J Sci Comput (2016) 67:581–601
DOI 10.1007/s10915-015-0094-4

A Barzilai–Borwein-Like Iterative Half Thresholding Algorithm for the $L_{1/2}$ Regularized Problem

Lei Wu[1] · **Zhe Sun**[1] · **Dong-Hui Li**[2]

Received: 16 July 2014 / Revised: 27 August 2015 / Accepted: 29 August 2015 /
Published online: 4 September 2015

Abstract In this paper, we propose a Barzilai–Borwein-like iterative half thresholding algorithm for the $L_{1/2}$ regularized problem. The algorithm is closely related to the iterative reweighted minimization algorithm and the iterative half thresholding algorithm. Under mild conditions, we verify that any accumulation point of the sequence of iterates generated by the algorithm is a first-order stationary point of the $L_{1/2}$ regularized problem. We also prove that any accumulation point is a local minimizer of the $L_{1/2}$ regularized problem when additional conditions are satisfied. Furthermore, we show that the worst-case iteration complexity for finding an ε scaled first-order stationary point is $O(\varepsilon^{-2})$. Preliminary numerical results show that the proposed algorithm is practically effective.

Keywords $L_{1/2}$ regularized problem · Sparse optimization · Prox-linear algorithms · Barzilai–Borwein steplength · Iterative half thresholding algorithm

Mathematics Subject Classification 65F22 · 90C26 · 90C06 · 49M05

The work was supported by the National Nature Science Foundation of P.R. China (No. 11501265, 11201197 and 11371154), the Nature Science Foundation of Jiangxi (No. 20132BAB211011), and the Foundation of Department of Education Jiangxi Province (No. GJJ13204).

✉ Zhe Sun
snzma@126.com

Lei Wu
wulei102@126.com

Dong-Hui Li
dhli@scnu.edu.cn

1 College of Mathematics and Information Science, Jiangxi Normal University, Nanchang, Jiangxi, People's Republic of China

2 School of Mathematical Sciences, South China Normal University, Guangzhou, Guangdong, People's Republic of China

1 Introduction

In this paper, we consider the following $L_{1/2}$ regularized problem

$$\min_{x\in R^n} F(x) \triangleq f(x) + \rho\|x\|_{1/2}^{1/2}, \tag{1.1}$$

where $\rho > 0$, $\|x\|_{1/2}^{1/2} = \sum_{i=1}^n |x_i|^{1/2}$, and f is bounded below and Lipschitz continuously differentiable in R^n, that is, there exist two constants L_{low} and $L_f > 0$ such that

$$f(x) \geq L_{\text{low}} \quad \text{and} \quad \|\nabla f(x) - \nabla f(y)\|_2 \leq L_f\|x - y\|_2, \qquad \forall x, y \in R^n. \tag{1.2}$$

The $L_{1/2}$ regularized problem is a nonconvex, nonsmooth and non-Lipschitz optimization problem, which has many applications in variable selection and compressed sensing [18,23]. When $f(x) = \|Ax - b\|_2^2/2$ with $A \in R^{m\times n}$ and $b \in R^m$ ($m < n$ or even $m \ll n$), problem (1.1) reduces to the following L_2–$L_{1/2}$ minimization problem

$$\min_{x\in R^n} F(x) \triangleq \frac{1}{2}\|Ax - b\|_2^2 + \rho\|x\|_{1/2}^{1/2}, \tag{1.3}$$

which is a penalty version of the following constrained $L_{1/2}$ minimization problem

$$\min_{x\in R^n} \|x\|_{1/2}^{1/2}, \qquad \text{s.t.} \qquad Ax = b.$$

Problem (1.3) is a special L_2–L_p $(0 < p < 1)$ minimization problem [5,6,15]. The purpose of solving this unconstrained (or constrained) optimization problem is to compute the sparse solution of underdetermined linear systems $Ax = b$ (see, e.g., [22,23], for details).

Because of nonconvexity and nonsmoothness, many efficient iterative algorithms for optimization problems cannot be directly applied to solve the $L_{1/2}$ regularized problem and the general L_p $(0 < p < 1)$ regularized problem. Recently, a great deal of effort was made to study iterative algorithms for solving this class of optimization problems (see, e.g., [3–6,15,17,18,21–23], for details).

Bian et al. [3] proposed first-order and second-order interior point algorithms to solve a class of nonsmooth, nonconvex, and non-Lipschitz minimization problems and proved that the worst-case iteration complexity for finding an ε scaled first-order stationary point is $O(\varepsilon^{-2})$ for first-order interior point algorithm and $O(\varepsilon^{-3/2})$ for second-order interior point algorithm. By the use of the first-order necessary condition of (1.3), Wu et al. [21] proposed a gradient based method to solve it and verified that the sequence of iterates converges to a stationary point of (1.3). Li et al. [17] proposed feasible direction algorithms to solve an equivalent smooth constrained reformulation of (1.1) and showed that the iterate sequence generated by the proposed algorithms converges to a stationary point of (1.1).

Iterative reweighted minimization algorithms are another class of important algorithms for solving the L_p $(0 < p < 1)$ regularized problem [4,6,15,18,23]. This class of algorithms was firstly proposed to solve the L_2–L_p minimization problem [4,6,15,16,23] and it consists of solving a sequence of weighted convex optimization problems. According to the type of unconstrained optimization problems solved at each iteration, they were called the iterative reweighted L_1 minimization algorithm (IRL$_1$) [4,6,18] and the iterative reweighted L_2 minimization algorithm (IRL$_2$) [15,16,18]. Lu [18] extended IRL$_1$ and IRL$_2$ algorithms and proposed a more general IRL$_\alpha$ algorithm to solve the L_p regularized problem. For the $L_{1/2}$ regularized problem (1.1), the IRL$_\alpha$ algorithm at each iteration solves a weighted optimization problem of the form:

$$x^{k+1} := \arg\min_{x \in R^n} f(x) + \rho \sum_{i=1}^{n} w_i^k |x_i|^{\alpha}, \tag{1.4}$$

where

$$w_i^k = \frac{1}{2\alpha}\left(|x_i^k|^{\alpha} + \xi\right)^{\frac{1}{2\alpha}-1}, \qquad \forall i = 1, 2, \ldots, n,$$

with $\xi > 0$ and $\alpha \geq 1$. When $\alpha = 1$ and $\alpha = 2$, respectively, the IRL$_\alpha$ algorithm reduces to the IRL$_1$ algorithm [4,6,18] and the IRL$_2$ algorithm [15,16,18]. It is verified [6,15,18] that the sequence $\{x^k\}$ converges to a stationary point of the approximation of the form

$$\min_{x \in R^n} F_{\alpha,\xi}(x) \triangleq f(x) + \rho \sum_{i=1}^{n} (|x_i|^{\alpha} + \xi)^{1/2\alpha}. \tag{1.5}$$

The iterative reweighted minimization algorithms need to solve a sequence of large-scale optimization problems. Moreover, they require ξ to be dynamically updated and approach zero (we refer the reader to [18] for details). This usually brings in lots of computational effort. In order to improve the computational efficiency, one prox-linear iteration was applied to approximately solve (1.4), which yielded the following variant [18]:

$$x^{k+1} := \arg\min_{x \in R^n} f\left(x^k\right) + \nabla f\left(x^k\right)^T\left(x - x^k\right) + \frac{\lambda_k}{2}\|x - x^k\|_2^2 + \rho \sum_{i=1}^{n} w_i^k |x_i|^{\alpha}, \tag{1.6}$$

where λ_k is chosen to ensure that the new iterate x^{k+1} satisfies the following inequality

$$F_{\alpha,\xi}\left(x^{k+1}\right) \leq F_{\alpha,\xi}\left(x^k\right) - \frac{c}{2}\|x^{k+1} - x^k\|_2^2,$$

with $c > 0$. We call the variant the accelerated iterative reweighted L_α minimization algorithm (abbreviated by AIRL$_\alpha$).

There is also an alternative way to derive (1.6). We first apply the prox-linear algorithm to (1.1) and obtain the iteration

$$x^{k+1} = \arg\min_{x \in R^n} f\left(x^k\right) + \nabla f\left(x^k\right)^T\left(x - x^k\right) + \frac{\lambda_k}{2}\|x - x^k\|_2^2 + \rho\|x\|_{1/2}^{1/2}. \tag{1.7}$$

The iteration (1.7) is equivalent to solving n one-dimensional L_2–$L_{1/2}$ minimization problems. Applying one iteration of the IRL$_\alpha$ algorithm to approximately solve these one-dimensional problems yields (1.6). In this way, the prox-linear algorithm can be regarded as outer iteration and the IRL$_\alpha$ algorithm is inner iteration.

The one-dimensional L_2–$L_{1/2}$ minimization problem has a closed-form solution [22]. Using the closed-form solution, Xu et al. [22] proposed an iterative half Thresholding algorithm (denoted by IHTA) to solve the L_2–$L_{1/2}$ minimization problem of the form (1.3). The IHTA iteration can be described as follows:

$$x^{k+1} = H_{\rho\mu_k,\frac{1}{2}}\left(x^k - \mu_k A^T\left(Ax^k - b\right)\right), \tag{1.8}$$

where μ_k is a steplength parameter and $H_{\alpha,\frac{1}{2}} : R^n \to R^n$ is called the half thresholding operator [22], defined by

$$\left(H_{\alpha,\frac{1}{2}}(x)\right)_i = \begin{cases} \frac{2}{3}x_i\left(1 + \cos\left(\frac{2}{3}\pi - \frac{2}{3}h(x_i,\alpha)\right)\right), & \text{if } |x_i| > \frac{3}{2}\left(\frac{\alpha}{2}\right)^{2/3}, \\ 0, & \text{otherwise,} \end{cases} \tag{1.9}$$

with

$$h(x_i, \alpha) = \arccos\left(\frac{\alpha}{8}\left(\frac{|x_i|}{3}\right)^{-3/2}\right).$$

The IHTA algorithm can be seen as an operator splitting algorithm. Based on the operator splitting technique, the iterative soft thresholding algorithm (denoted by ISTA) was also proposed to solve the L_1 regularized problem [7,9,13]. IHTA and ISTA algorithms require the steplength to be sufficiently small to guarantee convergence (see [7,9,13,22] for details). A small steplength usually results in slow convergence in practice (see, e.g., [2,10,14,20,21]). By combining the Barzilai–Borwein (BB) steplength [1] with the nonmonotone line search [11], accelerated ISTA methods were proposed to solve the L_1 regularized problem [14,20]. Numerical experiments show that the BB steplength significantly improve the speed of ISTA [14,20].

Motivated by [14,20], we shall propose a Barzilai–Borwein-like iterative half Thresholding algorithm (abbreviated by BBIHTA) to solve the $L_{1/2}$ regularized problem (1.1). We firstly apply the prox-linear algorithm to (1.1) and obtain the iteration (1.7), in which the λ_k is closely related to the BB steplength and is chosen such that the new iterate x^{k+1} satisfies the inequality

$$F\left(x^{k+1}\right) \leq \max_{0\leq j\leq \min(k,M)} F\left(x^{k-j}\right) - \gamma\|x^{k+1}-x^k\|_2^2,$$

where M is a given nonnegative integer, γ is a positive constant and F is defined by (1.1). Noting that the one-dimensional L_2–$L_{1/2}$ minimization problem has a closed-form solution, we then solve these one-dimensional L_2–$L_{1/2}$ minimization problems exactly. Under appropriate conditions, we study the convergence of the BBIHTA algorithm and show that the worst-case iteration complexity for finding an ε scaled first-order stationary point is $O(\varepsilon^{-2})$. Finally, we do some preliminary numerical experiments to show the efficiency of the proposed algorithm.

The remainder of the paper is organized as follows. In Sect. 2, we give some preliminaries and propose the BBIHTA algorithm to solve the $L_{1/2}$ regularized problem (1.1). In Sect. 3, we study the convergence of the proposed algorithm and verify that any accumulation point of the sequence of iterates is a first-order stationary point of (1.1) under mild conditions and is also a local minimizer when additional conditions are satisfied. In Sect. 4, we show that the worst-case iteration complexity for finding an ε scaled first-order stationary point is $O(\varepsilon^{-2})$. Finally in Sect. 5, we shall numerically compare the performance of the BBIHTA algorithm with that of the other two algorithms.

2 Preliminaries and the Algorithm

The BBIHTA algorithm is based on the iteration (1.7). It is readily seen that the iteration (1.7) can be equivalently reformulated as

$$x^{k+1} := \arg\min_{x\in R^n}\left\|x - \left(x^k - \frac{\nabla f(x^k)}{\lambda_k}\right)\right\|_2^2 + \frac{2\rho}{\lambda_k}\|x\|_{1/2}^{1/2}. \tag{2.1}$$

For simplicity, we let, in the latter part of the paper,

$$y^k = x^k - \nabla f\left(x^k\right)/\lambda_k \quad \text{and} \quad \beta_k = 2\rho/\lambda_k. \tag{2.2}$$

Then (2.1) can be decomposed into n one-dimensional L_2–$L_{1/2}$ minimization problems

$$x_i^{k+1} = \arg\min_{t \in R} \left(t - y_i^k\right)^2 + \beta_k |t|^{1/2}, \quad i = 1, 2, \cdots, n, \tag{2.3}$$

where x_i^{k+1} and y_i^k are, respectively, the ith components of x^{k+1} and y^k. Since the optimization problem (2.3) is nonconvex, it may have more than one local minimizer. The following lemma shows this.

Lemma 2.1 *The optimization problem of the form (2.3) has at most two local minimizers.*

Proof Without loss of generality, let us assume that $y_i^k > 0$. Define $\varphi : R \to R$ by

$$\varphi(t) \equiv \left(t - y_i^k\right)^2 + \beta_k |t|^{1/2}. \tag{2.4}$$

Then $\varphi(t)$ is a monotonically decreasing function of t on the interval $(-\infty, 0]$. Therefore, the optimization problem (2.3) has no local minimizer on the interval $(-\infty, 0)$. Let t^* be a local minimizer. Then $t^* = 0$ or it satisfies

$$t^* > 0 \quad \text{and} \quad \varphi'\left(t^*\right) = 2\left(t^* - y_i^k\right) + \beta_k \left(t^*\right)^{-1/2}/2 = 0. \tag{2.5}$$

From (2.5) we deduce that if $t^* > 0$, $(t^*)^{1/2}$ must be a root of the following equation

$$4t\left(t^2 - y_i^k\right) + \beta_k = 0.$$

Let t_i $(i = 1, 2, 3)$ be the solutions of the equation above and satisfy $t_1 \le t_2 \le t_3$. Then we get $t_1 t_2 t_3 = -\beta_k/4$ and $t_1 + t_2 + t_3 = 0$, which imply that $t_1 < 0$ and $t_2 > 0$. Therefore, $\varphi'((t_2)^2) = 0$ and $\varphi'((t_3)^2) = 0$. Note that $\lim_{t \to 0+} \varphi'(t) = +\infty$. By continuity of $\varphi'(t)$, we deduce that $\varphi'(t) > 0$ in the interval $(0, (t_2)^2)$, which implies that the function $\varphi(t)$ is monotonically increasing in this interval. Hence, $(t_2)^2$ is not a local minimizer of the optimization problem (2.3) and this problem has at most one local minimizer in the interval $(0, +\infty)$. On the other hand, since $\varphi(t)$ is monotonically decreasing on the interval $(-\infty, 0]$ and is monotonically increasing in the interval $(0, (t_2)^2)$, 0 is a local minimizer of the optimization problem (2.3).

The discussion above shows that the optimization problem of the form (2.3) has at most two local minimizers. The proof is complete. □

Generally, it is difficult to find an optimal solution (i.e., a global minimizer) of a nonconvex optimization problem. Xu et al. [22] proved that the optimization problem of the form (2.3) has a closed-form solution and derived its expression. We introduce the following two lemmas to show this. For completeness, we provide a simpler proof for Lemma 2.2.

Lemma 2.2 *Let $x_i^{k+1} \in R$ be an optimal solution of problem (2.3). Then $x_i^{k+1} \neq 0$ if and only if the following inequality holds*

$$\left|y_i^k\right| > \frac{3}{2}\left(\frac{\beta_k}{2}\right)^{2/3}.$$

Proof Necessary. Without loss of generality, let us assume that $x_i^{k+1} > 0$. Then from (2.3) we derive that

$$\left(x_i^{k+1} - y_i^k\right)^2 + \beta_k |x_i^{k+1}|^{1/2} < \left(y_i^k\right)^2. \tag{2.6}$$

Moreover, by the first-order necessary condition of (2.3) we know that x_i^{k+1} satisfies the following equality

$$2\left(x_i^{k+1} - y_i^k\right) + \frac{\beta_k}{2\left(x_i^{k+1}\right)^{1/2}} = 0.$$

This yields

$$y_i^k = x_i^{k+1} + \frac{\beta_k}{4\left(x_i^{k+1}\right)^{1/2}}. \tag{2.7}$$

Substituting (2.7) into (2.6), we get

$$x_i^{k+1} > \left(\frac{\beta_k}{2}\right)^{2/3}.$$

It is not difficult to verify that the function $\phi(t) \equiv t + \beta_k t^{-1/2}/4$ is monotonically increasing when $t \geq (\beta_k/2)^{2/3}$. Therefore, we deduce from (2.7) that

$$y_i^k > \frac{3}{2}\left(\frac{\beta_k}{2}\right)^{2/3}. \tag{2.8}$$

Sufficiency. Without loss of generality, we assume that (2.8) holds. Let

$$\bar{t} \equiv \left(\frac{\beta_k}{2}\right)^{2/3}. \tag{2.9}$$

Then it holds

$$\varphi\left(x_i^{k+1}\right) \leq \varphi(\bar{t}) = \left(\bar{t} - y_i^k\right)^2 + \beta_k \bar{t}^{1/2} = \left(y_i^k\right)^2 + \bar{t}^2 - 2y_i^k \bar{t} + \beta_k \bar{t}^{1/2}, \tag{2.10}$$

where φ is defined by (2.4). By combining (2.4), (2.8) and (2.9), we obtain

$$\bar{t}^2 - 2y_i^k \bar{t} + \beta_k \bar{t}^{1/2} = \bar{t}^{1/2}\left(\bar{t}^{3/2} - 2y_i^k \bar{t}^{1/2} + \beta_k\right) < \left(\frac{\beta_k}{2}\right)^{2/3}\left(\frac{\beta_k}{2} - \frac{3\beta_k}{2} + \beta_k\right) = 0.$$

This together with (2.10) yields

$$\varphi\left(x_i^{k+1}\right) < \left(y_i^k\right)^2 = \varphi(0).$$

Hence, $x_i^{k+1} \neq 0$. The proof is complete. □

The following lemma is from [22], which gives an explicit expression for the optimal solution of problem (2.3).

Lemma 2.3 *For any $i = 1, 2, \ldots, n$, define*

$$x_i^{k+1} \triangleq \left(H_{\beta_k, \frac{1}{2}}\left(y^k\right)\right)_i, \tag{2.11}$$

where y^k and β_k are defined by (2.2) and $(H_{\beta_k, \frac{1}{2}}(y^k))_i$ is calculated according to (1.9) with $\alpha = \beta_k$. Then x_i^{k+1} is an optimal solution of problem (2.3).

Now we give a first-order necessary condition for optimality for problem (1.1). Let $x^* = (x_1^*, x_2^*, \ldots, x_n^*)^T \in R^n$ be a local minimizer of (1.1). For any $i = 1, 2, \ldots, n$, if $x_i^* \neq 0$, then simple calculation shows that

$$\nabla_i f\left(x^*\right) + \rho \frac{\operatorname{sgn}\left(x_i^*\right)}{2|x_i^*|^{1/2}} = 0,$$

that is,

$$2|x_i^*|^{1/2} \nabla_i f\left(x^*\right) + \rho \operatorname{sgn}\left(x_i^*\right) = 0,$$

where $\nabla_i f(x)$ denotes the ith component of $\nabla f(x)$ and sgn : $R \to \{-1, 0, 1\}$ is defined by

$$\operatorname{sgn}(t) = \begin{cases} 1, & t > 0, \\ 0, & t = 0, \\ -1, & t < 0. \end{cases}$$

Define a mapping $\Psi(x) = (\Psi_1(x), \Psi_2(x), \ldots, \Psi_n(x))^T : R^n \to R^n$ with

$$\Psi_i(x) = \operatorname{mid}\left\{2|x_i|^{1/2}\nabla_i f(x) - \rho, \quad x_i, \quad 2|x_i|^{1/2}\nabla_i f(x) + \rho\right\}, \quad i = 1, 2, \ldots, n, \tag{2.12}$$

where mid{a, b, c} takes the medium value among three variables a, b, c. With the notions above, it is not difficult to verify the following proposition. The proof is omitted here.

Proposition 2.1 *If $x^* \in R^n$ is a local minimizer of problem (1.1), then $\Psi(x^*) = 0$.*

We call a vector $x \in R^n$ a first-order stationary point if it satisfies that $\Psi(x) = 0$. From Proposition 2.1 it follows that every local minimizer of problem (1.1) is a first-order stationary point. In [5,18], similar concepts and properties were also presented.

We now propose the Barzilai–Borwein-like iterative half thresholding algorithm with a nonmonotone line search strategy [11] to solve problem (1.1).

Algorithm 2.1 (Barzilai–Borwein-Like Iterative Half Thresholding (BBIHTA) Algorithm) Given $x^0 \in R^n$, $0 < \lambda_{\min} < \lambda_{\max} < +\infty$, $\bar{\lambda} \in [\lambda_{\min}, \lambda_{\max}]$, $\gamma \in (0, 1)$, $1 < \sigma_1 < \sigma_2 < +\infty$, and integer $M \geq 0$. Set $k := 0$.

Step 1. If $\|\Psi(x^k)\|_2 = 0$ stop.
Step 2. Set $\lambda_k \leftarrow \bar{\lambda}$.
Step 3. Compute y^k and β_k by

$$y^k = x^k - \nabla f\left(x^k\right)/\lambda_k \qquad \text{and} \qquad \beta_k = 2\rho/\lambda_k.$$

Step 4. For each $i = 1, 2, \ldots, n$, compute x_i^{k+1} by (2.11) and set

$$x^{k+1} = \left(x_1^{k+1}, x_2^{k+1}, \ldots, x_n^{k+1}\right)^T.$$

Step 5. (nonmonotone line search)
If

$$F\left(x^{k+1}\right) \leq \max_{0 \leq j \leq \min(k,M)} F\left(x^{k-j}\right) - \gamma\|x^{k+1} - x^k\|_2^2, \tag{2.13}$$

then set $s^k = x^{k+1} - x^k$, $d^k = \nabla f(x^{k+1}) - \nabla f(x^k)$ and go to Step 7; otherwise, go to Step 6.

Step 6. Choose

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

set $\lambda_k \leftarrow \lambda_{\text{new}}$, and go to Step 3.
Step 7. Compute $p^k = (s^k)^T d^k$ and $q^k = (s^k)^T s^k$. Define $\lambda_k^{BB} = p^k / q^k$ and

$$\bar{\lambda} = \min\{\lambda_{\max}, \max\{\lambda_{\min}, \lambda_k^{BB}\}\}. \tag{2.14}$$

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

Remark 2.1 (i) The λ_k^{BB} in Step 7 is the BB steplength and is introduced by Barzilai and Borwein in [1]. The BB gradient method possesses the R-superlinear convergence for two-dimensional convex quadratics. Raydan [19] extended the method to unconstrained optimization by incorporating the nonmonotone line search [11]. The resulting method is globally convergent and is competitive with some standard conjugate gradient codes. Due to the simplicity and efficiency, the BB-like methods have received much attentions in finding sparse approximation solutions to large underdetermined linear systems of equations from signal/image processing [12,14,20]. However, it is difficult to analyze the convergence rate of the BB-like methods even if the objective function is convex [8]. The object of (2.14) is to keep $\bar{\lambda}$ uniformly bounded and this $\bar{\lambda}$ is always accepted by the non-monotone line search.

(ii) From the previous discussion, we know that x^{k+1} generated by (2.11) satisfies

$$x^{k+1} = \arg\min_{x \in R^n} f(x^k) + \nabla f\left(x^k\right)^T \left(x - x^k\right) + \frac{\lambda_k}{2}\|x - x^k\|_2^2 + \rho\|x\|_{1/2}^{1/2}.$$

(iii) Denote

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

Since f is bounded below, it is not difficult to verify that Ω_0 is a closed and bounded set. That is, there exists a constant $\beta > 0$ such that

$$\|x\|_2 < \beta, \qquad \forall x \in \Omega_0.$$

Moreover, the sequence $\{x^k\}$ generated by Algorithm 2.1 is contained in Ω_0.

Proposition 2.2 *The BBIHTA algorithm cannot cycle indefinitely between Step 3 and Step 6.*

Proof Since f is Lipschitz continuously differentiable in R^n, it holds

$$\begin{aligned} F(x^{k+1}) &= f\left(x^{k+1}\right) + \rho\|x^{k+1}\|_{1/2}^{1/2} \\ &= f\left(x^k\right) + \int_0^1 \nabla f\left(x^k + t\left(x^{k+1} - x^k\right)\right)^T \left(x^{k+1} - x^k\right) dt + \rho\|x^{k+1}\|_{1/2}^{1/2} \\ &= f\left(x^k\right) + \nabla f\left(x^k\right)^T \left(x^{k+1} - x^k\right) \\ &\quad + \int_0^1 \left[\nabla f\left(x^k + t\left(x^{k+1} - x^k\right)\right) - \nabla f\left(x^k\right)\right]^T \left(x^{k+1} - x^k\right) dt + \rho\|x^{k+1}\|_{1/2}^{1/2}. \end{aligned} \tag{2.15}$$

From (1.2) it follows

$$\left|\left[\nabla f\left(x^k + t\left(x^{k+1} - x^k\right)\right) - \nabla f\left(x^k\right)\right]^T \left(x^{k+1} - x^k\right)\right| \leq t L_f \|x^{k+1} - x^k\|_2^2,$$

which together with (2.15) implies

$$F\left(x^{k+1}\right) \leq f\left(x^k\right) + \nabla f\left(x^k\right)^T\left(x^{k+1} - x^k\right) + L_f\|x^{k+1} - x^k\|_2^2 \int_0^1 t dt + \rho\|x^{k+1}\|_{1/2}^{1/2}$$
$$= f\left(x^k\right) + \nabla f\left(x^k\right)^T\left(x^{k+1} - x^k\right) + \frac{L_f}{2}\|x^{k+1} - x^k\|_2^2 + \rho\|x^{k+1}\|_{1/2}^{1/2}. \tag{2.16}$$

Assume that $\lambda_k \geq L_f + 2\gamma$. Then by (2.16) we have

$$F\left(x^{k+1}\right) \leq f\left(x^k\right) + \nabla f\left(x^k\right)^T\left(x^{k+1} - x^k\right)$$
$$+\frac{\lambda_k}{2}\|x^{k+1} - x^k\|_2^2 + \rho\|x^{k+1}\|_{1/2}^{1/2} - \gamma\|x^{k+1} - x^k\|_2^2. \tag{2.17}$$

From (ii) of Remark 2.1, it follows

$$f\left(x^k\right) + \nabla f\left(x^k\right)^T\left(x^{k+1} - x^k\right) + \frac{\lambda_k}{2}\|x^{k+1} - x^k\|_2^2 + \rho\|x^{k+1}\|_{1/2}^{1/2}$$
$$\leq f\left(x^k\right) + \rho\|x^k\|_{1/2}^{1/2} = F\left(x^k\right). \tag{2.18}$$

Combining (2.17) and (2.18), we obtain

$$F\left(x^{k+1}\right) \leq F\left(x^k\right) - \gamma\|x^{k+1} - x^k\|_2^2 \leq \max_{0\leq j\leq \min(k,M)} F\left(x^{k-j}\right) - \gamma\|x^{k+1} - x^k\|_2^2.$$

Therefore, when $\lambda_k \geq L_f + 2\gamma$, the x^{k+1} generated by (2.11) satisfies (2.13).

At kth iteration, we get, by Steps 2, 6 and 7, $\lambda_k = \bar{\lambda}\sigma_1^l \geq \lambda_{\min}\sigma_1^l$, after l cycles between Step 3 and Step 6. Since $\sigma_1 > 1$, it holds $\lambda_k \geq L_f + 2\gamma$ when $l \geq \frac{\log(L_f+2\gamma)-\log(\lambda_{\min})}{\log(\sigma_1)}$. Therefore, by the discussion above, we deduce that the BBIHTA algorithm cannot cycle indefinitely between Step 3 and Step 6. The proof is complete. □

The following proposition shows that the sequence $\{\lambda_k\}$ generated by Algorithm 2.1 is bounded. This proposition shall be used to prove Theorem 3.1 and Lemma 4.1.

Proposition 2.3 *Let $\{\lambda_k\}$ be a sequence generated by Algorithm 2.1. Then the following inequalities hold*

$$\lambda_{\min} \leq \lambda_k \leq \max\{\lambda_{\max}, \sigma_2(L_f + 2\gamma)\}, \qquad \forall k = 0, 1, \ldots.$$

Proof For any $k = 0, 1, \ldots$, by Algorithm 2.1 we obtain $\lambda_k \geq \bar{\lambda} \geq \lambda_{\min}$. If $\lambda_k = \bar{\lambda}$, by (2.14) we get $\lambda_k \leq \lambda_{\max}$. Otherwise, from Step 6 of Algorithm 2.1 and the proof of Proposition 2.2 it follows that

$$\lambda_k/\sigma_2 < L_f + 2\gamma.$$

This means that $\lambda_k < \sigma_2(L_f+2\gamma)$. Therefore, for any $k = 0, 1, \ldots$, the following inequalities hold

$$\lambda_{\min} \leq \lambda_k \leq \max\{\lambda_{\max}, \sigma_2(L_f + 2\gamma)\}.$$

The proof is complete. □

3 The Convergence of the BBIHTA Algorithm

In this section, we shall study the convergence of the BBIHTA algorithm. To this end, we first show the following lemma.

Lemma 3.1 *Suppose that the sequence* $\{x^k\}$ *generated by Algorithm 2.1 is infinite. Then we have*

$$\lim_{k\to\infty} \|x^{k+1} - x^k\|_2 = 0.$$

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

$$k - \min(k, M) \le l(k) \le k \quad \text{and} \quad F\left(x^{l(k)}\right) = \max_{0\le j\le \min(k,M)} F\left(x^{k-j}\right). \tag{3.1}$$

Then, we have

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

This together with (2.13) implies

$$F\left(x^{l(k+1)}\right) \le F\left(x^{l(k)}\right).$$

Hence, the sequence $\{F(x^{l(k)})\}$ is nonincreasing. Moreover we obtain from (2.13) for any $k > M$,

$$\begin{aligned} F\left(x^{l(k)}\right) &\le \max_{0\le j\le \min(l(k)-1,M)} F\left(x^{l(k)-1-j}\right) - \gamma\|x^{l(k)} - x^{l(k)-1}\|_2^2 \\ &= F\left(x^{l(l(k)-1)}\right) - \gamma\|x^{l(k)} - x^{l(k)-1}\|_2^2. \end{aligned} \tag{3.2}$$

Since f is bounded below, the function F is also bounded below. Consequently the sequence $\{F(x^{l(k)})\}$ admits a limit. This together with (3.2) implies

$$\lim_{k\to\infty} \|x^{l(k)} - x^{l(k)-1}\|_2 = 0. \tag{3.3}$$

In what follows, we shall prove that $\lim_{k\to\infty} \|x^{k+1} - x^k\|_2 = 0$. First we show, by induction, that for any given $j = 1, 2, \ldots, M+1$,

$$\lim_{k\to\infty} \|x^{l(k)-j+1} - x^{l(k)-j}\|_2 = 0 \tag{3.4}$$

and

$$\lim_{k\to\infty} F\left(x^{l(k)-j}\right) = \lim_{k\to\infty} F\left(x^{l(k)}\right). \tag{3.5}$$

(We assume, for the remainder of the proof, that the iteration index k is large enough to avoid the occurrence of negative subscripts.) If $j = 1$, (3.4) follows from (3.3). Moreover, (3.4) implies that (3.5) holds for $j = 1$ since $F(x)$ is uniformly continuous on Ω_0. Assume now that (3.4) and (3.5) hold for a given j. Then by (2.13) we get

$$F\left(x^{l(k)-j}\right) \le F\left(x^{l(l(k)-j-1)}\right) - \gamma\|x^{l(k)-j} - x^{l(k)-j-1}\|_2^2.$$

Taking limits in both sides of the last inequality as $k \to \infty$, we have by (3.5)

$$\lim_{k\to\infty} \|x^{l(k)-j} - x^{l(k)-j-1}\|_2 = 0,$$

which implies $\left\{\|x^{l(k)-j} - x^{l(k)-(j+1)}\|_2\right\} \to 0$. By (3.5) and the uniform continuity of F on Ω_0, we get

$$\lim_{k\to\infty} F\left(x^{l(k)-(j+1)}\right) = \lim_{k\to\infty} F\left(x^{l(k)-j}\right) = \lim_{k\to\infty} F\left(x^{l(k)}\right).$$

By the principle of induction, we have proved (3.4) and (3.5) for any given $j = 1, 2, \cdots, M+1$.

Now for any k, it holds that

$$x^{k+1} = x^{l(k+2+M)} - \sum_{j=1}^{l(k+2+M)-(k+1)} \left(x^{l(k+2+M)-j+1} - x^{l(k+2+M)-j}\right). \tag{3.6}$$

By (3.1), we have $l(k+2+M) - (k+1) \le k+2+M-(k+1) = M+1$ and $l(k+2+M)-(k+1) \ge k+2+M-M-(k+1) = 1$. It then follows from (3.4) and (3.6) that

$$\lim_{k\to\infty} \|x^{k+1} - x^{l(k+2+M)}\| = 0.$$

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

$$\lim_{k\to\infty} F(x^k) = \lim_{k\to\infty} F\left(x^{l(k)}\right). \tag{3.7}$$

By (2.13), we have

$$F(x^{k+1}) \le F(x^{l(k)}) - \gamma\|x^{k+1} - x^k\|_2^2.$$

Taking limits in both sides of the last inequality as $k \to \infty$, we obtain by (3.7)

$$\lim_{k\to\infty} \|x^{k+1} - x^k\|_2 = 0.$$

The proof is complete. □

The convergence of the BBIHTA algorithm is stated in the following theorem. This theorem shows that every accumulation point of the sequence of iterates generated by the BBIHTA algorithm is a first-order stationary point of (1.1).

Theorem 3.1 *Suppose that the sequence* $\{x^k\}$ *is generated by Algorithm 2.1. Then either* $\Psi(x^j) = 0$ *for some finite* j*, or every accumulation point* $\bar{x}$ *of* $\{x^k\}$ *satisfies* $\Psi(\bar{x}) = 0$.

Proof When $\{x^k\}$ is finite, from the termination condition, it is clear that there exists some finite j satisfying $\Psi(x^j) = 0$. Suppose that $\{x^k\}$ is infinite. Then by Lemma 3.1 we have

$$\lim_{k\to\infty} \|x^{k+1} - x^k\|_2 = 0. \tag{3.8}$$

Let $\bar{x}$ be an arbitrary accumulation point of $\{x^k\}$. Then there exists a subsequence $\{x^{k_l}\} \subset \{x^k\}$ such that

$$\lim_{l\to+\infty} x^{k_l} = \bar{x}. \tag{3.9}$$

For any i with $\bar{x}_i = 0$, by (2.12) we get

$$\Psi_i(\bar{x}) = 0.$$

For any i with $\bar{x}_i \neq 0$, by (3.9) we deduce that $x_i^{k_l} \neq 0$ for sufficiently large l. Without loss of generality, let us assume that $\bar{x}_i > 0$. Then it holds that $x_i^{k_l} > 0$ for sufficiently large l.

Note that $x_i^{k_l}$ is an optimal solution of (2.3). By the first-order necessary condition of (2.3), we know that for sufficiently large l, $x_i^{k_l}$ satisfies

$$2\left(x_i^{k_l}-y_i^{k_l-1}\right)+\frac{\beta_{k_l-1}}{2\left(x_i^{k_l}\right)^{1/2}}=0,$$

where $y_i^{k_l-1}=x_i^{k_l-1}-\nabla_i f(x^{k_l-1})/\lambda_{k_l-1}$ and $\beta_{k_l-1}=2\rho/\lambda_{k_l-1}$. Therefore, it holds

$$x_i^{k_l}-x_i^{k_l-1}+\frac{\nabla_i f\left(x^{k_l-1}\right)}{\lambda_{k_l-1}}+\frac{\rho}{2\lambda_{k_l-1}\left(x_i^{k_l}\right)^{1/2}}=0,$$

which implies

$$2\left(x_i^{k_l}\right)^{1/2}\nabla_i f\left(x^{k_l-1}\right)+\rho=-2\lambda_{k_l-1}\left(x_i^{k_l}\right)^{1/2}\left(x_i^{k_l}-x_i^{k_l-1}\right). \tag{3.10}$$

By Proposition 2.3 we know that the sequence $\{\lambda_{k-1}\}$ is bounded. From (3.8) and (3.9) we deduce that

$$\lim_{l\to+\infty}\left(x_i^{k_l}-x_i^{k_l-1}\right)=0 \quad \text{and} \quad \lim_{l\to+\infty}x^{k_l-1}=\bar{x}.$$

So using these facts together with (3.10) and (3.9), we obtain

$$\begin{aligned}2(\bar{x}_i)^{1/2}\nabla_i f(\bar{x})+\rho &= \lim_{l\to+\infty}\left(2\left(x_i^{k_l}\right)^{1/2}\nabla_i f\left(x^{k_l-1}\right)+\rho\right)\\ &= -\lim_{l\to+\infty}\left(2\lambda_{k_l-1}\left(x_i^{k_l}\right)^{1/2}\left(x_i^{k_l}-x_i^{k_l-1}\right)\right)=0.\end{aligned}$$

This together with (2.12) and the assumption $\bar{x}_i>0$ implies

$$\Psi_i(\bar{x})=0.$$

The discussion above shows

$$\Psi_i(\bar{x})=0, \qquad \forall i.$$

This completes the proof. □

Remark 3.1 Since the $L_{1/2}$ regularized problem (1.1) is a nonconvex optimization problem, its first-order stationary point may not be a local minimizer. To see this, consider the following one-dimensional $L_{1/2}$ regularized problem

$$\min_{t\in R}\frac{1}{3}t^3+2|t|^{1/2}.$$

Simple calculation shows that the point -1 is a first-order stationary point of the problem above but not a minimizer.

The following theorem gives a sufficient condition for an accumulation point $\bar{x}$ of $\{x^k\}$ to be a local minimizer of (1.1). Before giving the theorem, we introduce some notations on submatrix and subvector. Given index sets $I\subset\{1,2,\ldots,n\}$ and $J\subset\{1,2,\ldots,n\}$, $\nabla^2_{IJ}f(\bar{x})$ denotes the submatrix of the Hessian matrix $\nabla^2 f(\bar{x})$ consisting of rows and columns indexed by I and J respectively, $\nabla_I f(\bar{x})$ denotes the subvector of the gradient $\nabla f(\bar{x})$ consisting of components indexed by I and x_I denotes the subvector of x consisting of components indexed by I.

Theorem 3.2 *Let f be twice continuously differentiable. Assume that $\{x^k\}$ is a sequence of iterates generated by Algorithm 2.1 and $\bar{x}$ is an accumulation point of $\{x^k\}$. If it holds $\bar{I}_A = \emptyset$, or $\bar{I}_A \neq \emptyset$ and*

$$\sigma\left(\nabla^2_{\bar{I}_A\bar{I}_A} f(\bar{x})\right) > \frac{1}{4}\rho \max_{i\in\bar{I}_A}\{|\bar{x}_i|^{-3/2}\}, \tag{3.11}$$

then $\bar{x}$ is a local minimizer of the $L_{1/2}$ regularized problem (1.1). Here, $\bar{I}_A := \{i \mid \bar{x}_i \neq 0\}$, and $\sigma(\nabla^2_{\bar{I}_A\bar{I}_A} f(\bar{x}))$ denotes the smallest eigenvalue of $\nabla^2_{\bar{I}_A\bar{I}_A} f(\bar{x})$.

Proof For any $d \in R^n$ sufficiently small, it holds that

$$\begin{aligned} F(\bar{x}+d) - F(\bar{x}) &= \nabla f(\bar{x})^T d + \frac{1}{2}\int_0^1 d^T \nabla^2 f(\bar{x}+td)d dt + \rho\left(\|\bar{x}+d\|_{1/2}^{1/2} - \|\bar{x}\|_{1/2}^{1/2}\right) \\ &= \frac{1}{2}d^T\nabla^2 f(\bar{x})d + \nabla f(\bar{x})^T d + \rho\sum_{i=1}^n \left(|\bar{x}_i+d_i|^{1/2} - |\bar{x}_i|^{1/2}\right) + o\left(\|d\|_2^2\right) \\ &= \frac{1}{2}d^T\nabla^2 f(\bar{x})d + \sum_{i\in\bar{I}_A}\left(\frac{\partial f(\bar{x})}{\partial x_i} + \frac{\rho \operatorname{sgn}(\bar{x}_i)}{|\bar{x}_i+d_i|^{1/2}+|\bar{x}_i|^{1/2}}\right)d_i \\ &\quad + \sum_{i\in\bar{I}_C}\left(\frac{\partial f(\bar{x})}{\partial x_i}d_i + \rho|d_i|^{1/2}\right) + o\left(\|d\|_2^2\right), \end{aligned} \tag{3.12}$$

where $\bar{I}_C := \{i \mid \bar{x}_i = 0\}$.

Assuming that $\bar{I}_A = \emptyset$, by (3.12) we have

$$F(\bar{x}+d) - F(\bar{x}) = \rho\|d\|_{1/2}^{1/2} + \nabla f(\bar{x})^T d + \frac{1}{2}d^T\nabla^2 f(\bar{x})d + o\left(\|d\|_2^2\right),$$

which implies that for any d sufficiently small, it holds

$$F(\bar{x}+d) > F(\bar{x}).$$

Therefore, $\bar{x}$ is a local minimizer of the $L_{1/2}$ regularized problem (1.1).

Assume that $\bar{I}_A \neq \emptyset$. Since $\bar{x}$ is an accumulation point of $\{x^k\}$, by Theorem 3.1 and (2.12) we get

$$\frac{\partial f(\bar{x})}{\partial x_i} + \frac{\rho \operatorname{sgn}(\bar{x}_i)}{2|\bar{x}_i|^{1/2}} = 0, \qquad \forall i \in \bar{I}_A.$$

This together with (3.12) implies that

$$\begin{aligned} F(\bar{x}+d) - F(\bar{x}) &= \frac{1}{2}d^T\nabla^2 f(\bar{x})d + \sum_{i\in\bar{I}_A}\left(\frac{\rho \operatorname{sgn}(\bar{x}_i)}{|\bar{x}_i+d_i|^{1/2}+|\bar{x}_i|^{1/2}} - \frac{\rho \operatorname{sgn}(\bar{x}_i)}{2|\bar{x}_i|^{1/2}}\right)d_i \\ &\quad + \sum_{i\in\bar{I}_C}\left(\frac{\partial f(\bar{x})}{\partial x_i}d_i + \rho|d_i|^{1/2}\right) + o\left(\|d\|_2^2\right). \end{aligned} \tag{3.13}$$

Simple calculation shows that

$$\begin{aligned} d^T\nabla^2 f(\bar{x})d &= d_{\bar{I}_A}^T \nabla^2_{\bar{I}_A\bar{I}_A} f(\bar{x})d_{\bar{I}_A} + d_{\bar{I}_A}^T \nabla^2_{\bar{I}_A\bar{I}_C} f(\bar{x})d_{\bar{I}_C} + d_{\bar{I}_C}^T \nabla^2_{\bar{I}_C\bar{I}_A} f(\bar{x})d_{\bar{I}_A} \\ &\quad + d_{\bar{I}_C}^T \nabla^2_{\bar{I}_C\bar{I}_C} f(\bar{x})d_{\bar{I}_C} \\ &= d_{\bar{I}_A}^T \nabla^2_{\bar{I}_A\bar{I}_A} f(\bar{x})d_{\bar{I}_A} + g^T d_{\bar{I}_C}, \end{aligned} \tag{3.14}$$

where g is a bounded vector. For any $i \in \bar{I}_A$, it holds

$$\begin{aligned}
&\frac{\rho \operatorname{sgn}(\bar{x}_i)}{|\bar{x}_i + d_i|^{1/2} + |\bar{x}_i|^{1/2}} - \frac{\rho \operatorname{sgn}(\bar{x}_i)}{2|\bar{x}_i|^{1/2}} = \frac{\rho \operatorname{sgn}(\bar{x}_i)(|\bar{x}_i|^{1/2} - |\bar{x}_i + d_i|^{1/2})}{2|\bar{x}_i|^{1/2}(|\bar{x}_i + d_i|^{1/2} + |\bar{x}_i|^{1/2})} \\
&= \frac{-\rho d_i}{2|\bar{x}_i|^{1/2}(|\bar{x}_i + d_i|^{1/2} + |\bar{x}_i|^{1/2})^2} \\
&= \frac{-\rho d_i}{8|\bar{x}_i|^{3/2}} + \frac{\rho d_i}{2|\bar{x}_i|^{1/2}} \left(\frac{(|\bar{x}_i + d_i|^{1/2} + |\bar{x}_i|^{1/2})^2 - 4|\bar{x}_i|}{4|\bar{x}_i|(|\bar{x}_i + d_i|^{1/2} + |\bar{x}_i|^{1/2})^2} \right) \\
&= \frac{-\rho d_i}{8|\bar{x}_i|^{3/2}} + \frac{\rho d_i}{2|\bar{x}_i|^{1/2}} \left(\frac{|\bar{x}_i + d_i| + 2|\bar{x}_i + d_i|^{1/2}|\bar{x}_i|^{1/2} - 3|\bar{x}_i|}{4|\bar{x}_i|(|\bar{x}_i + d_i|^{1/2} + |\bar{x}_i|^{1/2})^2} \right).
\end{aligned} \tag{3.15}$$

Combining (3.13), (3.14) and (3.15), we get

$$\begin{aligned}
F(\bar{x} + d) - F(\bar{x}) &= \frac{1}{2} d_{\bar{I}_A}^T \nabla^2_{\bar{I}_A \bar{I}_A} f(\bar{x}) d_{\bar{I}_A} - \sum_{i \in \bar{I}_A} \frac{\rho d_i^2}{8|\bar{x}_i|^{3/2}} \\
&\quad + \rho \sum_{i \in \bar{I}_A} \left(\frac{|\bar{x}_i + d_i| + 2|\bar{x}_i + d_i|^{1/2}|\bar{x}_i|^{1/2} - 3|\bar{x}_i|}{8|\bar{x}_i|^{3/2}(|\bar{x}_i + d_i|^{1/2} + |\bar{x}_i|^{1/2})^2} \right) d_i^2 \\
&\quad + \sum_{i \in \bar{I}_C} \left(\frac{\partial f(\bar{x})}{\partial x_i} d_i + \rho |d_i|^{1/2} \right) + g^T d_{\bar{I}_C} + o(\|d\|_2^2) \\
&= \frac{1}{2} d_{\bar{I}_A}^T \nabla^2_{\bar{I}_A \bar{I}_A} f(\bar{x}) d_{\bar{I}_A} \\
&\quad - \sum_{i \in \bar{I}_A} \frac{\rho d_i^2}{8|\bar{x}_i|^{3/2}} + (\nabla_{\bar{I}_C} f(\bar{x}) + g)^T d_{\bar{I}_C} + \rho \|d_{\bar{I}_C}\|_{1/2}^{1/2} + o(\|d\|_2^2),
\end{aligned} \tag{3.16}$$

where the second equality follows because

$$\lim_{d \to 0} |\bar{x}_i + d_i| + 2|\bar{x}_i + d_i|^{1/2}|\bar{x}_i|^{1/2} - 3|\bar{x}_i| = 0.$$

From (3.11) and (3.16) we deduce that for any d sufficiently small, it holds

$$F(\bar{x} + d) > F(\bar{x}).$$

Therefore, $\bar{x}$ is a local minimizer of the $L_{1/2}$ regularized problem (1.1). The proof is complete.

Similar to Theorem 2.2 of [5], the following theorem derives a bound on the number of nonzero entries in any accumulation point $\bar{x}$ of $\{x^k\}$.

Theorem 3.3 *Suppose that the sequence $\{x^k\}$ is generated by Algorithm 2.1 and that x^0 is the initial iterate. Let $\bar{x}$ be an accumulation point of $\{x^k\}$. Then the number of nonzero entries in $\bar{x}$ is bounded by*

$$\|\bar{x}\|_0 \le \frac{\sqrt{L_f [2(F(x^0) - L_{\text{low}})]^3}}{\rho^2}.$$

Proof Since $\bar{x}$ is an accumulation point of $\{x^k\}$ and x^0 is the initial iterate, from Algorithm 2.1 it follows that $F(\bar{x}) \le F(x^0)$. Moreover, by Theorem 3.1 we know that $\bar{x}$ is a first-order stationary point. Let $\bar{x}_i \neq 0$. By Theorem 2.2 of [18], we get

$$|\bar{x}_i| \geq \left(\frac{\rho}{2\sqrt{2}\sqrt{L_f(F(x^0) - L_{\text{low}})}} \right)^2,$$

where L_f and L_{low} is defined by (1.2). Note that

$$\|\bar{x}\|_{1/2}^{1/2} = \frac{F(\bar{x}) - f(\bar{x})}{\rho} \leq \frac{F(x^0) - L_{\text{low}}}{\rho}.$$

Therefore, simple calculation shows that

$$\|\bar{x}\|_0 \leq \frac{\sqrt{L_f[2(F(x^0) - L_{\text{low}})]^3}}{\rho^2}.$$

The proof is complete. □

Theorem 3.3 shows that for any given $x^0 \in R^n$, a smaller parameter ρ may yield a denser approximation. Hence, both x^0 and ρ should be chosen appropriately.

4 Complexity of the BBIHTA Algorithm

In this section, we shall study the complexity of the BBIHTA algorithm. We verify that the worst-case complexity of the algorithm for generating an ε global minimizer or an ε scaled first-order stationary point of (1.1) is $O(\varepsilon^{-2})$. To this end, we introduce some notions and an important lemma.

Definition 4.1 (i) For $\varepsilon > 0$, x is called an ε global minimizer of (1.1) if

$$F(x) - F(x^*) \leq \varepsilon,$$

where x^* is a global minimizer of (1.1);

(ii) x is called an ε scaled first-order stationary point of (1.1) if it satisfies

$$\|\Psi(x)\|_2 \leq \varepsilon.$$

Lemma 4.1 *Let $\{x^k\}$ be a sequence of iterates generated by the BBIHTA algorithm. Then it holds*

$$\|x^{k+1} - x^k\|_2 \geq \frac{\|\Psi\left(x^{k+1}\right)\|_2}{2\beta^{1/2}(\lambda + L_f)}, \qquad k = 0, 1, \ldots,$$

where $\lambda = \max\{\lambda_{\max}, \sigma_2(L_f + 2\gamma)\}$, β and L_f are given by (iii) of Remark 2.1 and (1.2), respectively.

Proof Since the sequence $\{x^k\}$ is generated by the BBIHTA algorithm, from the discussion of Sect. 2 we derive that for any k and i, x_i^{k+1} is an optimal solution of problem (2.3). Assume that $x_i^{k+1} \neq 0$. Then by the first-order necessary condition for optimality, we get

$$2\left(x_i^{k+1} - y_i^k\right) + \frac{\beta_k \operatorname{sgn}\left(x_i^{k+1}\right)}{2|x_i^{k+1}|^{1/2}} = 0.$$

This together with (2.2) implies that

$$2\left(x_i^{k+1} - x_i^k + \frac{\nabla_i f(x^k)}{\lambda_k}\right) + \frac{\rho \operatorname{sgn}\left(x_i^{k+1}\right)}{\lambda_k |x_i^{k+1}|^{1/2}} = 0.$$

Simple calculation shows that

$$\begin{aligned} &2|x_i^{k+1}|^{1/2}\nabla_i f\left(x^{k+1}\right) + \rho \operatorname{sgn}\left(x_i^{k+1}\right) \\ &= -2|x_i^{k+1}|^{1/2}\left[\lambda_k\left(x_i^{k+1} - x_i^k\right) + \nabla_i f\left(x^k\right) - \nabla_i f\left(x^{k+1}\right)\right] \end{aligned}$$

and so by combining with (2.12), we obtain

$$\begin{aligned} |\Psi_i\left(x^{k+1}\right)| &\le |2|x_i^{k+1}|^{1/2}\nabla_i f\left(x^{k+1}\right) + \rho \operatorname{sgn}\left(x_i^{k+1}\right)| \\ &= 2|x_i^{k+1}|^{1/2}\left|\lambda_k\left(x_i^{k+1} - x_i^k\right) + \nabla_i f\left(x^k\right) - \nabla_i f\left(x^{k+1}\right)\right|. \end{aligned} \tag{4.1}$$

It is easy to see that (4.1) also holds when $x_i^{k+1} = 0$. Combining (4.1), (2.12), (1.2), (iii) of Remark 2.1 and Proposition 2.3 gives

$$\begin{aligned} \|\Psi(x^{k+1})\|_2 &\le \left[\sum_{i=1}^n \left(2|x_i^{k+1}|^{1/2}\left|\lambda_k\left(x_i^{k+1} - x_i^k\right) + \nabla_i f(x^k) - \nabla_i f(x^{k+1})\right|\right)^2\right]^{1/2} \\ &= 2\left[\sum_{i=1}^n |x_i^{k+1}|\left(\lambda_k\left(x_i^{k+1} - x_i^k\right) + \nabla_i f(x^k) - \nabla_i f(x^{k+1})\right)^2\right]^{1/2} \\ &\le 2\left[\sum_{i=1}^n \|x^{k+1}\|_2\left(\lambda_k\left(x_i^{k+1} - x_i^k\right) + \nabla_i f(x^k) - \nabla_i f(x^{k+1})\right)^2\right]^{1/2} \\ &= 2\left(\|x^{k+1}\|_2\right)^{1/2}\left\|\lambda_k(x^{k+1} - x^k) + \nabla f(x^k) - \nabla f\left(x^{k+1}\right)\right\|_2 \\ &\le 2\left(\|x^{k+1}\|_2\right)^{1/2}\left(\lambda\left\|x^{k+1} - x^k\right\|_2 + \left\|\nabla f(x^k) - \nabla f(x^{k+1})\right\|_2\right) \\ &\le 2\beta^{1/2}(\lambda + L_f)\|x^{k+1} - x^k\|_2, \end{aligned}$$

where $\lambda = \max\{\lambda_{\max}, \sigma_2(L_f + 2\gamma)\}$. Therefore,

$$\|x^{k+1} - x^k\|_2 \ge \frac{\|\Psi\left(x^{k+1}\right)\|_2}{2\beta^{1/2}(\lambda + L_f)}.$$

The proof is complete. □

By Lemma 4.1, we can verify the following theorem.

Theorem 4.1 *For any $\varepsilon \in (0, 1)$, the BBIHTA algorithm obtains an ε scaled first-order stationary point or an ε global minimizer of the $L_{1/2}$ regularized problem (1.1) in no more than $O(\varepsilon^{-2})$ iterations.*

Proof Let $\{x^k\}$ be a sequence of iterates generated by the BBIHTA algorithm. From (2.13) and Lemma 4.1, we obtain

$$\begin{aligned} F(x^{k+1}) &\le \max_{0\le j\le \min(k,M)} F\left(x^{k-j}\right) - \gamma\|x^{k+1} - x^k\|_2^2 \\ &\le \max_{0\le j\le \min(k,M)} F\left(x^{k-j}\right) - \gamma\left(\frac{\|\Psi(x^{k+1})\|_2}{2\beta^{1/2}(\lambda + L_f)}\right)^2, \qquad \forall k = 0, 1, 2, \cdots. \end{aligned}$$

Let $l(k)$ be an integer satisfying (3.1). Then the following inequalities hold

$$1 \leq k+1-l(k) \leq M+1 \tag{4.2}$$

and

$$F\left(x^{k+1}\right) - F\left(x^{l(k)}\right) \leq -\gamma \left(\frac{\|\Psi(x^{k+1})\|_2}{2\beta^{1/2}(\lambda + L_f)}\right)^2. \tag{4.3}$$

Assume that $\|\Psi(x^{k+1})\|_2 > \varepsilon$ for all k. From (4.3) it follows that

$$F\left(x^{k+1}\right) - F\left(x^{l(k)}\right) < -\gamma \left(\frac{\varepsilon}{2\beta^{1/2}(\lambda + L_f)}\right)^2, \qquad \forall k = 0, 1, 2, \ldots. \tag{4.4}$$

By combining (4.2) and (4.4), we deduce that for any k, there exist at least $J \triangleq \lceil k/(M+1) \rceil$ nonnegative integers (denoted by $k_1, k_2, \ldots, k_J$) such that

$$0 = k_1 < k_2 < \ldots < k_J = k$$

and

$$F\left(x^{k_{j+1}}\right) - F\left(x^{k_j}\right) < -\gamma \left(\frac{\varepsilon}{2\beta^{1/2}(\lambda + L_f)}\right)^2, \qquad \forall j = 1, 2, \ldots, J-1.$$

This yields

$$\begin{aligned} F\left(x^k\right) - F\left(x^*\right) &= F\left(x^{k_J}\right) - F\left(x^*\right) \\ &= \sum_{j=J-1}^{1} \left(F\left(x^{k_{j+1}}\right) - F\left(x^{k_j}\right)\right) + F\left(x^{k_1}\right) - F(x^*) \\ &< -(J-1)\gamma \left(\frac{\varepsilon}{2\beta^{1/2}(\lambda + L_f)}\right)^2 + F\left(x^0\right) - F\left(x^*\right), \end{aligned}$$

where x^* is a global minimizer of the $L_{1/2}$ regularized problem (1.1). Choose

$$k = (M+1) \left\lceil \frac{(2\beta^{1/2}(\lambda + f_L))^2 (F(x^0) - F(x^*))}{\gamma} \varepsilon^{-2} \right\rceil + 1.$$

Then by a simple calculation, we get

$$F\left(x^k\right) - F\left(x^*\right) < 0 < \varepsilon.$$

Therefore, the conclusion holds. The proof is complete. □

5 Numerical Experiments

In this section, we report some preliminary numerical results to demonstrate the performance of the BBIHTA algorithm. We shall compare the performance of the BBIHTA algorithm with the IHTA algorithm proposed by Xu et al. [22] and the AIRL$_1$ (i.e., $\alpha = 1$) algorithm proposed by Lu in Section 3.1 of [18]. All codes are written in MATLAB and all computations are performed on a Lenovo PC (2.53 GHz, 2.00 GB of RAM).

We test these algorithms on the sparse signal recovery problem, where the goal is to reconstruct a length-n sparse signal x_s from m observations b with $m < n$ via solving the

L_2–$L_{1/2}$ minimization problem (1.3). The signal x_s, the sensing matrix A and the observation b in the problem are generated by the following Matlab code.

$$\begin{aligned} \mathrm{x_s} = \mathrm{zeros(n,1)};\quad \mathrm{p} = \mathrm{randperm(n)};\quad \mathrm{x_s(p(1:T))} = \mathrm{sign(randn(T,1))};\\ \mathrm{A} = \mathrm{randn(m,n)};\quad \mathrm{A} = \mathrm{orth(A')'};\quad \mathrm{b} = \mathrm{A} * \mathrm{x_s}. \end{aligned} \tag{5.1}$$

Obviously, $\|x_s\|_0 = T$. Moreover, the vectors x_s and b satisfy: $b = Ax_s$. In order to enforce that $Ax \approx b$ in (1.3), we must choose ρ to be extremely small. Unfortunately, choosing a small value for ρ makes (1.3) extremely difficult to solve numerically. Xu et al. [22] present three schemes to adjust the parameter ρ. In these schemes, the parameter ρ is updated at each iteration. Following [10] and [13], we adopt a continuation strategy to decrease ρ for three algorithms. In particular, the ρ in (1.3) is set according to the following procedure:

(P1): set $\rho := 3 \times 10^{-2}$ and TOLA $= 10^{-8}$;
(P2): compute $\bar{x}$ via approximately solving (1.3) such that

$$\|\Psi(\bar{x})\|_\infty < \mathrm{TOLA},$$

where Ψ is defined by (2.12);
(P3): if $\|A\bar{x} - b\|_\infty < 10^{-3}$, stop; otherwise, let $\rho := 0.1 \times \rho$, and go to step (P2).

By (5.1) it is not difficult to verify that the eigenvalues of $A^T A$ are 0 and 1. This means that $\|A^T A\|_2 = 1$. Hence by [24], we set $\mu_k = 0.99$ in the IHTA iteration (1.8). The parameters in the BBIHTA algorithm are set as follows:

$$\lambda_{\min} = 10^{-8}, \quad \lambda_{\max} = 10^8, \quad \bar{\lambda} = 1, \quad M = 10, \quad \gamma = 10^{-4}, \quad \sigma_1 = 1.1, \quad \sigma_2 = 1.9.$$

The parameters in the AIRL$_1$ algorithm are chosen according to [18], that is,

$$L_{\min} = 10^{-8}, \quad L_{\max} = 10^8, \quad c = 10^{-4}, \quad \tau = 1.1, \quad L_0^0 = 1,$$

and

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

with $\Delta x = x^k - x^{k-1}$ and $\Delta g = \nabla f(x^k) - \nabla f(x^{k-1})$. We use the nonmonotone line search to improve the convergence of the AIRL$_1$ algorithm. Additionally, to ensure that the sequence of iterates generated by the AIRL$_1$ algorithm converges to a stationary point of (1.3), we define the sequences of $\{\varepsilon^r\}$ and $\{\delta^r\}$ by

$$\varepsilon^{r+1} = 0.1 \times \varepsilon^r \quad \text{and} \quad \delta^{r+1} = 0.1 \times \delta^r, \quad r = 0, 1, \ldots,$$

where ε^0 and δ^0 are given. For each ε^r, the AIRL$_1$ algorithm is applied to (1.5) for finding x^r such that

$$\left\| X^r \nabla f(x^r) + \frac{\rho}{2} |X^r|^{1/2} \left(|x^r|^{1/2} + \varepsilon^r\right)^{-1/2} \right\| < \delta^r,$$

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

The initial guesses for three algorithms are generated by the following three schemes: (1) x^0 being a solution from the L_2–L_1 minimization problem

$$\min_{x \in R^n} \frac{1}{2}\|Ax - b\|_2^2 + \rho\|x\|_1,$$

where $\rho = 3 \times 10^{-2}$; (2) $x^0 = (0, 0, \ldots, 0)^T$; and (3) $x^0 = A^T b$. The L_2–L_1 minimization problem is solved by the SpaRSA algorithm [20].

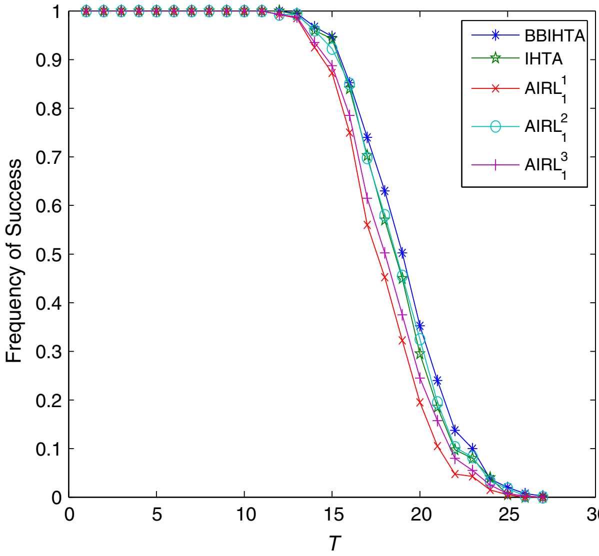

Fig. 1 Successful frequency for signal reconstruction

We first test the ability of three algorithms in recovering the sparse solutions. In this experiment, we fix $m = 2^6$ and $n = 2^8$ and vary the value of T from 1 to 27. For each T, we construct 400 random pairs (A, x_s), and run these algorithms to obtain an approximation $\bar{x}$ for each pair (A, x_s). For the AIRL$_1$ algorithm, we choose three parameter pairs $(\varepsilon^0, \delta^0)$. That is,

$$\varepsilon^0 = 10^{-1}, \quad \delta^0 = 10^{-1}; \qquad \varepsilon^0 = 10^{-1}, \quad \delta^0 = 10^{-4}; \qquad \varepsilon^0 = 10^{-2}, \quad \delta^0 = 10^{-1}.$$

For convenience of presentation, we name the corresponding algorithm as AIRL$_1^1$, AIRL$_1^2$ and AIRL$_1^3$. We consider the recovery a success if $\|x_s - \bar{x}\|_\infty \leq 10^{-2}$. For each T, Fig. 1 presents the percentages of successfully finding the sparse solutions for these algorithms.

From Fig. 1, we can see that the BBIHTA algorithm gives slightly better success rates than other algorithms. Figure 1 also shows that the success rate of the AIRL$_1$ algorithm is dependent on the parameters ε^0 and δ^0. When $\varepsilon^0 = 10^{-1}$ and $\delta^0 = 10^{-4}$, it gives better success rates.

We then compare the computational efficiency of the BBIHTA algorithm against the IHTA algorithm and the AIRL$_1^2$ algorithm, i.e., $\varepsilon^0 = 10^{-1}$ and $\delta^0 = 10^{-4}$. In this experiment, we choose different T, m and n. For each (T, m, n), we construct 400 random pairs (A, x_s), and we run these algorithms to generate an approximation. Table 1 lists average number of iterations and average CPU time. The results in this table show that, the BBIHTA algorithm takes fewer iterations and requires less CPU time than the IHTA algorithm and the AIRL$_1^2$ algorithm.

In previous experiments, BBIHTA, IHTA and AIRL$_1$ algorithms are used to recover the sparse signal by solving a sequence of L_2–$L_{1/2}$ minimization problems and the penalty parameter ρ is adjusted by a continuous strategy. In the rest of this section, we compare three algorithms by solving the L_2–$L_{1/2}$ minimization problem with fixed ρ, i.e., $\rho = 3 \times 10^{-2}$. The stopping condition for all the methods is $\|\Psi(x^k)\|_\infty < 10^{-8}$ and the initial iterate x^0 is a solution of the corresponding L_2–L_1 minimization problem. In the experiment, we fix

Table 1 The number of iterations and CPU time (seconds) for three algorithms

Prolem (T, m, n)	BBIHTA		IHTA		$AIRL_1^2$	
	ITER	Time	ITER	Time	ITER	Time
$(10, 2^6, 2^8)$	44.5125	0.0218	323.6050	0.1447	77.2125	0.0245
$(20, 2^7, 2^9)$	47.8700	0.0300	342.3900	0.2560	83.8850	0.0349
$(40, 2^8, 2^{10})$	49.8950	0.1434	350.1025	0.9180	87.7850	0.2050
$(80, 2^9, 2^{11})$	50.8025	0.4947	354.3900	3.1093	89.8775	0.7850
$(160, 2^{10}, 2^{12})$	51.4900	1.5010	360.0575	10.5531	91.4675	2.5641

Table 2 The number of iterations and CPU time (seconds) for three algorithms

Prolem (T, m, n)	BBIHTA		IHTA		$AIRL_1^2$	
	ITER	Time	ITER	Time	ITER	Time
$(10, 2^8, 2^{10})$	11.3275	0.0257	74.6200	0.1772	28.0950	0.0461
$(20, 2^8, 2^{10})$	13.8683	0.0266	91.3015	0.1972	37.2052	0.0500
$(30, 2^8, 2^{10})$	16.1297	0.0283	110.4008	0.2227	44.3230	0.0560
$(40, 2^8, 2^{10})$	18.0480	0.0297	129.9923	0.2388	51.7914	0.0639

$m = 2^8$ and $n = 2^{10}$, and choose $T = 10, 20, 30, 40$. For each (T, m, n), we construct 400 random pairs (A, x_s), and we run these algorithms to generate an approximation. Table 2 lists average number of iterations and average CPU time. The results show that the BBIHTA algorithm performs better than the other two algorithms.

6 Final Remarks

The $L_{1/2}$ regularized problem arises in many important applications and is a class of non-Lipschitz, nonconvex optimization problems. This paper proposes a Barzilai–Borwein-like iterative half thresholding algorithm for the $L_{1/2}$ regularized problem. We verify that any accumulation point of the sequence of iterates generated by the algorithm is a first-order stationary point of this problem under mild conditions and it is also a local minimizer when additional conditions are satisfied. Furthermore, we show that the worst-case iteration complexity for finding an ε scaled first-order stationary point is $O(\varepsilon^{-2})$. Preliminary numerical results show that the proposed algorithm is practically effective.

Acknowledgments The authors would like to thank two anonymous referees for their valuable suggestions and comments.

References

1. Barzilai, J., Borwein, J.: Two point step size gradient methods. IMA J. Numer. Anal. **8**, 141–148 (1988)
2. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**, 183–202 (2009)

3. Bian, W., Chen, X., Ye, Y.: Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization. Math. Program. **149**, 301–327 (2015)
4. Candès, E., Wakin, M., Boyd, S.: Enhancing sparsity by reweighted L_1 minimization. J. Fourier Anal. Appl. **14**, 877–905 (2008)
5. Chen, X., Xu, F., Ye, Y.: Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. SIAM J. Sci. Comput. **32**, 2832–2852 (2010)
6. Chen, X., Zhou, W.: Convergence of the reweighted L_1 minimization algorithm for L_2-L_p minimization. Comput. Optim. Appl. **59**, 47–61 (2014)
7. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward–backward splitting. SIAM J. Multiscale Model. Simul. **4**, 1168–1200 (2005)
8. Dai, Y.H.: A new analysis on the Barzilai–Borwein gradient method. J. Oper. Res. Soc. China **1**, 187–198 (2013)
9. Daubechies, I., De Friese, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. **57**, 1413–1457 (2004)
10. Figueiredo, M., Nowak, R., Wright, S.: Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. IEEE J. Sel. Top. Signal. **1**, 586–598 (2007)
11. Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for Newton's method. SIAM J. Numer. Anal. **23**, 707–716 (1986)
12. Hager, W.W., Phan, D.T., Zhang, H.: Gradient-based methods for sparse recovery. SIAM J. Imaging Sci. **4**, 146–165 (2011)
13. Hale, E., Yin, W., Zhang, Y.: Fixed-point continuation for ℓ_1-minimization: methodology and convergence. SIAM J. Optim. **19**, 1107–1130 (2008)
14. Hale, E., Yin, W., Zhang, Y.: Fixed-point continuation applied to compressed sensing: implementation and numerical experiments. J. Comput. Math. **28**, 170–194 (2010)
15. Lai, M., Wang, J.: An unconstrained L_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems. SIAM J. Optim. **21**, 82–101 (2011)
16. Lai, M., Xu, Y., Yin, W.: Improved iteratively reweighted least squares for unconstrained smoothed L_q minimization. SIAM J. Numer. Anal. **51**, 927–957 (2013)
17. Li, D.H., Wu, L., Sun, Z., Zhang, X.J.: A constrained optimization reformulation and a feasible descent direction method for $L_{1/2}$ regularization. Comput. Optim. Appl. **59**, 263–284 (2014)
18. Lu, Z.: Iterative reweighted minimization methods for L_p regularized unconstrained nonlinear programming. Math. Program. **147**, 277–307 (2014)
19. Raydan, M.: The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM J. Optim. **7**, 26–33 (1997)
20. Wright, S.J., Nowak, R., Figueiredo, M.: Sparse reconstruction by separable approximation. IEEE Trans. Signal Process. **57**, 2479–2493 (2009)
21. Wu, L., Sun, Z., Li, D.H.: Gradient based method for the L_2-$L_{1/2}$ minimization and application to compressive sensing. Pac. J. Optim. **10**, 401–414 (2014)
22. Xu, Z., Chang, X., Xu, F., Zhang, H.: $L_{1/2}$ regularization: a thresholding representation theory and a fast solver. IEEE Trans. Neural Netw. Learn. Syst. **23**, 1013–1027 (2012)
23. Xu, Z., Zhang, H., Wang, Y., Chang, X.: $L_{1/2}$ regularizer. Sci. China Ser. F **52**, 1–9 (2009)
24. Zeng, J., Lin, S., Wang, Y., Xu, Z.: $L_{1/2}$ regularization: convergence of iterative half thresholding algorithm. IEEE Trans. Signal Process. **62**, 2317–2329 (2014)