Strong Semismoothness of Eigenvalues of Symmetric Matrices and Its Application to Inverse Eigenvalue Problems[1]

Defeng Sun[2]

Department of Mathematics

National University of Singapore, Republic of Singapore

Jie Sun[3]

Faculty of Business Administration and Singapore-MIT Alliance

National University of Singapore, Republic of Singapore

August 16, 2001

Abstract. It is well known that the eigenvalues of a real symmetric matrix are not everywhere differentiable. A classical result of Ky Fan states that each eigenvalue of a symmetric matrix is the difference of two convex functions. This directly implies that the eigenvalues of a symmetric matrix are semismooth everywhere. Based on a very recent result of the authors, it is further proved in this paper that the eigenvalues of a symmetric matrix are strongly semismooth everywhere. As an application, it is demonstrated how this result can be used to analyze the quadratic convergence of Newton's methods for solving inverse eigenvalue problems (IEPs). This, in a systematic way, not only extends the quadratic convergence results of Friedland, Nocedal and Overton [SIAM Journal on Numerical Analysis, Vol. 24, 1987, 634–667] and others, but also gives an affirmative answer to a conjecture made by Dai and Lancaster [Numerical Linear Algebra with Applications, Vol. 4, 1997, 1–21] for the quadratic convergence of Newton's methods for solving generalized IEPs with multiple eigenvalues.

Keywords: Symmetric matrices, eigenvalues, strong semismoothness, Newton's methods, inverse eigenvalue problems, quadratic convergence

AMS subject classifications: 65F15, 65F18, 65H10, 65H17

---

1. *The research was partially supported by the grant R314-000-026-112 of National University of Singapore, and a grant from the Singapore-MIT Alliance.*

2. *E-mail: matsundf@nus.edu.sg. Fax: 65–779 5452.*

3. *E-mail: jsun@nus.edu.sg.*

1 Introduction

The theory of semismooth functions developed in the last decade has been successful in analyzing the quadratic convergence of Newton's method for non-differentiable (nonsmooth) equations; it is well received by the optimization community, but is perhaps not well known by researchers in numerical analysis. In this paper we take the inverse eigenvalue problem (IEP) as an example to show how this theory can be used in analyzing matrix-related equations. For applications of the IEP the interested reader is referred to the paper of Friedland, Nocedal, and Overton [9], the book of Xu [24], and the references therein.

Let $\mathcal{S}$ be the linear space of symmetric matrices of size n. Let $A : \Re^n \to \mathcal{S}$ be continuously differentiable. Given n real numbers $\{\lambda_i^*\}_{i=1}^n$, which are arranged in the decreasing order $\lambda_1^* \geq \cdots \geq \lambda_n^*$, the IEP is to find a vector $c^* \in \Re^n$ such that $\lambda_i(A(c^*)) = \lambda_i^*$ for $i = 1, \cdots, n$. A typical choice for $A(c)$ is

$$A(c) = A_0 + \sum_{j=1}^{n} c_j A_j, \tag{1}$$

where $A_0, A_1, \cdots, A_n \in \mathcal{S}$. In this case, $A(c)$ is an affine function of c.

Define $F : \Re^n \to \Re^n$ by

$$F(c) = \begin{bmatrix} \lambda_1(A(c)) - \lambda_1^* \\ \vdots \\ \lambda_n(A(c)) - \lambda_n^* \end{bmatrix}. \tag{2}$$

Then the IEP is equivalent to find a $c^* \in \Re^n$ to be a solution of the following equation

$$F(c) = 0. \tag{3}$$

Of course, there are other ways to formulate the IEP as a system of equations. For instance we may solve $F(c) = 0$ where

$$F(c) = \begin{bmatrix} \det(A(c) - \lambda_1^* I) \\ \vdots \\ \det(A(c) - \lambda_n^* I) \end{bmatrix}.$$

It should be noted that the convergence theory we are going to present does not depend on the specific form of $F(c)$. Rather it is derived based on a property of F called strong semismoothness (defined later). It is well known that for $X \in \mathcal{S}$ the eigenvalues of X, as functions of X, are not everywhere differentiable. However, we shall show that they are strongly semismooth and therefore quadratic convergence of Newton's methods is a natural result when applied to equations involving eigenvalues. In doing so, we also give a constructive proof for a difficult result of Chen and Tseng [3] on upper semicontinuity of a set-valued mapping of orthogonal matrices.

The concept of semismoothness of functionals was originally studied by Mifflin [10] while strong semismoothness was introduced by Qi and Sun in [15] for vector valued functions. Recently, it is further extended to matrix valued functions [20] for the study of quadratic convergence of Newton's methods. Generally speaking, *strong semismoothness* of an equation is tied with quadratic convergence of a Newton method applied to the equation and *semismoothness* corresponds to superlinear convergence. It was shown that smooth functions, piecewise smooth functions, and convex and concave functions are semismooth functions. They are not, however, necessarily strong semismooth functions.

To see the motivation of this paper more clearly, let us consider the following example

$$X = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix},$$

where x_1, x_2 and x_3 are parameters. In this case, we have

$$\lambda_1(X) = \frac{x_1 + x_2 + \sqrt{(x_1 - x_3)^2 + 4x_2^2}}{2} \quad \text{and} \quad \lambda_2(X) = \frac{x_1 + x_2 - \sqrt{(x_1 - x_3)^2 + 4x_2^2}}{2}. \tag{4}$$

Since $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$ are not differentiable at X with $x_1 = x_3$ and $x_2 = 0$, a gradient-dependent numerical method (e.g., Newton's method) may get into trouble when hitting those points. In addition, theoretical analysis gets tricky without differentiability. Further inspection reveals that $\lambda_1(\cdot)$ is a convex function and $\lambda_2(\cdot)$ is a concave function. Hence, both of them are semismooth functions [10] and some nonsmooth version of Newton's methods [15] might be applied to equations containing $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$. This should be not a coincidence. Let $f_m(X)$ be the sum of m largest eigenvalues of X. Then, Ky Fan's Maximum Principle [7] says that for each $i = 1, \cdots, n$, $f_i(\cdot)$ is a convex function. This result implies that

- $\lambda_1(\cdot)$ is a convex function and $\lambda_n(\cdot)$ is a concave function;
- For $i = 2, \cdots, n-1$, $\lambda_i(\cdot)$ is the difference of two convex functions.

Since convex and concave functions are semismooth and the difference of two semismooth functions is still a semismooth function [10], Ky Fan's result shows that $\lambda_1(\cdot), \cdots, \lambda_n(\cdot)$ are all semismooth functions. It is therefore expectable that, when applying a Newton method to IEPs, the convergence rate is at least superlinear. A more interesting question is: are all $\lambda_1(\cdot), \cdots, \lambda_n(\cdot)$ strongly semismooth functions (therefore implying quadratic convergence)? In this paper, based on a recent result of the authors [20], we will give an affirmative answer to the above question.

The organization of this paper is as follows. Some basic facts on semismoothness are presented in Section 2. Some nonsmooth versions of the Newton method, which we call *relative generalized*

Newton methods, are introduced in Section 3. Section 4 concentrates on the strong semismoothness of eigenvalues of a symmetric matrix. The quadratic convergence of Newton's methods for IEPs and generalized IEPs is proved in Section 5. Section 6 gives a summary and a few possible future research topics.

Some notations to be used are as follows.

- Usually, calligraphic letters denote matrix sets. Capital letters represent matrices or functions. Lower case letters are for vectors. Greek letters stand for scalars.
- $\mathcal{O}$ is the set of all $n \times n$ orthogonal matrices.
- A superscript "T" represents the transpose of matrices and vectors. For a matrix A, $A_{i\cdot}$ and $A_{\cdot j}$ represent the ith row and jth column of A, respectively.
- Unless otherwise specified, all vector norms are 2-norms and matrix norms are Frobenius norms: $\|A\| := \mathrm{trace}\,(A^T A)^{1/2}$.
- We write $X = O(\alpha)$ (respectively, $o(\alpha)$) if $\|X\|/|\alpha|$ is uniformly bounded (respectively, tends to zero) as $\alpha \to 0$.

2 Some Basic Facts on Semismoothness

2.1 Semismooth Vector Functions

Let $G : \Re^n \to \Re^n$ be a locally Lipschitz continuous function. Then $G : \Re^n \to \Re^n$ has a generalized Jacobian $\partial G(x)$ in the sense of Clarke [4]. Denote

$$\partial_B G(x) := \{V \in \Re^{n\times n} |\ V = \lim_{x^k \to x} G'(x^k), G \text{ is differentiable at } x^k\}.$$

Then the Clarke's generalized Jacobian is

$$\partial G(x) = \mathrm{conv}\{\partial_B G(x)\},$$

where "conv" stands for the convex hull in the usual sense of convex analysis [17].

G is said to be *semismooth* at $x \in \Re^n$, if

$$\lim_{\substack{V\in \partial G(x+th') \\ h' \to h,\ t\downarrow 0}} \{Vh'\}$$

exists for any $h \in \Re^n$.

It is shown that the composition of semismooth functions is still a semismooth function (see [10]). If G is semismooth at x, then $G'(x;h)$, the directional derivative of G at x in the direction of h, exists for any $h \in \Re^n$ and is equal to the above limit.

G is said to be *strongly semismooth* at x if G is semismooth at x and for any $V \in \partial G(x+h)$, $h \to 0$,

$$G(x+h) - G(x) - Vh = O(\|h\|^2).$$

A function G is said to be a (strongly) semismooth function if it is (strongly) semismooth everywhere on $\Re^n$. It is shown that the composition of strongly semismooth functions is strongly semismooth [8].

2.2 Generalized Newton Methods

Suppose that $G : \Re^n \to \Re^n$ is locally Lipschitz continuous. Based on $\partial G(x)$, Qi and Sun [15] proposed the following Newton's method for solving $G(x) = 0$.

Generalized Newton Method I: Given $x^0 \in \Re^n$, for $k = 0, 1, \cdots$,

$$x^{k+1} = x^k - V_k^{-1} G(x^k), \tag{5}$$

where $V_k \in \partial G(x^k)$.

The following convergence theorem for the generalized Newton method (5) is established in [15].

Theorem 2.1 *Suppose that $G(x^*) = 0$. If all $V \in \partial G(x^*)$ are nonsingular and G is semismooth at x^*, then there exists a neighborhood $N(x^*)$ of x^* such that for any $x^0 \in N(x^*)$ the generalized Newton method I is well defined and is Q-superlinearly convergent. Moreover, if G is strongly semismooth at x^*, then (5) converges Q-quadratically.*

To relax the nonsingularity assumption on $\partial G(x^*)$, Qi [13] introduced the following method based on the concept of $\partial_B G(x)$.

Generalized Newton Method II: Given $x^0 \in \Re^n$, for $k = 0, 1, \cdots$,

$$x^{k+1} = x^k - V_k^{-1} G(x^k), \tag{6}$$

where $V_k \in \partial_B G(x^k)$.

The convergence theorem for the generalized Newton method (6) is the same as Theorem 2.1 except that ∂G is replaced by $\partial_B G$.

Theorem 2.2 *Suppose that $G(x^*) = 0$. If all $V \in \partial_B G(x^*)$ are nonsingular and G is semismooth at x^*, then there exists a neighborhood $N(x^*)$ of x^* such that for any $x^0 \in N(x^*)$ the generalized Newton method II is well defined and is Q-superlinearly convergent. Moreover, if G is strongly semismooth at x^*, then (6) converges Q-quadratically.*

Now, let us consider the following composite nonsmooth equations

$$G(x) := \Phi(\Psi(x)) = 0, \tag{7}$$

where $\Phi : \Re^n \to \Re^m$ is nonsmooth but of special structure and $\Psi : \Re^m \to \Re^n$ is continuously differentiable. It is noted that neither $\partial G(x)$ nor $\partial G_B(x)$ is easy to compute even if $\partial\Phi(y)$ and $\partial_B\Phi(y)$ are available. To circumvent the difficulty in computing $\partial G(x)$ and $\partial G_B(x)$, Potra et al. [12] introduced the following concept of generalized Jacobian

$$\partial_Q G(x) = \partial_B \Phi(\Psi(x))\Psi'(x),$$

where "Q" stands for "Quasi". We shall see in the later discussion that $\partial_Q G(x)$ is more convenient to compute than $\partial G(x)$ and $\partial_B G(x)$ for IEPs.

Generalized Newton Method III: Given $x^0 \in \Re^n$, for $k = 0, 1, \cdots$,

$$x^{k+1} = x^k - V_k^{-1} G(x^k), \tag{8}$$

where $V_k \in \partial_Q G(x^k)$.

The following convergence theorem for the generalized Newton method (8) for solving (7) is proved in [12, Theorem 5.3].

Theorem 2.3 *Suppose that G is defined by (7) and $G(x^*) = 0$. If all $V \in \partial_Q G(x^*)$ are nonsingular and Φ is semismooth at $\Psi(x^*)$, then there exists a neighborhood $N(x^*)$ of x^* such that for any $x^0 \in N(x^*)$ the generalized Newton method III is well defined and is Q-superlinearly convergent. Moreover, if Φ is strongly semismooth at $\Psi(x^*)$ and Ψ' is Lipschitz continuous around x^*, then (8) converges Q-quadratically.*

2.3 Semismooth Matrix Functions

In [20], Sun and Sun extended the concepts of semismoothness and strong semismoothness from vector valued functions to matrix valued functions. Suppose that $G : \mathcal{S} \to \mathcal{S}$ is a locally Lipschitzian matrix valued function. Denote the set of points at which G is Frechét-differentiable by D_G and the Frechét derivative of G at Z (understood as a linear operator from $\mathcal{S}$ to $\mathcal{S}$) by G'_Z.

Definition 2.4 *The* generalized derivative *of G at X, denoted by ∂G_X, is the set defined as following:*

$$\partial G_X := \mathrm{conv}\{\lim G'_Z | \ Z \to X, \ Z \in D_G\}. \tag{9}$$

Definition 2.5 *Suppose that $G : \mathcal{S} \to \mathcal{S}$ is a locally Lipschitzian matrix valued function. G is said to be semismooth at $X \in \mathcal{S}^n$ if G is directionally differentiable at X and for any $V \in \partial G_{X+H}$ and $\Delta X \to 0$,*

$$G(X + \Delta X) - G(X) - V(\Delta X) = o(\|\Delta X\|).$$

G is said to be $p-$order $(0 < p < \infty)$ semismooth at X if G is semismooth at X and

$$G(X + \Delta X) - G(X) - V(\Delta X) = O(\|\Delta X\|^{1+p}). \tag{10}$$

In particular, G is called strongly semismooth at X if G is 1-order semismooth at X.

A function G is said to be a (strongly) semismooth function if it is (strongly) semismooth everywhere on $\mathcal{S}$.

The next result [20, Theorem 3.7] provides a different way of proving semismoothness.

Theorem 2.6 *Suppose that $G : \mathcal{S} \to \mathcal{S}$ is locally Lipschitzian and directionally differentiable in a neighborhood of X. Then for any $p \in (0, \infty)$ the following two statements are equivalent:*

(a) *for any $V \in \partial G_{X+\Delta X}, \Delta X \to 0$,*

$$G(X + \Delta X) - G(X) - V(\Delta X) = O(\|\Delta X\|^{1+p});$$

(b) *for any $X + \Delta X \in D_G, \Delta X \to 0$,*

$$G(X + \Delta X) - G(X) - G'(X + \Delta X; \Delta X) = O(\|\Delta X\|^{1+p}).$$

Composition preserves strong semismoothness in the matrix case as well.

Proposition 2.7 *[20, Lemma 3.9] Let $p \in (0, \infty)$. Suppose $G : \mathcal{S} \to \mathcal{S}$ is (p-order semismooth) semismooth at X and $W : \mathcal{S} \to \mathcal{S}$ is (p-order semismooth) semismooth at $G(X)$. Then the composite function $W \circ G$ is (p-order semismooth) semismooth at X.*

3 Relative Generalized Newton Methods

Let $G : \Re^n \to \Re^n$ be locally Lipschitz continuous. Apart from the semismoothness, another key assumption for the superlinear convergence of the generalized Newton methods I-III is the nonsingularity of $\partial G(x^*)$, $\partial_B G(x^*)$, or $\partial_Q G(x^*)$. However, this may not be satisfied in general for IEPs with multiple eigenvalues. In order to weaken the nonsingularity assumption on the generalized Jacobian, based on the concept of relative generalized gradient introduced by Clarke [4, Page 231], we shall introduce several relative generalized Jacobians and the corresponding generalized Newton methods.

Let S be a subset of $\Re^n$. The S-relative generalized Jacobian of G at x, is defined as follows

$$\partial|_S G(x) := \{V|\ V \text{ is a limit of } V_i \in \partial G(y_i),\ y_i \in S, y_i \to x\}.$$

By [4, Proposition 6.2.1], the following result can be proved in an analogous way. We omit the detail.

Lemma 3.1 *Let G be Lipschitz continuous near x. Then*

(a) $\partial|_S G(x)$ is a compact subset of $\partial G(x)$.

(b) $\partial|_S G(x) = \partial G(x)$ if x lies in the interior part of S; $\partial|_S G(x) = \emptyset$ if $(x + \varepsilon B) \cap S = \emptyset$ for some $\varepsilon > 0$; $\partial|_S G(x)$ is nonempty if $x \in \mathrm{cl}(S)$, the closure of S.

(c) $\partial|_S G(\cdot)$ is upper semicontinuous at x.

Now, we can introduce our first relative generalized Newton method for solving $G(x) = 0$.

Relative Generalized Newton Method I: Given $x^0 \in \Re^n$, for $k = 0, 1, \cdots$, and $x^k \in S$,

$$x^{k+1} = x^k - V_k^{-1} G(x^k), \tag{11}$$

where $V_k \in \partial|_S G(x^k)$.

In the following analysis, we assume that the relative generalized Newton method I does not find a solution in finite steps.

Theorem 3.2 *Suppose that $G(x^*) = 0$ and $x^* \in \mathrm{cl}(S)$. If all $V \in \partial|_S G(x^*)$ are nonsingular and G is semismooth at x^*, then there exists a neighborhood $N(x^*)$ of x^* such that for any $x^0 \in N(x^*) \cap S$*

the relative generalized Newton method I either stops in finite steps with some $x^k \notin S$ or generates an infinite sequence $\{x^k\} \in N(x^*) \cap S$ and the whole sequence converges Q-superlinearly to x^*. Moreover, if G is strongly semismooth at x^*, then the rate of convergence is Q-quadratic.

Proof. By using Lemma 3.1, there exist a neighborhood $N(x^*)$ of x^* and a positive number κ such that for any $x \in N(x^*) \cap S$, all $V \in \partial|_S G(x)$ are nonsingular and

$$\|V^{-1}\| \leq \kappa. \tag{12}$$

Since G is semismooth at x^*, by shrinking $N(x^*)$ if necessary, we have for all $x \in N(x^*) \cap S$ and $V \in \partial|_S G(x)$,

$$\|G(x) - G(x^*) - V(x - x^*)\| \leq \frac{1}{2\kappa}\|x - x^*\|. \tag{13}$$

By using (12) and (13), we have for $k = 0, 1, \cdots$ that

$$\begin{aligned} \|x^{k+1} - x^*\| &= \|x^k - V_k^{-1}G(x^k) - x^*\| \\ &= \|V_k^{-1}[G(x^k) - G(x^*) - V_k(x - x^*)]\| \\ &\leq \|V_k^{-1}\|\|G(x^k) - G(x^*) - V_k(x - x^*)\| \\ &\leq \tfrac{1}{2}\|x^k - x^*\|, \end{aligned}$$

which implies that if (11) does not stop at some step with $x^k \notin S$, then $\{x^k\} \in N(x^*) \cap S$ and the whole sequence converges to x^* linearly.

Next, suppose that (11) does not stop at some step with $x^k \notin S$. Since G is semismooth at x^* and $x^k \to x^*$, we have

$$G(x^k) - G(x^*) - V_k(x^k - x^*) = o(\|x^k - x^*\|),$$

which, together with (11), implies that

$$\begin{aligned} \|x^{k+1} - x^*\| &= \|x^k - V_k^{-1}G(x^k) - x^*\| \\ &= \|V_k^{-1}[G(x^k) - G(x^*) - V_k(x - x^*)]\| \\ &\leq O(\|G(x^k) - G(x^*) - V_k(x - x^*)\|) \\ &= o(\|x^k - x^*\|). \end{aligned}$$

This proves the superlinear convergence of $\{x^k\}$.

By the above argument, we can see that if G is strongly semismooth at x^*, then (11) either stops in finite steps with some $x^k \notin S$ or generates an infinite sequence $\{x^k\} \in N(x^*) \cap S$ and the whole sequence converges Q-quadratically to x^*. This completes the proof. $\square$

The proof of Theorem 3.2 might serve as an example to show the simplicity of the analysis of Newton's methods by using the concept of (strong) semismoothness. Parallel to the definition of $\partial_B G(x)$ and $\partial_Q G(x)$, we define

$$\partial_B|_S G(x) := \{V| \ V \text{ is a limit of } V_i \in \partial_B G(y_i), \ y_i \in S, \ y_i \to x\}$$

and

$$\partial_Q|_S G(x) := \{V| \ V \text{ is a limit of } V_i \in \partial_Q G(y_i), \ y_i \in S, \ y_i \to x\}.$$

Similar to Lemma 3.1, we have

Lemma 3.3 *Let G be Lipschitz continuous near x. Then*

(a) $\partial_B|_S G(x)$ and $\partial_Q|_S G(x)$ are compact subsets of $\partial_B G(x)$ and $\partial_Q G(x)$, respectively.

(b) $\partial_B|_S G(x) = \partial_B G(x)$ and $\partial_Q|_S G(x) = \partial_Q G(x)$, if x lies in the interior part of S; $\partial_B|_S G(x) = \partial_Q|_S G(x) = \emptyset$ if $(x + \varepsilon B) \cap S = \emptyset$ for some $\varepsilon > 0$; both $\partial_B|_S G(x)$ and $\partial_Q|_S G(x)$ are nonempty if $x \in \mathrm{cl}(S)$, the closure of S.

(c) $\partial_B|_S G(\cdot)$ and $\partial_Q|_S G(\cdot)$ are upper semicontinuous at x.

Analogously, we define the second and third relative generalized Newton method.

Relative Generalized Newton Method II (III): Given $x^0 \in \Re^n$, for $k = 0, 1, \cdots$, and $x^k \in S$,

$$x^{k+1} = x^k - V_k^{-1} G(x^k), \tag{14}$$

where $V_k \in \partial_B|_S G(x^k)$ ($V_k \in \partial_Q|_S G(x^k)$ in Method III).

The following theorem can be similarly proved by using Lemma 3.3 and the approach of proving Theorems 3.2. We omit the details.

Theorem 3.4 *Suppose that $G(x^*) = 0$ and $x^* \in \mathrm{cl}(S)$. If all $V \in \partial_B|_S G(x^*)$ ($V \in \partial_Q|_S G(x^*)$ in Method III) are nonsingular and G is semismooth at x^* (Φ is semismooth at $\Psi(x^*)$ in Method III), then there exists a neighborhood $N(x^*)$ of x^* such that for any $x^0 \in N(x^*) \cap S$ the relative generalized Newton methods II and III either stops in finite steps with some $x^k \notin S$ or generates an infinite sequence $\{x^k\} \in N(x^*) \cap S$ and the whole sequence converges Q-superlinearly to x^*. Moreover, if G (Φ in Method III) is strongly semismooth at x^* (at $\Psi(x^*)$ and Ψ' is Lipschitz continuous around x^* in Method III), then the rate of convergence is Q-quadratic.*

4 Strong Semismoothness of Eigenvalues

In this section, we shall prove the strong semismoothness of eigenvalues of symmetric matrices. Suppose $X \in \mathcal{S}$. Then, there exists an orthogonal matrix $Q \in \mathcal{O}$ such that X satisfies

$$Q^T X Q = \Lambda(X) := \operatorname{diag}(\lambda_1(X), \cdots, \lambda_n(X)), \tag{15}$$

where $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$.

We define a "configuration vector" K to distinguish different eigenvalues. Let

$$K := \{k_0, k_1, \cdots, k_l\} \tag{16}$$

with $1 = k_0 < k_1 < \cdots < k_l = n+1$ such that there is a change of eigenvalues at k_i. Namely for $t = 1, \cdots, l$,

$$\lambda_s(X) = \lambda_{k_{t-1}}(X), \;\; s \in [k_{t-1}, k_t - 1], \tag{17}$$

where we use the simple notation $[k_{t-1}, k_t - 1]$ to represent the index set $\{k_{t-1}, k_{t-1}+1, \cdots, k_t - 1\}$.

Let $H \in \mathcal{S}$ and let P (depending on H) be an orthogonal matrix such that

$$P^T(\Lambda(X) + H)P = \Lambda(Y) := \operatorname{diag}(\lambda_1(Y), \cdots, \lambda_n(Y)), \tag{18}$$

where $\lambda_1(Y) \geq \cdots \geq \lambda_n(Y)$ and $Y := \Lambda(X) + H$.

After the above preparation, we can state the following result, which was essentially proved in the proof of Lemma 4.2 of [20].

Lemma 4.1 *For any $H \in \mathcal{S}$ and $H \to 0$, we have*

$$P_{ij} = O(\|H\|), \quad i,j = 1, \cdots, n, \;\; (i,j) \notin \bigcup_{t=1}^{l} \{[k_{t-1}, k_t - 1] \times [k_{t-1}, k_t - 1]\}. \tag{19}$$

Proof. It has been proved in the proof of Lemma 4.2 of [20] that (19) is true for any $H \in \mathcal{S}$ such that $\Lambda(X) + H$ is nonsingular and $H \to 0$.

Next, we prove that (19) is also true for the case that $\Lambda(X) + H$ is singular and $H \to 0$. It is easy to check that the conclusion of this lemma holds if $H = 0$. Hence, we can assume $H \neq 0$. Define

$$\lambda_{\min}(|Y|) = \min_{\lambda_i(Y) \neq 0} |\lambda_i(Y)| \quad \text{and} \quad \tilde{\Lambda} = \operatorname{diag}(\tilde{\lambda}_1, \cdots, \tilde{\lambda}_n),$$

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

$$\tilde{\lambda}_i = \begin{cases} \lambda_i(Y) & \text{if } \lambda_i(Y) \neq 0 \\ \lambda_{\min}(|Y|) \min\{\frac{1}{2}, ||H||^2\} & \text{otherwise} \end{cases} .$$

Denote

$$\tilde{H} = P\tilde{\Lambda}P^T - \Lambda(X).$$

Hence, $P^T[\Lambda(X) + \tilde{H}]P = \tilde{\Lambda}$ is nonsingular. By noting the fact $\tilde{H} = H + O(\|H\|^2)$, it follows that (19) also holds for the case that $\Lambda(X) + H$ is singular and $H \to 0$. This completes the proof. $\square$

Define a "truncated" matrix $O \in \Re^{n\times n}$ as follows:

$$O_{ij} = \begin{cases} P_{ij} & \text{if } (i,j) \in \bigcup_{t=1}^{l} \{[k_{t-1}, k_t - 1] \times [k_{t-1}, k_t - 1]\} \\ 0 & \text{otherwise} \end{cases}, \quad i, j = 1, \cdots, n. \tag{20}$$

Hence, from Lemma 4.1, we know that for any $H \to 0$,

$$O = P + O(\|H\|). \tag{21}$$

It is noted, however, O may not be an orthogonal matrix but has a block-diagonal structure with each block corresponding to a set of identical eigenvalues of X. That is,

$$O = \text{ diag}\,(O_1, \cdots, O_l),$$

where

$$O_t = (P_{ij})_{i,j=k_{t-1}}^{k_t - 1}, \text{ for } t = 1, \cdots, l.$$

Since $P \in \mathcal{O}$, by using Lemma 4.1 and (20), for $t = 1, \cdots, l$ and $i, j = 1, \cdots, k_t - k_{t-1}$, we have for any $H \to 0$,

$$\|(O_t)_{\cdot j}\|^2 = 1 + O(\|H\|^2) \quad \text{and} \quad \langle (O_t)_{\cdot j}, (O_t)_{\cdot i}\rangle = O(\|H\|^2),\ i \neq j. \tag{22}$$

It is obvious from (22) that for any $H \in \mathcal{S}$ sufficiently close to 0 the columns of O_t are independent because

$$\sum_j \beta_j (O_t)_{\cdot j} = 0 \;\Rightarrow\; \beta_j[1 + O(\|H\|^2)] = O(\|H\|^2) \;\Rightarrow\; \beta_j = 0\ \forall\ j.$$

For each $t = 1, \cdots, l$, let $\tilde{P}_t$ be a matrix of the same order of O_t and be obtained by applying the Gram-Schmidt orthogonalization algorithm to each O_t, i.e. for $j = 1, \cdots, k_t - k_{t-1}$, let

$$(\tilde{O}_t)_{\cdot j} = (O_t)_{\cdot j} - \sum_{i=1}^{j-1} \langle (\tilde{P}_t)_{\cdot i}, (O_t)_{\cdot j}\rangle (\tilde{P}_t)_{\cdot i} \text{ and } (\tilde{P}_t)_{\cdot j} = (\tilde{O}_t)_{\cdot j}/\|(\tilde{O}_t)_{\cdot j}\|. \tag{23}$$

By (22) and (23), for $i, j = 1, \cdots, k_t - k_{t-1}$, $t = 1, \cdots, l$ we have for any $H \to 0$ that

$$\|(\tilde{P}_t)_{\cdot j}\|^2 = 1, \quad (\tilde{P}_t)_{\cdot j} = (O_t)_{\cdot j} + O(\|H\|^2) \text{ and } \langle (\tilde{P}_t)_{\cdot j}, (\tilde{P}_t)_{\cdot i}\rangle = 0,\ i \neq j. \tag{24}$$

Denote

$$\tilde{P} = \text{ diag}\,(\tilde{P}_1, \cdots, \tilde{P}_l). \tag{25}$$

Then, we have the following lemma.

Lemma 4.2 *For any $H \in \mathcal{S}$ sufficiently small, the matrix defined $\tilde{P}$ defined by (25) and (23) is an orthogonal matrix and satisfies*

$$\tilde{P}^T \Lambda(X) \tilde{P} = \Lambda(X). \tag{26}$$

Furthermore, for any $H \to 0$,

$$P = \tilde{P} + O(\|H\|). \tag{27}$$

Proof. By (24), we know that each $\tilde{P}_t$, $t = 1, \cdots, l$ is an orthogonal matrix. Since $\lambda_{k_{t-1}}(X) = \cdots = \lambda_{k_t - 1}(X)$, $t = 1, \cdots, l$, we have

$$\tilde{P_t}^T \operatorname{diag}(\lambda_{k_{t-1}}(X), \cdots, \lambda_{k_t-1}(X)) \tilde{P_t} = \operatorname{diag}(\lambda_{k_{t-1}}(X), \cdots, \lambda_{k_t-1}(X)).$$

Hence, $\tilde{P}$ is an orthogonal matrix and satisfies (26). By using (21) and (24), we directly obtain (27). This completes the proof. $\square$

For any $\Delta X \in \mathcal{S}$, let $U \in \mathcal{O}$ (depending on X and ΔX) be any orthogonal matrix such that

$$U^T(X + \Delta X)U = \Lambda(X + \Delta X) := \operatorname{diag}(\lambda_1(X + \Delta X)), \cdots, \lambda_n(X + \Delta X)),$$

where $\lambda_1(X + \Delta X) \geq \cdots \geq \lambda_n(X + \Delta X)$.

By using the above lemma, we have the following result.

Lemma 4.3 *For any $\Delta X \in \mathcal{S}$ sufficiently small, there exists a $V \in \mathcal{O}$ such that*

$$V^T X V = \Lambda(X)$$

and for any $\Delta X \to 0$,

$$U = V + O(\|\Delta X\|). \tag{28}$$

Proof. Let $P = Q^T U$ and $H = Q^T \Delta X Q$, where Q is defined in (15). Then, by Lemma 4.2, for any such defined P, there exists $\tilde{P} \in \mathcal{O}$ such that

$$\tilde{P}^T \Lambda(X) \tilde{P} = \Lambda(X)$$

and for any $\Delta X \to 0$,

$$P = \tilde{P} + O(\|H\|) = \tilde{P} + O(\|\Delta X\|).$$

Let $V = Q\tilde{P}$. Then $V \in \mathcal{O}$,

$$V^T X V = \tilde{P}^T Q^T X Q \tilde{P} = \tilde{P}^T \Lambda(X) \tilde{P} = \Lambda(X),$$

and for any $\Delta X \to 0$,

$$U = V + O(\|\Delta X\|).$$

This completes the proof. $\Box$

A similar result to Lemma 4.3 has also been proved in [3] based on a so called "sin(Θ)" theorem in [18, Theorem 3.4]. The proof provided here is due to a direct comparison between entries of P and $\tilde{P}$ and it indeed furnishes an algorithm for computing V. We hope the reader will find our proof more constructive.

One direct result of Lemma 4.3 is that the (normalized) eigenvectors of symmetric matrices, thought not continuous, are upper Lipschitz continuous. To see this, for any $Z \in \mathcal{S}$, let

$$\mathcal{U}(Z) := \{U \in \mathcal{O} |\ U^T Z U \text{ is diagonal}\}.$$

Let E be a matrix of $\mathcal{S}$ with $E_{ij} = 1$ for all $i, j = 1, \cdots, n$.

Proposition 4.4 *For any $X \in \mathcal{S}$, there exists a constant $\mu > 0$ such that*

$$\mathcal{U}(X + \Delta X) \subseteq \mathcal{U}(X) + \mu ||\Delta X|| E \tag{29}$$

for all ΔX sufficiently small.

Proof. For any $U \in \mathcal{U}(X + \Delta X)$, there exists a diagonal matrix $D(X + \Delta X)$ such that

$$U^T(X + \Delta X)U = D(X + \Delta X).$$

Let $R \in \Re^{n\times n}$ be a permutation matrix such that

$$RD(X + \Delta X)R^T = \Lambda(X + \Delta X)$$

with $\lambda_1(X + \Delta X) \geq \cdots \geq \lambda_n(X + \Delta X)$. Let $\tilde{U} = UR^T$. Then we obtain $\tilde{U}^T(X + \Delta X)\tilde{U} = \Lambda(X + \Delta X)$. Hence, by Lemma 4.3, there exists a $\tilde{V} \in \mathcal{O}$ such that $\tilde{V}^T X \tilde{V} = \Lambda(X)$ and

$$\tilde{U} = \tilde{V} + O(\|\Delta X\|),$$

i.e.,

$$U = \tilde{V}R + O(\|\Delta X\|)$$

because $R^T = R^{-1}$ and $\|R^T\| = 1$. Let $V = \tilde{V}R$. Then

$$V^T V = R^T \tilde{V}^T \tilde{V} R = R^T R = I \quad \text{and} \quad V^T X V = (\tilde{V}R)^T X \tilde{V} R = R^T \Lambda(X) R$$

is a diagonal matrix. Hence, we have proved $V \in \mathcal{U}(X)$ and for $\Delta X \to 0$

$$U = V + O(\|\Delta X\|).$$

This implies that there exists a $\mu > 0$ such that (29) holds. $\square$

In Section 1 we have seen from an example of a two by two matrix that the eigenvalues are not differentiable if X has multiple eigenvalues. This can be easily extended to the general case: $\Lambda(\cdot)$ is not differentiable at X if X has multiple eigenvalues. To see this, suppose that X has multiple eigenvalues. Let $Q \in \mathcal{O}$ be such that $Q^T X Q = \Lambda(X)$ with $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$. Without loss of generality, we assume $\lambda_1(X) = \lambda_2(X)$.

Define $Z_\varepsilon = \begin{bmatrix} \lambda_1(X) & \varepsilon \\ \varepsilon & \lambda_2(X) \end{bmatrix}$, where $\varepsilon \in \Re$ is a parameter. Let

$$X_\varepsilon = Q^T \operatorname{diag}(Z_\varepsilon, \lambda_3(X), \cdots, \lambda_n(X)) Q.$$

Then, for all ε sufficiently small, we have

$$\lambda_1(X_\varepsilon) = \lambda_1(X) + |\varepsilon|.$$

This clearly shows $\lambda_1(\cdot)$, and so $\Lambda(\cdot)$, is not differentiable at X.

On the other hand, by [23, pp. 66-68] and [22, Theorem 2.3] we know that if X has distinct eigenvalues then $\Lambda(\cdot)$ is analytic in a neighborhood of X. Hence, we have the following result.

Lemma 4.5 $\Lambda(\cdot)$ *is differentiable at X if and only if $X \in \mathcal{S}$ has distinct eigenvalues. Moreover, if X has distinct eigenvalues, $\Lambda(\cdot)$ is analytic in a neighborhood of X.*

By Ky Fan [7] and Mifflin [10], we have

Lemma 4.6 $\Lambda(\cdot)$ *is a locally Lipschitz continuous and semismooth function.*

Next, we will use Lemmas 4.3 and 4.6 to derive a useful formula for the derivative of $\Lambda(X)$ when $X \in \mathcal{S}$ has distinct eigenvalues.

For any $X \in \mathcal{S}$ and $\Delta X \in \mathcal{S}$, let $Q(X + \Delta X) \in \mathcal{O}$ be such that

$$Q(X + \Delta X)^T (X + \Delta X) Q(X + \Delta X) = \Lambda(X + \Delta X)$$

with $\lambda_1(X + \Delta X) \geq \cdots \geq \lambda_n(X + \Delta X)$. Define

$$q_i(X + \Delta X) = (Q(X + \Delta X))_{\cdot i}, \quad i = 1, \cdots, n.$$

Lemma 4.7 *For any $X \in \mathcal{S}$, if X has distinct eigenvalues, then $\Lambda(\cdot)$ is continuously differentiable at X and for any $\Delta X \in S$,*

$$(\lambda_i)'_X(\Delta X) = q_i(X)^T \Delta X q_i(X), \quad i = 1, \cdots, n. \tag{30}$$

Proof. By Lemma 4.5, $\Lambda(\cdot)$ is continuously differentiable at X. We now prove (30). By Weyl's theorem [1, p. 63], there exists a number $\delta > 0$ such that for all $\|\Delta X\| \leq \delta$, $X + \Delta X$ also has distinct eigenvalues. Then, for all $\|\Delta X\| \leq \delta$, by ignoring the choice of sign possible for each $q_i(X + \Delta X)$, the set $\{q_i(X + \Delta X)\}_{i=1}^n$ is unique and

$$\langle q_i(X + \Delta X), q_j(X + \Delta X) \rangle = 0, \quad i \neq j, \ i, j = 1, \cdots, n. \tag{31}$$

By reducing δ if necessary, from Lemma 4.3, there exists a $\mu > 0$ such that for all $||\Delta X|| \leq \delta$,

$$\|q_i(X + \Delta X) - q_i(X)\| \leq \mu \|\Delta X\|. \tag{32}$$

For each $i = 1, \cdots, n$, we have

$$\lambda_i(X + \Delta X) - \lambda_i(X) = q_i(X + \Delta X)^T (X + \Delta X) q_i(X + \Delta X) - q_i(X)^T X q_i(X),$$

which, together with (32) and Lemma 4.6, implies that

$$\begin{aligned}
&\lambda_i(X + \Delta X) - \lambda_i(X) - q_i^T(X) \Delta X q_i(X) \\
&= q_i(X + \Delta X)^T X q_i(X + \Delta X) - q_i(X)^T X q_i(X) + O(\|\Delta X\|^2) \\
&= [q_i(X + \Delta X) - q_i(X)]^T X [q_i(X) + q_i(X + \Delta X)] + O(\|\Delta X\|^2) \\
&= \lambda_i(X) [q_i(X + \Delta X) - q_i(X)]^T q_i(X) + [q_i(X + \Delta X) - q_i(X)]^T (X + \Delta X) q_i(X + \Delta X) \\
&\quad - [q_i(X + \Delta X) - q_i(X)]^T \Delta X q_i(X + \Delta X) + O(\|\Delta X\|^2) \\
&= \lambda_i(X) [q_i(X + \Delta X) - q_i(X)]^T q_i(X) \\
&\quad + [\lambda_i(X) + O(\|\Delta X\|)] [q_i(X + \Delta X) - q_i(X)]^T q_i(X + \Delta X) + O(\|\Delta X\|^2) \\
&= \lambda_i(X) [q_i(X + \Delta X) - q_i(X)]^T [q_i(X + \Delta X) + q_i(X)] + O(\|\Delta X\|^2) \\
&= \lambda_i(X) \left[\|q_i(X + \Delta X)\|^2 - \|q_i(X)\|^2\right] + O(\|\Delta X\|^2) = O(\|\Delta X\|^2),
\end{aligned}$$

where the last equality is due to the fact that the norms of q_i's are one. This implies that (30) holds. $\square$

By Lemma 4.5, we have the (dense) subsets of $\mathcal{S}$ on which Λ is continuously differentiable

$$D_\Lambda = \{Y \in \mathcal{S} |\ Y \text{ has distinct eigenvalues}\}.$$

The following result is our main theorem of this section.

Theorem 4.8 $\Lambda(\cdot)$ *is a strongly semismooth function.*

Proof. Suppose that $X_0 \in \mathcal{S}$ is a given matrix. For any $X \in D_\Lambda$, denote $\Delta X = X - X_0$. For $i = 1, \cdots, n$, from $Xq_i(X) = \lambda_i(X)q_i(X)$, we have

$$q_i(X)^T X_0 q_i(X) + q_i(X)^T \Delta X q_i(X) = \lambda_i(X). \tag{33}$$

By Lemma 4.3, there exists a $\mu > 0$ such that for any X sufficiently close to X_0 there exists a matrix $Q(X_0) \in \mathcal{O}$ (depending on the choice of X) such that $Q(X_0)^T X_0 Q(X_0) = \Lambda(X_0)$ and

$$\|q_i(X) - q_i(X_0)\| \leq \mu||X - X_0||, \tag{34}$$

where $q_i(X_0) := (Q(X_0))_{\cdot i}$, $i = 1, \cdots, n$. Hence, from (33), (34) and the local Lipschitz continuity of $\Lambda(\cdot)$ (Lemma 4.6), for $i = 1, \cdots, n$, $X \in D_\Lambda$ and $\Delta X \to 0$, we have

$$\begin{aligned}
\lambda_i(X) &= q_i(X)^T \left[X_0 q_i(X_0) + X_0(q_i(X) - q_i(X_0))\right] + q_i(X)^T \Delta X q_i(X) \\
&= \lambda_i(X_0) q_i(X)^T q_i(X_0) + q_i(X)^T X \left[q_i(X) - q_i(X_0)\right] + O(\|\Delta X\|^2) \\
&\quad + q_i(X)^T \Delta X q_i(X) \\
&= \lambda_i(X_0) q_i(X)^T q_i(X_0) + \lambda_i(X) q_i(X)^T \left[q_i(X) - q_i(X_0)\right] \\
&\quad + q_i(X)^T \Delta X q_i(X) + O(\|\Delta X\|^2) \\
&= \lambda_i(X_0) q_i(X)^T q_i(X_0) + \left[\lambda_i(X_0) + O(\|\Delta X\|)\right] q_i(X)^T \left[q_i(X) - q_i(X_0)\right] \\
&\quad + q_i(X)^T \Delta X q_i(X) + O(\|\Delta X\|^2) \\
&= \lambda_i(X_0) q_i(X)^T q_i(X) + q_i(X)^T \Delta X q_i(X) + O(\|\Delta X\|^2) \\
&= \lambda_i(X_0) + q_i(X)^T \Delta X q_i(X) + O(\|\Delta X\|^2),
\end{aligned} \tag{35}$$

which, according to Lemma 4.7, implies

$$\lambda_i(X) - \lambda_i(X_0) - (\lambda_i)'_X(\Delta X) = O(\|\Delta X\|)^2, \quad i = 1, \cdots, n.$$

This, together with Theorem 2.6, implies that for $X \to X_0$ and $V \in \partial\Lambda_X$,

$$\Lambda(X) - \Lambda(X_0) - V(X - X_0) = O(\|X - X_0\|^2).$$

Hence, by Lemma 4.6 and the definition of strong semismoothness, $\Lambda(\cdot)$ is strongly semismooth at X_0. Since $X_0 \in \mathcal{S}$ is chosen arbitrarily, $\Lambda(\cdot)$ is a strongly semimsooth function. $\square$

5 Newton's Methods for Inverse Eigenvalue Problems

In this section, we shall show how the strong semismoothness of eigenvalues of symmetric matrices can be used to analyze the quadratic convergence of some Newton's methods for solving inverse eigenvalue problems (IEPs). Unless stated otherwise, $A : \Re^n \to \mathcal{S}$ is assumed to be continuously differentiable everywhere and $F : \Re^n \to \Re^n$ is defined by (2), i.e.,

$$F(c) = \begin{bmatrix} \lambda_1(A(c)) - \lambda_1^* \\ \vdots \\ \lambda_n(A(c)) - \lambda_n^* \end{bmatrix},$$

where $\{\lambda^*\}_{i=1}^n$ are given n numbers and arranged in the decreasing order. Then the IEP is equivalent to find a $c^* \in \Re^n$ such that $F(c^*) = 0$.

For any $c \in \Re^n$, let $\mathcal{Q}(c) \subseteq \mathcal{O}$ be a subset of $\Re^{n\times n}$ such that for any $Q(c) \in \mathcal{Q}(c)$ we have

$$Q(c)^T A(c) Q(c) = \Lambda(A(c))$$

with $\lambda_1(A(c)) \geq \cdots \geq \lambda_n(A(c))$. For any $Q(c) \in \mathcal{Q}(c)$, define

$$q_i(c) = (Q(c))_{\cdot i}, \quad i = 1, \cdots, n.$$

Let $\partial A(c)/\partial c_j$ be the partial derivative of $A(c)$ with respect to c_j, $j = 1, \cdots, n$. Then for any $c \in \Re^n$

$$\partial_Q F(c) = \partial_B \Lambda_{A(c)}(A'(c))$$

is well defined. By using Lemmas 4.5 and 4.7 and [20, Theorem 2.5], we have the following result.

Proposition 5.1

(a). *For any $c \in \Re^n$, $V \in \partial_Q F(c)$ if and only if there exists a $Q(c) \in \mathcal{Q}(c)$ such that*

$$V_{i\cdot} = \Big[q_i(c)^T (\partial A(c)/\partial c_1) q_i(c), \cdots, q_i(c)^T (\partial A(c)/\partial c_n) q_i(c)\Big], \quad i = 1, \cdots, n. \tag{36}$$

(b). *If $c \in \Re^n$ is such that $A(c)$ has distinct eigenvalues, then F is continuously differentiable at c and for any $Q(c) \in \mathcal{Q}(c)$,*

$$F_i'(c) = \left[q_i(c)^T (\partial A(c)/\partial c_1) q_i(c), \cdots, q_i(c)^T (\partial A(c)/\partial c_n) q_i(c)\right]. \tag{37}$$

Hence, according to Proposition 5.1, a generalized Newton method for solving the IEP can be described as follows.

Algorithm 5.2 (A Generalized Newton Method)

Step 0. *Choose a starting point value c^0. $k := 0$.*

Step 1. *Compute a $Q(c^k) \in \mathcal{Q}(c^k)$ and form $V_k \in \partial_Q F(c^k)$ according to Proposition 5.1.*

Step 2. *Compute Δc^k by*

$$F(c^k) + V_k \Delta c^k = 0. \tag{38}$$

Define $c^{k+1} := c^k + \Delta c^k$.

Step 3. *Replace k by $k+1$ and go to Step 1.*

In the above generalized Newton method, at the k-th step one needs to compute eigenvectors $Q(c^k)$ and eigenvalues $\Lambda(A(c^k))$. Once they are computed, $F(c^k)$ and $V^k \in \partial_Q F(c^k)$ can be formulated easily. If $A(c)$ takes form (1) and at each step $A(c^k)$ has distinct eigenvalues, Algorithm 5.2 reduces to the Newton method considered by many authors, e.g., see [11, 9] and references therein.

Theorem 5.3 *Suppose that F is defined by (2) and $F(c^*) = 0$. If all $V \in \partial_Q F(c^*)$ are nonsingular and A' is Lipschitz continuous around c^*, then there exists a neighborhood $N(c^*)$ of c^* such that for any $c^0 \in N(c^*)$ Algorithm 5.2 is well defined and the iterates $\{c^k\}$ converge to c^* Q-quadratically.*

Proof. From Theorem 4.8, we know that $\Lambda(\cdot)$ is strongly semismooth everywhere. Hence, by Theorem 2.3 we obtain the conclusion of this theorem. □

Theorem 5.3 contains a very general convergence result for the quadratic convergence of Newton's methods for solving IEPs. However, the nonsingularity assumption on $\partial_Q F(c^*)$ is too strong for IEPs when $A(c^*)$ has multiple eigenvalues. To relax this condition, let $S \subseteq \Re^n$ be defined by

$$S = \{c \in \Re^n \mid A(c) \text{ has distinct eigenvalues}\}. \tag{39}$$

Then, by Lemma 4.5 and Proposition 5.1 for any $c \in S$, $F(\cdot)$ is continuously differentiable at c and

$$\partial_B F(c) = \partial_Q F(c) = \partial F(c) = \{F'(c)\}.$$

Theorem 5.4 *Suppose that F is defined by (2), $F(c^*) = 0$, and S is defined by (39). If (i) for each k, $c^k \in S$ and $c^* \in \mathrm{cl}S$; (ii) all $V \in \partial_B|_S F(c^*)$ are nonsingular; and (iii) A' is Lipschitz continuous around c^*, then there exists a neighborhood $N(c^*)$ of c^* such that for any $c^0 \in N(c^*)$ Algorithm 5.2 is well defined and the iterates $\{c^k\}$ converge to c^* Q-quadratically.*

Proof. By using Theorems 3.4 and 4.8, we obtain this theorem. □

In Theorem 5.4, we only need the nonsingularity of $\partial_B|_S F(c^*)$ rather than $\partial_Q F(c^*)$. The price to pay is that all the iterates must stay in S, where S is defined by (39). However, according to [9], this is a reasonable assumption for IEPs.

Theorem 5.5 *Suppose that F is defined by (2) and $F(c^*) = 0$. If (i) for each k, $A(c^k)$ has distinct eigenvalues and V_k is invertible; (ii) $\limsup_{k\to\infty} \|{V_k}^{-1}\| < \infty$; and (iii) A' is Lipschitz continuous around c^*, then there exists a neighborhood $N(c^*)$ of c^* such that for any $c^0 \in N(c^*)$, the iterates $\{c^k\}$ generated by Algorithm 5.2 converge to c^* Q-quadratically*

Proof. By using Theorem 3.4 with $S = \{c^0, c^1, \cdots\}$, we obtain the conclusion of this theorem. □

It was probably Nocedal and Overton [11] who firstly discussed the quadratic convergence of Newton's methods for solving IEPs with multiple eigvenvalues. In their proof a Theorem of Rellich [16] on analytic matrix functions was invoked. In [9], by using the eigenprojector, Friedland et al. presented a different elegant proof on the quadratic convergence of Newton's method for solving IEPs with multiple eigenvalues. The latter did not use Rellich's theorem. Our results in this paper could be thought of as a generalization of their Method I by explicitly exploring the strong semismoothness of the eigenvalue functions.

Next, we consider a generalized inverse eigenvalue problem (GIEP). Let $C : \Re^n \to \mathcal{S}$ and $D : \Re^n \to \mathcal{S}$ be continuously differentiable and $D(c)$ be positive definite whenever $c \in \Omega$, an open subset of $\Re^n$. Given n real numbers $\{\lambda_i^*\}_{i=1}^n$, which are arranged in the decreasing order $\lambda_1^* \geq \cdots \geq \lambda_n^*$, the GIEP is to find a vector $c^* \in \Omega$ such that the symmetric generalized eigenvalue problem $C(c^*)x = \lambda D(c^*)x$ has the prescribed eigenvalues $\lambda_1^*, \cdots, \lambda_n^*$. If $D(c) \equiv I$, then the GIEP is the IEP considered above. Dai and Lancaster [6] and Dai [5] considered another special case of the GIEP, i.e., $C(c)$ and $D(c)$ are defined by

$$C(c) = C_0 + \sum_{i=1}^{n} c_i C_i, \quad D(c) = D_0 + \sum_{i=1}^{n} c_i D_i, \tag{40}$$

where $C_0, C_1, \cdots, C_n, D_0, D_1, \cdots D_n \in \mathcal{S}$ and $D(c)$ is positive definite whenever $c \in \Omega$.

For any $c \in \Omega$, let

$$B(c) = D(c)^{\frac{1}{2}}, \tag{41}$$

which is well defined because for any $c \in \Omega$, $D(c)$ is symmetric and positive definite. Let

$$A(c) = B(c)^{-1}C(c)B(c)^{-1}, \quad c \in \Omega. \tag{42}$$

Define $F : \Omega \to \Re^n$ by

$$F(c) = \begin{bmatrix} \lambda_1(A(c)) - \lambda_1^* \\ \vdots \\ \lambda_n(A(c)) - \lambda_n^* \end{bmatrix}, \qquad c \in \Omega. \tag{43}$$

Then the GIEP is equivalent to find a $c^* \in \Omega$ to be a solution of the following equation

$$F(c) = 0, \qquad c \in \Omega. \tag{44}$$

Hence, generalized Newton method 5.2 can be applied to GIEPs. We omit the details here.

When $C(c)$ and $D(c)$ take the form (40), Dai and Lancaster [6] proposed the following Newton method for solving the GIEP:

For any $c \in \Omega$, let $P(c) \in \Re^{n\times n}$ be such that

$$P(c)^T C(c) P(c)^T = \Lambda(A(c)), \quad P(c)^T D(c) P(c) = I. \tag{45}$$

Let

$$Q(c) = B(c)P(c), \quad c \in \Omega.$$

Then $Q(c) \in \mathcal{O}$ and

$$Q(c)^T A(c) Q(c) = \Lambda(A(c)), \quad c \in \Omega. \tag{46}$$

Hence, we can always compute a $Q(c)$ to satisfy (46) first and then solve (45) by letting $P(c) = B(c)^{-1}Q(c)$. For $c \in \Omega$ and $i = 1, \cdots, n$, let

$$p_i(c) = (P(c))_{\cdot i}.$$

For $c \in \Omega$, define $W(c) \in \Re^{n\times n}$ by

$$W(c)_{ij} = p_i(c)^T (C_j - \lambda_i(A(c)) D_j) p_i(c), \quad i, j = 1, \cdots n. \tag{47}$$

Then, Dai and Lancaster's algorithm [6] can be described as

Algorithm 5.6 (A Newton Method of Dai and Lancaster [6])

Step 0. *Choose a starting point value* c^0. $k := 0$.

Step 1. *Compute* $C(c^k) = C_0 + \sum_{j=1}^n c_j^k C_j, \quad D(c^k) = D_0 + \sum_{j=1}^n c_j^k D_j$.

Step 2. *Find* $P(c^k)$ *such that (45) holds. Compute a matrix* $W(c^k)$ *according to (47).*

Step 3. *Compute* Δc^k *by*

$$F(c^k) + W(c^k)\Delta c^k = 0. \tag{48}$$

Define $c^{k+1} := c^k + \Delta c^k$.

Step 4. *Replace* k *by* $k+1$ *and go to Step 1.*

The following theorem gives an affirmative answer to a conjecture made in [6, p. 11] on the quadratic convergence of Algorithm 5.6, which was supported by numerical experiments.

Theorem 5.7 *Suppose that* $c^* \in \Omega$ *is such that* $F(c^*) = 0$ *and* $A(\cdot)$ *is defined by (42). If (i) for each* k, $A(c^k)$ *has distinct eigenvalues and* $W(c^k)$ *is invertible; and (ii)* $\limsup_{k\to\infty} \|W(c^k)^{-1}\| < \infty$, *then there exists a neighborhood* $N(c^*)$ *of* c^* *such that for any* $c^0 \in N(c^*)$, *the iterates* $\{c^k\}$ *generated by Algorithm 5.6 converge to* c^* *Q-quadratically.*

Proof. Since $A(c^k)$ has distinct eigenvalues, F is continuously differentiable at c^k. From [6, Equation (2.6)], we know that

$$W(c^k) = F'(c^k),$$

which, together with Proposition 5.1, implies that Algorithm 5.6 is a special case of Algorithm 5.2 under our conditions. By using Theorem 5.5, we get the conclusion of the theorem. $\square$

6 Summary and Possible Future Research Topics

In this paper we review basic concepts of semismoothness and several Newton's methods for semismooth equations. We show the strong semimsoothness of eigenvalues as functions of the symmetric matrices and then demonstrate how this result can be used to provide a unified analysis for the quadratic convergence of Newton's methods for IEPs and GIEPs.

We feel that the following several research topics can be further investigated. Firstly, to apply the strong semimsoothness of eigenvalues of symmetric matrices to the convergence analysis of Newton methods for solving other forms of IEPs, for example, the least square form of IEPs [9]. Secondly, to use the recently developed quasi-Newton methods for solving structured nonsmooth equations [19] to analyze the superlinear convergence of quasi-Newton methods for IEPs and GIEPs like Method II in [9]. When IEPs have distinct eigenvalues, Chan et al. [2] provided an approach towards this. It is still open if IEPs have multiple eigenvalues. Thirdly, it is proved in [21] that any nonsmooth function has approximate smoothing functions. By using these smoothing functions, globalized smoothing Newton methods developed in [14] can be applied to solve nonsmooth equations directly. It is then an interesting topic to construct computable smoothing functions for IEPs in order to globalize Newton's methods discussed in Section 5.

References

[1] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1997.

[2] R. H. Chan, S. -F. Xu and H. -M. Zhou, "On the convergence rate of a quasi-Newton method for inverse eigenvalue problems," *SIAM Journal on Numerical Analysis*, **36** (1999) 436–441.

[3] X. Chen and P. Tseng, "Non-interior continuation methods for solving semidefinite complementarity problems," Manuscript, Department of Mathematics, University of Washington, Seattle, Washington 98195, USA, 1999.

[4] F. Clarke, *Optimization and Nonsmooth Analysis,* John Wiley and Sons, New York, 1983.

[5] H. Dai, "An algorithm for symmetric generalized inverse eigenvalue problems," *Linear Algebra and its Applications*, **296** (1999) 79–98.

[6] H. Dai and P. Lancaster, "Newton's method for a generalized inverse eigenvalue problem," *Numerical Linear Algebra with Applications*, **4** (1997) 1–21.

[7] K. Fan, "On a theorem of Weyl concerning eigenvalues of linear transformations. I.", *Proc. Nat. Acad. Sci. U. S. A.*, **35** (1949) 652–655.

[8] A. Fischer, "Solution of monotone complementaity problems with locally Lipschitzian functions," *Mathematical Programming,* **76** (1997) 513–532.

[9] S. Friedland, J. Nocedal and M. L. Overton, "The formulation and analysis of numerical methods for inverse eigenvalue problems," *SIAM Journal on Numerical Analysis*, **24** (1987) 634–667.

[10] R. Mifflin, "Semismooth and semiconvex functions in constrained optimization", *SIAM Journal on Control and Optimization*, **15** (1977) 957–972.

[11] J. Nocedal and M. L. Overton, "Numerical methods for solving inverse eigenvalue problems," Lecture Notes in Math. No. 1005, V. Pereya and A. Reinoza, eds., Springer-Verlag, New York, Berlin, 1983, pp. 212–226.

[12] F. Potra, L. Qi and D. Sun, "Secant methods for semismooth equations," *Numerische Mathematik*, **80** (1998) 305–304.

[13] L. Qi, "Convergence analysis of some algorithms for solving nonsmooth equations", *Mathematics of Operations Research*, **18** (1993) 227–244.

[14] L. Qi, D. Sun and G. Zhou, "A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities," *Mathematical Programming*, **87** (2000) 1-35.

[15] L. Qi and J. Sun, "A nonsmooth version of Newton's method", *Mathematical Programming*, **58** (1993) 353–367.

[16] F. Rellich, *Perturbation Theory of Eigenvalue Problems*, Gordon and Breach Science Publishers, New York, London, Paris, 1969.

[17] R. T. Rockafellar, *Convex Analysis,* Princeton University Press, New Jersey, 1970.

[18] G. W. Stewart and J. -G. Sun, *Matrix Perturbation Theory*, Academic Press, New York, 1990.

[19] D. Sun and J. Han, “Newton and quasi-Newton methods for a class of nonsmooth equations and related problems”, *SIAM Journal on Optimization*, **7** (1997) 463–480.

[20] D. Sun and J. Sun,“Semismooth matrix valued functions”, Manuscript, Dept of Decision Sciences, National University of Singapore, http://www.fba.nus.edu.sg/depart/ds/dsstaff/fbasunj/index.htm, 1999. Accepted by Mathematics of Operations Research.

[21] D. Sun and L. Qi, “Solving variational inequality problems via smoothing-nonsmooth reformulations,” *Journal of Computational and Applied Mathematics*, **129** (2001) 37–62.

[22] J. -G. Sun, “Eigenvalues and eigenvectors of a matrix dependent on several parameters,” *Journal of Computational Mathematics*, **3** (1985) 351–364.

[23] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

[24] S. -F. Xu, *An Introduction to Inverse Algebraic Eigenvalue Problems*, Beijing University Press, Beijing, 1998.