SMOOTH CONVEX APPROXIMATION AND ITS APPLICATIONS

SHI SHENGYUAN

(B.Sc.(Hons.), ECNU)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

2004

Acknowledgements

I would like to thank my supervisor, Dr Sun Defeng, who has been helping me when I am in trouble, encouraging me when I lose confidence and sharing happiness with me when I make progress. This thesis would not come out without the invaluable suggestion and patient guidance from Dr Sun Defeng. If not for him, I would not have learned so much. My thanks also go out to the Department of Mathematics, National University of Singapore. Thanks to all staffs and friends who support me during these two years.

Many people have made important contributions to this thesis by providing me with insightful feedback and astute reviews. Without their contributions, I would have been unable to meet the demands and deadlines of this thesis.

Shi Shengyuan

Jul. 2004

Contents

Summary

It is well known that the eigenvalues of a real symmetric matrix are not everywhere differentiable. Ky Fan's classical result [11] states that each eigenvalue of a symmetric matrix is the difference of two convex functions, which implies that the eigenvalues are semismooth functions. Based on a recent result of Sun and Sun [30], it is further proved that the eigenvalues of symmetric matrix are strongly semismooth everywhere. The concept of semismoothness of functionals was originally studied by Mifflin [19]. Later Qi and Sun developed this idea to strong semismoothness [26] for vector valued functions. Recently, both concepts are further extended to matrix valued functions [29]. Generally speaking, *strong semismoothness* of an equation is tied with quadratic convergence of the 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 strongly semismooth functions.

In this thesis, we consider a smooth approximation function to the sum of the κ largest eigenvalues. Thus the κth largest eigenvalue function can be approximated by the difference of two smooth functions. To make it applicable to a wide class of applications, the study is conducted on the composite function of a smoothing function $f_\kappa(\varepsilon, \cdot)$ and the eigenvalue function $\lambda(\cdot)$. Namely, we find a smoothing function $f_\kappa(\varepsilon, \lambda(X))$ for $f_\kappa(\lambda(X))$, such that $f_\kappa(\varepsilon, \lambda(Y)) \to f_\kappa(\lambda(X))$, as $(\varepsilon, Y) \to (0^+, X)$. It is proved in [28] that via convolution any nonsmooth function has its approximate smoothing function. But the proof does not give any concrete smoothing function. The main aim of this thesis is to find a computable smooth function to approximate every eigenvalue function.

As applications, we can use this smooth convex approximation function to solve some minmax problems and inverse eigenvalue problems (IEPs).

The organization of this thesis is as follows. Some introduction of previous research works done in this area is presented in Chapter 1. Then in Chapter 2, we give the smoothing approximation function of the κth largest component which is the difference of two convex smooth functions. We use primal-dual excessive gap algorithm to test the computability and give the results. Chapter 3 concentrates on showing the strong semismoothness of $g_\kappa(\varepsilon, x)$. Chapter 4, we give out the most important discovery in this thesis: we find the smoothing approximate function for the sum of the κ largest eigenvalues. Therefore every eigenvalue function can be approximated by the difference of two smooth functions. In the last Chapter, we apply the smoothing approximate function to solve some special case of inverse eigenvalue problems.

List of Notation

- $A, B, \ldots$ denote matrices.
- $\mathcal{S}_n$ is the set of real symmetric matrices; $\mathcal{O}_n$ is the set of all $n \times n$ orthogonal matrices.
- A superscript "T" represents the transpose of matrices and vectors.
- For a matrix M, $M_{i\cdot}$ and $M_{\cdot j}$ represent the ith row and jth column of M, respectively. M_{ij} denotes the (i,j)th entry of M.
- A diagonal matrix is written as $\mathrm{Diag}(\beta_1, \ldots, \beta_n)$ and a block-diagonal matrix is denoted by $\mathrm{Diag}(B_1, \ldots, B_s)$ where $B_1, \ldots, B_s$ are matrices.
- We use $\circ$ to denote the Hadamard product between matrices, i.e.

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

- 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

$$\mathcal{A}y := \sum_{i=1}^{m} y_i A_i \quad \text{and} \quad A(y) := A_0 + \mathcal{A}y, \quad \forall y \in \mathbb{R}^m. \tag{1}$$

- 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 (1) 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 $D \in \mathcal{S}_n$,

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

- The eigenvalues of $X \in \mathcal{S}$ is designated by $\lambda_i(X)$, $i = 1, \ldots, n$.

- We write $X = O(\alpha)$ (respectively, $\mathrm{o}(\alpha)$) if $\|X\|/|\alpha|$ is uniformly bounded (respectively, tends to zero) as $\alpha \to 0$.

- F represents the scalar field of either real $\mathbb{R}$ or complex $\mathbb{C}$.

- $\mathcal{M}, \mathcal{N}, \ldots$ denote certain subsets of square matrices of which the size is clear from the context.

Chapter 1

Introduction

As we mentioned in the part of summary, the eigenvalue function is usually not differentiable, which inevitably gives rise to extreme difficulties in a gradient-dependent numerical method (e.g., Newton's method). To see this point more clearly, let us consider the following example

$$X = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \tag{1.1}$$

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

$$\lambda_1(X) = \frac{x_1 + x_3 + \sqrt{(x_1 - x_3)^2 + 4x_2^2}}{2} \tag{1.2}$$

and

$$\lambda_2(X) = \frac{x_1 + x_3 - \sqrt{(x_1 - x_3)^2 + 4x_2^2}}{2} \tag{1.3}$$

Since $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$ are not differentiable at X with $x_1 = x_3$ and $x_2 = 0$, the classical optimization methods (often using the information of gradient and Hessian of objective functions) may get into trouble. The works conducted recently by Lewis [16], Lewis and Sendov [17], Qi and Yang [25] within a very general framework of spectral functions open ways in such extensions. A function f on the space of $n-$by$-n$ real symmetric matrices is called *spectral* if it depends only on the

eigenvalues of its argument. Spectral functions are just symmetric functions of the eigenvalues. We can think of a spectral function 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, 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

$$\begin{aligned}(f \circ \lambda)(X) &= f(\lambda(X)) \\ &= f(\lambda_1(X), \lambda_2(X), \ldots, \lambda_n(X)) \quad \text{for any} \quad X \in \mathcal{S}_n.\end{aligned} \tag{1.4}$$

It seems that the spectral function, thought of as a composition of $\lambda(\cdot)$ and a symmetric function f, would inherit the nonsmoothness of the eigenvalue function. However, Lewis proved in [16] that $(f \circ \lambda)$ is indeed (strictly) differentiable at $X \in \mathcal{S}$ if and only if f is (strictly) differentiable at $\lambda(X)$. Moreover, it is further proved in [17] that $(f \circ \lambda)$ is twice (continuously) differentiable at $X \in \mathcal{S}$ if and only if f is twice (continuously) differentiable at $\lambda(X)$. These results play an important role in this thesis.

Spectral function is normally nondifferentiable. For example ,let

$$f_1(x) := \max\{x_1, \ldots, x_n\} \tag{1.5}$$

Then

$$\lambda_1(X) = (f_1 \circ \lambda)(X), \quad \forall X \in \mathcal{S}_n, \tag{1.6}$$

where $\lambda(X)$ is the vector function of eigenvalues of X, $\lambda_1(X)$ is the maximum eigenvalue function, i.e., $\lambda_1(X) \geq \lambda_2(X) \geq \ldots \geq \lambda_n(X)$. According to (1.2), we know spectral function $(f_1 \circ \lambda)(X)$ may not be differentiable.

A well known smoothing function to the maximum function (1.5) is the exponential penalty function:

$$f_1(\varepsilon, x) := \varepsilon \ln \left(\sum_{i=1}^{n} e^{x_i/\varepsilon} \right), \quad \text{on } \mathbb{R}_{++} \times \mathbb{R}^n. \tag{1.7}$$

It is a C^∞ convex function and has the following uniform approximation to f_1 [7]:

$$0 \leq f_1(\varepsilon, x) - f_1(x) \leq \varepsilon \ln n. \tag{1.8}$$

The penalty function, sometimes called the aggregation function, is used in a number of occasions [2, 14, 18, 23, 24, 32, 33].

It is easy to see that the exponential penalty function (1.7) is symmetric in $\mathbb{R}^n$ and the well defined spectral function $f_1(\varepsilon, \lambda(X))$ is a uniform approximation to $\lambda_1(\cdot)$, i.e.,

$$0 \leq f_1(\varepsilon, \lambda(X)) - \lambda_1(X) \leq \varepsilon \ln n, \qquad \forall (\varepsilon, X) \in \mathbb{R}_{++} \times \mathcal{S}_n. \tag{1.9}$$

According to [8, Lemma 3.1], we obtain

$$\nabla_X f_1(\varepsilon, \lambda(X)) = U Diag[\nabla_\varsigma f_1(\varepsilon, \varsigma)] U^T = U Diag[\mu(\varepsilon, \varsigma)] U^T, \tag{1.10}$$

with

$$\mu_i(\varepsilon, \varsigma) = \frac{e^{\varsigma_i/\varepsilon}}{\sum_{j=1}^{n} e^{\varsigma_j/\varepsilon}}, \tag{1.11}$$

where we denote $\varsigma := \lambda(X)$ for simplicity.

We can look back to the example (1.1). Since we have gradient form (1.10), we can immediately apply the classical optimization method (e.g., gradient method) by using the smooth approximate function $f_1(\varepsilon, \lambda(X))$ instead of $\lambda_1(X)$ to help solve some optimization problems.

According to (1.7), we have a method to smoothly approximate the maximum eigenvalue function. In the rest of this thesis, we will search for a smooth approximate function of every eigenvalue. And more importantly, this smoothed approximate function has a good property of computability.

Chapter 2

The Smoothing Function for the κth Largest Component

2.1 The Sum of the κ largest components

For $x \in \mathbb{R}^n$ we denote by $x^{[\kappa]}$ the κth largest component of x, i.e.,

$$x^{[1]} \geq x^{[2]} \geq \cdots \geq x^{[\kappa]} \geq \cdots \geq x^{[n]}$$

sorted in nonincreasing order. Define

$$f_\kappa(x) = \sum_{i=1}^{\kappa} x^{[i]}$$

as the sum of the κ largest components of x. Since

$$f_\kappa(x) = \sum_{i=1}^{\kappa} x^{[i]} = \max\{x_{i_1} + \cdots + x_{i_\kappa} | 1 \leq i_1 < i_2 < \cdots < i_\kappa \leq n\}$$

is the maximum of all possible sums of κ different components of x. It is the pointwise maximum of $n!/(\kappa!(n-\kappa)!)$ linear functions, which means $f_\kappa(x)$ is convex and strongly semismooth (we will give out the definition of *semismooth* in Chapter 3).

To characterize the components that achieve the maximum in the following results, information about the multiplicity of the components of $x = (x_1, \ldots, x_n)^T$ is needed. Let

$$\begin{aligned} &x^{[1]} \geq \cdots \geq x^{[r]} > \\ &x^{[r+1]} = \cdots = x^{[\kappa]} = \cdots = x^{[r+t]} > \\ &x^{[r+t+1]} \geq \cdots \geq x^{[n]}, \end{aligned} \tag{2.1}$$

where $t \geq 1$ and $r \geq 0$ are integers. The multiplicity of the κth component is t. The number of components larger than $x^{[\kappa]}$ is r. Here r may be zero; in particular this must be the case if $\kappa = 1$. Note that by definition

$$r + 1 \leq \kappa \leq r + t \leq n,$$

so $t \geq \kappa - r$. Also, $t = 1$ implies that $\kappa = r + 1$.

Clearly, we can express $f_\kappa(x)$ in the following way:

$$\begin{aligned} f_\kappa(x) = \quad & \max \quad x^T v \\ & \text{s.t.} \quad \sum_{i=1}^{n} v_i = \kappa \\ & \qquad 0 \leq v_i \leq 1, \ i = 1, 2, \ldots, n \end{aligned} \tag{2.2}$$

If the components of $x \in \mathbb{R}^n$ are arranged in the order of (2.1), then directly from the property of (2.2), we have

$$\begin{aligned} &\operatorname{argmax}\{x^T v : \sum_{i=1}^{n} v_i = \kappa, 0 \leq v_i \leq 1, \ i = 1, 2, \ldots, n\} \\ = \quad &\left\{ v \in \mathbb{R}^n : \begin{array}{ll} v_i = 1 & \text{if } i = [1], \ldots, [r], \\ 0 \leq v_i \leq 1 & \text{if } i = [r+1], \ldots, [r+t], \text{ and } \sum_{i=[r+1]}^{[r+t]} v_i = \kappa - r \\ v_i = 0 & \text{if } i = [r+t+1], \ldots, [n] \end{array} \right\}. \end{aligned} \tag{2.3}$$

From (2.3) we know $f_\kappa(x)$ may not be differentiable at any $x \in \mathbb{R}^n$. However, when $\kappa = n$, $f_\kappa(x) = f_n(x)$ is the sum of all components. Clearly, $f_n(x)$ is already

a continuously differentiable function. So in the following sections and chapters, we only need to find a smoothing function of a nonsmooth function $f_\kappa(x)$ when $\kappa \in \{1, 2, \ldots, n-1\}$.

2.2 The smoothing function of the sum of the κ largest components

In this section, we will give a smoothing function $g_\kappa(\varepsilon, x)$ of a nonsmooth function $f_\kappa(x)$, where $g_\kappa(\cdot,\cdot) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$, such that

$$g_\kappa(\varepsilon, y) \to f_\kappa(x), \text{ as } (\varepsilon, y) \to (0, x). \tag{2.4}$$

Here the function $g_\kappa(\cdot,\cdot)$ is required to be continuously differentiable around (ε, x) unless $\varepsilon = 0$.

We separate into two steps to obtain $g_\kappa(\varepsilon, x)$:

1. find a smoothing function $f_\kappa(\varepsilon, x)$ on $\mathbb{R}_{++} \times \mathbb{R}^n$,
2. then $g_\kappa(\varepsilon, x)$ is constructed by

$$g_\kappa(\varepsilon, x) = \begin{cases} f_\kappa(\varepsilon, x), & \varepsilon > 0 \\ f_\kappa(x), & \varepsilon = 0 \\ f_\kappa(-\varepsilon, x), & \varepsilon < 0. \end{cases} \tag{2.5}$$

2.2.1 Smoothing function $f_\kappa(\varepsilon, x)$

Denote by $\mathcal{Q}$ the convex set in $\mathbb{R}^n$:

$$\mathcal{Q} = \{v \in \mathbb{R}^n : \sum_{i=1}^{n} v_i = \kappa, 0 \le v_i \le 1,\ i = 1, 2, \ldots, n\}, \tag{2.6}$$

and

$$p(z) = \begin{cases} z \ln z, & z \in (0, 1] \\ 0, & z = 0 \end{cases} \tag{2.7}$$

then let

$$r(v) = \sum_{i=1}^{n} p(v_i) + \sum_{i=1}^{n} p(1 - v_i) + R,\ v \in \mathcal{Q} \tag{2.8}$$

where $R = n \ln n - \kappa \ln \kappa - (n - \kappa) \ln(n - \kappa)$. So $r(v)$ is continuous and strongly convex on $\mathcal{Q}$. Denote

$$v_0 = \operatorname{argmin}\{r(v) : v \in \mathcal{Q}\}. \tag{2.9}$$

By using the KKT condition, we calculate as follows

$$v_0 = (\frac{\kappa}{n}, \frac{\kappa}{n}, \ldots, \frac{\kappa}{n})^T, \tag{2.10}$$

and

$$r(v_0) = 0. \tag{2.11}$$

It is easy to check that the maximal value of $r(v)$ is R. So we have

$$0 \leq r(v) \leq R,\ v \in \mathcal{Q}. \tag{2.12}$$

Define $f_\kappa(\cdot, \cdot) : \mathbb{R}_{++} \times \mathbb{R}^n \to \mathbb{R}$ as:

$$\begin{aligned} f_\kappa(\varepsilon, x) = \quad & \max \quad x^T v - \varepsilon r(v) \\ & \text{s.t.} \quad \sum_{i=1}^{n} v_i = \kappa \\ & \qquad 0 \leq v_i \leq 1,\ i = 1, \ldots, n. \end{aligned} \tag{2.13}$$

Lemma 2.1. *$f_\kappa(\varepsilon, x)$ in (2.13) is equivalent to $\widetilde{f}_\kappa(\cdot, \cdot) : \mathbb{R}_{++} \times \mathbb{R}^n \to \mathbb{R}$ as:*

$$\begin{aligned} \tilde{f}_\kappa(\varepsilon, x) = \quad & \max \quad x^T v - \varepsilon r(v) \\ & \text{s.t.} \quad \sum_{i=1}^{n} v_i = \kappa \\ & \qquad 0 < v_i < 1,\ i = 1, \ldots, n. \end{aligned} \tag{2.14}$$

Proof. Since $r(v)$ in (2.8) is strongly convex, the optimal solution of (2.13) is unique. On the other hand, the first order necessary and sufficient optimality conditions for (2.14) look as follows:

$$\begin{cases} -x_i + \varepsilon(\ln v_i - \ln(1 - v_i)) + \alpha = 0, \ i = 1, \ldots, n \\ \sum_{i=1}^{n} v_i = \kappa \end{cases} \tag{2.15}$$

where α is the Lagrangian multiplier for $\sum_{i=1}^{n} v_i = \kappa$ in (2.14). Clearly, we obtain

$$v_i(\varepsilon, x) = \frac{1}{1 + e^{\frac{\alpha(\varepsilon,x) - x_i}{\varepsilon}}}, \ i = 1, \ldots, n, \tag{2.16}$$

where

$$\sum_{i=1}^{n} \frac{1}{1 + e^{\frac{\alpha(\varepsilon,x) - x_i}{\varepsilon}}} = \kappa. \tag{2.17}$$

By using numerical method such as Newton's method and bisection, we can solve $\alpha(\varepsilon, x)$ through (2.17). Substituting $\alpha(\varepsilon, x)$ to (2.16), we can obtain our optimal solution $v(\varepsilon, x)$. (2.16) and (2.17) also satisfy the first order necessary and sufficient optimality conditions for (2.13). Therefore $v(\varepsilon, x)$ in (2.16) is the optimal solution of (2.13). Since the optimal solution of (2.13) is unique, $v(\varepsilon, x)$ is the only optimal solution to (2.13), which means (2.13) and (2.14) are equivalent. □

Before proving $f_\kappa(\varepsilon, x)$ is continuously differentiable on $\mathbb{R}_{++} \times \mathbb{R}^n$, we will give the following lemma:

Lemma 2.2. *$v(\varepsilon, x)$ in (2.16), which is the optimal solution to (2.13), is continuously differentiable on $\mathbb{R}_{++} \times \mathbb{R}^n$, with*

$$\nabla v(\varepsilon, x) = -\frac{\gamma_i}{\sum_{j=1}^{n} \gamma_j} \Big(\sum_{j=1}^{n} \beta_j, \gamma_1, \gamma_2, \ldots, \gamma_n \Big)^T + \Big(\beta_i, \gamma_1, \gamma_2, \ldots, \gamma_n \Big)^T, \tag{2.18}$$

where

$$\beta_i = \frac{(\alpha(\varepsilon, x) - x_i) e^{(\alpha(\varepsilon,x) - x_i)/\varepsilon}}{\varepsilon^2 (1 + e^{(\alpha(\varepsilon,x) - x_i)/\varepsilon})^2} \tag{2.19}$$

and

$$\gamma_i = \frac{e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon}}{\varepsilon(1+e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon})^2}. \tag{2.20}$$

Proof. From (2.16), we know the continuity and differentiability of $v(\varepsilon, x)$ depend on $\alpha(\varepsilon, x)$. First we show $\alpha(\varepsilon, x)$ is continuously differentiable on $\mathbb{R}_{++} \times \mathbb{R}^n$. Let

$$h((\varepsilon, x), \alpha(\varepsilon, x)) := \sum_{i=1}^{n} \frac{1}{1+e^{\frac{\alpha(\varepsilon,x)-x_i}{\varepsilon}}} - \kappa. \tag{2.21}$$

From (2.17), we have the equation

$$h((\varepsilon, x), \alpha(\varepsilon, x)) = 0. \tag{2.22}$$

Taking derivatives on both sides of (2.22),

$$\nabla_\alpha h((\varepsilon, x), \alpha(\varepsilon, x))\nabla\alpha(\varepsilon, x) + \nabla_{(\varepsilon,x)} h((\varepsilon, x), \alpha(\varepsilon, x)) = 0. \tag{2.23}$$

where

$$\nabla_\alpha h((\varepsilon, x), \alpha(\varepsilon, x)) = -\sum_{i=1}^{n} \frac{e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon}}{\varepsilon(1+e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon})^2} < 0, \tag{2.24}$$

and

$$\nabla_{(\varepsilon,x)} h((\varepsilon, x), \alpha(\varepsilon, x)) = (\mu(\varepsilon, x), \nu_1(\varepsilon, x), \ldots, \nu_n(\varepsilon, x))^T, \tag{2.25}$$

with

$$\mu(\varepsilon, x) = \sum_{i=1}^{n} \frac{(\alpha(\varepsilon, x) - x_i)e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon}}{\varepsilon^2(1+e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon})^2} \quad \text{and} \quad \nu_i(\varepsilon, x) = \frac{e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon}}{\varepsilon(1+e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon})^2}. \tag{2.26}$$

Since $\nabla_{(\varepsilon,x)} h((\varepsilon, x), \alpha(\varepsilon, x))$ is continuous and $\nabla_\alpha h((\varepsilon, x), \alpha(\varepsilon, x)) < 0$, we have $\alpha(\varepsilon, x)$ is continuously differentiable. Moreover

$$\nabla\alpha(\varepsilon, x) = -\frac{\nabla_{(\varepsilon,x)} h((\varepsilon, x), \alpha(\varepsilon, x))}{\nabla_\alpha h((\varepsilon, x), \alpha(\varepsilon, x))}. \tag{2.27}$$

Now, we will show $v(\varepsilon, x)$ is continuously differentiable. Denote the right hand side of (2.16) by $\rho_i((\varepsilon, x), \alpha(\varepsilon, x)) := \dfrac{1}{1+e^{\frac{\alpha(\varepsilon,x)-x_i}{\varepsilon}}}$, Taking derivatives on both sides of

$$v_i(\varepsilon, x) = \rho_i((\varepsilon, x), \alpha(\varepsilon, x)), \tag{2.28}$$

we have

$$\nabla v_i(\varepsilon, x) = \nabla_\alpha \rho_i((\varepsilon, x), \alpha(\varepsilon, x))\nabla\alpha(\varepsilon, x) + \nabla_{(\varepsilon,x)}\rho_i((\varepsilon, x), \alpha(\varepsilon, x)), \tag{2.29}$$

where

$$\nabla_\alpha \rho_i((\varepsilon, x), \alpha(\varepsilon, x)) = -\frac{e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon}}{\varepsilon(1 + e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon})^2}, \tag{2.30}$$

$\nabla\alpha(\varepsilon, x)$ is of (2.27) and

$$\nabla_{(\varepsilon,x)}\rho_i = (\sigma_i(\varepsilon, x), \nu_1(\varepsilon, x), \ldots, \nu_n(\varepsilon, x))^T, \tag{2.31}$$

with

$$\sigma_i(\varepsilon, x) = \frac{(\alpha(\varepsilon, x) - x_i)e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon}}{\varepsilon^2(1 + e^{(\alpha(\varepsilon,x)-x_i)/\varepsilon})^2} \tag{2.32}$$

and $\nu_i(\varepsilon, x)$ of (2.26). According to equations from (2.28) to (2.32), we have showed $v(\varepsilon, x)$ is continuously differentiable. Directly from (2.29) to (2.32), we obtain (2.18) with (2.19) and (2.20) . □

Now we are ready to give the following Theorem,

Theorem 2.3. *$f_\kappa(\varepsilon, x)$ in (2.13) is continuously differentiable on $\mathbb{R}_{++} \times \mathbb{R}^n$.*

Proof. Sine $f_\kappa(\varepsilon, x) = x^T v(\varepsilon, x) - \varepsilon r(v(\varepsilon, x))$, where $v(\varepsilon, x)$ is the optimal solution, and directly from Lemma 2.2, we can obtain $f_\kappa(\varepsilon, x)$ is continuously differentiable.

□

Lemma 2.4. *$f_\kappa(\varepsilon, x)$ is convex on $\mathbb{R}_{++} \times \mathbb{R}^n$.*

Proof. For any $\lambda \in [0, 1]$ and $(\varepsilon, x), (\tau, y) \in \mathbb{R}_{++} \times \mathbb{R}^n$, we have

$$\begin{aligned}
& f_\kappa(\lambda\varepsilon + (1-\lambda)\tau, \lambda x + (1-\lambda)y) \\
= & \max_{v\in\mathcal{Q}}\{(\lambda x + (1-\lambda)y)^T v - (\lambda\varepsilon + (1-\lambda)\tau)r(v)\} \\
= & \max_{v\in\mathcal{Q}}\{\lambda(x^T v - \varepsilon r(v)) + (1-\lambda)(y^T v - \tau r(v))\} \\
\leq & \max_{v\in\mathcal{Q}}\{\lambda(x^T v - \varepsilon r(v))\} + \max_{v\in\mathcal{Q}}\{(1-\lambda)(y^T v - \tau r(v))\} \\
= & \lambda f_\kappa(\varepsilon, x) + (1-\lambda) f_\kappa(\tau, y)
\end{aligned} \tag{2.33}$$

□

Since $R = \max\{r(v) : v \in \mathcal{Q}\}$, we have

$$f_\kappa(\varepsilon, x) \leq f_\kappa(x) \leq f_\kappa(\varepsilon, x) + \varepsilon R, \quad \varepsilon > 0. \tag{2.34}$$

Thus, we have the following conclusion:

Theorem 2.5. *The function $f_\kappa(\varepsilon, \cdot)$ for each $\varepsilon > 0$ is a smooth convex approximation of the function $f_\kappa(\cdot)$.*

Proof. It is a direct result of Theorem 2.3, Lemma 2.4 and inequalities (2.34). □

In order to show the gradient of $f_\kappa(\varepsilon, x)$, let us introduce some basic concepts.

Definition 1. Let D be a nonempty convex set in $\mathbb{R}^n$, and let $f : D \to \mathbb{R}$ be convex. Then ξ is called a **subgradient of** f at $\bar{x} \in D$ if

$$f(x) \geq f(\bar{x}) + \xi^T(x - \bar{x}) \quad \text{for all } x \in D. \tag{2.35}$$

The collection of subgradients of f at $\bar{x}$ is called the **subdifferential of** f at $\bar{x}$, denoted by $\partial f(\bar{x})$.

Lemma 2.6. *[27, Theorem 25.1, Page 242] Let D be a nonempty convex set in $\mathbb{R}^n$, and let $f : D \to \mathbb{R}$ be convex. Suppose that f is differentiable at $\bar{x} \in \text{int}D$. Then $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.*

Theorem 2.7. *The gradient of $f_\kappa(\varepsilon, x)$ on $\mathbb{R}_{++} \times \mathbb{R}^n$ is*

$$\nabla f_\kappa(\varepsilon, x) = \begin{pmatrix} -r(v(\varepsilon, x)) \\ v(\varepsilon, x) \end{pmatrix}, \tag{2.36}$$

where $v(\varepsilon, x)$ is the optimal solution of (2.13).

Proof. $\forall(\tau, y) \in \mathbb{R}_{++} \times \mathbb{R}^n$, we have

$$\begin{aligned} f_\kappa(\tau, y) &= \max_{v \in \mathcal{Q}} (\tau, y^T) \begin{pmatrix} -r(v) \\ v \end{pmatrix} \\ &\geq (\tau, y^T) \begin{pmatrix} -r(v(\varepsilon, x)) \\ v(\varepsilon, x) \end{pmatrix} \\ &= f_\kappa(\varepsilon, x) + (-r(v(\varepsilon, x)), v(\varepsilon, x)^T) \begin{pmatrix} \tau - \varepsilon \\ y - x \end{pmatrix}, \end{aligned} \tag{2.37}$$

where $v(\varepsilon, x)$ is the optimal solution of $f_\kappa(\varepsilon, x)$. Since $f_\kappa(\varepsilon, x)$ is convex (by Theorem 2.4) and continuously differentiable (by Theorem 2.3), and according to Lemma 2.6, we have $\{\nabla f_\kappa(\varepsilon, x)\} = \partial f_\kappa(\varepsilon, x)$ on $\mathbb{R}_{++} \times \mathbb{R}^n$. □

2.2.2 Smoothing function $g_\kappa(\varepsilon, x)$

Now we are ready to define $g_\kappa(\cdot, \cdot) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ as:

$$g_\kappa(\varepsilon, x) = \begin{cases} f_\kappa(\varepsilon, x), & \varepsilon > 0 \\ f_\kappa(x), & \varepsilon = 0 \\ f_\kappa(-\varepsilon, x), & \varepsilon < 0. \end{cases} \tag{2.38}$$

According to the nice properties of $f_\kappa(\varepsilon, x)$, we know $g_\kappa(\varepsilon, x)$ is a smoothing function of a nonsmooth function $f_\kappa(x)$, with

$$g_\kappa(\varepsilon, y) \to f_\kappa(x), \text{ as } (\varepsilon, y) \to (0, x). \tag{2.39}$$

Here the function $g_\kappa(\cdot, \cdot)$ is continuously differentiable around (ε, x) unless $\varepsilon = 0$.

Function $g_\kappa(\varepsilon, x)$ is convex on $\mathbb{R}_+ \times \mathbb{R}^n$ and $\mathbb{R}_- \times \mathbb{R}^n$, but may not convex on $\mathbb{R} \times \mathbb{R}^n$. The gradient of $g_\kappa(\cdot, \cdot)$ is

$$\nabla g_\kappa(\varepsilon, x) = \nabla f_\kappa(\varepsilon, x), \text{ on } \mathbb{R}_{++} \times \mathbb{R}^n \tag{2.40}$$

and

$$\nabla g_\kappa(\varepsilon, x) = \nabla f_\kappa(|\varepsilon|, x), \text{ on } \mathbb{R}_{--} \times \mathbb{R}^n. \tag{2.41}$$

In this section we find a smoothing function of the sum of the κ largest components which is computable. In the next section, we will show some numerical results and discuss the complexity.

2.3 Computational results for minmax problems

In this section, we continue the research by Nesterov [20] and [21]. It is shown that some structured non-smooth problem can be solved with efficiency estimates $O(\frac{1}{\epsilon})$, where ϵ is the desired accuracy of the solution. We extend Nesterov's primal-dual symmetric technique to the sum of the κ largest components. Here we treat ε as a parameter.

Denote

$$Q_1 = \{x \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = \kappa_1, 0 \leq x_i \leq 1\},$$

and

$$Q_2 = \{v \in \mathbb{R}^m : \sum_{j=1}^{m} v_j = \kappa_2, 0 \leq v_j \leq 1\}.$$

Let $A : \mathbb{R}^n \to \mathbb{R}^m$, $x \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$. Consider the following minmax problem:

$$\min_{x \in Q_1} \max_{v \in Q_2} \{(Ax)^T v\}. \tag{2.42}$$

This problem is reduced to :

$$\min_{x \in Q_1} f(x), \ f(x) = \max_{v \in Q_2} \{(Ax)^T v\}, \tag{2.43}$$

$$\max_{v \in Q_2} g(v), \ g(v) = \min_{x \in Q_1} \{(A^T v)^T x\}. \tag{2.44}$$

Let's choose the *Entropy Distance*:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|, \ \|v\|_1 = \sum_{j=1}^{m} |v_j|,$$

and

$$R_1 = n \ln n - \kappa_1 \ln \kappa_1 - (n - \kappa_1) \ln(n - \kappa_1),$$

$$R_2 = m \ln m - \kappa_2 \ln \kappa_2 - (m - \kappa_2) \ln(m - \kappa_2).$$

We have primal form:

$$f^{\varepsilon_2}(x) = \max_{v \in Q_2} \{(Ax)^T v - \varepsilon_2 r_2(v)\}, \quad \varepsilon_2 > 0, \tag{2.45}$$

where

$$r_2(v) = \sum_{j=1}^{m} v_j \ln v_j + \sum_{j=1}^{m} (1 - v_j) \ln(1 - v_j) + R_2 \tag{2.46}$$

is a continuous and strongly convex. According to Nesterov [20, Theorem 1], we know

$$\nabla f^{\varepsilon_2}(x) = A^T v^{\varepsilon_2}(x), \tag{2.47}$$

where $v^{\varepsilon_2}(x)$ is the optimal solution of (2.45).

Similarly, we have dual form:

$$g^{\varepsilon_1}(v) = \min_{x \in Q_1} \{(A^T v)^T x + \varepsilon_1 r_1(x)\}, \quad \varepsilon_1 > 0, \tag{2.48}$$

where

$$r_1(x) = \sum_{i=1}^{n} x_i \ln x_i + \sum_{i=1}^{n} (1 - x_i) \ln(1 - x_i) + R_1 \tag{2.49}$$

is a continuous and strongly convex. According to Nesterov [20, Theorem 1], we know

$$\nabla g^{\varepsilon_1}(v) = A x^{\varepsilon_1}(v), \tag{2.50}$$

where $x^{\varepsilon_1}(v)$ is the optimal solution of (2.48).

2.3.1 Algorithm

In order to apply Nesterov primal-dual excessive gap technique [21], we need to introduce the *Bregman distance* and the *Bregman projection*.

Bregman distances were introduced in [3] as an extension to the usual metric discrepancy measure $(x, y) \to \|x - y\|^2$ and have since found numerous applications in optimization, convex feasibility, convex inequalities, variational inequalities, monotone inclusions, equilibrium problems; see [1, 4, 6] and the references therein. If f is a real convex differentiable function, then the *Bregman distance* between two parameters z and x is defined as

$$\xi(z, x) = f(x) - f(z) - \langle \nabla f(z), x - z \rangle, \quad x, z \in Q, \tag{2.51}$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product, $\nabla f(z)$ is the gradient of f at z, and Q is a convex set. When the function f has the form $f(z) = \sum_{i=1}^{n} g_i(z_i)$, with the $g_i(t) = t^2$, for all i. Then the function $f(z) = \sum_{i=1}^{n} g_i(z_i) = \sum_{i=1}^{n} z_i^2$ is a separable Bregman function and $\xi_(z, x)$ is the squared Euclidean distance between z and x. The appendix of [5] gives out detailed definitions of *Bregman functions, distances and projections.*

The problem under consideration in this thesis is the *Bregman distance* between z and x as

$$\xi_1(z, x) = r_1(x) - r_1(z) - \nabla r_1(z)^T (x - z), \quad x, z \in Q_1, \tag{2.52}$$

where $r_1(x)$ is differentiable for any x and z from Q_1. Define the *Bregman projection* of h as follows:

$$V_1(z, h) = \operatorname{argmin}\{h^T (x - z) + \xi_1(z, x) : x \in Q_1\}. \tag{2.53}$$

Similarly, we have

$$\xi_2(w, v) = r_2(v) - r_2(w) - \nabla r_2(w)^T (v - w), \quad w, v \in Q_2, \tag{2.54}$$

and

$$V_2(w, l) = \operatorname{argmax}\{l^T (v - w) - \xi_2(w, v) : v \in Q_2\}. \tag{2.55}$$

Now we are ready to give the algorithm [21]:

1. Initialization:

Choose an arbitrary $\varepsilon_2 > 0$, and any $\varepsilon_1 \geq \frac{1}{\varepsilon_2}$. Set

$$\bar{x}_0 = V_1(x_0, \varepsilon_2 \nabla f^{\varepsilon_2}(x_0)),\ \bar{v}_0 = v^{\varepsilon_2}(x_0),\ \varepsilon_{1,0} = \varepsilon_1,\ \varepsilon_{2,0} = \varepsilon_2, \tag{2.56}$$

where $x_0 = (\frac{\kappa_1}{n}, \frac{\kappa_1}{n}, \ldots, \frac{\kappa_1}{n})^T$.

2. Iterations ($k \geq 0$):

- Set $\tau_k = \frac{2}{k+3}$.
- If k is even then generate $(\bar{x}_{k+1}, \bar{v}_{k+1})$ from $(\bar{x}_k, \bar{v}_k)$ using:

  $\hat{x}_k = (1-\tau_k)\bar{x}_k + \tau_k x^{\varepsilon_{1,k}}(\bar{v}_k)$,

  $\bar{v}_{k+1} = (1-\tau_k)\bar{v}_k + \tau_k v^{\varepsilon_{2,k}}(\hat{x}_k)$,

  $\tilde{x}_k = V_1(x^{\varepsilon_{1,k}}(\bar{v}_k), \frac{\tau_k}{(1-\tau_k)\varepsilon_{1,k}} \nabla f^{\varepsilon_{2,k}}(\hat{x}_k))$,

  $\bar{x}_{k+1} = (1-\tau_k)\bar{x}_k + \tau_k \tilde{x}_k$,

  $\varepsilon_{1,k+1} = (1-\tau_k)\varepsilon_{1,k}$.

- If k is odd then generate $(\bar{x}_{k+1}, \bar{v}_{k+1})$ from $(\bar{x}_k, \bar{v}_k)$ using:

  $\hat{v}_k = (1-\tau_k)\bar{v}_k + \tau_k v^{\varepsilon_{2,k}}(\bar{x}_k)$,

  $\bar{x}_{k+1} = (1-\tau_k)\bar{x}_k + \tau_k x^{\varepsilon_{1,k}}(\hat{v}_k)$,

  $\tilde{v}_k = V_2(v^{\varepsilon_{2,k}}(\bar{x}_k), \frac{\tau_k}{(1-\tau_k)\varepsilon_{2,k}} \nabla g^{\varepsilon_{1,k}}(\hat{v}_k))$,

  $\bar{v}_{k+1} = (1-\tau_k)\bar{v}_k + \tau_k \tilde{v}_k$,

  $\varepsilon_{2,k+1} = (1-\tau_k)\varepsilon_{2,k}$.

According to Nesterov [21, Theorem 3], we have the following statement:

Theorem 2.8. *Let the sequences $\{\bar{x}_k\}_{k=0}^{\infty}$ and $\{\bar{v}_k\}_{k=0}^{\infty}$ be generated by the above method. We have*

$$f(\bar{x}_k) - g(\bar{v}_k) \leq \frac{4\|A\|_{1,2}}{k+1}\sqrt{R_1 R_2}, \tag{2.57}$$

where $\|A\|_{1,2} = \max_{x,v}\{(Ax)^T v : \|x\|_1 = 1, \|v\|_1 = 1\}$.

2.3.2 Computational complexity

Let's discuss the complexity of above algorithm. At each iteration we need to compute the following objects.

1. Computation of $v^{\varepsilon_2}(x)$ and $x^{\varepsilon_1}(v)$.

   $v^{\varepsilon_2}(x)$ is the optimal solution of:

$$\begin{aligned} f^{\varepsilon_2}(x) = \quad & \max \quad \{(Ax)^T v - \varepsilon_2 r_2(v)\} \\ & \text{s.t.} \quad \sum_{j=1}^{m} v_i = \kappa_2 \\ & \qquad 0 \le v_j \le 1,\ j = 1, 2, \ldots, m. \end{aligned} \tag{2.58}$$

   Using the KKT condition, we need to solve the following equations:

$$\begin{cases} c_j + \varepsilon_2(\ln v_j - \ln(1 - v_j)) + \alpha = 0,\ j = 1, \ldots, m \\ \sum_{j=1}^{m} v_j = \kappa_2 \end{cases} \tag{2.59}$$

   with $c = -Ax$. Clearly,

$$v_j = \frac{1}{1 + e^{\frac{\alpha + c_j}{\varepsilon_2}}},\ j = 1, \ldots, m, \tag{2.60}$$

   where

$$\sum_{j=1}^{m} \frac{1}{1 + e^{\frac{\alpha + c_j}{\varepsilon_2}}} = \kappa_2. \tag{2.61}$$

   We can use numerical method (e.g. Newton's method, bisection method, etc.) to solve α through (2.61). Since the dimension of α is one, it is quite easy to solve. By substituting α to (2.60), we can obtain our optimal solution $v^{\varepsilon_2}(x)$ which is unique.

   It is almost the same stroke to compute $x^{\varepsilon_1}(v)$, so we skip the discussion.

2. Computation of $V_1(z, h)$ and $V_2(w, l)$.

   Let's first study $V_1(z, h)$. Applying the KKT condition to (2.53), we have

the following equations:

$$\begin{cases} h_i + \ln x_i - \ln(1-x_i) - \ln z_i + \ln(1-z_i) + \beta = 0, \ i = 1, \ldots, n \\ \sum_{i=1}^{n} x_i = \kappa_1 \end{cases} \tag{2.62}$$

Clearly,

$$x_i = \frac{z_i}{e^{h_i} e^{\beta}(1-z_i) + z_i}, \ i = 1, \ldots, n, \tag{2.63}$$

where

$$\sum_{i=1}^{n} \frac{z_i}{e^{h_i} e^{\beta}(1-z_i) + z_i} = \kappa_1. \tag{2.64}$$

We can use numerical method (e.g. Newton's method, bisection method, etc.) to solve β through (2.64). Since the dimension of β is one, it is quite easy to solve. By substituting β to (2.63), we can obtain $V_1^{(i)}(z,h) = x_i(z,h)$. The computation of $V_2(w,l)$ is the same as $V_1(z,h)$.

Thus, we have shown that all computations at each iteration of our algorithm is very cheap.

2.3.3 Computational results

We will present the computational results of minmax problem (2.42):

$$\min_{x \in Q_1} \max_{v \in Q_2} \{(Ax)^T v\}.$$

The matrix A is generated randomly. Each of its entries is uniformly distributed in the interval $[-1, 1]$. Thus $\|A\|_{1,2} \leq 1$.

We want to test the stability of our algorithm and the rate of convergence namely the order $O(\frac{1}{k})$, where k is the iteration count.

Set ϵ as the desired accuracy of the solution, i.e., $f(\bar{x}_k) - g(\bar{v}_k) \leq \epsilon$. According to (2.57), we have the predicted iteration value N: $N = \lceil (\frac{4}{\epsilon}\sqrt{R_1 R_2}) \rceil$. It is the smallest integer which is larger than or equal to $\frac{4}{\epsilon}\sqrt{R_1 R_2}$.

We implement the algorithm exactly as it is presented in this thesis and choose different values of accuracy ϵ, dimension m, n and different values of κ_1, κ_2 respectively, to get different results.

Results for $\epsilon = 0.01,\ \kappa_1 = \kappa_2 = 1$.

m \ n	100	300	1000	3000
100	328	406	552	604
300	402	540	623	660
1000	460	620	689	720

(2.65)

Number of iterations: 15-25% of predicted values.

Results for $\epsilon = 0.001,\ \kappa_1 = \kappa_2 = 1$.

m \ n	100	300	1000	3000
100	2948	4104	4702	4952
300	4100	4560	5184	5860
1000	4512	5024	5610	6520

(2.66)

Number of iterations: 15-25% of predicted values.

Results for $\epsilon = 0.01,\ \kappa_1 = \kappa_2 = 2$.

m \ n	100	300	1000	3000
100	545	564	572	614
300	622	650	686	722
1000	716	740	765	810

(2.67)

Number of iterations: 10-20% of predicted values.

Results for $\epsilon = 0.01,\ \kappa_1 = 10,\ \kappa_2 = 20$.

m \ n	50	100	150	300
50	2322	3176	4496	4920
100	3962	5162	7546	8990
150	4914	7700	8930	10840

(2.68)

Number of iterations: 20-55% of predicted values.

From these tables, we conclude that the actual iterations are better than our predicted values. When the accuracy or dimension increased, iterations are also increased, but with a decelerating speed. For future studies, we can apply this primal dual method to other *minmax* problems, such as

$$\min_{x\in Q_1}\max_{v\in Q_2}\{(Ax)^T v + c^T x + b^T v\}.$$

2.4 The κth Largest Component

From previous sections, we already know the sum of the κ largest components $f_\kappa(x)$ and the smoothing function $f_\kappa(\varepsilon, x)$ of it. So the κth largest component of $x = (x_1, x_2, \cdots, x_n)^T$ can be expressed by

$$x^{[\kappa]} = f_\kappa(x) - f_{\kappa-1}(x). \tag{2.69}$$

Therefore, we denote $\phi_\kappa(\varepsilon, x)$ by the difference of following two functions:

$$\phi_\kappa(\varepsilon, x) = f_\kappa(\varepsilon, x) - f_{\kappa-1}(\varepsilon, x). \tag{2.70}$$

Clearly, $\phi_\kappa(\varepsilon, x)$ is a smooth function, which approximates to the κth largest component of x, as ε approaches *zero*.

2.5 Summary

In this chapter, we first give the function $f_\kappa(x)$ as the sum of the κ largest components of $x \in \mathbb{R}^n$, which is a convex function. After introducing the smooth

convex function $f_\kappa(\varepsilon, x)$, we give the gradient of $f_\kappa(\varepsilon, x)$. Then we find a smoothing function $g_\kappa(\varepsilon, x)$ on $\mathbb{R} \times \mathbb{R}^n$ unless $\varepsilon = 0$. According to primal-dual excessive gap algorithm, we use this smooth function to solve some minmax problem and test the results.

Since $f_\kappa(\varepsilon, x)$ is the smoothing approximation function of the sum of the κ largest components, we can use the difference of $f_\kappa(\varepsilon, x)$ and $f_{\kappa-1}(\varepsilon, x)$ to approximate to the κth largest component, i.e.,

$$\phi_\kappa(\varepsilon, y) = \big(f_\kappa(\varepsilon, y) - f_{\kappa-1}(\varepsilon, y)\big) \longrightarrow x^{[\kappa]}, \quad \text{as } (\varepsilon, y) \to (0^+, x). \tag{2.71}$$

Thus $\phi_\kappa(\varepsilon, x)$ is the smooth approximate function of the κth largest component.

Chapter 3

Semismoothness

In this chapter we first introduce some basic concepts and preliminary results used in our analysis.

3.1 Preliminaries

In order to establish superlinear convergence of generalized Newton methods for nonsmooth equations, we need the concept of semismoothness. Semismoothness was originally introduced by Mifflin [19] for functionals. Convex functions, smooth functions, and piecewise linear functions are examples of semismooth functions. The composition of semismooth functions is still a semismooth function [19]. Semismooth functionals play an important role in the global convergence theory of nonsmooth optimization. In [26], Qi and Sun extended the definition of semismooth functions to vector-valued functions. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous function. According to Rademacher's Theorem, F is differentiable almost everywhere. Let D_F be the set of differentiable points of F and let F' be the Jacobian of F whenever it exists. Denote

$$\partial_B F(x) := \{V \in \mathbb{R}^{m\times n} |\ V = \lim_{x^k \to x} F'(x^k),\ x^k \in D_F\}.$$

Then Clarke's generalized Jacobian [10] is

$$\partial F(x) = \text{conv}\{\partial_B F(x)\},$$

where "conv" stands for the convex hull in the usual sense of convex analysis [27].

Definition 2. Suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a locally Lipschitz continuous function. F is said to be semismooth at $x \in \mathbb{R}^n$ if F is directionally differentiable at x and for any $V \in \partial F(x + \Delta x)$,

$$F(x + \Delta x) - F(x) - V(\Delta x) = o(\|\Delta x\|). \tag{3.1}$$

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

$$F(x + \Delta x) - F(x) - V(\Delta x) = O(\|\Delta x\|^{1+p}). \tag{3.2}$$

In particular, F is called strongly semismooth at x if F is 1-*order* semismooth at x.

A function F is said to be a (strongly) semismooth function if it is (strongly) semismooth everywhere on $\mathbb{R}^n$. The next result [29, Theorem 3.7] provides a convenient tool for proving strong semismoothness.

Theorem 3.1. *Suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$ 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 $V \in \partial F(x + \Delta x)$,

$$F(x + \Delta x) - F(x) - V(\Delta x) = O(\|\Delta x\|^{1+p}); \tag{3.3}$$

(b) for any $x + \Delta x \in D_F$,

$$F(x + \Delta x) - F(x) - F'(x + \Delta x)(\Delta x) = O(\|\Delta x\|^{1+p}). \tag{3.4}$$

Later we will use (b) to prove the $p-order$ $(0 < p < \infty)$ semismoothness of $g_\kappa(\varepsilon, x)$.

3.2 Semismoothness of $g_\kappa(\varepsilon, x)$

We have

$$g_\kappa(\varepsilon, x) = \begin{cases} f_\kappa(\varepsilon, x), & \varepsilon > 0 \\ f_\kappa(x), & \varepsilon = 0 \\ f_\kappa(-\varepsilon, x), & \varepsilon < 0. \end{cases} \tag{3.5}$$

where $g_\kappa(\cdot,\cdot) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$, $f_\kappa(x)$ is in the form of (2.2) and $f_\kappa(\varepsilon, x)$ is in the form of (2.13). Before discussing semismoothness of $g_\kappa(\varepsilon, x)$, we will first introduce some lemmas.

Lemma 3.2. *$g_\kappa(\varepsilon, x)$ is Lipschitz continuous on $\mathbb{R} \times \mathbb{R}^n$.*

Proof. i) When $\varepsilon > 0$ and $\tau > 0$, we have

$$\begin{aligned} |g_\kappa(\varepsilon, x) - g_\kappa(\tau, y)| &= \Big| \int_0^1 \nabla g_\kappa((\varepsilon + \theta(\varepsilon - \tau)), (x + \theta(x - y)))d\theta \Big| \\ &\le \Big| (-r(v), v) \begin{pmatrix} \varepsilon - \tau \\ x - y \end{pmatrix} \Big| \\ &\le \|(-r(v), v)\| \Big\| \begin{pmatrix} \varepsilon - \tau \\ x - y \end{pmatrix} \Big\| \\ &\le M \Big\| \begin{pmatrix} \varepsilon - \tau \\ x - y \end{pmatrix} \Big\|, \end{aligned} \tag{3.6}$$

where $M = \sqrt{R^2 + 1}$.

ii) When $\varepsilon \ge 0$, $\tau \ge 0$ and at least one of them equals zero, we take limit on both sides of (3.6), inequality (3.6) still holds.

iii) When at least one of ε, τ is negative, we have

$$\begin{aligned} |g_\kappa(\varepsilon, x) - g_\kappa(\tau, y)| &= |g_\kappa(|\varepsilon|, x) - g_\kappa(|\tau|, y)| \\ &\le M \Big\| \begin{pmatrix} |\varepsilon| - |\tau| \\ x - y \end{pmatrix} \Big\| \\ &\le M \Big\| \begin{pmatrix} |\varepsilon - \tau| \\ x - y \end{pmatrix} \Big\|. \end{aligned} \tag{3.7}$$

□

Actually, $g_\kappa(\varepsilon, x)$ is globally Lipschitz continuous on $\mathbb{R} \times \mathbb{R}^n$.

Lemma 3.3. *$g_\kappa(\varepsilon, x)$ is directionally differentiable in a neighbourhood of $(0, x)$.*

Proof. Consider $(\Delta\varepsilon, \Delta x) \in \mathbb{R} \times \mathbb{R}^n$,

i) when $\Delta\varepsilon \geq 0$ and $t > 0$, denote by

$$\zeta(t) := \frac{g_\kappa(0 + t\Delta\varepsilon, x + t\Delta x) - g(0, x)}{t}. \tag{3.8}$$

According to the convexity of $g_\kappa(\cdot, \cdot)$ on $\mathbb{R}_+ \times \mathbb{R}^n$, we have

$$\zeta(t_1) \leq \zeta(t_2) \quad \forall\ 0 < t_1 \leq t_2. \tag{3.9}$$

From Lemma 3.2, there exists a constant C, such that $|\zeta(t)| \leq C$. Therefore $\lim\limits_{t \downarrow 0} \zeta(t)$ exists.

ii) When $\Delta\varepsilon < 0$ and $t > 0$, we have

$$\lim_{t \downarrow 0} \zeta(t) = \lim_{t \downarrow 0} \frac{g_\kappa(0 + t|\Delta\varepsilon|, x + t\Delta x) - g(0, x)}{t}. \tag{3.10}$$

According to case i), we know the existence of $\lim_{t \downarrow 0} \zeta(t)$. □

For the simplicity of notation, we assume that vector $x = (x_1, \ldots, x_n)^T$ is in the non-increasing order, i.e.,

$$\begin{aligned} &x_1 \geq \cdots \geq x_r > \\ &x_{r+1} = \cdots = x_\kappa = \cdots = x_{r+t} > \\ &x_{r+t+1} \geq \cdots \geq x_n, \end{aligned} \tag{3.11}$$

where $t \geq 1$ and $r \geq 0$ are integers. The multiplicity of the κth element is t. The number of elements larger than x_κ is r. Here r may be zero; in particular this must be the case if $\kappa = 1$. Note that by definition

$$r + 1 \leq \kappa \leq r + t \leq n,$$

so $t \geq \kappa - r$. Also, $t = 1$ implies that $\kappa = r + 1$.

Lemma 3.4. *If $x = (x_1, \ldots, x_n)^T$ is in the order of (3.11), then for any $(\Delta\varepsilon, \Delta x) \to 0$ with $\Delta\varepsilon > 0$, we have*

$$\lim_{(\Delta\varepsilon,\Delta x)\to(0^+,0)} \sup \alpha(\Delta\varepsilon, x + \Delta x) \leq x_1 \tag{3.12}$$

and

$$\lim_{(\Delta\varepsilon,\Delta x)\to(0^+,0)} \inf \alpha(\Delta\varepsilon, x + \Delta x) \geq x_n, \tag{3.13}$$

where α is in the form of (2.17).

Proof. Suppose by contrary that (3.12) does not hold. Then there exists a sequence $\{(\Delta\varepsilon^k, \Delta x^k)\}$ with $(\Delta\varepsilon^k, \Delta x^k) \to (0^+, 0)$ such that

$$\lim_{k\to\infty} \alpha(\Delta\varepsilon^k, x + \Delta x^k) > x_1. \tag{3.14}$$

According to (2.16), we have

$$v_i(\Delta\varepsilon^k, x + \Delta x^k) = \frac{1}{1 + e^{\frac{\alpha(\Delta\varepsilon^k, x+\Delta x^k) - (x_i + \Delta x_i^k)}{\Delta\varepsilon^k}}}, \quad \text{for } i = 1, \ldots, n. \tag{3.15}$$

By noting that $x = (x_1, \ldots, x_n)^T$ is in the order of (3.11), the inequality (3.14) and the equation(3.15), we have

$$\lim_{k\to\infty} v_i(\Delta\varepsilon^k, x + \Delta x^k) = 0, \quad \text{for } i = 1, \ldots, n, \tag{3.16}$$

which contradicts to

$$\sum_{i=1}^{n} v_i(\Delta\varepsilon^k, x + \Delta x^k) = \kappa, \text{ where } \kappa \in \{1, 2, \ldots, n-1\}. \tag{3.17}$$

Therefore, (3.12) holds.

Suppose by contrary that (3.13) does not hold. Then there exists a sequence $\{(\Delta\varepsilon^j, \Delta x^j)\}$ with $(\Delta\varepsilon^j, \Delta x^j) \to (0^+, 0)$ such that

$$\lim_{j\to\infty} \alpha(\Delta\varepsilon^j, x + \Delta x^j) < x_n. \tag{3.18}$$

According to (2.16), we have

$$v_i(\Delta\varepsilon^j, x+\Delta x^j) = \frac{1}{1+e^{\frac{\alpha(\Delta\varepsilon^j, x+\Delta x^j)-(x_i+\Delta x_i^j)}{\Delta\varepsilon^j}}}, \quad \text{for } i=1,\ldots,n. \tag{3.19}$$

By noting that $x=(x_1,\ldots,x_n)^T$ is in the order of (3.11), the inequality (3.18) and the equation(3.19), we have

$$\lim_{j\to\infty} v_i(\Delta\varepsilon^j, x+\Delta x^j) = 1, \quad \text{for } i=1,\ldots,n, \tag{3.20}$$

which contradicts to

$$\sum_{i=1}^{n} v_i(\Delta\varepsilon^j, x+\Delta x^j) = \kappa, \text{ where } \kappa \in \{1,2,\ldots,n-1\}. \tag{3.21}$$

Therefore, (3.13) holds. □

Now we are ready to give out the most important result of this chapter:

Theorem 3.5. *$g_\kappa(\varepsilon, x)$ is p-order $(0<p<\infty)$ semismooth at $(0,x)\in\mathbb{R}\times\mathbb{R}^n$.*

Proof. First we need to prove that for any $(\Delta\varepsilon, \Delta x)\to 0$ with $\Delta\varepsilon > 0$ we have

$$g_\kappa(0+\Delta\varepsilon, x+\Delta x) - g_\kappa(0,x) - \nabla g_\kappa(0+\Delta\varepsilon, x+\Delta x)^T \begin{pmatrix} \Delta\varepsilon \\ \Delta x \end{pmatrix} = O\left(\left\| \begin{pmatrix} \Delta\varepsilon \\ \Delta x \end{pmatrix} \right\|^{1+p}\right). \tag{3.22}$$

Suppose by contrary that (3.22) is not true. Then there exists a sequence $\{(\Delta\varepsilon^j, \Delta x^j)\}$ with $(\Delta\varepsilon^j, \Delta x^j)\to 0$ and $\Delta\varepsilon^j > 0$ for each j, such that

$$\lim_{(\Delta\varepsilon^j, \Delta x^j)\to(0^+,0)} \frac{\left\| g_\kappa(0+\Delta\varepsilon^j, x+\Delta x^j) - g_\kappa(0,x) - \nabla g_\kappa(0+\Delta\varepsilon^j, x+\Delta x^j)^T \begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix} \right\|}{\|(\Delta\varepsilon^j, (\Delta x^j)^T)\|^{1+p}} = +\infty. \tag{3.23}$$

By lemma 3.4, we obtain $\{\alpha(\Delta\varepsilon^j, x+\Delta x^j)\}$ is bounded from both sides. By taking a subsequence if necessary, we can assume that there exists $\bar{\alpha}$, such that

$$\lim_{j\to\infty} \alpha(\Delta\varepsilon^j, x+\Delta x^j) = \bar{\alpha}. \tag{3.24}$$

Since $\Delta\varepsilon > 0$, we have

$$g_\kappa(0+\Delta\varepsilon^j, x+\Delta x^j) = f_\kappa(0+\Delta\varepsilon^j, x+\Delta x^j), \tag{3.25}$$

and

$$\nabla g_\kappa(0+\Delta\varepsilon^j, x+\Delta x^j) = \nabla f_\kappa(0+\Delta\varepsilon^j, x+\Delta x^j) = \begin{pmatrix} -r(v(0+\Delta\varepsilon^j, x+\Delta x^j)) \\ v(0+\Delta\varepsilon^j, x+\Delta x^j) \end{pmatrix}. \tag{3.26}$$

By definition of $g_\kappa(\cdot,\cdot)$ (2.38), we know

$$g_\kappa(0, x) = f_\kappa(x). \tag{3.27}$$

By substituting (3.25), (3.26) and (3.27) to the left hand side of (3.22), we obtain

$$\begin{aligned} &f_\kappa(0+\Delta\varepsilon^j, x+\Delta x^j) - f_\kappa(x) - \nabla f_\kappa(0+\Delta\varepsilon^j, x+\Delta x^j)^T \begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix} \\ &= x^T v(\Delta\varepsilon^j, x+\Delta x^j) - x^T v(0, x), \end{aligned} \tag{3.28}$$

where $v(0,x)$ is in the form of (2.3). By using the equation (2.16) and (2.17), we have

$$v_i(\Delta\varepsilon^j, x+\Delta x^j) = \frac{1}{1+e^{\frac{\alpha(\Delta\varepsilon^j, x+\Delta x^j)-(x_i+\Delta x_i^j)}{\Delta\varepsilon^j}}},\ i=1,\ldots,n, \tag{3.29}$$

where

$$\sum_{i=1}^n v_i(\Delta\varepsilon^j, x+\Delta x^j) = \sum_{i=1}^n \frac{1}{1+e^{\frac{\alpha(\Delta\varepsilon^j, x+\Delta x^j)-(x_i+\Delta x_i^j)}{\Delta\varepsilon^j}}} = \kappa. \tag{3.30}$$

For the simplicity of notation, we assume vector $x = (x_1, \ldots, x_n)^T$ is in the order of (3.11).

- Case 1): $t = 1$, i.e., the multiplicity of the κth element is 1:

$$x_1 \geq \cdots \geq x_{\kappa-1} > x_\kappa > x_{\kappa+1} \geq x_{\kappa+2} \geq \cdots \geq x_n. \tag{3.31}$$

We shall prove that in this case $\bar{\alpha}$ must satisfy:

$$x_\kappa \geq \bar{\alpha} \geq x_{\kappa+1}. \tag{3.32}$$

If $\bar{\alpha} > x_\kappa$, then $\bar{\alpha} > x_\kappa > x_{\kappa+1} \geq \cdots \geq x_n$. From (3.29), we have

$$\lim_{j\to\infty} v_i(\Delta\varepsilon^j, x + \Delta x^j) = 0, \text{ for } i = \kappa, \ldots, n.$$

Since $\sum_{i=1}^{n} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa$, we obtain

$$\lim_{j\to\infty} \sum_{i=1}^{\kappa-1} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa - \lim_{j\to\infty} \sum_{i=\kappa}^{n} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa, \tag{3.33}$$

which contradicts to

$$0 < v_i(\Delta\varepsilon^j, x + \Delta x^j) < 1.$$

Therefore the left hand side inequality of (3.32) holds.

If $\bar{\alpha} < x_{\kappa+1}$, then $x_1 \geq \cdots \geq x_{\kappa-1} > x_\kappa > x_{\kappa+1} > \alpha$, we have

$$\lim_{j\to\infty} v_i(\Delta\varepsilon^j, x + \Delta x^j) = 1, \text{ for } i = 1, \ldots, \kappa + 1.$$

Therefore

$$\lim_{j\to\infty} \sum_{i=1}^{\kappa+1} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa + 1. \tag{3.34}$$

But on the other hand, we know

$$0 < v_i(\Delta\varepsilon^j, x + \Delta x^j) < 1, \ \sum_{i=1}^{n} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa,$$

which is contradictory to (3.34). Therefore the right hand side inequality of (3.32) holds. So the inequality (3.32) holds.

- Case 1.1): $\bar{\alpha} = x_\kappa$. From (3.29) and (3.31), we have

$$v_i(\Delta\varepsilon^j, x + \Delta x^j) = 1 - O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right), \text{ for } i = 1, \ldots, \kappa - 1.$$

and

$$v_i(\Delta\varepsilon^j, x + \Delta x^j) = O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right), \text{ for } i = \kappa + 1, \ldots, n.$$

From (3.30), we have

$$\sum_{i=1}^{\kappa-1} v_i(\Delta\varepsilon^j, x + \Delta x^j) + v_\kappa(\Delta\varepsilon^j, x + \Delta x^j) + \sum_{i=\kappa+1}^{n} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa.$$

Hence,

$$v_\kappa(\Delta\varepsilon^j, x + \Delta x^j) = 1 - O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right).$$

and

$$\begin{aligned} & \sum_{i=1}^{n} x_i v_i(\Delta\varepsilon^j, x + \Delta x^j) - \sum_{i=1}^{n} x_i v_i(0, x) \\ = & \sum_{i=1}^{\kappa} x_i (v_i(\Delta\varepsilon^j, x + \Delta x^j) - 1) + \sum_{i=\kappa+1}^{n} x_i (v_i(\Delta\varepsilon^j, x + \Delta x^j) - 0) \\ = & O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right), \end{aligned} \tag{3.35}$$

which contradicts to (3.23).

- Case 1.2): $x_\kappa > \bar{\alpha} > x_{\kappa+1}$. From (3.29) and (3.31), we have

$$v_i(\Delta\varepsilon^j, x + \Delta x^j) = 1 - O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right), \text{ for } i = 1, \ldots, \kappa,$$

and

$$v_i(\Delta\varepsilon^j, x + \Delta x^j) = O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right), \text{ for } i = \kappa + 1, \ldots, n.$$

. Thus,

$$\begin{aligned} & \sum_{i=1}^{n} x_i v_i(\Delta\varepsilon^j, x+\Delta x^j) - \sum_{i=1}^{n} x_i v_i(0,x) \\ = & \sum_{i=1}^{\kappa} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - 1) + \sum_{i=\kappa+1}^{n} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - 0) \\ = & O\left(\left\| \begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix} \right\|^{1+p} \right), \end{aligned} \tag{3.36}$$

which contradicts to (3.23).

- Case 1.3): $\bar{\alpha} = x_{\kappa+1}$. From (3.29), (3.30) and (3.31), we have

$$v_i(\Delta\varepsilon^j, x+\Delta x^j) = 1 - O\left(\left\| \begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix} \right\|^{1+p} \right), \text{ for } i = 1, \ldots, \kappa,$$

and

$$\sum_{i=1}^{\kappa} v_i(\Delta\varepsilon^j, x+\Delta x^j) + \sum_{i=\kappa+1}^{n} v_i(\Delta\varepsilon^j, x+\Delta x^j) = \kappa.$$

Thus,

$$\sum_{i=\kappa+1}^{n} v_i(\Delta\varepsilon^j, x+\Delta x^j) = \kappa - \sum_{i=1}^{\kappa} v_i(\Delta\varepsilon^j, x+\Delta x^j) = O\left(\left\| \begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix} \right\|^{1+p} \right).$$

Since $0 < v_i(\Delta\varepsilon^j, x+\Delta x^j) < 1$, we have

$$v_i(\Delta\varepsilon^j, x+\Delta x^j) = O\left(\left\| \begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix} \right\|^{1+p} \right), \text{ for } i = \kappa+1, \ldots, n.$$

Therefore,

$$\begin{aligned} & \sum_{i=1}^{n} x_i v_i(\Delta\varepsilon^j, x+\Delta x^j) - \sum_{i=1}^{n} x_i v_i(0,x) \\ = & \sum_{i=1}^{\kappa} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - 1) + \sum_{i=\kappa+1}^{n} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - 0) \\ = & O\left(\left\| \begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix} \right\|^{1+p} \right), \end{aligned} \tag{3.37}$$

which contradicts to (3.23).

- Case 2): $t > 1$, i.e., the multiplicity of the κth element is larger than 1:

$$\begin{aligned} &x_1 \geq \cdots \geq x_r > \\ &x_{r+1} = \cdots = x_\kappa = \cdots = x_{r+t} > \\ &x_{r+t+1} \geq \cdots \geq x_n. \end{aligned} \tag{3.38}$$

We shall prove that in this case $\bar{\alpha}$ must satisfy:

$$x_\kappa \geq \bar{\alpha} \geq x_{r+t+1}. \tag{3.39}$$

If $\bar{\alpha} > x_\kappa$, then $\bar{\alpha} > x_{r+1} = \cdots = x_\kappa = \cdots = x_{r+t} > x_{r+t+1} \geq \cdots \geq x_n$. From (3.29), we have

$$\lim_{j\to\infty} v_i(\Delta\varepsilon^j, x + \Delta x^j) = 0, \text{ for } i = r+1, \ldots, n.$$

Since $\sum_{i=1}^{n} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa$, we obtain

$$\lim_{j\to\infty} \sum_{i=1}^{r} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa - \lim_{j\to\infty} \sum_{i=r+1}^{n} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa. \tag{3.40}$$

From $r \leq \kappa - 1$, we know (3.40) contradicts to

$$0 < v_i(\Delta\varepsilon^j, x + \Delta x^j) < 1.$$

Therefore the left hand side inequality of (3.39) holds.

If $x_{r+t+1} > \bar{\alpha}$, then $x_1 \geq \cdots \geq x_r > x_{r+1} = \cdots = x_{r+t} > x_{r+t+1} > \bar{\alpha}$. From (3.29), we have

$$\lim_{j\to\infty} v_i(\Delta\varepsilon^j, x + \Delta x^j) = 1, \text{ for } i = 1, \ldots, r+t+1.$$

Therefore

$$\lim_{j\to\infty} \sum_{i=1}^{r+t+1} v_i(\Delta\varepsilon^j, x + \Delta x^j) \geq \kappa + 1. \tag{3.41}$$

But on the other hand, we know

$$0 < v_i(\Delta\varepsilon^j, x + \Delta x^j) < 1,\ \sum_{i=1}^{n} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa,$$

which is contradictory to (3.41). Therefore the right hand side inequality of (3.39) holds. So the inequality (3.39) holds.

- Case 2.1): $\kappa = r + t$, i.e.,

$$x_1 \geq \cdots \geq x_r > x_{r+1} = \cdots = x_\kappa > x_{\kappa+1} \geq \cdots \geq x_n. \tag{3.42}$$

According to (3.39), we have

$$x_\kappa \geq \bar{\alpha} \geq x_{\kappa+1}. \tag{3.43}$$

- Case 2.1.1): $\bar{\alpha} = x_\kappa$. From (3.29) and (3.43), we have

$$v_i(\Delta\varepsilon^j, x + \Delta x^j) = 1 - O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right),\ \text{for } i = 1, \ldots, r$$

and

$$v_i(\Delta\varepsilon^j, x + \Delta x^j) = O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right),\ \text{for } i = \kappa+1, \ldots, n.$$

Hence, from

$$\sum_{i=1}^{r} v_i(\Delta\varepsilon^j, x + \Delta x^j) + \sum_{i=r+1}^{\kappa} v_i(\Delta\varepsilon^j, x + \Delta x^j) + \sum_{i=\kappa+1}^{n} v_i(\Delta\varepsilon^j, x + \Delta x^j) = \kappa,$$

we get

$$\sum_{i=r+1}^{\kappa} v_\kappa(\Delta\varepsilon^j, x + \Delta x^j) = (\kappa - r) - O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right).$$

Thus,

$$\begin{aligned}
&\sum_{i=1}^{n} x_i v_i(\Delta\varepsilon^j, x+\Delta x^j) - \sum_{i=1}^{n} x_i v_i(0,x) \\
= &\sum_{i=1}^{r} x_i(v_i(\Delta\varepsilon^j, x+\Delta x^j)-1) + x_\kappa(\sum_{i=r+1}^{\kappa} v_i(\Delta\varepsilon^j, x+\Delta x^j) - \sum_{r+1}^{\kappa} v(0,x)) \\
&+ \sum_{i=\kappa+1}^{n} x_i(v_i(\Delta\varepsilon^j, x+\Delta x^j)-0) \\
= &\; O\left(\left\|\begin{pmatrix}\Delta\varepsilon^j\\ \Delta x^j\end{pmatrix}\right\|^{1+p}\right),
\end{aligned} \tag{3.44}$$

which contradicts to (3.23).

- Case 2.1.2): $x_\kappa > \bar{\alpha} > x_{\kappa+1}$. From (3.29), we have

$$v_i(\Delta\varepsilon^j, x+\Delta x^j) = 1 - O\left(\left\|\begin{pmatrix}\Delta\varepsilon^j\\ \Delta x^j\end{pmatrix}\right\|^{1+p}\right), \text{ for } i=1,\ldots,\kappa$$

and

$$v_i(\Delta\varepsilon^j, x+\Delta x^j) = O\left(\left\|\begin{pmatrix}\Delta\varepsilon^j\\ \Delta x^j\end{pmatrix}\right\|^{1+p}\right), \text{ for } i=\kappa+1,\ldots,n.$$

Thus

$$\begin{aligned}
&\sum_{i=1}^{n} x_i v_i(\Delta\varepsilon^j, x+\Delta x^j) - \sum_{i=1}^{n} x_i v_i(0,x) \\
= &\sum_{i=1}^{\kappa} x_i(v_i(\Delta\varepsilon^j, x+\Delta x^j)-1) + \sum_{i=\kappa+1}^{n} x_i(v_i(\Delta\varepsilon^j, x+\Delta x^j)-0) \\
= &\; O\left(\left\|\begin{pmatrix}\Delta\varepsilon^j\\ \Delta x^j\end{pmatrix}\right\|^{1+p}\right),
\end{aligned} \tag{3.45}$$

which contradicts to (3.23).

- Case 2.1.3): $\bar{\alpha} = x_{\kappa+1}$. From (3.29) and (3.30), we have

$$v_i(\Delta\varepsilon^j, x+\Delta x^j) = 1 - O\left(\left\|\begin{pmatrix}\Delta\varepsilon^j\\ \Delta x^j\end{pmatrix}\right\|^{1+p}\right), \text{ for } i=1,\ldots,\kappa.$$

Since

$$\sum_{i=1}^{\kappa} v_i(\Delta\varepsilon^j, x+\Delta x^j) + \sum_{i=\kappa+1}^{n} v_i(\Delta\varepsilon^j, x+\Delta x^j) = \kappa,$$

we obtain

$$\sum_{i=\kappa+1}^{n} v_i(\Delta\varepsilon^j, x+\Delta x^j) = \kappa - \sum_{i=1}^{\kappa} v_i(\Delta\varepsilon^j, x+\Delta x^j) = O\left(\left\|\begin{pmatrix}\Delta\varepsilon^j\\ \Delta x^j\end{pmatrix}\right\|^{1+p}\right).$$

From $0 < v_i(\Delta\varepsilon^j, x+\Delta x^j) < 1$, we have

$$v_i(\Delta\varepsilon^j, x+\Delta x^j) = O\left(\left\|\begin{pmatrix}\Delta\varepsilon^j\\ \Delta x^j\end{pmatrix}\right\|^{1+p}\right), \text{ for } i = \kappa+1, \ldots, n.$$

Thus,

$$\begin{aligned} & \sum_{i=1}^{n} x_i v_i(\Delta\varepsilon^j, x+\Delta x^j) - \sum_{i=1}^{n} x_i v_i(0, x) \\ = & \sum_{i=1}^{\kappa} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - 1) + \sum_{i=\kappa+1}^{n} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - 0) \\ = & O\left(\left\|\begin{pmatrix}\Delta\varepsilon^j\\ \Delta x^j\end{pmatrix}\right\|^{1+p}\right), \end{aligned} \tag{3.46}$$

which contradicts to (3.23).

- Case 2.2): $\kappa < r+t$, i.e.,

$$\begin{aligned} & x_1 \geq \cdots \geq x_r > \\ & x_{r+1} = \cdots = x_\kappa = \cdots = x_{r+t} > \\ & x_{r+t+1} \geq \cdots \geq x_n. \end{aligned} \tag{3.47}$$

We shall prove that in this case

$$x_\kappa \geq \bar{\alpha} > x_{r+t+1}. \tag{3.48}$$

According to (3.43), we only need to prove that $\bar{\alpha} > x_{r+t+1}$.

If $\bar{\alpha} = x_{r+t+1}$, then

$$x_1 \geq \cdots \geq x_r > x_{r+1} = \cdots = x_\kappa = \cdots = x_{r+t} > \bar{\alpha}.$$

Hence, from (3.29) we have

$$\lim_{j\to\infty} v_i(\Delta\varepsilon^j, x+\Delta x^j) = 1, \text{ for } i = 1, \ldots, r+t.$$

Therefore,

$$\lim_{j\to\infty} \sum_{i=1}^{n} v_i(\Delta\varepsilon^j, x+\Delta x^j) \geq r+t > \kappa \tag{3.49}$$

which is contradictory to (3.30).

From (3.29), (3.47) and (3.48), we have

$$v_i(\Delta\varepsilon^j, x+\Delta x^j) = 1 - O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right), \text{ for } i = 1, \ldots, r,$$

and

$$v_i(\Delta\varepsilon^j, x+\Delta x^j) = O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right), \text{ for } i = r+t+1, \ldots, n.$$

According to (3.30),

$$\sum_{i=1}^{r} v_i(\Delta\varepsilon^j, x+\Delta x^j) + \sum_{i=r+1}^{r+t} v_i(\Delta\varepsilon^j, x+\Delta x^j) + \sum_{i=r+t+1}^{n} v_i(\Delta\varepsilon^j, x+\Delta x^j) = \kappa.$$

Hence,

$$\sum_{i=r+1}^{r+t} v_i(\Delta\varepsilon^j, x+\Delta x^j) = (\kappa - r) - O\left(\left\|\begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix}\right\|^{1+p}\right). \tag{3.50}$$

Thus, by (3.29), (3.47), (3.48) and (3.50),

$$\begin{aligned}
&\sum_{i=1}^{n} x_i v_i(\Delta\varepsilon^j, x+\Delta x^j) - \sum_{i=1}^{n} x_i v_i(0,x) \\
=\ & \sum_{i=1}^{r} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - 1) + \sum_{i=r+1}^{r+t} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - v_i(0,x)) \\
&+ \sum_{i=r+t+1}^{n} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - 0) \\
=\ & \sum_{i=1}^{r} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - 1) + x_\kappa \Big(\sum_{i=r+1}^{r+t} v_i(\Delta\varepsilon^j, x+\Delta x^j) - \sum_{i=r+1}^{r+t} v_i(0,x) \Big) \\
&+ \sum_{i=r+t+1}^{n} x_i (v_i(\Delta\varepsilon^j, x+\Delta x^j) - 0) \\
=\ & O\left(\left\| \begin{pmatrix} \Delta\varepsilon^j \\ \Delta x^j \end{pmatrix} \right\|^{1+p} \right),
\end{aligned} \tag{3.51}$$

which contradicts to (3.23).

We have proved all situations for $\Delta\varepsilon > 0$ that (3.22) holds. Then we will show that in the following two cases, (3.22) still holds.

Next, by (3.22), for any $(\Delta\varepsilon, \Delta x) \to 0$ with $\Delta\varepsilon < 0$ and the definition of $g_\kappa(\cdot,\cdot)$, we have

$$g_\kappa(0+\Delta\varepsilon, x+\Delta x) - g_\kappa(0,x) - \nabla g_\kappa(0+\Delta\varepsilon, x+\Delta x)^T \begin{pmatrix} \Delta\varepsilon \\ \Delta x \end{pmatrix} \tag{3.52}$$

$$= g_\kappa(0+|\Delta\varepsilon|, x+\Delta x) - g_\kappa(0,x) - \nabla g_\kappa(0+|\Delta\varepsilon|, x+\Delta x)^T \begin{pmatrix} -\Delta\varepsilon \\ \Delta x \end{pmatrix} \tag{3.53}$$

$$\begin{aligned}
&= g_\kappa(0+|\Delta\varepsilon|, x+\Delta x) - g_\kappa(0,x) - \nabla g_\kappa(0+|\Delta\varepsilon|, x+\Delta x)^T \begin{pmatrix} |\Delta\varepsilon| \\ \Delta x \end{pmatrix} \\
&= O\Big(\Big\| \begin{pmatrix} \Delta\varepsilon \\ \Delta x \end{pmatrix} \Big\|^{p+1} \Big).
\end{aligned} \tag{3.54}$$

Thus, the equation (3.22) holds for any $(\Delta\varepsilon, \Delta x) \to (0,0)$ with $\Delta\varepsilon < 0$.

Finally, we consider the case that $(\Delta\varepsilon, \Delta x) \to (0, 0)$ with $\Delta\varepsilon = 0$.

Suppose that at the point $(0, x + \Delta x)$, $g_\kappa(\cdot, \cdot)$ is differentiable (in the sense of Fréchet). Denote by $y := x + \Delta x$. Since $g_\kappa(\cdot, \cdot)$ is differentiable at $(0, y)$, $\forall\ (\Delta\tau, \Delta y) \in \mathbb{R} \times \mathbb{R}^n$, we have

$$g_\kappa(\Delta\tau, y + \Delta y) - g_\kappa(0, y) - \begin{pmatrix} \nabla_\tau g_\kappa(0, y) \\ \nabla_y g_\kappa(0, y) \end{pmatrix}^T \begin{pmatrix} \Delta\tau \\ \Delta y \end{pmatrix} = o\left(\left\| \begin{pmatrix} \Delta\tau \\ \Delta y \end{pmatrix} \right\|\right). \tag{3.55}$$

In particular, we set $\Delta\tau = 0$, then the left hand side of (3.55) is

$$\begin{aligned} & g_\kappa(0, y + \Delta y) - g_\kappa(0, y) - \begin{pmatrix} \nabla_\tau g_\kappa(0, y) \\ \nabla_y g_\kappa(0, y) \end{pmatrix}^T \begin{pmatrix} 0 \\ \Delta y \end{pmatrix} \\ = \ & g_\kappa(0, y + \Delta y) - g_\kappa(0, y) - \nabla_y g_\kappa(0, y)^T \Delta y \\ = \ & f_\kappa(y + \Delta y) - f_\kappa(y) - \nabla_y g_\kappa(0, y)^T \Delta y. \end{aligned} \tag{3.56}$$

Thus, we have

$$f_\kappa(y + \Delta y) - f_\kappa(y) - \nabla_y g_\kappa(0, y)^T \Delta y = o(\|\Delta y\|), \tag{3.57}$$

which means $f_\kappa(y)$ is differentiable (in the sense of Fréchet) at y, with $\nabla f_\kappa(y) = \nabla_y g_\kappa(0, y)$, i.e.,

$$\nabla f_\kappa(x + \Delta x) = \nabla_x g_\kappa(0, x + \Delta x). \tag{3.58}$$

Thus, for $\Delta\varepsilon = 0$, we have

$$\begin{aligned} & g_\kappa(\Delta\varepsilon, x + \Delta x) - g_\kappa(0, x) - \begin{pmatrix} \nabla_\varepsilon g_\kappa(\Delta\varepsilon, x + \Delta x) \\ \nabla_x g_\kappa(\Delta\varepsilon, x + \Delta x) \end{pmatrix}^T \begin{pmatrix} \Delta\varepsilon \\ \Delta x \end{pmatrix} \\ = \ & g_\kappa(0, x + \Delta x) - g_\kappa(0, x) - \nabla_x g_\kappa(0, x + \Delta x)^T \Delta x \\ = \ & f_\kappa(x + \Delta x) - f_\kappa(x) - \nabla f_\kappa(x + \Delta x)^T \Delta x. \end{aligned} \tag{3.59}$$

Since $f_\kappa(x)$ is a piecewise linear function, it is p-order semismooth, i.e.,

$$f_\kappa(x + \Delta x) - f_\kappa(x) - \nabla f_\kappa(x + \Delta x)^T \Delta x = O(\|\Delta x\|^{1+p}) = O\left(\left\| \begin{pmatrix} 0 \\ \Delta x \end{pmatrix} \right\|^{1+p}\right). \tag{3.60}$$

We obtain

$$g_\kappa(0, x+\Delta x) - g_\kappa(0, x) - \nabla g_\kappa(0, x+\Delta x)^T \begin{pmatrix} 0 \\ \Delta x \end{pmatrix} = O\left(\left\| \begin{pmatrix} 0 \\ \Delta x \end{pmatrix} \right\|^{1+p} \right). \quad (3.61)$$

Overall, we have proved that (3.22) holds at $(\Delta\varepsilon, \Delta x) \to 0$. Hence by Lemma 3.2, 3.3, equation (3.22) and Theorem 3.1, we obtain $g_\kappa(\varepsilon, x)$ is p-order semismooth at $(0, x) \in \mathbb{R} \times \mathbb{R}^n$. □

Chapter 4

Smoothing Approximation to Eigenvalues

4.1 Spectral functions

4.1.1 Introduction

A function F on the space of $n-$by$-n$ real symmetric matrices is called *spectral* if it depends only on the eigenvalues of its argument. Spectral functions are just symmetric functions of the eigenvalues. In this thesis we are interested in functions F of a symmetric matrix argument that are invariant under orthogonal similarity transformations:

$$F(U^T AU) = F(A), \quad \text{for all } U \in \mathcal{O} \text{ and } A \in \mathcal{S},$$

where $\mathcal{O}$ denotes the set of orthogonal matrices and $\mathcal{S}$ denotes the set of symmetric matrices. Every such function can be decomposed as $F(A) = (f \circ \lambda)(A)$, where λ is the map that gives the eigenvalues of the matrix A and f is a symmetric function. We call such functions F *spectral functions* (or just functions of eigenvalues) because they depend only on the spectrum of the operator A. Therefore, we can regard *a spectral function* as a composition of a symmetric function $f : \mathbb{R}^n \to \mathbb{R}$ and the eigenvalue function $\lambda(\cdot) : \mathcal{S} \to \mathbb{R}^n$; that is, the spectral function $(f \circ \lambda) : \mathcal{S} \to \mathbb{R}$

is given by

$$(f \circ \lambda)(X) := f(\lambda(X)) \quad X \in \mathcal{S}.$$

4.1.2 Preliminary results

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

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

Clearly $\mathcal{O}_X$ is nonempty for each $X \in \mathcal{S}_n$.

Now we refer to the formula for the gradient of a differential spectral function [16].

Proposition 4.1. *Let f be a symmetric function from $\mathbb{R}^n$ to $\mathbb{R}$ and $X \in \mathcal{S}_n$. Then the following holds:*

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

$$\nabla(f \circ \lambda)(X) = U\text{Diag}[\nabla f(\lambda(X))]U^T, \ \forall U \in \mathcal{O}_X. \tag{4.1}$$

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

Lewis and Sendov [17] 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 numerical methods which need use second-order derivatives. Suppose that f is twice differentiable at $\mu \in \mathbb{R}^n$. Define the matrix $\mathcal{C}(\mu) \in \mathbb{R}^{n \times n}$:

$$(\mathcal{C}(\mu))_{ij} := \begin{cases} 0 & \text{if } i = j \\ (\nabla^2 f(\mu))_{ii} - (\nabla^2 f(\mu))_{ij} & \text{if } i \neq j \text{ and } \mu_i = \mu_j \\ \dfrac{(\nabla f(\mu))_i - (\nabla f(\mu))_j}{\mu_i - \mu_j} & \text{else.} \end{cases} \tag{4.2}$$

It is easy to see that $C(\mu)$ is symmetric due to the symmetry of f. The following result is proved by Lewis and Sendov [17, Theorem 3.3, 4.2].

Proposition 4.2. *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*

$$\nabla^2(f \circ \lambda)(X)[H] = U(\mathrm{Diag}[\nabla^2 f(\lambda(X))\mathrm{diag}[\widetilde{H}]] + \mathcal{C}(\lambda(X)) \circ \widetilde{H})U^T, \quad \forall H \in \mathcal{S}_n, \tag{4.3}$$

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

Remark. $U \in \mathcal{O}_X$ in formulae (4.1) and (4.3) can be any choice, such that $U^T X U = \mathrm{Diag}[\lambda(X)]$, and doesn't depend on the particular choice.

4.2 Smoothing approximation

In chapter 2, we give the form

$$g_\kappa(\varepsilon, x) = \begin{cases} f_\kappa(\varepsilon, x), & \varepsilon > 0 \\ f_\kappa(x), & \varepsilon = 0 \\ f_\kappa(-\varepsilon, x), & \varepsilon < 0. \end{cases} \tag{4.4}$$

to smoothing approximate to the sum of the κ largest components of $x \in \mathbb{R}^n$, i.e.,

$$\begin{aligned} \lim_{\varepsilon \to 0, y \to x} g_\kappa(\varepsilon, y) &= f_\kappa(x) \\ &= x^{[1]} + \cdots + x^{[\kappa]}. \end{aligned}$$

We define function $g_\kappa(\varepsilon, \lambda(\cdot))$ as a composite function of $g_\kappa(\varepsilon, \cdot) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ and the eigenvalue function $\lambda(\cdot) : \mathcal{S}_n \to \mathbb{R}^n$, i.e.,

$$g_\kappa(\varepsilon, \lambda(X)), \quad \text{for any } X \in \mathcal{S}_n. \tag{4.5}$$

Since we have (2.34), i.e.,

$$0 \leq f_\kappa(x) - g_\kappa(\varepsilon, x) \leq \varepsilon R,$$

we can easily get the well defined function $g_\kappa(\varepsilon, \lambda(X))$ is an approximation to the sum of the κ largest eigenvalues

$$0 \leq (\lambda^{[1]}(X) + \lambda^{[2]}(X) + \cdots + \lambda^{[\kappa]}(X)) - g_\kappa(\varepsilon, \lambda(X)) \leq \varepsilon R \tag{4.6}$$

where $\lambda(X) \in \mathbb{R}^n$. We denote by $\lambda^{[\kappa]}(X)$ the κth largest eigenvalue of $X \in \mathcal{S}_n$, i.e.,

$$\lambda^{[1]}(X) \geq \lambda^{[2]}(X) \geq \cdots \geq \lambda^{[\kappa]}(X) \geq \cdots \geq \lambda^{[n]}(X)$$

are the eigenvalues of X sorted in nonincreasing order.

Let

$$\chi_\kappa(\varepsilon, X) := g_\kappa(\varepsilon, \lambda(X)), \tag{4.7}$$

we have the following results.

Theorem 4.3. *Let $\varepsilon > 0$ be given. The function $\chi_\kappa(\varepsilon, \cdot) : \mathcal{S}_n \to \mathbb{R}$ is continuously differentiable, and the gradient of $\chi_\kappa(\varepsilon, \cdot)$ at $X \in \mathcal{S}_n$ is given by*

$$\nabla_X \chi_\kappa(\varepsilon, X) = Q\mathrm{Diag}[\nabla_\varsigma \chi_\kappa(\varepsilon, \varsigma)]Q^T = Q\mathrm{Diag}[v(\varepsilon, \varsigma)]Q^T, \tag{4.8}$$

with $\varsigma := \lambda(X)$, $Q \in \mathcal{O}_X$, and $v(\varepsilon, \varsigma)$ is the optimal solution to $f_\kappa(\varepsilon, \varsigma)$, where

$$v_i(\varepsilon, \varsigma) = \frac{1}{1 + e^{\frac{\alpha(\varepsilon,\varsigma) - \varsigma_i}{\varepsilon}}}, \text{ for } i = 1, \ldots, n, \tag{4.9}$$

and

$$\sum_{i=1}^{n} \frac{1}{1 + e^{\frac{\alpha(\varepsilon,\varsigma) - \varsigma_i}{\varepsilon}}} = \kappa. \tag{4.10}$$

Proof. It follows from Theorem 2.3 that $g_\kappa(\varepsilon, \cdot)$ is continuous differentiable on $\mathbb{R}_{++} \times \mathbb{R}^n$. Then we use Proposition 4.1, equation (4.1) to get the first equality of (4.8). According to (2.36), we know

$$\nabla_x f_\kappa(\varepsilon, x) = v(\varepsilon, x),$$

so we get the second equality of (4.8). (4.9) and (4.10) are direct results. □

Theorem 4.4. *The function $\chi_\kappa(\cdot,\cdot)$ is continuously differentiable around (ε, X) with $\varepsilon \neq 0$ and strongly semismooth at $(0, X)$.*

Proof. From Theorem 4.3, we know $\chi_\kappa(\varepsilon,\cdot)$ is continuously differentiable around X when $\varepsilon > 0$ is fixed. According to the symmetric property of $\chi_\kappa(\varepsilon,\cdot)$, we can easily get that $\chi_\kappa(\varepsilon,\cdot)$ is continuously differentiable around X when $\varepsilon < 0$ is fixed. By Theorem 2.3, we know that $\chi_\kappa(\cdot, X)$ is continuously differentiable around any $\varepsilon \neq 0$ for any fixed X. So $\chi_\kappa(\varepsilon, X)$ is continuously differentiable around (ε, X) with $\varepsilon \neq 0$. From Theorem 3.5, we know $g_\kappa(\cdot,\cdot)$ is p-order semismooth at $(0, x)$. The recent result of Sun and Sun [30] shows that the eigenvalue function $\lambda(\cdot)$ is strongly semismooth. Since $\chi_\kappa(\varepsilon, X)$ is the composite of $g_\kappa(\varepsilon,\cdot)$ and eigenvalue function $\lambda(X)$, and the composite of p-order semismooth functions is p-order semismooth [12], we obtain that $\chi_\kappa(\varepsilon, X)$ is strongly semismooth at $(0, X)$. □

Theorem 4.4 is one of the most important results in this thesis. It shows $g_\kappa(\varepsilon, \lambda(X))$ is not only a smooth approximate function to the sum of the κ largest eigenvalue functions but also strongly semismooth at $(0, X)$. Let

$$\phi_\kappa(\varepsilon, X) := g_\kappa(\varepsilon, \lambda(X)) - g_{\kappa-1}(\varepsilon, \lambda(X)) \tag{4.11}$$

which is a smooth approximate function to the κth largest eigenvalue function. Here (4.11) is also continuously differentiable around (ε, X) with $\varepsilon \neq 0$ and strongly semismooth at $(0, X)$.

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

$$\mathcal{A}y := \sum_{i=1}^{m} y_i A_i, \ \forall y \in \mathbb{R}^m, \tag{4.12}$$

and

$$A(y) := A_0 + \mathcal{A}y. \tag{4.13}$$

Definition 3. Define $\theta_\kappa(\varepsilon,\cdot):\mathbb{R}_{++}\times\mathbb{R}^m\to\mathbb{R}$ by

$$\theta_\kappa(\varepsilon,y):=f_\kappa(\varepsilon,\lambda(A(y))),\ \forall y\in\mathbb{R}^m. \tag{4.14}$$

According to Theorem 2.7, the following result holds:

Let $\varepsilon>0$ be given, the function

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

is continuously differentiable, and the gradient of $\theta_\kappa(\varepsilon,\cdot)$ at $y\in\mathbb{R}^m$ is given by

$$\begin{aligned}\nabla_y\theta_\kappa(\varepsilon,y) &= \mathcal{A}^*(U(\mathrm{Diag}[\nabla_\varsigma f_\kappa(\varepsilon,\varsigma)])U^T)\\ &= \mathcal{A}^*(U(\mathrm{Diag}[v(\varepsilon,\varsigma)])U^T),\end{aligned} \tag{4.15}$$

where $U\in\mathcal{O}_{A(y)}$, $\mathcal{A}^*D=(\langle A_1,D\rangle,\ldots,\langle A_m,D\rangle)^T$, $\varsigma:=\lambda(A(y))$ and

$$v_i(\varepsilon,\varsigma)=\frac{1}{1+e^{\frac{\alpha(\varepsilon,\varsigma)-\varsigma_i}{\varepsilon}}},\ \text{for } i=1,\ldots,n \tag{4.16}$$

$$\sum_{i=1}^{n}\frac{1}{1+e^{\frac{\alpha(\varepsilon,\varsigma)-\varsigma_i}{\varepsilon}}}=\kappa. \tag{4.17}$$

We will apply above results in the next chapter to solve the inverse eigenvalue problems.

Chapter 5

Application in Inverse Eigenvalue Problems

5.1 Introduction

5.1.1 Objective

An inverse eigenvalue problem concerns the reconstruction of a matrix from prescribed spectral data. The spectral data involved may consist of the complete or only partial information of eigenvalues or eigenvectors. The objective of an inverse eigenvalue problem is to construct a matrix that maintains a certain specific structure as well as that given spectral property.

Associated with any inverse eigenvalue problem are two fundamental questions– the theoretic issue on *solvability* and the practical issue on *computability*. A major effort in solvability has been to determine a necessary or a sufficient condition under which an inverse eigenvalue problem has a solution. The main concern in computability, on the other hand, has been to develop a procedure by which, knowing a priori that the given spectral data are feasible, a matrix can be constructed numerically. Both questions are difficult and challenging.

5.1.2 Application

Inverse eigenvalue problems arise in a remarkable variety of applications. The list includes but is not limited to control design, system identification, seismic tomography, principal component analysis, exploration and remote sensing, antenna array processing, geophysics, molecular spectroscopy, particle physics, structure analysis, circuit theory, mechanical system simulation, and so on.

5.1.3 Diversity

Depending on the application, inverse eigenvalue problems may be described in several different forms. Translated into mathematics, it is often necessary in order that the inverse eigenvalue problem be meaningful, to restrict the construction to special classes of matrices, especially to those with specified structures. A problem without any restriction on the matrix is generally of little interest. The solution to an inverse eigenvalue problem therefore should satisfy two constraints–the *spectral constraint* referring to the prescribed spectral data and the *structural constraint* referring to the desirable structure.

5.1.4 Overview

A collection of inverse eigenvalue problems are discussed by Moody T. Chu [9], who discusses explicitly 37 inverse eigenvalue problems, current developments in both the theoretic and the algorithmic aspects.

5.2 Parameterized Inverse Eigenvalue Problem

5.2.1 Generic form

A generic Parameterized Inverse Eigenvalue Problem (PIEP) can be described as follows:

Given a family of matrices $A(c) \in \mathcal{M}$ *with* $c = [c_1, \ldots, c_m] \in \mathrm{F}^m$ *and scalars* $\{\lambda_1, \ldots, \lambda_n\} \subset \mathrm{F}$*, find a parameter* c *such that* $\lambda(A(c)) = \{\lambda_1, \ldots, \lambda_n\}$*, where* F *represents the scalar field of either real* $\mathbb{R}$ *or complex* $\mathbb{C}$*, and* $\mathcal{M}$ *denotes certain subsets of square matrices.*

Note that the number m of parameters in c may be different from n. Depending upon how the family of matrices $A(c)$ is specifically defined in terms of c, the PIEP can appear and be solved very differently. Inverse eigenvalue problems in the above PIEP format arise frequently in discrete modeling [13, 15, 22] and factor analysis [15]. A common feature of PIEP is that the parameter c is used as a "control" that modulates to the underlying problem in a certain specific, predestined way.

5.2.2 Special case

The inclusion of PIEP is quite broad. We only discuss a special case. Let $\mathcal{S}$ be the linear space of symmetric matrices of size n. Let $A : \mathbb{R}^n \to \mathcal{S}$ be continuously differentiable. Given n real numbers $\{\lambda_i^*\}_{i=1}^n$, which are arranged in the nonincreasing order

$$\lambda_1^* \geq \lambda_2^* \geq \cdots \geq \lambda_n^*.$$

The Inverse Eigenvalue Problems (IEPs) is to find a vector $c^* \in \mathbb{R}^n$ such that $\lambda_i(A(c^*)) = \lambda_i^*$, for $i = 1, \cdots, n$. A typical choice for $A(c)$ is

$$A(c) = A_0 + \sum_{j=1}^{n} c_j A_j, \tag{5.1}$$

where $A_0, A_1, \cdots, A_n \in \mathcal{S}$. In this case, $A(c)$ is an affine function of c. Define $F : \mathbb{R}^n \to \mathbb{R}^n$ by

$$F(c) = \begin{bmatrix} \lambda_1(A(c)) - \lambda_1^* \\ \vdots \\ \lambda_n(A(c)) - \lambda_n^* \end{bmatrix} \tag{5.2}$$

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

$$F(c) = 0.$$

It is well known that the eigenvalues of a symmetric matrix are strongly semi-smooth everywhere. Let $f_\kappa(\lambda(A(c))) = \sum_{i=1}^{\kappa} \lambda_i(A(c))$ be the sum of the κ largest eigenvalues of symmetric matrix $A(c)$, i.e.,

$$\begin{aligned} f_1(\lambda(A(c))) &= \lambda_1^*; \\ f_2(\lambda(A(c))) &= \lambda_1^* + \lambda_2^*; \\ &\vdots \\ f_n(\lambda(A(c))) &= \lambda_1^* + \lambda_2^* + \cdots + \lambda_n^* \end{aligned} \tag{5.3}$$

We have a smooth function $g_\kappa(\varepsilon, \lambda(A(c)))$ which approaches $f_\kappa(\lambda(A(c)))$ when $\varepsilon \to 0$, i.e.,

$$g_\kappa(\varepsilon, \lambda(A(c))) \longrightarrow \lambda_1^* + \cdots \lambda_\kappa^*; \text{ when } \varepsilon \to 0 \tag{5.4}$$

Let

$$G(\varepsilon, c) = \begin{bmatrix} g_1(\varepsilon, \lambda(A(c))) - \lambda_1^* \\ g_2(\varepsilon, \lambda(A(c))) - (\lambda_1^* + \lambda_2^*) \\ \vdots \\ g_{n-1}(\varepsilon, \lambda(A(c))) - \sum_{i=1}^{n-1} \lambda_i^* \\ f_n(\lambda(A(c))) - \sum_{i=1}^{n} \lambda_i^* \end{bmatrix} = 0, \tag{5.5}$$

and define the auxiliary equation

$$E(\varepsilon, \lambda(A(c))) := \begin{bmatrix} \varepsilon \\ G(\varepsilon, c) \end{bmatrix}. \tag{5.6}$$

Remark. In the last equation of (5.5), we use f_n instead of g_n, because $f_n(\lambda(A(c)))$ is the sum of all eigenvalues of symmetric matrix $A(c)$. It is already a smooth function.

We can use some numerical method (eg. *squared smoothing Newton method* [31]) to solve equation (5.6). The following table shows the numerical results of IEPs by using *squared smoothing Newton method* with the matrix A_j, for $j = 0, \ldots, n$, generated randomly. Each entry of A_j is uniformly distributed in the interval $[-1, 1]$.

n	$\|\lambda - \lambda^*\|$	iteration
4	0.0048	4
8	0.0051	47
10	0.0033	187
16	0.0050	856
30	0.0052	4396

(5.7)

Here n is dimension, λ is our computation result and λ^* is given.

Bibliography

[1] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, "Bregman monotone optimization algorithms", *SIAM Journal on Control and Optimization*, vol. 42, pp. 596-636, 2003.

[2] A. Ben-Tal and M. Teboulle, "A smoothing technique for nondifferentiable optimization problems", in Optimization. Fifth French German Conference, Lecture Notes in Math. 1405, Springer-Verlag, New York, 1989, 1–11.

[3] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming", U.S.S.R. *Computational Mathematics and Mathematical Physics*, vol. 7, pp. 200-217, 1967.

[4] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, Kluwer, Boston, MA, 2000.

[5] C. Byrne and Y. Censor, "Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization", *Annals of Operations Research*, 105, pp. 77-98, 2001.

[6] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, New York, 1997.

[7] P.-L. Chang, "A minimax approach to nonlinear programming", Doctoral Dissertation, Univeristy of Washington, Department of Mathematics, 1980.

[8] X. Chen, H.D. Qi, L. Qi and K. Teo, "Smooth convex approximation to the maximum eigenvalue function" to appear in *J. of Global Optimization.*

[9] Moody T. Chu, "Inverse eigenvalue problems", *SIAM Rev.*, Vol. 40, No. 1, pp1-39, March 1998.

[10] F. Clarke, "Optimization and nonsmooth analysis", John Wiley and Sons, New York, 1983.

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

[12] A. Fischer, "Solution of monotone complementarity problems with locally Lips- chitzian functions", *Math. Programming, 76* (1997), pp. 513-532.

[13] G.M.L. Gladwell, "The inverse problem for the vibrating beam", Proc. Roy. Soc. Ser. A, 393 (1984), 277–295.

[14] A.A. Goldstein, "Chebyshev approximation and linear inequalities via exponentials", Department of Mathematics, Univesity of Washington, Seattle, 1997.

[15] O.H. Hald, "On discrete and numerical inverse Sturm-Liouville problems", Ph.D. thesis, New York University, New York, 1972.

[16] A.S. Lewis, "Derivatives of spectral funtions", *Mathematics of Operations Research, 21* (1996), 576–588.

[17] A.S. Lewis and H.S. Sendov, "Twice differentiable spectral functions", *SIAM J. Matrix Anal. Appl., 22* (2001), 368–386.

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

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

[20] Yu. Nesterov, "Smooth minimization of non-smooth functions", CORE DP 2003/12, February 2003. Accepted by *Mathematcal Programming.*

[21] Yu. Nesterov, "Excessive gap technique in non-smooth convex minimizarion", CORE DP 2003/35, May 2003. Accepted by *SIAM J. Optimization.*

[22] M.R. Osborne, "On the inverse eigenvalue problem for matrices and related problems for diffenence and differential equations", Lecure Notes in Mathematics 228, Springer-Verlag, New York, 1971, pp. 155–168.

[23] J. Peng and Z. Lin, "A non-interior continuation method for generalized linear complementarity problems", *Math. Programming, 86* (1999), 533–563.

[24] H.D. Qi and L. Liao, "A smoothing Newton method for extended vertical linear complementarity problems", *SIAM J. Matrix Anal. Appl., 21* (1999), 45–66.

[25] H.D. Qi and X. Yang, "Semismooth of spectral functions", *SIAM J. Matrix Anal. Appl., 25* (2003), 766–783.

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

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

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

[29] D. Sun and J. Sun, "Semismooth matrix valued functions", *Mathematics of Operations Research, 27* (2002), 150–169.

[30] D. Sun and J. Sun, "Strong semismoothness of eigenvalues of symmetric matrices and its application to inverse eigenvalue problems", *SIAM J. Numer. Anal., 40* (2003), 2352–2367.

[31] J. Sun, D. Sun and L. Qi, "A smoothing Newton method for nonsmooth matrix equations and its applications in semidefinite optimization problems", *SIAM J. Optimization, 14* (2004), 783–806.

[32] H. Tangand and L. Zhang, "A maximum entropy method for convex programming", *Chinese Sci. Bull., 39* (1994), 682–684.

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

Name: Shi Shengyuan

Degree: Master of Science

Department: Mathematics

Thesis Title: Smooth Convex Approximation and Its Applications

Abstract

In this thesis, we consider a smooth convex approximation to the sum of the κ largest components. To make it applicable to a wide class of applications, the study is conducted on some minmax problems. Based on a special smoothing technique, we give an efficient scheme for nonsmooth convex function. By using the composite property of $g_\kappa(\varepsilon, \cdot)$ and eigenvalue function $\lambda(X)$, we find the smooth approximate function to the sum of the κ largest eigenvalue function.

Keywords:

Convex optimization, non-smooth optimization, semismooth function, sum of eigenvalues

SMOOTH CONVEX APPROXIMATION AND ITS APPLICATIONS

SHI SHENGYUAN

NATIONAL UNIVERSITY OF SINGAPORE

2004