

Complexity Analysis of Interior Point Algorithms for Non-Lipschitz and Nonconvex Minimization

Wei Bian · Xiaojun Chen · Yinyu Ye

July 25, 2012, Received: date / Accepted: date

Abstract We propose a first order interior point algorithm for a class of non-Lipschitz and nonconvex minimization problems with box constraints, which arise from applications in variable selection and regularized optimization. The objective functions of these problems are continuously differentiable typically at interior points of the feasible set. Our algorithm is easy to implement and the objective function value is reduced monotonically along the iteration points. We show that the worst-case complexity for finding an ϵ scaled first order stationary point is $O(\epsilon^{-2})$. Moreover, we develop a second order interior point algorithm using the Hessian matrix, and solve a quadratic program with ball constraint at each iteration. Although the second order interior point algorithm costs more computational time than that of the first order algorithm in each iteration, its worst-case complexity for finding an ϵ scaled second order stationary point is reduced to $O(\epsilon^{-3/2})$. An ϵ scaled second order stationary point is an ϵ scaled first order stationary point.

Keywords constrained non-Lipschitz optimization · complexity analysis · interior point method · first order algorithm · second order algorithm

This work was supported partly by Hong Kong Research Council Grant PolyU5003/10p, The Hong Kong Polytechnic University Postdoctoral Fellowship Scheme and the NSF foundation (11101107,10971043) of China.

Wei Bian

Department of Mathematics, Harbin Institute of Technology, Harbin, China. Current address: Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong. E-mail: bianweilvse520@163.com.

Xiaojun Chen

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China. E-mail: maxjchen@polyu.edu.hk

Yinyu Ye

Department of Management Science and Engineering, Stanford University, Stanford, CA 94305. E-mail: yinyu-ye@stanford.edu

Mathematics Subject Classification (2000) 90C30 · 90C26 · 65K05 · 49M37

1 Introduction

In this paper, we consider the following optimization problem:

$$\begin{aligned} \min \quad & f(x) = H(x) + \lambda \sum_{i=1}^n \varphi(x_i^p) \\ \text{s.t.} \quad & x \in \Omega = \{x : 0 \leq x \leq b\}, \end{aligned} \quad (1)$$

where $H \in \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable, $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ is continuous and concave, $\lambda > 0$, $0 < p < 1$, $b = (b_1, b_2, \dots, b_n)^T$ with $b_i > 0$, $i = 1, 2, \dots, n$. Moreover, φ is continuously differentiable in \mathbb{R}_{++} and $\varphi(0) = 0$. Without loss of generality, we assume that a minimizer of (1) exists and $\min_{\Omega} f(x) \geq 0$.

Problem (1) is nonsmooth, nonconvex, and non-Lipschitz, which has been extensively used in image restoration, signal processing and variable selection; see, e.g., [8, 11, 13, 19]. The function $H(x)$ is often used as a data fitting term, while the function $\sum_{i=1}^n \varphi(x_i^p)$ is used as a regularization term. Numerical experiments indicate these type of problems could be solved effectively for finding a local minimizer or stationary point. But little theoretical complexity or convergence speed analysis of the problems is known, which is in contrast to complexity study of convex optimization in the past thirty years.

There were few results on complexity analysis of nonconvex optimization problems. Using an interior-point algorithm, Ye [17] proved that an ϵ scaled KKT or first order stationary point of general quadratic programming can be computed in $O(\epsilon^{-1} \log(\epsilon^{-1}))$ iterations where each iteration would solve a ball-constrained or trust-region quadratic program that is equivalent to a simplex convex quadratic minimization problem. He also proved that, as $\epsilon \rightarrow 0$, the iterative sequence converges to a point satisfying the scaled second order necessary optimality condition. The same complexity result was extended to linearly constrained concave minimization by Ge et al. [10].

Cartis, Gould and Toint [2] estimated the worst-case complexity of a first order trust-region or quadratic regularization method for solving the following unconstrained nonsmooth, nonconvex minimization problem

$$\min_{x \in \mathbb{R}^n} \Phi_h(x) := H(x) + h(c(x)),$$

where $h : \mathbb{R}^m \rightarrow \mathbb{R}$ is convex but may be nonsmooth and $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is continuously differentiable. They show that their method takes at most $O(\epsilon^{-2})$ steps to reduce the size of a first order criticality measure below ϵ , which is the same in order as the worst-case complexity of steepest-descent methods applied to unconstrained, nonconvex smooth optimization. However, f in (1) cannot be defined in the form of Φ_h .

Garmanjani and Vicente [9] proposed a class of smoothing direct-search methods for the unconstrained optimization of nonsmooth functions by applying a direct-search method to the smoothing function \tilde{f} of the objective function f [3]. When f is locally Lipschitz, the smoothing direct-search method [9] took at most $O(-\epsilon^{-3} \log \epsilon)$ iterations to find an x such that $\|\nabla \tilde{f}(x, \mu)\| \leq \epsilon$ and $\mu \leq \epsilon$, where μ is the smoothing parameter. When $\mu \rightarrow 0$, $\tilde{f}(x, \mu) \rightarrow f(x)$ and $\nabla \tilde{f}(x, \mu) \rightarrow v$ with $v \in \partial f(x)$.

Recently, Bian and Chen [1] proposed a smoothing sequential quadratic programming (SSQP) algorithm for solving the following non-Lipshchitz unconstrained minimization problem

$$\min_{x \in \mathbb{R}^n} f_0(x) := H(x) + \lambda \sum_{i=1}^n \varphi(|x_i|^p). \quad (2)$$

At each step, the SSQP algorithm solves a strongly convex quadratic minimization problem with a diagonal Hessian matrix, which has a simple closed-form solution. The SSQP algorithm is easy to implement and its worst-case complexity of reaching an ϵ scaled stationary point is $O(\epsilon^{-2})$.

Obviously, the objective functions f of (1) and f_0 of (2) are identical in \mathbb{R}_+^n . Moreover, f is smooth in the interior of \mathbb{R}_+^n . Note that problem (2) includes the l_2 - l_p problem

$$\min_{x \in \mathbb{R}^n} \|Ax - c\|^2 + \lambda \sum_{i=1}^n |x_i|^p \quad (3)$$

as a special case, where $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^m$, and $p \in (0, 1)$.

In this paper, we analyze the worst-case complexity of interior point methods for solving problem (1). We propose a first order interior point algorithm with the worst-case complexity for finding an ϵ scaled first order stationary point being $O(\epsilon^{-2})$, and develop a second order interior point algorithm with the worst-case complexity of it for finding an ϵ scaled second order stationary point being $O(\epsilon^{-3/2})$.

Our paper is organized as follows. In Section 2, a first order interior point algorithm is proposed for solving (1), which only uses ∇f and a Lipschitz constant of H on Ω and is easy to implement. Any iteration point $x^k > 0$ belongs to Ω and the objective function is monotonically decreasing along the generated sequence $\{x^k\}$. Moreover, the algorithm produces an ϵ scaled first order stationary point of (1) in at most $O(\epsilon^{-2})$ steps. In Section 3, a second order interior point algorithm is given to solve a special case of (1), where H is twice continuously differentiable, $\varphi := t$ and $\Omega = \{x : x \geq 0\}$. By using the Hessian of H , the second order interior point algorithm can generate an ϵ interior scaled second order stationary point in at most $O(\epsilon^{-3/2})$ steps.

Throughout this paper, $\mathbb{K} = \{0, 1, 2, \dots\}$, $I = \{1, 2, \dots, n\}$, $I_b = \{i \in \{1, 2, \dots, n\} : b_i < +\infty\}$ and $e_n = (1, 1, \dots, 1)^T \in \mathbb{R}^n$. For $x \in \mathbb{R}^n$, $A = (a_{ij})_{m \times n} \in \mathbb{R}^{m \times n}$ and $q > 0$, $|A|^q = (|a_{ij}|^q)_{m \times n}$, $x^q = (x_1^q, x_2^q, \dots, x_n^q)^T$, $[x_i]_{i=1}^n := x$, $|x| = (|x_1|, |x_2|, \dots, |x_n|)^T$, $\|x\| = \|x\|_2 := (x_1^2 + x_2^2 + \dots + x_n^2)^{\frac{1}{2}}$ and $\text{diag}(x) = \text{diag}(x_1, x_2, \dots, x_n)$. For two matrices $A, B \in \mathbb{R}^{n \times n}$, we denote $A \succeq B$ when $A - B$ is positive semi-definite.

2 First Order Method

In this section, we propose a first order interior point algorithm for solving (1), which uses the first order reduction technique and keeps all iterates $x^k > 0$ in the feasible set Ω . We show that the objective function value $f(x^k)$ is monotonically decreasing along the sequence $\{x^k\}$ generated by the algorithm, and the worst-case complexity of the algorithm for generating an ϵ interior scaled first order stationary point of (1) is $O(\epsilon^{-2})$, which is the same in the worst-case complexity order of the steepest-descent methods for nonconvex smooth optimization problems. Moreover, it is worth noting that the proposed first order interior point algorithm is easy to implement, and computation cost at each step is little.

Throughout this section, we need the following assumptions.

Assumption 2.1: ∇H is globally Lipschitz on Ω with a Lipschitz constant β .

Specially, when $I_b \neq \emptyset$ we choose β such that $\beta \geq \max_{i \in I_b} \frac{1}{b_i}$ and $\beta \geq 1$.

Assumption 2.2: For any given $x^0 \in \Omega$, there is $R \geq 1$ such that $\sup\{\|x\|_\infty : f(x) \leq f(x_0), x \in \Omega\} \leq R$.

When $H(x) = \frac{1}{2}\|Ax - c\|^2$, Assumption 2.1 holds with $\beta = \|A^T A\|$. Assumption 2.2 holds, if Ω is bounded or $H(x) \geq 0$ for all $x \in \Omega$ and $\varphi(t) \rightarrow \infty$ as $t \rightarrow \infty$.

2.1 First Order Necessary Conditions

Note that for problem (1), when $0 < p < 1$, the Clarke generalized gradient of $\varphi(|s|^p)$ does not exist at 0. Inspired by the scaled first and second order necessary conditions for local minimizers of unconstrained non-Lipschitz optimization in [5,6], we give the scaled first and second order necessary condition for local minimizers of the constrained non-Lipschitz optimization (1) in this section. Then, for any $\epsilon \in (0, 1]$, the ϵ scaled first order and second order necessary stationary point of (1) can be deduced directly.

First, for $\epsilon > 0$, an ϵ global minimizer of (1) is defined as a feasible solution $0 \leq x_\epsilon \leq b$ and

$$f(x_\epsilon) - \min_{0 \leq x \leq b} f(x) \leq \epsilon.$$

It has been proved that finding a global minimizer of the unconstrained l_2 - l_p minimization problem (3) is strongly NP hard in [4]. For the unconstrained l_2 - l_p optimization (3), any local minimizer x satisfies the first order necessary condition [6]

$$XA^T(Ax - c) + \lambda p|x|^p = 0, \quad (4)$$

and the second order necessary condition

$$XA^TAX + \lambda p(p-1)|X|^p \succeq 0, \quad (5)$$

where $|X|^p = \text{diag}(|x_1|^p, \dots, |x_n|^p)$.

For (1), if x is a local minimizer of (1) at which f is differentiable, then $x \in \Omega$ satisfies

- (i) $[\nabla f(x)]_i = 0$ if $x_i \neq b_i$;
- (ii) $[\nabla f(x)]_i \leq 0$ if $x_i = b_i$.

Although $[\nabla f(x)]_i$ does not exist when $x_i = 0$, one can see that, as $x_i \rightarrow 0+$, $[\nabla f(x)]_i \rightarrow +\infty$.

Denote

$$X\nabla f(x) = X\nabla H(x) + \lambda p[\nabla\varphi(s)_{s=x_i^p x_i^p}]_{i=1}^n.$$

Similarly, using X as a scaling matrix, if x is a local minimizer of (1), then $x \in \Omega$ satisfies the scaled first order necessary condition

$$[X\nabla f(x)]_i = 0 \quad \text{if } x_i \neq b_i; \quad (6a)$$

$$[\nabla f(x)]_i \leq 0 \quad \text{if } x_i = b_i. \quad (6b)$$

Now we can define an ϵ scaled first order stationary point of (1).

Definition 1 For a given $0 < \epsilon \leq 1$, we call $x \in \Omega$ an ϵ scaled first order stationary point of (1), if there is $\delta > 0$ such that

- (i) $|[X\nabla f(x)]_i| \leq \epsilon$ if $x_i < b_i - \delta\epsilon$;
- (ii) $[\nabla f(x)]_i \leq \epsilon$ if $x_i \geq b_i - \delta\epsilon$.

Definition 1 is consistent with the first order necessary conditions in (6a)-(6b) with $\epsilon = 0$. Moreover, Definition 1 is consistent with the first order necessary conditions given in [6] with $\epsilon = 0$ for unconstrained optimization.

2.2 First Order Interior Point Algorithm

Note that for any $x, x^+ \in (0, b]$, Assumption 2.1 implies that

$$H(x^+) \leq H(x) + \langle \nabla H(x), x^+ - x \rangle + \frac{\beta}{2} \|x^+ - x\|^2. \quad (7)$$

Since φ is concave on $[0, +\infty)$, then for any $s, t \in (0, +\infty)$,

$$\varphi(t) \leq \varphi(s) + \langle \nabla\varphi(s), t - s \rangle. \quad (8)$$

Thus, for any $x, x^+ \in (0, b]$, we obtain

$$f(x^+) \leq f(x) + \langle \nabla f(x), x^+ - x \rangle + \frac{\beta}{2} \|x^+ - x\|^2. \quad (9)$$

Let $Xd_x = x^+ - x$. We obtain

$$f(x^+) \leq f(x) + \langle X\nabla f(x), d_x \rangle + \frac{\beta}{2} \|Xd_x\|^2. \quad (10)$$

We now use the reduction idea to solve the constrained non-Lipschitz optimization problem (1). To achieve a reduction of the objective function, we

minimize a quadratic function subject to a box constraint at each step when $x > 0$, which is

$$\begin{aligned} \min \quad & \langle X \nabla f(x), d_x \rangle + \frac{\beta}{2} d_x^T X^2 d_x \\ \text{s.t.} \quad & d_x^2 \leq \frac{1}{4} e_n, \quad d_x \leq X^{-1}(b - x). \end{aligned} \quad (11)$$

For any fixed $x \in (0, b]$, the objective function in (11) is strongly convex and separable about every element of x , then the unique solution of (11) has a close form as

$$d_x = P_{\mathcal{D}_x} \left[-\frac{1}{\beta} X^{-1} \nabla f(x) \right],$$

where $\mathcal{D}_x = [-\frac{1}{2}e_n, \min\{\frac{1}{2}e_n, X^{-1}(b - x)\}]$ and $P_{\mathcal{D}_x}$ is the orthogonal project operator on the box \mathcal{D}_x .

For simplicity, we use d_k and \mathcal{D}_k to denote d_{x^k} and \mathcal{D}_{x^k} , respectively. Denote $X_{k+1} = \text{diag}(x^{k+1})$ and $X_k = \text{diag}(x^k)$.

First Order Interior Point Algorithm

Give $\epsilon \in (0, 1]$ and choose $x^0 \in (0, b]$.

For $k \geq 0$, set

$$d_k = P_{\mathcal{D}_k} \left[-\frac{1}{\beta} X_k^{-1} \nabla f(x^k) \right] \quad (12a)$$

$$x^{k+1} = x^k + X_k d_k. \quad (12b)$$

Lemma 1 *The proposed First Order Interior Point Algorithm is well defined, which means that $x^k \in (0, b]$, $\forall k \in \mathbb{K}$.*

Proof We only need to prove that if $0 < x^k \leq b$, then $0 < x^{k+1} \leq b$.

On the one hand, by $d_k \leq X_k^{-1}(b - x^k)$, we have

$$x^{k+1} = x^k + X_k d_k \leq x^k + (b - x^k) = b.$$

On the other hand, using $d_k \geq -\frac{1}{2}e_n$, we obtain

$$x^{k+1} = x^k + X_k d_k \geq x^k - \frac{1}{2}x^k = \frac{1}{2}x^k > 0.$$

Hence, $0 < x^{k+1} \leq b$.

Lemma 2 *Let $\{x^k\}$ be the sequence generated by the First Order Interior Point Algorithm, then the sequence $\{f(x^k)\}$ is monotonely decreasing and satisfies*

$$f(x^{k+1}) - f(x^k) \leq -\frac{\beta}{2} \|X_k d_k\|^2 = -\frac{\beta}{2} \|x^{k+1} - x^k\|^2. \quad (13)$$

Moreover, we have $\|x^k\|_\infty \leq R$.

Proof From the KKT condition of (11), the solution d_k of quadratic programming (11) satisfies the necessary and sufficient condition as follows

$$\begin{aligned} \beta X_k^2 d_k + X_k \nabla f(x^k) + M_k d_k + \nu_k &= 0, \quad d_k^2 \leq \frac{1}{4} e_n, \quad d_k \leq X_k^{-1}(b - x^k), \\ M_k(d_k^2 - \frac{1}{4} e_n) &= 0, \quad N_k(d_k - X_k^{-1}(b - x^k)) = 0, \end{aligned} \quad (14)$$

where $M_k = \text{diag}(\mu_k)$ and $N_k = \text{diag}(\nu_k)$ with $M_k, N_k \succeq 0$.

Moreover, from (10), we obtain

$$\begin{aligned} f(x^{k+1}) - f(x^k) &\leq \langle X_k \nabla f(x), d_k \rangle + \frac{\beta}{2} \|X_k d_k\|^2 \\ &= \langle -\beta X_k^2 d_k - M_k d_k - \nu_k, d_k \rangle + \frac{\beta}{2} \|X_k d_k\|^2 \quad (15) \\ &= -\frac{\beta}{2} \|X_k d_k\|^2 - d_k^T M_k d_k - \nu_k^T d_k. \end{aligned}$$

Fix $i \in I$. If $[d_k]_i < \frac{b-x^k}{x^k}$, then $[\nu_k]_i = 0$. On the other hand, if $[d_k]_i = \frac{b-x^k}{x^k} > 0$, then $[\nu_k]_i \geq 0$. Thus,

$$[d_k]_i [\nu_k]_i \geq 0, \quad \forall i \in I. \quad (16)$$

From (15), (16) and (12b), we get

$$f(x^{k+1}) - f(x^k) \leq -\frac{\beta}{2} \|X_k d_k\|^2 = -\frac{\beta}{2} \|x^{k+1} - x^k\|^2.$$

Thus, $f(x^{k+1}) \leq f(x^k)$, which implies that $f(x^k) \leq f(x^0)$, $k \in \mathbb{K}$.

From Assumption 2.2, we obtain $\|x^k\|_\infty \leq R$.

Different from some other potential reduction methods, the objective function is monotonely decreasing along the sequence generated by the First Order interior Point Algorithm.

Theorem 1 *For any $\epsilon \in (0, 1]$, the First Order Interior Point Algorithm obtains an ϵ scaled first order stationary point or ϵ global minimizer of (1) in no more than $O(\epsilon^{-2})$ steps.*

Proof Let $\{x^k\}$ be the sequence generated by the proposed First Order Interior Point Algorithm. Then $x^k \in (0, b]$ and $\|x^k\|_\infty \leq R$, $\forall k \in \mathbb{K}$. Without loss of generality, we suppose that $R \geq 1$.

In the following, we will consider four cases.

Case 1: $\|X_k d_k\| \geq \frac{1}{4\beta R} \epsilon$.

From (13), we obtain that

$$f(x^{k+1}) - f(x^k) \leq -\frac{1}{2 \times 4^2 R^2 \beta} \epsilon^2 = -\frac{1}{32 R^2 \beta} \epsilon^2.$$

Case 2: $\|\mu_k\|_\infty \geq \frac{1}{4\beta} \epsilon^2$.

From (14), if $\|\mu_k\|_\infty \geq \frac{1}{4\beta}\epsilon^2$, then there is $i \in I$ such that $[\mu_k]_i \geq \frac{1}{4\beta}\epsilon^2$ and $[d_k]_i^2 = \frac{1}{4}$.

Using the above analysis into (15), we get

$$f(x^{k+1}) - f(x^k) \leq -\frac{1}{16\beta}\epsilon^2. \quad (17)$$

Case 3: $\nu_k^T d_k \geq \frac{1}{24}\epsilon^2$.

From (15), we get

$$f(x^{k+1}) - f(x^k) \leq -\frac{1}{24}\epsilon^2. \quad (18)$$

Case 4: $\|X_k d_k\| < \frac{1}{4\beta R}\epsilon$, $\|\mu^k\|_\infty < \frac{1}{4\beta}\epsilon^2$ and $\nu_k^T d_k < \frac{1}{24}\epsilon^2$.

From the first condition in (14), we obtain that

$$\beta X_k d_k + \nabla f(x^k) + M_k X_k^{-1} d_k + X_k^{-1} \nu_k = 0. \quad (19)$$

From $d_k^2 \leq \frac{1}{4}e_n$, we have

$$x^{k+1} = x^k + X_k d_k \leq \frac{3}{2}x^k. \quad (20)$$

Then, we get

$$\|X_{k+1} X_k^{-1} d_k\|_\infty \leq \|d_k\|_\infty \|X_{k+1} X_k^{-1}\|_\infty \leq 2\|d_k\|_\infty \leq 1. \quad (21)$$

By mean value theorem and $\|X_{k+1}\|_\infty \leq R$, there is $\tau \in [0, 1]$ such that

$$\begin{aligned} & \|X_{k+1}(\nabla f(x^{k+1}) - \beta X_k d_k - \nabla f(x^k))\|_\infty \\ &= \|X_{k+1}(\nabla^2 f(\tau x^{k+1} + (1-\tau)x^k)(x^{k+1} - x^k) - \beta X_k d_k)\|_\infty \\ &\leq \beta \|X_{k+1} X_k d_k\|_\infty + \beta \|X_{k+1} X_k d_k\|_\infty \\ &\leq 2\beta R \|X_k d_k\|_\infty < \frac{1}{2}\epsilon. \end{aligned} \quad (22)$$

By virtue of $\|\mu_k\|_\infty < \frac{1}{4\beta}\epsilon^2$ and (20), we obtain

$$\|X_{k+1} M_k X_k^{-1} d_k\|_\infty \leq 2\|\mu_k\|_\infty \|d_k\|_\infty < \frac{1}{4\beta}\epsilon^2 < \frac{\epsilon}{4}. \quad (23)$$

Then, from (14), (20), (22) and (23), we obtain

$$\begin{aligned} & |[X_{k+1} \nabla f(x^{k+1})]_i| \\ &= |[X_{k+1}(\nabla f(x^{k+1}) - \beta X_k d_k - \nabla f(x^k) - M_k X_k^{-1} d_k - X_k^{-1} \nu_k)]_i| \\ &\leq \|X_{k+1}(\nabla f(x^{k+1}) - \beta X_k d_k - \nabla f(x^k))\|_\infty + \|X_{k+1} M_k X_k^{-1} d_k\|_\infty \\ &\quad + |[X_{k+1} X_k^{-1} \nu_k]_i| \\ &\leq \frac{3}{4}\epsilon + \frac{3}{2}|[\nu_k]_i|. \end{aligned} \quad (24)$$

Fix $i \in I$. We consider two subclasses in this case.

Case 4.1: $x_i^{k+1} < b_i - \frac{1}{2\beta}\epsilon$. From $\|X_k d_k\| \leq \frac{1}{4\beta R}\epsilon \leq \frac{1}{4\beta}\epsilon$, then $x_i^k \leq b_i - \frac{\epsilon}{4\beta}$. Then,

$$\frac{b_i - x_i^k}{x_i^k} \geq \frac{\epsilon}{4\beta b_i} \geq \frac{\epsilon}{4}.$$

When $[\nu_k]_i = 0$, then

$$|[X_{k+1} \nabla f(x^{k+1})]_i| \leq \frac{3}{4}\epsilon.$$

On the other hand, when $[\nu_k]_i \neq 0$, then $[d_k]_i = \frac{b_i - x_i^k}{x_i^k} \geq \frac{\epsilon}{4}$. Combining with (16) and $\nu_k^T d_k \leq \frac{\epsilon^2}{24}$, it follows that

$$[\nu_k]_i \leq \frac{\epsilon}{6}. \quad (25)$$

Then,

$$|[X_{k+1} \nabla f(x^{k+1})]_i| \leq \epsilon.$$

Case 4.2: $x_i^{k+1} \geq b_i - \frac{\epsilon}{2\beta}$.

Then, $x_i^{k+1} \geq \frac{b_i}{2}$. By (20), $x_i^k \geq \frac{b_i}{4}$. Then,

$$|[M_k X_k^{-1} d_k]_i| \leq \frac{\epsilon^2}{2\beta b_i} \leq \frac{\epsilon}{2}.$$

Similar to the calculation method in (22) and by $[\nu_k]_i \geq 0$, we obtain

$$\begin{aligned} [\nabla f(x^{k+1})]_i &= [\nabla f(x^{k+1}) - \beta X_k d_k - \nabla f(x^k) - M_k X_k^{-1} d_k - X_k^{-1} \nu_k]_i \\ &\leq \|\nabla f(x^{k+1}) - \beta X_k d_k - \nabla f(x^k)\|_\infty + |[M_k X_k^{-1} d_k]_i| - [X_k^{-1} \nu_k]_i \\ &\leq 2\beta \|X_k d_k\|_\infty + \frac{1}{2}\epsilon - [X_k^{-1} \nu_k]_i \leq \epsilon - [X_k^{-1} \nu_k]_i \leq \epsilon. \end{aligned} \quad (26)$$

Therefore, from the analysis in Cases 4.1 - 4.2, x^{k+1} is an ϵ scaled first order stationary point of (1).

Basing on the above analysis in Cases 1 - 3, at least one of the following two facts holds at the k th iteration:

- (i) $f(x^{k+1}) - f(x^k) \leq -\frac{1}{32R^2\beta}\epsilon^2$;
- (ii) x^{k+1} is an ϵ scaled first order stationary point of (1).

Therefore, we would produce an ϵ global minimizer or an ϵ scaled first order stationary point of (1) in at most $32f(x^0)R^2\beta\epsilon^{-2}$.

3 Second Order Interior Point Algorithm

In this section, we consider a special case of (1) as follows

$$\begin{aligned} \min \quad & f(x) = H(x) + \lambda \sum_{i=1}^n x_i^p \\ \text{s.t.} \quad & x \in \Omega = \{x : x \geq 0\}, \end{aligned} \quad (27)$$

where λ and p are defined as in (1). In this section, we need Assumption 2.2 and the following assumption on H .

Assumption 3.1: H is twice continuously differentiable and $\nabla^2 H$ is globally Lipschitz on Ω with Lipschitz constant γ .

We will propose a second order interior point algorithm for solving (27), by using the Hessian of H in the algorithm. We show that the worst-case complexity of the second order interior point algorithm for finding an ϵ scaled second order stationary point of (27) is $O(\epsilon^{-\frac{3}{2}})$. Comparing with the first order interior point algorithm proposed in Section 2, the worst-case complexity of the second order interior point algorithm is better and the generated point satisfies stronger optimality conditions. However, a quadratic program with ball constraint has to be solved at each step, which may be nonconvex.

3.1 Second Order Necessary Condition for (27)

Besides the scaled second order necessary condition for unconstrained $l_2 - l_p$ optimization given in [6], Chen, Niu and Yuan[5] present a scaled second order necessary condition and a scaled second order sufficient condition for more general unconstrained non-Lipschitz optimization. Based on these scaled necessary conditions in [5, 6], we define a second order necessary condition for (27).

First, if x is a local minimizer of (27), then $x \in \Omega$ must satisfy

- (i) $X \nabla f(x) = 0$;
- (ii) $X \nabla^2 f(x) X \succeq 0$.

For $i \in I$, note that when $x_i = 0$, then $X_{ii} = 0$ and

$$[X \nabla^2 f(x) X]_{ii} = [X \nabla^2 H(x) X]_{ii}.$$

Now we give the definition of the ϵ scaled second order stationary point of (27) as follows.

Definition 2 For a given $\epsilon \in (0, 1]$, we call $x \in \Omega$ an ϵ scaled second order stationary point of (27), if

$$\|X \nabla f(x)\|_{\infty} \leq \epsilon \quad (28a)$$

$$X \nabla^2 f(x) X \succeq -\sqrt{\epsilon} I_n. \quad (28b)$$

Definition 2 is consistent with the scaled second order necessary conditions given above when $\epsilon = 0$. Moreover, Definition 2 is consistent with the scaled second order necessary conditions given in [6] for unconstrained optimization.

3.2 Second Order Interior Point Algorithm

From Taylor expansion, for any $x, x^+ \in \Omega$, Assumption 3.1 implies that

$$\begin{aligned} H(x^+) &\leq H(x) + \langle \nabla H(x), x^+ - x \rangle \\ &\quad + \frac{1}{2} \langle \nabla^2 H(x)(x^+ - x), x^+ - x \rangle + \frac{1}{6} \gamma \|x^+ - x\|^3. \end{aligned} \quad (29)$$

Similarly, for any $t, s \in \mathbb{R}_{++}$,

$$\begin{aligned} t^p &\leq s^p + \langle ps^{p-1}, t - s \rangle \\ &\quad + \frac{p(p-1)}{2} s^{p-2} (t-s)^2 + \frac{p(p-1)(p-2)}{6} s^{p-3} (t-s)^3. \end{aligned} \quad (30)$$

Thus, for any $x, x^+ \in \text{int}(\Omega)$, we obtain

$$\begin{aligned} f(x^+) - f(x) &\leq \langle \nabla f(x), x^+ - x \rangle + \frac{1}{2} \langle \nabla^2 f(x)(x^+ - x), x^+ - x \rangle \\ &\quad + \frac{1}{6} \gamma \|x^+ - x\|^3 + \lambda \frac{p(p-1)(p-2)}{6} \sum_{i=1}^n x_i^{p-3} (x_i^+ - x_i)^3, \end{aligned} \quad (31)$$

which can also be expressed by

$$\begin{aligned} f(x^+) - f(x) &\leq \langle X \nabla f(x), d_x \rangle + \frac{1}{2} \langle X \nabla^2 f(x) X d_x, d_x \rangle \\ &\quad + \frac{1}{6} \gamma \|X d_x\|^3 + \frac{\lambda p(p-1)(p-2)}{6} \sum_{i=1}^n x_i^p [d_x]_i^3. \end{aligned} \quad (32)$$

Since $p(1-p) \leq \frac{1}{4}$, then $p(1-p)(2-p) \leq \frac{1}{2}$. Combining this estimation with $\sum_{i=1}^n [d_k]_i^3 \leq \|d_k\|^3$, if $\|x\|_\infty \leq R$, then (32) implies

$$\begin{aligned} f(x^+) - f(x) &\leq \langle X \nabla f(x), d_x \rangle + \frac{1}{2} \langle X \nabla^2 f(x) X d_x, d_x \rangle \\ &\quad + \frac{1}{6} (\gamma R^3 + \frac{1}{2} \lambda R^p) \|d_x\|^3. \end{aligned} \quad (33)$$

In this section, we minimize a quadratic function subject to a ball constraint. For a given $\epsilon \in (0, 1]$ and $x \in \text{int}(\Omega)$, at each iteration, we solve the following problem

$$\begin{aligned} \min \quad & q(d_x) = \langle X \nabla f(x), d_x \rangle + \frac{1}{2} \langle X \nabla^2 f(x) X d_x, d_x \rangle \\ \text{s.t.} \quad & \|d_x\|^2 \leq \vartheta^2 \epsilon, \end{aligned} \quad (34)$$

where $\vartheta = \frac{1}{2} \min\{\frac{1}{\gamma R^3 + \frac{1}{2} \lambda R^p}, 1\}$ and R is the constant in Assumption 2.2.

To solve (34), we consider two cases.

Case 1: $X \nabla^2 f(x) X$ is positive semi-definite.

From the KKT condition of (34), the solution d_x of (34) satisfies the necessary and sufficient conditions as follows

$$X\nabla^2 f(x)Xd_x + X\nabla f(x) + \rho_x d_x = 0, \quad (35a)$$

$$\rho_x \geq 0, \quad \|d_x\|^2 \leq \vartheta^2 \epsilon, \quad \rho_x(\|d_x\|^2 - \vartheta^2 \epsilon) = 0. \quad (35b)$$

In this case, (34) is a convex quadratic programming with ball constraint, which can be solved effectively in polynomial time, see [15] and references therein.

Case 2: $X\nabla^2 f(x)X$ has at least one negative eigenvalue.

From [14], the solution d_x of (34) satisfies the following necessary and sufficient conditions

$$X\nabla^2 f(x)Xd_x + X\nabla f(x) + \rho_x d_x = 0, \quad (36a)$$

$$\|d_x\|^2 = \vartheta^2 \epsilon, \quad X\nabla^2 f(x)X + \rho_x I_n \text{ is positive semi-definite.} \quad (36b)$$

In this case, (34) is a nonconvex quadratic programming with ball constraint. However, by the analysis results in [17, 18], (36a)-(36b) can also be effectively solved in polynomial time with the worst-case complexity $O(\log(\log(1/\epsilon)))$.

Therefore, we can assume that the subproblem (34) can be solved effectively.

From (35) and (36), if d_x solves (34), then

$$\begin{aligned} q(d_x) &= \langle X\nabla f(x), d_x \rangle + \frac{1}{2} \langle X\nabla^2 f(x)Xd_x, d_x \rangle \\ &= \langle -X\nabla^2 f(x)Xd_x - \rho_x d_x, d_x \rangle + \frac{1}{2} \langle X\nabla^2 f(x)Xd_x, d_x \rangle \quad (37) \\ &= -\rho_x \|d_x\|^2 - \frac{1}{2} d_x^T X\nabla^2 f(x)Xd_x. \end{aligned}$$

Since $X\nabla^2 f(x)X + \rho_x I_n$ is always positive semi-definite,

$$d_x^T X\nabla^2 f(x)Xd_x \geq -\rho_x \|d_x\|^2,$$

and thus

$$q(d_x) \leq -\frac{1}{2} \rho_x \|d_x\|^2.$$

Therefore, from (33), we have

$$f(x^+) - f(x) \leq -\frac{1}{2} \rho_x \|d_x\|^2 + \frac{1}{6} (\gamma R^3 + \frac{1}{2} \lambda R^p) \|d_x\|^3. \quad (38)$$

Denote $X_{k+1} = \text{diag}(x^{k+1})$ and $X_k = \text{diag}(x^k)$. And we also use d_k and ρ_k to denote d_{x^k} and ρ_{x^k} for simplicity in this section.

Second Order Interior Point Algorithm

Choose $x^0 \in \text{int}(\Omega)$ and $\epsilon \in (0, 1]$.

For $k \geq 0$,

$$\text{Solve (34) with } x = x^k \text{ for } d_k \quad (39a)$$

$$x^{k+1} = x^k + X_k d_k. \quad (39b)$$

From the definition of ϑ in (34), $\|d_k\|_\infty \leq \frac{1}{2}$, similar to the analysis in Lemma 1, the Second Order Interior Point Algorithm is also well defined. Let $\{x^k\}$ be the sequence generated by it, then $x^k \in \text{int}(\Omega)$. In the following theorem, we will prove that there is $\kappa > 0$ such that either $f(x^{k+1}) - f(x^k) \leq -\kappa\epsilon^{\frac{3}{2}}$ or x^{k+1} is an ϵ scaled second order stationary point of (27).

Lemma 3 *If $\rho_k \geq \frac{4}{9}(\gamma R^3 + \frac{1}{2}\lambda R^p)\|d_k\|$ holds for all $k \in \mathbb{K}$, then the Second Order Interior Point Algorithm produces an ϵ global minimizer of (1) in at most $O(\epsilon^{-\frac{3}{2}})$ steps. Moreover, $\|x^k\|_\infty \leq R, \forall k \in \mathbb{K}$.*

Proof If $\rho_k \geq \frac{4}{9}(\gamma R^3 + \frac{1}{2}\lambda R^p)\|d_k\|$, we have that

$$\frac{1}{6}(\gamma R^3 + \frac{1}{2}\lambda R^p)\|d_k\| \leq \frac{3}{8}\rho_k, \quad (40)$$

and from (35b) and (36b), we obtain

$$\|d_k\|^2 = \vartheta^2\epsilon.$$

From (38) and (40), we have

$$\begin{aligned} f(x^{k+1}) - f(x^k) &\leq -\frac{1}{2}\rho_k\|d_x\|^2 + \frac{1}{6}(\gamma R^3 + \frac{1}{2}\lambda R^p)\|d_k\|^3 \\ &\leq -\frac{1}{2}\rho_k\|d_x\|^2 + \frac{3}{8}\rho_k\|d_x\|^2 = -\frac{1}{8}\rho_k\|d_k\|^2, \end{aligned}$$

which means $f(x^{k+1}) \leq f(x^k)$.

If this always holds, then $f(x^k)$ is monotonely decreasing. By Assumption 2.2, $\|x^k\|_\infty \leq R, k \in \mathbb{K}$. Moreover,

$$\begin{aligned} f(x^k) - f(x^0) &\leq -\frac{k}{8}\rho_k\|d_k\|^2 \\ &\leq -\frac{k}{18}(\gamma R^3 + \frac{1}{2}\lambda R^p)\|d_k\|^3 \\ &= -\frac{k}{18}(\gamma R^3 + \frac{1}{2}\lambda R^p)\vartheta^3\epsilon^{\frac{3}{2}}, \end{aligned}$$

which follows that we would produce an ϵ global minimizer of (1) in at most $18f(x^0)(\gamma R^3 + \frac{1}{2}\lambda R^p)^{-1}\vartheta^{-3}\epsilon^{-\frac{3}{2}}$ steps.

In what follows, we prove that x^{k+1} is an ϵ scaled second order stationary point of (27) when $\rho_k < \frac{4}{9}(\gamma R^3 + \frac{1}{2}\lambda R^p)\|d_k\|$ for some k .

Lemma 4 *If there is $k \in \mathbb{K}$ such that $\rho_k < \frac{4}{9}(\gamma R^3 + \frac{1}{2}\lambda R^p)\|d_k\|$, then x^{k+1} satisfies*

$$\|X_{k+1}\nabla f(x^{k+1})\|_\infty \leq \epsilon.$$

Proof From (35a) and (36a), the following relation always holds

$$\begin{aligned} -\rho_k d_k &= X_k \nabla^2 f(x^k) X_k d_k + X_k \nabla f(x^k) \\ &= X_k \nabla^2 H(x^k) X_k d_k + \lambda p(p-1) X_k^{p-2} X_k^2 d_k + X_k \nabla H(x^k) + \lambda p X_k (x^k)^{p-1} \\ &= X_k (\nabla^2 H(x^k) X_k d_k + \lambda p(p-1) X_k^{p-2} X_k d_k + \nabla H(x^k) + \lambda p (x^k)^{p-1}). \end{aligned}$$

Then,

$$\nabla^2 H(x^k) X_k d_k + \lambda p(p-1) X_k^{p-2} X_k d_k + \nabla H(x^k) + \lambda p (x^k)^{p-1} + \rho_k X_k^{-1} d_k = 0.$$

Thus, there is $\tau \in [0, 1]$ such that

$$\begin{aligned} &\nabla H(x^{k+1}) + \lambda p (x^{k+1})^{p-1} \\ &= \nabla H(x^{k+1}) + \lambda p (x^{k+1})^{p-1} - \nabla^2 H(x^k) X_k d_k - \lambda p(p-1) X_k^{p-2} X_k d_k - \nabla H(x^k) \\ &\quad - \lambda p (x^k)^{p-1} - \rho_k X_k^{-1} d_k \\ &= \nabla^2 H(\tau x^k + (1-\tau)x^{k+1}) X_k d_k - \nabla^2 H(x^k) X_k d_k \\ &\quad + \lambda p (x^{k+1})^{p-1} - \lambda p(p-1) X_k^{p-2} X_k d_k - \lambda p (x^k)^{p-1} - \rho_k X_k^{-1} d_k. \end{aligned} \tag{41}$$

Therefore,

$$\begin{aligned} &\|X_{k+1} (\nabla H(x^{k+1}) + \lambda p (x^{k+1})^{p-1})\|_\infty \\ &\leq \|X_{k+1} (\nabla^2 H(\tau x^k + (1-\tau)x^{k+1}) X_k d_k - \nabla^2 H(x^k) X_k d_k)\|_\infty \\ &\quad + \|X_{k+1} (\lambda p (x^{k+1})^{p-1} - \lambda p(p-1) X_k^{p-2} X_k d_k - \lambda p (x^k)^{p-1})\|_\infty \\ &\quad + \rho_k \|X_{k+1} X_k^{-1} d_k\|_\infty. \end{aligned} \tag{42}$$

From Assumption 3.1, we estimate the first term after the inequality in (42) as follows

$$\begin{aligned} &\|X_{k+1}\|_\infty \|\nabla^2 H(\tau x^k + (1-\tau)x^{k+1}) X_k d_k - \nabla^2 H(x^k) X_k d_k\|_\infty \\ &\leq \gamma(1-\tau) \|X_{k+1}\|_\infty \|X_k d_k\|_\infty^2 \leq \gamma \|X_{k+1}\|_\infty \|X_k\|_\infty^2 \|d_k\|_\infty^2 \leq \gamma R^3 \|d_k\|_\infty^2. \end{aligned} \tag{43}$$

From $X_k D_k = X^{k+1} - X^k$, we have

$$X_k^{-1} X_{k+1} = D_k + I_n, \tag{44}$$

with $D_k = \text{diag}(d_k)$, which implies

$$X_k^{-1} x^{k+1} = d_k + e_n.$$

Then, we consider the second term in (42),

$$\begin{aligned} &X_{k+1} (\lambda p (x^{k+1})^{p-1} - \lambda p(p-1) X_k^{p-2} X_k d_k - \lambda p (x^k)^{p-1}) \\ &= \lambda p X_k^p (X_k^{-p} (x^{k+1})^p + (1-p) X_k^{-1} X_{k+1} d_k - X_k^{-1} x^{k+1}) \\ &= \lambda p X_k^p ((d_k + e_n)^p + (1-p)(D_k + I_n) d_k - (d_k + e_n)) \\ &= \lambda p X_k^p ((d_k + e_n)^p + (1-p)d_k^2 - p d_k - e_n). \end{aligned} \tag{45}$$

Using the Taylor expansion that $1 + pt - \frac{p(1-p)}{2}t^2 \leq (1+t)^p \leq 1 + pt$, we obtain

$$e_n + pd_k - \frac{p(1-p)}{2}d_k^2 \leq (d_k + e_n)^p \leq e_n + pd_k. \quad (46)$$

Adding $(1-p)d_k^2 - pd_k - e_n$ into (46), we have

$$0 \leq (d_k + e_n)^p + (1-p)d_k^2 - pd_k - e_n \leq (1-p)d_k^2.$$

Thus,

$$\|(d_k + e_n)^p + (1-p)d_k^2 - pd_k - e_n\|_\infty \leq (1-p)\|d_k^2\|_\infty. \quad (47)$$

Then, from (45) and (47), we get

$$\begin{aligned} & \|X_{k+1}(\lambda p(x^{k+1})^{p-1} - \lambda p(p-1)X_k^{p-2}X_k d_k - \lambda p(x^k)^{p-1})\|_\infty \\ & \leq \lambda p \|X_k^p\|_\infty \|(d_k + e_n)^p + (1-p)d_k^2 - pd_k - e_n\|_\infty \\ & \leq \lambda p(1-p) \|X_k^p\|_\infty \|d_k^2\|_\infty \leq \frac{1}{2} \lambda R^p \|d_k\|_\infty^2. \end{aligned} \quad (48)$$

Similar to the analysis in (21), we have

$$\rho_k \|X_{k+1}X_k^{-1}d_k\|_\infty \leq \rho_k \|d_k\|_\infty (1 + \|d_k\|_\infty) \leq \frac{3}{2} \rho_k \|d_k\|_\infty. \quad (49)$$

Therefore, from (42), (43), (48) and (49), we obtain

$$\begin{aligned} & \|X_{k+1} \nabla H(x^{k+1}) + \lambda p(x^{k+1})^p\|_\infty \\ & \leq (\gamma R^3 + \frac{1}{2} \lambda R^p) \|d_k\|_\infty^2 + \frac{3}{2} \rho_k \|d_k\|_\infty \\ & \leq (\gamma R^3 + \frac{1}{2} \lambda R^p) \|d_k\|_\infty^2 + \frac{2}{3} (\gamma R^3 + \frac{1}{2} \lambda R^p) \|d_k\|_\infty^2 \\ & = \frac{5}{3} (\gamma R^3 + \frac{1}{2} \lambda R^p) \vartheta^2 \epsilon \leq \frac{5}{12} \min\{\gamma R^3 + \frac{1}{2} \lambda R^p, \frac{1}{\gamma R^3 + \frac{1}{2} \lambda R^p}\} \epsilon \leq \frac{5}{12} \epsilon. \end{aligned}$$

Lemma 5 Under the assumptions in Lemma 4, x^{k+1} satisfies

$$X_{k+1} \nabla^2 f(x^{k+1}) X_{k+1} \succeq -\sqrt{\epsilon} I_n.$$

Proof From (35b) and (36b), we know

$$X_k \nabla^2 f(x^k) X_k + \rho_k I_n \text{ is positive semi-definite.}$$

Then,

$$\nabla^2 H(x^k) + \lambda p(p-1)X_k^{p-2} \succeq -\rho_k X_k^{-2}; \quad (50)$$

From Assumption 3.1, we obtain

$$\|\nabla^2 H(x^{k+1}) - \nabla^2 H(x^k)\| \leq \gamma \|x^k - x^{k+1}\| \leq \gamma \|X_k\| \|d_k\| \leq \gamma R \|d_k\|. \quad (51)$$

Note that $\nabla^2 H(x^{k+1})$ and $\nabla^2 H(x^k)$ are symmetric, (51) gives

$$\nabla^2 H(x^{k+1}) - \nabla^2 H(x^k) \succeq -\gamma R \|d_k\| I_n. \quad (52)$$

Adding (50) and (52), we get

$$\nabla^2 H(x^{k+1}) \succeq -\lambda p(p-1)X_k^{p-2} - \rho_k X_k^{-2} - \gamma R \|d_k\| I_n. \quad (53)$$

Adding $\lambda p(p-1)X_{k+1}^{p-2}$ into the both sides of (53), we obtain

$$\begin{aligned} & X_{k+1} \nabla^2 H(x^{k+1}) X_{k+1} + \lambda p(p-1)X_{k+1} X_{k+1}^{p-2} X_{k+1} \\ & \succeq -\lambda p(p-1)X_{k+1} X_k^{p-2} X_{k+1} - \rho_k X_{k+1} X_k^{-2} X_{k+1} \\ & \quad - \gamma R \|d_k\| X_{k+1}^2 + \lambda p(p-1)X_{k+1} X_{k+1}^{p-2} X_{k+1}. \end{aligned} \quad (54)$$

Using (44) again, we get

$$-\rho_k X_{k+1} X_k^{-2} X_{k+1} = -\rho_k (D_k + I_n)^2 \succeq -\frac{9}{4} \rho_k I_n. \quad (55)$$

On the other hand, using (44), we have

$$\begin{aligned} & X_{k+1} X_{k+1}^{p-2} X_{k+1} - X_{k+1} X_k^{p-2} X_{k+1} = X_{k+1}^p - X_{k+1}^2 X_k^{p-2} \\ & = X_{k+1}^p (I_n - (X_{k+1}^{2-p} X_k^{p-2})) = X_{k+1}^p (I_n - (I_n + D_k)^{2-p}). \end{aligned} \quad (56)$$

From the Taylor expansion that $(1+t)^{2-p} \leq 1 + (2-p)t + \frac{(2-p)(1-p)}{2} t^2$, we get

$$(I_n + D_k)^{2-p} \preceq I_n + (2-p)D_k + \frac{1}{2}(2-p)(1-p)D_k^2. \quad (57)$$

Applying (57), $D_k \preceq \frac{1}{2}I_n$ and $0 < p < 1$ to (56), we derive

$$\begin{aligned} & \lambda p(1-p)(X_{k+1} X_{k+1}^{p-2} X_{k+1} - X_{k+1} X_k^{p-2} X_{k+1}) \\ & = \lambda p(1-p)X_{k+1}^p (I_n - (I_n + D_k)^{2-p}) \\ & \succeq -\lambda p(1-p)X_{k+1}^p ((2-p)D_k + \frac{1}{2}(2-p)(1-p)D_k^2) \\ & \succeq -\lambda p(1-p)\|x_{k+1}\|_\infty^p ((2-p)\|d_k\|_\infty + \frac{1}{2}(2-p)(1-p)\|d_k\|_\infty^2) I_n \\ & \succeq -\lambda p(1-p)R^p (2-p + \frac{(2-p)(1-p)}{2}) \|d_k\|_\infty I_n \\ & \succeq -\lambda R^p \frac{p(1-p)(2-p)(3-p)}{2} \|d_k\|_\infty I_n \\ & \succeq -\frac{\lambda}{2} R^p \|d_k\|_\infty I_n. \end{aligned} \quad (58)$$

From (54), (55), (56), (58), and $\rho_k < \frac{4}{9}(\gamma R^3 + \frac{\lambda}{2} R^p) \|d_k\|$, we obtain

$$\begin{aligned} & X_{k+1} \nabla^2 f(x^{k+1}) X_{k+1} \\ & = X_{k+1} \nabla^2 H(x^{k+1}) X_{k+1} + \lambda p(p-1)X_{k+1} X_{k+1}^{p-2} X_{k+1} \\ & \succeq -(\frac{9}{4}\rho_k + \gamma R^3 \|d_k\|_\infty + \frac{1}{2}\lambda R^p \|d_k\|_\infty) I_n \\ & \succeq -2(\gamma R^3 + \frac{\lambda}{2} R^p) \|d_k\|_\infty I_n \\ & \succeq -\min\{1, \frac{1}{\gamma R^3 + \frac{1}{2}\lambda R^p}\} \sqrt{\epsilon} I_n \succeq -\sqrt{\epsilon} I_n. \end{aligned} \quad (59)$$

According to Lemmas 3 - 5, we can obtain the complexity of the Second Order Interior Point Algorithm for finding an ϵ scaled second order stationary point of (27).

Theorem 2 *For any $\epsilon \in (0, 1]$, the proposed Second Order Interior Point Algorithm obtains an ϵ scaled second order stationary point or ϵ global minimizer of (27) in no more than $O(\epsilon^{-3/2})$ steps.*

Remark 1 When $H(x) = \frac{1}{2}\|Ax - c\|^2$ in (27), then $\gamma = 0$ and ϑ in (34) turns to be

$$\vartheta = \min\left\{\frac{1}{\lambda R^p}, \frac{1}{2}\right\}.$$

The proposed Second Order Interior Point Algorithm obtains an ϵ scaled second order stationary point or ϵ global minimizer of (27) in no more than $36f(x^0) \max\{\lambda^2 R^{2p}, 8\lambda R^p\} \epsilon^{-\frac{3}{2}}$ steps.

4 Final remarks

This paper proposes two interior point methods for solving constrained non-Lipschitz, nonconvex optimization problems arising in many important applications. The first order interior point method is easy to implement and its worst-case complexity is $O(\epsilon^{-2})$ which is the same in order as the worst-case complexity of steepest-descent methods applied to unconstrained, nonconvex smooth optimization, and the trust region methods and SSQP methods applied to unconstrained, nonconvex nonsmooth optimization [1, 2]. The second order interior method has a better complexity order $O(\epsilon^{-\frac{3}{2}})$ for finding an ϵ scaled second order stationary point. To the best of our knowledge, this is the first method for finding an ϵ scaled second order stationary point with complexity analysis.

Assumptions in this paper are standard and applicable to many regularization models in practice. For example, $H(x) = \|Ax - c\|^2$ and φ is one of the following six penalty functions

- i) soft thresholding penalty function [11, 12]: $\varphi_1(s) = s$
- ii) logistic penalty function [13]: $\varphi_2(s) = \log(1 + \alpha s)$
- iii) fraction penalty function [7, 13]: $\varphi_3(s) = \frac{\alpha s}{1 + \alpha s}$
- iv) hard thresholding penalty function[8]: $\varphi_4(s) = \lambda - (\lambda - s)_+^2 / \lambda$
- v) smoothly clipped absolute deviation penalty function[8]:

$$\varphi_5(s) = \int_0^s \min\left\{1, \frac{(\alpha - t/\lambda)_+}{\alpha - 1}\right\} dt$$

- vi) minimax concave penalty function [19]:

$$\varphi_6(s) = \int_0^s \left(1 - \frac{t}{\alpha\lambda}\right)_+ dt.$$

Here α and λ are two positive parameters, especially, $\alpha > 2$ in $\varphi_5(s)$ and $\alpha > 1$ in $\varphi_6(s)$. These six penalty functions are concave in R_+ and continuously differentiable in R_{++} , which are often used in statistics and sparse reconstruction.

References

1. W. Bian and X. Chen, Smoothing SQP algorithm for non-Lipschitz optimization with complexity analysis, Preprint, 2012.
2. C. Cartis, N. I. M. Gould and P. Toint, On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming, *SIAM J. Optim.*, 21, 1721-1739 (2011).
3. X. Chen, Smoothing methods for nonsmooth, nonconvex minimization, *Math. Program.*, 134, 71-99 (2012).
4. X. Chen, D. Ge, Z. Wang and Y. Ye, Complexity of unconstrained L_2 - L_p minimization, submitted to *Math. Program.*, under minor revision.
5. X. Chen, L. Niu and Y. Yuan, Optimality conditions and smoothing trust region Newton method for non-Lipschitz optimization, Preprint, 2012.
6. X. Chen, F. Xu and Y. Ye, Lower bound theory of nonzero entries in solutions of l_2 - l_p minimization, *SIAM J. Sci. Comput.*, 32, 2832-2852 (2010).
7. X. Chen and W. Zhou, Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization, *SIAM J. Imaging Sci.*, 3, 765-790 (2010).
8. J. Fan, Comments on 'Wavelets in statistics: a review' by A. Antoniadis, *Stat. Method. Appl.*, 6, 131-138 (1997).
9. R. Garmanjani and L.N. Vicente, Smoothing and worst case complexity for direct-search methods in nonsmooth optimization, *IMA J. Numer. Anal.*, to appear.
10. D. Ge, X. Jiang and Y. Ye, A note on the complexity of L_p minimization, *Math. Program.*, 21, 1721-1739 (2011).
11. J. Huang, J. L. Horowitz and S. Ma, Asymptotic properties of bridge estimators in sparse high-dimensional regression models, *Ann. Statist.*, 36, 587-613 (2008).
12. R. Tibshirani, Shrinkage and selection via the Lasso, *J. Roy. Statist. Soc. Ser. B*, 58, 267-288 (1996).
13. M. Nikolova, M.K. Ng, S. Zhang and W.-K. Ching, Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization, *SIAM J. Imaging Sci.*, 1, 2-25 (2008).
14. S.A. Vavasis, R. Zippel, Proving polynomial time for sphere-constrained quadratic programming, Technical Report 90-1182, Department of Computer Science, Cornell University, Ithaca, NY, 1990.
15. S.A. Vavasis, *Nonlinear Optimization: Complexity Issues*, Oxford Sciences, New York, 1991.
16. Y. Ye, *Interior point algorithms: theory and analysis*, John Wiley & Sons, Inc., New York, 1997.
17. Y. Ye, On the complexity of approximating a KKT point of quadratic programming, *Math. Program.*, 80, 195-211 (1998).
18. Y. Ye, A new complexity result on minimization of a quadratic function with a sphere constraint, in *Recent Advances in Global Optimization*, C. Floudas and P.M. Pardalos, eds., Princeton University Press, Princeton, NJ (1992).
19. C.-H Zhang, Nearly unbiased variable selection under minimax concave penalty, *Ann. Statist.*, 38, 894-942, (2010).