Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL
Pacific Journal of Optimization
Vol. 10, No. 2, pp. 401–414, April 2014

A GRADIENT BASED METHOD FOR THE L_2-$L_{1/2}$ MINIMIZATION AND APPLICATION TO COMPRESSIVE SENSING

Lei Wu*, Zhe Sun† and Dong-Hui Li‡

Abstract: We study the L_2-$L_{1/2}$ minimization problem arising from compressive sensing. We first derive an equivalent nonsmooth equation to the first order necessary condition of the problem. Based on this equation, we construct a descent direction for the objective function and then develop a gradient based method to solve the problem. The method is very easy to implement. A good property of the method is that the generated sequence of the function evaluations is decreasing. Under mild conditions, we prove the global convergence of the method. Our preliminary numerical results show that the proposed method is practically effective.

Key words: *compressive sensing, L_2-$L_{1/2}$ minimization, optimality condition, gradient based method*

Mathematics Subject Classification: *5A09, 90C26, 90C46, 49M05*

1 Introduction

In the last few years, algorithms for finding sparse solutions of underdetermined linear systems have been extensively studied, largely because solving such problems is one of critical problems in compressive sensing (CS) research. Let x be an original signal that we wish to capture. Without loss of generality, we assume that x is sparse under the canonical basis; i.e., the number of nonzero components in x, denoted by $\|x\|_0$, is far fewer than its length. Instead of sampling x directly, in CS one first obtains a set of linear measurements

$$b = Ax \in R^m, \tag{1.1}$$

where $A \in R^{m\times n}$ $(m < n)$ is an encoding matrix. The original signal x is then reconstructed from the underdetermined linear system $Ax = b$ via a certain reconstruction technique. Basic CS theory presented in [5, 12] states that it is extremely probable to reconstruct x accurately or even exactly from b provided that x is sufficiently sparse relative to the number of measurements, and the encoding matrix A possesses certain desirable attributes [29].

*The first and third author were supported by the NNSF of China under the grant 11071087, and the major project of the Ministry of Education of China under the grant 309023.
†The second author was supported by the NNSF of China under the grant 11126147 and 11201197.
‡Corresponding author

In order to recover the sparse signal x from the underdetermined system (1.1), one could naturally consider seeking among all solutions of (1.1) the sparsest one, which results in the optimization problem

$$\min_{x \in R^n} \{\|x\|_0 : Ax = b\}. \tag{1.2}$$

Problem (1.2) is called the L_0 minimization problem. The solution is called the sparse solution of $Ax = b$. Unfortunately, problem (1.2) is a discrete minimization problem which is very hard to deal with. In terms of computational complexity, it is NP-hard [19]. In order to overcome such difficulty, a typical relaxation problem is the L_1 minimization problem

$$\min_{x \in R^n} \{\|x\|_1 : Ax = b\}, \tag{1.3}$$

where $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ is the standard L_1 norm. This problem was studied independently by Tibshirani [24] and Chen, Donoho and Saunders [8], known respectively as LASSO [24] and Basis Pursuit [8]. It is a convex quadratic optimization problem, and therefore, can be very efficiently solved. Many researchers have made a lot of contributions to tackle the problem (1.3), see, e.g., [2, 9, 11, 13, 16, 17, 23, 25, 29, 30]. However, the L_1 minimization problem often leads to sub-optimal sparsity in reality [31]. For many practical applications, the solutions yielded from L_1 minimization problems are less sparse than those of L_0 minimization problems.

Another approach for the computation of the sparse solutions is based on the L_p minimization problem

$$\min_{x \in R^n} \{\|x\|_p^p : Ax = b\}, \tag{1.4}$$

where $\|x\|_p^p = \sum_{i=1}^n |x_i|^p$ and $0 < p < 1$. This L_p minimization problem is motivated by the fact

$$\|x\|_0 = \lim_{p \to 0+} \|x\|_p^p.$$

From the theory of optimization, the solution of the L_p minimization problem (1.4) can be found via solving the L_2-L_p minimization problem

$$\min_{x \in R^n} \frac{1}{2}\|Ax - b\|_2^2 + \rho\|x\|_p^p, \tag{1.5}$$

where $\|\cdot\|_2^2$ stands for the Euclidian norm of vectors and $\rho > 0$ is a parameter. Compared with the L_1 minimization problem, the L_p minimization problem with $0 < p < 1$ can find sparser solutions, which was evidenced in extensive computational studies [6, 7, 27]. Moreover, Xu, Zhang, Wang and Chang [27] numerically justified that the sparsity-promotion capability of the $L_{1/2}$ minimization problem was strongest among the L_p minimization problems with all $p \in [1/2, 1)$ and the L_p minimization problem performed similarly when $0 < p \leq 1/2$. So the $L_{1/2}$ minimization problem can be taken as a representative of L_p $(0 < p < 1)$ minimization problems. However, the L_p $(0 < p < 1)$ minimization problem is a nonconvex, nonsmooth and non-Lipschitz optimization problem. It is very difficult in general to have a thorough theoretical understanding and efficient algorithms for its solutions. Ge, Jiang and Ye [14] proved that finding the global minimal value was strongly NP-hard, while computing a local minimizer could be done in polynomial time. They developed an interior-point potential reduction algorithm with a provable complexity bound to solve this problem.

Recently, there have been developed some iterative methods for solving the L_2-L_p minimization problem. Chen, Xu and Ye in [10] developed a lower bound theorem to classify zero and nonzero entries in every local solution. By combining the orthogonal matching pursuit (OMP) algorithm [4] and the smoothing gradient (SG) algorithm, they proposed a hybrid OMP-SG algorithm to solve the L_2-L_p minimization problem (1.5). Based on a smoothing technique, Lai and Wang [18] developed an iterative algorithm to solve problem (1.5). However, this algorithm requires solving a series of systems of nonlinear equations. Recently, Xu, Chang, Xu and Zhan [28] developed a threshoding representation theory for the L_2-$L_{1/2}$ minimization problem and proved an alternative feature theorem on its solutions. They proposed an iterative half thresholding algorithm for solving it and verified the convergence of the iterative half thresholding algorithm.

In this paper, our goal is to develop a first order method to solve the following L_2-$L_{1/2}$ minimization problem

$$\min_{x\in R^n} f(x) \triangleq \frac{1}{2}\|Ax-b\|_2^2 + \rho\|x\|_{1/2}^{1/2}. \tag{1.6}$$

The method is a descent algorithm in the sense that the generated sequence of function evaluations is decreasing. We show that the method is globally convergent under mild conditions. Finally, we shall apply the proposed method to solve some practical problems from the sparse signal recovery.

The remainder of the paper is organized as follows. In Section 2, we derive an equivalent nonsmooth equation to the first order necessary condition of the L_2-$L_{1/2}$ minimization problem (1.6) and construct a descent direction of the objective function f defined by (1.6). In Section 3, we propose a gradient based method for solving (1.6) and establish its global convergence. In Section 4, we do some preliminary numerical experiments to test the efficiency of the proposed method. The numerical results show that the proposed method can recover the original signal well.

2 Preliminaries

For simplicity, we let $L(x) \triangleq \frac{1}{2}\|Ax-b\|_2^2$ in the latter part of this paper. Suppose that $x^* = (x_1^*, x_2^*, \dots, x_n^*)^T \in R^n$ is a solution of (1.6). Then for $x_i^* \neq 0$, it holds that

$$\nabla_i L(x^*) + \frac{\rho \operatorname{sign}(x_i^*)}{2|x_i^*|^{1/2}} = 0.$$

Here and throughout, $\nabla_i L(x)$ denotes the ith component of $\nabla L(x)$ and $\nabla L(x)$ denotes the gradient of L at x. The last equation can be further written as

$$\begin{cases} \nabla_i L(x^*) + \dfrac{\rho}{2|x_i^*|^{1/2}} = 0, & x_i^* > 0, \\ \nabla_i L(x^*) - \dfrac{\rho}{2|x_i^*|^{1/2}} = 0, & x_i^* < 0, \end{cases}$$

or equivalently,

$$\begin{cases} 2|x_i^*|^{1/2}\nabla_i L(x^*) + \rho = 0, & x_i^* > 0, \\ 2|x_i^*|^{1/2}\nabla_i L(x^*) - \rho = 0, & x_i^* < 0. \end{cases}$$

Define $h(x) = (h_1(x), h_2(x), \dots, h_n(x))^T : R^n \to R^n$ with elements

$$h_i(x) = \operatorname{mid}\left\{2|x_i|^{1/2}\nabla_i L(x) - \rho,\ \ x_i,\ \ 2|x_i|^{1/2}\nabla_i L(x) + \rho\right\}, \quad i = 1, 2, \dots, n, \tag{2.1}$$

where for $a, b, c \in R$, the function mid{a, b, c} takes the medium value among a, b and c. Then it is not difficult to see that a solution of (1.6) must be a solution of the nonlinear equation $h(x) = 0$. Specifically, we have the following proposition.

Proposition 2.1. *If $x^* \in R^n$ is a minimizer of problem (1.6), then*

$$h(x^*) = 0. \tag{2.2}$$

The condition $h(x^*) = 0$ is called the first order necessary condition for $x^* \in R^n$ to be a solution of (1.6). A point is called a stationary point of f if it satisfies (2.2). If some second order condition holds at the stationary point x^*, then it will be a local minimizer of the problem [14].

We are not intended to solve the nonsmooth equation (2.2) directly. Instead, we are going to find a descent direction for $f(x)$ based on the function $h(x)$ defined by (2.1) and then develop a decent method for solving (1.6). The following proposition shows that if x is not a stationary point, then $-h(x)$ provides a descent direction of f at x.

Proposition 2.2. *Let $x \in \mathbb{R}^n$ satisfy $h(x) \neq 0$. Then $d = -h(x)$ is a descent direction of f at x, i.e., there is a constant $\bar{\alpha} > 0$ such that*

$$f(x + \alpha d) < f(x), \quad \forall \alpha \in (0, \bar{\alpha}).$$

Moreover, for all $\alpha > 0$ sufficiently small it holds that

$$f(x + \alpha d) \leq f(x) - \gamma\alpha \frac{\|d\|^2}{\mu(x)}, \tag{2.3}$$

where $\gamma \in (0, 1)$ and $\mu(x) = 2 \max\limits_{1 \leq i \leq n} |x_i|^{1/2}$.

Proof. First we note that the condition $h(x) \neq 0$ implies $x \neq 0$. By Taylor's expansion, for any $\alpha > 0$ sufficiently small, we get by the definition of $f(x)$

$$\begin{aligned} f(x + \alpha d) &= L(x + \alpha d) + \rho\|x + \alpha d\|_{1/2}^{1/2} \\ &= L(x) + \alpha \nabla L(x)^T d + o(\alpha) + \rho\|x\|_{1/2}^{1/2} + \rho\left(\|x + \alpha d\|_{1/2}^{1/2} - \|x\|_{1/2}^{1/2}\right) \\ &= f(x) + \sum_{i=1}^{n} \left[\alpha \nabla_i L(x) d_i + \rho(|x_i + \alpha d_i|^{1/2} - |x_i|^{1/2})\right] + o(\alpha). \end{aligned} \tag{2.4}$$

For i with $x_i \neq 0$, it holds that

$$|x_i + \alpha d_i|^{1/2} - |x_i|^{1/2} = \frac{\alpha d_i \operatorname{sign}(x_i)}{2|x_i|^{1/2}} + o(\alpha),$$

which implies

$$\begin{aligned} &\alpha \nabla_i L(x) d_i + \rho\left(|x_i + \alpha d_i|^{1/2} - |x_i|^{1/2}\right) \\ &= \alpha\left(\nabla_i L(x) + \rho \operatorname{sign}(x_i)/2|x_i|^{1/2}\right) d_i + o(\alpha) \\ &= \frac{\alpha\left(2|x_i|^{1/2} \nabla_i L(x) + \rho \operatorname{sign}(x_i)\right) d_i}{2|x_i|^{1/2}} + o(\alpha). \end{aligned} \tag{2.5}$$

For i with $x_i = 0$, we deduce from (2.1) and the definition of d that $d_i = 0$ and hence

$$\alpha \nabla_i L(x) d_i + \rho(|x_i + \alpha d_i|^{1/2} - |x_i|^{1/2}) = 0. \tag{2.6}$$

Combining (2.4)-(2.6), we obtain

$$f(x+\alpha d) = f(x) + \alpha \left[\sum_{x_i \neq 0} \Big(2|x_i|^{1/2} \nabla_i L(x) + \rho \operatorname{sign}(x_i) \Big) d_i / 2|x_i|^{1/2} \right] + o(\alpha). \tag{2.7}$$

For index i with $x_i > 0$, we have either $h_i(x) = 2|x_i|^{1/2} \nabla_i L(x) + \rho$ and

$$\Big(2|x_i|^{1/2} \nabla_i L(x) + \rho \operatorname{sign}(x_i) \Big) d_i = -h_i^2(x),$$

or $h_i(x) < 2|x_i|^{1/2} \nabla_i L(x) + \rho$ and

$$\Big(2|x_i|^{1/2} \nabla_i L(x) + \rho \operatorname{sign}(x_i) \Big) d_i < -h_i^2(x).$$

This implies that for each i with $x_i > 0$, it holds that

$$\Big(2|x_i|^{1/2} \nabla_i L(x) + \rho \operatorname{sign}(x_i) \Big) d_i \leq -h_i^2(x). \tag{2.8}$$

For index i with $x_i < 0$, by a similar argument we get

$$\Big(2|x_i|^{1/2} \nabla_i L(x) + \rho \operatorname{sign}(x_i) \Big) d_i \leq -h_i^2(x). \tag{2.9}$$

Since $h(x) \neq 0$ and $h_i(x) = 0$ for each i satisfying $x_i = 0$, it must hold

$$\sum_{x_i \neq 0} h_i^2(x) > 0.$$

This together with (2.8) and (2.9) implies

$$\sum_{x_i \neq 0} \Big(2|x_i|^{1/2} \nabla_i L(x) + \rho \operatorname{sign}(x_i) \Big) d_i / 2|x_i|^{1/2} \leq -\frac{\|d\|^2}{\mu(x)}. \tag{2.10}$$

Therefore we claim by (2.7) that (2.3) is satisfied for all $\alpha > 0$ sufficiently small. In particular, d is a descent direction of f at x. The proof is complete. $\square$

Proposition 2.2 gives a descent direction of the objective function f at x. Based on this descent direction, we will propose a gradient based method to solve the L_2-$L_{1/2}$ minimization problem (1.6) in the next section.

3 A Gradient Based Method and Its Convergence

In this section, we present an algorithm with a nonmonotone line search strategy [15] for solving problem (1.6), and establish its global convergence. Let x^k be the current iterate. For simplicity, in the following, we denote $d^k \triangleq -h(x^k)$.

Algorithm 3.1 (Gradient Based Method)**.** Given $x^0 \neq 0$, $0 < \lambda_{\min} < \lambda_{\max} < \infty$, $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$, integer $M \geq 0$, $\gamma \in (0,1)$, $0 < \sigma_1 < \sigma_2 < 1$. Set $k := 0$.

Step 1. If $d^k = 0$, stop. Otherwise, compute

$$\mu^k \triangleq \max_{i=1,\cdots,n} 2|x_i^k|^{1/2}.$$

Step 2. Backtracking line search

Step 2.1 Set $\lambda \leftarrow \lambda_k$.

Step 2.2 (nonmonotone line search)
If

$$f(x^k + \lambda d^k) \leq \max_{0 \leq j \leq \min(k,M)} f(x^{k-j}) - \gamma\lambda \frac{\|d^k\|^2}{\mu^k}, \tag{3.1}$$

then set $\alpha^k = \lambda$, $x^{k+1} = x^k + \alpha^k d^k$, $s^k = x^{k+1} - x^k$, $y^k = d^k - d^{k+1}$. Go to Step 3.

Step 2.3 Choose

$$\lambda_{\text{new}} \in [\sigma_1\lambda, \sigma_2\lambda].$$

Set $\lambda \leftarrow \lambda_{\text{new}}$, and go to Step 2.2.

Step 3. Compute $p^k = \langle s^k, y^k\rangle$. If $p^k \leq 0$, set $\lambda_{k+1} = \lambda_{\max}$; else, compute $q^k = \langle y^k, y^k\rangle$ and

$$\lambda_{k+1} = \min\{\lambda_{\max}, \max\{\lambda_{\min}, p^k/q^k\}\}.$$

Set $k := k+1$, and go to Step 1.

Remark 3.2. (i) The determination of the steplength α_k in Steps 2 and 3 is based on the steplength rule in [1] and [15].

(ii) It is not difficult to see from Proposition 2.2 that the cycle between Steps 2.2 and 2.3 is finite. Consequently, Algorithm 3.1 is well defined.

(iii) Denote

$$\Omega_0 \triangleq \{x : f(x) \leq f(x^0)\}.$$

It is easy to see that Ω_0 is a bounded set. Moreover, the sequence $\{x^k\}$ generated by Algorithm 3.1 is contained in Ω_0.

The latter part of this section is dedicated to the global convergence of Algorithm 3.1. To this end, we first show the following lemma.

Lemma 3.3. *Suppose that the sequence $\{x^k\}$ generated by Algorithm 3.1 is infinite. Then we have*

$$\lim_{k\to\infty} \alpha^k d^k = 0.$$

Proof. Let $l(k)$ be an integer such that

$$k - \min(k,M) \leq l(k) \leq k \qquad \text{and} \qquad f(x^{l(k)}) = \max_{0 \leq j \leq \min(k,M)} f(x^{k-j}). \tag{3.2}$$

Then, we have

$$\begin{aligned} f(x^{l(k+1)}) &= \max_{0 \leq j \leq \min(k+1,M)} f(x^{k+1-j}) \leq \max_{0 \leq j \leq \min(k,M)+1} f(x^{k+1-j}) \\ &= \max\left\{(f(x^{k+1}), f(x^{l(k)}))\right\}. \end{aligned}$$

This together with (3.1) implies

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

Hence, the sequence $\{f(x^{l(k)})\}$ is nonincreasing. Moreover we obtain from (3.1) for any $k > M$,

$$\begin{aligned} f(x^{l(k)}) &= f(x^{l(k)-1} + \alpha^{l(k)-1} d^{l(k)-1}) \\ &\le \max_{0 \le j \le \min(l(k)-1, M)} f(x^{l(k)-1-j}) - \gamma \alpha^{l(k)-1} \frac{\|d^{l(k)-1}\|^2}{\mu^{l(k)-1}} \\ &= f(x^{l(l(k)-1)}) - \gamma \alpha^{l(k)-1} \frac{\|d^{l(k)-1}\|^2}{\mu^{l(k)-1}}. \end{aligned} \tag{3.3}$$

Since $f(x^k) \le f(x^0)$ for all k, we have $\{x^k\} \subset \Omega_0$. Consequently the sequence $\{f(x^{l(k)})\}$ admits a limit. Also since $\alpha^k > 0$, it follows from (3.3) that

$$\lim_{k \to \infty} \alpha^{l(k)-1} \frac{\|d^{l(k)-1}\|^2}{\mu^{l(k)-1}} = 0, \tag{3.4}$$

and hence it must hold

$$\lim_{k \to \infty} \alpha^{l(k)-1} \frac{\|d^{l(k)-1}\|}{\mu^{l(k)-1}} = 0. \tag{3.5}$$

In fact, if (3.5) does not hold, there exists a subsequence

$$\lim_{k' \to \infty} \alpha^{l(k')-1} \frac{\|d^{l(k')-1}\|}{\mu^{l(k')-1}} = c, \tag{3.6}$$

where $0 < c \le +\infty$. Equation (3.6) together with (3.4) implies $\lim_{k' \to \infty} \|d^{l(k')-1}\| = 0$. Since $0 < \alpha^{l(k)-1} \le \lambda_{\max}$, it follows from (3.6) that $\lim_{k' \to \infty} \mu^{l(k')-1} = 0$. By the definition of μ^k, d^k and $h(x)$, we have $d_i^{l(k')-1} = -h_i(x^{l(k')-1}) = -x_i^{l(k')-1}$ for any i and all k' sufficiently large. This yields $\lim_{k' \to \infty} \frac{\|d^{l(k')-1}\|}{\mu^{l(k')-1}} = 0$. Consequently, we obtain

$$\lim_{k' \to \infty} \alpha^{l(k')-1} \frac{\|d^{l(k')-1}\|}{\mu^{l(k')-1}} = 0.$$

This is a contradiction to (3.6). The contradiction proves (3.5).

In what follows, we shall prove $\lim_{k \to \infty} \alpha^k \|d^k\| / \mu^k = 0$. Let

$$\hat{l}(k) \triangleq l(k + M + 2).$$

Here and in the sequel, we assume, without loss of generality, that the iteration index k is large enough to avoid the occurrence of negative subscripts, that is $k \ge j - 1$.

First we show, by induction, that for any given $j \ge 1$

$$\lim_{k \to \infty} \alpha^{\hat{l}(k)-j} \frac{\|d^{\hat{l}(k)-j}\|}{\mu^{\hat{l}(k)-j}} = 0 \tag{3.7}$$

and

$$\lim_{k \to \infty} f(x^{\hat{l}(k)-j}) = \lim_{k \to \infty} f(x^{l(k)}). \tag{3.8}$$

If $j = 1$, since $\{\hat{l}(k)\} \subset \{l(k)\}$, (3.7) follows from (3.5). Since Ω_0 is bounded, $\mu^{\hat{l}(k)-j}$ must be bounded from above. This and (3.7) in turn imply $\left\{ \|x^{\hat{l}(k)} - x^{\hat{l}(k)-1}\| \right\} \to 0$. Since

$f(x)$ is uniformly continuous on Ω_0, we obtain (3.8) for $j = 1$. Assume now that (3.7) and (3.8) hold for a given j. Then by (3.1) one can write

$$f(x^{\hat{l}(k)-j}) \le f(x^{l(\hat{l}(k)-j-1)}) - \gamma\alpha^{\hat{l}(k)-j-1}\frac{\|d^{\hat{l}(k)-j-1}\|^2}{\mu^{\hat{l}(k)-j-1}}.$$

Taking limits in both sides of the last inequality as $k \to \infty$, we derive by (3.8)

$$\lim_{k\to\infty} \alpha^{\hat{l}(k)-(j+1)}\frac{\|d^{\hat{l}(k)-(j+1)}\|^2}{\mu^{\hat{l}(k)-(j+1)}} = 0.$$

Similar to the proof of (3.5), it is not difficult to derive

$$\lim_{k\to\infty} \alpha^{\hat{l}(k)-(j+1)}\frac{\|d^{\hat{l}(k)-(j+1)}\|}{\mu^{\hat{l}(k)-(j+1)}} = 0.$$

This implies $\left\{\|x^{\hat{l}(k)-j} - x^{\hat{l}(k)-(j+1)}\|\right\} \to 0$. By (3.8) and the uniform continuity of L on Ω_0, we get

$$\lim_{k\to\infty} f(x^{\hat{l}(k)-(j+1)}) = \lim_{k\to\infty} f(x^{\hat{l}(k)-j}) = \lim_{k\to\infty} f(x^{l(k)}).$$

By the principle of induction, we have proved (3.7) and (3.8) for any given $j \ge 1$.

Now for any k, it holds that

$$x^{k+1} = x^{\hat{l}(k)} - \sum_{j=1}^{\hat{l}(k)-k-1} \alpha^{\hat{l}(k)-j} d^{\hat{l}(k)-j}. \tag{3.9}$$

By (3.2), we have $\hat{l}(k) - k - 1 = l(k + M + 2) - k - 1 \le M + 1$. It then follows from (3.7) and (3.9) that

$$\lim_{k\to\infty} \|x^{k+1} - x^{\hat{l}(k)}\| = 0.$$

Since $\{f(x^{l(k)})\}$ admits a limit, by the uniform continuity of f on Ω_0, it holds that

$$\lim_{k\to\infty} f(x^k) = \lim_{k\to\infty} f(x^{\hat{l}(k)}). \tag{3.10}$$

By (3.1), we have

$$f(x^{k+1}) \le f(x^{l(k)}) - \gamma\alpha^k\frac{\|d^k\|^2}{\mu^k}.$$

Taking limits in both sides of the last inequality as $k \to \infty$, we obtain by (3.10) $\lim_{k\to\infty} \alpha^k\frac{\|d^k\|^2}{\mu^k} = 0$. Following a similar argument to the proof of (3.5), it is not difficult to get

$$\lim_{k\to\infty} \alpha^k\frac{\|d^k\|}{\mu^k} = 0.$$

Since Ω_0 is bounded, the positive sequence $\{\mu^k\}$ must be bounded from above. By the definition of μ^k, we get $\lim\limits_{k\to\infty} \alpha^k\|d^k\| = 0$ as desired.

$\square$

The global convergence of Algorithm 3.1 is stated in the following theorem.

Theorem 3.4. *Suppose that the sequence $\{x^k\}$ is generated by Algorithm 3.1. Then either $h(x^j) = 0$ for some finite j, or every limit point $\bar{x}$ of $\{x^k\}$ satisfies $h(\bar{x}) = 0$.*

Proof. When $\{x^k\}$ is finite, from the termination condition, it is clear that there exists some finite j satisfying $h(x^j) = 0$. Suppose that $\{x^k\}$ is infinite. Then by Lemma 3.3, we have

$$\lim_{k\to\infty} \alpha^k \|h^k\| = 0. \tag{3.11}$$

Let $\bar{x}$ be an arbitrary limit point of $\{x^k\}$ and relabel $\{x^k\}$ a subsequence converging to $\bar{x}$. It is easy to see from (3.11) that either $\lim_{k\to\infty} \|h(x^k)\| = 0$, which implies, by continuity, $h(\bar{x}) = 0$, or there exists a subsequence $\{\alpha^k\}_K \subset \{\alpha^k\}$ such that

$$\lim_{k\to\infty, k\in K} \alpha^k = 0.$$

In the latter case, by the line search rule, there exists a positive integer $\bar{k}$ satisfying for all $k \geq \bar{k}$, $k \in K$ there is a constant $\sigma^k \in [\sigma_1, \sigma_2]$ such that

$$f\left(x^k - \frac{\alpha^k}{\sigma^k} h^k\right) > \max_{0\leq j\leq \min(k,M)} f(x^{k-j}) - \gamma \frac{\alpha^k}{\sigma^k} \frac{\|h^k\|^2}{\mu^k} \geq f(x^k) - \gamma \frac{\alpha^k}{\sigma^k} \frac{\|h^k\|^2}{\mu^k}.$$

This implies,

$$f\left(x^k - \frac{\alpha^k}{\sigma^k} h^k\right) - f(x^k) \geq -\gamma \frac{\alpha^k}{\sigma^k} \frac{\|h^k\|^2}{\mu^k}. \tag{3.12}$$

By the mean value theorem, for any $k \geq \bar{k}$, $k \in K$, there exists an $\omega_k \in (0, 1)$ such that

$$\begin{aligned}
f(x^k - \frac{\alpha^k}{\sigma^k} h^k) - f(x^k) &= L(x^k - \frac{\alpha^k}{\sigma^k} h^k) - L(x^k) + \rho(\|x^k - \frac{\alpha^k}{\sigma^k} h^k\|_{1/2}^{1/2} - \|x^k\|_{1/2}^{1/2}) \\
&= -\frac{\alpha^k}{\sigma^k} \nabla L(x^k - \omega_k \frac{\alpha^k}{\sigma^k} h^k)^T h^k + \rho \sum_{x_i^k \neq 0} (|x_i^k - \frac{\alpha^k}{\sigma^k} h_i^k|^{1/2} - |x_i^k|^{1/2}) \\
&= -\frac{\alpha^k}{\sigma^k} \sum_{x_i^k \neq 0} (\nabla_i L(x^k - \omega_k \frac{\alpha^k}{\sigma^k} h^k) + \frac{\rho \text{sign } (x_i^k)}{|x_i^k - \frac{\alpha^k}{\sigma^k} h_i^k|^{1/2} + |x_i^k|^{1/2}}) h_i^k.
\end{aligned}$$

This together with (3.12) implies

$$\sum_{x_i^k \neq 0} \left(\nabla_i L(x^k - \omega_k \frac{\alpha^k}{\sigma^k} h^k) + \frac{\rho \text{sign } (x_i^k)}{|x_i^k - \frac{\alpha^k}{\sigma^k} h_i^k|^{1/2} + |x_i^k|^{1/2}}\right) h_i^k \leq \gamma \frac{\|h^k\|^2}{\mu^k}.$$

Let $\{x^k\}_{K_1} \subset \{x^k\}_K$ be a subsequence such that $\lim\limits_{k\to\infty, k\in K_1} x^k = \bar{x}$. Since $\sigma^k \in [\sigma_1, \sigma_2]$ and $\lim\limits_{k\to\infty} \alpha^k h^k = 0$, taking limits in the last inequality, we obtain

$$\sum_{\bar{x}_i \neq 0} \left(\nabla_i L(\bar{x}) + \frac{\rho \text{sign } (\bar{x}_i)}{2|\bar{x}_i|^{1/2}}\right) h_i(\bar{x}) \leq \gamma \frac{\|h(\bar{x})\|^2}{\bar{\mu}}$$

with $\bar{\mu} = \max\limits_{i=1,\dots,n} \{2|\bar{x}_i|^{1/2}\}$, which yields

$$\sum_{\bar{x}_i \neq 0} \frac{(2|\bar{x}_i|^{1/2} \nabla_i L(\bar{x}) + \rho \text{sign } (\bar{x}_i)) h_i(\bar{x})}{2|\bar{x}_i|^{1/2}} \leq \gamma \frac{\|h(\bar{x})\|^2}{\bar{\mu}},$$

and hence

$$\begin{aligned}\frac{\|h(\bar{x})\|^2}{\bar{\mu}} &= \frac{\sum\limits_{\bar{x}_i=0}(h_i(\bar{x}))^2+\sum\limits_{\bar{x}_i\neq 0}(h_i(\bar{x}))^2}{\bar{\mu}}\\ &\le \sum_{\bar{x}_i\neq 0}\frac{(2|\bar{x}_i|^{1/2}\nabla_i L(\bar{x})+\rho\text{sign }(\bar{x}_i))h_i(\bar{x})}{2|\bar{x}_i|^{1/2}}\le \gamma\frac{\|h(\bar{x})\|^2}{\bar{\mu}},\end{aligned}$$

where the first inequality can be obtained in a way similar to the proof of (2.10). Since $\gamma \in (0,1)$ and $\bar{\mu} > 0$, we get $h(\bar{x}) = 0$. This completes the proof. □

4 Applications to Compressive Sensing

In this section, we provide some applications to demonstrate the high performance of the proposed algorithm (abbreviated as GBM).

We apply the GBM method to a typical compressive sensing problem, i.e., sparse signal recovery, where the goal is to reconstruct a length-n sparse signal (in the canonical basis) from m observations with $m < n$. Our purpose is to assess the efficiency, accuracy and the ability in recovering the sparsity of the algorithm. In particular, we compare the performance of the GBM with five other existing algorithms, namely the orthogonal greedy algorithm (OGA, see [21]), the regularized orthogonal matching pursuit algorithm (OMP, see [20]), the spg_bp algorithm (spg_bp, see [26]), the iterative half thresholding algorithm (abbreviated as IHTA), which was proposed recently by Xu et al. [28] to find sparse solutions via solving the L_2-$L_{1/2}$ minimization problem (1.6), and the iterative algorithm (abbreviated as IALW) proposed by Lai and Wang [18].

While do numerical experiments, we chose the parameters in the GBM method as the same as those in [3] and [22]. Specifically, we took $\lambda_{\min} = 10^{-3}$, $\lambda_{\max} = 10^3$, $\lambda_0 = 1$, $M = 5$, $\gamma = 10^{-4}$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.6$. The code of the iterative half thresholding algorithm (IHTA) was provided by Dr. Chang, one author of [28]. The parameters μ_n and λ_n in the IHTA were chosen according to Scheme 3 with the values suggested by the authors. We wrote the code of the IALW method with the parameters as given in [18]. The Matlab codes for other compared algorithms were obtained via internet. Specifically, the OGA code was from http : //www.math.drexel.edu/ ∼ foucart/software.htm, the OMP code from http : //www − personal.umich.edu/ ∼ romanv/software/romp.m and the spg_bp code came from http : //www.cs.ubc.ca/ ∼ mpf/software.html. All the codes were run on a Lenovo PC (2.53GHz, 2.00GB of RAM) with the use of Matlab 7.8.

We applied the above six algorithms to the sparse signal recovery problem. We say the original signal x_s to be T-sparse if only T of the signal coefficients are nonzero and the others are zero. In the experiments, we used the following Matlab code to generate the original signal x_s, the sensing matrix A and the observation b.

$$\begin{gathered}\mathrm{x_s = zeros(n,1);\quad p = randperm(n);\quad x_s(p(1:T)) = sign(randn(T,1));}\\ \mathrm{A = randn(m,n);\quad A = orth(A')';\quad b = A * x_s.}\end{gathered} \tag{4.1}$$

We first compared the GBM algorithm with OGA, OMP, spg_bp and IHTA in the ability of recovering the sparse solutions. We constructed 400 random pairs (A, x_s) with matrices $A \in R^{64\times 256}$ and vectors $x_s \in R^{256}$ using (4.1) for sparsity $\|x_s\|_0 = T$ with $T = 1, 2, \ldots, 30$. For each T and each pair, we run these algorithms to obtain a vector $\tilde{x}$. We regard the recovery a success if $\|x_s - \tilde{x}\|_\infty < 10^{-5}$. We plot the percentages of successfully finding

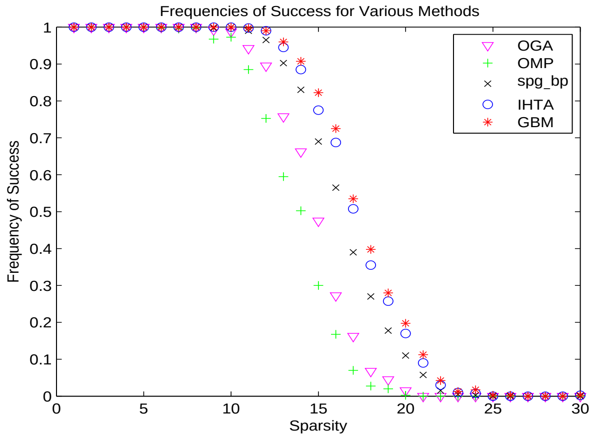

Fig. 1: Successful sparse recovery rates of the five algorithms.

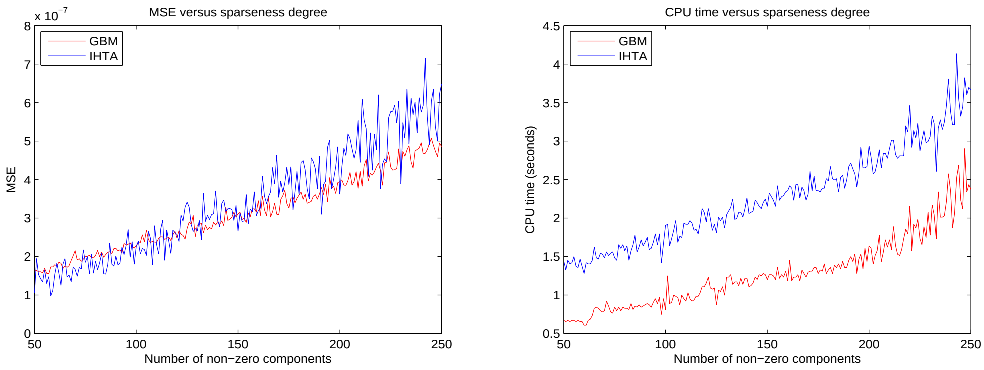

Fig. 2: MSE and CPU times for GBM and IHTA, as a function of the number of non-zero components of x_s.

the sparse solutions for each method in Fig. 1. From Fig. 1, we can see that GBM gives somewhat better success rates than the other four algorithms considered.

We then compared the computational efficiency of the GBM algorithm against the IHTA algorithm, which is regarded as a highly efficient method for the L_2-$L_{1/2}$ minimization problem. In the experiments, we fixed the matrix size ($m = 1024$ and $n = 4096$) and considered a range of degrees of sparseness: the number T of non-zeros spikes in x_s (randomly located values of ± 1) ranges from 50 to 250. For each value of T, we generated the random data set $(\mathrm{x_s}, \mathrm{A}, \mathrm{b})$ by (4.1). For each data set, we used GBM and IHTA to reconstruct the signal. We measured the quality of the reconstructed signal $\hat{x}$ using the mean-square error

$$\mathrm{MSE} = \|\hat{x} - x_s\|_2.$$

Fig. 2 plots the reconstruction MSE and the CPU time, as a function of the number of nonzero components in x_s. We observe that both methods obtain exact reconstructions for T from 50 up to 250. Concerning computational efficiency, our main focus in this experiment, we can observe that GBM performed faster than IHTA.

Prolem	GBM		IHTA	
$n = 4096, m = 1024$	MSE	time (seconds)	MSE	time (seconds)
$T = 50$	1.2025e-001	0.5623	1.5823e-001	1.3659
$T = 60$	1.7591e-001	0.5562	1.8044e-001	1.2997
$T = 70$	1.9563e-001	0.6014	2.1009e-001	1.3514
$T = 80$	2.0486e-001	0.5827	2.0778e-001	1.3383
$T = 90$	1.9742e-001	0.8591	1.9882e-001	1.7846
$T = 100$	2.2087e-001	0.8160	2.3341e-001	1.6472

Table 1: Results for signal reconstruction with noisy.

We then added noisy signals to the problem, i.e.,

$$\mathrm{b} = \mathrm{A} * \mathrm{x_s} - \mathrm{w},$$

where $w = \sigma\eta$ is independent identically distributed Gaussian noise with zero mean and variance $\sigma^2 = 10^{-4}$. We applied GBM and IHTA to reconstruct the signal. To compare both algorithms under noisy circumstance, we listed the MSE and the CPU time in Table 1. The results in Table 1 show the out-performance of the GBM algorithm, as compared with the IHTA algorithm.

We also compared our algorithm with the IALW algorithm. We fixed the degree of sparseness, i.e., $T = 10$, and considered a range of the matrix size. For each pair of (n, m), we generated the random data set $(\mathrm{x_s}, \mathrm{A}, \mathrm{b})$ by (4.1) and used GBM and IALW to reconstruct the signal. The results were listed in Table 2.

Prolem	GBM		IALW	
$T = 10$	MSE	time (seconds)	MSE	time (seconds)
$n = 256, m = 64$	8.5814e-008	0.0268	2.0438e-007	0.0468
$n = 512, m = 128$	9.3586e-008	0.0468	1.2125e-007	0.1560
$n = 1024, m = 256$	6.1938e-008	0.0468	7.9912e-008	0.7332
$n = 2048, m = 512$	6.3102e-008	0.1716	7.9845e-008	4.0872
$n = 4096, m = 1024$	7.1798e-008	0.4524	8.4210e-008	28.1738

Table 2: Results for signal reconstruction by GBM and IALW.

The results in Table 2 show that the GBM used much less CPU time to solve the L_2-$L_{1/2}$ minimization problem especially when $n = 2048, m = 512$ and $n = 4096, m = 1024$.

Acknowledgments

The authors would like to thank the two anonymous referees and the editor for their valuable suggestions and comments. We are grateful to Doctor Xiangyu Chang for providing us the code of the iterative half thresholding algorithm (IHTA).

References

[1] J. Barzilai and J. Borwein, Two point step size gradient methods, *IMA J. Numer. Anal.* 8 (1988) 141–148.

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

[3] E.G. Birgin, J.M. Martinez and M. Raydan, Nonmonotone spectral projected gradient methods on convex sets, *SIAM J. Optim.* 10 (2000) 1196–1211.

[4] A.M. Bruckstein, D.L. Donoho and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Rev.* 51 (2009) 34–81.

[5] E. Candès, J. Romberg and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory* 52 (2006) 489–509.

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

[7] R. Chartrand, Nonconvex regularization for shape preservation, *IEEE International Conference on Image Processing (ICIP)*, 2007.

[8] S.S. Chen, D.L. Donoho and M.A. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.* 20 (1998) 33–61.

[9] X. Chen and W. Zhou, Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization, *SIAM J. Imaging Sci.* 3 (2010) 765–790.

[10] 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 (2010) 2832–2852.

[11] I. Daubechies, M. De Friese and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Commun. Pure Appl. Math.* 57 (2004) 1413–1457.

[12] D.L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory*, 52 (2006) 1289–1306.

[13] M. Figueiredo, R. Nowak and S. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, *IEEE J. Sel. Top. Signa.* 1 (2007) 586–598.

[14] D. Ge, X. Jiang and Y. Ye, A note on the complexity of L_p minimization, *Math. Program., Ser. B* 129 (2011) 285–299.

[15] L. Grippo, F. Lampariello and S. Lucidi, A nonmonotone line search technique for Newton's method, *SIAM J. Numer. Anal.* 23 (1986) 707–716.

[16] E. Hale, W. Yin and Y. Zhang, Fixed-point continuation for ℓ_1-minimization: Methodology and convergence, *SIAM J. Optim.* 19 (2008) 1107–1130.

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

[18] M. Lai and J. Wang, An unconstrained L_q minimization with $0 < q \leq 1$ for sparse solution of under-determined linear systems, *SIAM J. Optim.* 21 (2011) 82–101.

[19] B.K. Natarajan, Sparse approximate solutions to linear systems, *SIAM J. Comput.* 24 (1995) 227–234.

[20] D. Needell, R. Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit, *Found. Comput. Math.* 9 (2009) 317–334.

[21] A. Petukhov, Fast implementation of orthogonal greedy algorithm for tight wavelet frames, *Signal Process* 86 (2006) 471–479.

[22] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM J. Optim.* 7 (1997) 26–33.

[23] M. Tao and X.M. Yuan, Recovering lower-rank and sparse components of matrices from incomplete and noisy observations, *SIAM J. Optim.* 21 (2011) 57–81.

[24] R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Stat. Soc., Ser. B* 58 (1996) 267–288.

[25] S. J. Wright, R. Nowak and M. Figueiredo, Sparse reconstruction by separable approximation, *IEEE Trans. Signal Process.* 57 (2009) 2479–2493.

[26] E. Van den Berg and M. Friedlander, Probing the pareto frontier for basis pursuit solutions, *SIAM J. Sci. Comput.* 31 (2008) 890–912.

[27] Z. Xu, H. Zhang, Y. Wang and X. Chang, $L_{1/2}$ regularizer, *Sci. China, Ser. F* 52 (2009) 1–9.

[28] Z. Xu, X. Chang, F. Xu and H. Zhang, $L_{1/2}$ regularization: A thresholding representation theory and a fast solver, *IEEE Trans. Neural Netw. Learn. Syst.* 23 (2012) 1013–1027.

[29] J.F. Yang and Y. Zhang, Alternating direction algorithms for ℓ_1-problems in compressive sensing, *SIAM J. Sci. Comput.* 33 (2011) 250–278.

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

[31] T. Zhang, Analysis of multi-stage convex relaxation for sparse regularization, *J. Mach. Learn. Res.* 11 (2010) 1087–1107.

Manuscript received 4 June 2012
revised 18 October 2012, 9 January 2013
accepted for publication 15 January 2013

Lei Wu
College of Mathematics and Information Science
Jiangxi Normal University, Nanchang, 330022 China
E-mail address: `wulei102@126.com`

Zhe Sun
College of Mathematics and Information Science
Jiangxi Normal University, Nanchang, 330022 China
E-mail address: `snzma@126.com`

Dong-Hui Li
School of Mathematical Sciences
South China Normal University, Guangzhou, 510631 China
E-mail address: `dhli@scnu.edu.cn`