Iterative Reweighted Minimization Methods for l_p Regularized Unconstrained Nonlinear Programming

Zhaosong Lu*

October 6, 2012

Abstract

In this paper we study general l_p regularized unconstrained minimization problems. In particular, we derive lower bounds for nonzero entries of first- and second-order stationary points, and hence also of local minimizers of the l_p minimization problems. We extend some existing iterative reweighted l_1 (IRL$_1$) and l_2 (IRL$_2$) minimization methods to solve these problems and proposed new variants for them in which each subproblem has a closed form solution. Also, we provide a unified convergence analysis for these methods. In addition, we propose a novel Lipschitz continuous ϵ-approximation to $\|x\|_p^p$. Using this result, we develop new IRL$_1$ methods for the l_p minimization problems and showed that any accumulation point of the sequence generated by these methods is a first-order stationary point, provided that the approximation parameter ϵ is below a computable threshold value. This is a remarkable result since all existing iterative reweighted minimization methods require that ϵ be dynamically updated and approach zero. Our computational results demonstrate that the new IRL$_1$ method is generally more stable than the existing IRL$_1$ methods [21, 18] in terms of objective function value and CPU time.

Key words: l_p minimization, iterative reweighted l_1 minimization, iterative reweighted l_2 minimization

1 Introduction

Recently numerous optimization models and methods have been proposed for finding sparse solutions to a system or an optimization problem (e.g., see [28, 14, 8, 7, 24, 9, 11, 10, 13, 29, 21, 5, 1, 23, 30, 32, 26, 33]). In this paper we are interested in one of those models, namely, the l_p regularized unconstrained nonlinear programming model

$$\min_{x\in\Re^n}\{F(x) := f(x) + \lambda\|x\|_p^p\}, \qquad (1)$$

*Department of Mathematics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada. (email: zhaosong@sfu.ca). This work was supported in part by NSERC Discovery Grant.

for some $\lambda > 0$ and $p \in (0, 1)$, where f is a smooth function with L_f-Lipschitz-continuous gradient in $\Re^n$, that is,

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L_f \|x - y\|_2, \quad \forall x, y \in \Re^n,$$

and f is bounded below in $\Re^n$. Here, $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$ for any $x \in \Re^n$. One can observe that as $p \downarrow 0$, problem (1) approaches the l_0 minimization problem

$$\min_{x \in \Re^n} f(x) + \lambda \|x\|_0, \tag{2}$$

which is an exact formulation of finding a sparse vector to minimize the function f. Some efficient numerical methods such as iterative hard thresholding [5] and penalty decomposition methods [26] have recently been proposed for solving (2). In addition, as $p \uparrow 1$, problem (1) approaches the l_1 minimization problem

$$\min_{x \in \Re^n} f(x) + \lambda \|x\|_1, \tag{3}$$

which is a widely used convex relaxation for (2). When f is a convex quadratic function, model (3) is shown to be extremely effective in finding a sparse vector to minimize f. A variety of efficient methods were proposed for solving (3) over last few years (e.g., see [29, 1, 23, 30, 32]). Since problem (1) is intermediate between problems (2) and (3), one can expect that it is also capable of seeking out a sparse vector to minimize f. As demonstrated by extensive computational studies in [11, 31], problem (1) can even produce a sparser solution than (3) does while both achieve similar values of f.

A great deal of effort was recently made by many researchers (e.g., see [11, 12, 13, 21, 31, 25, 17, 18, 20, 15, 22, 27, 2, 16]) for studying problem (1) or its related problem

$$\min_{x \in \Re^n} \{\|x\|_p^p : Ax = b\}. \tag{4}$$

In particular, Chartrand [11], Chartrand and Staneva [12], Foucart and Lai [21], and Sun [27] established some sufficient conditions for recovering the sparest solution to a undetermined linear system $Ax = b$ by the model (4). Efficient iterative reweighted l_1 (IRL_1) and l_2 (IRL_2) minimization algorithms were also proposed for finding an approximate solution to (4) by Foucart and Lai [21] and Daubechies et al. [20], respectively. Though problem (4) is generally NP hard (see [15, 22]), it is shown in [21, 20] that under some assumptions, the sequence generated by IRL_1 and IRL_2 algorithms converges to the sparest solution to the above linear system, which is also the global minimizer of (4). In addition, Chen et al. [17] considered a special case of problem (1) with $f(x) = \frac{1}{2}\|Ax - b\|_2^2$, namely, the problem

$$\min_{x \in \Re^n} \frac{1}{2}\|Ax - b\|_2^2 + \lambda \|x\|_p^p. \tag{5}$$

They derived lower bounds for nonzero entries of local minimizers of (5) and also proposed a hybrid orthogonal matching pursuit-smoothing gradient method for solving (5). Since $\|x\|_p^p$ is non-Lipschitz continuous, Chen and Zhou [18] recently considered the following approximation to (5):

$$\min_{x\in\Re^n} \frac{1}{2}\|Ax-b\|_2^2 + \lambda \sum_{i=1}^n (|x_i|+\epsilon)^p$$

for some small $\epsilon > 0$. And they also proposed an IRL_1 algorithm to solve this approximation problem. Recently, Lai and Wang [25] considered another approximation to (5), which is

$$\min_{x\in\Re^n} \frac{1}{2}\|Ax-b\|_2^2 + \lambda \sum_{i=1}^n (|x_i|^2+\epsilon)^{p/2},$$

and proposed an IRL_2 algorithm for solving this approximation. Very recently, Bian and Chen [2] and Chen et al. [16] proposed a smoothing sequential quadratic programming (SQP) algorithm and a smoothing trust region Newton (TRN) method, respectively, for solving a class of nonsmooth nonconvex problems that include (1) as a special case. When applied to problem (1), their methods first approximate $|x|_p^p$ by a suitable smooth function and then apply an SQP or a TRN algorithm to solve the resulting approximation problem. Lately, Bian et al. [3] proposed first- and second-order interior point algorithms for solving a class of non-Lipschitz and nonconvex minimization problems with *bounded* box constraints, which can be suitably applied to l_p regularized minimization problems over a *compact* box.

In this paper we consider general l_p regularized unconstrained optimization problem (1). In particular, we first derive lower bounds for nonzero entries of first- and second-order stationary points, and hence also of local minimizers of (1). We then extend the aforementioned IRL_1 and IRL_2 methods [21, 20, 25, 18] to solve (1) and propose some new variants for them. We also provide a unified convergence analysis for these methods. Finally, we propose a novel Lipschitz continuous ϵ-approximation to $\|x\|_p^p$ and also propose a locally Lipschitz continuous function $F_\epsilon(x)$ to approximate $F(x)$. Subsequently, we develop IRL_1 minimization methods for solving the resulting approximation problem $\min_{x\in\Re^n} F_\epsilon(x)$. We show that any accumulation point of the sequence generated by these methods is a first-order stationary point of problem (1), provided that ϵ is below a computable threshold value. This is a remarkable result since all existing iterative reweighted minimization methods for l_p minimization problems require that ϵ be dynamically updated and approach zero.

The outline of this paper is as follows. In Subsection 1.1 we introduce some notations that are used in the paper. In Section 2 we derive lower bounds for nonzero entries of stationary points, and hence also of local minimizers of problem (1). We also propose a locally Lipschitz continuous function $F_\epsilon(x)$ to approximate $F(x)$ and study some properties of the approximation problem $\min_{x\in\Re^n} F_\epsilon(x)$. In Section 3, we extend the existing IRL_1 and IRL_2 minimization methods from problems (4) and (5) to general problems (1) and propose new variants for them. We also provide a unified convergence analysis for these methods. In Section 4 we propose new IRL_1 methods for solving (1) and establish their convergence. In Section 5, we conduct numerical experiments to compare the performance of several IRL_1 minimization methods that are studied in this paper for (1). Finally, in Section 6 we present some concluding remarks.

1.1 Notation

The set of all n-dimensional positive vectors is denoted by $\Re^n_{++}$. Given any $x \in \Re^n$ and a scalar τ, $|x|^\tau$ denotes an n-dimensional vector whose ith component is $|x_i|^\tau$. In addition, $\mathrm{Diag}(x)$ denotes an $n \times n$ diagonal matrix whose diagonal is formed by the vector x. Given an index set $\mathcal{B} \subseteq \{1, \ldots, n\}$, $x_\mathcal{B}$ denotes the sub-vector of x indexed by $\mathcal{B}$. Similarly, $X_{\mathcal{B}\mathcal{B}}$ denotes the sub-matrix of X whose rows and columns are indexed by $\mathcal{B}$. In addition, if a matrix X is positive semidefinite, we write $X \succeq 0$. The sign operator is denoted by sgn, that is,

$$\mathrm{sgn}(t) = \begin{cases} 1 & \text{if } t > 0, \\ [-1, 1] & \text{if } t = 0, \\ -1 & \text{otherwise.} \end{cases}$$

Finally, for any $\beta < 0$, we define $0^\beta = \infty$.

2 Technical results

In this section we derive lower bounds for nonzero entries of stationary points, and hence also of local minimizers of problem (1). We also propose a nonsmooth but locally Lipschitz continuous function $F_\epsilon(x)$ to approximate $F(x)$. Moreover, we show that when ϵ is below a computable threshold value, a certain stationary point of the corresponding approximation problem $\min_{x \in \Re^n} F_\epsilon(x)$ is also that of (1). This result plays a crucial role in developing new IRL_1 methods for solving (1) in Section 4.

2.1 Lower bounds for nonzero entries of stationary points of (1)

The first- and second-order stationary points of problem (1) are defined in [17]. We first review these definitions. Then we derive lower bounds for nonzero entries of the stationary points, and hence also of local minimizers of problem (1).

Definition 1 *Let x^* be a vector in $\Re^n$ and $X^* = \mathrm{Diag}(x^*)$. $x^* \in \Re^n$ is a first-order stationary point of* (1) *if*

$$X^* \nabla f(x^*) + \lambda p |x^*|^p = 0. \tag{6}$$

In addition, $x^* \in \Re^n$ is a second-order stationary point of (1) *if*

$$(X^*)^T \nabla^2 f(x^*) X^* + \lambda p(p-1) \mathrm{Diag}(|x^*|^p) \succeq 0. \tag{7}$$

Similar to general unconstrained smooth optimization, we can show that any local minimizer of (1) is also a stationary point that is defined above.

Proposition 2.1 *Let x^* be a local minimizer of* (1) *and $X^* = \mathrm{Diag}(x^*)$. The following statements hold:*

(i) x^* is a first-order stationary point, that is, (6) *holds at x^*.*

(ii) Further, if f is twice continuously differentiable in a neighborhood of x^*, then x^* is a second-order stationary point, that is, (7) holds at x^*.

Proof. (i) Let $\mathcal{B} = \{i : x_i^* \neq 0\}$. Since x^* is a local minimizer of (1), one can observe that x^* is also a local minimizer of

$$\min_{x \in \Re^n} \{f(x) + \lambda \|x_{\mathcal{B}}\|_p^p : x_i = 0,\ i \notin \mathcal{B}\}. \tag{8}$$

Note that the objective function of (8) is differentiable at x^*. The first-order optimality condition of (8) yields

$$\frac{\partial f(x^*)}{\partial x_i} + \lambda p |x_i^*|^{p-1} \operatorname{sgn}(x_i^*) = 0,\ \forall i \in \mathcal{B}.$$

Multiplying by x_i^* both sides of the above equality, we have

$$x_i^* \frac{\partial f(x^*)}{\partial x_i} + \lambda p |x_i^*|^p = 0,\ \forall i \in \mathcal{B}.$$

Since $x_i^* = 0$ for $i \notin \mathcal{B}$, we observe that the above equality also holds for $i \notin \mathcal{B}$. Hence, (6) holds.

(ii) By the assumption, we observe that the objective function of (8) is twice continuously differentiable at x^*. The second-order optimality condition of (8) yields

$$\nabla^2_{\mathcal{B}\mathcal{B}} f(x^*) + \lambda p(p-1) \mathrm{Diag}(|x^*_{\mathcal{B}}|^{p-2}) \ \succeq\ 0,$$

which, together with the fact that $X^* = \mathrm{Diag}(x^*)$ and $x_i^* = 0$ for $i \notin \mathcal{B}$, implies that (7) holds. ∎

Recently, Chen et al. [17] derived some interesting lower bounds for the nonzero entries of local minimizers of problem (1) for the special case where $f(x) = \frac{1}{2}\|Ax - b\|^2$ for some $A \in \Re^{m \times n}$ and $b \in \Re^m$. We next establish similar lower bounds for the nonzero entries of stationary points, and hence also of local minimizers of general problem (1).

Theorem 2.2 *Let x^* be a second-order stationary point of (1) and $\mathcal{B} = \{i : x_i^* \neq 0\}$. Suppose that f is twice continuously differentiable in a neighborhood of x^*. Then the following statement holds:*

$$|x_i^*| \ \geq\ \left(\frac{\lambda p(1-p)}{L_f}\right)^{\frac{1}{2-p}}, \quad \forall i \in \mathcal{B}. \tag{9}$$

Proof. Since f is twice continuously differentiable in a neighborhood of x^* and f has L_f-Lipschitz-continuous gradient in $\Re^n$, we see that $\|\nabla^2 f(x^*)\|_2 \leq L_f$. In addition, since x^* satisfies (7), we have

$$e_i^T[(X^*)^T \nabla^2 f(x^*) X^*] e_i + \lambda p(p-1) e_i^T \mathrm{Diag}(|x^*|^{p-2})] e_i \geq 0,$$

where e_i is the ith coordinate vector. It then follows that for each $i \in \mathcal{B}$,

$$[\nabla^2 f(x^*)]_{ii} + \lambda p(p-1)|x_i^*|^{p-2} \geq 0,$$

which yields

$$|x_i^*| \geq \left(\frac{\lambda p(1-p)}{[\nabla^2 f(x^*)]_{ii}}\right)^{\frac{1}{2-p}} \geq \left(\frac{\lambda p(1-p)}{\|\nabla^2 f(x^*)\|_2}\right)^{\frac{1}{2-p}} \geq \left(\frac{\lambda p(1-p)}{L_f}\right)^{\frac{1}{2-p}}, \quad \forall i \in \mathcal{B}.$$

■

Theorem 2.3 *Let x^* be a first-order stationary point satisfying $F(x^*) \leq F(x^0) + \epsilon$ for some $x^0 \in \Re^n$ and $\epsilon \geq 0$, and let $\underline{f} = \inf_{x \in \Re^n} f(x)$ and $\mathcal{B} = \{i : x_i^* \neq 0\}$. Then the following statement holds:*

$$|x_i^*| \;\geq\; \left(\frac{\lambda p}{\sqrt{2L_f[F(x^0)+\epsilon-\underline{f}]}}\right)^{\frac{1}{1-p}}, \quad \forall i \in \mathcal{B}. \tag{10}$$

Proof. Since f has L_f-Lipschitz-continuous gradient in $\Re^n$, it is well known that

$$f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{L_f}{2}\|y-x\|_2^2, \quad \forall x, y \in \Re^n.$$

Letting $x = x^*$ and $y = x^* - \nabla f(x^*)/L_f$, we obtain that

$$f(x^* - \nabla f(x^*)/L_f) \;\leq\; f(x^*) - \frac{1}{2L_f}\|\nabla f(x^*)\|_2^2. \tag{11}$$

Note that

$$f(x^* - \nabla f(x^*)/L_f) \geq \inf_{x \in \Re^n} f(x) = \underline{f}, \qquad f(x^*) \leq F(x^*) \leq F(x^0) + \epsilon.$$

Using these relations and (11), we have

$$\|\nabla f(x^*)\|_2 \;\leq\; \sqrt{2L_f[f(x^*) - f(x^* - \nabla f(x^*)/L_f)]} \;\leq\; \sqrt{2L_f[F(x^0)+\epsilon-\underline{f}]}. \tag{12}$$

Since x^* satisfies (6), we obtain that for every $i \in \mathcal{B}$,

$$|x_i^*| = \left(\frac{1}{\lambda p}\left|\frac{\partial f(x^*)}{\partial x_i}\right|\right)^{\frac{1}{p-1}} \;\geq\; \left(\frac{\|\nabla f(x^*)\|_2}{\lambda p}\right)^{\frac{1}{p-1}},$$

which together with (12) yields

$$|x_i^*| \;\geq\; \left(\frac{\lambda p}{\sqrt{2L_f[F(x^0)+\epsilon-\underline{f}]}}\right)^{\frac{1}{1-p}}, \quad \forall i \in \mathcal{B}.$$

■

2.2 Locally Lipschitz continuous approximation to (1)

It is known that for $p \in (0,1)$, the function $\|x\|_p^p$ is not locally Lipschitz continuous at some points in $\Re^n$ and the Clarke subdifferential does not exist there (see, for example, [17]). This brings a great deal of challenge for designing algorithms for solving problem (1). In this subsection we propose a nonsmooth but Lipschitz continuous ϵ-approximation to $\|x\|_p^p$ for every $\epsilon > 0$. As a consequence, we obtain a nonsmooth but locally Lipschitz continuous ϵ-approximation $F_\epsilon(x)$ to $F(x)$. Furthermore, we show that when ϵ is below a computable threshold value, a certain stationary point of the corresponding approximation problem $\min_{x\in\Re^n} F_\epsilon(x)$ is also that of (1).

Lemma 2.4 *Let $u > 0$ be arbitrarily given, and let q be such that*

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{13}$$

Define

$$h_u(t) := \min_{0\le s\le u} p\left(|t|s - \frac{s^q}{q}\right), \quad \forall t \in \Re. \tag{14}$$

Then the following statements hold:

(i) $0 \ \le\ h_u(t) - |t|^p \ \le\ u^q$ *for every* $t \in \Re$.

(ii) h_u *is pu-Lipschitz continuous in* $(-\infty,\infty)$*, i.e.,*

$$|h_u(t_1) - h_u(t_2)| \ \le pu|t_1 - t_2|, \quad \forall t_1, t_2 \in \Re.$$

(iii) The Clarke subdifferential of h_u*, denoted by* ∂h_u*, exists everywhere, and it is given by*

$$\partial h_u(t) = \begin{cases} p|t|^{p-1}\operatorname{sgn}(t) & \text{if } |t| > u^{q-1}, \\ pu\operatorname{sgn}(t) & \text{if } |t| \le u^{q-1}. \end{cases} \tag{15}$$

Proof. (i) Let $g_t(s) = p(|t|s - s^q/q)$ for $s > 0$. Since $p \in (0,1)$, we observe from (13) that $q < 0$. It then implies that $g_t(s) \to \infty$ as $s \downarrow 0$. This together with the continuity of g_t implies that $h_u(t)$ is well-defined for all $t \in \Re$. In addition, it is easy to show that $g_t(\cdot)$ is convex in $(0,\infty)$, and moreover, $\inf\limits_{s>0} g_t(s) \ = \ |t|^p$. Hence, we have

$$h_u(t) \ = \ \min_{0\le s\le u} g_t(s) \ \ge \ \inf_{s>0} g_t(s) \ = \ |t|^p, \quad \forall t \in \Re.$$

We next show that $h_u(t) - |t|^p \ \le \ u^q$ by dividing its proof into two cases.

1) Assume that $|t| > u^{q-1}$. Then, the optimal value of (14) is achieved at $s^* = |t|^{\frac{1}{q-1}}$ and hence,

$$h_u(t) \ = \ p\left(|t|s^* - \frac{(s^*)^q}{q}\right) \ = \ |t|^p.$$

2) Assume that $|t| \le u^{q-1}$. It can be shown that the optimal value of (14) is achieved at $s^* = u$. Using this result and the relation $|t| \le u^{q-1}$, we obtain that

$$h_u(t) \;=\; p\left(|t|u - \frac{u^q}{q}\right) \;\le\; p\left(u^{q-1}u - \frac{u^q}{q}\right) \;=\; u^q,$$

which implies that $h_u(t) - |t|^p \;\le\; h_u(t) \;\le\; u^q$.

Combining the above two cases, we conclude that statement (i) holds.

(ii) Let $\phi : [0, \infty) \to \Re$ be defined as follows:

$$\phi(t) = \begin{cases} t^p & \text{if } t > u^{q-1}, \\ p(tu - u^q/q) & \text{if } 0 \le t \le u^{q-1}. \end{cases}$$

It is not hard to see that

$$\phi'(t) = \begin{cases} pt^{p-1} & \text{if } t > u^{q-1}, \\ pu & \text{if } 0 \le t \le u^{q-1}. \end{cases} \tag{16}$$

Hence, $0 \le \phi'(t) \le pu$ for every $t \in [0, \infty)$, which implies that ϕ is pu-Lipschitz continuous on $[0, \infty)$. In addition, one can observe from the proof of (i) that $h_u(t) = \phi(|t|)$ for all t. By the chain rule, we easily conclude that h_u is pu-Lipschitz continuous in $(-\infty, \infty)$.

(iii) Since h_u is Lipschitz continuous everywhere, it follows from Theorem 2.5.1 of [19] that

$$\partial h_u(t) = \operatorname{cov}\left\{ \lim_{t_k \in D \to t} h_u'(t_k) \right\}, \tag{17}$$

where cov denotes convex hull and D is the set of points at which h_u is differentiable. Recall that $h_u(t) = \phi(|t|)$ for all t. Hence, $h_u'(t) = \phi'(|t|)\,\operatorname{sgn}(t)$ for every $t \neq 0$. Using this relation, (16) and (17), we immediately see that statement (iii) holds. ∎

Corollary 2.5 *Let $u > 0$ be arbitrarily given, and let $h(x) = \sum_{i=1}^n h_u(x_i)$ for every $x \in \Re^n$, where h_u is defined in* (14). *Then the following statements hold:*

(i) $0 \;\le\; h(x) - \|x\|_p^p \;\le\; nu^q$ *for every* $x \in \Re^n$.

(ii) h *is* $\sqrt{n}pu$*-Lipschitz continuous in* $\Re^n$*, i.e.,*

$$\|h(x) - h(y)\|_2 \;\le \sqrt{n}pu\|x - y\|_2, \qquad \forall x, y.$$

(iii) The Clark subdifferential of h *exists at every* $x \in \Re^n$.

We are now ready to propose a nonsmooth but locally Lipschitz continuous ϵ-approximation to $F(x)$.

Proposition 2.6 *Let $\epsilon > 0$ be arbitrarily given. Define*

$$F_\epsilon(x) := f(x) + \lambda \sum_{i=1}^{n} h_{u_\epsilon}(x_i), \tag{18}$$

where

$$h_{u_\epsilon}(t) := \min_{0 \le s \le u_\epsilon} p\left(|t|s - \frac{s^q}{q}\right), \qquad u_\epsilon := \left(\frac{\epsilon}{\lambda n}\right)^{\frac{1}{q}}. \tag{19}$$

Then the following statements hold:

(i) $0 \ \le \ F_\epsilon(x) - F(x) \ \le \ \epsilon$ *for every* $x \in \Re^n$.

(ii) F_ϵ *is locally Lipschitz continuous in* $\Re^n$. *Furthermore, if* f *is Lipschitz continuous, so is* F_ϵ.

(iii) The Clark subdifferential of F_ϵ *exists at every* $x \in \Re^n$.

Proof. Using the definitions of F_ϵ and F, we have $F_\epsilon(x) - F(x) = \lambda(\sum_{i=1}^n h_{u_\epsilon}(x_i) - \|x\|_p^p)$, which, together with Corollary 2.5 (i) with $u = u_\epsilon$, implies that statement (i) holds. Since f is differentiable in $\Re^n$, it is known that f is locally Lipschitz continuous. In addition, we know from Corollary 2.5 (ii) that $\sum_{i=1}^n h_{u_\epsilon}(x_i)$ is Lipschitz continuous in $\Re^n$. These facts imply that statement (ii) holds. Statement (iii) immediately follows from Corollary 2.5 (iii). ∎

From Proposition 2.6, we know that F_ϵ is a nice ϵ-approximation to F. It is very natural to find an approximate solution of (1) by solving the corresponding ϵ-approximation problem

$$\min_{x \in \Re^n} F_\epsilon(x), \tag{20}$$

where F_ϵ is defined in (18). Strikingly, we can show that when ϵ is below a computable threshold value, a certain stationary point of problem (20) is also that of (1).

Theorem 2.7 *Let $x^0 \in \Re^n$ be an arbitrary point, and let ϵ be such that*

$$0 < \epsilon < n\lambda \left[\frac{\sqrt{2L_f[F(x^0) + \epsilon - \underline{f}]}}{\lambda p}\right]^q, \tag{21}$$

where $\underline{f} = \inf_{x \in \Re^n} f(x)$. *Suppose that* x^* *is a first-order stationary point of* (20) *such that that* $F_\epsilon(x^*) \le F_\epsilon(x^0)$. *Then,* x^* *is also a first-order stationary point of* (1), *i.e.,* (6) *holds at* x^*. *Moreover, the nonzero entries of* x^* *satisfy the first-order lower bound* (10).

Proof. Let $\mathcal{B} = \{i : x_i^* \neq 0\}$. Since x^* is a first-order stationary point of (20), we have $0 \in \partial F_\epsilon(x^*)$. Hence, it follows that

$$\frac{\partial f(x^*)}{\partial x_i} + \lambda \partial h_{u_\epsilon}(x_i^*) \ = \ 0, \quad \forall i \in \mathcal{B}. \tag{22}$$

In addition, we notice that

$$f(x^*) \ \le \ F(x^*) \ \le \ F_\epsilon(x^*) \ \le \ F_\epsilon(x^0) \ \le \ F(x^0) + \epsilon. \tag{23}$$

This relation together with (22) and (12) implies that

$$|\partial h_{u_\epsilon}(x_i^*)| \ = \ \frac{1}{\lambda}\left|\frac{\partial f(x^*)}{\partial x_i}\right| \ \le \ \frac{1}{\lambda}\|\nabla f(x^*)\|_2 \ \le \ \frac{\sqrt{2L_f[F(x^0)+\epsilon-\underline{f}]}}{\lambda}, \quad \forall i \in \mathcal{B}. \tag{24}$$

We now claim that $|x_i^*| > {u_\epsilon}^{q-1}$ for all $i \in \mathcal{B}$, where u_ϵ is defined in (19). Suppose for contradiction that there exists some $i \in \mathcal{B}$ such that $0 < |x_i^*| \le {u_\epsilon}^{q-1}$. It then follows from (15) that $|\partial h_{u_\epsilon}(x_i^*)| = pu_\epsilon$. Using this relation, (21) and the definition of u_ϵ, we obtain that

$$|\partial h_{u_\epsilon}(x_i^*)| \ = \ pu_\epsilon = p\left(\frac{\epsilon}{\lambda n}\right)^{1/q} \ > \ \frac{\sqrt{2L_f[F(x^0)+\epsilon-\underline{f}]}}{\lambda},$$

which contradicts (24). Therefore, $|x_i^*| > {u_\epsilon}^{q-1}$ for all $i \in \mathcal{B}$. Using this fact and (15), we see that $\partial h_{u_\epsilon}(x_i^*) \ = \ p|x_i^*|^{p-1}\operatorname{sgn}(x_i^*)$ for every $i \in \mathcal{B}$. Substituting it into (22), we obtain that

$$\frac{\partial f(x^*)}{\partial x_i} + \lambda p|x_i^*|^{p-1}\operatorname{sgn}(x_i^*) = 0, \quad \forall i \in \mathcal{B}.$$

Multiplying by x_i^* both sides of the above equality, we have

$$x_i^*\frac{\partial f(x^*)}{\partial x_i} + \lambda p|x_i^*|^p = 0, \quad \forall i \in \mathcal{B}.$$

Since $x_i^* = 0$ for $i \notin \mathcal{B}$, we observe that the above equality also holds for $i \notin \mathcal{B}$. Hence, (6) holds. In addition, recall from (23) that $F(x^*) \ \le \ F(x^0)+\epsilon$. Using this relation and Theorem 2.3, we immediately see that the second part of this theorem also holds. ∎

Corollary 2.8 *Let $x^0 \in \Re^n$ be an arbitrary point, and let ϵ be such that* (21) *holds. Suppose that x^* is a local minimizer of* (20) *such that $F_\epsilon(x^*) \le F_\epsilon(x^0)$. Then the following statements hold:*

i) x^* is a first-order stationary point of (1)*, i.e.,* (6) *holds at x^*. Moreover, the nonzero entries of x^* satisfy the first-order lower bound* (10)*.*

ii) Suppose further that f is twice continuously differentiable in a neighborhood of x^*. Then, x^* is a second-order stationary point of (1)*, i.e.,* (7) *holds at x^*. Moreover, the nonzero entries of x^* satisfy the second-order lower bound* (9)*.*

Proof. (i) Since x^* is a local minimizer of (20), we know that x^* is a stationary point of (20). Statement (i) then immediately follows from Theorem 2.7.

(ii) Let $\mathcal{B} = \{i | x_i^* \neq 0\}$. Since x^* is a local minimizer of (20), we observe that x^* is also a local minimizer of

$$\min_{x \in \Re^n} \left\{ f(x) + \lambda \sum_{i \in \mathcal{B}} h_{u_\epsilon}(x_i) : x_i = 0,\ i \notin \mathcal{B} \right\}. \tag{25}$$

Notice that x^* is a first-order stationary point of (20). In addition, $F(x^*) \le F(x^0) + \epsilon$ and ϵ satisfies (21). Using the same arguments as in the proof of Theorem 2.7, we have $|x_i^*| > {u_\epsilon}^{q-1}$ for all $i \in \mathcal{B}$. Recall from the proof of Lemma 2.4 (i) that $h_{u_\epsilon}(t) = |t|^p$ if $|t| > {u_\epsilon}^{q-1}$. Hence, $\sum_{i \in \mathcal{B}} h_{u_\epsilon}(x_i) = \sum_{i \in \mathcal{B}} |x_i|^p$ for all x in a neighborhood of x^*. This, together with the fact that x^* is a local minimizer of (25), implies that x^* is also a local minimizer of (8). The rest of the proof is similar to that of Proposition 2.1 and Theorem 2.2. ∎

3 A unified analysis for some existing iterative reweighted minimization methods

Recently two types of IRL_1 and IRL_2 methods have been proposed in the literature [21, 20, 25, 18] for solving problem (4) or (5). In this section we extend these methods to solve (1) and also propose a variant of them in which each subproblem has a closed form solution. Moreover, we provide a unified convergence analysis for them.

3.1 The first type of IRL_α methods and its variant for (1)

In this subsection we consider the iterative reweighted minimization methods proposed in [25, 18] for solving problem (5), which apply an IRL_1 or IRL_2 method to solve a sequence of problems $\min_{x \in \Re^n} Q_{1,\epsilon^k}(x)$ or $\min_{x \in \Re^n} Q_{2,\epsilon^k}(x)$, where $\{\epsilon^k\}$ is a sequence of positive vectors approaching zero as $k \to \infty$ and

$$Q_{\alpha,\epsilon}(x) := \frac{1}{2}\|Ax - b\|_2^2 + \lambda \sum_{i=1}^n (|x_i|^\alpha + \epsilon_i)^{\frac{p}{\alpha}}. \tag{26}$$

In what follows, we extend the above methods to solve (1) and also propose a variant of them in which each subproblem has a closed form solution. Moreover, we provide a unified convergence analysis for them. Our key observation is that problem

$$\min_{x \in \Re^n} \{F_{\alpha,\epsilon}(x) := f(x) + \lambda \sum_{i=1}^n (|x_i|^\alpha + \epsilon_i)^{\frac{p}{\alpha}}\} \tag{27}$$

for $\alpha \ge 1$ and $\epsilon > 0$ can be suitably solved by an iterative reweighted l_α (IRL_α) method. Problem (1) can then be solved by applying the IRL_α method to a sequence of problems (27) with $\epsilon = \epsilon^k \to 0$ as $k \to \infty$.

We start by presenting an IRL$_\alpha$ method for solving problem (27) as follows.

An IRL$_\alpha$ **minimization method for** (27)**:**

Choose an arbitrary x^0. Set $k = 0$.

1) Solve the weighted l_α minimization problem

$$x^{k+1} \in \operatorname{Arg\,min} \left\{ f(x) + \frac{\lambda p}{\alpha} \sum_{i=1}^{n} s_i^k |x_i|^\alpha \right\}, \tag{28}$$

where $s_i^k = (|x_i^k|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}$ for all i.

2) Set $k \leftarrow k+1$ and go to step 1).

end

We next show that any accumulation point of $\{x^k\}$ generated above is a first-order stationary point of (27).

Theorem 3.1 *Let the sequence* $\{x^k\}$ *be generated by the above* IRL$_\alpha$ *minimization method. Suppose that* x^* *is an accumulation point of* $\{x^k\}$*. Then* x^* *is a first-order stationary point of* (27)*.*

Proof. Let q be such that

$$\frac{\alpha}{p} + \frac{1}{q} = 1. \tag{29}$$

It is not hard to show that for any $\delta > 0$,

$$(|t|^\alpha + \delta)^{\frac{p}{\alpha}} = \frac{p}{\alpha} \min_{s \geq 0} \left\{ (|t|^\alpha + \delta)s - \frac{s^q}{q} \right\}, \qquad \forall t \in \Re, \tag{30}$$

and moreover, the minimum is achieved at $s = (|t|^\alpha + \delta)^{\frac{1}{q-1}}$. Using this result, the definition of s^k, and (29), one can observe that for $k \geq 0$,

$$s^k = \arg\min_{s \geq 0} G_{\alpha,\epsilon}(x^k, s), \qquad x^{k+1} \in \operatorname{Arg\,min}_x G_{\alpha,\epsilon}(x, s^k), \tag{31}$$

where

$$G_{\alpha,\epsilon}(x,s) = f(x) + \frac{\lambda p}{\alpha} \sum_{i=1}^{n} \left[(|x_i|^\alpha + \epsilon_i)s_i - \frac{s_i^q}{q} \right]. \tag{32}$$

In addition, we see that $F_{\alpha,\epsilon}(x^k) = G_{\alpha,\epsilon}(x^k, s^k)$. It then follows that

$$F_{\alpha,\epsilon}(x^{k+1}) = G_{\alpha,\epsilon}(x^{k+1}, s^{k+1}) \leq G_{\alpha,\epsilon}(x^{k+1}, s^k) \leq G_{\alpha,\epsilon}(x^k, s^k) = F_{\alpha,\epsilon}(x^k). \tag{33}$$

Hence, $\{F_{\alpha,\epsilon}(x^k)\}$ is non-increasing. Since x^* is an accumulation point of $\{x^k\}$, there exists a subsequence K such that $\{x^k\}_K \to x^*$. By the continuity of $F_{\alpha,\epsilon}$, we have $\{F_{\alpha,\epsilon}(x^k)\}_K \to$

$F_{\alpha,\epsilon}(x^*)$, which together with the monotonicity of $F_{\alpha,\epsilon}(x^k)$ implies that $F_{\alpha,\epsilon}(x^k) \to F_{\alpha,\epsilon}(x^*)$. In addition, by the definition of s^k, we have $\{s^k\}_K \to s^*$, where $s_i^* = (|x_i^*|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}$ for all i. Also, we observe that $F_{\alpha,\epsilon}(x^*) = G_{\alpha,\epsilon}(x^*, s^*)$. Using (33) and $F_{\alpha,\epsilon}(x^k) \to F_{\alpha,\epsilon}(x^*)$, we see that $G_{\alpha,\epsilon}(x^{k+1}, s^k) \to F_{\alpha,\epsilon}(x^*) = G_{\alpha,\epsilon}(x^*, s^*)$. Further, it follows from (31) that

$$G_{\alpha,\epsilon}(x, s^k) \;\geq\; G_{\alpha,\epsilon}(x^{k+1}, s^k) \quad \forall x \in \Re^n.$$

Upon taking limits on both sides of this inequality as $k \in K \to \infty$, we have

$$G_{\alpha,\epsilon}(x, s^*) \;\geq\; G_{\alpha,\epsilon}(x^*, s^*) \quad \forall x \in \Re^n,$$

that is, $x^* \in \operatorname{Arg}\min\limits_{x\in\Re^n} G_{\alpha,\epsilon}(x, s^*)$, which, together with the first-order optimality condition and the definition of x^*, yields

$$0 \in \frac{\partial f(x^*)}{\partial x_i} + \lambda p(|x_i^*|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}|x_i^*|^{\alpha-1}\operatorname{sgn}(x_i^*), \quad \forall i. \tag{34}$$

Hence, x^* is a stationary point of (27). ■

The above IRL$_\alpha$ method needs to solve a sequence of reweighted l_α minimization problems (47) whose solution may not be cheaply computable. We next propose a variant of this method in which each subproblem is much simpler and has a closed form solution for some α's (e.g., $\alpha = 1$ or 2).

A variant of IRL$_\alpha$ minimization method for (27):

Let $0 < L_{\min} < L_{\max}$, $\tau > 1$ and $c > 0$ be given. Choose an arbitrary x^0 and set $k = 0$.

1) Choose $L_k^0 \in [L_{\min}, L_{\max}]$ arbitrarily. Set $L_k = L_k^0$.

   1a) Solve the weighted l_α minimization problem

$$x^{k+1} \in \operatorname{Arg}\min_x \left\{ f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{L_k}{2}\|x - x^k\|_2^2 + \frac{\lambda p}{\alpha}\sum_{i=1}^n s_i^k |x_i|^\alpha \right\}, \tag{35}$$

   where $s_i^k = (|x_i^k|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}$ for all i.

   1b) If

$$F_{\alpha,\epsilon}(x^k) - F_{\alpha,\epsilon}(x^{k+1}) \geq \frac{c}{2}\|x^{k+1} - x^k\|_2^2 \tag{36}$$

   is satisfied, where $F_{\alpha,\epsilon}$ is given in (27), then go to step 2).

   1c) Set $L_k \leftarrow \tau L_k$ and go to step 1a).

2) Set $k \leftarrow k + 1$ and go to step 1).

end

We first show that for each outer iteration, the number of its inner iterations is finite.

Theorem 3.2 *For each $k \geq 0$, the inner termination criterion* (36) *is satisfied after at most* $\left\lceil \frac{\log(L_f+c)-\log(2L_{\min})}{\log\tau} + 2 \right\rceil$ *inner iterations.*

Proof. Let $\bar{L}_k$ denote the final value of L_k at the kth outer iteration. Since the objective function of (35) is strongly convex with modulus L_k, we have

$$f(x^k) + \frac{\lambda p}{\alpha}\sum_{i=1}^n s_i^k |x_i^k|^\alpha \ \geq \ f(x^k) + \nabla f(x^k)^T(x^{k+1}-x^k) + \frac{\lambda p}{\alpha}\sum_{i=1}^n s_i^k |x_i^{k+1}|^\alpha + L_k\|x^{k+1}-x^k\|_2^2.$$

Recall that ∇f is L_f-Lipschitz continuous. We then have

$$f(x^{k+1}) \ \leq \ f(x^k) + \nabla f(x^k)^T(x^{k+1}-x^k) + \frac{L_f}{2}\|x^{k+1}-x^k\|_2^2. \tag{37}$$

Combining the above two inequalities, we obtain that

$$f(x^k) + \frac{\lambda p}{\alpha}\sum_{i=1}^n s_i^k |x_i^k|^\alpha \ \geq \ f(x^{k+1}) + \frac{\lambda p}{\alpha}\sum_{i=1}^n s_i^k |x_i^{k+1}|^\alpha + (L_k - \frac{L_f}{2})\|x^{k+1}-x^k\|_2^2,$$

which together with (32) yields

$$G_{\alpha,\epsilon}(x^k, s^k) \ \geq \ G_{\alpha,\epsilon}(x^{k+1}, s^k) + (L_k - \frac{L_f}{2})\|x^{k+1}-x^k\|_2^2.$$

Recall that $F_{\alpha,\epsilon}(x^k) = G_{\alpha,\epsilon}(x^k, s^k)$. In addition, it follows from (30) that $F_{\alpha,\epsilon}(x) = \min\limits_{s\geq 0} G_{\alpha,\epsilon}(x,s)$. Using these relations and the above inequality, we obtain that

$$\begin{aligned} F_{\alpha,\epsilon}(x^{k+1}) &= G_{\alpha,\epsilon}(x^{k+1}, s^{k+1}) \ \leq \ G_{\alpha,\epsilon}(x^{k+1}, s^k) \ \leq \ G_{\alpha,\epsilon}(x^k, s^k) - (L_k - \tfrac{L_f}{2})\|x^{k+1}-x^k\|_2^2 \\ &= F_{\alpha,\epsilon}(x^k) - (L_k - \tfrac{L_f}{2})\|x^{k+1}-x^k\|_2^2. \end{aligned}$$

Hence, (36) holds whenever $L_k \geq (L_f + c)/2$, which together with the definition of $\bar{L}_k$ implies that $\bar{L}_k/\tau < (L_f+c)/2$, that is, $\bar{L}_k < \tau(L_f+c)/2$. Let n_k denote the number of inner iterations for the kth outer iteration. Then, we have

$$L_{\min}\tau^{n_k-1} \leq L_k^0 \tau^{n_k-1} = \bar{L}_k < \tau(L_f+c)/2.$$

Hence, $n_k \leq \left\lceil \frac{\log(L_f+c)-\log(2L_{\min})}{\log\tau} + 2 \right\rceil$ and the conclusion holds. ■

We next establish that any accumulation point of the sequence $\{x^k\}$ generated above is a first-order stationary point of problem (27).

Theorem 3.3 *Let the sequence $\{x^k\}$ be generated by the above variant of* IRL_α *method. Suppose that x^* is an accumulation point of $\{x^k\}$. Then x^* is a first-order stationary point of* (27).

Proof. It follows from (36) that $\{F_{\alpha,\epsilon}(x^k)\}$ is non-increasing. Since x^* is an accumulation point of $\{x^k\}$, there exists a subsequence K such that $\{x^k\}_K \to x^*$. By the continuity of $F_{\alpha,\epsilon}$, we have $\{F_{\alpha,\epsilon}(x^k)\}_K \to F_{\alpha,\epsilon}(x^*)$, which together with the monotonicity of $\{F_{\alpha,\epsilon}(x^k)\}$ implies that $F_{\alpha,\epsilon}(x^k) \to F_{\alpha,\epsilon}(x^*)$. Using this result and (36), we can conclude that $\|x^{k+1} - x^k\| \to 0$. Let $\bar{L}_k$ denote the final value of L_k at the kth outer iteration. From the proof of Theorem 3.3, we know that $\bar{L}_k \in [L_{\min}, \tau(L_f + c)/2)$. The first-order optimality condition of (35) with $L_k = \bar{L}_k$ yields

$$0 \in \frac{\partial f(x^k)}{\partial x_i} + \bar{L}_k(x_i^{k+1} - x_i^k) + \lambda p s_i^k |x_i^{k+1}|^{\alpha-1} \operatorname{sgn}(x_i^{k+1}) = 0, \quad \forall i.$$

Upon taking limits on both sides of the above equality as $k \in K \to \infty$, we have

$$0 \in \frac{\partial f(x^*)}{\partial x_i} + \lambda p s_i^* |x_i^*|^{\alpha-1} \operatorname{sgn}(x_i^*), \quad \forall i,$$

where $s_i^* = (|x_i^*| + \epsilon_i)^{\frac{p}{\alpha}-1}$ for all i. Hence, x^* is a first-order stationary point of (27). ■

Corollary 3.4 *Let $\delta > 0$ be arbitrarily given, and let the sequence $\{x^k\}$ be generated by the above* IRL_α *method or its variant. Suppose that $\{x^k\}$ has at least one accumulation point. Then, there exists some x^k such that*

$$\|X^k \nabla f(x^k) + \lambda p |X^k|^\alpha (|x^k|^\alpha + \epsilon)^{\frac{p}{\alpha}-1}\| \;\leq\; \delta,$$

where $X^k = \mathrm{Diag}(x^k)$ and $|X^k|^\alpha = \mathrm{Diag}(|x^k|^\alpha)$.

Proof. Let x^* be an arbitrary accumulation point of $\{x^k\}$. It follows from Theorem 3.1 that x^* satisfies (34). Multiplying by x_i^* both sides of (34), we have

$$x_i^* \frac{\partial f(x^*)}{\partial x_i} + \lambda p (|x_i^*|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1} |x_i^*|^\alpha \;=\; 0 \quad \forall i,$$

which, together with the continuity of $\nabla f(x)$ and $|x|^\alpha$, implies that the conclusion holds. ■

We are now ready to present the first type of IRL_α methods and its variant for solving problem (1) in which each subproblem is in the form of (27) and solved by the IRL_α or its variant described above. The IRL_1 and IRL_2 methods proposed in [25, 18] can be viewed as the special cases of the following general IRL_α method (but not its variant) with $f(x) = \frac{1}{2}\|Ax-b\|_2^2$ and $\alpha = 1$ or 2.

The first type of IRL_α minimization methods and its variant for (1):

Let $\{\delta_k\}$ and $\{\epsilon^k\}$ be a sequence of positive scalars and vectors, respectively. Set $k = 0$.

1) Apply the IRL_α method or its variant to problem (27) with $\epsilon = \epsilon^k$ for finding x^k satisfying

$$\|X^k \nabla f(x^k) + \lambda p |X^k|^\alpha (|x^k|^\alpha + \epsilon^k)^{\frac{p}{\alpha}-1}\| \;\le\; \delta_k, \tag{38}$$

where $X^k = \text{Diag}(x^k)$ and $|X^k|^\alpha = \text{Diag}(|x^k|^\alpha)$.

2) Set $k \leftarrow k+1$ and go to step 1).

end

The convergence of the above IRL_α method and its variant is established as follows.

Theorem 3.5 *Let $\{\delta_k\}$ and $\{\epsilon^k\}$ be a sequence of positive scalars and vectors such that $\{\delta_k\} \to 0$ and $\{\epsilon^k\} \to 0$, respectively. Suppose that $\{x^k\}$ is a sequence of vectors generated above satisfying* (38)*, and that x^* is an accumulation point of $\{x^k\}$. Then x^* is a first-order stationary point of* (1)*, i.e.,* (6) *holds at x^*.*

Proof. Let $\mathcal{B} = \{i : x_i^* \neq 0\}$. It follows from (38) that

$$\left| x_i^k \frac{\partial f(x^k)}{\partial x_i} + \lambda p |x_i^k|^\alpha (|x_i^k|^\alpha + \epsilon_i^k)^{\frac{p}{\alpha}-1} \right| \le \delta_k \quad \forall i \in \mathcal{B}. \tag{39}$$

Since x^* is an accumulation point of $\{x^k\}$, there exists a subsequence K such that $\{x^k\}_K \to x^*$. Upon taking limits on both sides of (39) as $k \in K \to \infty$, we obtain that

$$x_i^* \frac{\partial f(x^*)}{\partial x_i} + \lambda p |x_i^*|^p \;=\; 0 \quad \forall i \in \mathcal{B}.$$

Since $x_i^* = 0$ for $i \notin \mathcal{B}$, we observe that the above equality also holds for $i \notin \mathcal{B}$. Hence, x^* satisfies (6) and it is a first-order stationary point of (1). ∎

3.2 The second type of IRL_α methods and its variant for (1)

In this subsection we are interested in the IRL_1 and IRL_2 methods proposed in [21, 20] for solving problem (4). Given $\{\epsilon^k\} \subset \Re^n_{++} \to 0$ as $k \to \infty$, these methods solve a sequence of problems $\min_{x \in \Re^n} Q_{1,\epsilon^k}(x)$ or $\min_{x\in\Re^n} Q_{2,\epsilon^k}(x))$ extremely "roughly" by executing IRL_1 or IRL_2 method only for one iteration for each ϵ^k, where $Q_{\alpha,\epsilon}$ is defined in (26).

We next extend the above methods to solve (1) and also propose a variant of them in which each subproblem has a closed form solution. Moreover, we provide a unified convergence analysis for them. We start by presenting the second type of IRL_α methods for solving (1) as follows. They evidently become an IRL_1 or IRL_2 method when $\alpha = 1$ or 2.

The second type of IRL_α minimization method for (1):

Let $\{\epsilon^k\}$ be a sequence of positive vectors in $\Re^n$. Choose an arbitrary x^0. Set $k = 0$.

1) Solve the weighted l_α minimization problem

$$x^{k+1} \in \operatorname{Arg}\min \left\{ f(x) + \frac{\lambda p}{\alpha} \sum_{i=1}^{n} s_i^k |x_i|^\alpha \right\}, \tag{40}$$

where $s_i^k = (|x_i^k|^\alpha + \epsilon_i^k)^{\frac{p}{\alpha}-1}$ for all i.

2) Set $k \leftarrow k+1$ and go to step 1).

end

We next establish that any accumulation point of $\{x^k\}$ is a stationary point of (1).

Theorem 3.6 *Suppose that $\{\epsilon^k\}$ is a sequence of non-increasing positive vectors in $\Re^n$ and $\epsilon^k \to 0$ as $k \to \infty$. Let the sequence $\{x^k\}$ be generated by the above* IRL_α *method. Suppose that x^* is an accumulation point of $\{x^k\}$. Then, x^* is a stationary point of* (1).

Proof. Let $G_{\alpha,\epsilon}(\cdot,\cdot)$ be defined in (32). By the definition of x^{k+1}, one can observe that $G_{\alpha,\epsilon^k}(x^{k+1}, s^k) \le G_{\alpha,\epsilon^k}(x^k, s^k)$. Also, by the definition of s^{k+1} and a similar argument as in the proof of Theorem 3.1, we have $G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^{k+1}) = \inf\limits_{s\ge 0} G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s)$. Hence, $G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^{k+1}) \le G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^k)$. In addiiton, since $s^k > 0$ and $\{\epsilon^k\}$ is componentwise non-increasing, we observe that $G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^k) \le G_{\alpha,\epsilon^k}(x^{k+1}, s^k)$. Combining these inequalities, we have

$$G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^{k+1}) \le G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^k) \le G_{\alpha,\epsilon^k}(x^{k+1}, s^k) \le G_{\alpha,\epsilon^k}(x^k, s^k), \quad \forall k \ge 0. \tag{41}$$

Hence, $\{G_{\alpha,\epsilon^k}(x^k, s^k)\}$ is non-increasing. Since x^* is an accumulation point of $\{x^k\}$, there exists a subsequence K such that $\{x^k\}_K \to x^*$. By the definition of s^k, one can verify that $G_{\alpha,\epsilon^k}(x^k, s^k) = f(x^k) + \lambda \sum_{i=1}^n (|x_i^k|^\alpha + \epsilon_i^k)^{\frac{p}{\alpha}}$. It then follows from $\{x^k\}_K \to x^*$ and $\epsilon^k \to 0$ that $\{G_{\alpha,\epsilon^k}(x^k, s^k)\}_K \to f(x^*) + \lambda \|x^*\|_p^p$. This together with the monotonicity of $\{G_{\alpha,\epsilon^k}(x^k, s^k)\}$ implies that $G_{\alpha,\epsilon^k}(x^k, s^k) \to f(x^*) + \lambda \|x^*\|_p^p$. Using this relation and (41), we further have

$$G_{\alpha,\epsilon^k}(x^{k+1}, s^k) \to f(x^*) + \lambda \|x^*\|_p^p. \tag{42}$$

Let $\mathcal{B} = \{i : x_i^* \neq 0\}$ and $\bar{\mathcal{B}}$ be its complement in $\{1, \ldots, n\}$. We claim that

$$x^* \in \operatorname{Arg}\min_{x_{\bar{\mathcal{B}}}=0} \left\{ f(x) + \frac{\lambda p}{\alpha} \sum_{i\in\mathcal{B}} |x_i|^\alpha |x_i^*|^{p-\alpha} \right\}. \tag{43}$$

Indeed, using the definition of s^k, we see that $\{s_i^k\}_K \to |x_i^*|^{p-\alpha}$, $\forall i \in \mathcal{B}$. Further, due to $s^k > 0$ and $q < 0$, we observe that

$$0 \le \frac{p}{\alpha} \sum_{i\in\bar{\mathcal{B}}} \left[\epsilon_i^k s_i^k - \frac{(s_i^k)^q}{q} \right] \le \frac{p}{\alpha} \sum_{i\in\bar{\mathcal{B}}} \left[(|x_i^k|^\alpha + \epsilon_i^k) s_i^k - \frac{(s_i^k)^q}{q} \right] = \sum_{i\in\bar{\mathcal{B}}} (|x_i^k|^\alpha + \epsilon_i^k)^{\frac{p}{\alpha}},$$

which, together with $\epsilon^k \to 0$ and $\{x_i^k\}_K \to 0$ for $i \in \bar{\mathcal{B}}$, implies that

$$\lim_{k \in K \to \infty} \sum_{i \in \bar{\mathcal{B}}} \left[\epsilon_i^k s_i^k - \frac{(s_i^k)^q}{q} \right] = 0. \tag{44}$$

In addition, by the definition of x^{k+1}, we know that $G_{\alpha,\epsilon^k}(x, s^k) \geq G_{\alpha,\epsilon^k}(x^{k+1}, s^k)$. Then for every $x \in \Re^n$ such that $x_{\bar{\mathcal{B}}} = 0$, we have

$$f(x) + \frac{\lambda p}{\alpha} \sum_{i \in \mathcal{B}} \left[(|x_i|^\alpha + \epsilon_i^k) s_i^k - \frac{(s_i^k)^q}{q} \right] + \frac{\lambda p}{\alpha} \sum_{i \in \bar{\mathcal{B}}} \left[\epsilon_i^k s_i^k - \frac{(s_i^k)^q}{q} \right] \; = \; G_{\alpha,\epsilon^k}(x, s^k) \; \geq \; G_{\alpha,\epsilon^k}(x^{k+1}, s^k).$$

Upon taking limits on both sides of this inequality as $k \in K \to \infty$, and using (42), (44) and the fact that $\{s_i^k\}_K \to |x_i^*|^{p-\alpha}$, $\forall i \in \mathcal{B}$, we obtain that

$$f(x) + \frac{\lambda p}{\alpha} \sum_{i \in \mathcal{B}} \left[|x_i|^\alpha |x_i^*|^{p-\alpha} - \frac{|x_i^*|^{q(p-\alpha)}}{q} \right] \; \geq \; f(x^*) + \lambda \|x^*\|_p^p$$

for all $x \in \Re^n$ such that $x_{\bar{\mathcal{B}}} = 0$. This inequality and (29) immediately yield (43). It then follows from (29) and the first-order optimality condition of (43) that

$$x_i^* \frac{\partial f(x^*)}{\partial x_i} + \lambda p |x_i^*|^p \; = \; 0 \quad \forall i \in \mathcal{B}.$$

Since $x_i^* = 0$ for $i \in \bar{\mathcal{B}}$, we observe that the above equality also holds for $i \in \bar{\mathcal{B}}$. Hence, x^* satisfies (6) and it is a stationary point of (1). ∎

Notice that the above IRL$_\alpha$ method requires solving a sequence of reweighted l_α minimization problems (40) whose solution may not be cheaply computable. We next propose a variant of this method in which each subproblem is much simpler and has a closed form solution for some α's (e.g., $\alpha = 1$ or 2).

A variant of the second type of IRL$_\alpha$ **minimization method for** (1)**:**

Let $\{\epsilon^k\}$ be a sequence of positive vectors in $\Re^n$, and let $0 < L_{\min} < L_{\max}$, $\tau > 1$ and $c > 0$ be given. Choose an arbitrary x^0. Set $k = 0$.

1) Choose $L_k^0 \in [L_{\min}, L_{\max}]$ arbitrarily. Set $L_k = L_k^0$.

   1a) Solve the weighted l_α minimization problem

$$x^{k+1} \in \operatorname{Arg}\min_x \left\{ f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{L_k}{2} \|x - x^k\|_2^2 + \frac{\lambda p}{\alpha} \sum_{i=1}^n s_i^k |x_i|^\alpha \right\}, \tag{45}$$

   where $s_i^k = (|x_i^k|^\alpha + \epsilon_i^k)^{\frac{p}{\alpha}-1}$ for all i.

1b) If

$$F_{\alpha,\epsilon^k}(x^k) - F_{\alpha,\epsilon^{k+1}}(x^{k+1}) \geq \frac{c}{2}\|x^{k+1} - x^k\|_2^2 \tag{46}$$

is satisfied, then go to step 2).

1c) Set $L_k \leftarrow \tau L_k$ and go to step 1a).

2) Set $k \leftarrow k+1$ and go to step 1).

end

We first show that for each outer iteration, the number of its inner iterations is finite.

Theorem 3.7 *For each $k \geq 0$, the inner termination criterion* (46) *is satisfied after at most* $\left\lceil \frac{\log(L_f+c)-\log(2L_{\min})}{\log\tau} + 2 \right\rceil$ *inner iterations.*

Proof. Let $G_{\alpha,\epsilon}(\cdot,\cdot)$ be defined in (32). Since the objective function of (45) is strong convex with modulus L_k, we have

$$\begin{aligned} f(x^k) + \tfrac{\lambda p}{\alpha}\sum_{i=1}^n s_i^k|x_i^k|^\alpha &\geq f(x^k) + \nabla f(x^k)^T(x^{k+1}-x^k) + \tfrac{\lambda p}{\alpha}\sum_{i=1}^n s_i^k|x_i^{k+1}|^\alpha + L_k\|x^{k+1}-x^k\|_2^2, \\ &\geq f(x^{k+1}) + \tfrac{\lambda p}{\alpha}\sum_{i=1}^n s_i^k|x_i^{k+1}|^\alpha + (L_k - \tfrac{L_f}{2})\|x^{k+1}-x^k\|_2^2, \end{aligned}$$

where the last inequality is due to (37). This inequality together with the definition of $G_{\alpha,\epsilon}$ implies that

$$G_{\alpha,\epsilon^k}(x^k, s^k) \;\geq\; G_{\alpha,\epsilon^k}(x^{k+1}, s^k) + (L_k - \frac{L_f}{2})\|x^{k+1}-x^k\|_2^2.$$

In addition, by the same arguments as in the proof of Theorem 3.6, we have $G_{\epsilon^{k+1}}(x^{k+1}, s^{k+1}) \leq G_{\alpha,\epsilon^k}(x^{k+1}, s^k)$. By the definitions of s^k and $F_{\alpha,\epsilon}$, one can easily verify that $G_{\alpha,\epsilon^k}(x^k, s^k) = F_{\alpha,\epsilon^k}(x^k)$ for all k. Combining these relations with the above inequality, we obtain that

$$\begin{aligned} F_{\alpha,\epsilon^{k+1}}(x^{k+1}) &= G_{\epsilon^{k+1}}(x^{k+1}, s^{k+1}) \;\leq\; G_{\alpha,\epsilon^k}(x^{k+1}, s^k) \;\leq\; G_{\alpha,\epsilon^k}(x^k, s^k) - (L_k - \tfrac{L_f}{2})\|x^{k+1}-x^k\|_2^2 \\ &= F_{\alpha,\epsilon^k}(x^k) - (L_k - \tfrac{L_f}{2})\|x^{k+1}-x^k\|_2^2. \end{aligned}$$

Hence, (46) holds whenever $L_k \geq (L_f+c)/2$. The rest of the proof is similar to that of Theorem 4.2. ∎

We next show that any accumulation point of the sequence $\{x^k\}$ generated above is a first-order stationary point of problem (1).

Theorem 3.8 *Suppose that $\{\epsilon^k\}$ is a sequence of non-increasing positive vectors in $\Re^n$ and $\epsilon^k \to 0$ as $k \to \infty$. Let the sequence $\{x^k\}$ be generated by the above* IRL_α *method. Suppose that x^* is an accumulation point of $\{x^k\}$. Then, x^* is a stationary point of* (1)*, i.e.,* (6) *holds at x^*.*

Proof. Since $F_{\alpha,\epsilon}(x) \geq F(x) \geq \underline{f}$ for every $x \in \Re^n$, we see that $\{F_{\alpha,\epsilon^k}(x^k)\}$ is bounded below. In addition, $\{F_{\alpha,\epsilon^k}(x^k)\}$ is non-increasing due to (46). Hence, $\{F_{\alpha,\epsilon^k}(x^k)\}$ converges, which together with (46) implies that $\|x^{k+1} - x^k\| \to 0$. Let $\bar{L}_k$ denote the final value of L_k at the kth outer iteration. By a similar argument as in the proof of Theorem 3.3, we can show that $\bar{L}_k \in [L_{\min}, \tau(L_f + c)/2)$. Let $\mathcal{B} = \{i | x_i^* \neq 0\}$. Since x^* is an accumulation point of $\{x^k\}$, there exists a subsequence K such that $\{x^k\}_K \to x^*$. By the definition of s^k, we see that $\lim\limits_{k \in K \to \infty} s_i^k = |x_i^*|^{p-\alpha}$. The first-order optimality condition of (45) with $L_k = \bar{L}_k$ yields

$$\frac{\partial f(x^{k+1})}{\partial x_i} + \bar{L}_k(x_i^{k+1} - x_i^k) + \lambda p s_i^k |x_i^{k+1}|^{\alpha-1} \operatorname{sgn}(x_i^{k+1}) = 0, \quad \forall i \in \mathcal{B}.$$

Upon taking limits on both sides of the above equality as $k \in K \to \infty$, and using the relation $\lim\limits_{k \in K \to \infty} s_i^k = |x_i^*|^{p-\alpha}$, we have

$$\frac{\partial f(x^*)}{\partial x_i} + \lambda p |x_i^*|^{p-1} \operatorname{sgn}(x_i^*) = 0, \quad \forall i \in \mathcal{B}.$$

Using this relation and a similar argument as in the proof of Theorem 2.1 (i), we can conclude that x^* satisfies (6). ■

4 New iterative reweighted l_1 minimization for (1)

The IRL_1 and IRL_2 methods studied in Section 3 require that the parameter ϵ be dynamically adjusted and approach zero. One natural question is whether an iterative reweighted minimization method can be proposed for (1) that shares a similar convergence with those methods but does not need to adjust ϵ. We will address this question by proposing a new IRL_1 method and its variant.

As shown in Subsection 2.2, problem (20) has a locally Lipschitz continuous objective function and it is an ϵ-approximation to (1). Moreover, when ϵ is below a computable threshold value, a certain stationary point of (20) is also that of (1). In this section we propose new IRL_1 methods for solving (1), which can be viewed as the IRL_1 methods directly applied to problem (20). The novelty of these methods is in that the parameter ϵ is chosen only once and then fixed throughout all iterations. Remarkably, we are able to establish that any accumulation point of the sequence generated by these methods is a first-order stationary point of (1).

New IRL_1 minimization method for (1):

Let q be defined in (13). Choose an arbitrary $x^0 \in \Re^n$ and ϵ such that (21) holds. Set $k = 0$.

1) Solve the weighted l_1 minimization problem

$$x^{k+1} \in \operatorname{Arg\,min}\left\{ f(x) + \lambda p \sum_{i=1}^n s_i^k |x_i| \right\}, \tag{47}$$

where $s_i^k = \min\left\{ (\frac{\epsilon}{\lambda n})^{\frac{1}{q}}, |x_i^k|^{\frac{1}{q-1}} \right\}$ for all i.

2) Set $k \leftarrow k+1$ and go to step 1).

end

We next establish that any accumulation point of $\{x^k\}$ generated by the above method is a first-order stationary point of (1).

Theorem 4.1 *Let the sequence* $\{x^k\}$ *be generated by the above* IRL$_1$ *method. Assume that* ϵ *satisfies* (21)*. Suppose that* x^* *is an accumulation point of* $\{x^k\}$*. Then* x^* *is a first-order stationary point of* (1)*, i.e.,* (6) *holds at* x^**. Moreover, the nonzero entries of* x^* *satisfy the first-order bound* (10)*.*

Proof. Let $u_\epsilon = (\frac{\epsilon}{\lambda n})^{1/q}$ and

$$G(x,s) = f(x) + \lambda p \sum_{i=1}^n \left[|x_i| s_i - \frac{s_i^q}{q} \right]. \qquad (48)$$

By the definition of $\{s^k\}$, one can observe that for $k \geq 0$,

$$s^k = \arg\min_{0 \leq s \leq u_\epsilon} G(x^k, s), \quad x^{k+1} \in \operatorname{Arg}\min_x G(x, s^k). \qquad (49)$$

In addition, we observe that $F_\epsilon(x) = \min\limits_{0 \leq s \leq u_\epsilon} G(x,s)$ and $F_\epsilon(x^k) = G(x^k, s^k)$ for all k, where F_ϵ is defined in (18). It then follows that

$$F_\epsilon(x^{k+1}) \; = \; G(x^{k+1}, s^{k+1}) \; \leq \; G(x^{k+1}, s^k) \; \leq \; G(x^k, s^k) \; = \; F_\epsilon(x^k). \qquad (50)$$

Hence, $\{F_\epsilon(x^k)\}$ is non-increasing. Since x^* is an accumulation point of $\{x^k\}$, there exists a subsequence K such that $\{x^k\}_K \to x^*$. By the continuity of F_ϵ, we have $\{F_\epsilon(x^k)\}_K \to F_\epsilon(x^*)$, which together with the monotonicity of $\{F_\epsilon(x^k)\}$ implies that $F_\epsilon(x^k) \to F_\epsilon(x^*)$. Let $s_i^* = \min\{u_\epsilon, |x_i^*|^{\frac{1}{q-1}}\}$ for all i. We then observe that $\{s^k\}_K \to s^*$ and $F_\epsilon(x^*) = G(x^*, s^*)$. Using (50) and $F_\epsilon(x^k) \to F_\epsilon(x^*)$, we see that $G(x^{k+1}, s^k) \to F_\epsilon(x^*) = G(x^*, s^*)$. In addition, it follows from (49) that

$$G(x, s^k) \; \geq \; G(x^{k+1}, s^k) \quad \forall x \in \Re^n.$$

Upon taking limits on both sides of this inequality as $k \in K \to \infty$, we have

$$G(x, s^*) \; \geq \; G(x^*, s^*) \quad \forall x \in \Re^n,$$

that is,

$$x^* \in \operatorname{Arg}\min \left\{ f(x) + \lambda p \sum_{i=1}^n s_i^* |x_i| \right\}. \qquad (51)$$

The first-order optimality condition of (51) yields

$$0 \in \frac{\partial f(x^*)}{\partial x_i} + \lambda p s_i^* \operatorname{sgn}(x_i^*), \quad \forall i. \qquad (52)$$

Recall that $s_i^* = \min\{u_\epsilon, |x_i^*|^{\frac{1}{q-1}}\}$, which together with (13) implies that for all i,

$$s_i^* = \begin{cases} |x_i^*|^{p-1}, & \text{if } |x_i^*| > u_\epsilon^{q-1}, \\ u_\epsilon, & \text{if } |x_i^*| \le u_\epsilon^{q-1}. \end{cases}$$

Substituting it into (52) and using (15), we obtain that

$$0 \in \frac{\partial f(x^*)}{\partial x_i} + \lambda \partial h_{u_\epsilon}(x_i^*), \quad \forall i.$$

It then follows from (18) that x^* is a first-order stationary point of F_ϵ. In addition, by the monotonicity of $\{F_\epsilon(x^k)\}$ and $F_\epsilon(x^k) \to F_\epsilon(x^*)$, we know that $F_\epsilon(x^*) \le F_\epsilon(x^0)$. Using these results and Theorem 2.7, we conclude that x^* is a first-order stationary point of (1). The rest of conclusion immediately follows from Theorem 2.3. ∎

The above IRL_1 method needs to solve a sequence of reweighted l_1 minimization problems (47) whose solution may not be cheaply computable. We next propose a variant of this method in which each subproblem has a closed form solution.

A variant of new IRL_1 **minimization method for** (1)**:**

Let $0 < L_{\min} < L_{\max}$, $\tau > 1$ and $c > 0$ be given. Let q be defined in (13). Choose an arbitrary x^0 and ϵ such that (21) holds. Set $k = 0$.

1) Choose $L_k^0 \in [L_{\min}, L_{\max}]$ arbitrarily. Set $L_k = L_k^0$.

   1a) Solve the weighted l_1 minimization problem

$$x^{k+1} \in \operatorname{Arg}\min_x \left\{ f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{L_k}{2}\|x - x^k\|_2^2 + \lambda p \sum_{i=1}^n s_i^k |x_i| \right\}, \tag{53}$$

   where $s_i^k = \min\left\{ (\frac{\epsilon}{\lambda n})^{\frac{1}{q}}, |x_i^k|^{\frac{1}{q-1}} \right\}$ for all i.

   1b) If

$$F_\epsilon(x^k) - F_\epsilon(x^{k+1}) \ge \frac{c}{2}\|x^{k+1} - x^k\|_2^2 \tag{54}$$

   is satisfied, where F_ϵ is defined in (18), then go to step 2).

   1c) Set $L_k \leftarrow \tau L_k$ and go to step 1a).

2) Set $k \leftarrow k + 1$ and go to step 1).

end

We first show that for each outer iteration, the number of its inner iterations is finite.

Theorem 4.2 *For each $k \ge 0$, the inner termination criterion* (54) *is satisfied after at most* $\left\lceil \frac{\log(L_f + c) - \log(2L_{\min})}{\log \tau} + 2 \right\rceil$ *inner iterations.*

Proof. Let $\bar{L}_k$ denote the final value of L_k at the kth outer iteration. Since the objective function of (53) is strongly convex with modulus L_k, we have

$$\begin{aligned} f(x^k) + \lambda p \textstyle\sum_{i=1}^n s_i^k |x_i^k| &\geq f(x^k) + \nabla f(x^k)^T (x^{k+1} - x^k) + \lambda p \textstyle\sum_{i=1}^n s_i^k |x_i^{k+1}| + L_k \|x^{k+1} - x^k\|_2^2, \\ &\geq f(x^{k+1}) + \lambda p \textstyle\sum_{i=1}^n s_i^k |x_i^{k+1}| + (L_k - \frac{L_f}{2}) \|x^{k+1} - x^k\|_2^2, \end{aligned}$$

where the last inequality is due to (37). This inequality together with (48) yields

$$G(x^k, s^k) \geq G(x^{k+1}, s^k) + (L_k - \frac{L_f}{2}) \|x^{k+1} - x^k\|_2^2.$$

Recall that $F_\epsilon(x) = \min_{0 \leq s \leq u_\epsilon} G(x, s)$ and $F_\epsilon(x^k) = G(x^k, s^k)$, where $u_\epsilon = (\frac{\epsilon}{\lambda n})^{1/q}$. Using these relations and the above inequality, we obtain that

$$\begin{aligned} F_\epsilon(x^{k+1}) &= G(x^{k+1}, s^{k+1}) \leq G(x^{k+1}, s^k) \leq G(x^k, s^k) - (L_k - \tfrac{L_f}{2}) \|x^{k+1} - x^k\|_2^2 \\ &= F_\epsilon(x^k) - (L_k - \tfrac{L_f}{2}) \|x^{k+1} - x^k\|_2^2. \end{aligned}$$

Hence, (54) holds whenever $L_k \geq (L_f + c)/2$. The rest of the proof is similar to that of Theorem 4.2. ∎

We next establish that any accumulation point of the sequence $\{x^k\}$ generated above is a first-order stationary point of problem (1).

Theorem 4.3 *Let the sequence $\{x^k\}$ be generated by the above variant of new* IRL$_1$ *method. Assume that ϵ satisfies* (21)*. Suppose that x^* is an accumulation point of $\{x^k\}$. Then x^* is a first-order stationary point of* (1)*, i.e.,* (6) *holds at x^*. Moreover, the nonzero entries of x^* satisfy the first-order bound* (10)*.*

Proof. It follows from (54) that $\{F_\epsilon(x^k)\}$ is non-increasing. Since x^* is an accumulation point of $\{x^k\}$, there exists a subsequence K such that $\{x^k\}_K \to x^*$. By the continuity of F_ϵ, we have $\{F_\epsilon(x^k)\}_K \to F_\epsilon(x^*)$, which together with the monotonicity of $\{F_\epsilon(x^k)\}$ implies that $F_\epsilon(x^k) \to F_\epsilon(x^*)$. Using this result and (54), we can conclude that $\|x^{k+1} - x^k\| \to 0$. Let $\bar{L}_k$ denote the final value of L_k at the kth outer iteration. By a similar argument as in the proof of Theorem 3.3, one can show that $\bar{L}_k \in [L_{\min}, \tau(L_f + c)/2)$. The first-order optimality condition of (53) with $L_k = \bar{L}_k$ yields

$$0 \in \frac{\partial f(x^k)}{\partial x_i} + \bar{L}_k (x_i^{k+1} - x_i^k) + \lambda p s_i^k \operatorname{sgn}(x_i^{k+1}) = 0, \quad \forall i.$$

Upon taking limits on both sides of the above equality as $k \in K \to \infty$, we have

$$0 \in \frac{\partial f(x^*)}{\partial x_i} + \lambda p s_i^* \operatorname{sgn}(x_i^*), \quad \forall i,$$

where $s_i^* = \min\{(\frac{\epsilon}{\lambda n})^{1/q}, |x_i^*|^{\frac{1}{q-1}}\}$ for all i. The rest of the proof is similar to that of Theorem 4.1. ∎

5 Computational results

In this section we conduct numerical experiment to compare the performance of the variants of IRL_1 methods proposed in Subsection 3.1 and 3.2 and Section 4. In particular, we apply these methods to problem (5) whose data are randomly generated. For convenience of presentation, we name these variants as IRL_1-1, IRL_1-2 and IRL_1-3, respectively. All codes are written in MATLAB and all computations are performed on a MacBook Pro running with Mac OS X Lion 10.7.4 and 4GB memory.

For all three methods, we choose $L_{\min} = $ 1e-8, $L_{\max} = $ 1e+8, $c = $ 1e-4, $\tau = 1.1$, and $L_0^0 = 1$. And we update L_k^0 by the similar strategy as used in spectral projected gradient method [4], that is,

$$L_k^0 = \max\left\{L_{\min}, \min\left\{L_{\max}, \frac{\Delta x^T \Delta g}{\|\Delta x\|^2}\right\}\right\},$$

where $\Delta x = x^k - x^{k-1}$ and $\Delta g = \nabla f(x^k) - \nabla f(x^{k-1})$. In addition, we choose $\epsilon^k = 0.1^k e$ and $\delta_k = 0.1^k$ for IRL_1-1 and $\epsilon^k = 0.995^k e$ for IRL_1-2, respectively, where e is the all-ones vector. For IRL_1-3, ϵ is chosen to be the one satisfying (21) but within 10^{-6} to the supremum of all ϵ's satisfying (21). The same initial point x^0 is used for IRL_1-1, IRL_1-2 and IRL_1-3. In particular, we choose x^0 to be

$$x^0 \in \operatorname{Arg\,min}\left\{\frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1\right\},$$

which can be computed by a variety of methods (e.g., [29, 1, 23, 30, 32]). And all methods terminate according to the following criterion

$$\|X\nabla f(x) + \lambda p|x|^p\|_\infty \le 1\text{e} - 4,$$

where $X = \text{Diag}(x)$.

In the first experiment, we set $\lambda = $ 3e-3 for problem (5). And the data A and b are randomly generated in the same manner as described in l_1-magic [6]. In particular, given $\sigma > 0$ and positive integers m, n, T with $m < n$ and $T < n$, we first generate a matrix $W \in \Re^{n\times m}$ with entries randomly chosen from a normal distribution with mean zero, variance one and standard deviation one. Then we compute an orthonormal basis, denoted by B, for the range space of W, and set $A = B^T$. We also randomly generate a vector $\tilde{x} \in \Re^n$ with only T nonzero components that are ± 1, and generate a vector $v \in \Re^m$ with entries randomly chosen from a normal distribution with mean zero, variance one and standard deviation one. Finally, we set $b = A\tilde{x} + \sigma v$. Especially, we choose $\sigma = 0.005$ for all instances.

The results of these methods for the above randomly generated instances with $p = 0.1$ and 0.5 are presented in Tables 1 and 2, respectively. In detail, the parameters m and n of each instance are listed in the first two columns, respectively. The objective function value of problem (5) for these methods is given in columns three to five, and CPU times (in seconds) are given in the last three columns, respectively. We shall mention that the CPU time reported here does not include the time for obtaining initial point x^0. For $p = 0.1$, we observe from

Table 1: Comparison of three IRL_1 methods for problem (5) with $p = 0.1$

Problem		Objective Value			CPU Time		
m	n	IRL_1-1	IRL_1-2	IRL_1-3	IRL_1-1	IRL_1-2	IRL_1-3
120	512	0.061371	0.061371	0.061371	0.02	0.29	0.01
240	1024	0.122579	0.122579	0.122579	0.01	0.47	0.01
360	1536	0.183595	0.183595	0.183595	0.01	0.77	0.01
480	2048	0.245253	0.245253	0.245253	0.02	1.45	0.02
600	2560	0.305575	0.305575	0.305575	0.03	2.30	0.03
720	3072	0.367497	0.367496	0.347697	0.04	3.11	0.04
840	3584	0.429549	0.429548	0.429549	0.05	3.83	0.06
960	4096	0.489512	0.489512	0.489512	0.06	5.32	0.08
1080	4608	0.550911	0.550911	0.554911	0.07	6.59	0.10
1200	5120	0.611896	0.611896	0.611896	0.10	7.51	0.13

Table 2: Comparison of three IRL_1 methods for problem (5) with $p = 0.5$

Problem		Objective Value			CPU Time		
m	n	IRL_1-1	IRL_1-2	IRL_1-3	IRL_1-1	IRL_1-2	IRL_1-3
120	512	0.061298	0.062003	0.061298	0.02	0.17	0.01
240	1024	0.122412	0.123449	0.122412	0.01	0.26	0.01
360	1536	0.183376	0.184881	0.183376	0.01	0.43	0.01
480	2048	0.244745	0.247495	0.244745	0.02	0.90	0.02
600	2560	0.304945	0.306632	0.304945	0.03	1.55	0.03
720	3072	0.366621	0.370576	0.366621	0.03	2.07	0.04
840	3584	0.429043	0.433426	0.429043	0.04	2.57	0.06
960	4096	0.488704	0.492537	0.488704	0.05	3.54	0.08
1080	4608	0.550031	0.554057	0.550031	0.06	4.40	0.10
1200	5120	0.610850	0.615399	0.610850	0.07	5.26	0.12

Table 1 that all three methods produce similar objective function values. The CPU time of IRL_1-1 and IRL_1-3 is very close, which is much less than that of IRL_1-2. For $p = 0.5$, we see from Table 2 that IRL_1-1 and IRL_1-3 achieve better objective function values than IRL_1-2 while the former two methods require less CPU time.

In the second experiment, we also randomly generate all instances for problem (5). In particular, we generate matrix A and vector b with entries randomly chosen from standard uniform distribution. In addition, we set $\lambda = 3\text{e-}3$ for (5). The results of these methods for the above randomly generated instances with $p = 0.1$ and 0.5 are presented in Tables 3 and 4, respectively. Same as above, the CPU time reported here does not include the time for obtaining initial point x^0. For $p = 0.1$, we observe from Table 3 that among IRL_1-1, IRL_1-2 and IRL_1-3 achieves best objective function values over 4, 4 and 3 instances out of total 10 instances, respectively. The average CPU time of IRL_1-3 is much less than that of IRL_1-1 and

Table 3: Comparison of three IRL_1 methods for problem (5) with $p = 0.1$

Problem		Objective Value			CPU Time		
m	n	IRL_1-1	IRL_1-2	IRL_1-3	IRL_1-1	IRL_1-2	IRL_1-3
120	512	0.6557	0.6007	0.6011	1.25	1.86	0.93
240	1024	1.1916	1.2090	1.2108	1.96	3.99	2.16
360	1536	1.7047	1.6955	1.7253	3.46	7.34	2.74
480	2048	2.3025	2.3112	2.3270	9.68	14.91	7.86
600	2560	2.7888	2.7432	2.7432	13.50	27.29	20.90
720	3072	3.3639	3.4051	3.4296	19.96	36.75	21.51
840	3584	3.7613	3.7614	3.7085	24.26	46.68	36.10
960	4096	4.4721	4.2879	4.2980	60.26	58.77	47.98
1080	4608	5.0258	4.8848	4.8649	72.45	69.51	39.35
1200	5120	5.2228	5.3789	5.3561	83.99	91.26	57.97

Table 4: Comparison of three IRL_1 methods for problem (5) with $p = 0.5$

Problem		Objective Value			CPU Time		
m	n	IRL_1-1	IRL_1-2	IRL_1-3	IRL_1-1	IRL_1-2	IRL_1-3
120	512	0.2408	0.2415	0.2405	2.12	1.57	0.99
240	1024	0.4096	0.4127	0.4140	7.10	3.12	2.42
360	1536	0.5361	0.5336	0.5336	21.50	5.31	4.04
480	2048	0.6900	0.6900	0.6934	34.93	14.07	9.95
600	2560	0.7725	0.7772	0.7739	61.08	21.68	25.49
720	3072	0.9393	0.9405	0.9406	259.72	34.94	35.55
840	3584	1.0113	1.007	1.007	313.30	47.24	39.43
960	4096	1.1403	1.1297	1.1280	533.36	54.50	52.90
1080	4608	1.2178	1.2186	1.2220	348.94	77.42	80.55
1200	5120	1.3291	1.3375	1.3375	835.89	104.27	114.99

IRL_1-2. For $p = 0.5$, all three methods achieve similar objective function values. The overall CPU time of IRL_1-2 and IRL_1-3 is very close, which is much less than that of IRL_1-1.

From the above two experiments, we observe that IRL_1-3 is generally more stable than IRL_1-1 and IRL_1-2 in terms of objective function value and CPU time.

6 Concluding remarks

In this paper we studied iterative reweighted minimization methods for l_p regularized unconstrained minimization problems (1). In particular, we derived lower bounds for nonzero entries of first- and second-order stationary points, and hence also of local minimizers of (1). We extended some existing IRL_1 and IRL_2 methods to solve (1) and proposed new variants for them. Also, we provided a unified convergence analysis for these methods. In addition, we proposed a novel Lipschitz continuous ϵ-approximation to $\|x\|_p^p$. Using this result, we developed new IRL_1 methods for (1) and showed that any accumulation point of the sequence generated by these methods is a first-order stationary point of problem (1), provided that the approximation parameter ϵ is below a computable threshold value. This is a remarkable result since all existing iterative reweighted minimization methods require that ϵ be dynamically updated and approach zero. Our computational results demonstrate that the new IRL_1 method is generally more stable than the existing IRL_1 methods [21, 18] in terms of objective function value and CPU time.

Recently, Zhao and Li [33] proposed an IRL_1 minimization method to identify sparse solutions to undetermined linear systems based on a class of regularizers. When applied to the l_p regularizer, their method becomes one of the first type of IRL_1 methods discussed in Subsection 3.1. Though we only studied the l_p regularized minimization problems, the techniques developed in our paper can be useful for analyzing the iterative reweighted minimization methods for the optimization problems with other regularizers. In addition, most of the results in this paper can be easily generalized to l_p regularized matrix optimization problems.

References

[1] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2:183–202, 2009.

[2] W. Bian and X. Chen. Smoothing SQP algorithm for non-Lipschitz optimization with complexity analysis. Preprint, February 2012.

[3] W. Bian, X. Chen and Y. Ye. Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization. Preprint, July 2012.

[4] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optimiz*, 4:1196–1211, 2000.

[5] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.*, 27(3):265–274, 2009.

[6] E. Candès and J. Romberg. 1_1-magic : Recovery of sparse signals via convex programming. User's guide, Applied & Computational Mathematics, California Institute of Technology, Pasadena, CA 91125, USA, October 2005. Available at www.l1-magic.org.

[7] E. J. Candés, J. Romberg and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE T. Inform. Theory*, 52:489-509, 2006.

[8] E. Candès and T. Tao. Decoding by linear programming. *IEEE T. Inform. Theory*, 51(12):4203–4215, 2005.

[9] E. Candès and T. Tao. The Dantzig selector: statistical estimation when p is much smaller than n. *Ann. Statist.*, 35(6):2313–2351, 2007.

[10] E. Candès, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted l_1 minimization. *J. Fourier Anal. Appl.*, 14:877–905, 2008.

[11] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Process. Lett.*, 14:707–710, 2007.

[12] R. Chartrand and V. Staneva. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problem*, 24:1–14, 2008.

[13] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. *33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[14] S. Chen, D. Donoho and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20:33-61, 1998.

[15] X. Chen, D. Ge, Z. Wang, and Y. Ye. Complexity of unconstrained L_2-L_p minimization. Submitted to *Math. Program.*, May 2011.

[16] X. Chen, L. Niu, and Y. Yuan. Optimality conditions and smoothing trust region Newton method for non-Lipschitz optimization. Preprint, March 2012.

[17] 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.

[18] X. Chen and W. Zhou. Convergence of reweighted l_1 minimization algorithms and unique solution of truncated l_p minimization. Preprint, April 2010.

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

[20] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gunturk. Iteratively reweighted least squares minimization for sparse recovery. *Commun. Pure. Appl. Math.*, 63:1–38, 2010.

[21] S. Foucart and M. Lai. Sparsest solutions of underdetermined linear systems via l_p-minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.*, 26:395–407, 2009.

[22] D. Ge, X. Jiang, and Y. Ye. A note on the complexity of L_p minimization. *Math. Program.*, 21: 1721–1739, 2011.

[23] E. T. Hale, W. Yin and Y. Zhang. Fixed-point continuation applied to compressed sensing: Implementation and numerical experiments. *J. Comput. Math*, 28(2):170-194, 2010.

[24] K. K. Herrity, A. C. Gilbert and J. A. Tropp. Sparse approximation via iterative thresholding. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2006.

[25] M. Lai and J. Wang. An unconstrained l_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems. *SIAM J. Optim.*, 21:82–101, 2010.

[26] Z. Lu and Y. Zhang. Sparse approximation via penalty decomposition methods. Preprint, May 2012.

[27] Q. Sun. Recovery of sparsest signals via l_q minimization. *Appl. Comput. Harmon. Anal.*, 32:329–341, 2012.

[28] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B*, 58(1):267-288, 1996.

[29] E. Van Den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comp.*, 31(2):890-912, 2008.

[30] S. J. Wright, R. Nowak, and M. Figueiredo. Sparse reconstruction by separable approximation. *IEEE T. Image Process.*, 57:2479–2493, 2009.

[31] Z. Xu, H. Zhang, Y. Wang, and X. Chang. $L_{\frac{1}{2}}$ regularizer. *Science in China*, 52:1–9, 2009.

[32] S. Yun and K.-C. Toh. A coordinate gradient descent method for l_1-regularized convex minimization. *Comput. Optim. Appl.*,48:273–307, 2011.

[33] Y. Zhao and D. Li. Reweighted l_1-minimization for sparse solutions to underdetermined linear systems. To appear in *SIAM J. Optim.*, 2012.