis, and M. L. Overton, *Robust stability and a criss-cross algorithm for pseudospectra*, IMA J. Numer. Anal., 23 (2003), pp. 359–375.
[6] E. W. Cheney and A. A. Goldstein, *Newton's method for convex programming and Tchebycheff approximation*, Numer. Math., 1 (1959), pp. 253–268.
[7] F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc., 205 (1975), pp. 247–262.
[8] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.
[9] S. J. Cox and M. L. Overton, *On the optimal design of columns against buckling*, SIAM J. Math. Anal., 23 (1992), pp. 287–325.
[10] N. Guglielmi and M. L. Overton, *Fast algorithms for the approximation of the pseudospectral abscissa and pseudospectral radius of a matrix*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1166–1192.
[11] W. Hare and C. Sagastizabal, *A redistributed proximal bundle method for nonconvex optimization*, SIAM J. Optim., 20 (2010), pp. 2442–2473.
[12] R. A. Horn and C. A. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1990.
[13] J. E. Kelley, *The cutting-plane method for solving convex programs*, J. SIAM, 8 (1960), pp. 703–712, http://www.jstor.org/stable/2099058.
[14] K. C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Math. 1133, Springer-Verlag, New York, 1985.
[15] P. Lancaster, *On eigenvalues of matrices dependent on a parameter*, Numer. Math., 6 (1964), pp. 377–387.
[16] C. Lemaréchal, A. Nemirovskii, and Y. Nesterov, *New variants of bundle methods*, Math. Program., 69 (1995), pp. 111–147.
[17] M. Mäkelä, *Survey of bundle methods for nonsmooth optimization*, Optim. Methods Softw., 17 (2002), pp. 1–29, https://doi.org/10.1080/10556780290027828.
[18] E. Mengi and M. L. Overton, *Algorithms for the computation of the pseudospectral radius and the numerical radius of a matrix*, IMA J. Numer. Anal., 25 (2005), pp. 648–669.
[19] E. Mengi, E. A. Yildirim, and M. Kilic, *Numerical optimization of eigenvalues of Hermitian matrix functions*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 699–724.
[20] J. E. Mitchell, *Cutting Plane Methods and Subgradient Methods*, INFORMS Tutorials in Operations Research, INFORMS, 2009.
[21] F. Rellich, *Perturbation Theory of Eigenvalue Problems*, Gordon and Breach, London, 1969.
[22] J. Shen and J. Lam, *On static output-feedback stabilization for multi-input multi-output positive systems*, Internat. J. Robust Nonlinear Control, 25 (2015), pp. 3154–3162, https://doi.org/10.1002/rnc.3256.
[23] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra. The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, Princeton, NJ, 2005.