Journal of Global Optimization **30:** 253–270, 2004.
© 2004 *Kluwer Academic Publishers. Printed in the Netherlands.*

Smooth Convex Approximation to the Maximum Eigenvalue Function

XIN CHEN[1], HOUDUO QI[3], LIQUN QI[2] and KOK-LAY TEO[2]
[1]*Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, 224 Mechanical Engineering Building, mc-244, 1206 West Green Street, Urbana, IL 61801; Corresponding author: E-mail: xinchen@uiuc.edu;*
[2]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong; E-mail: {maqilq, mateokl }@polyu.edu.hk*
[3]*School of Mathematics, University of Southampton, Highfield, Southampton, S017 1BJ, Great Britain; E-mail: hdqi@soton.ac.uk*

(Received 17 December 2002; accepted in revised form 6 January 2003)

Abstract. In this paper, we consider smooth convex approximations to the maximum eigenvalue function. To make it applicable to a wide class of applications, the study is conducted on the composite function of the maximum eigenvalue function and a linear operator mapping $\mathbb{R}^m$ to $\mathcal{S}_n$, the space of n-by-n symmetric matrices. The composite function in turn is the natural objective function of minimizing the maximum eigenvalue function over an affine space in $\mathcal{S}_n$. This leads to a sequence of smooth convex minimization problems governed by a smoothing parameter. As the parameter goes to *zero*, the original problem is recovered. We then develop a computable Hessian formula of the smooth convex functions, matrix representation of the Hessian, and study the regularity conditions which guarantee the nonsingularity of the Hessian matrices. The study on the well-posedness of the smooth convex function leads to a regularization method which is globally convergent.

AMS subject classifications: 49M45, 90C25, 90C33

Key words: Matrix representation, spectral function, Symmetric function, Tikhonov regularization

1. Introduction

Let $\mathcal{S}_n$ denote the space of n-by-n real symmetric matrices endowed with the inner product $\langle X, Y\rangle := \mathrm{tr}(XY)$ for any $X, Y \in \mathcal{S}_n$. The maximum eigenvalue function is often defined as the first component of the eigenvalue function $\lambda : \mathcal{S}_n \to \mathbb{R}^n$, where for any $X \in \mathcal{S}_n$, $\lambda(X)$ is the vector of eigenvalues of X in nonincreasing order, i.e., $\lambda_1(X) \geqslant \lambda_2(X) \geqslant \cdots \geqslant \lambda_n(X)$. The minimization of the maximum eigenvalue function over various sets gives rise to probably the most important class of eigenvalue optimization problems, see Lewis and Overton (1996). In particular, the following problem which is to minimize the maximum eigenvalue function in an affine subspace of $\mathcal{S}_n$ is an equivalent reformulation of the semidefinite programming relaxation of some combinatorial problems. Let $A_0, A_1, \ldots, A_m \in \mathcal{S}_n$

be given, and define an operator $\mathcal{A}:\mathbb{R}^m \to \mathcal{S}_n$ by

$$\mathcal{A}y := \sum_{i=1}^{m} y_i A_i, \qquad \forall y \in \mathbb{R}^m.$$

Then the basic eigenvalue optimization problem we mentioned is given by

$$\inf_{y\in\mathbb{R}^m} \lambda_1(A(y)) \tag{1}$$

where

$$A(y) := A_0 + \mathcal{A}y \tag{2}$$

For more discussion on the problem, see Lewis and Overton (1996), Oustry (1999) and Shapiro and Fan (1995).

It is well known that the eigenvalue function is usually not differentiable, which inevitably gives rise to extreme difficulties in extending classical optimization methods (which often make use of information of the gradient and Hessian of objective functions) to eigenvalue optimization problems, see Overton and Womersley (1995). Pioneering works conducted recently by Lewis within a very general framework of spectral functions open ways in such extensions. A spectral function is usually defined as a composite function of a symmetric function $f:\mathbb{R}^n \to \mathbb{R}$ and the eigenvalue function $\lambda(\cdot)$. A function $f:\mathbb{R}^n \to \mathbb{R}$ is symmetric if f is invariant under coordinate permutations, i.e., $f(P\mu) = f(\mu)$ for any $\mu \in \mathbb{R}^n$ and $P \in \mathcal{P}$, the set of all permutation matrices. Hence the spectral function defined by f and λ can be written as $(f \circ \lambda):\mathcal{S}_n \to \mathbb{R}$ with $(f \circ \lambda)(X) = f(\lambda(X))$ for any $X \in \mathcal{S}_n$. It seems that the spectral function, thought as a composition of $\lambda(\cdot)$ and a symmetric function, would inherit the nonsmoothness of the eigenvalue function. However, according to the study of Lewis (1996), the smoothness of a spectral function depends only on the smoothness of f. In fact, it is proved that $(f \circ \lambda)$ is differentiable if and only if f is differentiable (differentiability means Fréchet differentiability throughout the paper.) Furthermore, Lewis and Sendov (2001) proved that $(f \circ \lambda)$ is twice (continuously) differentiable if and only if f is twice (continuously) differentiable.

Although most of known spectral functions are not twice differentiable (not even differentiable), we will soon see that the smooth approximation denoted by f_ε ($\varepsilon > 0$ in general) to the nonsmooth function f gives rise to a smooth spectral function $(f_\varepsilon \circ \lambda)$, which is a smooth approximation to the spectral function $(f \circ \lambda)$. For example, let

$$f(x) := \max\{x_1, \ldots, x_n\} \tag{3}$$

Then

$$\lambda_1(X) = (f \circ \lambda)(X), \quad \forall X \in \mathcal{S}_n.$$

A well known smoothing function to the maximum function (3) is the exponential penalty function:

$$f_\varepsilon(x) := \varepsilon \ln\left(\sum_{i=1}^{n} e^{x_i/\varepsilon}\right), \qquad \varepsilon > 0. \tag{4}$$

It is a C^∞ convex function and has the following uniform approximation to f, see Chang (1980):

$$0 \leqslant f_\varepsilon(x) - f(x) \leqslant \varepsilon \ln n.$$

The penalty function, sometimes called the aggregation function, has been used in a number of settings: Ben-Tal and Teboulle (1989), Bertsekas (1982), Goldstein (1997), Li (1991), Peng and Lin (1999), Qi and Liao (1999), Qi and Tseng (2002), Tang and Zhang (1994), and Tseng and Bertsekas (1993). In particular, Ben-Tal and Teboulle (1989) studied this function together with its recession function (3). Further studies along this line were conducted recently by Auslender (1999) in a more general framework. In fact, this function is one of nonseparable penalty functions in Auslender (1999).

It is easy to see that the exponential penalty function is symmetric in $\mathbb{R}^n$ and the well defined spectral function $(f_\varepsilon \circ \lambda)$ is a uniform approximation to $\lambda_1(\cdot)$, i.e.,

$$0 \leqslant (f_\varepsilon \circ \lambda)(X) - \lambda_1(X) \leqslant \varepsilon \ln n, \qquad \forall\, \varepsilon > 0,\ X \in \mathcal{S}_n. \tag{5}$$

Moreover, $(f_\varepsilon \circ \lambda)$ is twice continuously differentiable. It remains to consider its composition with the affine mapping $A(\cdot)$. For $\varepsilon > 0$, define $\theta_\varepsilon : \mathbb{R}^m \to \mathbb{R}$ by

$$\theta_\varepsilon(y) := (f_\varepsilon \circ \lambda \circ A)(y) = f_\varepsilon(\lambda(A(y))), \qquad \forall\, y \in \mathbb{R}^m.$$

The uniform approximation (5) implies

$$\lim_{\varepsilon \to 0} \theta_\varepsilon(y) = \lambda_1(A(y)), \qquad \forall\, y \in \mathbb{R}^m.$$

Therefore it is natural to consider the following convex minimization problem with twice continuously differentiable objective function:

$$\min_{y \in \mathbb{R}^m} \theta_\varepsilon(y). \tag{6}$$

Hence we obtain a sequence of smooth functions which uniformly approximates the maximum eigenvalue function with the accuracy controlled by a smoothing parameter. As the parameter goes to *zero*, the function $\lambda_1(\cdot)$ is recovered. The efficiency of numerical methods for solving (6) depends on the availability and the efficient computation of its gradient and Hessian. A related question is the well-posedness of the problem (6) as defined in the book by Dontchev and Zolezzi (1993). This leads us to consider a regularized minimization method for solving (1).

The paper is organized as follows: In Section 2, we derive formulae for the gradient as well as for the Hessian of function θ_ε, and study their properties. In particular, we address the regularity conditions ensuring the nonsingularity of the Hessian. We also develop a more explicit matrix representation of the Hessian. The study on the well-posedness of the problem (6) leads to a regularization method for solving (1) with convergence analysis in Section 3. Finally, conclusions are drawn in Section 4. Frequently used basic gradient and Hessian formulae for continuously differentiable spectral functions due to Lewis (1996) and Lewis and Sendov (2001) are included in Section 5 as an Appendix.

Notation: Vectors in $\mathbb{R}^n$ are viewed as columns and capital letters such as X, Y et.al. always denote matrices in $\mathcal{S}_n$. For $X \in \mathcal{S}_n$, we denote by X_{ij} the (i,j)th entry of X. We use $\circ$ to denote the Hadamard product between matrices, i.e.,

$$X \circ Y = [X_{ij} Y_{ij}]_{i,j=1}^n.$$

Let the operator $\mathrm{diag}: \mathcal{S}_n \to \mathbb{R}^n$ be defined by $\mathrm{diag}[X] := (X_{11}, \ldots, X_{nn})^T$, while for $\mu \in \mathbb{R}^n$, $\mathrm{Diag}[\mu_1, \ldots, \mu_n]$ will denote the diagonal matrix with its ith diagonal entry μ_i. Sometimes we write $\mathrm{Diag}[\mu]$ instead of $\mathrm{Diag}[\mu_1, \ldots, \mu_n]$ for simplicity. Throughout, $\|\cdot\|$ denotes the Frobenius norm for matrices and the 2-norm for vectors. We let $\mathcal{A}^*: \mathcal{S}_n \to \mathbb{R}^m$ be the adjoint operator of the linear operator $\mathcal{A}: \mathbb{R}^m \to \mathcal{S}_n$ defined by (2) and satisfies for all $(d, D) \in \mathbb{R}^m \times \mathcal{S}_n$

$$d^T \mathcal{A}^* D := \langle D, \mathcal{A} d \rangle.$$

Hence, for all $D \in \mathcal{S}_n$,

$$\mathcal{A}^* D = (\langle A_1, D \rangle, \ldots, \langle A_m, D \rangle)^T.$$

2. Smooth convex minimization problems

Through the rest of the paper, we let f be the maximum function defined by (3), f_ε be the penalty function defined by (4), and θ_ε be the merit function in the convex minimization problem (6). The primary task in this section is to derive explicit formulae for the gradient and the Hessian of θ_ε which are used by many numerical methods for solving (6). We then discuss the regularity conditions ensuring the nonsingularity of the Hessian of θ_ε. Finally, we give a matrix representation for the Hessian.

2.1. BASIC CALCULATION

The following result is specialization of the results of [19, Lemma 7] and [22, Prop. 3.1] on a general form of f_ε to the maximum function (3), and can also be calculated directly.

LEMMA 2.1. *The following results hold*:

(*a*) *For any* $\varepsilon_1 > \varepsilon_2 > 0$, $f_{\varepsilon_1}(x) \geqslant f_{\varepsilon_2}(x)$ *for all* $x \in \mathbb{R}^n$.

(*b*) *The gradient of* f_ε *is given by*

$$\nabla f_\varepsilon(x) = \mu(\varepsilon, x) := (\mu_1(\varepsilon, x), \ldots, \mu_n(\varepsilon, x))^T \tag{7}$$

where for $i = 1, \ldots, n$

$$\mu_i(\varepsilon, x) = \frac{e^{x_i/\varepsilon}}{\sum_{j=1}^{n} e^{x_j/\varepsilon}} \in (0, 1] \quad \textit{and} \quad \sum_{i=1}^{n} \mu_i(\varepsilon, x) = 1. \tag{8}$$

(*c*) *The Hessian of* f_ε *is given by*

$$\nabla^2 f_\varepsilon(x) = \frac{1}{\varepsilon}\left(\mathrm{Diag}[\mu(\varepsilon, x)] - \mu(\varepsilon, x)(\mu(\varepsilon, x))^T\right). \tag{9}$$

Let $\mathcal{O}$ denote the group of $n \times n$ real orthogonal matrices. For each $X \in \mathcal{S}_n$, define the set of orthonormal eigenvectors of X by

$$\mathcal{O}_X := \{P \in \mathcal{O} |\ P^T X P = \mathrm{Diag}[\lambda(X)]\}.$$

Clearly $\mathcal{O}_X$ is nonempty for each $X \in \mathcal{S}_n$. It follows Proposition 4.1 that the spectral function $(f_\varepsilon \circ \lambda)$ is continuously differentiable in $\mathcal{S}_n$. We now relate the gradient of $(f_\varepsilon \circ \lambda)$ to the subdifferential of $\lambda_1(\cdot)$ (it is well known that $\lambda_1(\cdot)$ is convex and hence its subdifferential is well defined at any point $X \in \mathcal{S}_n$.) In fact, it is shown in [17, Thm. 1] that

$$\partial \lambda_1(X) = \left\{Q_1 Y Q_1^T : Y \in \mathcal{C}_r\right\} \tag{10}$$

where Q_1 is an $n \times r$ matrix whose columns form an orthogonal basis of the eigenspace associated with $\lambda_1(X)$ (it has dimension r), and $\mathcal{C}_r$ is the so-called spectraplex of $\mathcal{S}_r$:

$$\mathcal{C}_r := \{V \in \mathcal{S}_r : V \text{ is positive semidefinite, } \mathrm{tr}(V) = 1\}. \tag{11}$$

The following result means that the gradient $\nabla(f_\varepsilon \circ \lambda)(X)$ is an approximate element with respect to $\partial \lambda_1(X)$.

PROPOSITION 2.2. *For any* $X \in \mathcal{S}_n$, *we have*

$$\lim_{\varepsilon \to 0} \nabla(f_\varepsilon \circ \lambda)(X) \in \partial \lambda_1(X).$$

Proof. Let $X \in \mathcal{S}_n$ and $X = Q\mathrm{Diag}[\lambda(X)]Q^T$, $Q \in \mathcal{O}_X$. Denote $\lambda := \lambda(X)$ for simplicity. Assume that X has multiplicity r of the largest eigenvalue λ_1. Then the

first r columns of Q must form an orthogonal basis of the eigenspace associated with λ_1. These columns form a matrix, say $Q_1 \in \mathbb{R}^{n\times r}$.

It follows from (23) and (7) that

$$\nabla(f_\varepsilon \circ \lambda)(X) = Q\mathrm{Diag}[\nabla f_\varepsilon(\lambda)]Q^T = Q\mathrm{Diag}[\mu(\epsilon,\lambda)]Q^T,$$

with

$$\mu_i(\varepsilon,\lambda) = \frac{e^{\lambda_i/\varepsilon}}{\sum_{j=1}^{n} e^{\lambda_j/\varepsilon}} = \frac{e^{(\lambda_i-\lambda_1)/\varepsilon}}{\sum_{j=1}^{n} e^{(\lambda_j-\lambda_1)/\varepsilon}}.$$

Noticing that λ has multiplicity r of λ_1 and λ_1 is the largest element in λ, we obtain

$$\lim_{\varepsilon\to 0}\mu_i(\varepsilon,\lambda) = \begin{cases} \frac{1}{r} & i=1,\ldots,r \\ 0 & i=r+1,\ldots,n. \end{cases}$$

Therefore

$$\lim_{\varepsilon\to 0}\nabla(f_\varepsilon\circ\lambda)(X) = Q\mathrm{Diag}\left[\left(\frac{1}{r},\ldots,\frac{1}{r},0,\ldots,0\right)\right]Q^T = \frac{1}{r}Q_1 I_r Q_1^T,$$

where I_r is the identity matrix in $\mathcal{S}_r$. Our result then follows from the characterization (10) and (11) of subdifferential of $\lambda_1(\cdot)$ at X. □

The next step is to develop the formulae for the gradient and the Hessian of θ_ε, which, viewed as the composition of the twice differentiable spectral function $(f_\varepsilon \circ \lambda)$ and the affine mapping $A(\cdot):\mathbb{R}^m \to \mathcal{S}_n$, follow from the standard chain rule.

PROPOSITION 2.3. *Let $\varepsilon > 0$ be given. Then the followings hold*:

(*a*) *The function $\theta_\varepsilon(\cdot):\mathbb{R}^m\to\mathbb{R}$ is continuously differentiable, and the gradient of $\theta_\varepsilon(\cdot)$ at $y\in\mathbb{R}^m$ is given by*

$$\nabla\theta_\varepsilon(y) = \mathcal{A}^*\left(U(\mathrm{Diag}[\nabla f_\varepsilon(\lambda(A(y)))])U^T\right), \qquad \forall\, U\in\mathcal{O}_{A(y)}.$$

(*b*) *The function $\theta_\varepsilon(\cdot):\mathbb{R}^m\to\mathbb{R}$ is twice continuously differentiable, and the Hessian of $\theta_\varepsilon(\cdot)$ at $y\in\mathbb{R}^m$ is given by*

$$\nabla^2\theta_\varepsilon(y)[h] = \mathcal{A}^*\left(\nabla^2(f_\varepsilon\circ\lambda)(A(y))[\mathcal{A}h]\right), \qquad \forall\, h\in\mathbb{R}^m$$

where

$$\begin{aligned}&\nabla^2(f_\varepsilon\circ\lambda)(A(y))[\mathcal{A}h] \\ &= U\left(\mathrm{Diag}[\nabla^2 f_\varepsilon(\lambda(A(y)))\mathrm{diag}[\tilde{H}]] + \mathcal{C}\circ\tilde{H}\right)U^T \end{aligned} \tag{12}$$

and U is any orthogonal matrix in $\mathcal{O}_{A(y)}$, $H=\mathcal{A}h$, $\tilde{H}=U^THU$, and $\mathcal{C}=\mathcal{C}(\lambda(A(y)))$ is defined as in (25) *with f replaced by f_ε.*

We now develop a more explicit formula for calculating $\nabla\theta_\varepsilon(y)$. Let $U \in \mathcal{O}_{A(y)}$ and define

$$\tilde{A}_i := U^T A_i U, \qquad i = 1, \ldots, m.$$

Let $z := \lambda(A(y))$ and $\mu := \mu(\varepsilon, z)$ (cf. (8)), then

$$\langle A_i, U(\mathrm{Diag}[\nabla f_\varepsilon(\lambda(A(y)))])U^T\rangle = \langle U^T A_i U, \mathrm{Diag}[\mu]\rangle = \mu^T \mathrm{diag}[\tilde{A}_i].$$

We further let $W \in \mathbb{R}^{n\times m}$ with its ith column given by the vector $\mathrm{diag}[\tilde{A}_i]$, i.e.,

$$W(:, i) := \mathrm{diag}[\tilde{A}_i], \qquad i = 1, \ldots, m. \tag{13}$$

Then taking into account of the formula of $\nabla\theta_\varepsilon(y)$ in Proposition 2.3 (a), we have

$$\nabla\theta_\varepsilon(y) = W^T\mu.$$

In the next two subsections, we will pay attention to the Hessian of θ_ε, study its nonsingularity and develop a matrix representation for it.

2.2. NONSINGULARITY

The nonsingularity of the Hessian matrix $\nabla^2\theta_\varepsilon(y)$ plays a very important role in Newton-type methods for solving the (smooth) convex problem (6). It is clear to see that the Hessian matrix is split into two parts, namely

$$\nabla^2\theta_\varepsilon(y) = \mathcal{T}_1(y) + \mathcal{T}_2(y)$$

where $\mathcal{T}_1, \mathcal{T}_2 : \mathbb{R}^m \to \mathcal{S}_n$ are defined respectively by

$$\mathcal{T}_1(y)[h] := \mathcal{A}^*\left(U(\mathrm{Diag}(\nabla^2 f_\varepsilon(\lambda(A(y)))\mathrm{diag}[\tilde{H}]))U^T\right), \qquad \forall\, h \in \mathbb{R}^m$$
$$\mathcal{T}_2(y)[h] := \mathcal{A}^*\left(U(\mathcal{C}\circ\tilde{H})U^T\right), \qquad \forall\, h \in \mathbb{R}^m$$

where $U \in \mathcal{O}_{A(y)}$, $H = \mathcal{A}h$, $\tilde{H} = U^T H U$ and $\mathcal{C}$ is defined by $\mathcal{C}(\lambda(A(y)))$ as in (25) with f being replaced by f_ε. An implicit fact used in the above definitions is the independence of the choice of $U \in \mathcal{O}_{A(y)}$ on h. Since θ_ε is convex, $\nabla^2\theta_\varepsilon$ is positive semidefinite. We will see that it can be split into two positive semidefinite operators, i.e., $\mathcal{T}_1$ and $\mathcal{T}_2$. If one of them is positive definite, so is $\nabla^2\theta_\varepsilon$. This provides a way to study the nonsingularity of $\nabla^2\theta_\varepsilon$.

PROPOSITION 2.4. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be defined as above, then both of them are positive semidefinite operators. And $\nabla^2\theta_\epsilon(y)$ is positive definite iff the matrices $A_1, \ldots, A_m$ are linearly independent and $I \notin Range(\mathcal{A})$, where $Range(\mathcal{A}) = \{\mathcal{A}h | h \in \mathbb{R}^m\}$ and I is the identity matrix.*

Proof. We first prove the positive semidefiniteness of $\mathcal{T}_1(\cdot)$. It suffices to show that for any given $y \in \mathbb{R}^m$

$$\langle h, \mathcal{T}_1(y)[h] \rangle \geqslant 0, \qquad \forall\, h \in \mathbb{R}^m.$$

It is straightforward to see

$$\begin{aligned}
\langle h, \mathcal{T}_1(y)[h] \rangle &= \langle h, \mathcal{A}^*\big(U(\mathrm{Diag}(\nabla^2 f_\varepsilon(\lambda(A(y)))\mathrm{diag}[\tilde{H}]))U^T\big)\rangle \\
&= \langle \mathcal{A}h, U(\mathrm{Diag}(\nabla^2 f_\varepsilon(\lambda(A(y)))\mathrm{diag}[\tilde{H}]))U^T\rangle \\
&= \langle U^T(\mathcal{A}h)U, \mathrm{Diag}(\nabla^2 f_\varepsilon(\lambda(A(y)))\mathrm{diag}[\tilde{H}])\rangle \\
&= (\mathrm{diag}[\tilde{H}])^T \nabla^2 f_\varepsilon(\lambda(A(y)))\mathrm{diag}[\tilde{H}] \\
&\geqslant 0.
\end{aligned}$$

The last inequality used the semidefiniteness of $\nabla^2 f_\varepsilon(\cdot)$ as f_ε is a convex function.

Now we prove $\mathcal{T}_2$ is positive semidefinite. For simplicity, we let $z := \lambda(A(y))$. Then taking into account of Lemma 2.1 and (25), we obtain

$$\mathcal{C}_{ij} = (\mathcal{C}(z))_{ij} = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{\varepsilon}\mu_i(\varepsilon, z) & \text{if } i \neq j,\ z_i = z_j \\ \frac{\mu_i(\varepsilon, z) - \mu_j(\varepsilon, z)}{z_i - z_j} & \text{if } z_i \neq z_j. \end{cases} \tag{14}$$

Here we used the relation

$$(\nabla^2 f_\varepsilon(z))_{ii} = \frac{1}{\varepsilon}(\mu_i(\varepsilon, z) - (\mu_i(\varepsilon, z))^2), \quad (\nabla^2 f_\varepsilon(z))_{ij} = -\frac{1}{\varepsilon}\mu_i(\varepsilon, z)\mu_j(\varepsilon, z)$$

and $\mu_i(\varepsilon, z) = \mu_j(\varepsilon, z)$ if $z_i = z_j$ and $\mu_i(\varepsilon, z) > \mu_j(\varepsilon, z)$ if $z_i > z_j$. Thus, $\mathcal{C}_{ij} \geqslant 0$ for all $i, j = 1, \ldots, n$. We also note that for any $B, C, D \in \mathcal{S}_n$ it is easy to verify that

$$\langle B, C \circ D \rangle = \langle C, B \circ D \rangle = \langle D, B \circ C \rangle.$$

Using this and the nonnegativity of all elements of $\mathcal{C}$, we have for any $h \in \mathbb{R}^m$

$$\begin{aligned}
\langle h, \mathcal{T}_2(y)[h] \rangle &= \langle h, \mathcal{A}^*\big(U(\mathcal{C} \circ \tilde{H})U^T\big)\rangle \\
&= \langle \mathcal{A}h, U(\mathcal{C} \circ \tilde{H})U^T\rangle \\
&= \langle U^T(\mathcal{A}h)U, \mathcal{C} \circ \tilde{H}\rangle \\
&= \langle \tilde{H}, \mathcal{C} \circ \tilde{H}\rangle \\
&= \langle \mathcal{C}, \tilde{H} \circ \tilde{H}\rangle \\
&= \sum_{i=1}^n \sum_{j=1}^n (\mathcal{C}_{ij})(\tilde{H}_{ij})^2 \geqslant 0.
\end{aligned}$$

This proves the semidefiniteness of $\mathcal{T}_2$.

From the above argument, we see that $\nabla^2\theta_\varepsilon(y)$ is positive semidefinite and it is positive definite if and only if there does not exist nonzero h such that $\langle h, \mathcal{T}_1[h]\rangle = \langle h, \mathcal{T}_2[h]\rangle = 0$. Now note that $\langle h, \mathcal{T}_1[h]\rangle = 0$ if and only if $\tilde{H} = \rho I$ for some $\rho \in \mathbb{R}$ due to the special structure of $\nabla^2 f_\epsilon$, and $\langle h, \mathcal{T}_2[h]\rangle = 0$ if and only if $\tilde{H}_{ij} = 0$ for $i \neq j$ since $\mathcal{C}_{ij} > 0$ for $i \neq j$. Hence $\nabla^2\theta_\varepsilon[h] = 0$ for nonzero $h \in \mathbb{R}^m$ if and only if $\tilde{H} = \rho I$ for some $\rho \neq 0$. The possibility of $\rho = 0$ is removed by the linear independence of the matrices $A_1, \ldots, A_m$. But $\tilde{H} = \rho I$ is equivalent to $I \in Range(\mathcal{A})$. Thus we complete the proof. □

2.3. MATRIX REPRESENTATION

It is ideal to have a matrix representation $V \in \mathcal{S}_m$ for $\nabla^2\theta_\varepsilon(y)$ so that for any $h \in \mathbb{R}^m$,

$$\nabla^2\theta_\varepsilon(y)[h] = Vh. \tag{15}$$

Therefore, the Newton equation

$$Vh = -\nabla\theta_\varepsilon(y) \tag{16}$$

can be solved in a number of ways (e.g., Golub and Val Loan, 1996). To this end, let $U \in \mathcal{O}_{A(y)}$ be used in the definitions of $\mathcal{T}_1$ and $\mathcal{T}_2$, and recall that

$$\tilde{A}_i := U^T A_i U \qquad i = 1, \ldots, m.$$

Note that

$$\tilde{H} = U^T(\mathcal{A}h)U = U^T\left(\sum_{j=1}^m A_j h_j\right)U = \sum_{j=1}^m \tilde{A}_j h_j.$$

Let matrix $W \in \mathbb{R}^{n\times m}$ be defined by (13). Then

$$\mathrm{diag}[\tilde{H}] = \mathrm{diag}\left[\sum_{j=1}^m \tilde{A}_j h_j\right] = Wh.$$

For simplicity, further let $B := \nabla^2 f_\varepsilon(\lambda(A(y)))$. It follows that for $i = 1, \ldots, m$

$$\begin{aligned}(\mathcal{T}_1(y)[h])_i &= \langle A_i,\, U(\mathrm{Diag}(B\mathrm{diag}[\tilde{H}]))U^T\rangle \\ &= \langle U^T A_i U,\, \mathrm{Diag}[BWh]\rangle \\ &= \langle \tilde{A}_i,\, \mathrm{Diag}[BWh]\rangle \\ &= \sum_{j=1}^n (\tilde{A}_i)_{jj}(BWh)_j.\end{aligned}$$

Putting in the matrix form, we have

$$\mathcal{T}_1(y)[h] = \left(W^T B W\right) h. \tag{17}$$

Now we consider the linear operator $\mathcal{T}_2$. For $i = 1, \ldots, m$, we have

$$\begin{aligned}
(\mathcal{T}_2(y)[h])_i &= \langle A_i, U(\mathcal{C} \circ \tilde{H}) U^T \rangle \\
&= \langle U^T A_i U, \mathcal{C} \circ \tilde{H} \rangle \\
&= \langle \tilde{A}_i, \mathcal{C} \circ \tilde{H} \rangle \\
&= \langle \tilde{H}, \tilde{A}_i \circ \mathcal{C} \rangle \\
&= \sum_{j=1}^{m} \langle \tilde{A}_j, \tilde{A}_i \circ \mathcal{C} \rangle h_j \\
&= \sum_{j=1}^{m} \langle \mathcal{C}, \tilde{A}_i \circ \tilde{A}_j \rangle h_j.
\end{aligned}$$

Putting in matrix form, we then have

$$\mathcal{T}_2(y)[h] = Mh, \tag{18}$$

where $M \in \mathcal{S}_m$ is defined by

$$M_{ij} := \langle \mathcal{C}, \tilde{A}_i \circ \tilde{A}_j \rangle \qquad i, j = 1, \ldots, m. \tag{19}$$

Putting (17) and (18) together, we have the matrix representation (15) with $V \in \mathcal{S}_m$ given by

$$V := W^T B W + M.$$

Proposition 2.4 implies that both the matrix $W^T B W$ and M are positive semidef -inite. In fact the positive semidefiniteness of $W^T B W$ follows directly from that of B since B is always positive semidefinite, see the structure of B in (8). We restate Proposition 2.4 in terms of V.

PROPOSITION 2.5. *Let V be defined as above depending on given $y \in \mathbb{R}^m$. Then*

$$\nabla^2 \theta_\varepsilon(y) = V.$$

Moreover, V is positive definite if and only if the matrices $A_1, \ldots, A_m$ are linearly independent and $I \notin Range(\mathcal{A})$.

Compared with the formula in Proposition 2.3 (b), the advantage of the matrix representation of the Hessian matrix V is the separation of h from V, which makes the Newton equation (see, (16)) into matrix–vector formulation, and hence paves ways for applying numerical algorithm directly to it. To see clearly the implication of Proposition 2.5, let us consider the simplest case that $A(y)$ has multiplicity one

of the largest eigenvalue, i.e., $\lambda_1(A(y)) > \lambda_2(A(y))$. In this case, the function $\lambda_1(A(\cdot))$ is twice continuously differentiable at y. After calculation by using the formula (26) and repeating similar arguments as for the operator $\mathcal{T}_2$, we see that

$$\nabla^2 \lambda_1(A(y)) = M$$

where M is defined as in (19) with $\mathcal{C}$ replaced by

$$\mathcal{C} := \begin{pmatrix} 0 & \frac{1}{\lambda_1-\lambda_2} & \frac{1}{\lambda_1-\lambda_3} & \cdots & \frac{1}{\lambda_1-\lambda_n} \\ \frac{1}{\lambda_1-\lambda_2} & 0 & 0 & \cdots & 0 \\ \frac{1}{\lambda_1-\lambda_3} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\lambda_1-\lambda_n} & 0 & 0 & \cdots & 0 \end{pmatrix}$$

and $\lambda_i := \lambda_i(A(y)), i = 1, \ldots, n$. We stress that M is independent of choice of $U \in \mathcal{O}_{A(y)}$ due to the structure of $\mathcal{C}$. Then conditions ensuring the nonsingularity of M become the same conditions which guarantee the nonsingularity of the Hessian of $\lambda_1(A(\cdot))$ at y, which is the case for optimization problems with twice continuously differentiable data.

However, for the degenerate case, i.e., $A(y)$ has multiplicity r $(r > 1)$ of the largest eigenvalue, the situation becomes a bit complicated. For simplicity, let again $z := \lambda(A(y))$. Hence, $z_i = \lambda_1(A(y))$ for $i = 1, \ldots, r$. We are interested in the case where ε is small. It follows from (8) that

$$\lim_{\varepsilon \to 0} \mu_i(\varepsilon, z) = \begin{cases} \frac{1}{r} & i = 1, \ldots, r \\ 0 & i = r+1, \ldots, n. \end{cases}$$

We observe from (14) that some elements of $\mathcal{C}$ grow to infinity as ε approaches to zero. The same problem occurs with the matrix $\nabla^2 f_\varepsilon(y)$. Fortunately, those two matrices are well scaled, i.e., their magnitude is of $1/\varepsilon$. Let $\mathcal{C}_0, \mathcal{D}_0 \in \mathcal{S}_n$ be defined by

$$\mathcal{C}_0 := \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathcal{D}_0 := \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

where $C := \frac{1}{r}(E_r - I_r)$, $D := \frac{1}{r^2}(rI_r - E_r)$, and E_r is the matrix of all ones in $\mathcal{S}_r$ and I_r is the identity matrix in $\mathcal{S}_r$. It is easy to see from (14) and (9) that

$$\lim_{\varepsilon \to 0} \varepsilon \mathcal{C} = \mathcal{C}_0 \quad \text{and} \quad \lim_{\varepsilon \to 0} \varepsilon \nabla^2 f_\varepsilon(y) = \mathcal{D}_0. \tag{20}$$

The limits in (20) mean that eventually the eigenvectors corresponding to the largest eigenvalue of $A(y)$ dominate the Hessian of $\theta_\varepsilon(y)$ as ε goes to zero. This observation may be of vital importance in practically calculating the Hessian.

3. A regularization method

In previous sections, we have seen that the function $\theta_\varepsilon(\cdot)$ is well conditioned in the sense that it is twice continuously differentiable and strictly convex given the linear independence condition. Moreover, we have

$$\lim_{\varepsilon\to 0}\theta_\varepsilon(y)=\lambda_1(A(y)),\qquad \forall y\in\mathbb{R}^m.$$

This suggests that, instead of minimizing the nondifferentiable function $\lambda_1(A(\cdot))$, we can minimize the well conditioned function $\theta_\varepsilon(\cdot)$, and let ε go to zero. A crucial question to this approach is that if any selection of the minimizers of $\theta_\varepsilon(\cdot)$ converges to the set of minimizers of $\lambda_1(A(\cdot))$ as ε goes to zero. To answer this question, we apply Auslender's general penalty approach [1] to our problem (1).

Recall that the recession function of $h(\cdot)$ defined below is the max-function (3) (see Example (ii) of [1, Introduction])

$$h(x):=\ln\left(\sum_{i=1}^{n}e^{x_i}\right),\qquad \forall x\in\mathbb{R}^n.$$

Then the smoothing function f_ε defined by (4) can be obtained by

$$f_\varepsilon(x)=\varepsilon h(x/\varepsilon).$$

For simplicity, let $g:\mathbb{R}^m\to\mathbb{R}^n$

$$g(y):=(\lambda\circ A)(y)=\lambda(A(y)),$$

and the components of g be given respectively by

$$g_i(y):=\lambda_i(A(y))=(\lambda_i\circ A)(y),\quad i=1,\dots,n.$$

We then have

$$\theta_\varepsilon(y)=\varepsilon h(g(y)/\varepsilon).$$

For each $i=1,\dots,n$ and $y\in\mathbb{R}^m$, the recession function of g_i, denoted by $(g_i)_\infty$, is defined by (see [1, (2.1)])

$$(g_i)_\infty(y):=\inf\left\{\liminf_{k\to\infty}\frac{g_i(t_k y_k)}{t_k}:t_k\to+\infty,\ y_k\to y\right\}.$$

Let

$$g_\infty(y):=((g_1)_\infty(y),\dots,(g_n)_\infty(y)),\qquad \forall y\in\mathbb{R}^m.$$

A general convergence result of Auslender [1, Thm. 2.3], which in our case concerns the convergence of any selection of the minimizers of $\theta_\varepsilon(\cdot)$ to the solution set of the minimizers of $\lambda_1(A(\cdot))$, involves the condition

$$h(g_\infty(y))>0,\qquad \forall 0\neq y\in\mathbb{R}^m. \tag{21}$$

The other assumptions of [1, Thm. 2.3] are automatically satisfied in our case.

LEMMA 3.1. *Suppose that the solution set of (1), denoted by* Ω_0, *is nonempty and bounded. Then (21) is satisfied for all* $0 \neq y \in \mathbb{R}^m$.

Proof. Let $0 \neq y \in \mathbb{R}^m$ be given, and let $t_k \to +\infty$ and $y_k \to y$. We have

$$(\lambda_i \circ A)(t_k y_k)/t_k = \lambda_i(A(t_k y_k))/t_k = \lambda_i(A_0 + t_k \mathcal{A} y_k)/t_k = \lambda_i(A_0/t_k + \mathcal{A} y_k).$$

Hence

$$(g_i)_\infty(y) = \lim_{k\to\infty} \lambda_i(A_0/t_k + \mathcal{A} y_k) = \lambda_i(\mathcal{A} y).$$

Therefore, we have

$$h(g_\infty(y)) = \lambda_1(\mathcal{A} y).$$

Now we prove (21) holds for all $0 \neq y \in \mathbb{R}^m$. We first observe that $\lambda_1(\mathcal{A} y) \geqslant 0$ for all $y \in \mathbb{R}^m$. To see this, suppose there exists one $y \in \mathbb{R}^m$ such that $\lambda_1(\mathcal{A} y) < 0$, then $\lambda_1(A(ty)) \to -\infty$ as $t \to \infty$, which contradicts the nonemptiness of Ω_0. Let $y^* \in \Omega$. Note that we have proved $\lambda_1(\mathcal{A} y) \geqslant 0$ for all $y \in \mathbb{R}^m$. If $\lambda_1(\mathcal{A} y) = 0$ for some $0 \neq y \in \mathbb{R}^m$, then for any $t \geqslant 0$,

$$\lambda_1(A(y^* + ty)) = \lambda_1(A(y^*) + t\mathcal{A} y) \leqslant \lambda_1(A(y^*)) + t\lambda_1(\mathcal{A} y) = \lambda_1(A(y^*)).$$

If the equality holds for all $t \geqslant 0$, then $y^* + ty \in \Omega_0$ for all $t \geqslant 0$, contradicting the boundedness of Ω_0. Therefore, we must have

$$\lambda_1(A(y^* + t^* y) < \lambda_1(A(y^*))$$

for some large $t^* > 0$. That is, $y^* + t^* y$ is a point which yields a less value than that at y^*, and hence y^* could not be a solution of (1), a contradiction. This proves (21), completing our proof. □

Now we are ready to specialize to our problem a general convergence result for penalty and barrier methods due to Auslender [1, Thm. 2.3] (all assumptions except (2.7) of this theorem are automatically satisfied with our problem, and (2.7) is verified in Lemma 3.1).

THEOREM 3.2. *Suppose that the solution set* Ω_0 *of problem (1) is nonempty and compact. Then the solution set* Ω_ε *of problem (6) is also nonempty and compact for any* $\varepsilon > 0$. *Moreover, every selection* $y_\varepsilon \in \Omega_\varepsilon$ *stays bounded with all its limit points in* Ω_0. *In particular, if the matrices* $A_1, \ldots, A_m$ *are linearly independent, then for every* $\varepsilon > 0$ *the minimization problem (6) is strictly convex, admits a unique solution* y_ε, *and*

$$\lim_{\varepsilon\to 0} \operatorname{dist}(y_\varepsilon, \Omega_0) = 0$$

where dist *denotes the distance of* y_ε *to* Ω_0.

The boundedness condition on the solution set cannot be dropped in the theorem. Here is an example: Let $n=2, m=1$, $A_0=\mathrm{Diag}[0,0]$ and $A_1=\mathrm{Diag}[0,1]$. Then the solution set of the eigenvalue optimization problem is $\{x\in\mathbb{R}\,|\,x\leqslant 0\}$, which is obviously unbounded. However, for any $\varepsilon>0$,

$$\theta_\varepsilon(x)=\varepsilon\ln(1+e^{x/\varepsilon}),$$

and hence the convex minimization problem (6) has no solution. In this example, the linear independence condition holds. On the one hand, the linear independence increases the solvability of the smooth problem (6) as it is strictly convex. However, sometimes it is very difficult to check the linear independence. On the other hand, the linear independence is sometimes not sufficient to guarantee the solvability of the smooth problem. To overcome these difficulties, we consider the so-called Tikhonov regularization of the problem (6):

$$\min_{y\in\mathbb{R}^m}\hat{\theta}_\varepsilon(y), \tag{22}$$

where

$$\hat{\theta}_\varepsilon(y):=\theta_\varepsilon(y)+\frac{1}{2}\|y\|^2,\qquad \forall\, y\in\mathbb{R}^m.$$

Since $\theta_\varepsilon(\cdot)$ is convex for any $\varepsilon>0$, $\hat{\theta}_\varepsilon(\cdot)$ is strongly convex, and hence the problem (22) always has a (unique) solution and is well-posed in the sense of Dontchev and Zolezzi (1993). The Tikhonov regularization has been extensively studied in the book Dontchev and Zolezzi (1993) and been successfully used to study complementarity problems and variational inequalities by a number of authors recently, see Facchinei (1998), Facchinei and Kanzow (1999), Qi (1999, 2000), Ravindran and Gowda (2000) and Sun (1999). The following result concerning the convergence of the solution sequence of the regularized problem (22) can be proved similarly as that for [1, Thm. 2.3] by using Lemma 3.1.

THEOREM 3.3. *The smooth problem (22) has always a unique solution y_ε for any $\varepsilon>0$, and the solution sequence $\{y_\varepsilon\}_{\varepsilon>0}$ remains bounded and the distance between y_ε and Ω_0 approaches zero as $\varepsilon\to 0$ providing Ω_0 is nonempty and bounded.*

Our regularization method based on Theorem 3.3 can be stated as follows.

ALGORITHM 3.4 (A regularization method).

(*S*.1) *Let $\{\varepsilon_1,\varepsilon_2,\ldots\}$ be a given sequence decreasing to zero. Let $y^0\in\mathbb{R}^m$ be given. Set $k:=1$.*

(*S*.2) *Using an unconstrained convex minimization method with initial point y^{k-1} to find the unique solution of the problem (22) with $\varepsilon=\varepsilon_k$.*

(*S*.3) *Repeat* (*S*.2) *until a termination criterion is satisfied.*

REMARK. Although the problem (22) is well-posed for each $\varepsilon_k > 0$, we stress that sometimes it is not an easy task to find the exact solution of (22). Fortunately, it is possible to design an iterative algorithm based on Algorithm *3.4* by exploiting the regularization term of (22) as done in Qi (2000) and Sun (1999) for box variational inequalities and complementarity problems. It is beyond the scope of the current paper to design such an algorithm with convergence analysis.

4. Conclusions

In this paper, we studied smooth convex approximations to the maximum eigenvalue function. To make it applicable to a wide class of applications, the study are conducted on the composite function of the maximum eigenvalue function and a linear operator mapping $\mathbb{R}^m$ to $\mathcal{S}_n$, which in turn is the natural objective function of minimizing the maximum eigenvalue function over an affine space in $\mathcal{S}_n$. This leads us to a sequence of smooth convex minimization problems governed by a smoothing parameter. As the parameter goes to *zero*, the original problem is recovered. Efforts are then made on deriving a computable Hessian formula of the smooth convex functions, and on the regularity conditions which guarantee nonsingularity of the Hessian matrices. We also proposed a regularization technique ensuring the well-posedness of the smooth convex problems. This could lead to a class of numerical methods for minimizing the maximum eigenvalue function over an affine space. As observed by a referee, the arguments in the paper do not depend much on the special structure of the max function nor the log-exponential function and could possibly be expanded to a more general class of convex symmetric functions and their smoothing counterparts. For example, choose a smooth convex function $h: \mathbb{R}^n \to \mathbb{R}^n$ and define $f_\epsilon(x) := \epsilon h(\frac{x}{\epsilon})$. Then, f_ϵ forms a smoothing function for the recession function $h_\infty(x) := \lim_{\epsilon \to 0^+} f_\epsilon(x)$. Accordingly,

$$\min (f_\epsilon \circ \lambda \circ A)(y)$$

forms a smooth approximation of the nonsmooth optimization problem

$$\min (h_\infty \circ \lambda \circ A)(y).$$

Some additional assumptions on h might be needed to obtain results like Proposition 2.2 and Lemma 3.1, but most of the other results in the paper hold in this more general context.

Acknowledgment

We thank two anonymous referees for their detailed comments and constructive suggestions on the organization of the paper. In particular, we thank one referee for

explicitly pointing out a straightforward extension to the more general case which is briefly discussed above and may be interesting to wider audience. This work was supported by the Research Grant Council of Hong Kong.

Appendix

In this appendix, we recall formulae for the gradient and the Hessian of the twice continuously differentiable spectral functions $(f \circ \lambda) : \mathcal{S}_n \to \mathbb{R}$, which have played important roles in this paper.

PROPOSITION 4.1 [12, Thm. 1.1] (formula for gradient). *Let f be a symmetric function from $\mathbb{R}^n$ to $\mathbb{R}$ and $X \in \mathcal{S}_n$. Then the following hold*:

(*a*) *$(f \circ \lambda)$ is differentiable at point X if and only if f is differentiable at point $\lambda(X)$. In this case the gradient of $(f \circ \lambda)$ at X is given by*

$$\nabla(f \circ \lambda)(X) = U \mathrm{Diag}[\nabla f(\lambda(X))] U^T, \quad \forall\, U \in \mathcal{O}_X. \tag{23}$$

More generally, the gradient of $(f \circ \lambda)$ has the following formula

$$\nabla(f \circ \lambda)(X) = V \mathrm{Diag}[\nabla f(\mu)] V^T, \tag{24}$$

for any orthogonal matrix $V \in \mathcal{O}$ and $\mu \in \mathbb{R}^n$ satisfying $X = V \mathrm{Diag}[\mu] V^T$.

(*b*) *$(f \circ \lambda)$ is continuously differentiable at point X if and only if f is continuously differentiable at point $\lambda(X)$.*

In a recent paper, Lewis and Sendov (2001) found a formula for calculating the Hessian of the spectral function $(f \circ \lambda)$, when it exists, via calculating the Hessian of f. This facilitates the possibility of using second-order methods for solving the convex minimization problem. Suppose that f is twice differentiable at $\mu \in \mathbb{R}^n$. Define the matrix $\mathcal{C}(\mu) \in \mathbb{R}^{n \times n}$:

$$(\mathcal{C}(\mu))_{ij} := \begin{cases} 0 & \text{if } \ i = j \\ \left(\nabla^2 f(\mu)\right)_{ii} - \left(\nabla^2 f(\mu)\right)_{ij} & \text{if } \ i \neq j \text{ and } \mu_i = \mu_j \\ \dfrac{(\nabla f(\mu))_i - (\nabla f(\mu))_j}{\mu_i - \mu_j} & \text{else.} \end{cases} \tag{25}$$

It is easy to see that $\mathcal{C}(\mu)$ is symmetric due to the symmetry of f.

PROPOSITION 4.2 [14, Thm. 3.3, 4.2] (formula for Hessian). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be symmetric. Then for any $X \in \mathcal{S}_n$, it holds that $(f \circ \lambda)$ is twice (continuously) differentiable at X if and only if f is twice (continuously) differentiable at $\lambda(X)$. Moreover, in this case the Hessian of the spectral function at X is*

$$\nabla^2(f \circ \lambda)(X)[H] = U\left(\mathrm{Diag}[\nabla^2 f(\lambda(X)) \mathrm{diag}[\tilde{H}]] + \mathcal{C}(\lambda(X)) \circ \tilde{H}\right) U^T, \quad \forall H \in \mathcal{S}_n \tag{26}$$

where U is any orthogonal matrix in $\mathcal{O}_X$ and $\tilde{H} = U^T H U$.

REMARK. We stress that the formulae (23) and (26) do not depend on the particular choice of $U \in \mathcal{O}_X$.

References

Auslender, A. (1999), Penalty and barrier methods: a unified framework, *SIAM J. Optimization* 10, 211–230.

Ben-Tal, A. and Teboulle, M. (1989), A smoothing technique for nondifferentiable optimization problems. In: *Optimization*, 1–11. Fifth French German Conference, Lecture Notes in Math. 1405, Springer, New York.

Bertsekas, D.P. (1982), *Constrained Optimization and Lagrangian Multiplier Methods*, Academic Press, New York.

Chang, P.-L. (1980), A minimax approach to nonlinear programming, Doctoral Dissertation, University of Washington, Department of Mathematics.

Chen, X. and Tseng, P. (2003), Non-interior continuation methods for solving semidefinite complementarity problems, *Mathematical Programming* 95, 431–474.

Dontchev, A.L. and Zolezzi, T. (1993), Well-posed optimization problems. In: *Lecture Notes in Mathematics*, 1543. Springer, Berlin.

Facchinei, F. (1998), Structural and stability properties of P_0 nonlinear complementarity problems, *Mathematics of Operations Research* 23, 735–745.

Facchinei, F. and Kanzow, C. (1999), Beyond monotonicity in regularization methods for nonlinear complementarity problems, *SIAM J. Control and Optimization* 37, 1150–1161.

Goldstein, A.A. (1997), Chebyshev approximation and linear inequalities via exponentials, Report, Department of Mathematics, University of Washington, Seattle.

Golub, G.H. and Val Loan, C.F. (1996), *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD.

Horn, R.A. and Johnson, C.R. (1985), *Matrix Analysis*, Cambridge Press, Cambridge, United Kingdom.

Lewis, A.S. (1996), Derivatives of spectral functions, *Mathematics of Operations Research* 21, 576–588.

Lewis, A.S. and Overton, M.L. (1996), Eigenvalue optimization, *Acta Numerica* 5, 149–190.

Lewis, A.S. and Sendov, H.S. (2001), Twice differentiable spectral functions, *SIAM J. Matrix Analysis and Application* 23, 368–386.

Li, X.-S. (1991), An aggregation function method for nonlinear programming, *Science in China (Series A)* 34, 1467–1473.

Oustry, F. (1999), The $\mathcal{U}$-Lagrangian of the maximum eigenvalue function, *SIAM J. Optimization* 9, 526–549.

Oustry, F. (2000), A second-order bundle method to minimize the maximum eigenvalue function, *Mathematical Programming* 89, 1–33.

Overton, M.L. and Womersley, R.S. (1995), Second derivatives for optimizing eigenvalues of symmetric matrices, *SIAM J. Matrix Analysis and Application* 16, 667–718.

Peng, J.-M. and Lin, Z.-H. (1999), A non-interior continuation method for generalized linear complementarity problems, *Mathematical Programming* 86, 533–563.

Qi, H.-D. (1999), Tikhonov regularization methods for variational inequality problems, *J. Optimization Theory and Application* 102, 193–201.

Qi, H.-D. (2000), A regularized smoothing Newton method for box constrained variational inequality problems with P_0-functions, *SIAM J. Optimization* 10, 315–330.

Qi, H.-D. and Liao, L.-Z. (1999), A smoothing Newton method for extended vertical linear complementarity problems, *SIAM J. Matrix Analysis and Application* 21, 45-66.

Qi L. and Tseng, P. (2002), An analysis of piecewise smooth functions and almost smooth functions, Department of Applied Mathematics, The Hong Kong Polytechnic University.

Ravindran, G. and Gowda, M.S. (2000), Regularization of P_0-functions in box variational inequality problems, *SIAM J. Optimization* 11, 748–760.

Shapiro, A. and Fan, M.K.H. (1995), On eigenvalue optimization, *SIAM J. Optimization* 5, 552–568.

Stewart, G.W. and Sun, J.-G. (1990), *Matrix Perturbation Theory*, Academic Press, New York.

Sun, D. (1990), A regularization Newton method for solving nonlinear complementarity problems, *Applied Mathematics and Optimization* 40, 315–339.

Tang, H. and Zhang, L. (1994), A maximum entropy method for convex programming, *Chinese Science Bulletin* 39, 682–684.

Tseng, P. and Bertsekas, D.P. (1993), On the convergence of the exponential multiplier method for convex programming, *Mathematical Programming* 60, 1–19.

Zhang, L. and Tang, H. (1998), A further study on a penalty function of Bertsekas. Advances in nonlinear programming (Beijing, 1996), Appl. Optim., 14, Kluwer Academic Publishers, Dordrecht, pp. 345–351.