AN INEXACT SQP NEWTON METHOD FOR CONVEX SC^1 MINIMIZATION PROBLEMS

CHEN YIDI

(Bsc., ECNU)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

2008

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Sun Defeng, for his insightful instructions and patience throughout my master candidature. I could not have completed the thesis without him. Furthermore, I would like to thank Ms. Gao Yan and Dr. Liu Yongjin at the National University of Singapore for discussions about the implementation of the inexact smoothing Newton method and the convergence analysis of the inexact SQP Newton method developed in this thesis. Last but not least, I would like to express my gratitude to my family and friends who have given me support when I was in difficulty.

Chen Yidi/July 2008

Contents

Summary

In this thesis, we introduce an inexact SQP Newton method for solving general convex SC^1 minimization problems

$$\begin{aligned} \min \quad & \theta(x) \\ \text{s.t.} \quad & x \in X, \end{aligned}$$

where X is a closed convex set in a finite dimensional Hilbert space Y and $\theta(\cdot)$ is a convex SC^1 function defined on an open convex set $\Omega \subseteq Y$ containing X.

The general convex SC^1 minimization problems model many problems as special cases. One particular example is the dual problem of the least squares covariance matrix (LSCM) problems with inequality constraints.

The purpose of this thesis is to introduce an efficient inexact SQP Newton method for solving the general convex SC^1 minimization problems under realistic assumptions. In Chapter 2, we introduce our method and conduct a complete convergence analysis including the superlinear (quadratic) rate of convergence. Numerical results conducted in Chapter 3 show that our inexact SQP Newton method is competitive when it is applied to the LSCM problems with many lower and upper bounds constraints. We make our final conclusions in Chapter 4.

Chapter 1

Introduction

In this thesis, we consider the following convex minimization problem:

$$\begin{aligned} \min \quad & \theta(x) \\ \text{s.t.} \quad & x \in X, \end{aligned} \tag{1.1}$$

where the objective function θ and the feasible set X satisfy the following assumptions:

(A1) X is a closed convex set in a finite dimensional Hilbert space Y;

(A2) $\theta(\cdot)$ is a convex LC^1 function defined on an open convex set $\Omega \subseteq Y$ containing X.

The LC^1 property of θ means that θ is Fréchet differentiable at all points in Ω and its gradient function $\nabla\theta : \Omega \to Y$ is locally Lipschitz in Ω. Furthermore, an LC^1 function θ defined on the open set $\Omega \subseteq Y$ is said to be SC^1 at a point $x \in \Omega$ if $\nabla\theta$ is semismooth at x (the definition of semismoothness will be given in Chapter 2).

There are many examples that can be modeled as SC^1 minimization problems [10]. One particular example is the following least squares covariance matrix

(LSCM) problem:

$$\begin{aligned} \min \quad & \frac{1}{2}\|X - C\|^2 \\ \text{s.t.} \quad & \langle A_i, X\rangle = b_i, \quad i = 1, \ldots, p, \\ & \langle A_i, X\rangle \geq b_i, \quad i = p+1, \ldots, m, \\ & X \in \mathcal{S}^n_+, \end{aligned} \tag{1.2}$$

where $\mathcal{S}^n$ and $\mathcal{S}^n_+$ are, respectively, the space of $n \times n$ symmetric matrices and the cone of positive semidefinite matrices in $\mathcal{S}^n$, $\|\cdot\|$ is the Frobenius norm induced by the standard trace inner product $\langle \cdot, \cdot \rangle$ in $\mathcal{S}^n$, C and A_i, $i = 1, \ldots, m$ are given matrices in $\mathcal{S}^n$, and $b \in \Re^m$.

Let $q = m - p$ and $Q = \{0\}^p \times \Re^q_+$. Denote $\mathcal{A} : \mathcal{S}^n \to \Re^m$ by

$$\mathcal{A}(X) := \begin{bmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{bmatrix}, \quad X \in \mathcal{S}^n.$$

For any symmetric $X \in \mathcal{S}^n$, we write $X \succeq 0$ and $X \succ 0$ to represent that X is positive semidefinite and positive definite, respectively. Then the feasible set of problem (1.2) can be written as follows:

$$\mathcal{F} = \{X \in \mathcal{S}^n \mid \mathcal{A}(X) \in b + Q, \ X \succeq 0\}.$$

The Lagrangian function $l : \mathcal{S}^n_+ \times Q^+ \to \Re$ for problem (1.2) is defined by

$$l(X, y) := \frac{1}{2}\|X - C\|^2 + \langle y, b - \mathcal{A}(X)\rangle,$$

where $(X, y) \in \mathcal{S}^n_+ \times Q^+$ and $Q^+ = \Re^p \times \Re^q_+$ is the dual cone of Q. Define $\theta(y) := -\inf_{X \in \mathcal{S}^n_+} l(X, y)$. Then the dual problem of (1.2) takes the following form (cf. [2, 16]):

$$\begin{aligned} \min \quad & \theta(y) := \frac{1}{2}\|\Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^* y)\|^2 - \langle b, y\rangle - \frac{1}{2}\|C\|^2 \\ \text{s.t.} \quad & y \in Q^+, \end{aligned} \tag{1.3}$$

where $\Pi_{\mathcal{S}^n_+}(\cdot)$ is the metric projector onto $\mathcal{S}^n_+$ and the adjoint $\mathcal{A}^* : \Re^m \to \mathcal{S}^n$ takes the form

$$\mathcal{A}^*(y) = \sum_{i=1}^{m} y_i A_i, \quad y \in \Re^m . \tag{1.4}$$

It is not difficult to see that the objective function $\theta(\cdot)$ in the dual problem (1.3) is a continuously differentiable convex function with

$$\nabla\theta(y) = \mathcal{A}\Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^* y) - b, \quad y \in \Re^m .$$

For any given $y \in \Re^m$, both $\theta(y)$ and $\nabla\theta(y)$ can be computed explicitly as the metric projector $\Pi_{\mathcal{S}^n_+}(\cdot)$ admits an analytic formula [17]. Furthermore, since the metric projection operator $\Pi_{\mathcal{S}^n_+}(\cdot)$ over the cone $\mathcal{S}^n_+$ has been proved to be strongly semismooth in [18], the dual problem (1.3) belongs to the class of the SC^1 minimization problems. Thus, applying any dual based methods to solve the least squares covariance matrix problem (1.2) means that eventually we have to solve a convex SC^1 minimization problem. In this thesis we focus on solving such general convex SC^1 problems.

The general convex SC^1 minimization problem (1.1) can be solved by many kinds of methods, such as the projected gradient method and BFGS method. In [10], Pang and Qi proposed a globally and superlinearly convergent SQP Newton method for convex SC^1 minimization problems under a BD-regularity assumption at the solution point, which is equivalent to the local strong convexity assumption on the objective function. This BD-regularity assumption is too restrictive. For example, the BD-regularity assumption fails to hold for the dual problem (1.3). For the details, see [7].

The purpose of this thesis is twofold. First we modify the SQP Newton method of Pang and Qi with a much less restrictive assumption than the BD-regularity. Secondly we introduce an inexact technique to improve the performance of the SQP Newton method. As the SQP Newton method in Pang and Qi [10], at each step, we need to solve a strictly convex program. We will apply the inexact smoothing Newton method recently proposed by Gao and Sun in [7] to solve it.

The following part of this thesis is organized as follows. In Chapter 2, we introduce a general inexact SQP Newton method for solving convex SC^1 minimization problems and provide a complete convergence analysis. In Chapter 3, we apply the inexact SQP Newton method to the dual problem (1.3) of the LSCM problem (1.2) and report our numerical results. We make our final conclusions in Chapter 4.

Chapter 2

An inexact SQP Newton method

In this chapter, we introduce an inexact SQP Newton method for solving the general convex SC^1 minimization problems (1.1).

Since $\theta(\cdot)$ is a convex function, $\bar{x} \in X$ solves problem (1.1) if and only if it satisfies the following variational inequality

$$\langle x - \bar{x}, \nabla\theta(\bar{x})\rangle \geq 0 \quad \forall\, x \in X. \tag{2.1}$$

Define $F : Y \to Y$ by

$$F(x) := x - \Pi_X(x - \nabla\theta(x)), \quad x \in Y, \tag{2.2}$$

where for any $x \in Y$, $\Pi_X(x)$ is the metric projection of x onto X, i.e., $\Pi_X(x)$ is the unique optimal solution to the following problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|y - x\|^2 \\ \text{s.t.} & y \in X. \end{array}$$

Then one can easily check that $\bar{x} \in X$ solves (1.1) if and only if $F(\bar{x}) = 0$ (cf. [4]).

2.1 Preliminaries

In order to design our inexact SQP Newton algorithm and analyze its convergence, we next recall some essential results related to semismooth functions.

Let $\mathcal{Z}$ be an arbitrary finite dimensional real vector space. Let $\mathcal{O}$ be an open set in $\mathcal{Y}$ and $\Xi : \mathcal{O} \subseteq \mathcal{Y} \to \mathcal{Z}$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. Then, by Rademacher's theorem [16, Chapter 9.J] we know that Ξ is almost everywhere Fréchet differentiable in $\mathcal{O}$. Let $\mathcal{O}_\Xi$ denote the set of points in $\mathcal{O}$ where Ξ is Fréchet differentiable. Let $\Xi'(y)$ denote the Jacobian of Ξ at $y \in \mathcal{O}_\Xi$. Then Clarke's generalized Jacobian of Ξ at $y \in \mathcal{O}$ is defined by [3]

$$\partial\Xi(y) := \mathrm{conv}\{\partial_B\Xi(y)\},$$

where "conv" denotes the convex hull and the B-subdifferential $\partial_B\Xi(y)$ is defined by Qi in [11]

$$\partial_B\Xi(y) := \left\{ V \,:\, V = \lim_{j\to\infty} \Xi'(y^j)\,,\ y^j \to y\,,\ y^j \in \mathcal{O}_\Xi \right\}.$$

The concept of semismoothness was first introduced by Mifflin [9] for functionals and was extended to vector-valued functions by Qi and Sun [12].

Definition 2.1.1. Let $\Xi : \mathcal{O} \subseteq \mathcal{Y} \to \mathcal{Z}$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. We say that Ξ is *semismooth* at a point $y \in \mathcal{O}$ if

(i) Ξ is directionally differentiable at y; and

(ii) for any $x \to y$ and $V \in \partial\Xi(x)$,

$$\Xi(x) - \Xi(y) - V(x-y) = o(||x-y||)\,. \tag{2.3}$$

The function $\Xi : \mathcal{O} \subseteq \mathcal{Y} \to \mathcal{Z}$ is said to be *strongly semismooth* at a point $y \in \mathcal{O}$ if Ξ is semismooth at y and for any $x \to y$ and $V \in \partial\Xi(x)$,

$$\Xi(x) - \Xi(y) - V(x-y) = O(\|x-y\|^2)\,. \tag{2.4}$$

Throughout this thesis, we assume that the metric projection operator $\Pi_X(\cdot)$ is strongly semismooth. The assumption is reasonable because it is satisfied when X is a symmetric cone including the cone of nonnegative orthant, the second-order cone, and the cone of symmetric and semidefinite matrices (cf. [19]).

We summarize some useful properties in the next proposition.

Proposition 2.1.1. Let F be defined by (2.2). Let $y \in Y$. Suppose that $\nabla\theta$ is semismooth at y. Then,

(i) F is semismooth at y;

(ii) for any $h \in Y$,

$$\partial_B F(y)h \subseteq h - \partial_B \Pi_X(y - \nabla\theta(y))(h - \partial_B \nabla\theta(y)(h)).$$

Moreover, if $I - S(I - V)$ is nonsingular for any $S \in \partial_B \Pi_X(y - \nabla\theta(y))$ and $V \in \partial_B \nabla\theta(y)$, then

(iii) all W in $\partial_B F(y)$ are nonsingular;

(iv) there exist $\overline{\sigma} > \underline{\sigma} > 0$ such that

$$\underline{\sigma}\|x - y\| \leq \|F(x) - F(y)\| \leq \overline{\sigma}\|x - y\| \tag{2.5}$$

holds for all x sufficiently close to y.

Proof. (i) Since the composite of semismooth functions is also semismooth (cf. [6]), F is semismooth at y.

(ii) The proof can be done by following that in [7, Proposition 2.3].

(iii) The conclusion follows easily from (ii) and the assumption.

(iv) Since all $W \in \partial_B F(y)$ are nonsingular, from [11] we know that $\|(W_x)^{-1}\| = O(1)$ for any $W_x \in \partial_B F(x)$ and any x sufficiently close to y. Then, the semismoothness of F at y easily implies that (2.5) holds (cf. [11]). We complete the proof. □

2.2 Algorithm

Algorithm 2.2.1. (An inexact SQP Newton method)

Step 0. Initialization. Select constants $\mu \in (0, 1/2)$ and $\gamma, \rho, \eta, \tau_1, \tau_2 \in (0, 1)$. Let $x^0 \in X$ and $f^{pre} := \|F(x^0)\|$. Let $\text{Ind}_1 = \text{Ind}_2 = \{0\}$. Set $k := 0$.

Step 1. Direction Generation. Select $V_k \in \partial_B \nabla\theta(x^k)$ and compute

$$\epsilon_k := \tau_2 \min\{\tau_1, \|F(x^k)\|\}. \tag{2.6}$$

Solve the following strictly convex program:

$$\begin{array}{ll} \min & \langle\, \nabla\theta(x^k),\, \Delta x\rangle + \dfrac{1}{2}\langle\, \Delta x,\, (V_k + \epsilon_k I)\Delta x\rangle \\ \text{s.t.} & x^k + \Delta x \in X \end{array} \tag{2.7}$$

approximately such that $x^k + \Delta x^k \in X$,

$$\langle \nabla\theta(x^k), \Delta x^k\rangle + \frac{1}{2}\langle \Delta x^k, (V_k + \epsilon_k I)\Delta x^k\rangle \leq 0 \tag{2.8}$$

and

$$\|R_k\| \leq \eta_k \|F(x^k)\|\,, \tag{2.9}$$

where R_k is defined by

$$R_k := x^k + \Delta x^k - \Pi_X\big(x^k + \Delta x^k - \big(\nabla\theta(x^k) + (V_k + \epsilon_k I)\Delta x^k\big)\big) \tag{2.10}$$

and

$$\eta_k := \min\{\eta, \|F(x^k)\|\}.$$

Step 2. Check Unit Steplength. If Δx^k satisfies the following condition:

$$\|F(x^k + \Delta x^k)\| \leq \gamma f^{pre}, \tag{2.11}$$

then set $x^{k+1} := x^k + \Delta x^k$, $\text{Ind}_1 = \text{Ind}_1 \cup \{k+1\}$, $f^{pre} = \|F(x^{k+1})\|$ and go to Step 4; otherwise, go to Step 3.

Step 3. Armijo Line Search. Let l_k be the smallest nonnegative integer l such that

$$\theta(x^k + \rho^l \Delta x^k) \le \theta(x^k) + \mu\rho^l \langle \nabla\theta(x^k), \Delta x^k \rangle \,. \tag{2.12}$$

Set $x^{k+1} := x^k + \rho^{l_k}\Delta x^k$. If $\|F(x^{k+1})\| \le \gamma f^{pre}$, then set $\mathrm{Ind}_2 = \mathrm{Ind}_2 \cup \{k+1\}$ and $f^{pre} = \|F(x^{k+1})\|$.

Step 4. Check Convergence. If x^{k+1} satisfies a prescribed stopping criteria, terminate; otherwise, replace k by $k+1$ and return to Step 1.

Before proving the convergence of Algorithm 2.2.1, we make some remarks to illustrate the algorithm.

(a). A stopping criterion has been omitted, and it is assumed without loss of generality that $\Delta x^k \ne 0$ and $F(x^k) \ne 0$ (otherwise, x^k is an optimal solution to problem (1.1)).

(b). In Step 1, we approximately solve the strictly convex problem (2.7) in order to obtain the search direction such that (2.8) and (2.9) hold. It is easy to see that the conditions (2.8) and (2.9) can be ensured because x^k is not optimal to (2.7) and $R_k = 0$ with Δx^k being chosen as the exact solution to (2.7).

(c). By using (2.8) and (2.9), we know that the search direction Δx^k generated by Algorithm 2.2.1 is always a descent direction. Since

$$\lim_{l\to\infty} [\theta(x^k + \rho^l \Delta x^k) - \theta(x^k)]/\rho^l = \nabla\theta(x^k)^T \Delta x^k < \mu \nabla\theta(x^k)^T \Delta x^k,$$

a simple argument shows that the integer l_k in Step 2 is finite and hence Algorithm 2.2.1 is well defined.

(d). The convexity of X implies that $\{x^k\} \subset X$.

2.3 Convergence Analysis

2.3.1 Global Convergence

In this subsection, we shall analyze the global convergence of Algorithm 2.2.1. We first denote the solution set by $\overline{X}$, i.e., $\overline{X} = \{x \in Y \mid x \text{ solves problem } (1.1)\}$.

In order to discuss the global convergence of Algorithm 2.2.1, we need the following assumption.

Assumption 2.3.1. The solution set $\overline{X}$ is nonempty and bounded.

The following result will be needed in the analysis of global convergence of Algorithm 2.2.1.

Lemma 2.3.1. Suppose that Assumption 2.3.1 is satisfied. Then there exists a positive number $c > 0$ such that $L_c = \{x \in Y \mid \|F(x)\| \leq c\}$ is bounded.

Proof. Since $\nabla\theta$ is monotone, the conclusion follows directly from the weakly univalent function theorem of [13, Theorem 2.5]. □

We are now ready to state our global convergence results of Algorithm 2.2.1.

Theorem 2.3.1. Suppose that X and θ satisfy Assumptions (A1) and (A2). Let Assumption 2.3.1 be satisfied. Then, Algorithm 2.2.1 generates an infinite bounded sequence $\{x^k\}$ such that

$$\lim_{k\to\infty} \theta(x^k) = \bar{\theta}, \tag{2.13}$$

where $\bar{\theta} := \theta(\bar{x})$ for any $\bar{x} \in \overline{X}$.

Proof. Let $\text{Ind} := \text{Ind}_1 \cup \text{Ind}_2$. We prove the theorem by considering the following two cases.

<u>**Case 1**</u>. $|\text{Ind}| = +\infty$.

Since the sequence $\{\|F(x^k)\| : k \in \text{Ind}\}$ is strictly decreasing and bounded from below, we know that

$$\lim_{k(\in \text{Ind})\to\infty} \|F(x^k)\| = 0. \tag{2.14}$$

By using Lemma 2.3.1, we easily obtain that the sequence $\{x^k : k \in \text{Ind}\}$ is bounded. Since any infinite subsequence of $\{\theta(x^k) : k \in \text{Ind}\}$ converges to $\bar{\theta}$ (cf. (2.14)), we conclude that $\lim_{k(\in \text{Ind})\to\infty} \theta(x^k) = \bar{\theta}$.

Next, we show that $\lim_{k\to\infty} \theta(x^k) = \bar{\theta}$. For this purpose, let $\{x^{k_j}\}$ be an arbitrary infinite subsequence of $\{x^k\}$. Then, there exist two sequence $\{k_{j,1}\} \subset \text{Ind}$ and $\{k_{j,2}\} \subset \text{Ind}$ such that $k_{j,1} \le k_j \le k_{j,2}$ and

$$\theta(x^{k_{j,2}}) \le \theta(x^{k_j}) \le \theta(x^{k_{j,1}}),$$

which implies that $\theta(x^{k_j}) \to \bar{\theta}$ as $k_j \to \infty$. Combining with Assumption 2.3.1, we know that the sequence $\{x^{k_j}\}$ must be bounded. The arbitrariness of $\{x^{k_j}\}$ implies that $\{x^k\}$ is bounded and $\lim_{k\to\infty} \theta(x^k) = \bar{\theta}$.

Case 2. $|\text{Ind}| < +\infty$.

After a finite number step, the sequence $\{x^k\}$ is generated by Step 3. Hence, we assume without loss of generality that $\text{Ind} = \{0\}$. It follows from [14, Corollary 8.7.1] that Assumption 2.3.1 implies that the set $\{x \in X : \theta(x) \le \theta(x_0)\}$ is bounded and hence $\{x^k\}$ is bounded. Therefore, there exists a subsequence $\{x^k : k \in K\}$ such that $x^k \to \bar{x}$ as $k(\in K) \to \infty$. Suppose for the purpose of a contradiction that $\bar{x}$ is not an optimal solution to problem (1.1). Then, by the definition of F (cf. (2.2)), we know that it holds $\|F(\bar{x})\| \ne 0$ and hence $\bar{\epsilon} := \tau_2 \min\{\tau_1, \|F(\bar{x})\|/2\} > 0$. Hence, it follows from (2.8) that we have that for all large k,

$$-\langle \nabla\theta(x^k), \Delta x^k\rangle \ge \frac{\bar{\epsilon}}{2}\|\Delta x^k\|^2, \tag{2.15}$$

which implies that the sequence $\{\Delta x^k\}$ is bounded.

Since $\{\theta(x^k)\}$ is a decreasing sequence and bounded from below, we know that the sequence $\{\theta(x^k)\}$ is convergent and hence $\{\theta(x^{k+1}) - \theta(x^k)\} \to 0$. The stepsize rule (2.12) implies that

$$\lim_{k\to\infty} \alpha_k \langle \nabla\theta(x^k), \Delta x^k \rangle = 0, \tag{2.16}$$

where $\alpha_k := \rho^{l_k}$.

There are two cases: (i) $\liminf_{k(\in K)\to\infty} \alpha_k > 0$ and (ii) $\liminf_{k(\in K)\to\infty} \alpha_k = 0$. In the first case, by (2.16), we can easily know that

$$\lim_{k(\in K)\to\infty} \langle \nabla\theta(x^k), \Delta x^k \rangle = 0.$$

In the latter case, without loss of generality, we assume that $\lim_{k(\in K)\to\infty} \alpha_k = 0$. Then, by the definition of α_k (cf. (2.12)), it follows that for each k,

$$\theta(x^k + \alpha'_k \Delta x^k) - \theta(x^k) > \mu\alpha'_k \langle\, \nabla\theta(x^k), \Delta x^k \rangle\,, \tag{2.17}$$

where $\alpha'_k := \alpha_k/\rho$. Note that we also have

$$\lim_{k(\in K)\to\infty} \alpha'_k = 0.$$

Dividing both sides in the expression (2.17) by α'_k, passing $k \in K$ to ∞, we can easily derive that

$$\lim_{k(\in K)\to\infty} \langle \nabla\theta(x^k), \Delta x^k \rangle \geq \mu \lim_{k(\in K)\to\infty} \langle \nabla\theta(x^k), \Delta x^k \rangle,$$

which, together with $\mu \in (0, 1/2)$ and (2.8), yields

$$\lim_{k(\in K)\to\infty} \langle \nabla\theta(x^k), \Delta x^k \rangle = 0.$$

Consequently, in both cases (i) and (ii), we have that

$$\lim_{k(\in K)\to\infty} \langle \nabla\theta(x^k), \Delta x^k \rangle = 0.$$

Hence, by (2.15), we obtain that

$$\lim_{k(\in K)\to\infty} \Delta x^k = 0.$$

Then, we deduce by passing to the limit $k(\in K) \to \infty$ in (2.9) that

$$\|F(\bar{x})\| \leq \bar{\eta}\|F(\bar{x})\|, \tag{2.18}$$

where $\bar{\eta} := \min\{\eta, \|F(\bar{x})\|\}$. Note that $\bar{\eta} < 1$, by (2.18), we easily obtain that $\|F(\bar{x})\| = 0$, which is a contradiction. Hence, we can conclude that $F(\bar{x}) = 0$ and hence $\bar{x} \in \overline{X}$.

By using the fact that $\lim_{k\to\infty} \theta(x^k) = \theta(\bar{x})$, together with Assumption 2.3.1, we know that $\{x^k\}$ is bounded and (2.13) holds. The proof is completed. □

2.3.2 Superlinear Convergence

The purpose of this subsection is to discuss the (quadratic) superlinear convergence of Algorithm 2.2.1 by assuming the (strong) semismoothness property of $\nabla\theta(\cdot)$ at a limit point $\bar{x}$ of the sequence $\{x^k\}$ and the nonsingularity of $I - S(I - V)$ with $S \in \partial_B \Pi_X(\bar{x} - \nabla\theta(\bar{x}))$ and $V \in \partial_B \nabla\theta(\bar{x})$.

Theorem 2.3.2. Suppose that $\bar{x}$ is an accumulation point of the infinite sequence $\{x^k\}$ generated by Algorithm 2.2.1 and $\nabla\theta$ is semismooth at $\bar{x}$. Suppose that for any $S \in \partial_B \Pi_X(\bar{x} - \nabla\theta(\bar{x}))$ and $V \in \partial_B \nabla\theta(\bar{x})$, $I - S(I - V)$ is nonsingular. Then the whole sequence $\{x^k\}$ converges to $\bar{x}$ superlinearly, i.e.,

$$\|x^{k+1} - \bar{x}\| = o(\|x^k - \bar{x}\|). \tag{2.19}$$

Moreover, if $\nabla\theta$ is strongly semismooth at $\bar{x}$, then the rate of convergence is quadratic, i.e.,

$$\|x^{k+1} - \bar{x}\| = O(\|x^k - \bar{x}\|^2). \tag{2.20}$$

We only prove the semismooth case. One may apply the similar arguments to prove the case when $\nabla\theta$ is strongly semismooth at $\bar{x}$. We omit the details. In order to prove Theorem 2.3.2, we first establish several lemmas.

Lemma 2.3.2. Assume that the conditions of Theorem 2.3.2 are satisfied. Then, for any given $V \in \partial_B \nabla\theta(\bar{x})$, the origin is the unique optimal solution to the following problem:

$$\begin{aligned} \min \quad & \langle \nabla\theta(\bar{x}),\, \Delta x\rangle + \frac{1}{2}\langle \Delta x,\, V\Delta x\rangle \\ \text{s.t.} \quad & \bar{x} + \Delta x \in X. \end{aligned} \tag{2.21}$$

Proof. By [4], we easily obtain that Δx solves (2.21) if and only if

$$G(\Delta x) = 0, \tag{2.22}$$

where

$$G(\Delta x) := \bar{x} + \Delta x - \Pi_X(\bar{x} + \Delta x - (\nabla\theta(\bar{x}) + V\Delta x)).$$

Since $\bar{x}$ is an optimal solution to problem (1.1), we know that $\bar{x} - \Pi_X(\bar{x} - \nabla\theta(\bar{x})) = 0$, which, together with (2.22), implies that the origin is an optimal solution to problem (2.21).

Next, we show the uniqueness of solution of problem (2.21). Suppose that $\Delta\bar{x} \neq 0$ is also an optimal solution to problem (2.21). Then, since problem (2.21) is convex, for any $t \in [0,1]$, we know that $t\Delta\bar{x} \neq 0$ is an optimal solution to problem (2.21). However, by Proposition 2.1.1, we know that the nonsingularity of $I - S(I - V)$ with $S \in \partial_B \Pi_X(\bar{x} - \nabla\theta(\bar{x}))$ and $V \in \partial_B \nabla\theta(\bar{x})$ implies that $G(\Delta x) = 0$ has only one unique solution in a neighborhood of the origin. Hence, we have obtained a contradiction. The contradiction shows that the origin is the unique optimal solution to problem (2.21). □

Lemma 2.3.3. Assume that the conditions of Theorem 2.3.2 are satisfied. Then, the sequence $\{\Delta x^k\}$ generated by Algorithm 2.2.1 converges to 0.

Proof. Suppose on the contrary that there exists a subsequence of $\{\Delta x^k\}$ which does not converge to 0. Without loss of generality, we may assume that $\{\Delta x^k\}$

does not converge to 0. Let $t_k := 1/\max\{1, \|\Delta x^k\|\}(\in (0,1])$ and $\Delta\hat{x}^k := t_k \Delta x^k$. Denote

$$\phi_k(\Delta x) := \langle \nabla\theta(x^k), \Delta x\rangle + \frac{1}{2}\langle \Delta x, (V_k + \epsilon_k I)\Delta x\rangle.$$

Then, by the convexity of ϕ_k, we obtain that

$$\begin{aligned} \phi_k(\Delta\hat{x}^k) &= \phi_k((1-t_k)\cdot 0 + t_k \Delta x^k) \\ &\leq (1-t_k)\phi_k(0) + t_k\phi_k(\Delta x^k) \\ &= 0 + t_k\phi_k(\Delta x^k) < 0, \end{aligned} \tag{2.23}$$

where the strict inequality follows from (2.8). Since the sequence $\{\Delta\hat{x}^k\}$ satisfies $\|\Delta\hat{x}^k\| \leq 1$, by passing to a subsequence, if necessary, we may assume that there exists a constant $\hat{\delta} \in (0,1]$ such that the sequence $\{\Delta\hat{x}^k\} \to \Delta\hat{x}$ with $\|\Delta\hat{x}\| = \hat{\delta}$. Hence, since the matrices in $\partial_B\nabla\theta(x^k)$ are uniformly bounded, from (2.23), by passing to the limit $k \to \infty$ and taking a subsequence if necessary, we can easily deduce that

$$\langle \nabla\theta(\bar{x}), \Delta\hat{x}\rangle + \frac{1}{2}\langle \Delta\hat{x}, V\Delta\hat{x}\rangle \leq 0 \tag{2.24}$$

for some $V \in \partial_B\nabla\theta(\bar{x})$ since $\partial_B\nabla\theta(\cdot)$ is upper semicontinuous.

On the other hand, since $x^k + \Delta x^k \in X$ and $x^k \in X$, we know that $x^k + \Delta\hat{x}^k = t_k(x^k + \Delta x^k) + (1-t_k)x^k \in X$, which, due to the fact that X is closed, implies that $\bar{x} + \Delta\hat{x} \in X$. This, together with (2.24), means that $\Delta\hat{x}$ is an optimal solution to problem (2.21), which is a contradiction to Lemma 2.3.2 since $\|\Delta\hat{x}\| = \hat{\delta}$. Hence, the sequence $\{\Delta x^k\}$ generated by Algorithm 2.2.1 converges to 0. The proof is completed. □

Lemma 2.3.4. Assume that the conditions of Theorem 2.3.2 are satisfied. Then $\bar{x}$ is the unique optimal solution to problem (1.1) and $\{x^k\}$ converges to $\bar{x}$ such that

$$\|x^k + \Delta x^k - \bar{x}\| = o(\|x^k - \bar{x}\|). \tag{2.25}$$

Proof. By Theorem 2.3.1 and Proposition 2.1.1 we know that $\bar{x}$ is the unique optimal solution to problem (1.1) and $\{x^k\}$ converges to $\bar{x}$.

It follows from Lemma 2.3.3 that $\Delta x^k \to 0$ as $k \to \infty$. Let us denote $\hat{x}^k := x^k + \Delta x^k - (\nabla\theta(x^k) + (V_k + \epsilon_k I)\Delta x^k)$. Then, we first obtain that

$$
\begin{aligned}
\hat{x}^k - \bar{x} + \nabla\theta(\bar{x}) &= x^k + \Delta x^k - (\nabla\theta(x^k) + (V_k + \epsilon_k I)\Delta x^k) - \bar{x} + \nabla\theta(\bar{x}) \\
&= x^k + \Delta x^k - \bar{x} - (\nabla\theta(x^k) - \nabla\theta(\bar{x}) - V_k(x^k - \bar{x})) - \\
&\quad -(V_k + \epsilon_k I)(x^k + \Delta x^k - \bar{x}) + \epsilon_k(x^k - \bar{x}) \\
&= (I - V_k)(x^k + \Delta x^k - \bar{x}) + o(\|x^k + \Delta x^k - \bar{x}\|) + o(\|x^k - \bar{x}\|),
\end{aligned}
$$

where the third equality follows from the semismoothness of $\nabla\theta$ at the point $\bar{x}$ and $\epsilon_k \to 0$ as $k \to \infty$.

By noting the definition of R_k (cf. (2.10)), we further obtain that

$$
\begin{aligned}
x^k + \Delta x^k - \bar{x} &= R_k + \Pi_X(\hat{x}^k) - \bar{x} \\
&= R_k + \Pi_X(\hat{x}^k) - \Pi_X(\bar{x} - \nabla\theta(\bar{x})) \\
&= R_k + S_k(\hat{x}^k - \bar{x} + \nabla\theta(\bar{x})) + O(\|\hat{x}^k - \bar{x} + \nabla\theta(\bar{x})\|^2) \\
&= R_k + S_k(I - V_k)(x^k + \Delta x^k - \bar{x}) + o(\|x^k + \Delta x^k - \bar{x}\|) + \\
&\quad + o(\|x^k - \bar{x}\|), \qquad (2.26)
\end{aligned}
$$

where $S_k \in \partial_B \Pi_X(\hat{x}^k)$ and the third equality comes from the strong semismoothness of $\Pi_X(\cdot)$.

Since $I - S(I - V)$ is nonsingular for any $S \in \partial_B \Pi_X(\bar{x} - \nabla\theta(\bar{x}))$ and $V \in \partial_B \nabla\theta(\bar{x})$, $I - S_k(I - V_k)$ is also nonsingular for all k sufficiently large. This, together with (2.26), implies that all x^k sufficiently close to $\bar{x}$,

$$\|x^k + \Delta x^k - \bar{x}\| \le O(\|R_k\|) + o(\|x^k - \bar{x}\|).$$

By combining (iv) of Proposition 2.1.1 with (2.9), we obtain that $\|R_k\| \le O(\|x^k - \bar{x}\|^2)$. It follows that (2.25) holds. This completes the proof. □

We are now ready to prove Theorem 2.3.2.

Proof of Theorem 2.3.2. By Lemma 2.3.4 we know that $\{x^k\}$ converges to $\bar{x}$. In virtue of Lemma 2.3.4, it then remains to show that the unit steplength in Algorithm 2.2.1 can be always chosen for sufficiently large k. By virtue of Proposition 2.1.1, by using the fact that $F(\bar{x}) = 0$, we know that there exist $\overline{\sigma} > \underline{\sigma} > 0$ satisfying for sufficiently large k,

$$\underline{\sigma}\|x^k - \bar{x}\| \leq \|F(x^k)\| \leq \overline{\sigma}\|x^k - \bar{x}\|.$$

Since $x^k + \Delta x^k$ is closer to $\bar{x}$ than x^k (cf. (2.25)), we further obtain that for sufficiently large k,

$$\underline{\sigma}\|x^k + \Delta x^k - \bar{x}\| \leq \|F(x^k + \Delta x^k)\| \leq \overline{\sigma}\|x^k + \Delta x^k - \bar{x}\|,$$

which implies that for sufficiently large k,

$$\|F(x^k + \Delta x^k)\| \leq \frac{\overline{\sigma}}{\underline{\sigma}} \frac{\|x^k + \Delta x^k - \bar{x}\|}{\|x^k - \bar{x}\|}\|F(x^k)\| \leq o(1)\|F(x^k)\|, \tag{2.27}$$

where the second inequality follows from (2.25).

Next, we prove that that for sufficiently large k, the unit steplength is always satisfied by considering the following two cases:

Case I. If $|\text{Ind}_1| = +\infty$. Then, there exists sufficiently large k such that at the $(k-1)$-th iteration, $k \in \text{Ind}_1$ and $f^{pre} = \|F(x^k)\|$. It follows from (2.27) that the condition (2.11) is always satisfied for sufficiently large k and hence $x^{k+1} = x^k + \Delta x^k$.

Case II. If $|\text{Ind}_1| < +\infty$. Then, since $\lim_{x\to\bar{x}} \theta(x^k) = \theta(\bar{x})$ (cf. Theorem 2.3.1), we know that $\liminf_{k\to\infty} \|F(x^k)\| = 0$ and hence $|\text{Ind}_2| = +\infty$. This means that there exists a sufficiently large k such that at the $(k-1)$-th iteration, $k \in \text{Ind}_2$ and $f^{pre} = \|F(x^k)\|$. The same arguments as in Case I) lead to $x^{k+1} = x^k + \Delta x^k$.

Thus, by using (2.25) in Lemma 2.3.4, we know that (2.19) holds. The proof is completed. □

Chapter 3

Numerical Experiments

In this chapter, we shall take the following special least squares covariance matrix problem (3.1) as an example to demonstrate the efficiency of our inexact SQP Newton method:

$$\begin{array}{ll} \min & \frac{1}{2}\|X - C\|^2 \\ \text{s.t.} & X_{ij} = e_{ij}, \ (i,j) \in \mathcal{B}_e \,, \\ & X_{ij} \geq l_{ij}, \ (i,j) \in \mathcal{B}_l \,, \\ & X_{ij} \leq u_{ij}, \ (i,j) \in \mathcal{B}_u \,, \\ & X \in \mathcal{S}^n_+ \,, \end{array} \tag{3.1}$$

where $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ are three index subsets of $\{(i,j) \,|\, 1 \leq i \leq j \leq n\}$ satisfying $\mathcal{B}_e \cap \mathcal{B}_l = \emptyset$, $\mathcal{B}_e \cap \mathcal{B}_u = \emptyset$, and $l_{ij} < u_{ij}$ for any $(i,j) \in \mathcal{B}_l \cap \mathcal{B}_u$. Denote the cardinalities of $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ by p, q_l, and q_u, respectively. Let $m := p + q_l + q_u$. For any $(i,j) \in \{1, \ldots, n\} \times \{1, \ldots, n\}$, define $\mathcal{E}^{ij} \in \Re^{n \times n}$ by

$$(\mathcal{E}^{ij})_{st} := \begin{cases} 1 & \text{if } (s,t) = (i,j)\,, \\ 0 & \text{otherwise}\,, \end{cases} \qquad s,t = 1, \ldots, n\,.$$

Thus, problem (3.1) can be written as a special case of (1.2) with

$$\mathcal{A}(X) := \begin{bmatrix} \{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_e} \\ \{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_l} \\ -\{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_u} \end{bmatrix}, \quad X \in \mathcal{S}^n \tag{3.2}$$

and

$$b := \begin{pmatrix} \{e_{ij}\}_{(i,j)\in\mathcal{B}_e} \\ \{l_{ij}\}_{(i,j)\in\mathcal{B}_l} \\ -\{u_{ij}\}_{(i,j)\in\mathcal{B}_u} \end{pmatrix},$$

where $A^{ij} := \frac{1}{2}(\mathcal{E}^{ij} + \mathcal{E}^{ji})$. Then, its dual problem takes the same form as (1.3) with $q := q_l + q_u$.

In our numerical experiments, we compare our inexact SQP Newton method, which is referred as `Inexact-SQP` in the numerical results, with the exact SQP Newton and the inexact smoothing Newton method of Gao and Sun [7], which are referred as `Exact-SQP` and `Smoothing`, respectively, for solving the least squares covariance matrix problem with simple constraints (3.1). We also use `Smoothing` to solve our subproblems (2.7) approximately.

We implemented all algorithms in `MATLAB 7.3` running on a Laptop of Intel Core Duo CPU and 3.0GB of RAM. The testing examples are given below.

Example 3.0.1. Let $n = 387$. The matrix C is the $n \times n$ 1-day correlation matrix from the lagged datasets of RiskMetrics (`www.riskmetrics.com/stddownload_edu.html`). For the test purpose, we perturb C to

$$C := (1-\alpha)C + \alpha R,$$

where $\alpha \in (0,1)$ and R is a randomly generated symmetric matrix with entries in $[-1,1]$. The `MATLAB` code for generating the random matrix R is: `R = 2.0*rand(n,n)-ones(n,n); R = triu(R)+triu(R,1)'; for i=1:n; R(i,i) = 1;`

`end`. Here we take $\alpha = 0.1$ and

$$\mathcal{B}_e := \{(i,i) \,|\, i = 1, \ldots, n\}\,.$$

The two index sets $\mathcal{B}_l$, $\mathcal{B}_u \subset \{(i,j) \,|\, 1 \leq i < j \leq n\}$ consist of the indices of $\min(\hat{n}_r, n-i)$ randomly generated elements at the ith row of X, $i = 1, \ldots, n$ with $\hat{n}_r$ taking the following values: 1, 5, 10, 20, 50, 100, and 150. We take $l_{ij} \in [-0.5, 0.5]$ for $(i,j) \in \mathcal{B}_l$ randomly and set $u_{ij} = 0.5$ for $(i,j) \in \mathcal{B}_u$.

Example 3.0.2. The matrix C is a randomly generated $n \times n$ symmetric matrix with entries in $[-1, 1]$. The index sets $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ are generated in the same as in Example 3.0.1 with $\hat{n}_r = 1$, 5, 10, 20, 50, 100, and 150. We test for $n = 500$ and $n = 1000$, respectively.

We report the numerical results in Tables 3.1-3.2, where "*Iter*" and "*Res*" stand for the number of total iterations and the residue at the final iterate of an algorithm, respectively. The cputime is reported in the `hour:minute:second` format.

		Example 3.0.1		
Method	$\hat{n}_r$	*Iter*	*cputime*	*Res*
`Exact-SQP`	1	9	0:44	1.1e-8
	5	10	1:21	4.0e-8
	10	10	2:01	8.1e-9
	20	10	3:01	2.0e-8
	50	10	11:20	2.4e-7
	100	11	25:07	7.7e-7
	150	12	48:59	1.3e-8
`Inexact-SQP`	1	9	0:22	9.3e-9
	5	10	0:43	2.0e-8
	10	10	1:06	5.7e-8
	20	10	1:36	3.9e-8
	50	10	4:21	2.7e-7
	100	12	13:36	5.3e-8
	150	12	19:49	1.9e-8
`Smoothing`	1	8	0:17	1.2e-8
	5	10	0:27	5.2e-9
	10	10	0:32	2.1e-7
	20	12	0:52	1.9e-7
	50	22	6:05	6.4e-8
	100	23	26:01	5.0e-8
	150	22	14:07	9.9e-8

Table 3.1: Numerical results for Example 3.0.1

Example 3.0.2		n=500			n=1000		
Method	$\hat{n}_r$	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*
Exact-SQP	1	7	0:29	3.5e-7	8	4:06	1.8e-8
	5	8	0:53	5.1e-7	9	6:56	9.3e-8
	10	9	1:29	7.5e-8	10	9:20	6.0e-8
	20	10	4:05	2.0e-8	11	16:46	2.4e-7
	50	10	8:55	1.0e-7	13	39:33	1.6e-8
	100	11	28:39	7.7e-8	13	1:49:13	2.2e-7
	150	12	57:27	4.8e-8	13	3:17:41	1.5e-7
Inexact-SQP	1	7	0:16	3.6e-7	8	2:01	5.3e-8
	5	8	0:27	7.4e-7	9	3:31	1.5e-7
	10	9	0:47	1.2e-7	10	4:04	1.3e-7
	20	10	1:35	4.9e-8	11	8:44	2.6e-7
	50	11	5:16	1.0e-8	13	25:01	2.0e-8
	100	11	13:39	7.7e-8	13	57:46	2.4e-7
	150	12	25:29	2.4e-8	13	1:24:08	1.7e-7
Smoothing	1	7	0:13	1.5e-7	7	1:31	5.4e-7
	5	9	0:20	1.2e-7	9	2:26	5.0e-8
	10	9	0:28	1.6e-7	9	2:49	9.5e-7
	20	12	1:11	1.3e-8	11	4:18	1.6e-8
	50	12	1:42	1.9e-7	15	9:23	9.3e-8
	100	19	12:31	1.2e-8	18	17:00	1.2e-8
	150	24	46:33	6.0e-8	21	27:36	8.1e-8

Table 3.2: Numerical results for Example 3.0.2

Chapter 4

Conclusions

In this paper, we introduced a globally and superlinearly convergent inexact SQP Newton method – Algorithm 2.2.1 for solving convex SC^1 minimization problems. Our method much relaxes the restrictive BD-regularity assumption made by Pang and Qi in [10]. The conducted numerical results for solving the least squares covariance matrix problem with simple constraints (3.1) showed that Algorithm 2.2.1 is more effective than its exact version. For most of the tested examples, Algorithm 2.2.1 is less efficient than the inexact smoothing Newton method of Gao and Sun [7]. Nevertheless, it is very competitive when the number of the constraints is large, i.e., when m is huge. Further study is needed in order to fully disclose the behavior of our introduced inexact SQP Newton method. This is, however, beyond the scope of this thesis.

Bibliography

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

[2] J. Borwein and A. S. Lewis, *Partially finite convex programming, part II: Explicit lattice models*, Mathematical Programming 57 (1992), pp. 49–83.

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

[4] B.C. Eaves, *On the basic theorem for complemenarity*, Mathematical Programming 1 (1971), pp. 68–75.

[5] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems: Volumes I & II*, Springer, New York, 2003.

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

[7] Yan Gao and Defeng Sun, *Calibrating Least Squares Covariance Matrix Problems with Equality and Inequality Constraints*, Technical Report, Department of Mathematics, National University of Singapore, 2008.

[8] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[9] R. Mifflin, *Semismoothness and semiconvex functions in constrained optimization*, SIAM Journal on Control Optimization 15 (1977), pp. 959–972.

[10] J.S. Pang and L. Qi, *A globally convergent Newton method for convex SC^1 minimization problems*, Journal of Optimization Theory and Applications, 85 (1995), pp. 633–648.

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

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

[13] G. Ravindran and M.S. Gowda, *Regularization of P_0-functions in box variational inequality problems*, SIAM Journal on Optimization 11 (2000/01), pp. 748–760.

[14] R.T. Rockafellar, *Convex Analyis*, Princeton University Press, Princeton, 1970.

[15] R.T. Rockafellar, *Conjugate Duality and Optimization*, SIAM, Philadelphia, 1974.

[16] R.T. Rockafellar and R.J.B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[17] N.C. Schwertman and D.M. Allen, *Smoothing an indefinite variance-covariance matrix*, Journal of Statistical Computation and Simulation 9 (1979), pp. 183–194.

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

[19] D.F. Sun and J. Sun, *Löwner's operator and spectral functions in Euclidean Jordan algebras*, Mathematics of Operations Research 33 (2008), pp. 421–445.

[20] H.A. van der Vorst, *BI-CGSTAB: a fast and smoothly converging variant of BI-CG for the solution of nonsymmetric linear systems*, SIAM Journal on Scientific and Statistical Computing 13 (1992), pp. 631-644.

[21] E.H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory I and II.* In E.H. Zarantonello, editor, Contributions to Nonlinear Functional Analysis, pp. 237–424, Academic Press, New York, 1971.

Name: Yidi Chen

Degree: Master of Science

Department: Mathematics

Thesis Title: An Inexact SQP Newton Method for Convex SC^1 Minimization Problems

Abstract

The convex SC^1 minimization problems model many problems as special cases. One particular example is the dual problem of the least squares covariance matrix (LSCM) problems with inequality constraints. The purpose of this thesis is to introduce an efficient inexact SQP Newton method for solving the general convex SC^1 minimization problems under realistic assumptions. In Chapter 2, we introduce our method and conduct a complete convergence analysis including the super-linear (quadratic) rate of convergence. Numerical results conducted in Chapter 3 show that our inexact SQP Newton method is competitive when it is applied to the LSCM problems with many lower and upper bounds constraints. We make our final conclusions in Chapter 4.

Keywords:

SC^1 minimization, inexact SQP Newton method, super-linear convergence

AN INEXACT SQP NEWTON METHOD FOR CONVEX SC^1 MINIMIZATION PROBLEMS

CHEN YIDI

NATIONAL UNIVERSITY OF SINGAPORE

2008

An Inexact SQP Newton Method for Convex SC^1 Minimization Problems **Chen Yidi** **2008**