SIAM J. MATRIX ANAL. APPL.
Vol. 35, No. 2, pp. 699–724
© 2014 Society for Industrial and Applied Mathematics

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

# NUMERICAL OPTIMIZATION OF EIGENVALUES OF HERMITIAN MATRIX FUNCTIONS*

EMRE MENGI†, E. ALPER YILDIRIM‡, AND MUSTAFA KILIÇ†

**Abstract.** This work concerns the global minimization of a prescribed eigenvalue or a weighted sum of prescribed eigenvalues of a Hermitian matrix-valued function depending on its parameters analytically in a box. We describe how the analytical properties of eigenvalue functions can be put into use to derive piecewise quadratic functions that underestimate the eigenvalue functions. These piecewise quadratic underestimators lead us to a global minimization algorithm, originally due to Breiman and Cutler. We prove the global convergence of the algorithm and show that it can be effectively used for the minimization of extreme eigenvalues, e.g., the largest eigenvalue or the sum of the largest specified number of eigenvalues. This is particularly facilitated by the analytical formulas for the first derivatives of eigenvalues, as well as analytical lower bounds on the second derivatives that can be deduced for extreme eigenvalue functions. The applications that we have in mind also include the $H_\infty$-norm of a linear dynamical system, numerical radius, distance to uncontrollability, and various other nonconvex eigenvalue optimization problems, for which, generically, the eigenvalue function involved is simple at all points.

**Key words.** Hermitian eigenvalues, analytic, global optimization, perturbation of eigenvalues, quadratic programming

**AMS subject classifications.** 65F15, 90C26

**DOI.** 10.1137/130933472

## 1. Introduction.

The main object of this work is a matrix-valued function $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times n}$ that is *analytic* and *Hermitian* at all $\omega \in \mathbb{R}^d$. Here, we consider the numerical *global* minimization of a prescribed eigenvalue $\lambda(\omega)$ of $\mathcal{A}(\omega)$ over $\omega \in \mathcal{B} \subseteq \mathbb{R}^d$, where $\mathcal{B}$ denotes a box. From an application point of view, a prescribed eigenvalue typically refers to the $j$th largest eigenvalue, i.e., $\lambda(\omega) := \lambda_j(\mathcal{A}(\omega))$, or a weighted sum of $j$ largest eigenvalues, i.e., $\lambda(\omega) := \sum_{k=1}^{j} d_k \lambda_k(\mathcal{A}(\omega))$ for given real numbers $d_1, \ldots, d_j$. However, it may as well refer to a particular eigenvalue with respect to a different criterion as long as the (piecewise) analyticity properties discussed below and in section 3 are satisfied.

The literature from various engineering fields and applied sciences is rich with eigenvalue optimization problems that fit into the setting of the previous paragraph. There are problems arising in structural design and vibroacoustics, for which the minimization of the largest eigenvalue or maximization of the smallest eigenvalue of a matrix-valued function is essential, e.g., the problem of designing the strongest column which originated from Euler in the 18th century [30]. In control theory, various quantities regarding dynamical systems can be posed as eigenvalue optimization problems. For instance, the distance from a linear dynamical system to a nearest

*Received by the editors August 19, 2013; accepted for publication (in revised form) by P.-A. Absil February 24, 2014; published electronically June 5, 2014.
http://www.siam.org/journals/simax/35-2/93347.html

†Department of Mathematics, Koç University, 34450 Sarıyer, İstanbul, Turkey (emengi@ku.edu.tr, mukilic@ku.edu.tr). The work of the first author was supported in part by European Commission grant PIRG-GA-268355 and TÜBİTAK (Scientific and Technological Research Council of Turkey) career grant 109T660. The work of the third author was partly supported by European Commission grant PIRG-GA-268355.

‡Department of Industrial Engineering, Koç University, 34450 Sarıyer, İstanbul, Turkey (alperyildirim@ku.edu.tr). This author was supported in part by TÜBİTAK grant 112M870 and by TÜBA-GEBİP (Turkish Academy of Sciences Young Scientists Award Program).



Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

unstable system [47] and the $H_\infty$-norm of a linear dynamical system have nonconvex eigenvalue optimization characterizations [3]. In graph theory, relaxations of some NP-hard graph partitioning problems give rise to optimization problems in which the sum of the $j$ largest eigenvalues is to be minimized [10].

In this paper, we offer a generic algorithm based on the analytical properties of eigenvalues of an analytic and Hermitian matrix-valued function, which is applicable for any eigenvalue optimization problem whenever lower bounds on the second derivatives of the eigenvalue function can be calculated analytically or numerically. All the existing global eigenvalue optimization algorithms in the nonconvex setting are designed for specific problems, e.g., [3, 5, 6, 7, 15, 17, 19, 20, 21, 22, 32], while widely adopted techniques such as interior point methods [34]—when it is possible to pose an eigenvalue optimization problem as a semidefinite program—or a bundle method [31] are effective in the convex setting. We foresee nonconvex eigenvalue optimization problems that depend on a few parameters as the typical setting for the use of the algorithm here.

For the optimization of nonconvex eigenvalue functions, it appears essential to benefit from the global properties of eigenvalue functions, such as their global Lipschitzness or global bounds on their derivatives. Such global properties lead us to approximate $\lambda(\omega)$ globally with underestimating functions, which we call support functions. Furthermore, the derivatives of the eigenvalue functions can be evaluated effectively at no cost once the eigenvalue function is evaluated (due to analytic expressions for the derivatives of eigenvalues in terms of eigenvectors as discussed in section 3.2.1). Therefore, the incorporation of the derivatives into the support functions yields *quadratic support functions* on which our algorithm relies. The quadratic support functions for eigenvalue functions are derived exploiting the analytical properties of eigenvalues and presume the availability of a lower bound $\gamma$ on the second derivatives of the eigenvalue function that is obtained either analytically or numerically.

*Example.* Consider the minimization of the largest eigenvalue $\lambda_1(\omega) = \lambda_1(\mathcal{A}(\omega))$ of

$$\mathcal{A} : \mathbb{R} \to \mathbb{R}^{n\times n}, \quad \mathcal{A}(\omega) := A_0 + \omega A_1 + \omega^2 A_2,$$

where $A_0, A_1, A_2 \in \mathbb{R}^{n\times n}$ are given symmetric matrices. It can be deduced from the expressions in section 3.2.2 that $\lambda_1''(\omega) \geq \gamma := 2\lambda_{\min}(A_2)$ for all $\omega$ such that $\lambda_1(\omega)$ is simple. Furthermore, due to expressions in section 3.2.1, at all such $\omega$, we have $\lambda_1'(\omega) = v_1(\omega)^T(A_1 + 2\omega A_2)\, v_1(\omega)$, where $v_1(\omega)$ is a unit eigenvector associated with $\lambda_1(\omega)$. Consequently, it turns out that, for any $\omega_k \in \mathbb{R}$ where $\lambda_1(\omega_k)$ is simple, there is a support function

$$q(\omega) := \lambda_1(\omega_k) + \lambda_1'(\omega_k)(\omega - \omega_k) + \frac{\gamma}{2}(\omega - \omega_k)^2$$

satisfying $q(\omega) \leq \lambda_1(\omega)$ for all $\omega \in \mathbb{R}$; see section 5.2 for the details.

Support functions were earlier explored by the global optimization community. The Piyavskii–Shubert algorithm [40, 45] is derivative-free and constructs conic support functions based on Lipschitz continuity with a known global Lipschitz constant. It converges sublinearly in practice. Sophisticated variants that make use of several Lipschitz constants simultaneously appeared in the literature [24, 43]. The idea of using derivatives in the context of global optimization yields powerful algorithms. Breiman and Cutler [4] developed an algorithm that utilizes quadratic support functions depending on the derivatives. Some variants of the Breiman–Cutler algorithm

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

are also suggested for functions with Lipschitz-continuous derivatives; for instance [18, 26, 27] benefit from multiple Lipschitz constants for the derivatives, [42] estimates Lipschitz constants for the derivatives locally, while [29] modifies the support functions of the Breiman–Cutler algorithm in the univariate case so that the subproblems become smooth; however, all these variants in the multivariate case end up working on a mesh as a downside. The quadratic support functions that we derive for $\lambda(\omega)$ coincide with the quadratic support functions on which the Breiman–Cutler algorithm is built. Consequently, our approach is a variant of the algorithm due to Breiman and Cutler [4].

At every iteration of the algorithm, a global minimizer of a piecewise quadratic model defined as the maximum of a set of quadratic support functions is determined. A new quadratic support function is constructed around this global minimizer, and the piecewise quadratic model is refined with the addition of this new support function. In practice, we observe a linear rate of convergence to a global minimizer.

The algorithm appears applicable especially to extremal eigenvalue functions of the form

$$\lambda(\omega) = \sum_{k=1}^{j} d_k \lambda_k \left(\mathcal{A}(\omega)\right),$$

where $d_k$ are given real numbers such that $d_1 \geq d_2 \geq \cdots \geq d_j \geq 0$. This is facilitated by the simple quadratic support functions derived in section 5.2 and expressions for the lower bound $\gamma$ on the second derivatives derived in section 6. The algorithm is also applicable if the eigenvalue function $\lambda(\omega)$ is simple over all $\omega \in \mathcal{B}$, which holds for various eigenvalue optimization problems of interest.

**Outline.** We start in the next section with a list of eigenvalue optimization problems to which our proposed algorithm fits well. In section 3, the basic results concerning the analyticity and derivatives of the eigenvalues of a Hermitian matrix-valued function $\mathcal{A}(\omega)$ that depends analytically on $\omega$ are reviewed. In section 4, for a general eigenvalue function, the piecewise quadratic support functions that are defined as the minimum of $n$ quadratic functions are derived. In section 5, it is shown that these piecewise quadratic support functions simplify to smooth quadratic support functions for the extremal eigenvalue functions, as well as for the eigenvalue functions that are simple for all $\omega \in \mathcal{B}$. Global lower bounds $\gamma$ on the second derivatives of an extremal eigenvalue function are deduced in section 6. The algorithm based on the quadratic support functions is presented in section 7. We establish the global convergence of the proposed algorithm in section 8. Finally, comprehensive numerical experiments are provided in section 9. The examples indicate the superiority of the algorithm over the Lipschitz continuity based algorithms, e.g., [24, 40, 45], as well as the level-set based approaches devised for particular nonconvex eigenvalue optimization problems, e.g., [20, 32]. The reader who prefers to avoid technicalities at first could glance at the algorithm in section 7, then go through sections 3–6 for the theoretical foundation.

## 2. Applications.

**2.1. Quantities related to dynamical systems.** The *numerical radius* $r(A)$ of $A \in \mathbb{C}^{n\times n}$ is the modulus of the outermost point in its field of values [23] and is defined by

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

This quantity gives information about the powers of $A$, e.g.,$\|A^k\| \leq 2r(A)^k$, and is used in the literature to analyze the convergence of iterative methods for the solution of linear systems [1, 11]. An eigenvalue optimization characterization is given by [23]:

$$r(A) = -\Big[\min_{\theta\in[0,2\pi]} \lambda_n\left(\mathcal{A}(\theta)\right)\Big], \quad \mathcal{A}(\theta) := -(Ae^{i\theta} + A^*e^{-i\theta})/2.$$

The $\mathrm{H}_\infty$-norm is one of the most widely used norms in practice for the descriptor system

$$Ex'(t) = Ax(t) + Bu(t) \ \text{ and } \ y(t) = Cx(t) + Du(t),$$

where $u(t)$ and $y(t)$ are the input and output functions, respectively, and $E, A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}, C \in \mathbb{C}^{p\times n}, D \in \mathbb{C}^{p\times m}$ with $m, p \leq n$ are the system matrices. The $\mathrm{H}_\infty$-norm of the transfer function for this system is defined as

$$\|H\|_\infty := \frac{1}{\inf_{\omega\in\mathbb{R}} \ \sigma_n\left[H(i\omega)^\dagger\right]}, \quad H(s) := \left[C(sE - A)^{-1}B + D\right].$$

Here and elsewhere, $\sigma_j(\cdot)$ represents the $j$th largest singular value, and $H(i\omega)^\dagger$ denotes the pseudoinverse of $H(i\omega)$. Also above, with zero initial conditions for the descriptor system, the transfer function $H(s)$ reveals the linear relation between the input and output, as $Y(s) = H(s)U(s)$, with $U(s)$ and $Y(s)$ denoting the Laplace transformations of $u(t)$ and $y(t)$, respectively. Note that the $\mathrm{H}_\infty$-norm above is ill-posed (i.e., the associated operator is unbounded) if the pencil $L(\lambda) = A - \lambda E$ has an eigenvalue on the imaginary axis or to the right of the imaginary axis. Therefore, when the $\mathrm{H}_\infty$-norm is well-posed, the matrix-valued function $\mathcal{A}(\omega) := H(i\omega)$ is analytic at all $\omega \in \mathbb{R}$. A relevant quantity is the (continuous) distance to instability from a matrix $A \in \mathbb{C}^{n\times n}$; the eigenvalue optimization characterization for the $\mathrm{H}_\infty$-norm with $E = B = C = I_n$ and $D = 0$ reduces to that for the distance to instability [47] from $A$ with respect to the $\ell_2$-norm.

Paige [39] suggested the distance to uncontrollability, for a given $A \in \mathbb{C}^{n\times n}$ and $B \in \mathbb{C}^{n\times m}$ with $m \leq n$, defined by

$$\tau(A, B) := \inf\left\{\left\|\begin{bmatrix}\Delta A & \Delta B\end{bmatrix}\right\|_2 \mid (A + \Delta A, B + \Delta B) \ \text{ is uncontrollable}\right\},$$

as a robust measure of controllability. Here, the controllability of a linear control system $(A, B)$ of the form $x'(t) = Ax(t) + Bu(t)$ means that the function $x(t)$ can be driven into any state at a particular time by some input $u(t)$ and could be equivalently characterized as $\operatorname{rank}\left(\begin{bmatrix}A - zI & B\end{bmatrix}\right) = n$ for all $z \in \mathbb{C}$. Therefore, the eigenvalue optimization characterization for the distance to uncontrollability takes the form [12]

$$\tau(A, B) = \min_{z\in\mathbb{C}} \sigma_n\left(\mathcal{A}(z)\right), \quad \mathcal{A}(z) := \begin{bmatrix}A - zI & B\end{bmatrix}.$$

**2.2. Minimizing the largest or maximizing the smallest eigenvalues.** In the 18th century, Euler considered the design of the strongest column with a given volume with respect to the radii of the cross sections [30, 37]. The problem can be formulated as finding the parameters, representing the radii of cross sections, maximizing

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

the smallest eigenvalue of a fourth order differential operator. The analytical solution of the problem was considered in several studies in the 1970s and 1980s [2, 33, 35], which were motivated by the earlier work of Tadjbakhsh and Keller [46]. Later, the problem was treated numerically [8] by means of the finite-element discretization, giving rise to the problem

$$\min_{\omega\in\mathbb{R}^d} \lambda_1\left(\mathcal{A}(\omega)\right). \tag{2.1}$$

The treatment in [8] yields $\mathcal{A}(\omega) := A_0 + \sum_{j=1}^{d} \omega_j A_j$. In this affine setting, the minimization of the largest eigenvalue is a convex optimization problem (immediate from Theorem 6.1 below) and received considerable attention [14, 16, 36].

In the general setting, when the dependence of the matrix function $\mathcal{A}(\omega)$ on the parameters is not affine, the problem in (2.1) is nonconvex. Such nonconvex problems are significant (though they are not studied much, excluding a few studies such as [38] that offer only local analysis) in robust control theory, for instance, to ensure robust stability. The dual form that concerns the maximization of the smallest eigenvalue is of interest in vibroacoustics.

**2.3. Minimizing the sum of the $j$ largest eigenvalues.** In graph theory, relaxations of the NP-hard partitioning problems lead to eigenvalue optimization problems that require the minimization of the sum of the $j$ largest eigenvalues. For instance, given a weighted graph with $n$ vertices and nonnegative integers $d_1 \geq d_2 \geq \cdots \geq d_j$ summing up to $n$, consider finding a partitioning of the graph such that the $\ell$th partition contains exactly $d_\ell$ vertices for $\ell = 1, \ldots, j$ and the sum of the weights of the edges within each partition is maximized. The relaxation of this problem suggested in [10] is of the form

$$\min_{\omega\in\mathbb{R}^d} \sum_{k=1}^{j} d_k \lambda_k\left(\mathcal{A}(\omega)\right). \tag{2.2}$$

The problem (2.2) is convex if $\mathcal{A}(\omega)$ is an affine function of $\omega$, as in the case considered by [10]; see also [9].

Once again, in general, the minimization of the sum of the $j$ largest eigenvalues is not a convex optimization problem, and there are a few studies in the literature that attempted to analyze the problem locally, for instance, around the points where the eigenvalues coalesce [44].

**3. Background on perturbation theory of eigenvalues.** In this section, we first briefly summarize the analyticity results, mostly borrowed from [41, Chapter 1], related to the eigenvalues of matrix-valued functions. Then, expressions [28] are provided for the derivatives of Hermitian eigenvalues in terms of eigenvectors and the derivatives of matrix-valued functions. Finally, we elaborate on the analyticity of singular value problems as special Hermitian eigenvalue problems.

**3.1. Analyticity of eigenvalues.**

**3.1.1. Univariate matrix functions.** For a univariate matrix-valued function $\mathcal{A}(\omega)$ that depends on $\omega$ analytically, which may or may not be Hermitian, the characteristic polynomial is of the form

$$g(\omega, \lambda) := \det(\lambda I - \mathcal{A}(\omega)) = a_n(\omega)\lambda^n + \cdots + a_1(\omega)\lambda + a_0(\omega),$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where $a_0(\omega), \ldots, a_n(\omega)$ are analytic functions of $\omega$. It follows from the Puiseux theorem (see, e.g., [48, Chapter 2]) that each root $\tilde{\lambda}_j(\omega)$ such that $g(\omega, \tilde{\lambda}_j(\omega)) = 0$ has a Puiseux series of the form

$$\tilde{\lambda}_j(\omega) = \sum_{k=0}^{\infty} c_{k,j}\omega^{k/r} \tag{3.1}$$

for all small $\omega$, where $r$ is the multiplicity of the root $\tilde{\lambda}_j(0)$.

Now suppose $\mathcal{A}(\omega)$ is Hermitian for all $\omega$, and let $\ell$ be the smallest integer such that $c_{\ell,j} \neq 0$. Then, we have

$$\lim_{\omega \to 0^+} \frac{\tilde{\lambda}_j(\omega) - \tilde{\lambda}_j(0)}{\omega^{\ell/r}} = c_{\ell,j},$$

which implies that $c_{\ell,j}$ is real, since $\tilde{\lambda}_j(\omega)$ and $\omega^{\ell/r}$ are real numbers for each $\omega$. Furthermore,

$$\lim_{\omega \to 0^-} \frac{\tilde{\lambda}_j(\omega) - \tilde{\lambda}_j(0)}{(-\omega)^{\ell/r}} = (-1)^{\ell/r} c_{\ell,j}$$

is real, which implies that $(-1)^{\ell/r}$ is real, or equivalently that $\ell/r$ is integer. This observation reveals that the first nonzero term in the Puiseux series of $\tilde{\lambda}_j(\omega)$ is an integer power of $\omega$. The same argument applied to the derivatives of $\tilde{\lambda}_j(\omega)$ and the associated Puiseux series indicates that only integer powers of $\omega$ can appear in the Puiseux series (3.1), that is, the Puiseux series reduces to a power series. This establishes that $\tilde{\lambda}_j(\omega)$ is an analytic function of $\omega$. Indeed, it can also be deduced that, associated with $\tilde{\lambda}_1(\omega), \ldots, \tilde{\lambda}_n(\omega)$, there is an orthonormal set $\{v_1(\omega), \ldots, v_n(\omega)\}$ of eigenvectors, where each of $v_1(\omega), \ldots, v_n(\omega)$ varies analytically with respect to $\omega$ (see [41] for details).

THEOREM 3.1 (Rellich). *Let $\mathcal{A}(\omega) : \mathbb{R} \to \mathbb{C}^{n\times n}$ be a Hermitian matrix-valued function that depends on $\omega$ analytically.*

(i) *The $n$ roots of the characteristic polynomial of $\mathcal{A}(\omega)$ can be arranged so that each root $\tilde{\lambda}_j(\omega)$ for $j = 1, \ldots, n$ is an analytic function of $\omega$.*

(ii) *There exists an eigenvector $v_j(\omega)$ associated with $\tilde{\lambda}_j(\omega)$ for $j = 1, \ldots, n$ that satisfies the following:*
   (1) $(\tilde{\lambda}_j(\omega)I - \mathcal{A}(\omega))v_j(\omega) = 0$ for all $\omega \in \mathbb{R}$,
   (2) $\|v_j(\omega)\|_2 = 1$ for all $\omega \in \mathbb{R}$,
   (3) $v_j^*(\omega)v_k(\omega) = 0$ for all $\omega \in \mathbb{R}$ *for $k \neq j$, and*
   (4) $v_j(\omega)$ *is an analytic function of $\omega$.*

**3.1.2. Multivariate matrix functions.** The eigenvalues of a multivariate matrix-valued function $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times n}$ that depends on $\omega$ analytically do not have a power series representation in general even when $\mathcal{A}(\omega)$ is Hermitian. As an example, consider

$$\mathcal{A}(\omega) = \begin{bmatrix} \omega_1 & \frac{\omega_1+\omega_2}{2} \\ \frac{\omega_1+\omega_2}{2} & \omega_2 \end{bmatrix} \quad \text{with} \quad \tilde{\lambda}_{1,2}(\omega) = \frac{\omega_1+\omega_2}{2} \pm \sqrt{\frac{\omega_1^2+\omega_2^2}{2}}.$$

On the other hand, it follows from Theorem 3.1 that there are underlying eigenvalue functions $\tilde{\lambda}_j(\omega)$, $j = 1, \ldots, n$, of $\mathcal{A}(\omega)$, each of which is analytic along every line in

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$\mathbb{R}^d$, when $\mathcal{A}(\omega)$ is Hermitian. This analyticity property along lines in $\mathbb{R}^d$ implies the existence of the first partial derivatives of $\tilde{\lambda}_j(\omega)$ everywhere. Expressions for the first partial derivatives will be derived in the next subsection, indicating their continuity. As a consequence of the continuity of the first partial derivatives, each $\tilde{\lambda}_j(\omega)$ must be differentiable.

THEOREM 3.2. *Let $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times n}$ be a Hermitian matrix-valued function that depends on $\omega$ analytically. Then, the $n$ roots of the characteristic polynomial of $\mathcal{A}(\omega)$ can be arranged so that each root $\tilde{\lambda}_j(\omega)$ is* (i) *analytic on every line in $\mathbb{R}^d$, and* (ii) *differentiable on $\mathbb{R}^d$.*

### 3.2. Derivatives of eigenvalues.

**3.2.1. First derivatives of eigenvalues.** Consider a univariate Hermitian matrix-valued function $\mathcal{A}(\omega)$ that depends on $\omega$ analytically. An analytic eigenvalue $\tilde{\lambda}_j(\omega)$ and the associated eigenvector $v_j(\omega)$ as described in Theorem 3.1 satisfy

$$\mathcal{A}(\omega)v_j(\omega) = \tilde{\lambda}_j(\omega)v_j(\omega).$$

Taking the derivatives of both sides, we obtain

$$\frac{d\mathcal{A}(\omega)}{d\omega}v_j(\omega) + \mathcal{A}(\omega)\frac{dv_j(\omega)}{d\omega} = \frac{d\tilde{\lambda}_j(\omega)}{d\omega}v_j(\omega) + \tilde{\lambda}_j(\omega)\frac{dv_j(\omega)}{d\omega}. \tag{3.2}$$

Multiplying both sides by $v_j(\omega)^*$ and using the identities $v_j(\omega)^*\mathcal{A}(\omega) = v_j(\omega)^*\tilde{\lambda}_j(\omega)$ as well as $v_j(\omega)^*v_j(\omega) = \|v_j(\omega)\|_2^2 = 1$, we get

$$\frac{d\tilde{\lambda}_j(\omega)}{d\omega} = v_j(\omega)^*\frac{d\mathcal{A}(\omega)}{d\omega}v_j(\omega). \tag{3.3}$$

**3.2.2. Second derivatives of eigenvalues.** By differentiating both sides of (3.3), it is possible to deduce the formula (the details are omitted for brevity)

$$\frac{d^2\tilde{\lambda}_j(\omega)}{d\omega^2} = v_j(\omega)^*\frac{d^2\mathcal{A}(\omega)}{d\omega^2}v_j(\omega) + 2\sum_{k=1,k\neq j}^{n}\frac{1}{\tilde{\lambda}_j(\omega)-\tilde{\lambda}_k(\omega)}\left|v_k(\omega)^*\frac{d\mathcal{A}(\omega)}{d\omega}v_j(\omega)\right|^2 \tag{3.4}$$

for the second derivatives assuming that the (algebraic) multiplicity of $\tilde{\lambda}_j(\omega)$ is one.

If, on the other hand, the eigenvalues repeat at a given $\hat{\omega}$, specifically when the (algebraic) multiplicity of $\tilde{\lambda}_j(\hat{\omega})$ is greater than one, formula (3.4) generalizes as

$$\begin{aligned}\frac{d^2\tilde{\lambda}_j(\hat{\omega})}{d\omega^2} &= v_j(\hat{\omega})^*\frac{d^2\mathcal{A}(\hat{\omega})}{d\omega^2}v_j(\hat{\omega}) \\ &\quad + 2\sum_{k=1,k\neq j,k\notin\alpha}^{n}\lim_{\tilde{\omega}\to\hat{\omega}}\left(\frac{1}{\tilde{\lambda}_j(\tilde{\omega})-\tilde{\lambda}_k(\tilde{\omega})}\left|v_k(\tilde{\omega})^*\frac{d\mathcal{A}(\tilde{\omega})}{d\omega}v_j(\tilde{\omega})\right|^2\right).\end{aligned} \tag{3.5}$$

Here, $\alpha$ denotes the set of indices of the analytic eigenvalues (specified in Theorem 3.1) that are identical to $\tilde{\lambda}_j(\omega)$ at all $\omega$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

**3.2.3. Derivatives of eigenvalues for multivariate Hermitian matrix functions.** Let $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times n}$ be Hermitian and analytic. It follows from (3.3) that

$$\frac{\partial \tilde{\lambda}_j(\omega)}{\partial \omega_k} = v_j^*(\omega)\frac{\partial \mathcal{A}(\omega)}{\partial \omega_k} v_j(\omega). \tag{3.6}$$

Since $\mathcal{A}(\omega)$ and $v_j(\omega)$ are analytic with respect to $\omega_\ell$ for $\ell = 1, \ldots, n$, this implies the continuity, and the analyticity with respect to $\omega_\ell$, of each partial derivative $\partial\tilde{\lambda}_j(\omega)/\partial\omega_k$, and hence the existence of $\partial^2\tilde{\lambda}_j(\omega)/(\partial\omega_k\partial\omega_\ell)$, everywhere. If the multiplicity of $\tilde{\lambda}_j(\omega)$ is one, differentiating both sides of (3.6) with respect to $\omega_\ell$ would yield the following expressions for the second partial derivatives:

$$\begin{aligned}
&\frac{\partial^2 \tilde{\lambda}_j(\omega)}{\partial\omega_k\,\partial\omega_\ell} \\
&\quad = v_j^*(\omega)\frac{\partial^2 \mathcal{A}(\omega)}{\partial\omega_k\,\partial\omega_l} v_j(\omega) \\
&\qquad + 2\cdot\Re\left(\sum_{m=1,m\neq j}^{n} \frac{1}{\tilde{\lambda}_j(\omega)-\tilde{\lambda}_m(\omega)}\left(v_j(\omega)^*\frac{\partial \mathcal{A}(\omega)}{\partial\omega_k}v_m(\omega)\right)\left(v_m(\omega)^*\frac{\partial \mathcal{A}(\omega)}{\partial\omega_\ell}v_j(\omega)\right)\right).
\end{aligned}$$

Expressions similar to (3.5) can be obtained for the second partial derivatives when $\tilde{\lambda}_j(\omega)$ has multiplicity greater than one.

**3.3. Analyticity of singular values.** Some of the applications (see section 2.1) concern the optimization of the $j$th largest singular value of an analytic matrix-valued function. The singular value problems are special Hermitian eigenvalue problems. In particular, denoting the $j$th largest singular value of an analytic matrix-valued function $\mathcal{B}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times m}$ (not necessarily Hermitian) by $\sigma_j(\omega)$, the set of eigenvalues of the Hermitian matrix-valued function

$$\mathcal{A}(\omega) := \begin{bmatrix} 0 & \mathcal{B}(\omega) \\ \mathcal{B}(\omega)^* & 0 \end{bmatrix}$$

is $\{\sigma_j(\omega), -\sigma_j(\omega) : j = 1, \ldots, n\}$. In the univariate case $\sigma_j(\omega)$ is the $j$th largest of the $2n$ analytic eigenvalues, $\tilde{\lambda}_1(\omega), \ldots, \tilde{\lambda}_{2n}(\omega)$, of $\mathcal{A}(\omega)$. The multivariate $d$-dimensional case is similar, with the exception that each eigenvalue $\tilde{\lambda}_j(\omega)$ is differentiable and analytic along every line in $\mathbb{R}^d$. Let us focus on the univariate case throughout the rest of this section. Extensions to the multivariate case are similar to the previous sections. Suppose $v_j(\omega) := \left[\begin{smallmatrix} u_j(\omega) \\ w_j(\omega)\end{smallmatrix}\right]$, with $u_j(\omega) \in \mathbb{C}^n$, $w_j(\omega) \in \mathbb{C}^m$, is the analytic eigenvector function as specified in Theorem 3.1 of $\mathcal{A}(\omega)$ associated with $\tilde{\lambda}_j(\omega)$, that is,

$$\begin{bmatrix} 0 & \mathcal{B}(\omega) \\ \mathcal{B}(\omega)^* & 0 \end{bmatrix}\begin{bmatrix} u_j(\omega) \\ w_j(\omega)\end{bmatrix} = \tilde{\lambda}_j(\omega)\begin{bmatrix} u_j(\omega) \\ w_j(\omega)\end{bmatrix}.$$

The above equation implies

$$\mathcal{B}(\omega)w_j(\omega) = \tilde{\lambda}_j(\omega)u_j(\omega) \quad \text{and} \quad \mathcal{B}(\omega)^*u_j(\omega) = \tilde{\lambda}_j(\omega)w_j(\omega). \tag{3.7}$$

In other words, $u_j(\omega)$, $w_j(\omega)$ are analytic and consist of a pair of consistent left and right singular vectors associated with $\tilde{\lambda}_j(\omega)$. To summarize, in the univariate case,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$\tilde{\lambda}_j(\omega)$ can be considered as a signed analytic singular value of $\mathcal{B}(\omega)$, and there is a consistent pair of analytic left and right singular vector functions, $u_j(\omega)$ and $w_j(\omega)$, respectively.

Next, in the univariate case, we derive expressions for the first derivative of $\tilde{\lambda}_j(\omega)$, in terms of the corresponding left and right singular vectors. It follows from the singular value equations (3.7) above that $\|u_j(\omega)\| = \|w_j(\omega)\| = 1/\sqrt{2}$. (If $\tilde{\lambda}_j(\omega) = 0$, this equality follows from analyticity.) Now, the application of the expression (3.3) yields

$$\begin{aligned}\frac{d\tilde{\lambda}_j(\omega)}{d\omega} &= \begin{bmatrix} u_j(\omega)^* & w_j(\omega)^* \end{bmatrix} \begin{bmatrix} 0 & d\mathcal{B}(\omega)/d\omega \\ d\mathcal{B}(\omega)^*/d\omega & 0 \end{bmatrix} \begin{bmatrix} u_j(\omega) \\ w_j(\omega) \end{bmatrix} \\ &= u_j(\omega)^* \frac{d\mathcal{B}(\omega)}{d\omega} w_j(\omega) + w_j(\omega)^* \frac{d\mathcal{B}(\omega)^*}{d\omega} u_j(\omega) \\ &= 2 \cdot \Re \left( u_j(\omega)^* \frac{d\mathcal{B}(\omega)}{d\omega} w_j(\omega) \right).\end{aligned}$$

In terms of the unit left and right singular vectors $\hat{u}_j(\omega) := \sqrt{2} \cdot u_j(\omega)$ and $\hat{w}_j(\omega) := \sqrt{2} \cdot w_j(\omega)$, respectively, associated with $\tilde{\lambda}_j(\omega)$, we obtain

$$\frac{d\tilde{\lambda}_j(\omega)}{d\omega} = \Re \left( \hat{u}_j(\omega)^* \frac{d\mathcal{B}(\omega)}{d\omega} \hat{w}_j(\omega) \right). \tag{3.8}$$

**Notation.** Throughout the rest of the text, we denote the eigenvalues of $\mathcal{A}(\omega)$ that are analytic in the univariate case (stated in Theorem 3.1) and differentiable and analytic along every line in the multivariate case (stated in Theorem 3.2) with $\tilde{\lambda}_1(\omega), \ldots, \tilde{\lambda}_n(\omega)$. On the other hand, $\lambda_j(\omega)$ or $\lambda_j(\mathcal{A}(\omega))$ denotes the $j$th largest eigenvalue, and $\sigma_j(\omega)$ or $\sigma_j(\mathcal{A}(\omega))$ denotes the $j$th largest singular value of $\mathcal{A}(\omega)$.

**4. Piecewise quadratic support functions.** Let $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n : \mathbb{R}^d \to \mathbb{R}$ be eigenvalue functions of a Hermitian matrix-valued function $\mathcal{A}(\omega)$ that are analytic along every line in $\mathbb{R}^d$ and differentiable on $\mathbb{R}^d$, and let $\mathcal{B} \subset \mathbb{R}^d$ be the box defined by

$$\mathcal{B} := \mathcal{B}\big(\omega_1^{(l)}, \omega_1^{(u)}, \ldots, \omega_d^{(l)}, \omega_d^{(u)}\big) := \big\{ \omega \in \mathbb{R}^d \,\big|\, \omega_j \in \big[\omega_j^{(l)}, \omega_j^{(u)}\big] \text{ for } j = 1, \ldots, d \big\}. \tag{4.1}$$

Consider the closed and connected subsets $\mathcal{P}_1, \ldots, \mathcal{P}_q$ of $\mathcal{B}$, with $q$ as small as possible, such that $\cup_{k=1}^q \mathcal{P}_k = \mathcal{B}$, and $\mathcal{P}_k \cap \mathcal{P}_\ell = \partial\mathcal{P}_k \cap \partial\mathcal{P}_\ell$ for each $k$ and $\ell$, and such that in the interior of $\mathcal{P}_k$ none of the eigenvalue functions $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ intersects another. Define $\lambda : \mathcal{B} \to \mathbb{R}$ as follows:

$$\lambda(\omega) := f\big(\tilde{\lambda}_{s_{k1}}(\omega), \ldots, \tilde{\lambda}_{s_{kj}}(\omega)\big) \text{ for all } \omega \in \operatorname{int}(\mathcal{P}_k), \tag{4.2}$$

where $f$ is analytic, and $s_k = [s_{k1} \quad \ldots \quad s_{kj}]^T \in \mathbb{Z}_+^j$ is a vector of indices such that

$$\tilde{\lambda}_{s_{ki}}(\omega) = \tilde{\lambda}_{s_{\ell i}}(\omega) \text{ for } i = 1, \ldots, j,$$

and for all $\omega \in \partial\mathcal{P}_k \cap \partial\mathcal{P}_\ell$ in order to ensure the continuity of $\lambda(\omega)$ on $\mathcal{B}$. The extremal eigenvalue function $\lambda(\omega) = \sum_{k=1}^j d_k \lambda_k(\omega)$ fits into the framework.

We derive a piecewise quadratic support function $q_k(\omega)$ about a given point $\omega_k \in \mathcal{B}$ bounding $\lambda(\omega)$ from below for all $\omega \in \mathcal{B}$ and such that $q_k(\omega_k) = \lambda(\omega_k)$. Let us focus

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

on the direction $p := (\omega - \omega_k)/\|\omega - \omega_k\|$, the univariate function $\phi(\alpha) := \lambda(\omega_k + \alpha p)$, and the analytic univariate functions $\tilde{\phi}_j(\alpha) := \tilde{\lambda}_j(\omega_k + \alpha p)$ for $j = 1, \ldots, n$. Also, let us denote the isolated points in the interval $[0, \|\omega - \omega_k\|]$, where two distinct functions among $\tilde{\phi}_1(\alpha), \ldots, \tilde{\phi}_n(\alpha)$ intersect each other by $\alpha^{(1)}, \ldots, \alpha^{(m)}$. At these points, $\phi(\alpha)$ may not be differentiable. We have

$$\lambda(\omega) = \lambda(\omega_k) + \sum_{\ell=0}^{m} \int_{\alpha^{(\ell)}}^{\alpha^{(\ell+1)}} \phi'(t)dt, \tag{4.3}$$

where $\alpha^{(0)} := 0$ and $\alpha^{(m+1)} := \|\omega - \omega_k\|$. Due to the existence of the second partial derivatives of $\tilde{\lambda}_j(\omega)$ (since the expression (3.6) implies the analyticity of the first partial derivatives with respect to each parameter disjointly), there exists a constant $\gamma$ that satisfies

$$\lambda_{\min}\big(\nabla^2 \tilde{\lambda}_j(\omega)\big) \geq \gamma \quad \text{for all } \omega \in \mathcal{B} \quad j = 1, \ldots, n. \tag{4.4}$$

Furthermore, $\tilde{\phi}_j''(\alpha) = p^T \nabla^2 \tilde{\lambda}_j(\omega_k + \alpha p)\ p \geq \lambda_{\min}\big(\nabla^2 \tilde{\lambda}_j(\omega_k + \alpha p)\big) \geq \gamma$ for all $\alpha \in [0, \|\omega - \omega_k\|]$. Thus, applying the mean value theorem to the analytic functions $\tilde{\phi}_j'(\alpha)$ for $j = 1, \ldots, n$ and since $\phi'(t) \geq \min_{j=1,\ldots,n} \tilde{\phi}_j'(t)$, we obtain

$$\phi'(t) \geq \min_{j=1,\ldots,n} \tilde{\phi}_j'(0) + \gamma t.$$

By substituting the last inequality in (4.3), integrating the right-hand side of (4.3), and using $\tilde{\phi}_j'(0) = \nabla \tilde{\lambda}_j(\omega_k)^T p$ (since $\tilde{\lambda}_j$ is differentiable), we arrive at the following.

THEOREM 4.1. *Suppose $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n \times n}$ is an analytic and Hermitian matrix-valued function, the eigenvalue function $\lambda(\omega)$ is defined as in* (4.2) *in terms of the eigenvalues $\tilde{\lambda}_1(\omega), \ldots, \tilde{\lambda}_n(\omega)$ of $\mathcal{A}(\omega)$ that are differentiable and analytic on every line in $\mathbb{R}^d$, and $\gamma$ is a lower bound as in* (4.4). *Then the following inequality holds for all $\omega \in \mathcal{B}$:*

$$\lambda(\omega) \geq \left[ q_k(\omega) := \lambda(\omega_k) + \big(\min_{j=1,\ldots,n} \nabla \tilde{\lambda}_j(\omega_k)^T (\omega - \omega_k)\big) + \frac{\gamma}{2}\|\omega - \omega_k\|^2 \right]. \tag{4.5}$$

## 5. Simplified piecewise quadratic support functions.

**5.1. Support functions under generic simplicity.** In various instances, the eigenvalue functions $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ do not intersect each other at any $\omega \in \mathcal{B}$ generically. In such cases, for some $j$, we have $\lambda(\omega) = \tilde{\lambda}_j(\omega)$ for all $\omega \in \mathcal{B}$; therefore $\lambda(\omega)$ is analytic in the univariate case and analytic along every line in the multivariate case. For instance, the singular values of the matrix function $\mathcal{A}(\omega) := C(\omega i I - A)^{-1} B + D$ involved in the definition of the $\mathrm{H}_\infty$-norm do not coalesce at any $\omega \in \mathbb{R}$ on a dense subset of the set of quadruples $(A, B, C, D)$. Similar remarks apply to all the specific eigenvalue optimization problems in section 2.1.

Under the generic simplicity assumption, the piecewise quadratic support function (4.5) simplifies to

$$q_k(\omega) = \lambda(\omega_k) + \nabla \lambda(\omega_k)^T (\omega - \omega_k) + \frac{\gamma}{2}\|\omega - \omega_k\|^2. \tag{5.1}$$

Here, $\gamma$ is a lower bound on $\lambda_{\min}\left(\nabla^2 \lambda(\omega)\right)$ for all $\omega \in \mathcal{B}$. In many cases, it may be possible to obtain a rough lower bound $\gamma$ numerically by means of the expressions for the second derivatives in sections 3.2.2 and 3.2.3 and by exploiting the Lipschitz continuity of the eigenvalue $\lambda(\omega)$ and other eigenvalues.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

**5.2. Support functions for extremal eigenvalues.** Consider the extremal eigenvalue function

$$\lambda(\omega) = \sum_{k=1}^{j} d_k \lambda_k(\omega) \tag{5.2}$$

for given real numbers $d_1 \geq d_2 \geq \cdots \geq d_j \geq 0$. A special case when $d_1, d_2, \ldots d_j$ are integers is discussed in section 2.3. When $d_1 = 1, d_2 = \cdots = d_j = 0$, this reduces to the maximal eigenvalue function $\lambda(\omega) = \lambda_1(\omega)$ in section 2.2.

For simplicity, let us suppose that $\lambda(\omega)$ is differentiable at $\omega_k$, about which we derive a support function below. This is generically the case. In the unlikely case of two eigenvalues coalescing at $\omega_k$, the nondifferentiability is isolated at this point. Therefore, $\lambda(\omega)$ is differentiable at all nearby points. For a fixed $\omega \in \mathcal{B}$, as in the previous section, define $\phi : \mathbb{R} \to \mathbb{R}$, $\phi(\alpha) := \lambda(\omega_k + \alpha p)$ for $p = (\omega - \omega_k)/\|\omega - \omega_k\|$. Denote the points $\alpha \in (0, \|\omega - \omega_k\|]$, where either one of $\lambda_1(\omega_k + \alpha p), \ldots, \lambda_j(\omega_k + \alpha p)$ is not simple by $\alpha^{(1)}, \ldots, \alpha^{(m)}$ in increasing order. These are the points where $\phi(\alpha)$ is possibly not analytic. At any $\alpha \in (0, \|\omega - \omega_k\|)$, by $\phi_{+,\alpha}$ we refer to the analytic function satisfying $\phi_{+,\alpha}(\tilde{\alpha}) = \phi(\tilde{\alpha})$ for all $\tilde{\alpha} > \alpha$ sufficiently close to $\alpha$. Similarly, $\phi_{-,\alpha}$ refers to the analytic function satisfying $\phi_{-,\alpha}(\tilde{\alpha}) = \phi(\tilde{\alpha})$ for all $\tilde{\alpha} < \alpha$ sufficiently close to $\alpha$. Furthermore, $\phi'_+(\alpha)$ and $\phi'_-(\alpha)$ represent the right-hand and left-hand derivatives of $\phi(\alpha)$, respectively.

LEMMA 5.1. *The following relation holds for all* $\alpha \in (0, \|\omega - \omega_k\|)$*:* (i) $\phi_{+,\alpha}(\tilde{\alpha}) \geq \phi_{-,\alpha}(\tilde{\alpha})$ *for all* $\tilde{\alpha} > \alpha$ *sufficiently close to* $\alpha$*;* (ii) *consequently* $\phi'_+(\alpha) \geq \phi'_-(\alpha)$.

*Proof.* The functions $\phi_{-,\alpha}(\tilde{\alpha})$ and $\phi_{+,\alpha}(\tilde{\alpha})$ are of the form

$$\phi_{-,\alpha}(\tilde{\alpha}) = \sum_{k=1}^{j} d_k \tilde{\lambda}_{n_k}(\omega_k + \tilde{\alpha} p) \quad \text{and} \quad \phi_{+,\alpha}(\tilde{\alpha}) = \sum_{k=1}^{j} d_k \lambda_k(\omega_k + \tilde{\alpha} p) \tag{5.3}$$

for some indices $n_1, \ldots, n_j$ and for all $\tilde{\alpha} > \alpha$ sufficiently close to $\alpha$. In (5.3), the latter equality follows from $\phi_{+,\alpha}(\tilde{\alpha}) = \phi(\tilde{\alpha}) = \lambda(\omega_k + \tilde{\alpha} p)$ for all $\tilde{\alpha} > \alpha$ sufficiently close to $\alpha$ by definition. The former equality is due to $\phi_{-,\alpha}(\tilde{\alpha}) = \phi(\tilde{\alpha}) = \lambda(\omega_k + \tilde{\alpha} p)$ for all $\tilde{\alpha} < \alpha$ sufficiently close to $\alpha$ implying $\phi_{-,\alpha}(\tilde{\alpha})$ is a weighted sum of $j$ of the analytic eigenvalues $\tilde{\lambda}_1(\omega_k + \tilde{\alpha} p), \ldots, \tilde{\lambda}_n(\omega_k + \tilde{\alpha} p)$ with weights $d_1, \ldots, d_j$. We rephrase the inequality $\phi_{+,\alpha}(\tilde{\alpha}) \geq \phi_{-,\alpha}(\tilde{\alpha})$ as $\phi_{+,\alpha}(\tilde{\alpha}) - \phi_{-,\alpha}(\tilde{\alpha}) = \sum_{k=1}^{j} d_k \cdot a_k \geq 0$, where $a_k = \lambda_k(\omega_k + \tilde{\alpha} p) - \tilde{\lambda}_{n_k}(\omega_k + \tilde{\alpha} p)$, where $k = 1, \ldots, j$.

Note that $\sum_{k=1}^{q} a_k \geq 0$ for each $q = 1, \ldots, j$ since $\lambda_1(\omega_k + \tilde{\alpha} p), \ldots, \lambda_q(\omega_k + \tilde{\alpha} p)$ are the largest $q$ eigenvalues. In particular, their sum cannot be less than the sum of $\tilde{\lambda}_{n_1}(\omega_k + \tilde{\alpha} p), \ldots, \tilde{\lambda}_{n_q}(\omega_k + \tilde{\alpha} p)$. Therefore,

$$\begin{aligned} \sum_{k=1}^{j} d_k \cdot a_k &= d_j \sum_{k=1}^{j} a_k + (d_{j-1} - d_j) \sum_{k=1}^{j-1} a_k \\ &\quad + (d_{j-2} - d_{j-1}) \sum_{k=1}^{j-2} a_k + \cdots + (d_1 - d_2) a_1 \\ &\geq 0 \end{aligned}$$

since $d_1 \geq d_2 \geq \cdots \geq d_j \geq 0$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Part (ii) is immediate from part (i) due to $\phi'_+(\alpha) = \phi'_{+,\alpha}(\alpha) \geq \phi'_{-,\alpha}(\alpha) = \phi'_-(\alpha)$, where the inequality follows from an application of the Taylor's theorem to $\phi_{+,\alpha}(\tilde{\alpha})$ and $\phi_{-,\alpha}(\tilde{\alpha})$ for $\tilde{\alpha} > \alpha$ and around $\alpha$. ∎

THEOREM 5.2. *Let $\omega_k \in \mathbb{R}^d$ be such that all the eigenvalues $\lambda_1(\omega_k), \ldots, \lambda_j(\omega_k)$ are simple, and let $\gamma$ satisfy $\lambda_{\min}\left(\nabla^2\lambda(\omega)\right) \geq \gamma$ for all $\omega \in \mathcal{B}$ such that $\lambda(\omega)$ is simple. Then, the following inequality holds for the extreme eigenvalue function $\lambda(\omega)$ given by* (5.2) *and for all $\omega \in \mathcal{B}$:*

$$\lambda(\omega) \geq q_k(\omega) := \lambda(\omega_k) + \nabla\lambda(\omega_k)^T(\omega - \omega_k) + \frac{\gamma}{2}\|\omega - \omega_k\|^2.$$

*Proof.* It follows from the Taylor's theorem that for each $k = 0, \ldots, m$ and $\alpha$ that belongs to the closure of $(\alpha^{(k)}, \alpha^{(k+1)})$ (here we define $\alpha^{(0)} = 0$ and $\alpha^{(m+1)} = \|\omega - \omega_k\|$), we have

$$\phi(\alpha) = \phi(\alpha^{(k)}) + \phi'_+(\alpha^{(k)})(\alpha - \alpha^{(k)}) + \frac{\phi''(\eta)}{2}(\alpha - \alpha^{(k)})^2,$$

where $\eta \in (\alpha^{(k)}, \alpha)$. By observing $\phi''(\eta) = p^T\nabla^2\lambda(\omega_k + \eta p)p \geq \lambda_{\min}[\nabla^2(\lambda(\omega_k + \eta p))] \geq \gamma$, we deduce

$$\phi(\alpha) \geq \phi(\alpha^{(k)}) + \phi'_+(\alpha^{(k)})(\alpha - \alpha^{(k)}) + \frac{\gamma}{2}(\alpha - \alpha^{(k)})^2, \tag{5.4}$$

and by differentiating we also deduce

$$\phi'_-(\alpha) \geq \phi'_+(\alpha^{(k)}) + \gamma(\alpha - \alpha^{(k)}). \tag{5.5}$$

Next, we prove the inequalities

$$\phi(\alpha^{(\ell)}) \geq \phi(\alpha^{(0)}) + \phi'_+(\alpha^{(0)})(\alpha^{(\ell)} - \alpha^{(0)}) + \frac{\gamma}{2}(\alpha^{(\ell)} - \alpha^{(0)})^2 \tag{5.6}$$

$$\phi'_-(\alpha^{(\ell)}) \geq \phi'_+(\alpha^{(0)}) + \gamma(\alpha^{(\ell)} - \alpha^{(0)}) \tag{5.7}$$

for each $\ell = 1, 2, \ldots, m+1$ by induction. The inequalities (5.6) and (5.7) for $\ell = 1$ hold by applications of (5.4) and (5.5) with $\alpha = \alpha^{(1)}, \alpha^{(k)} = \alpha^{(0)}$. Let us suppose that the inequalities indeed hold for $\ell = 2, \ldots, k$ as the inductive hypothesis. Now, by another application of (5.5) and since $\phi'_+(\alpha^{(k)}) \geq \phi'_-(\alpha^{(k)})$ (see Lemma 5.1), we obtain

$$\begin{aligned}
\phi'_-(\alpha^{(k+1)}) &\geq \phi'_+(\alpha^{(k)}) + \gamma(\alpha^{(k+1)} - \alpha^{(k)}) \\
&\geq \phi'_-(\alpha^{(k)}) + \gamma(\alpha^{(k+1)} - \alpha^{(k)}) \\
&\geq \left[\phi'_+(\alpha^{(0)}) + \gamma(\alpha^{(k)} - \alpha^{(0)})\right] + \gamma(\alpha^{(k+1)} - \alpha^{(k)}) \\
&= \phi'_+(\alpha^{(0)}) + \gamma(\alpha^{(k+1)} - \alpha^{(0)}),
\end{aligned}$$

where we use the inductive hypothesis with $\ell = k$ in the third inequality. Furthermore, by (5.4), the inequality $\phi'_+(\alpha^{(k)}) \geq \phi'_-(\alpha^{(k)})$, and exploiting the inductive hypothesis with $\ell = k$, we end up with

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$\begin{aligned}\phi(\alpha^{(k+1)}) &\geq \phi(\alpha^{(k)}) + \phi'_+(\alpha^{(k)})(\alpha^{(k+1)} - \alpha^{(k)}) + \frac{\gamma}{2}(\alpha^{(k+1)} - \alpha^{(k)})^2 \\ &\geq \phi(\alpha^{(k)}) + \phi'_-(\alpha^{(k)})(\alpha^{(k+1)} - \alpha^{(k)}) + \frac{\gamma}{2}(\alpha^{(k+1)} - \alpha^{(k)})^2 \\ &= \left[\phi(\alpha^{(0)}) + \phi'_+(\alpha^{(0)})(\alpha^{(k)} - \alpha^{(0)}) + \frac{\gamma}{2}(\alpha^{(k)} - \alpha^{(0)})^2\right] \\ &\quad + \left[\phi'_+(\alpha^{(0)}) + \gamma(\alpha^{(k)} - \alpha^{(0)})\right](\alpha^{(k+1)} - \alpha^{(k)}) + \frac{\gamma}{2}(\alpha^{(k+1)} - \alpha^{(k)})^2 \\ &= \phi(\alpha^{(0)}) + \phi'_+(\alpha^{(0)})(\alpha^{(k+1)} - \alpha^{(0)}) + \frac{\gamma}{2}(\alpha^{(k+1)} - \alpha^{(0)})^2,\end{aligned}$$

proving the validity of (5.6) and (5.7) for each $\ell = 1, \ldots, m+1$.

The inequality (5.6) with $\ell = m+1$ yields

$$\phi(\alpha^{(m+1)}) \geq \phi(\alpha^{(0)}) + \phi'_+(\alpha^{(0)})(\alpha^{(m+1)} - \alpha^{(0)}) + \frac{\gamma}{2}(\alpha^{(m+1)} - \alpha^{(0)})^2,$$

from which the result follows by noting $\phi(\alpha^{(m+1)}) = \lambda(\omega)$, $\phi(\alpha^{(0)}) = \lambda(\omega_k)$, $\phi'_+(\alpha^{(0)}) = \nabla\lambda(\omega_k)^T p$, and $(\alpha^{(m+1)} - \alpha^{(0)}) = \|\omega - \omega_k\|$. ∎

A lower bound $\gamma$ as stated in Theorem 5.2 can be deduced analytically in various cases. This is discussed next.

**6. Lower bound $\gamma$ for the extremal eigenvalue functions.** The quadratic support functions discussed so far rely on the lower bound $\gamma$. Such bounds can be derived analytically for the extremal eigenvalue function (5.2) with $d_1 \geq d_2 \geq \cdots \geq d_j \geq 0$ in various cases.

THEOREM 6.1. *Suppose $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times n}$ is analytic and Hermitian at all $\omega \in \mathbb{R}^d$. Then, $\lambda(\omega)$ defined as in* (5.2) *satisfies*

$$\lambda_{\min}\left(\nabla^2\lambda(\omega)\right) \geq \left(\sum_{k=1}^{j} d_k\right) \cdot \lambda_{\min}\left(\nabla^2\mathcal{A}(\omega)\right)$$

*for each $\omega$ such that $\lambda_1(\omega), \ldots, \lambda_j(\omega)$ are simple, where $\nabla^2\mathcal{A}(\omega) \in \mathbb{C}^{nd\times nd}$ is given by*

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

*Proof.* First observe that, by the formulas in section 3.2.3, we obtain

$$\begin{aligned}&\frac{\partial^2\lambda(\omega)}{\partial\omega_\ell\,\partial\omega_i} \\ &= \sum_{k=1}^{j} d_k \cdot \left(v_k(\omega)^* \frac{\partial^2\mathcal{A}(\omega)}{\partial\omega_\ell\,\partial\omega_i} v_k(\omega)\right) \\ &\quad + 2\cdot\Re\left[\sum_{k=1}^{j}\left(\sum_{m=k+1}^{n} \frac{d_k - d_m}{\lambda_k(\omega) - \lambda_m(\omega)}\left(v_k(\omega)^* \frac{\partial\mathcal{A}(\omega)}{\partial\omega_\ell} v_m(\omega)\right)\left(v_m(\omega)^* \frac{\partial\mathcal{A}(\omega)}{\partial\omega_i} v_k(\omega)\right)\right)\right],\end{aligned}$$

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where we define $d_{j+1} = \cdots = d_n = 0$. As for the Hessian, this yields

$$\nabla^2\lambda(\omega) = \sum_{k=1}^{j} d_k \mathcal{H}^{(k)}(\omega) + 2 \cdot \sum_{k=1}^{j} \sum_{m=k+1}^{n} \frac{d_k - d_m}{\lambda_k(\omega) - \lambda_m(\omega)} \Re\big(\mathcal{H}^{(k,m)}(\omega)\big), \tag{6.1}$$

where $\mathcal{H}^{(k)}(\omega), \mathcal{H}^{(k,m)}(\omega) \in \mathbb{R}^{d\times d}$ are such that the $(\ell, i)$ entries of $\mathcal{H}^{(k)}(\omega)$ and $\mathcal{H}^{(k,m)}(\omega)$ are given by

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

respectively. Furthermore, it can easily be verified that $\mathcal{H}^{(k,m)}(\omega)$ is positive semidefinite for each $k$ and $m$. Indeed, denoting $h_i^{(k,m)} := v_m(\omega)^*(\partial\mathcal{A}(\omega)/\partial\omega_i)v_k(\omega)$, for each $u \in \mathbb{C}^d$, observe that

$$u^T \mathcal{H}^{(k,m)}(\omega) u = \left|\sum_{i=1}^{d} u_i h_i^{(k,m)}\right|^2 \geq 0.$$

This in turn implies that $\Re(\mathcal{H}^{(k,m)}(\omega))$ is positive semidefinite, i.e., since $\mathcal{H}^{(k,m)}(\omega) = \Re(\mathcal{H}^{(k,m)}(\omega)) + i\Im(\mathcal{H}^{(k,m)}(\omega))$, for all $u \in \mathbb{R}^d$ we have $u^T\Re(\mathcal{H}^{(k,m)}(\omega))u = u^T\mathcal{H}^{(k,m)}(\omega)u \geq 0$. Consequently, it follows from (6.1) that

$$\begin{aligned}\lambda_{\min}\left(\nabla^2\lambda(\omega)\right) \geq \lambda_{\min}\left(\sum_{k=1}^{j} d_k \mathcal{H}^{(k)}(\omega)\right) &\geq \sum_{k=1}^{j} d_k \lambda_{\min}\big(\mathcal{H}^{(k)}(\omega)\big) \\ &\geq \sum_{k=1}^{j} d_k \lambda_{\min}\left(\nabla^2\mathcal{A}(\omega)\right).\end{aligned}$$

To see the last inequality, note that $\mathcal{H}^{(k)}(\omega) = [I_d \otimes v_k(\omega)^*] \cdot \nabla^2\mathcal{A}(\omega) \cdot [I_d \otimes v_k(\omega)]$, where $\otimes$ denotes the Kronecker product, and therefore for some $v(\omega) \in \mathbb{C}^d$ of unit length

$$\begin{aligned}\lambda_{\min}\big(\mathcal{H}^{(k)}(\omega)\big) = v(\omega)^* \cdot \mathcal{H}^{(k)}(\omega) \cdot v(\omega) &= [v(\omega)^* \otimes v_k(\omega)^*] \cdot \nabla^2\mathcal{A}(\omega) \cdot [v(\omega) \otimes v_k(\omega)] \\ &\geq \lambda_{\min}\left(\nabla^2\mathcal{A}(\omega)\right). \qquad \square\end{aligned}$$

There are various instances when the theorem above reveals an immediate lower bound $\gamma$ on $\lambda_{\min}\left(\nabla^2\lambda(\omega)\right)$ over all $\omega \in \mathcal{B}$. A particular instance is given by the corollary below when $\mathcal{A}(\omega)$ is a quadratic function of $\omega$.

COROLLARY 6.2. *Suppose that $\mathcal{A}(\omega) : \mathbb{R}^d \to \mathbb{C}^{n\times n}$ is of the form*

$$\mathcal{A}(\omega) = A_0 + \sum_{\ell=1}^{d} \omega_\ell A_\ell + \frac{1}{2}\sum_{\ell=1}^{d}\sum_{i=1}^{d} \omega_\ell\, \omega_i A_{\ell i},$$

*where $A_\ell$ and $A_{\ell i}$ are Hermitian and $A_{\ell i} = A_{i\ell}$. Then $\lambda(\omega)$ defined as in* (5.2) *satisfies*

$$\lambda_{\min}\left(\nabla^2\lambda(\omega)\right) \geq \left(\sum_{k=1}^{j} d_k\right) \cdot \lambda_{\min}\left(\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1d} \\ A_{21} & A_{22} & \dots & A_{2d} \\ & & \ddots & \\ A_{d1} & A_{d2} & \dots & A_{dd} \end{bmatrix}\right)$$

*for each $\omega$ such that $\lambda_1(\omega), \dots, \lambda_j(\omega)$ are simple.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

**7. The algorithm.** We consider the minimization of an eigenvalue function $\lambda : \mathcal{B} \to \mathbb{R}$ formally defined in section 4 over a box $\mathcal{B} \subset \mathbb{R}^d$. We will utilize the quadratic function given by (5.1) under the assumption that this is indeed a support function. This is certainly the case for $\lambda(\omega) = \sum_{k=1}^{j} d_k \lambda_k(\omega)$ with weights $d_k$ such that $d_1 \geq d_2 \geq \cdots \geq d_j \geq 0$ as discussed in section 5.2 (specifically see Theorem 5.2), and generically the case for all eigenvalue or singular value functions that arise from various applications in section 2.1, since these eigenvalue and singular value functions are simple at all $\omega \in \mathcal{B}$ generically, as discussed in section 5.1. Thus the algorithm is applicable to all the problems in section 2.

The algorithm starts by constructing a support function $q_0$ about an arbitrary point $\omega_0 \in \mathcal{B}$. The next step is to find the global minimizer $\omega_1$ of $q_0$ in $\mathcal{B}$ and to construct the quadratic support function $q_1$ about $\omega_1$. In general, suppose there are $s+1$ support functions $q_0, q_1, \ldots, q_s$ about the points $\omega_0, \omega_1, \ldots, \omega_s$. A new quadratic support function $q_{s+1}$ is constructed about $\omega_{s+1}$, which is a global minimizer of

$$\overline{q}_s(\omega) = \max_{k=0,\ldots,s} q_k(\omega). \tag{7.1}$$

The details of the algorithm are formally presented in Algorithm 1. In the algorithm, the computationally challenging task is the determination of a global minimizer of $\overline{q}_s$ defined by (7.1). For this purpose, we partition the box $\mathcal{B}$ into regions $\mathcal{R}_0, \ldots, \mathcal{R}_s$ such that the quadratic function $q_k$ takes the largest value inside the region $\mathcal{R}_k$ (see Figure 1). Therefore, the minimization of $\overline{q}_s$ over the region $\mathcal{R}_k$ is equivalent to the

ALGORITHM 1. Support-based eigenvalue optimization.

**Require:** A box $\mathcal{B} \subset \mathbb{R}^d$ and an eigenvalue function $\lambda : \mathcal{B} \to \mathbb{R}$ satisfying either the generic simplicity assumption or of the form $\lambda(\omega) = \sum_{k=1}^{j} d_k \lambda_k(\omega)$ with $d_1 \geq \cdots \geq d_j \geq 0$.

Pick an arbitrary $\omega_0 \in \mathcal{B}$.
$u_1 \leftarrow \lambda(\omega_0)$; $\omega best \leftarrow \omega_0$.
$\overline{q}_0(\omega) := q_0(\omega) := \lambda(\omega_0) + \nabla\lambda(\omega_0)^T(\omega - \omega_0) + (\gamma/2)\|\omega - \omega_0\|^2$.
$\omega_1 \leftarrow \arg\min_{\omega\in\mathcal{B}} \overline{q}_0(\omega)$; $l_1 \leftarrow \overline{q}_0(\omega_1)$.
**if** $\lambda(\omega_1) < u_1$ **then**
  $u_1 \leftarrow \lambda(\omega_1)$; $\omega best \leftarrow \omega_1$.
**end if**
$s \leftarrow 1$.
While $u_s - l_s > \epsilon$ do
**loop**
  $q_s(\omega) := \lambda(\omega_s) + \nabla\lambda(\omega_s)^T(\omega - \omega_s) + (\gamma/2)\|\omega - \omega_s\|^2$.
  $\overline{q}_s(\omega) := \max_{k=0,\ldots,s}\{q_k(\omega)\}$.
  $\omega_{s+1} \leftarrow \arg\min_{\omega\in\mathcal{B}} \overline{q}_s(\omega)$; $l_{s+1} \leftarrow \overline{q}_s(\omega_{s+1})$.
  **if** $\lambda(\omega_{s+1}) < u_s$ **then**
    $u_{s+1} \leftarrow \lambda(\omega_{s+1})$; $\omega best \leftarrow \omega_{s+1}$.
  **else**
    $u_{s+1} \leftarrow u_s$.
  **end if**
  $s \leftarrow s + 1$.
**end loop**
**Output:** $l_s, u_s, \omega best$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

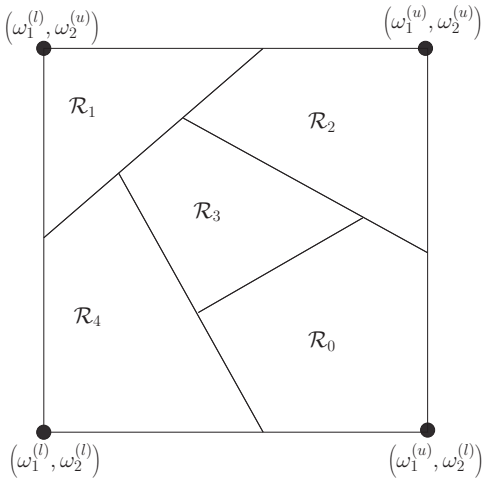



FIG. 1. *To minimize the piecewise quadratic function $\overline{q}_s$ over the box $\mathcal{B}$, the box is split into regions $\mathcal{R}_k$, $k = 0, \ldots, s$, such that $q_k$ is the largest inside the region $\mathcal{R}_k$. Above, a possible partitioning with $s = 4$ is illustrated in the two-dimensional case.*

minimization of $q_k$ over the same region. This problem can be posed as the following quadratic programming problem:

$$\begin{aligned} &\text{minimize}_{\omega \in \mathbb{R}^d} && q_k(\omega) \\ &\text{subject to} && q_k(\omega) \geq q_\ell(\omega), \quad \ell \neq k, \\ &&& \omega_j \in [\omega_j^{(l)}, \omega_j^{(u)}], \qquad j = 1, \ldots, d. \end{aligned} \tag{7.2}$$

Here, we remark that the seemingly quadratic inequalities $q_k(\omega) \geq q_\ell(\omega)$ are in fact linear since the quadratic terms cancel out. The feasibility of the algorithm largely relies on the efficiency with which we can solve subproblems (7.2). These subproblems are nonconvex whenever $\gamma < 0$. However, since they involve the optimization of a concave function over a polytope, the solution for each subproblem must be attained at one of the vertices of the polytope.

Breiman and Cutler introduced the notion of quadratic support functions of the form (5.1) for global optimization [4].[1] They described how the subproblems of the form (7.2) can be solved efficiently. At each iteration, when a new quadratic support function is added, a new polytope is introduced associated with it, and some points—call these dead vertices—that used to be vertices of some polytope are no longer vertices (i.e., they now lie strictly inside the new polytope). A vertex $v$ is dead after the introduction of $q_{s+1}(\omega)$ if and only if $q_{s+1}(\omega) > \overline{q}_s(\omega)$. Breiman and Cutler observed that these dead vertices form a connected graph. Therefore, they can be identified efficiently. Furthermore, a new vertex $v$ (on the boundary of the new polytope) appears between a dead vertex $v_d$ and each vertex $v_a$—each existing vertex that is still a vertex after the addition of $q_{s+1}(\omega)$ and adjacent to $v_d$—given by the formula

$$\begin{aligned} v = &\left[ \frac{q_{s+1}(v_a) - \overline{q}_s(v_a)}{(q_{s+1}(v_a) - \overline{q}_s(v_a)) - (q_{s+1}(v_d) - \overline{q}_s(v_d))} \right] v_d \\ &+ \left[ 1 - \frac{q_{s+1}(v_a) - \overline{q}_s(v_a)}{(q_{s+1}(v_a) - \overline{q}_s(v_a)) - (q_{s+1}(v_d) - \overline{q}_s(v_d))} \right] v_a. \end{aligned}$$

[1] We are grateful to an anonymous referee who pointed us to this reference.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

These new vertices, possibly with the addition of some of the vertices where only box constraints are active, form the set of vertices for the new polytope. The edges for the new polytope can be determined by the common active constraints of its vertices; two vertices are adjacent if and only if there are $d-1$ active constraints common to these vertices. All these observations accompanied by appropriate data structures (e.g., a heap of vertices indexed based on the values of $\overline{q}_s$ after the formation of $q_s$ and an adjacency list for each vertex) lead to an efficient algorithm when the dimension $d$ is small. We refer to [4] for further details.

Algorithm 1 could be based on the more general piecewise quadratic support functions (4.5) by adjusting lines 3 and 11 accordingly. This would make the algorithm applicable for the optimization of more general eigenvalue functions, i.e., those that do not involve the sum of the largest eigenvalues and violating generic simplicity everywhere. Subproblems analogous to (7.2) could be devised; however, their solutions appear prohibitively expensive.

**8. Convergence analysis.** In this section, we analyze the convergence of Algorithm 1—in the general setting when the support functions (4.5) are used on lines 3 and 11—for the following optimization problem:

$$\text{(P)} \quad \lambda^* := \min_{\omega \in \mathcal{B}} \lambda(\omega).$$

Recall that the algorithm starts off by picking an arbitrary point $\omega_0 \in \mathcal{B}$. At iteration $s$, the algorithm picks $\omega_{s+1}$ to be a global minimizer of $\overline{q}_s(x)$ over $\mathcal{B}$, where $\overline{q}_s(\omega)$ is the maximum of the functions $q_k(\omega)$ constructed at the points $\omega_k,\ k = 0, \ldots, s$, in $\mathcal{B}$. Note that $\{l_s\}$ is a nondecreasing sequence of lower bounds on $\lambda^*$, while $\{u_s\}$ is a nonincreasing sequence of upper bounds on $\lambda^*$.

We require that $\lambda : \mathcal{B} \to \mathbb{R}$ be a continuous and piecewise function defined in terms of the differentiable functions $\tilde{\lambda}_j : \mathbb{R}^d \to \mathbb{R},\ j = 1, \ldots, n$, as described in section 4. The differentiability of each $\tilde{\lambda}_j$ on $\mathbb{R}^d$ implies the boundedness of $\|\nabla \tilde{\lambda}_j(x)\|$ on $\mathcal{B}$. Consequently, we define

$$\mu := \max_{j=1,\ldots,n} \max_{\omega \in \mathcal{B}} \|\nabla \tilde{\lambda}_j(\omega)\|.$$

We furthermore require each piece $\tilde{\lambda}_j$ to be analytic along every line in $\mathbb{R}^d$, and we exploit the existence of a scalar $\gamma$ that satisfies (4.4). Our convergence analysis depends on the scalars $\mu$ and $\gamma$. We now establish the convergence of Algorithm 1 to a global minimizer of (P).

THEOREM 8.1. *Let $\{\omega_s\}$ be the sequence of iterates generated by Algorithm* 1 *in the general case when the support functions* (4.5) *are used on lines* 3 *and* 11. (i) *Every limit point of this sequence is a global minimizer of the problem* (P). (ii) *Furthermore* $\lim_{s\to\infty} u_s = \lim_{s\to\infty} l_s = \lambda^*$.

*Proof.* Since $\mathcal{B}$ is a bounded subset of $\mathbb{R}^d$, it follows that the sequence $\{\omega_s\}$ has at least one limit point $\omega^* \in \mathcal{B}$. By passing to a subsequence if necessary, we may assume that $\{\omega_s\}$ itself is a convergent sequence. Let $l^*$ denote the limit of the bounded nondecreasing sequence $\{l_s\}$. Since $l_s \le \lambda^* \le \lambda(\omega_s)$ for each $s \ge 0$, it suffices to show that $l^* = \lambda(\omega^*)$.

Suppose, for a contradiction, that there exists a real number $\delta > 0$ such that

$$\lambda(\omega^*) \ge l^* + \delta. \tag{8.1}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

By the continuity of $\lambda$, there exists $s_1 \in \mathbb{N}$ such that

$$\lambda(\omega_s) \geq l^* + \frac{\delta}{2} \quad \text{for all } s \geq s_1. \tag{8.2}$$

Since $\omega^*$ is the limit of the sequence $\{\omega_s\}$, there exists $s_2 \in \mathbb{N}$ such that

$$\|\omega_{s'} - \omega_{s''}\| < \min\left\{\sqrt{\frac{\delta}{6|\gamma|}}, \frac{\delta}{12\mu}\right\} \quad \text{for all } s' \geq s'' \geq s_2, \tag{8.3}$$

where we define $1/\mu := +\infty$ if $\mu = 0$ and $1/|\gamma| := +\infty$ if $\gamma = 0$. Let $s_* = \max\{s_1, s_2\}$. For each $s \geq s_*$, it follows from the definition of the functions $\overline{q}_s(\omega)$ that

$$\begin{aligned}
\overline{q}_s(\omega_{s+1}) &\geq \overline{q}_{s_*}(\omega_{s+1}) \\
&\geq q_{s_*}(\omega_{s+1}) \\
&= \lambda(\omega_{s_*}) + \nabla\tilde{\lambda}_{j_*}(\omega_{s_*})^T(\omega_{s+1} - \omega_{s_*}) + \frac{\gamma}{2}\|\omega_{s+1} - \omega_{s_*}\|^2,
\end{aligned}$$

where $j_* \in \{1, \ldots, n\}$ is the index of the function that determines the value of $q_{s_*}(w_{s+1})$ (see (4.5)). Now, by applying the Cauchy–Schwarz and triangle inequalities, and then using the inequalities (8.2) and (8.3), we arrive at

$$\begin{aligned}
\overline{q}_s(\omega_{s+1}) &\geq \lambda(\omega_{s_*}) - \|\nabla\tilde{\lambda}_{j_*}(x^*)\|\|\omega_{s+1} - \omega_{s_*}\| - \frac{|\gamma|}{2}\|\omega_{s+1} - \omega_{s_*}\|^2 \\
&\geq \left(l^* + \frac{\delta}{2}\right) - \left(\mu \cdot \frac{\delta}{12\mu}\right) - \left(\frac{|\gamma|}{2} \cdot \frac{\delta}{6|\gamma|}\right) \\
&= l^* + \frac{\delta}{3}.
\end{aligned}$$

Using the definition $l_{s+1} = \overline{q}_s(\omega_{s+1})$, it follows that

$$l_{s+1} \geq l^* + \frac{\delta}{3} \quad \text{for all } s \geq s_*.$$

Since $\delta > 0$, this contradicts our assumption that $l^*$ is the limit of the nondecreasing sequence $\{l_s\}$. Therefore, we have $\lambda(\omega^*) < l^* + \delta$ for all $\delta > 0$, or equivalently $\lambda(\omega^*) \leq l^*$. Since $l_s \leq \lambda(\omega)$ for all $s \in \mathbb{N}$ and $\omega \in \mathcal{B}$, it follows that $l^* \leq \lambda(\omega^*)$, which establishes that $\lambda(\omega^*) = l^* \leq \lambda(\omega)$ for all $\omega \in \mathcal{B}$. Therefore, $\omega^*$ is a global minimizer of (P). Moreover, $\lim_{s\to\infty} l_s = \lambda^*$. The sequence $\{u_s\}$ must also converge to $\lambda^*$, which can be deduced by observing $\lambda^* \leq u_s \leq \lambda(\omega_s)$ for each $s \in \mathbb{N}$ and taking the limit as $s \to \infty$. The proof of assertion (i) is completed by repeating the same argument for any other limit point of the sequence $\{\omega_s\}$.

Assertion (ii) can be concluded by noting that the monotone, bounded sequences $\{l_s\}$ and $\{u_s\}$ must converge, and they have subsequences converging to $\lambda^*$. ∎

**9. Numerical experiments.** We compare Algorithm 1 with the following algorithms:

(1) a brute force approach;
(2) the Piyavskii–Shubert algorithm [40, 45] (only one-dimensional case);
(3) DIRECT method [24];
(4) the specialized level-set based algorithms whenever possible.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The brute force approach (1) splits the box $\mathcal{B}$ into subboxes of equal side-lengths and the eigenvalue function is computed at the corners of the subboxes. Algorithms (2) and (3) are global optimization techniques based on Lipschitz continuity of the function. The latter method (3) benefits from several Lipschitz constant estimates simultaneously, while the former one (2) utilizes a global Lipschitz constant. Algorithms that fall into the category (4) are level-set based approaches. They typically converge fast, but each iteration is costly. Each of the algorithms in (4) is devised for a particular eigenvalue optimization problem.

*Example* 1 (numerical radius). This one-dimensional example concerns the calculation of the numerical radius (defined and motivated in section 2.1) of an $n \times n$ matrix $A$. The matrix-valued function involved is $\mathcal{A}(\theta) = -(Ae^{i\theta} + A^*e^{-i\theta})/2$ and $\lambda_n(\theta) := \lambda_n(\mathcal{A}(\theta))$ is sought to be minimized over all $\theta \in [0, 2\pi]$. The numerical radius of $A$ corresponds to the negative of this globally minimal value of $\lambda_n(\theta)$. Here, we assume that the generic analyticity holds, that is, the eigenvalue $\lambda_n(\theta)$ is simple for all $\theta$. The derivative

$$\frac{d\lambda_n(\theta)}{d\theta} = \Im\left(v_n^*(\theta)Ae^{i\theta}v_n(\theta)\right) \tag{9.1}$$

can be deduced from (3.3). We numerically observe that typically $\lambda_n''(\theta) \geq -2\|A\|$ holds for all $\theta$, and set $\gamma = -2\|A\|$. We specifically focus on matrices

$$A_n = P_n - (n/20) \cdot iR_n$$

of various sizes, where $P_n$ is an $n \times n$ matrix obtained from a finite difference discretization of the Poisson operator, and $R_n$ is a random $n \times n$ matrix with entries selected from a normal distribution with zero mean and unit variance. This is a carefully chosen challenging example, as $\lambda_n(\theta)$ has many local minima (see Figure 2).

In Table 1, the function evaluations of algorithms (1)–(3) are given along with the function evaluations of Algorithm 1 for the matrix $A_n$ with $n = 400$ and with respect to absolute accuracy. Algorithm 1—called eigopt in the table—converges linearly; this is evident from roughly the same number of function evaluations required for every two-decimal-digit accuracy. All other algorithms, including the DIRECT method, converge sublinearly. The computed global minimizer is marked (with an asterisk) on a plot of $\lambda_n(\theta)$ with respect to $\theta \in [0, 2\pi]$ in Figure 2 on the left. In the same figure on the right, the first five quadratic support functions are shown for the same example. The level-set based algorithm (4) for the numerical radius [32] requires the solutions of eigenvalue problems twice the size of $\mathcal{A}(\theta)$. It is not included in Table 1, because these larger eigenvalue problems dominate the computation time rather than the calculation of $\lambda_n(\theta)$. Instead we compare the CPU times (in seconds) of Algorithm 1 and this specialized algorithm on Poisson matrices $A_n$ of various sizes $n$ in Table 2. Algorithm 1 is run to retrieve the results with at least 10-decimal-digit accuracy. The table displays the superiority of the running times of Algorithm 1 as compared to those of the level-set approach.

*Example* 2 (distance to uncontrollability). Next, we consider the distance to uncontrollability defined and motivated in section 2.1. Here, for a given linear system $x'(t) = Ax(t) + Bu(t)$, where $A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{n\times m}$ are such that $n \geq m$, the smallest singular value of $\mathcal{A}(z) = \begin{bmatrix} A - zI & B \end{bmatrix}$ is sought to be minimized over $z \in \mathbb{C}$. We again assume generic simplicity, that is, the multiplicity of $\sigma_n(z) := \sigma_n(\mathcal{A}(z))$ is one for all $z$. This property holds on a dense subset of all pairs $(A, B)$. In this case, expressions for the gradient is given by (3.8). In particular, denoting a consistent pair of unit left and

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

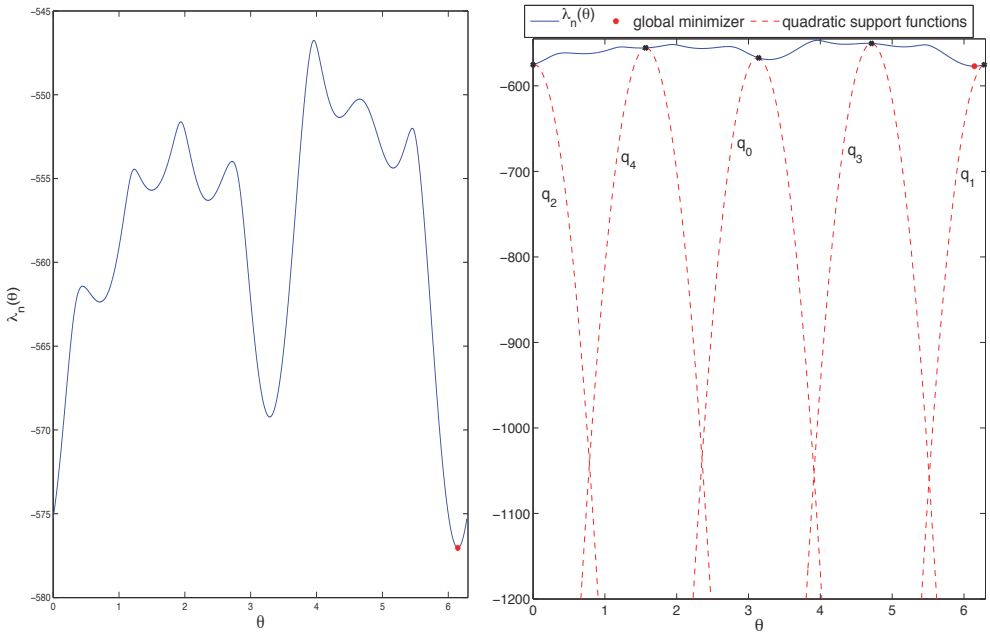



FIG. 2. *On the left, the plot of $\lambda_n(\mathcal{A}(\theta))$ for the $400 \times 400$ Poisson example (Example 1) with respect to $\theta$; On the right, the plot of $\lambda_n(\mathcal{A}(\theta))$ together with the first five quadratic support functions $q_j$, $j = 0, \ldots, 4$, formed. In both figures, the asterisk represents the global minimizer; the squares on the right mark the iterates of the algorithm around which the support functions are constructed.*

TABLE 1

*Number of function evaluations by various algorithms to compute the numerical radius of the $400 \times 400$ Poisson example (Example 1) with respect to absolute accuracy $\epsilon$. Note that eigopt makes use of the derivatives in addition to the eigenvalues evaluated. However, their derivatives come essentially at no cost once the eigenvalues are evaluated.*

| $\epsilon$ | $10^{-2}$ | $10^{-4}$ | $10^{-6}$ | $10^{-8}$ | $10^{-10}$ | $10^{-12}$ |
|---|---|---|---|---|---|---|
| eigopt | 46 | 59 | 69 | 79 | 89 | 98 |
| Brute force | 881 | 8812 | 88125 | 881249 | 8815191 | 86070462 |
| Piyavskii–Shubert | 1907 | 18817 | – | – | – | – |
| DIRECT | 25 | 51 | 61 | 105 | 245 | 597 |

TABLE 2

*CPU times (in seconds) required by Algorithm 1 (eigopt) and the specialized level-set approach in [32] to compute the numerical radius of Poisson matrices of varying size $n$.*

| $n$ | 400 | 900 | 1600 | 2500 | 3600 |
|---|---|---|---|---|---|
| eigopt | 14 | 103 | 328 | 1079 | 2788 |
| Level-set | 17 | 181 | 1477 | – | – |

right singular vectors associated with $\sigma_n(z)$ by $u_n(z) \in \mathbb{C}^n, v_n(z) = \begin{bmatrix} \tilde{v}_n(z) \\ \hat{v}_n(z) \end{bmatrix} \in \mathbb{C}^{n+m}$ where $\tilde{v}_n(z) \in \mathbb{C}^n, \hat{v}_n(z) \in \mathbb{C}^m$, we have

$$\nabla \sigma_n(z) = \left( \frac{\partial \sigma_n(z)}{\partial \Re z}, \frac{\partial \sigma_n(z)}{\partial \Im z} \right) = \left( -\Re \left( u_n^*(z) \tilde{v}_n(z) \right), \Im \left( u_n^*(z) \tilde{v}_n(z) \right) \right).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 3

*Number of function evaluations by various algorithms to compute the distance to uncontrollability for the linear system $(A, B)$ arising from the heat equation* [25, *Example* 3.2], *where $A \in \mathbb{R}^{30\times 30}$, $B \in \mathbb{R}^{30\times 1}$ with respect to absolute accuracy $\epsilon$.*

| $\epsilon$ | $10^{-2}$ | $10^{-4}$ | $10^{-6}$ | $10^{-8}$ | $10^{-10}$ | $10^{-12}$ |
|---|---|---|---|---|---|---|
| eigopt | 520 | 531 | 543 | 556 | 572 | 585 |
| Brute force | $4.2 \times 10^6$ | $4.2 \times 10^{10}$ | $4.2 \times 10^{14}$ | $4.2 \times 10^{18}$ | $4.2 \times 10^{22}$ | $4.2 \times 10^{26}$ |
| DIRECT | 37781 | 37867 | 37867 | 38233 | 38441 | 38805 |

Furthermore, $\gamma = -4$ appears to be a good lower bound for $\lambda_{\min}\left(\nabla^2 \sigma_n(z)\right)$ numerically. We perform tests on linear systems $(A, B)$ arising from a discretization of the heat equation, taken from the SLICOT library; see [25, Example 3.2]. The matrix $A$ is real, symmetric, tridiagonal, and $n \times n$, whereas $B$ is real and $n \times 1$.

In Table 3, the number of function evaluations required by global optimization algorithms for Lipschitz continuous functions, and Algorithm 1, are presented when the order of the system satisfies $n = 30$. The Piyavskii–Shubert algorithm is omitted, because it would be based on locating global minimizers of piecewise cones, and it is not immediate how one would locate these minimizers. Once again, the number of function evaluations for Algorithm 1 seems to suggest linear convergence.

The progress of the algorithm can be traced from the graph associated with it. Recall that the box is split into subregions (indeed polytopes). Inside each subregion, one of the quadratic support functions dominates the others. For the heat equation example of order $n = 30$, these graphs are provided after 60, 360, and 580 iterations in Figure 3 in the top three plots. The bottom plot in Figure 3 is an illustration of the level sets of $\sigma_n(z)$ in the complex plane along with the computed global minimizer (accurate up to 13 decimal digits after 589 function evaluations) marked with an asterisk. The box is split into subregions more or less uniformly initially, e.g., after 60 iterations. However, later iterations form finer subregions around minimizers of $\sigma_n(z)$.

We also compare Algorithm 1 with the level-set approach in [20] on the heat equation examples of varying order. The level set approaches become prohibitively expensive as the number of optimization parameters increases. Here, with the minimization over two parameters, it requires the solutions of eigenvalue problems of size $n^2$ for a system of order $n$. For small systems, the level-set approach works very well; however, even for medium-scale systems it becomes computationally infeasible. This is illustrated in Table 4. On the other hand, Algorithm 1 is capable of solving even a problem of order 1000 in a reasonable amount of time.

*Example* 3 (minimizing the largest eigenvalue, convex). This example is taken from [13] and concerns the minimization of the largest eigenvalue of an affine matrix function of the form

$$\mathcal{A}(\omega) = A_0 + \sum_{j=1}^{5} \omega_j A_j,$$

depending on five parameters, where $A_j \in \mathbb{R}^{5\times 5}$, $j = 0, \ldots, 5$, are real, symmetric, and as in [13], where the minimal value of $\lambda_1(\omega) := \lambda_1(\mathcal{A}(\omega))$ is cited as 0.708882597. Whenever $\mathcal{A}(\omega)$ is affine, the largest eigenvalue $\lambda_1(\omega)$ is convex by Theorem 6.1 (see also [14, Theorem A.1]), so we set $\gamma = 0$. Thus, this eigenvalue optimization problem can be posed as a semidefinite program and solved by means of interior-point methods [34]. The purpose here is to compare the performances of Algorithm 1 and the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

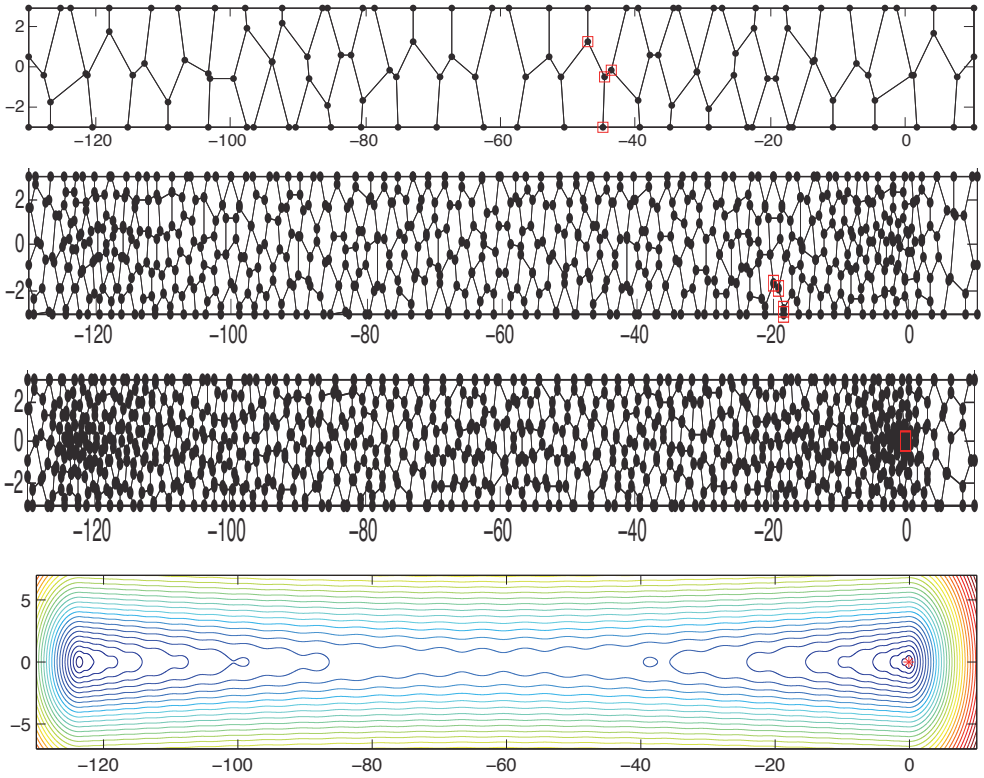



FIG. 3. *The top three plots display the graph associated with Algorithm* 1 *after* 60, 360, *and* 580 *iterations to compute the distance to uncontrollability for the heat equation example* [25, *Example* 3.2] *of order* $n = 30$ *(see Example* 2*). The red squares correspond to dead vertices. The bottom plot displays the level sets of* $\sigma_n(z)$ *for the same heat equation example together with the computed global minimizer marked with an asterisk.*

TABLE 4

*CPU times (in seconds) required by Algorithm* 1 *(eigopt) and the level-set approach in* [20] *to compute the distance to uncontrollability for the heat equation examples (Example* 2*) with respect to the order of the system* $n$*.*

| $n$ | 30 | 60 | 100 | 200 | 400 | 1000 |
|---|---|---|---|---|---|---|
| eigopt | 45 | 46 | 56 | 85 | 119 | 384 |
| Level-set | 20 | 393 | – | – | – | – |

DIRECT method on a multidimensional example for which the solution is known (even though our algorithm is really devised for nonconvex problems, and in all likelihood an interior-point method would outperform it in the convex case).

This comparison is provided in Table 5, where we again observe linear convergence of Algorithm 1. The DIRECT method is not suitable even for a few decimal-digits precision. Indeed, even after 500,000 function evaluations and 3732 seconds of CPU time, it cannot achieve six-decimal-digit accuracy. In contrast to the one-dimensional and two-dimensional examples, keeping the graph structure properly, in particular forming the adjacencies between the new vertices, takes almost all the computational time rather than the function evaluations. Even if the matrices $A_j$ were much bigger

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 5

*Number of function evaluations by Algorithm* 1 *(eigopt) and the DIRECT method for minimizing the largest eigenvalue of the affine matrix function (Example* 3*) with respect to absolute accuracy* $\epsilon$*. The CPU times for eigopt (in seconds) are also provided in parentheses.*

| $\epsilon$ | $10^{-2}$ | $10^{-4}$ | $10^{-6}$ | $10^{-8}$ | $10^{-10}$ | $10^{-12}$ |
|---|---|---|---|---|---|---|
| eigopt | 17 (34) | 34 (203) | 50 (501) | 67 (1021) | 87 (2040) | 108 (3725) |
| DIRECT | 459 | 45937 | – | – | – | – |

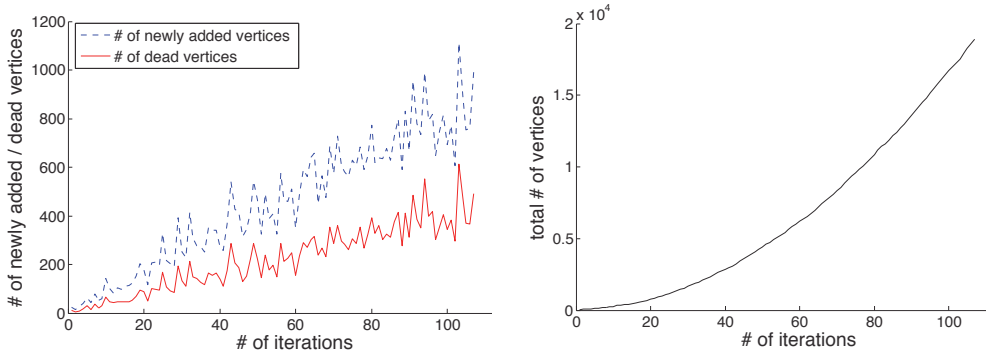



FIG. 4. *On the left, the number of dead vertices (solid line) and the number of newly added vertices (dashed line) with respect to the iteration number are displayed, whereas on the right, the total number of vertices with respect to the iteration number is displayed for Example* 3.

than $5 \times 5$, the computation times would not be affected significantly. Therefore, in the table, the CPU times (in seconds) are also provided in parenthesis.

The later iterations are more expensive, since the number of new vertices created increases at the later iterations. Up to four dimensions, the increase in the number of vertices at every iteration seems more or less fixed. In contrast, this does not hold when the dimension is five or more. (These observations are solely based on numerical experiments; we do not have a clear understanding of this phenomenon at the moment.) This is illustrated for the affine example with five parameters in Figure 4. In the figure on the left, the solid and dashed lines represent the number of dead vertices and the number of newly added vertices, respectively, at iterations $1, 2, \ldots, 107$. The graph indicates an increase in both the number of dead vertices and the number of new vertices with respect to the iteration. However, the increase in the number of new vertices is larger. On the right, the total number of vertices is displayed with respect to the iteration number. The graph reveals that the number of vertices seems to increase superlinearly.

*Example* 4 (minimizing the largest eigenvalue, nonconvex). This is a nonconvex example depending on four parameters, involving the minimization of the largest eigenvalue of the matrix function

$$\mathcal{A}(\omega) = A_0 + \sum_{j=1}^{4} \omega_j A_j + \frac{1}{2} \sum_{j=1}^{4} \sum_{k=1}^{4} \omega_j \omega_k A_{jk}, \tag{9.2}$$

where $A_0, A_1, \ldots, A_4$ are as in the previous affine example taken from [13], while $A_{jk}$ are randomly chosen $5 \times 5$ symmetric and real matrices such that $A_{jk} = A_{kj}$. By Corollary 6.2, a lower bound for $\lambda_{\min}\left(\nabla^2 \lambda_1\left(\mathcal{A}(\omega)\right)\right)$ is given by $\gamma = \lambda_{\min}\left(\nabla^2 \mathcal{A}\right)$,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 6
*Number of function evaluations by Algorithm* 1 *(eigopt) and DIRECT method for minimizing the largest eigenvalue of the matrix function* (9.2) *with respect to absolute accuracy* $\epsilon$.

| $\epsilon$ | $10^{-2}$ | $10^{-4}$ | $10^{-6}$ | $10^{-8}$ | $10^{-10}$ | $10^{-12}$ |
|---|---|---|---|---|---|---|
| eigopt | 52 | 141 | 225 | 309 | 391 | 472 |
| DIRECT | 45 | 161 | 369 | 939 | 1367 | 2049 |

where $\nabla^2\mathcal{A} \in \mathbb{R}^{20\times 20}$, and $A_{jk}$ constitutes the $5 \times 5$ submatrix of $\nabla^2\mathcal{A}$ at rows $5(j-1)+1 : 5j$ and columns $5(k-1)+1 : 5k$. Table 6 displays a comparison of the number of function evaluations for Algorithm 1 and the DIRECT method with respect to accuracy. As usual, Algorithm 1 seems to exhibit linear convergence.

**10. Software.** A MATLAB implementation of Algorithm 1 is available on the Web. The implementation makes use of heaps, adjacency lists, and stacks and updates the underlying graph structure efficiently. The user is expected to write down a MATLAB routine that calculates the eigenvalue function and its gradient at a given point. The name of this routine and $\gamma$, a global lower bound on the minimum eigenvalues of the Hessians of the eigenvalue functions, must be supplied by the user. We refer to the Web page associated with this implementation,[2] where a user guide is also provided.

**11. Conclusion.** The analytical properties of eigenvalues of matrix-valued functions facilitate the use of so-called quadratic support functions that globally underestimate the eigenvalue curves for their optimization. This observation motivates the idea of adapting support function based global optimization approaches, especially the approach due to Breiman and Cutler [4], for nonconvex eigenvalue optimization. In this paper, we illustrated how such global optimization approaches based on the derivative information could be realized in the context of nonconvex eigenvalue optimization. We derived the necessary quadratic support functions, elaborated on deducing analytical global lower bounds $\gamma$ for the second derivatives of the extreme eigenvalue functions, which are essential for the algorithm, and provided a global convergence proof. The algorithm is especially applicable for the optimization of extreme eigenvalues, for instance, for the minimization of the largest eigenvalue, and those eigenvalue functions that exhibit generic simplicity.

**Acknowledgments.** We thank two anonymous referees for their invaluable comments, which improved this paper considerably. We are also grateful to Michael Karow and Melina Freitag for helpful discussions and feedback.

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] D. BARNES, *The shape of the strongest column is arbitrarily close to the shape of the weakest columns*, Quart. Appl. Math., 44 (1986), pp. 605–609.
[3] S. BOYD AND V. BALAKRISHNAN, *A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its* $\mathbf{L}_\infty$*-norm*, Systems Control Lett., 15 (1990), pp. 1–7.
[4] L. BREIMAN AND A. CUTLER, *A deterministic algorithm for global optimization*, Math. Program., 58 (1993), pp. 179–199.

[2] http://home.ku.edu.tr/∼emengi/software/eigopt.html.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

[5] N. A. Bruinsma and M. Steinbuch, *A fast algorithm to compute the $H_\infty$-norm of a transfer function matrix*, Systems Control Lett., 14 (1990), pp. 287–293.
[6] R. Byers, *A bisection method for measuring the distance of a stable matrix to the unstable matrices*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 875–881.
[7] R. Byers, *The descriptor controllability radius*, in Proceedings of the International Symposium on Numerical Methods, 2, U. Helmke, R. Mennicken, and H. Saurer, eds., Akademie Verlag, Berlin, 1993, pp. 85–88.
[8] S. J. Cox and M. L. Overton, *On the optimal design of columns against buckling*, SIAM J. Math. Anal., 23 (1992), pp. 287–325.
[9] J. Cullum, W. E. Donath, and P. Wolfe, *The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices*, Math. Program Study, 3 (1975), pp. 35–55.
[10] W. E. Donath and A. J. Hoffman, *Lower bounds for the partitioning of graphs*, IBM J. Res. Dev., 17 (1973), pp. 420–425.
[11] M. Eiermann, *Field of values and iterative methods*, Linear Algebra Appl., 180 (1993), pp. 167–197.
[12] R. Eising, *Between controllable and uncontrollable*, Systems Control Lett., 4 (1984), pp. 263–264.
[13] M. K. H. Fan and B. Nekooie, *On minimizing the largest eigenvalue of a symmetric matrix*, Linear Algebra Appl., 214 (1995), pp. 225–246.
[14] R. Fletcher, *Semidefinite matrix constraints in optimization*, SIAM J. Control Optim., 23 (1985), pp. 493–513.
[15] 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.
[16] S. Friedland, J. Nocedal, and M. L. Overton, *The formulation and analysis of numerical methods for inverse eigenvalue problems*, SIAM J. Numer. Anal., 24 (1987), pp. 634–667.
[17] M. Gao and M. Neumann, *A global minimum search algorithm for estimating the distance to uncontrollability*, Linear Algebra Appl., 188–189 (1993), pp. 305–350.
[18] V. P. Gergel, *A global optimization algorithm for multivariate functions with Lipschitzian first derivatives*, J. Global Optim., 10 (1997), pp. 257–281.
[19] M. Gu, *New methods for estimating the distance to uncontrollability*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 989–1003.
[20] M. Gu, E. Mengi, M. L. Overton, J. Xia, and J. Zhu, *Fast methods for estimating the distance to uncontrollability*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 477–502.
[21] C. He and G. A. Watson, *An algorithm for computing the distance to instability*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 101–116.
[22] D. Hinrichsen and M. Motscha, *Optimization problems in the robustness analysis of linear state space systems*, in Proceedings of the International Seminar on Approximation and Optimization, Springer-Verlag, New York, 1988, pp. 54–78.
[23] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.
[24] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, *Lipschitzian optimization without the Lipschitz constant*, J. Optim. Theory Appl., 79 (1993), pp. 157–181.
[25] D. Kressner, V. Mehrmann, and T. Penzl, *Ctdsx—A Collection of Benchmarks for State-Space Realizations of Continuous-Time Dynamical Systems*, SLICOT Working Note, 1998.
[26] D. E. Kvasov and Ya. D. Sergeyev, *A univariate global search working with a set of Lipschitz constants for the first derivative*, Optim. Lett., 3 (2009), pp. 303–318.
[27] D. E. Kvasov and Ya. D. Sergeyev, *Lipschitz gradients for global optimization in a one-point-based partitioning scheme*, J. Comput. Appl. Math., 236 (2012), pp. 4042–4054.
[28] P. Lancaster, *On eigenvalues of matrices dependent on a parameter*, Numer. Math., 6 (1964), pp. 377–387.
[29] D. Lera and Ya. D. Sergeyev, *Acceleration of univariate global optimization algorithms working with Lipschitz functions and Lipschitz first derivatives*, SIAM J. Optim., 23 (2013), pp. 508–529.
[30] A. S. Lewis and M. L. Overton, *Eigenvalue optimization*, Acta Numer., 5 (1996), pp. 149–190.
[31] M. Mäkelä, *Survey of bundle methods for nonsmooth optimization*, Optim. Methods Softw., 17 (2002), pp. 1–29.
[32] 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.
[33] M. Myers and W. Spillers, *A note on the strongest fixed-fixed column*, Quart. Appl. Math., XLIV (1986), pp. 583–588.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[34] Y. Nesterov and A. Nemirovski, *Interior Point Polynomial Methods in Convex Programming*, SIAM, Philadelphia, 1994.
[35] N. Olhoff and S. Rasmussen, *On single and bimodal optimum buckling loads of clamped columns*, Int. J. Solids Struct., 13 (1977), pp. 605–614.
[36] M. L. Overton, *On minimizing the maximum eigenvalue of a symmetric matrix*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 256–268.
[37] M. L. Overton, *Large-scale optimization of eigenvalues*, SIAM J. Optim., 2 (1991), pp. 88–120.
[38] M. L. Overton and R. S. Womersley, *Second derivatives for optimizing eigenvalues of symmetric matrices*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 697–718.
[39] C. C. Paige, *Properties of numerical algorithms relating to computing controllability*, IEEE Trans. Automat. Control, 26 (1981), pp. 130–138.
[40] S. A. Piyavskii, *An algorithm for finding the absolute extremum of a function*, USSR Comput. Math. and Math. Phys., 12 (1972), pp. 57–67.
[41] F. Rellich, *Perturbation Theory of Eigenvalue Problems*, Gordon and Breach, London, 1969.
[42] Ya. D. Sergeyev, *Multidimensional global optimization using the first derivatives*, Comput. Math. Math. Phys., 39 (1999), pp. 743–752.
[43] Y. D. Sergeyev and D. E. Kvasov, *Global search based on efficient diagonal partitions and a set of Lipschitz constants*, SIAM J. Optim., 16 (2006), pp. 910–937.
[44] A. Shapiro and M. K. H. Fan, *On eigenvalue optimization*, SIAM J. Optim., 5 (1995), pp. 552–569.
[45] B. Shubert, *A sequential method seeking the global maximum of a function*, SIAM J. Numer. Anal., 9 (1972), pp. 379–388.
[46] I. Tadjbakhsh and J. B. Keller, *Strongest columns and isoperimetric inequalities for eigenvalues*, J. Appl. Mech., 29 (1962), pp. 159–164.
[47] C. F. Van Loan, *How near is a matrix to an unstable matrix*? Lin. Alg. Role Systems Theory, 47 (1984), pp. 465–479.
[48] C. T. C. Wall, *Singular Points of Plane Curves*, Cambridge University Press, Cambridge, UK, 2004.

Downloaded 01/22/15 to 212.175.32.130. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.