NORTH-HOLLAND

On Minimizing the Largest Eigenvalue of a Symmetric Matrix*

Michael K. H. Fan† and Batool Nekooie

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Submitted by Richard A. Brualdi

ABSTRACT

Optimization problems involving eigenvalues arise in many engineering problems. In this paper, we consider the problem of minimizing the largest eigenvalue over an affine family of symmetric matrices. This problem has a variety of applications, such as the stability analysis of dynamic systems or the computation of structured singular values. Given $\epsilon \geq 0$, we give an optimality condition which ensures that the largest eigenvalue is within ϵ error bound of the solution. Also, a new line search rule is proposed, and it is shown to have good descent properties. When the multiplicity of the largest eigenvalue the solution is known, a new algorithm for the optimization problem under consideration is proposed. Some numerical experiments on the proposed algorithm are presented.

0. NOTATION

x^T	Transpose of vector x
$\|x\|$	Euclidean norm of vector x
I	Identity matrix of appropriate size
$A > 0$	Matrix A is positive definite (similar definitions for $A \geq 0$ and $A \leq 0$)

*Some of the results in this paper were given without proofs in [17].

†Corresponding author.

LINEAR ALGEBRA AND ITS APPLICATIONS 214:225–246 (1995)

655 Avenue of the Americas, New York, NY 10010

0024-3795/95/$9.50
SSDI 0024-3795(93)00068-B

$\xi_1(A)$	Largest eigenvalue of symmetric matrix A
co S	Convex hull of set S
nr S	Nearest point to the origin in set S
tr A	Trace of matrix A
∂B_k	Unit sphere in $\mathbb{R}^k$
π'	Derivative of function π
$A^\dagger$	Moore-Penrose generalized inverse of matrix A

Let $\eta(\cdot):\mathbb{R} \to \mathbb{R}$. The notation $\eta(x) = O(x^k)$ means that $\eta(x)$ goes to zero at least as fast as x^k does. More precisely, it means that there exists $K \geq 0$ such that $|\eta(x)|/|x^k| \leq K$ as $x \to 0$.

1. INTRODUCTION

Optimization problems involving eigenvalues arise in many engineering problems. In this paper, given symmetric matrices A_i, $i = 0, \ldots, m$, we consider the convex (nondifferentiable) optimization problem.

$$\lambda^* = \inf_{x \in \mathbb{R}^m} \lambda_1(x), \tag{1.1}$$

where $\lambda_1(x)$ denotes the largest eigenvalues of

$$A(x) = A_0 + \sum_{i=1}^{m} x_i A_i.$$

An application of (1.1) is in the following storage problem [6]. Suppose a company manufactures n chemical products. Certain pairs of these products are incompatible and would cause explosions if brought into contact with each other. As a precautionary measure the company wishes to partition its warehouse into compartments, and store incompatible chemical products in different compartments. The question is the least number of the compartments, denoted by χ, into which the warehouse should be partitioned. This problem is actually a special case of the vertex coloring problem in graph theory. In [26] Lovasz showed that the following inequality is satisfied:

$$\chi \geq \min_{A \in \mathcal{A}} \xi_1(A), \tag{1.2}$$

where $\mathcal{A}$ denotes the set of all real symmetric $n \times n$ matrices $A = (a_{ij})$ for which $A_{ij} = 1$ if $i = j$ or if ith and jth chemical products are incompatible. It is easy to see that the right hand side of (1.2) can be recast into (1.1). Other applications of (1.1) and its variations can be found in, e.g., [3, 7, 9, 10, 13–15, 18, 19, 34, 35]. For study of (1.1) and related problems, see, e.g., [1, 2, 4, 8, 20, 23, 28–33, 36] and the references therein.

$$A_0 = \begin{bmatrix} 0.2190 & 0.3632 & 0.5991 \\ 0.3632 & 0.9347 & 0.6072 \\ 0.5991 & 0.6072 & 0.0346 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} 0.1235 & 0.2687 & 0.5443 \\ 0.2687 & 0.1334 & 0.3768 \\ 0.5443 & 0.3768 & 0.1390 \end{bmatrix}$$

$$W = \left\{ [v^H A_0 v \quad v^H A_1 v]^T : v \in \partial B_3 \right\}$$

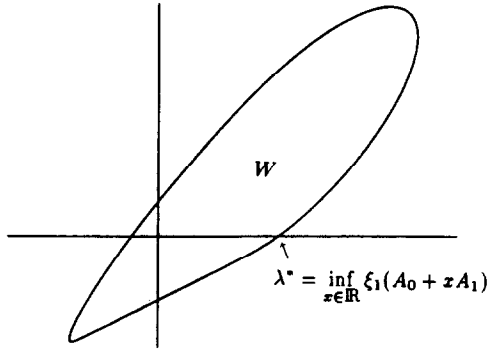

FIG. 1. Geometric interpretation of λ^*.

There is a simple geometric interpretation for the problem (1.1). Consider the intersection of the first coordinate axis and convex hull of the set

$$W = \{[v^H A_0 v \cdots v^H A_m v]^T : v \in \partial B_n\}.$$

Then it can be easily checked that λ^* is equal to the maximal value of the intersection. We illustrate this interpretation in Figure 1, using an example with $n = 3$ and $m = 1$. It has been shown that W is convex if $m = 1$ or if $m = 2$ and $n \geq 3$, and not necessarily convex otherwise [5, 21]. Notice that λ^* is finite if and only if there exists no x such that $\Sigma_{i=1}^m x_i A_i$ is positive definite. Furthermore, if λ^* is finite, it must lie between the smallest and largest eigenvalues of A_0.

It is well known that the eigenvalues of a matrix are not differentiable at points where they coalesce. Typically, the process of minimization tends to make them coalesce at the solution. Therefore, smooth optimization methods usually cannot be directly applied to this type of problems. In this paper we explore some properties of the problem (1.1). In Section 2, given $\epsilon \geq 0$ and $x \in \mathbb{R}^m$, we derive a sufficient condition on x to satisfy

$$\lambda_1(x) - \epsilon \leq \lambda^* < \lambda_1(x). \tag{1.3}$$

This condition reduces to the usual optimality condition for (1.1) when $\epsilon = 0$. Also we show how to construct descent directions for $\lambda_1(x)$ when the sufficient condition fails to hold. In Section 3, given a descent direction, we propose a line search rule for (1.1). Under some mild assumptions, it is shown that the resulting step size is a very good estimate of the one achieving the exact line search. In Section 4, under the assumption that the multiplicity of $\lambda_1(x)$ at solution is known, we propose an algorithm to solve (1.1). We first show that, locally, (1.1) is equivalent to a smooth constrained optimization problem. We then consider a sequence of subproblems which are constructed using second order approximations of the

objective and the constraint. A special method is derived to estimate the solution of the subproblem. It only involves essentially solving two linear equations of size m, and yet, according to our numerical experiments, gives very good estimates when the solution of the original problem is approached. Finally, some numerical experiments on the proposed algorithm are given in Section 5.

In the sequel, an eigenvalue decomposition of a symmetric matrix means one with decreasing eigenvalues and orthonormal eigenvectors.

2. OPTIMALITY CONDITIONS

We begin with a sufficient condition for (1.3). Given $x \in \mathbb{R}^m$, let $A(x) = U(x)\Lambda(x)U^T(x)$ be an eigenvalue decomposition of $A(x)$, where $\Lambda(x) = \operatorname{diag}(\lambda_1(x), \ldots, \lambda_n(x))$ and $U(x) = [u_1(x) \cdots u_n(x)]$. Given any $\epsilon \geq 0$, define q_ϵ to be the largest integer between 1 and n satisfying $\lambda_{q_\epsilon}(x) \geq \lambda_1(x) - \epsilon$. Also define $\Lambda_\epsilon(x) = \operatorname{diag}(\lambda_1(x), \ldots, \lambda_{q_\epsilon}(x))$ and $U_\epsilon(x) = [u_1(x) \cdots u_{q_\epsilon}(x)]$. We now introduce the notation of ϵ-optimality condition.

DEFINITION 2.1. Given $\epsilon \geq 0$ and $x \in \mathbb{R}^m$. We say that the *ϵ-optimality condition* holds at x if there exists a $q_\epsilon \times q_\epsilon$ (symmetric) positive semidefinite matrix Q satisfying $\operatorname{tr} Q = 1$ and, for $i = 1, \ldots, m$, $\operatorname{tr}[\widehat{A}_i(x)Q] = 0$, where $\widehat{A}_i(x)$ is defined by $\widehat{A}_i(x) = U_\epsilon^T(x)A_iU_\epsilon(x)$. Furthermore, any matrix Q satisfying these properties is called a *multiplier matrix*.

Notice that the multiplier matrix is in general nonunique. The result below shows that, among other things, the ϵ-optimality condition is sufficient for (1.3).

THEOREM 2.1. *Let $\epsilon \geq 0$ and $x \in \mathbb{R}^m$. Suppose that the ϵ-optimal condition holds at x with a multiplier matrix Q. Let $Q = VSV^T$ be an eigenvalue decomposition of Q with $S = \operatorname{diag}(s_1, \ldots, s_{q_\epsilon})$ and $V = [v_1 \cdots v_{q_\epsilon}]$. Then*

$$\lambda_1(x) - \epsilon \leq \lambda_{q_\epsilon}(x) \leq \min_{k=1,\ldots,q_\epsilon} v_k^T \Lambda_\epsilon(x) v_k \leq \lambda^* \leq \lambda_1(x). \tag{2.1}$$

In view of Theorem 2.1, the ϵ-optimality condition may be used as the stopping criterion for any algorithm which looks for $x \in \mathbb{R}^m$ to satisfy (1.3). Although the ϵ-optimality condition is in general not necessary for (1.3), we will show below that if it does not hold, one can always construct a descent direction of $\lambda_1(x)$ at x. It is interesting to notice that the ϵ-optimality condition reduces to the usual optimality condition for (1.1) when ϵ is zero (see, e.g., [29] in a slightly different context on the optimality condition for (1.1)). Moreover, it can be shown that in this case the condition is necessary and sufficient.

Verification of the ϵ-optimality condition using its definition is not an easy task. Difficulty arises because the matrix Q needs to be constrained in the set of

positive semidefinite matrices. Below we give two equivalent characterizations of the ϵ-optimality condition. Theorem 2.2 shows that the ϵ-optimality condition can be equivalently stated as the origin inclusion property of some convex set. Theorem 2.3 shows that the ϵ-optimality condition holds if and only if the solution of some optimization problem in the form of (1.1), possibly with much smaller matrices, is not less than zero. Both results also provide means to construct descent directions when the ϵ-optimality condition fails to hold.

THEOREM 2.2. *Define the set* $W_\epsilon(x) \subset \mathbb{R}^m$ *by*

$$W_\epsilon(x) = \left\{[v^T \widehat{A}_1(x)v \cdots v^T \widehat{A}_m(x)v]^T : v \in \partial B_{q_\epsilon}\right\}$$

where $\widehat{A}_i(x)$ *is defined in Definition* 2.1. *Then the* ϵ*-optimality condition holds if and only if the set* co $W_\epsilon(x)$ *contains the origin. Furthermore, if* co $W_\epsilon(x)$ *does not contain the origin, then* $-\text{nr co } W_\epsilon(x)$ *is a descent direction for* $\lambda_1(x)$ *at* x.

Existing methods for computing nr co $W_\epsilon(x)$ can be found in, e.g., [15, 22]. Notice that it can be obtained analytically when $q_\epsilon \leq 2$ (see [16] in another context).

THEOREM 2.3. *Let* $u \in \partial B_n$, *and define* $z = [u^T \widehat{A}_1(x)u \cdots u^T \widehat{A}_m(x)u]^T$. *Assume* $z \neq 0$. *Define* $z_1 = -z/\|z\|$, *and let* $z_2, \ldots, z_m$ *be such that* $Z = [z_1 \cdots z_m]$ *is orthogonal. For* $i = 1, \ldots, m$, *define* $\widetilde{A}_i(x) = \sum_{j=1}^m z_{ji}\widehat{A}_j(x)$. *Then the* ϵ*-optimality condition holds if and only if*

$$\inf_{y \in \mathbb{R}^{m-1}} \xi_1 \left(\widetilde{A}_1(x) + \sum_{i=1}^{m-1} y_i \widetilde{A}_{i+1}(x) \right) \geq 0.$$

Furthermore, let $y \in \mathbb{R}^{m-1}$ *be such that*

$$\xi_1 \left(\widetilde{A}_1(x) + \sum_{i=1}^{m-1} y_i \widetilde{A}_{i+1}(x) \right) \geq 0. \tag{2.2}$$

Then the vector

$$h = Z \begin{bmatrix} 1 \\ y \end{bmatrix} \tag{2.3}$$

is a descent direction for $\lambda_1(x)$ *at* x.

3. A LINE SEARCH RULE

Let h be a descent direction for $\lambda_1(x)$ at x. A line search rule is to find a step size $s > 0$ such that $\lambda_1(x + sh) < \lambda_1(x)$. Define $\pi(s) = \lambda_1(x + sh)$, and

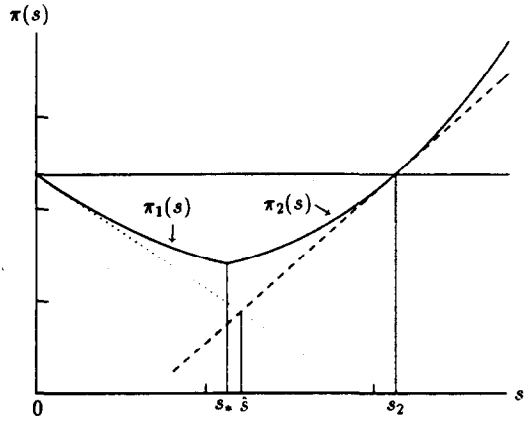

FIG. 2. Definition of $\hat{s}$.

suppose that $\inf_{s>0} \pi(s)$ is achieved at s_*. It is well known that $\pi(s)$ is analytic everywhere except at finitely many points [24]. Without loss of generality, we assume $\pi(s_*) = 0$. Therefore, in the neighborhood of s_*, $\pi(s)$ can be expressed as $\pi(s) = \max\{\pi_1(s), \pi_2(s)\}$, where $\pi_1(s)$ and $\pi_2(s)$ are analytic functions with Taylor series expansions about s_* given by, for $i = 1, 2$,

$$\pi_i(s) = a_i(s - s_*) + b_i(s - s_*)^2 + c_i(s - s_*)^3 + O(|s - s_*|^4). \tag{3.1}$$

Notice that the coefficients in (3.1) are not arbitrary; they satisfy $a_1 a_2 \leq 0$, $b_1 \geq 0$, and $b_2 \geq 0$ (without loss of generality, we may assume $a_1 \leq 0$ and $a_2 \geq 0$).

In this section, given a search direction h, we propose a line search rule. We first compute $s_2 > 0$ that satisfies

$$\pi_1(0) = \pi_2(s_2). \tag{3.2}$$

Then, the step size $\hat{s}$ is defined as the point where the tangent lines of $\pi_1(s)$ at $s = 0$ and $\pi_2(s)$ at $s = s_2$ intersect, i.e., $\hat{s}$ satisfies

$$\hat{s} = \frac{\pi_2'(s_2)s_2}{\pi_2'(s_2) - \pi_1'(0)}. \tag{3.3}$$

See also Figure 2 for an illustration of $\hat{s}$. The following theorem shows that, under some mild assumptions, $\hat{s}$ is a very good estimate of s_*.

THEOREM 3.1. *Suppose* (i) $a_2 > 0$ *or* (ii) $a_1 = a_2 = 0, b_1 > 0$, *and* $b_2 > 0$.

Let $\widehat{s}$ be defined by (3.3). *Then there exists $M > 0$ such that*

$$\widehat{s} - s_* = Ms_*^2 + O(s_*^3). \tag{3.4}$$

The following proposition addresses the question how to find s_2 to satisfy (3.2). This result has also been observed in [38].

PROPOSITION 3.1. *Let h be a descent direction for $\lambda_1(x)$ at x. Define $A = A(x) - \lambda_1(x)I$ and $B = -\sum_{i=1}^m h_i A_i$. Suppose that B is nonsingular, and let $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n$ be the roots of the polynomial* $\det(A - \gamma B)$. *Then $\lambda^* = -\infty$ if $\gamma_1 = 0$. Otherwise, s_2 defined by* (3.2) *is the smallest positive element in the set $\{\gamma_1, \ldots, \gamma_n\}$.*

4. A LOCAL ALGORITHM

The basic idea behind Newton's method on minimizing a smooth convex function is first to construct a second order model around the current estimate. Then the new estimate is defined as the global minimizer of the model. Newton's method cannot be directly applied to minimization of a nondifferentiable function, as the second order model, if it exists, may not approximate the function well in the neighborhood of the current estimate. In the nondifferentiable case, it is necessary that a good model itself be nondifferentiable. In general, $\lambda_1(x)$ is nondifferentiable at x when the corresponding multiplicity is more than one. In this section, we propose a model of $\lambda_1(x)$ around x when x is sufficiently close to x^*. We also propose a method to minimize the model. Under a certain assumption on the multiplicity at the solution (see below), a new algorithm is then derived for the problem (1.1). Numerical experiments show that the proposed algorithm has good convergence behavior. Moreover, it will be seen that the computational complexity at each iteration is not more than a few times that of computing the eigenvalue decomposition of an $n \times n$ matrix.

Given $x \in \mathbb{R}^m$, the multiplicity of $\lambda_1(x)$ is said to be q if $\lambda_1(x) = \lambda_q(x) > \lambda_{q+1}(x)$. Suppose that the solution of (1.1) is bounded, and let x^* be the minimizer. Throughout this section, we assume that the multiplicity of $\lambda_1(x^*)$ is given. In the neighborhood of x^*, by the continuity property of eigenvalues, it holds that $\lambda_q(x) > \lambda_{q+1}(x)$. The following theorem characterizes the problem (1.1).

THEOREM 4.1. *Let q be the multiplicity of $\lambda_1(x)$ at x^*. Then*

$$\lambda^* = \inf_{x \in \mathbb{R}^m} \{f_1(x) : f_2(x) = 0\}, \tag{4.1}$$

where $f_1(x)$ and $f_2(x)$ are defined by

$$f_1(x) = \frac{1}{q}\sum_{i=1}^{q}\lambda_i(x)$$

and

$$f_2(x) = \frac{1}{2}\sum_{i=1}^{q-1}\sum_{j=i+1}^{q}\left[\lambda_i(x) - \lambda_j(x)\right]^2$$

respectively.

Using a result in [14], it can be shown that $f_1(x)$ defined above is convex in x. We also have the following result.

THEOREM 4.2. *Suppose that $\lambda_q(x) > \lambda_{q+1}(x)$ for some q. Then $f_1(x)$ and $f_2(x)$ defined in Theorem* 4.1 *are analytic at x.*

Theorem 4.1 and Theorem 4.2 together reveal that, in a neighborhood of x^*, (1.1) can be equivalently formulated as a *smooth constrained* optimization problem. An immediate approach that comes to mind in solving (4.1) is sequential quadratic programming (SQP). That is, one solves a sequence of QP problems, each of which is constructed using the second order approximation of $f_1(x)$ and the first order approximation of $f_2(x)$. In our case, however, that will not work well, since the first order approximation of $f_2(x)$ is degenerate at all feasible x [i.e., $f_2(x) = 0$]. Instead, we take second order approximations for both $f_1(x)$ and $f_2(x)$ to form the subproblem. The resulting optimization becomes nonconvex, since the constraint set is no longer convex. We will propose a method to estimate the solution of the subproblem. Essentially it only requires solving two linear equations. Yet, according to our numerical experiments, the estimate seems to be very good. Below we derive the gradients and Hessians of $f_1(x)$ and $f_2(x)$.

THEOREM 4.3. *Suppose that $\lambda_q(x^*) > \lambda_{q+1}(x^*)$ for some q. Also, suppose that both $x - x^*$ and $h \in \mathbb{R}^m$ are sufficiently small. Then*

$$f_i(x+h) = f_i(x) + g_i^T(x)h + \tfrac{1}{2}h^T H_i(x)h + O(\|h\|^3), \qquad i = 1, 2,$$

where

$$g_1(x) = \frac{1}{q}\sum_{k=1}^{q} t_k,$$

$$H_1(x) = \frac{1}{q}\sum_{k=1}^{q} T_k,$$

$$
\begin{aligned}
g_2(x) &= \sum_{k,l}^{q} [\lambda_k(x) - \lambda_l(x)]\,(t_k - t_l),\\
H_2(x) &= H_{21}(x) + H_{22}(x),\\
H_{21}(x) &= \sum_{k,l}^{q} \left[(t_k - t_l)(t_k - t_l)^T + 2q t_{kl} t_{kl}^T\right],\\
H_{22}(x) &= \sum_{k,l}^{q} [\lambda_k(x) - \lambda_l(x)]\,(T_k - T_l),\\
t_k &= t_{kk},\\
T_k &= T_{kk},\\
(t_{kl})_i &= u_k^T(x) A_i u_l(x),\\
(T_{kl})_{ij} &= u_k^T(x) A_i S_l A_j u_l(x) + u_l^T(x) A_i S_k A_j u_k(x),\\
S_k &= U_\perp [\lambda_k(x) I - \Lambda_\perp]^{-1} U_\perp^T,\\
U_\perp &= \left[u_{q+1}(x) \cdots u_n(x)\right],\\
\Lambda_\perp &= \operatorname{diag}\left(\lambda_{q+1}(x), \ldots, \lambda_n(x)\right).
\end{aligned}
$$

Here, we denote by $(t)_i$ the ith element of the vector t, and by $(T)_{ij}$ the ijth element of the matrix T. Also, we denote $\sum_{k=1}^{q-1} \sum_{i=k+1}^{q}$ by $\sum_{k,l}^{q}$.

The subproblem we will solve is

$$\min_h g_1^T(x)h + \tfrac{1}{2} h^T H_1(x) h \tag{4.2}$$

subject to the constraint

$$f_2(x) + g_2^T(x)h + \tfrac{1}{2} h^T H_2(x) h = 0. \tag{4.3}$$

We propose to estimate the solution of the subproblem as $h = \beta\widehat{h}$, where $\widehat{h} = h_a + h_b$, and h_a, h_b, and β are computed as follows:

(1) The vector h_a is to reduce the error for the constraint (4.3). Notice that the original constraint $f_2(x + h) = 0$ is the same as $f_2(x + h) \leq 0$. Also, notice that $H_2(x) = H_{21}(x) + H_{22}(x)$, $H_{21}(x) \geq 0$, and if $f_2(x) = 0$ then $H_{22}(x) = 0$ (see Theorem 4.3 above). Therefore, we define $h_a = -H_{21}^{\dagger}(x) g_2(x)$, which is essentially equal to the minimizer of

$$\min_h g_2^T(x)h + \tfrac{1}{2} h^T H_2(x) h.$$

(2) The vector h_b is to reduce the objective function in (4.2) while keeping the constraint (4.3) from changing too much. Notice that in the case of $f_2(x) = 0$

[which implies $H_2(x) = H_{21}(x) \geq 0$], choosing h_b as a linear combination of vectors in the null space of $H_{21}(x)$ will force the error of the constraint to stay on the order of $O(\|h_b\|^3)$. We adopt this idea in defining h_b even when $f_2(x) \neq 0$. Therefore, let the columns of N form a basis of the null space of $H_{21}(x)$. Define $h_b = N_z$ and choose z to minimize

$$g_1^T(x)(h_a + Nz) + \tfrac{1}{2}(h_a + Nz)^T H_1(x)(h_a + Nz).$$

It is easy to show that z is equal to

$$z = -[N^T H_1(x) N]^{-1} N^T [H_1(x) h_a + g_1(x)].$$

Thus,

$$h_b = -N[N^T H_1(x) N]^{-1} N^T [H_1(x) h_a + g_1(x)].$$

(3) We then perform line search along $\widehat{h}$ to ensure monotone decrease of the largest eigenvalue $\lambda_1(x + \beta\widehat{h})$ [not $f_1(x + \beta\widehat{h})$], where β is the step size.

The proposed algorithm is summarized as follows.

ALGORITHM 4.1. *Step 0: Estimate the multiplicity q at solution. Choose initial x^0. Set $k = 0$.*

Step 1: Perform an eigenvalue decomposition of $A(x^k)$. Compute $g_1(x^k)$, $g_2(x^k)$, $H_1(x^k)$, and $H_2(x^k)$ defined in Theorem 4.3. Let the columns of N form the null space of $H_{21}(x^k)$.

Step 2: Compute the search direction $\widehat{h}$:

$$\begin{aligned} h_a &= -H_{21}^{\dagger}(x^k) g_2(x^k), \\ h_b &= -N\left[N^T H_1(x^k) N\right]^{-1} N^T \left[H_1(x^k) h_a + g_1(x^k)\right], \\ \widehat{h} &= h_a + h_b. \end{aligned}$$

Step 3: Perform line search. Set $\beta = 1$. If

$$\lambda_1(x^k) I - A(x^k + \beta\widehat{h}) > 0, \tag{4.4}$$

then set $x^{k+1} = x^k + \beta\widehat{h}$, $k \leftarrow k + 1$, and go to step 1. Otherwise, set $\beta \leftarrow \beta/2$ and repeat the test in (4.4).

The stopping criterion of the algorithm can be based on the size of the search direction $\widehat{h}$. We don't have a convergence analysis of the proposed algorithm yet; this is under investigation. However, our numerical experiments suggest that the proposed algorithm is promising.

REMARK 1. This remark concerns some implementation details.

(1) The matrix $H_{21}(x)$ defined in Theorem 4.3 may be of full rank. In this case, it indicates that the minimizer x^* can be defined solely by the constraint. We then define $\hat{h} = h_a$ in step 2.

(2) The null space of $H_{21}(x)$ can be computed by a QR decomposition.

(3) The matrix $N^T H_1(x) N$ may be singular or nearly singular. In this case, we solve h_b in step 2 in the least squares sense. Alternatively, we may add a small number to its diagonal elements to avoid the singularity.

(4) The matrices $H_1(x)$ and $H_{21}(x)$ need not be formed explicitly. The matrix $H_1(x)$ is positive semidefinite, and its square root factor $H_1^{1/2}(x)$ [i.e, $H_1(x) = H_1^{T/2}(x) H_1^{1/2}(x)$] is readily available. The Cholesky decomposition of $N^T H_1(x)^N$ (in solving h_b) can be obtained by a QR decomposition of $H_1^{1/2}(x)N$. Similar arguments can be applied to the matrix $H_{21}(x)$.

(5) Checking the positive definiteness of a matrix in (4.4) can be done by a Cholesky decomposition.

REMARK 2. This remark concerns the computational effort. Each iteration requires an eigenvalue decomposition of an n by n symmetric matrix [$O(n^3)$ flops], computing various quantities in steps 1 and 2. [$O(qmn^2) + O(m^3)$ flops], and performing line search in step 3. It is observed that the step size is one when approaching to the solution. Also, checking (4.4) by Cholesky decomposition, only needs a small fraction of the number of flops that is needed for an eigenvalue decomposition. The overall computational effort for each iteration then depends polynomially upon the size of the problem.

REMARK 3. This remark concerns the estimated multiplicity of $\lambda_1(x^*)$. When the multiplicity is chosen too small, it is observed that the step size becomes very small after a few iterations. This is quite reasonable, since other eigenvalues may become active along the search direction. When the multiplicity is chosen too large, it is possible for the algorithm to converge, with step size 1, but not to the solution. In this case, the result in Section 2 and some work of Overton's [29] can be used to split the eigenvalues. Methods to guess the true multiplicity at the solution and to update it dynamically are necessary to have a more general algorithm. This is also under investigation.

We shall close this section with the following result.

THEOREM 4.4. *Let the multiplicity q given in Algorithm* 4.1 *be the multiplicity*

of $\lambda_1(x)$. *Also let* $\widehat{h}$ *be defined as in step* 2 *of Algorithm* 4.1, *and assume* $\widehat{h} \neq 0$. *Then* $\widehat{h}$ *is a descent direction for* $\lambda_1(x)$ *at* x.

5. NUMERICAL TESTS

Algorithm 4.1 has been implemented in MATLAB [27]. In this section, we demonstrate it by two numerical examples. The first example uses randomly generated matrices, and the second is borrowed from [29].

EXAMPLE 1. $n = 5$ and $m = 5$. The six matrices are [17]

$$A_0 = \begin{bmatrix} -0.69 & -0.32 & 0.34 & 0.43 & -0.05 \\ -0.32 & -0.11 & -0.11 & -0.45 & -0.34 \\ 0.34 & -0.11 & -0.71 & -0.33 & -0.08 \\ 0.43 & -0.45 & -0.33 & 0.65 & 0.27 \\ -0.05 & -0.34 & -0.08 & 0.27 & 0.39 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} -0.66 & 0.31 & 0.57 & -0.06 & -0.44 \\ 0.31 & -0.23 & -0.12 & -0.35 & 0.28 \\ 0.57 & -0.12 & -0.26 & -0.06 & -0.37 \\ -0.06 & -0.35 & -0.06 & 0.64 & 0.34 \\ -0.44 & 0.28 & -0.37 & 0.34 & 0.61 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} -0.31 & 0.35 & 0.06 & -0.23 & 0.17 \\ 0.35 & 0.24 & -0.19 & 0.21 & -0.12 \\ 0.06 & -0.19 & -0.34 & 0.00 & -0.36 \\ -0.23 & 0.21 & 0.00 & 0.16 & -0.24 \\ 0.17 & -0.12 & -0.36 & -0.24 & 0.00 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 0.27 & -0.14 & 0.13 & -0.32 & -0.08 \\ -0.14 & -0.20 & -0.29 & -0.05 & -0.64 \\ 0.13 & -0.29 & -0.45 & -0.20 & -0.59 \\ -0.32 & -0.05 & -0.20 & -0.27 & -0.46 \\ -0.08 & -0.64 & -0.59 & -0.46 & -0.39 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} -0.57 & -0.38 & -0.09 & 0.31 & 0.22 \\ -0.38 & 0.66 & 0.17 & -0.03 & 0.51 \\ -0.09 & 0.17 & 0.23 & 0.12 & -0.21 \\ 0.31 & -0.03 & 0.12 & -0.56 & -0.21 \\ 0.22 & 0.51 & -0.21 & -0.21 & 0.59 \end{bmatrix},$$

$$A_5 = \begin{bmatrix} 0.22 & 0.28 & 0.14 & 0.03 & 0.09 \\ 0.28 & 0.69 & -0.12 & 0.10 & 0.30 \\ 0.14 & -0.12 & -0.77 & -0.21 & 0.13 \\ 0.03 & 0.10 & -0.21 & -0.42 & -0.15 \\ 0.09 & 0.30 & 0.13 & -0.15 & 0.22 \end{bmatrix},$$

TABLE 1. TEST RESULTS FOR EXAMPLE 1

Iter k	$\lambda_1(x^k)$	β	Iter k	$\lambda_1(x^k)$	β
0	1.204260662	–	10	0.710958694	1.00
1	1.040300888	1.00	11	0.709448235	1.00
2	1.024891334	1.00	12	0.708991355	1.00
3	0.875331883	1.00	13	0.708898846	1.00
4	0.836961646	0.13	14	0.708884719	1.00
5	0.772514993	0.25	15	0.708882872	1.00
6	0.744923566	0.13	16	0.708882634	1.00
7	0.731179722	0.13	17	0.708882602	1.00
8	0.717824578	0.50	18	0.708882598	1.00
9	0.715879585	1.00	19	0.708882597	1.00

First, we apply Kelley's cutting plane method [25] and obtain

$$x^* = \begin{bmatrix} -0.613628903 \\ 0.614530809 \\ -0.343726107 \\ -0.606777187 \\ 0.646451427 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \lambda_1(x^*) \\ \lambda_2(x^*) \\ \lambda_3(x^*) \\ \lambda_4(x^*) \\ \lambda_5(x^*) \end{bmatrix} = \begin{bmatrix} 0.708882597 \\ 0.708882597 \\ 0.061408961 \\ -0.603422793 \\ -1.454430904 \end{bmatrix}.$$

Thus the multiplicity of $\lambda_1(x^*)$ is 2. We then choose the initial guess $x = 0$ and run Algorithm 4.1. The result is summarized in Table 1.

EXAMPLE 2. This example is due to Overton [29]. The problem is to find the diagonal of a 10 by 10 matrix to maximize its largest eigenvalue in absolute value. The off-diagonal elements of the matrix are given (see [29]). The proposed algorithm can be easily modified to take advantage of the sparse structure of various matrices A_i, or to minimize the largest eigenvalue is absolute value. However, as was done in [23], here we simply apply our algorithm with the matrices

$$\tilde{A}_i = \begin{bmatrix} A_i & O \\ O & -A_i \end{bmatrix}, \qquad 0 \le i \le 10.$$

The resulting problem has 11 matrices of size 20. The optimal value λ^* for this problem is 22.3661216, and the associated multiplicity is 3. We set the multiplicity $q = 3$ and run the proposed algorithm. The result is summarized in Table 2. As pointed out by Overton [29], this problem is quite difficult to solve using any

TABLE 2. TEST RESULTS FOR EXAMPLE 1

Iter k	$\lambda_1(x^k)$	β	Iter k	$\lambda_1(x^k)$	β
0	37.6974815	—	11	22.4051515	0.015
1	30.4555145	0.500	12	22.4013878	0.004
2	27.5295344	0.125	13	22.3962849	0.004
3	24.0245400	0.125	14	22.3916726	0.016
4	23.7125700	0.062	15	22.3836499	0.250
5	23.4261984	0.031	16	22.3807503	0.500
6	23.3857604	0.031	17	22.3703151	1.000
7	22.7222438	0.031	18	22.3669657	1.000
8	22.6109781	0.031	19	22.3662088	1.000
9	22.5727067	0.008	20	22.3661253	1.000
10	22.4882827	0.004	21	22.3661216	1.000

algorithm which relies on the multiplicity at the solution. This is because several other eigenvalues are very close to the first three at the solution. We have also experienced some difficulty with this problem. As appears in Table 2, the step sizes of some early iterations are quite small. In fact, reducing the multiplicity at early iterations may give larger step size and consequently may increase the rate of convergence. When the algorithm starts to have step size 1 at the 17th iteration, it improves the accuracy of the result to six digits within four iterations.

APPENDIX

The proof of Theorem 2.1 is facilitated by the following result.

LEMMA A.1. *Let* $c_k \in \mathbb{R}$ *and* $d_k \in \mathbb{R}^m, k = 1, \ldots, q$. *Suppose* $0 \in \operatorname{co}\{d_1, \ldots, d_q\}$. *Then for any* $\widehat{x} \in \mathbb{R}^m$,

$$\min_{x \in \mathbb{R}^m} \max_{k=1,\ldots,q} c_k + \langle d_k, x \rangle \geq \min_{k=1,\ldots,q} c_k + \langle d_k, \widehat{x} \rangle.$$

Proof. It is equivalent to show that, for any $x, \widehat{x} \in \mathbb{R}^m$,

$$\max_{k=1,\ldots,q} c_k + \langle d_k, x \rangle \geq \min_{k=1,\ldots,q} c_k + \langle d_k, \widehat{x} \rangle.$$

Given $x \in \mathbb{R}^m$, define $u(x) = \max_{k=1,\ldots,q}(c_k + \langle d_k, x\rangle)$ and $l(x) = \min_{k=1,\ldots,q}(c_k + \langle d_k, x\rangle)$. Thus, for any $x, \hat{x} \in \mathbb{R}^m$ and $k = 1, \ldots, q$, we have

$$u(x) \geq c_k + \langle d_k, x\rangle \geq c_k + \langle d_k, \hat{x}\rangle - \langle d_k, \hat{x}\rangle + \langle d_k, x\rangle \geq l(\hat{x}) + \langle d_k, x - \hat{x}\rangle.$$

The hypothesis $0 \in \operatorname{co}\{d_1, \ldots, d_q\}$ implies that there exist $\alpha_k \geq 0$, $k = 1, \ldots, q$, such that $\sum_{k=1}^q \alpha_k = 1$ and $\sum_{k=1}^q \alpha_k d_k = 0$. Therefore,

$$u(x) = \sum_{k=1}^q \alpha_k u(x) \geq \sum_{k=1}^q \alpha_k [l(\hat{x}) + \langle d_k, x - \hat{x}\rangle] = l(\hat{x}).$$

■

Proof of Theorem 2.1. The last inequality in (2.1) is obvious. Define $\Omega = \{U_\epsilon(x)v_1, \ldots, U_\epsilon(x)v_{q_\epsilon}\}$. Then

$$\begin{aligned}\lambda^* &= \min_{x\in\mathbb{R}^n} \max_{w\in\partial B} w^H A(x) w \geq \min_{x\in\mathbb{R}^n} \max_{w\in\Omega} w^H A(x) w \\ &= \min_{x\in\mathbb{R}^n} \max_{k=1,\ldots,q_\epsilon} c_k + \langle d_k, x\rangle,\end{aligned}$$

where $c_k = v_k^H \hat{A}_0(x) v_k$ and $d_k = [v_k^H \hat{A}_1(x) v_k \cdots v_k^H \hat{A}_m(x) v_k]^T \in \mathbb{R}^m$ Notice that, for $i = 1, \ldots, m$,

$$0 = \hat{A}(x) : Q = \operatorname{tr}[\hat{A}_i(x) V S V^H] = \operatorname{tr}[V^H \hat{A}_i(x) V S] = \sum_{k=1}^{q_\epsilon} s_k v_k^H \hat{A}_i(x) v_k.$$

Therefore, $\sum_{k=1}^{q_\epsilon} s_k d_k = 0$, i.e., $0 \in \operatorname{co}\{d_1, \ldots, d_{q_\epsilon}\}$. In view of Lemma A.1, we have

$$\begin{aligned}\lambda^* &\geq \min_{k=1,\ldots,q_\epsilon} c_k + \langle d_k, x\rangle = \min_{k=1,\ldots,q_\epsilon} v_k^H U_\epsilon^H(x) U(x) \Lambda(x) U(x)^H U_\epsilon(x) v_k \\ &= \min_{k=1,\ldots,q_\epsilon} v_k^H \Lambda_\epsilon(x) v_k \geq \lambda_{q_\epsilon}(x) \leq \lambda_1(x) - \epsilon.\end{aligned}$$

Above, the next to the last inequality holds because that $v_k^H \Lambda_\epsilon(x) v_k$ is a convex combination of the diagonal elements in $\Lambda_\epsilon(x)$. ■

Proof of Theorem 2.2. The first claim follows from the proof of Theorem 2.1. It has been shown that $\operatorname{co} W_0(x)$ is the generalized gradient of $\lambda_1(x)$ (see, e.g., [11, 33]). Thus, if $\operatorname{co} W_0(x)$ does not contain the origin, then the vector $-\operatorname{nr} \operatorname{co} W_0(x)$ is the steepest descent direction for $\lambda_1(x)$. Notice that $W_{\epsilon_1}(x) \subset W_{\epsilon_2}(x)$ and therefore $\operatorname{co} W_{\epsilon_1}(x) \subset \operatorname{co} W_{\epsilon_2}(x)$ whenever

$0 \leq \epsilon_1 \leq \epsilon_2$. Using properties of convex sets, the second claim follows. ■

Proof of Theorem 2.3. First, we show that first claim.
Sufficiency: Define

$$\widetilde{W}(x) = \{[v^H \widetilde{A}_1(x)v \cdots v^H \widetilde{A}_m(x)v]^T : v \in \partial B_{q_\epsilon}\}$$

Given $v \in \partial B_{q_\epsilon}$, it is easy to check that

$$[v^H \widetilde{A}_1(x)v \cdots v^H \widetilde{A}_m(x)v] = [v^H \widehat{A}_1(x)v \cdots v^H \widehat{A}_m(x)v]Z.$$

Therefore,

$$\operatorname{co} \widetilde{W}(x) = Z^T \operatorname{co} W_\epsilon(x). \tag{A.1}$$

By the geometric interpretation given in Section 1, $\widetilde{\lambda}$ is equal to the maximal value of the intersection of $\operatorname{co} \widetilde{W}(x)$ and the first coordinate axis. In view of (A.1) and since Z is orthogonal, it holds that $\widetilde{\lambda}_{z_1} \in \operatorname{co} W_\epsilon(x)$. On the other hand, by the definition of z, we have $z \in \operatorname{co} W_\epsilon(x)$. Since $\widetilde{\lambda} \geq 0$, it is easy to check that the origin lies between z and $\widetilde{\lambda}_{z_1}$. Hence $0 \in \operatorname{co} W_\epsilon(x)$. In view of Theorem 2.2, the ϵ-optimality condition holds.

Necessity: Follow the discussion given in the sufficiency part; $\widetilde{\lambda} < 0$ implies $0 \notin \operatorname{co} \widetilde{W}(x)$. Hence $0 \notin \operatorname{co} W_\epsilon(x)$. In view of Theorem 2.2, the ϵ-optimality condition doesn't hold.

Second, we show that the vector h defined by (2.3) is a descent direction for $\lambda_1(x)$. It suffices to show $w^T h < 0$ for any $w \in W_0(x)$. In view of (2.2), it holds that, for any $v \in \partial B_{q_\epsilon}$,

$$v^H \left(\widetilde{A}_1(x) + \sum_{i=1}^{m-1} y_i \widetilde{A}_{i+1}(x) \right) v < 0$$

i.e., $[v^H \widehat{A}_1 v \cdots v^H \widehat{A}_m v]h < 0$. Since $\operatorname{co} W_0(x) \subset \operatorname{co} W_\epsilon(x)$, we then have $w^T h < 0$ for all $w \in W_0(x)$. ■

Proof of Theorem 3.1. Define $t_1 = -s_*$ and $t_2 = s_2 - s_*$. It can be checked that under the given assumptions, t_2 can be expressed as

$$t_2 = \alpha t_1 + \beta t_1^2 + O(t_1^3) \tag{A.2}$$

for some $\alpha < 0$ and β. Assume $a_2 > 0$. Substitute (A.2) into (3.2) to obtain

$$a_1 t_1 + b_1 t_1^2 = \pi_1(0) + O(t_1^3)$$

$$
\begin{aligned}
&= \pi_2(s_2) + O(t_1^3) \\
&= a_2 t_2 + b_2 t_2^2 + O(t_2^3) + O(t_1^3) \\
&= a_2(\alpha t_1 + \beta t_1^2) + b_2(\alpha t_1 + \beta t_1^2)^2 + O(t_1^3) \\
&= a_2\alpha t_1 + (a_2\beta + b_2\alpha^2)t_1^2 + O(t_1^3).
\end{aligned}
$$

Therefore, we have $a_1 = a_2\alpha$ and $b_1 = a_2\beta + b_2\alpha^2$, i.e.,

$$
\begin{aligned}
\alpha &= a_1 a_2^{-1}, \\
\beta &= (b_1 a_2^2 - b_2 a_1^2) a_2^{-3}.
\end{aligned}
$$

Substituting (3.2) and the above expressions for α and β into (3.3), we have

$$
\widehat{s} - s_* = \frac{b_1 a_2^2 - b_2 a_1^2}{a_2^2(a_1 - a_2)} s_*^2 + O(s_*^3).
$$

Thus (3.4) holds with

$$
M = \frac{b_1 a_2^2 - b_2 a_1^2}{a_2^2(a_1 - a_2)}.
$$

Now we assume that $a_1 = a_2 = 0, b_1 > 0$, and $b_2 > 0$. Again, substitute (A.2) into (3.2) and obtain

$$
\begin{aligned}
b_1 t_1^2 + c_1 t_1^3 &= \pi_1(0) + O(|t_1|^4) \\
&= \pi_2(s_2) + O(|t_1|^4) \\
&= b_2 t_2^2 + c_2 t_2^3 + O(|t_2|^4) + O(|t_1|^4) \\
&= b_2(\alpha t_1 + \beta t_1^2)^2 + c_2(\alpha t_1 + \beta t_1^2)^3 + O(|t_1|^4) \\
&= b_2\alpha^2 t_1^2 + (2b_2\alpha\beta + c_2\alpha^3)t_1^3 + O(|t_1|^4).
\end{aligned}
$$

Therefore, we have $b_1 = b_2\alpha^2$ and $c_1 = 2b_2\alpha\beta + c_2\alpha^3$, i.e.,

$$
\begin{aligned}
\alpha &= -\sqrt{b_1 b_2^{-1}}, \\
\beta &= -\tfrac{1}{2}\left(c_1\sqrt{b_1 b_2}^{-1} + c_2 b_1 b_2^{-2}\right).
\end{aligned}
$$

Substituting (3.2) and the above expressions for α and β into (3.3), we have

$$
\widehat{s} - s_* = \frac{c_1 b_2\sqrt{b_2} + c_2 b_1\sqrt{b_1}}{2b_2\sqrt{b_2 b_1}(\sqrt{b_1} + \sqrt{b_2})} s_*^2 + O(s_*^3)
$$

Thus, (3.4) also holds with

$$
M = \frac{c_1 b_2\sqrt{b_2} + c_2 b_1\sqrt{b_1}}{2b_2\sqrt{b_2 b_1}(\sqrt{b_1} + \sqrt{b_2})}.
$$

■

Proof of Proposition 3.1. The claim that the polynomial $\det(A - \gamma B)$ has n real roots follows from the assumptions that A, B are Hermitian, B is nonsingular, and $A \leq 0$. Also, since $\det A = 0$, zero is then one of the roots. Notice that, if s_2 exists, it will satisfy $\lambda_1(x + s_2 h) = \lambda_1(x)$. Therefore,

$$\det\left(A(x) + \sum_{i=1}^{m} s_2 h_i A_i - \lambda_1(x) I\right) = \det(A - s_2 B) = 0,$$

i.e., s_2 is equal to a positive element in $\{\gamma_1, \ldots, \gamma_n\}$. Thus $\gamma_1 = 0$ implies that s_2 does not exist. On the other hand, using the assumption that h is a descent direction, it can be easily checked that s_2 does exist if $\gamma_1 > 0$. Now, suppose that $\gamma_1 = 0$. Then using the assumptions that B is nonsingular and (again) h is a descent direction, it follows easily that $\lim_{s \to \infty} \lambda_1(x + sh) = -\infty$. Therefore, $\lambda^* = -\infty$. To conclude the result, it remains to show that, if $\gamma_1 > 0$, then s_2 is equal to the smallest positive element in $\{\gamma_1, \ldots, \gamma_n\}$. Notice that, since $\lambda_1(x + s_2 h) = \lambda_1(x)$, we have $A - s_2 B \leq 0$, i.e., $\phi^H(A - s_2 B)\phi \leq 0$ for all $\phi \in \partial B_n$. For $i = 1, \ldots, n$, let ϕ_i be the unit length generalized eigenvector corresponding to γ_i. Then, for $i = 1, \ldots, n$, we have $\phi_i^H A \phi_i - \gamma_i \phi_i^H B \phi_i = 0$ and $\phi_i^H A \phi_i - s_2 \phi_i^H B \phi_i \leq 0$. In view of the fact that $A \leq 0$, this implies that, when $\gamma_i > 0$, we have $\phi_i^H B \phi_i < 0$ and $s_2 \leq \gamma_i$. ■

Proof of Theorem 4.1. It is obvious by using the definition of the multiplicity q of $\lambda_1(x)$ at x^*. ■

Let $P = \{s_1, \ldots, s_q\}$ be a partition of $\{1, \ldots, n\}$, i.e., P is a collection of nonempty, pairwise disjoint subsets whose union is $\{1, \ldots, n\}$. A function $\eta : \mathbb{R}^n \to \mathbb{R}$ is said to be *symmetric w.r.t. a partition P* if the value of η is invariant under permutations of its arguments within the groups $s_1, \ldots, s_q$. For any $a \in \mathbb{R}^m$, we define the partition $P_{\lambda(a)}$ to be the partition that groups the indices i for which $\lambda_i(a)$ are of a same value. We will employ the following result in proving Theorem 4.2.

Fact A.2. [37] Suppose that $\eta : \mathbb{R}^n \to \mathbb{R}$ is analytic and is symmetric w.r.t. $P_{\lambda(a)}$ for some $a \in \mathbb{R}^m$. Then the composite function $f(x) = \eta(\lambda_1(x), \ldots, \lambda_n(x))$ is analytic at a.

Proof of Theorem 4.2. Let $\eta_1(z), \eta_2(z) : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$\eta_1(z) = \frac{1}{q} \sum_{i=1}^{q} z_i \quad \text{and} \quad \eta_2(z) = \frac{1}{2} \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} (z_i - z_j)^2$$

respectively. Therefore, $f_i(x) = \eta_i(\lambda_1(x), \ldots, \lambda_n(x))$, $i = 1, 2$. Using the definition of the multiplicity q of $\lambda_1(x)$ at x^*, it is straightforward to check that both $\eta_1(z)$ and $\eta_2(z)$ are analytic and are symmetric w.r.t. the partition $P_{\lambda(x^*)}$. The claim then follows from Fact A.2. ■

Proof of Theorem 4.3. It is straightforward to verify the expressions by using a result in [37]. ■

Proof of Theorem 4.4. It is easily checked that $\widehat{h}$ is a descent direction for $\lambda_1(x)$ at x when $q = 1$. Thus, assume $q > 1$. Let $g_i(x)$ and $H_i(x)$, $i = 1, 2$, be defined as in Theorem 4.3, and let h_a and h_b be defined as in Algorithm 4.1. Under the assumption of the multiplicity, it is easy to see that $f_2(x) = 0$, $g_2(x) = 0$, $h_a = 0$, $H_2(x)\widehat{h} = 0$, $g_1^T(x)\widehat{h} = g_1^T(x)h_b = -h_b^T H_1(x)h_b$, and $N^T H_1(x)h_b = -N^T g_1(x)$. Then

$$f_2(x + \beta\widehat{h}) = \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} [\lambda_i(x + \beta\widehat{h}) - \lambda_j(x + \beta\widehat{h})]^2$$

$$= f_2(x) + \beta g_2^T(x)\widehat{h} + \frac{\beta^2}{2}\widehat{h}^T H_2(x)\widehat{h} + O(\beta^3) = O(\beta^3).$$

Since $h_b \neq 0$ (by assumption) and $H_1(x) \geq 0$, in view of the definitions of h_b and N, it holds that $g_1^T(x)\widehat{h} < 0$. Let $M = -g_1^T(x)\widehat{h} > 0$. In view of the property of $O(\cdot)$, there exists $\beta_1 > 0$ such that

$$f_2(x + \beta\widehat{h}) \leq \frac{M^2\beta^2 q^2}{9(q-1)} \qquad \forall\, \beta \in (0, \beta_1).$$

Thus,

$$\lambda_1(x + \beta\widehat{h}) - \frac{1}{q}\sum_{i=1}^{q}(x + \beta\widehat{h})$$

$$= \frac{1}{q}\sqrt{\left(\sum_{i=2}^{q}\left[\lambda_1(x + \beta\widehat{h}) - \lambda_i(x + \beta\widehat{h})\right]\right)^2}$$

$$\leq \frac{1}{q}\sqrt{(q-1)\sum_{i=2}^{q}[\lambda_1(x + \beta\widehat{h}) - \lambda_i(x + \beta\widehat{h})]^2}$$

$$\leq \frac{1}{q}\sqrt{(q-1)\frac{M^2\beta^2 q^2}{9(q-1)}} = \tfrac{1}{3}M\beta \qquad \forall \beta \in (0, \beta_1). \quad \text{(A.3)}$$

Above, to obtain the first inequality, we use the fact that any $c \in \mathbb{R}^k$ satisfies $\|c\|_1 \leq \sqrt{k}\|c\|_2$, where $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the l_1 and l_2 norms, respectively. On the other hand, we also have

$$f_1(x+\beta\hat{h}) = \frac{1}{q}\sum_{i=1}^{q}\lambda_i(x+\beta\hat{h}) = f_1(x) + \beta g_1{}^T(x)\hat{h} + O(\beta^2)$$
$$= \lambda_1(x) - M\beta + O(\beta^2),$$

i.e., $(1/q)\sum_{i=1}^{q}\lambda_i(x+\beta\hat{h}) - \lambda_1(x) + M\beta = O(\beta^2)$. In view of the property of $O(\cdot)$ again, there exists $\beta_2 > 0$ such that

$$\frac{1}{q}\sum_{i=1}^{q}\lambda_i(x+\beta\hat{h}) - \lambda_1(x) + M\beta \leq \tfrac{1}{3}M\beta \qquad \forall\,\beta \in (0,\beta_2), \tag{A.4}$$

Let $\beta_0 = \min\{\beta_1, \beta_2\}$. Adding (A.3) and (A.4) together yields

$$\lambda_1(x+\beta\hat{h}) - \lambda_1(x) \leq -M\beta + \tfrac{1}{3}M\beta + \tfrac{1}{3}M\beta = -\tfrac{1}{3}M\beta < 0 \qquad \forall\,\beta \in (0,\beta_0).$$

Hence, $\hat{h}$ is a descent direction for $\lambda_1(x)$ at x. ∎

The authors wish to thank Dr. N. K. Tsing, Dr. A. L. Tits, and the anonymous referee for many helpful suggestions.

REFERENCES

1 F. Alizadeh, Optimization over the positive definite cone: Interior-point methods and combinatorial applications, in *Advances in Optimization and Parallel Computing* (Panos Pardalos, Ed.), North-Holland, 1991.

2 F. Alizadeh, Combinatorial optimization with semidefinite matrices, in *Proceedings of Second Annual Integer Programming and Combinatorial Optimization Conference*, Carnegie-Mellon Univ., 1991.

3 J. C. Allwright, LQP: Dominant output feedback, *IEEE Trans. Automat. Control* 27:915–921 (1982).

4 J. C. Allwright, On maximizing the minimum eigenvalue of a linear combination of symmetric matrices, *SIAM J. Matrix Anal. Appl.* 10(3):347–382 (July 1989).

5 Y.-H. Au-Yeung and N.-K. Tsing, An extension of the Hausdorff-Toeplitz theorem on the numerical range, *Proc. Amer. Math. Soc.* 89:215–218 (1983).

6 J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, American Elsevier, 1973.

7 S. P. Boyd and C. H. Barratt, *Linear Controller Design: Limits of Performance*, Prentice-Hall, 1991.

8 S. P. Boyd and L. El Ghaoui, Method of centers for minimizing generalized eigenvalues, *Linear Algebra Appl.*, to appear.
9 S. P. Boyd and Q. Yang, Structured and simultaneous Lyapunov functions for system stability problems, *Internat J. Control* 49(6):2215–2240 (1990).
10 J. Chen, C. N. Nett, and M. K. H. Fan, Worst-case system identification in H_∞: Essentially optimal algorithms, error bounds, and validation of a priori information, in *Proceedings of 1992 American Control Conference*, Chicago, June 1992, pp. 251–257.
11 F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, 1983.
12 J. Cullum, W. E. Donath, and P. Wolfe, The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices, *Math. Programming Study* 3:35–55 (1975).
13 W. E. Donath and A. J. Hoffman, Algorithms for partitioning graphs and computer logic based on eigenvectors of connection matrices, *IBM Tech. Disclosures Bull.* 15 (1972).
14 W. E. Donath and A. J. Hoffman, Lower bounds for the partitioning of graphs, *IBM J. Res. Develop.* 17 (1973).
15 J. C. Doyle, Analysis of feedback systems with structured uncertainties, *Proc. IEE—D* 129(6):242–250 (1982).
16 M. K. H. Fan, An Algorithm to Compute the Structured Singular Value, Technical Report, SRC TR-86-8, Systems Research Center, Univ. of Maryland, 1986.
17 M. K. H. Fan and B. Nekooie, On minimizing the largest eigenvalue of a symmetric matrix, in *Proceedings of 31th IEEE Conference on Decision and Control*, Tucson, Ariz., Dec. 1992, pp. 134–139.
18 M. K. H. Fan and A. L. Tits, A measure of worst-case H_∞ performance and of largest acceptable uncertainty, *Systems Control Lett.* 18:409–421 (June 1992).
19 M. K. H. Fan, A. L. Tits, and J. C. Doyle, Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics, *IEEE Trans. Automat. Control* 36(1):25–38 (1991).
20 R. Fletcher, Semi-definite matrix constraints in optimization, *SIAM J. Control Optim.* 23:493–513 (1985).
21 F. Hausdorff, Der Wertvorrat einer bilinearform, *Math. Z.* 3:314–316 (1919).
22 J. E. Hauser, Proximity Algorithms: Theory and Implementation, Memo UCB/ERL M86/53, Electronics Research Lab. Univ. of California, Berkeley, Calif., May 1986.
23 F. Jarre, An Interior-Point Method for Minimizing the Maximum Eigenvalue of a Linear Combination of Matrices, Technical Report SCL 91–8, June 1991.
24 T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed. Springer, Berlin, 1976.
25 J. E. Kelley, The cutting-plane method for solving convex programs, *J. Soc. Indust. Appl. Math.* 8(2):703–712 (Dec. 1960).
26 L. Lovasz, *An Algorithmic Theory of Numbers, Graphs, and Convexity*, CBMS 50, SIAM, Philadelphia, 1986.
27 C. Moler, J. Little, and S. Bangert, *PRO-METLAB User's Guide*, The Math Works, Inc., Sherborn, Mass., 1987.
28 Y. E. Nesterov and A. S. Nemirovsky, *Self-Concordant Functions and Polynomial Time Methods in Convex Programming*, Moscow, 1990.
29 M. L. Overton, On minimizing the maximum eigenvalue of a symmetric matrix,

SIAM J. Matrix Anal. Appl. 9:256–268 (1988).
30 M. L. Overton, Large-scale optimization of eigenvalues, *SIAM J. Optim.* 2:88–120 (1992).
31 M. L. Overton and R. S. Womersley, On the sum of the largest eigenvalues of a symmetric matrix, *SIAM J. Matrix Anal. Appl.*, 1992, pp. 41–45.
32 E. R. Panier, On the need for special purpose algorithms for minimax eigenvalue problems, *J. Optim. Theory Appl.* 62(2):279–287 (Aug. 1989).
33 E. Polak and Y. Wardi, Nondifferentiable optimization algorithm for the design of control systems subject to singular value inequalities, *Automatica* 18(3):267–283 (1982).
34 U. T. Ringertz, Optimal Design of Nonlinear Shell Structures, Report, Flygtekniska Forsoksantalten, Aeronautical Research Inst. of Sweden, 1991.
35 M. G. Safonov, Stability margins of diagonally perturbed multivariable feedback systems, *Proc. IEE—D* 129:251–256 (1982).
36 A. Shapiro, Extremal problems on the set of nonnegative definite matrices, *Linear Algebra Appl.* 67:7–18 (1985).
37 N.-K. Tsing, M. K. H. Fan, and E. I. Verriest, On analyticity of functions involving eigenvalues, *Linear Algebra Appl.*, to appear.
38 P. Young, private communications, 1991.

Received 28 April 1992; final manuscript accepted 25 March 1993