NOMONOTONE SPECTRAL GRADIENT METHOD FOR SPARSE RECOVERY

WANYOU CHENG AND ZIXIN CHEN
College of Computer
Dongguan University of Technology
Dongguan, 523000, China

DONGHUI LI
School of Mathematical Sciences
South China Normal University
Guangzhou, 510631, China

(Communicated by Xuecheng Tai)

Abstract. In the paper, we present an algorithm framework for the more general problem of minimizing the sum \( f(x) + \psi(x) \), where \( f \) is smooth and \( \psi \) is convex, but possible nonsmooth. At each step, the search direction of the algorithm is obtained by solving an optimization problem involving a quadratic term with diagonal Hessian and Barzilai-Borwein steplength plus \( \psi(x) \). The method with the nomonotone line search techniques is showed to be globally convergent. In particular, if \( f \) is convex, we show that the method shares a sublinear global rate of convergence. Moreover, if \( f \) is strongly convex, we prove that the method converges R-linearly. Numerical experiments with compressive sense problems show that our approach is competitive with several known methods for some standard \( \ell_2 - \ell_1 \) problems.

1. Introduction. Consider the solution of the optimization problem

\[
\min \phi(x) := f(x) + \psi(x),
\]

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is a smooth function and \( \psi : \mathbb{R}^n - \mathbb{R} \) is convex. The function \( \psi \), usually called the regularizer or regularization function, is finite for all \( x \in \mathbb{R}^n \), but possibly nonsmooth. A special case of (1.1) that has attracted much interest in signal/image denoising and data mining/classification is the well-known \( \ell_2 - \ell_1 \) problem

\[
\min_{x \in \mathbb{R}^n} \frac{1}{2} \| Ax - b \|_2^2 + \mu \| x \|_1,
\]

where \( A \in \mathbb{R}^{m \times n} \) is dense (usually \( m \leq n \)), \( b \in \mathbb{R}^m \), \( \mu > 0 \) and \( n \) is large.

There have been many approaches [20, 23, 27, 29] for the solution of (1.1). The earlier work is due to Fukushima and Mine [20], who proposed a proximal gradient descent method with stepsize chosen by an Armijo-type rule. Further extension of this method [27] and trust-region versions of the method were studied in [23]. The

2010 Mathematics Subject Classification. 90C06, 90C25, 65Y20, 94A08.

Key words and phrases. Iterative shrinkage thresholding algorithm, nonsmooth optimization, \( \ell_1 \) − regularization, Barzilai-Borwein method.

Supported by the NSF of China via grant 11371154, 11101081, by the Guangdong Natural Science Funds S201310011809 and the Guangdong Province outstanding young teacher training program 3XZ150603.
816 Wanyou Cheng, Zixin Chen and Donghui Li

\[ \epsilon- \text{subgradient method} \ [32] \text{is developed to solve} \ (1.1) \text{if} \ f \text{is convex. Recently,} \]

Tseng and Yun [34] proposed a block coordinate gradient descent (CGD) method for minimizing (1.1) where \( \psi(x) \) is a separable convex function. Wright, Nowak and Figueiredo [38] introduced the sparse reconstruction by separable approximation (SPaRSA) algorithm for solving (1.1). Hager, Phan and Zhang [26] analyzed the convergence rate of the SPaRSA and proposed an improved version of SpaRSA based on a cyclic version of the BB iteration and an adaptive line search.

About the special case of (1.2), one of the most popular methods for solving problem (1.2) is the class of iterative shrinkage-thresholding algorithms (ISTA) [12, 21], or improved versions of the ISTA [4, 6, 19, 25, 26, 36, 37, 38], where each iteration involves a matrix-vector multiplication involving \( A \) and \( A^T \), followed by a shrinkage/soft-threshold step. Another way to compute (1.2) is that one can transform (1.2) to a quadratic problem with linear inequality constraint and solve it by interior point method. Some high-quality implementations of interior point methods such as \( \ell_1 \)-magic [10], primal-dual interior point method for convex objectives [33] and \( \ell_1 J_\gamma \) [28] have been developed. Figueiredo, Nowak and Wright [22] reformulated (1.1) as a box-constrained quadratic program, to which they applied the gradient projection method with Barzilai-Borwein steps. Other algorithms for the \( \ell_1 \) minimization include alternating direction method of multipliers SALSA [1, 8]; matching pursuit (MP) method and orthogonal MP method [15, 17, 35]; Bergman iterative regularization based methods [39]; gradient methods [30] for minimizing the more general function \( J(x) + H(x) \), where \( J \) is nonsmooth, \( H \) is smooth, and both are convex; a smoothed penalty algorithm (SPA) [2] that solves problem (1.2) with the quadratic penalty replaced by an “exact” \( \ell_2 - \ell_1 \) penalty. We refer to papers [16, 5, 9] for a review on recent advances in this area.

Recently, Tseng and Yun [34] considered the problem of minimizing the sum of a smooth of function and a separable convex function and developed a block coordinate gradient descent (CGD) method for minimizing (1.1) where \( \psi(x) \) is a separable convex function. The basic idea of the CGD is that, at each iteration they approximate \( f \) by a strictly convex quadratic approximation involving the Hessian of \( f \) and apply block coordinate descent to generate a feasible descent direction. Then, they perform an inexact line search along this direction and re-iterate. However, numerical results in [34] show the practical efficiency of the CGD method only for small and medium scale problem (1.2). In the paper, following the idea of the CGD method, we present an algorithm framework for large-scale problem (1.1). Numerical experiments show that it is more effective for large-scale problem (1.2). More precisely, at each step, the search direction of the algorithm is obtained by solving an optimization subproblem involving a quadratic term with diagonal Hessian and Barzilai-Borwein steplength plus \( \psi(x) \). Then, we perform a nonmonotone line search along this direction and re-iterate. We prove that the method with the nonmonotone line search techniques is globally convergent. In particular, if \( f \) is convex, we show that the method shares a sublinear global rate of convergence. Moreover, if \( f \) is strongly convex, we prove that the method converges R-linearly. Numerical experiments with compressive sense problems show that our approach is competitive with several known methods for some standard \( \ell_2 - \ell_1 \) problems.

The rest of the the paper is organized as follows. We propose the algorithm in Section 2. In Section 3, we establish the global convergence of the algorithm.
Convergence rate estimate for convex functions and strongly convex functions are given in Section 4. Some numerical results are reported in Section 5.

Throughout the paper, \( \| \cdot \| \) denotes the Euclidean norm of vectors. \( \partial \psi(y) \) is the subdifferential at \( y \), a set of column vectors.

2. Motivation and properties. In this section, we give some useful properties of the search direction and propose the algorithm. The search direction of the method is mostly closely related to that of the coordinate gradient descent method for separable minimization [34].

Before stating the method, let us simply recall the CGD method. Let \( x_k \) be \( k \)-th iteration. In the CGD method, they choose a nonempty index subset \( J \subseteq N \) and a symmetric and positive definite matrix \( H_k \) approximating the Hessian \( \nabla^2 f(x_k) \), and move \( x_k \) along the direction \( d_{H_k}(x_k, J) \), where

\[
(2.1) \quad d_{H_k}(x_k, J) := \arg \min_{v \in R^n} \{ \nabla f(x_k)^T v + \frac{1}{2} v^T H_k v + \psi(x_k + v) \mid v_j = 0 \ \forall j \notin J \},
\]

where \( v_j \) denotes the \( j \)-th component of vector \( v \). In the paper, we focus on the case that \( J = N \) and \( H_k = \lambda_k I \) where \( I \) denotes the identity matrix and \( \lambda_k \) is the Barzilai-Borwein (BB) step length [3, 11, 13, 14] with safeguard. That is, we use the diagonal matrix \( \lambda_k I \) to approximate the Hessian \( \nabla^2 f(x_k) \) of \( f \) at \( x_k \). The BB step length is defined by \( \lambda_k = \frac{\|s_k\|^2}{s_k^T y_k} \) where \( s_k = x_k - x_{k-1} \) and \( y_k = \nabla f(x_k) - \nabla f(x_{k-1}) \). A good property of \( \lambda_k I \) is that it satisfies the quasi-Newton condition

\[
\lambda_k := \arg \min_{\lambda \in R} \{ \|\lambda s_k - y_k\| \}.
\]

To avoid small or large values of \( \lambda_k \), we take

\[
(2.2) \quad \lambda_k := \min \{ \lambda_{\max}, \max \{ \lambda_{\min}, \frac{\|s_k\|^2}{s_k^T y_k} \} \},
\]

where \( 0 < \lambda_{\min} < \lambda_{\max} \) are fixed constants. In the rest of the paper, without other statement, we assume that \( J = N \) and \( H_k = \lambda_k I \), where \( \lambda_k \) is determined by (2.2). In the case, we abbreviate \( d_{H_k}(x_k, N) \) as \( d_k \).

Remark.

(1) Suppose that \( \bar{x}_k \) denotes a solution of the following problem. That is,

\[
(2.3) \quad \bar{x}_k \in \arg \min_{z \in R^n} \{ \nabla f(x_k)^T z + \frac{1}{2} \lambda_k \|z - x_k\|^2 + \psi(z) \},
\]

where \( \lambda_k \) appears in (2.2). An equivalent form of subproblem (2.3) is

\[
\bar{x}_k \in \arg \min_{z \in R^n} \left\{ \frac{1}{2} \|z - u_k\|^2 + \frac{1}{\lambda_k} \psi(z) \right\},
\]

where \( u_k = x_k - \frac{1}{\lambda_k} \nabla f(x_k) \). This form is considered frequently in the literature often under the name of ISTA algorithms [12, 21]. Letting \( x_k + v = z \), we have

\[
\nabla f(x_k)^T v + \frac{\lambda_k}{2} \|v\|^2 + \psi(x_k + v) = \nabla f(x_k)^T (z - x_k) + \frac{\lambda_k}{2} \|z - x_k\|^2 + \psi(z).
\]

Hence, we get \( d_k = \bar{x}_k - x_k \). Hence, our method is related to ISTA algorithms [12, 21] or variants of ISTA algorithms [4, 6, 19, 25, 26, 36, 37, 38].
(2) If \( \psi(x) = 0 \), then the solution of (2.1) is
\[
d_k = -\frac{1}{\lambda_k} \nabla f(x_k).
\]
In this case, the method reduces to steepest descent method with adjustment of the BB step length.

(3) Notice that if \( \psi \) is given by
\[
\psi(x) = \begin{cases} 
0 & \text{if } x \in \Omega \\
\infty & \text{else},
\end{cases}
\]
where \( \Omega = \{x : l \leq x \leq u\} \) and \( l \leq u \) (possibly with \( -\infty \) or \( \infty \) components) and \( g_k \) denotes the gradient of \( f \) at \( x_k \). By directly calculating, we get that
\[
d_k = P_{\Omega}(x_k - \frac{1}{\lambda_k} \nabla f(x_k)) - x_k,
\]
where \( P_{\Omega} \) is the projection operator on the set \( \Omega \). That is, \( d_k \) is exactly the spectral gradient-projection direction for bound-constrained minimization [7].

(4) If \( \psi(x) = \tau \|x\|_1 \), by directly calculating, we get that
\[
d_k = S(x_k - \frac{1}{\lambda_k} \nabla f(x_k), \frac{\tau}{\lambda_k}) - x_k,
\]
where \( S \) is the shrinkage operator and defined by
\[
S(y, \nu) := \text{sgn}(y) \odot \max\{\|y\| - \nu, 0\}.
\]
where \( y \in \mathbb{R}^n \), \( \nu > 0 \) and \( \text{sgn} \) is the sign function. In the case, the search direction \( d_k \) is identical to the search direction of the nonmonotone line search algorithm [36, 37].

The following Lemmas 2.2-2.4 are from Lemma 1, Lemma 2 and Lemma 5 in [34] respectively. They give some useful properties of \( d_k \). For completeness, we give the proof. The following Lemma 2.1 plays an important role in establishing Lemmas 2.2-2.4 and some convergence properties of our method.

**Lemma 2.1.** (i) Suppose that \( \psi \) is convex. Then, for any \( t \geq 0 \), we have
\[
\psi(x + td) - \psi(x) \leq t(\psi(x + d) - \psi(x)), \ \forall x, d \in \mathbb{R}^n.
\]

(ii) Suppose that \( f \) is Lipschitz continuously with Lipschitz constant \( L \) and \( \psi \) is convex. Then we have
\[
f(x) \leq f(y) + \nabla f(y)^T(x - y) + \frac{L}{2}\|x - y\|^2, \ \forall x, y \in \mathbb{R}^n.
\]

**Proof.** By the use of the convexity of \( \psi \), we get
\[
\psi(x + td) - \psi(x) = \psi(t(x + d) + (1 - t)x) - \phi(x) \leq t(\psi(x + d) - \psi(x)).
\]
By the tailor expand of $f(x)$ at $y$, there exists a constant $\theta \in (0,1)$ such that
\[
 f(x) - f(y) = \int_0^1 \nabla f(y + \theta(x - y))^T(x - y) d\theta \\
 = \int_0^1 (\nabla f(y + \theta(x - y)) - \nabla f(y))^T(x - y) d\theta + \nabla f(y)^T(x - y) \\
\leq \nabla f(y)^T(x - y) + \int_0^1 L\|x - y\|^2 d\theta \\
= \nabla f(y)^T(x - y) + \frac{L}{2}\|x - y\|^2.
\]
\[\Box\]

**Lemma 2.2.** Suppose that $\psi$ is convex. For any $x_k \in \mathbb{R}^n$ and $d_k$ determined by (2.1), we have
\[
 \phi'(x_k, d_k) \leq \nabla f(x_k)^T d_k + \psi(x_k + d_k) - \psi(x_k) \leq -\frac{\lambda_k}{2}\|d_k\|^2,
\]
where $\phi'(x, d)$ is the one-sided directional derivative of $\phi$ at $x$ in the direction $d$.

**Proof.** By the definition of $\phi'$ and (i) of Lemma 2.1, we get
\[
 \phi'(x_k, d_k) = \lim_{t \to 0} \frac{f(x_k + td_k) - f(x_k) + \phi(x_k + td_k) - \phi(x_k)}{t} \\
\leq \lim_{t \to 0} \frac{f(x_k + td_k) - f(x_k) + t(\phi(x_k + d_k) - \phi(x_k))}{t} \\
= \nabla f(x_k)^T d + \psi(x_k + d_k) - \psi(x_k).
\]
On the other hand, by the definition of $d_k$ in (2.1), we get
\[
 0 \in \{\nabla f(x_k) + \frac{\lambda_k}{2}d_k + \partial \psi(x_k + d_k)\}
\]
which implies that
\[
 d_k \in \arg \min_{v \in \mathbb{R}^n} \{\langle \nabla f(x_k) + \frac{\lambda_k}{2}d_k \rangle v + \psi(x_k + v)\}.
\]
Thus, we have
\[
 (\nabla f(x_k) + \frac{\lambda_k}{2}d_k)d_k + \psi(x_k + d_k) \leq \psi(x_k).
\]
That is,
\[
 \nabla f(x_k)^T d_k + \psi(x_k + d_k) - \psi(x_k) \leq -\frac{\lambda_k}{2}\|d_k\|^2.
\]
\[\Box\]

**Lemma 2.3.** Suppose that $\psi$ is convex. Then, for any $x_k \in \mathbb{R}^n$ is a stationary point of (1.1) if and only if $d_k = 0$.

**Proof.** Suppose that $x_k \in \mathbb{R}^n$ is a stationary point of (1.1). Then we have $\phi'(x_k, d_k) \geq 0$. Following from Lemma 2.2, we get $d_k = 0$. On the other hand, suppose that $d_k = 0$. By the definition of $d_k$ in (2.1), we get
\[
 0 \in \{\nabla f(x_k) + \frac{\lambda_k}{2}d_k + \partial \psi(x_k + d_k)\}.
\]
Namely,
\[
 0 \in \{\nabla f(x_k) + \partial \psi(x_k)\},
\]
which shows that $x_k$ is a stationary point of (1.1). \[\Box\]
Lemma 2.4. If $f$ is Lipschitz continuous with Lipschitz constant $L$ and $\psi$ is convex, then the descent condition

\begin{equation}
\phi(x_k + \alpha d_k) - \phi(x_k) \leq \delta \Delta_k
\end{equation}

is satisfied for any $\delta \in (0, 1)$ whenever $0 \leq \alpha \leq \min\{1, 2\lambda_{\min}(1-\delta)\}$ and $\Delta_k = \nabla f(x_k)^T d_k + \psi(x_k + d_k) - \psi(x_k)$.

Proof. First, by Lemma (2.2), we know $\Delta_k \leq 0$. By Lemma (2.1) and (2.2), we get that

\begin{align*}
\phi(x_k + \alpha d_k) - \phi(x_k) &= f(x_k + \alpha d_k) - f(x_k) + \psi(x_k + \alpha d_k) - \psi(x_k) \\
&\leq \alpha^2 \frac{L}{2} \|d_k\|^2 + \alpha \nabla f(x_k)^T d_k + \psi(x_k + \alpha d_k) - \psi(x_k) \\
&\leq \alpha^2 \frac{L}{2} \|d_k\|^2 + \alpha \nabla f(x_k)^T d_k + \alpha (\psi(x_k + d_k) - \psi(x_k)) \\
&= \alpha^2 \frac{L}{2} \|d_k\|^2 + \alpha (\nabla f(x_k)^T d_k + (\psi(x_k + d_k) - \psi(x_k))) \\
&= \alpha^2 \frac{L}{2} \|d_k\|^2 + \alpha \Delta_k \\
&\leq \alpha \Delta_k + \alpha^2 L \frac{\Delta_k}{-\lambda_k} \\
&\leq \alpha \Delta_k (1 - \frac{\alpha L}{\lambda_{\min}}) \\
&\leq \alpha \delta \Delta_k
\end{align*}

whenever $0 \leq \alpha \leq \min\{1, 2\lambda_{\min}(1-\delta)\}$.

To enlarge the possibly small stepsize generated by the line search in (2.5), the nonmonotone line search technique [24] is used in our method. We call our method as the nonmonotone BB method and abbreviate it as NBB. Specifically, the method for solving (1.1) is stated as follows.

Algorithm 2.1. (NBB)

Step 0. Given an initial point $x_0 \in \mathbb{R}$ and positive constants $\rho$, $M$, $\lambda_{\min}$, $\lambda_{\max}$ and $\delta \in (0, 1)$. Set $k := 0$.

Step 1. Perform convergence test and terminate with an approximate solution $x_k$ if the stopping condition is satisfied; otherwise go to Step 2.

Step 2. Choose $\lambda_k \in [\lambda_{\min}, \lambda_{\max}]$, compute the solution of (2.1) and obtain $d_k$.

Step 3. Determine $\alpha_k := \max\{\rho^j, j = 0, 1, \cdots\}$ satisfying

\begin{equation}
\phi(x_k + \alpha_k d_k) \leq \phi_k^{\max} + \delta \alpha_k \Delta_k,
\end{equation}

where $\phi_k^{\max} = \max\{\phi(x_{k-j}) : 0 \leq j \leq \min(k, M-1)\}$.

Step 4. Let the next iterate be $x_{k+1} := x_k + \alpha_k d_k$.

Step 5. Set $k := k + 1$ and go to Step 1.

Since $\phi_k^{\max} \geq \phi(x_k)$, Lemma 2.4 implies that the condition (2.6) must hold after a finite number of reduction of $\alpha_k$. Consequently, Algorithm 2.1 is well-defined. In
addition, if \( \psi \) is defined by (2.4), Algorithm 2.1 is identical to the nonmonotone spectral projected gradient method on convex sets [7]. More precisely, please see the Algorithm 2.2 in [7].

3. Convergence analysis. In this section, we analyze the convergence of Algorithm 2.1. To this end, we make the following assumptions.

Assumption 3.1.
(i) The level set \( \Omega := \{ x \in \mathbb{R}^n : \phi(x) \leq \phi(x_0) \} \) is contained in the interior of a compact, convex set \( K \).
(ii) \( f \) is Lipschitz continuously differentiable on \( K \). That is, there exists a constant \( L > 0 \) such that

\[
\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \|, \quad \forall x, y \in K.
\]

The following Lemma 3.1 is from the Corollary 24.5.1 in [31] which will used in proof of Theorem 3.1.

Lemma 3.1. If \( \psi \) is a proper convex function on \( \mathbb{R}^n \), \( \psi'(x,y) \) is an upper semi-continuous function of

\[
(x,y) \in [\text{int}(\text{dom}\psi) \times \mathbb{R}^n).
\]

Moreover, given any \( x \in \text{int}(\text{dom}\psi) \) and any \( \epsilon > 0 \), there exists a \( \delta > 0 \) such that

\[
\partial \psi(z) \subset \partial \psi(x) + \epsilon B, \quad \forall z \in (x + \delta B),
\]

where \( \text{dom}\psi = \{ x | \psi(x) < \infty \} \), \( \text{int}(\text{dom}\psi) \) denotes the interior of the set \( \text{dom}\psi \) and \( B \) is the Euclidean unit ball of \( \mathbb{R}^n \).

The following theorem together with Lemma 2.4 shows that every accumulation point of \( \{ x^k \} \) is a stationary point of (1.1).

Theorem 3.1. Assume that Assumption 3.1 holds. Let \( \{ x_k \} \) be the sequence generated by Algorithm 2.1.

(1) Then

\[
\lim_{k \to \infty} \phi(x_k) = \bar{\phi},
\]

and

\[
\lim_{k \to \infty} \| d_k \| = 0
\]

where \( \bar{\phi} \) is a constant.

(2) Then every limit point of \( \{ x_k \} \) is a stationary point of \( \phi \).

Proof. By Lemma 2.4, we have

\[
\frac{\alpha_k}{\rho} \geq \min \{ 1, 2 \frac{\lambda_{\min}(1 - \delta)}{L} \}.
\]

That is

\[
\alpha_k \geq \rho \min \{ 1, 2 \frac{\lambda_{\min}(1 - \delta)}{L} \} := c.
\]

By Lemma 2.2 and (2.6), we have

\[
\phi(x_k) \leq \phi_{k-1}^{\max} - \delta c \frac{\lambda_{\min}}{2} \| d_{k-1} \|^2.
\]

Define

\[
\phi(x_{i(i)}) = \max \{ \phi(x_k); \ (i-1)M < k \leq iM \} = \phi_{iM}^{\max}.
\]
Since $\phi(x_{k+1}) \leq \phi^\text{max}_k$ in Step 3 of NBB, it follows that $\phi^\text{max}_k$ is a nonincreasing function sequence of $k$. Moreover, since $\phi$ is bounded below, there exists a constant $\tilde{\phi}$ such that

$$
\lim_{i \to \infty} \phi(x_{l(i)}) = \tilde{\phi}.
$$

By (3.9) with $k = l(i)$ and the monotonicity of $\phi^\text{max}_k$, we have

$$
\phi(x_{l(i)}) \leq \phi(x_{l(i-1)}) - \delta c \frac{\lambda_{\min}}{2} \|d_{l(i)-1}\|^2.
$$

By rearranging this expression and using (3.10), we obtain

$$
\lim_{i \to \infty} d_{l(i)-1} = 0.
$$

By the continuity of $\phi$ and (3.10), we have

$$
\lim_{i \to \infty} \phi(x_{l(i)-1}) = \lim_{i \to \infty} \phi(x_{l(i)} - \alpha_{l(i)-1}d_{l(i)-1}) = \tilde{\phi}.
$$

We will now prove, by induction that the following limits are satisfied for $j = 1, 2, \ldots, M$:

$$
\lim_{i \to \infty} d_{l(i)-j} = 0.
$$

By rearranging the expression and using (3.9), we have

$$
\lim_{i \to \infty} \phi(x_{l(i)-j}) = \tilde{\phi}.
$$

We have already shown that the above results hold for $j = 1$. Assume that the results hold for a given $1 < j < M$. Let $k = l(i) - j$ in (3.9). Noting $(i - 2)M \leq l(i) - j - 1 \leq iM$ and the monotonicity of $\phi^\text{max}_k$, we have

$$
\phi(x_{l(i)-j}) \leq \phi(x_{l(i-2)}) - \delta c \frac{\lambda_{\min}}{2} \|d_{l(i)-1}\|^2.
$$

By rearranging the expression and using (3.9), we have

$$
\lim_{i \to \infty} d_{l(i)-j-1} = 0.
$$

For $j + 1$, by (3.11), we have

$$
\lim_{i \to \infty} \phi(x_{l(i)-j-1}) = \lim_{i \to \infty} \phi(x_{l(i)-j} - \alpha_{l(i)-j-1}d_{l(i)-j-1}) = \tilde{\phi}.
$$

For any $k \in ((i - 1)M, l(i))$, we have

$$
1 \leq l(i) - k \leq M
$$

and

$$
x_{l(i)} = x_k + \sum_{j=1}^{l(i)-k+1} \alpha_{l(i)-j}d_{l(i)-j}.
$$

By the continuity of $\phi$ and (3.10), we have

$$
\lim_{k \to \infty} \phi(x_k) = \lim_{i \to \infty} \phi(x_{l(i)} - \sum_{j=1}^{l(i)-k+1} \alpha_{l(i)-j}d_{l(i)-j}) = \tilde{\phi}.
$$

The last equality together with (3.9) and (3.10) implies

$$
\lim_{k \to \infty} d_k = 0.
$$
Let $\bar{x}$ be arbitrary limit point of $\{x_k\}$. By the definition of $d_k$, we have

$$0 \in \nabla f(x_k) + \bar{\lambda}_k d_k + \partial \psi(x_k + d_k).$$

It follows from the boundedness of $\bar{\lambda}_k$, (3.8) and Lemma 3.1 that

$$0 \in \nabla f(\bar{x}) + \partial \psi(\bar{x}).$$

The proof is completed. \hfill $\Box$

4. Linear convergence rate.

4.1. Sublinear convergence for convex functions. In this section, we give a sublinear convergence estimate for the error between the objective function value $\phi(x_k)$ and the optimal function value, assuming $f$ is convex and Assumption 3.1 holds.

By the assumptions, (1.1) has a solution $x^* \in \Omega$ and an associated objective function value $\phi^* := \phi(x^*)$.

**Lemma 4.1.** Suppose that Assumption 3.1 holds and $f$ is convex. Then

$$\lim_{k \to \infty} \phi(x_k) = \phi^*.$$

**Proof.** By Theorem 3.1, there exists a constant $\tilde{\phi}$ such that

$$\lim_{k \to \infty} \phi(x_k) = \tilde{\phi}$$

and all accumulation points of the iterates $x_k$ are stationary points. An accumulation point exists since $K$ is compact and the iterations are all contained in $\Omega \subset K$. Since $f$ and $\psi$ are both convex, a stationary point is a global minimizer of $\phi$. Hence, $\tilde{\phi} = \phi^*$. \hfill $\Box$

Our sublinear convergence result is the following.

**Theorem 4.1.** Suppose that Assumption 3.1 holds and $f$ is convex. Then, there exists a positive constant $a$ such that

$$\phi(x_k) - \phi^* \leq \frac{a}{k}$$

for all $k$, where $\phi^*$ is the optimal objective function values for (1.1).

**Proof.** Let $\Phi_{k-1}$ be defined by

$$\Phi_{k-1}(v) = f(x_{k-1}) + \nabla f(x_{k-1})^T v + \frac{\lambda_{k-1}}{2} \|v\|^2 + \psi(x_{k-1} + v).$$
By (3.7) and Lemma (2.1), we have

\[
\phi(x_k) - \Phi_{k-1}(d_{k-1}) = f(x_k) + \psi(x_k) - f(x_{k-1}) - \nabla f(x_{k-1})^T d_{k-1} - \frac{\lambda_{k-1}}{2} \|d_{k-1}\|^2 - \psi(x_{k-1} + d_{k-1}) \leq f(x_k) - f(x_{k-1}) - \nabla f(x_{k-1})^T d_{k-1} + \psi(x_{k-1} + \alpha_{k-1}d_{k-1}) - \psi(x_{k-1} + d_{k-1}) \leq \alpha_{k-1}\frac{L}{2} \|d_{k-1}\|^2 + \alpha_{k-1}\nabla f(x_{k-1})^T d_{k-1} - \nabla f(x_{k-1})^T d_{k-1} + \psi(x_{k-1} + \alpha_{k-1}d_{k-1}) - \psi(x_{k-1} + d_{k-1}) \leq (\alpha_{k-1} - 1)\nabla f(x_{k-1})^T d_{k-1} + \psi(x_{k-1} + d_{k-1}) - \psi(x_{k-1}) + \frac{L}{2} \|d_{k-1}\|^2 \]

\[
= -(1 - \alpha_{k-1}) \Delta_{k-1} + \frac{L}{2} \|d_{k-1}\|^2 \leq -(1 - \alpha_{k-1} + \frac{L}{\lambda_{\min}}) \Delta_{k-1} \leq -(1 - c + \frac{L}{\lambda_{\min}}) \Delta_{k-1}
\]

(4.12)

where the second inequality follows from (ii) of Lemma 2.1, the third inequality is due to (i) of Lemma 2.1 and the fourth inequality follows from Lemma 2.2. By the definition of \(d_{k-1}\), we have

\[
d_{k-1} = \arg \min_{v \in \mathbb{R}^n} \{\nabla f(x_{k-1})^T v + \frac{\lambda_{k-1}}{2} \|v\|^2 + \psi(x_{k-1} + v)\} = \arg \min_{v \in \mathbb{R}^n} \{f(x_{k-1}) + \nabla f(x_{k-1})^T v + \frac{\lambda_{k-1}}{2} \|v\|^2 + \psi(x_{k-1} + v)\} = \arg \min_{v \in \mathbb{R}^n} \{\Phi_{k-1}(v)\}.
\]

It follows that

\[
\Phi_{k-1}(d_{k-1}) = \min_{v \in \mathbb{R}^n} \{f(x_{k-1}) + \nabla f(x_{k-1})^T v + \frac{\lambda_{k-1}}{2} \|v\|^2 + \psi(x_{k-1} + v)\} \leq \min_{v \in \mathbb{R}^n} \{f(x_{k-1} + v) + \psi(x_{k-1} + v) + \frac{\lambda_{k-1}}{2} \|v\|^2\} \leq \min_{v \in \mathbb{R}^n} \{\phi(x_{k-1} + v) + \frac{\lambda_{k-1}}{2} \|v\|^2\}
\]

(4.13)

where the first inequality is due to the convexity of \(f\). Combining (4.12) and (4.13) yields

\[
\phi(x_k) \leq \min_{v \in \mathbb{R}^n} \{\phi(x_{k-1} + v) + \frac{\lambda_{\max}}{2} \|v\|^2\} - b_1 \Delta_{k-1}.
\]

(4.14)
We take 

\[ v = \lambda (x^* - x_{k-1}), \]

where \( \lambda \in [0, 1] \) and \( x^* \) is a solution of (1.1). By the convexity of \( \phi \), we have

\[
\min_{v \in \mathbb{R}^n} \{ \phi(x_{k-1} + v) + \frac{\lambda_{\max}}{2} \|v\|^2 \} \leq \phi((1 - \lambda)x_{k-1} + \lambda x^*) + \frac{\lambda_{\max}}{2} \|x_{k-1} - x^*\|^2
\]

\[
\leq (1 - \lambda) \phi(x_{k-1}) + \lambda \phi^* + \frac{\lambda_{\max}}{2} \|x_{k-1} - x^*\|^2
\]

(4.15)

\[
= (1 - \lambda) \phi(x_{k-1}) + \lambda \phi^* + \beta_k \lambda^2,
\]

where \( \beta_k := \frac{\lambda_{\max}}{2} \|x_{k-1} - x^*\|^2 \). Since both \( x_{k-1} \) and \( x^* \) lie in \( \Omega \), it follows that

\[
\beta_k := \frac{\lambda_{\max}}{2} \|x_{k-1} - x^*\|^2 \leq \frac{\lambda_{\max}}{2} (\text{diameter of } K)^2 := b_2 < +\infty.
\]

Combining the bound for \( \beta_k \) with (4.14) and (4.15) gives

(4.16) \[ \phi(x_k) \leq (1 - \lambda) \phi(x_{k-1}) + \lambda \phi^* + b_2 \lambda^2 - b_1 \Delta_{k-1} \]

for any \( \lambda \in [0, 1] \). Again, define

\[ \phi(x_{l(i)}) = \max \{ \phi(x_k) \colon (i - 1)M < i \leq iM \} = \phi_{lM}^{\max}. \]

Since \( \phi(x_{k+l}) \leq \phi_{lM}^{\max} \) in Step 3 of NBB, it follows that \( \phi_{lM}^{\max} \) is a nonincreasing function sequence of \( k \). By (4.16) with \( k = l(i) \) and the monotonicity of \( \phi_{lM}^{\max} \), we have

(4.17) \[ \phi(x_{l(i)}) \leq (1 - \lambda) \phi(x_{l(i-1)}) + \lambda \phi^* + b_2 \lambda^2 - b_1 \Delta_{l(i-1)}. \]

Step 3 of NBB implies that

\[ -b_1 \Delta_{k-1} \leq b_3 (\phi_{lM}^{\max} - \phi(x_k)), \]

where \( b_3 = \frac{b_1}{a_0} \). Again, we take \( k = l(i) \) and exploit the monotonicity of \( \phi_{lM}^{\max} \) to obtain

(4.18) \[ -b_1 \Delta_{l(i)-1} \leq b_3 (\phi(x_{l(i-1)}) - \phi(x_{l(i)})) \]

Letting \( \phi_i \) denote \( \phi(x_{l(i)}) \) and combining (4.17)-(4.18) gives

\[
(1 + b_3) \phi_i \leq b_2 \lambda^2 + \lambda (\phi^* - \phi_{i-1}) + (1 + b_3) \phi_{i-1}.
\]

for every \( \lambda \in [0, 1] \). The minimum value on the right-hand side is attained with the choice

\[ \bar{\lambda} = \min \{ 1, \frac{\phi_{i-1} - \phi^*}{2b_2} \}. \]

By Lemma 4.2, there exists an positive integer \( i_0 \) such that \( \bar{\lambda} = \frac{\phi_{i-1} - \phi^*}{2b_2} \) for all \( i > i_0 \). Thus, for all \( i > i_0 \), we get

\[
(1 + b_3) \phi_i \leq (1 + b_3) \phi_{i-1} - \frac{(\phi^* - \phi_{i-1})^2}{4b_2}.
\]

That is,

\[
\phi_i \leq \phi_{i-1} - \frac{(\phi^* - \phi_{i-1})^2}{4b_2(1 + b_3)}, \quad \forall i > i_0.
\]

Denote \( \phi_i - \phi^* \) by \( e_i \). From above inequality, we get that for all \( i > i_0 \)

\[
e_i \leq e_{i-1} - \frac{(e_{i-1})^2}{4b_2(1 + b_3)}.
\]
which implies that \(\{e_i\}_{i > i_0}\) is a nondecreasing sequence and
\[
e_i \leq e_{i-1} - \frac{e_{i-1} e_i}{4b_2(1+b_3)}, \quad \forall i > i_0.
\]
By dividing both side of the above inequality by \(e_{i-1} e_i\), we get
\[
\frac{1}{e_i} \geq \frac{1}{e_{i-1}} + b_4, \quad \forall i > i_0
\]
where \(b_4 = \frac{1}{4b_2(1+b_3)}\). Applying this inequality recursively gives
\[
\frac{1}{e_i} \geq \frac{1}{e_{i_0}} + \frac{1}{(i - i_0)b_4}, \quad \forall i > i_0.
\]
Simplifying the last inequality, we obtain
\[
e_i \leq \frac{1}{(i - i_0)b_4}
\]
for all \(i > i_0\). If \(i \geq 2i_0\), then we have
\[
e_i \leq \frac{1}{(i - i_0)b_4} = \frac{1}{i(i - i_0)b_4} \leq \frac{2}{ib_4},
\]
Since there is a finite number of positive integer \(i \in [1, 2i_0]\), we can choose \(b_5 > \frac{2}{b_4}\) large enough such that \(e_i \leq \frac{b_5}{i}\) for all integers \(i \in [1, 2i_0]\). Then for all \(i\), we have
\[
e_i \leq \frac{b_5}{i}.
\]
For all \(k \in ((i - 1)M, iM]\), we get \(i \geq \frac{k}{M}\) and
\[
\phi(x_k) - \phi^* \leq e_i \leq \frac{b_5}{i} \leq \frac{Mb_5}{k}.
\]
The proof is completed by taking \(a = Mb_5\). \(\square\)

### 4.2. Linear convergence for strongly convex functions.

In this section, we establish R-linear convergence for our method when \(f\) is strongly convex. Recall that \(f\) is strongly convex if there exists a scalar \(\mu > 0\) such that
\[
f(x) \geq f(y) + \nabla f(y)^T (x - y) + \frac{\mu}{2} \|x - y\|^2
\]
or
\[
(\nabla f(x) - \nabla f(y))^T (x - y) \geq \mu \|x - y\|^2
\]
for all \(x\) and \(y \in \mathbb{R}^n\).

Let \(d(x)\) satisfy
\[
d(x) = \arg \min_{v \in \mathbb{R}^n} \{\nabla f(x)^T v + \lambda \|v\|^2 + \psi(x + v)\}
\]
where \(\lambda \in [\lambda_{\text{min}}, \lambda_{\text{max}}]\). The following Lemmas 4.2 and 4.3 play an important role on establishing the R-linear convergence of the method.

**Lemma 4.2.** Suppose that Assumption 3.1 holds and \(f\) is strongly convex. Then we have
\[
\|x - x^*\| \leq \frac{L + \lambda_{\text{max}}}{\mu} \|d(x)\|.
\]
Lemma 4.3. Suppose that Assumption 3.1 holds and \( \phi \) is strongly convex. Then, for any \( x^* \in X \) (in fact, \( X \) is a singleton), we have

\[
(\nabla f(x) + \lambda d(x))^T d(x) + \psi(x + d(x)) \leq (\nabla f(x) + \lambda d(x))^T (x^* - x) + \psi(x^*),
\]

which shows that

\[
0 \in \{\nabla f(x^*) + \partial \psi(x^*)\}.
\]

This implies that

\[
x^* \in \arg \min_{v \in \mathbb{R}^n} \{\nabla f(x^*)^T v + \psi(v)\}.
\]

Hence, for \( v = x + d(x) \), we have

\[
(4.21) \quad \nabla f(x^*)^T x^* + \psi(x^*) \leq \nabla f(x^*)^T (x + d(x)) + \psi(x + d(x)).
\]

Adding the inequalities (4.20) and (4.21) and simplifying yield

\[
(\nabla f(x) + \lambda d(x))^T (x + d(x) - x^*) \leq \nabla f(x^*)^T (x + d(x) - x^*)
\]

which shows that

\[
(\nabla f(x) - \nabla f(x^*))^T (x - x^*) \leq -\lambda \|d(x)\|^2 + (\nabla f(x^*) - \nabla f(x))^T d(x) + \lambda d(x)^T (x^* - x)
\]

By (3.7) and (4.19), we get

\[
\mu \|x - x^*\|^2 \leq L \|x - x^*\| \|d(x)\| + \lambda \|d(x)\| \|x - x^*\| \leq L \|x - x^*\| \|d(x)\| + \lambda_{\text{max}} \|d(x)\| \|x - x^*\|.
\]

Rearranging the term of the above inequality, we get the conclusion.

Lemma 4.3. Suppose that Assumption 3.1 holds and \( f \) is strongly convex. Then, for any \( \bar{x} \in \mathbb{R}^n \), \( \alpha \in (0, 1] \) and \( x' = x + \alpha d(x) \), we have

\[
\nabla f(x)(x' - \bar{x}) + \psi(x') - \psi(\bar{x}) \leq -c_2 \triangle(x)
\]

where \( \triangle(x) = \nabla f(x)^T d(x) + \psi(x + \alpha d(x)) - \psi(x) \) and \( c_2 \) is a positive constant.

Proof. Again, by the definition of \( d(x) \) and the convexity of \( \phi \), we have

\[
d(x) \in \arg \min_{v \in \mathbb{R}^n} \{\nabla f(x) + \lambda d(x))^T v + \psi(x + v)\}.
\]

Thus, we have

\[
(\nabla f(x) + \lambda d(x))^T d(x) + \psi(x + d(x)) \leq (\nabla f(x) + \lambda d(x))^T (\bar{x} - x) + \psi(\bar{x}).
\]

Namely,

\[
(4.22) \quad (\nabla f(x) + \lambda d(x))^T (x + d(x) - \bar{x}) + \psi(x + d(x)) \leq \psi(\bar{x}).
\]
Suppose that Assumption 3.1 holds and the function is strongly convex. Then, following from (4.22), the four inequality follows from Lemma 4.2 and the fifth inequality follows from Lemma 4.2.

\[
\begin{align*}
\nabla f(x)^T(x' - \bar{x}) + \psi(x') - \psi(\bar{x}) &= (\nabla f(x) + \lambda d(x))^T(x' - \bar{x}) + \psi(x') - \psi(\bar{x}) - \lambda d(x)^T(x' - \bar{x}) \\
&= (\nabla f(x) + \lambda d(x))^T(x + \alpha d(x) - \bar{x}) + \psi(x + \alpha d(x)) - \psi(\bar{x}) - \lambda d(x)^T(x' - \bar{x}) \\
&= (\alpha - 1)(\nabla f(x) + \lambda d(x))^T d(x) + (\nabla f(x) + \lambda d(x))^T(x + d(x) - \bar{x}) + \psi(x + d(x)) - \lambda d(x)^T(x' - \bar{x}) \\
&= (\alpha - 1)(\nabla f(x)^T d(x) + \psi(x + d(x)) - \psi(x) + \lambda \|d(x)\|^2) + (\nabla f(x) + \lambda d(x))^T(x + d(x) - \bar{x}) \\
&+ \psi(x + d(x)) - \psi(\bar{x}) - \lambda d(x)^T(x' - \bar{x}) \\
&\leq (\alpha - 1)(\nabla f(x)^T d(x) + \psi(x + d(x)) - \psi(x) + \lambda \|d(x)\|^2) - \lambda d(x)^T(x' - \bar{x}) \\
&\leq -(1 - \alpha) \Delta(x) + \lambda \|d(x)\| \|x + \alpha d(x) - \bar{x}\| \\
&\leq -(1 - \alpha) \Delta(x) + \lambda (1 + \frac{\lambda_{\max} + L}{\mu}) \|d(x)\|^2 \\
&\leq -(1 - \alpha) \Delta(x) - 2 \frac{\lambda_{\max}}{\lambda_{\min}} (1 + \frac{\lambda_{\max} + L}{\mu}) \Delta(x) \\
&\overset{\text{def}}{=} -(1 - \alpha + 2 \frac{\lambda_{\max}}{\lambda_{\min}} (1 + \frac{\lambda_{\max} + L}{\mu})) \Delta(x)
\end{align*}
\]

where the first inequality follows from the convexity of $\psi$, the second inequality follows from (4.22), the four inequality follows from Lemma 4.2 and the fifth inequality follows from Lemma 2.2.

The following theorem shows that the method converges R-linearly if $f$ is strongly convex.

**Theorem 4.2.** Suppose that Assumption 3.1 holds and $f$ is strongly convex. Then there exist constant $\theta \in (0, 1)$ and $\bar{c}$ such that

\[
\phi(x_k) - \phi^* \leq \bar{c} \theta^k (\phi(x_1) - \phi^*)
\]

for every $k$.

**Proof.** As in the proof of Theorem 3.1, we have

\[
\phi(x_{l(i)}) - \phi(x_{l(i-1)}) \leq \delta c \Delta_{l(i-1)}.
\]
On the other hand, for some \( \xi_l(i) \) lying on the line segment jointing \( x_l(i) \) and \( x^* \), we have
\[
\phi(x_l(i)) - \phi^* = f(x_l(i)) - f(x^*) + \psi(x_l(i)) - \psi(x^*) \\
= \nabla f(x_l(i))^T (x_l(i) - x^*) + \psi(x_l(i)) - \psi(x^*) \\
= (\nabla f(x_l(i)) - \nabla f(x_l(i-1)))^T (x_l(i) - x^*) \\
+ \nabla f(x_l(i-1))^T (x_l(i-1) - x^*) + \psi(x_l(i)) - \psi(x^*) \\
\leq L(\|x_l(i) - x_l(i-1)\| + \|x_l(i-1) - x^*\|)\|x_l(i) - x^*\| - c_2 \Delta_l(i-1) \\
\leq L(\|x_l(i) - x_l(i-1)\| + \|x_l(i-1) - x^*\|)\|x_l(i) - x_l(i-1) + x_l(i-1) - x^*\| \\
- c_2 \Delta_l(i-1) \\
\leq L(1 + \frac{\lambda_{\text{max}} + L}{\mu})^2 d_{l(i-1)}^2 - c_2 \Delta_l(i-1) \\
\leq -L(1 + \frac{\lambda_{\text{max}} + L}{\mu})^2 \frac{2}{\lambda_{\text{min}}} \Delta_l(i-1) - c_2 \Delta_l(i-1) \\
\tag{4.24} \overset{\text{def}}{=} -c_3 \Delta_l(i-1)
\]
where the first inequality follows from Lemma 4.3 and the fifth inequality follows from Lemma 4.2. Combining (4.23) and (4.24) yields
\[
\phi(x_l(i)) - \phi^* \leq c_4 (\phi(x_l(i-1)) - \phi^*) \leq c_4^{i-1} (\phi(x_1) - \phi^*)
\]
where \( c_4 = \frac{c_3}{c_3 + 2\delta} \). For \( k \in (i-1)M,iM \), we have \( i \geq \frac{k}{M} \) and
\[
\phi(x_k) - \phi^* \leq \phi(x_l(i)) - \phi^* \leq \frac{1}{c_4} (c_4^{\frac{k}{M}}) (\phi(x_1) - \phi^*).
\]
Letting \( \hat{c} = \frac{1}{c_4} \) and \( \theta = c_4^{\frac{k}{M}} \), we get the conclusion. \( \square \)

5. Numerical experiments. In this section, we do some numerical experiments to illustrate the performance of the NBB method. We compare the performance of the NBB method with that of \( \ell_1 \) ls [28], FPCBB [25], IST [21] and FISTA [4]. All codes are written in MATLAB 7.0 and all tests described in this section were performed on a PC with Intel I5-3230 2.6GHZ CPU processor and 4G RAM memory with a Windows operating system. We implemented the NBB method with the following parameters: \( \lambda_{\text{min}} = 10^{-10} \), \( \lambda_{\text{max}} = 10^{10} \), \( \delta = 10^{-4} \), \( M = 5 \) and \( \rho = 0.5 \). The other four algorithms with default parameters were performed. The test problems are associated with applications in the areas of signal processing and image reconstruction. The initial point for all tested algorithms is the zero vector. For the stopping condition in subsections 5.1 - 5.3, we first run the \( \ell_1 \) ls algorithm to set a benchmark objective value and then ran other algorithms until they reaches the benchmark value of the objective function.

5.1. \( \ell_2 - \ell_1 \) problem. In this subsection, we do some numerical experiments to show the performance of the NBB method by solving \( \ell_2 - \ell_1 \) problems of form (1.2). We consider a typical compressed sensing scenario, where the goal is to reconstruct a length-\( n \) sparse signal from \( m \) observations, where \( m < n \). In this case, the \( m \times n \) matrix \( A \) is obtained by first filling it with independent samples of a standard Gaussian distribution and then orthonormalizing the rows. These random matrices
are generated by using MATLAB command randn. The observed vector is all both
\( b = Ax + \epsilon \), where the noise \( \epsilon \) is sampled from a Gaussian distribution with mean zero and variance \( 10^{-4} \) and the original signal \( x_{\text{ture}} \) contains \( T \) randomly placed
\( \pm 1 \) spikes, with zeros in other components. The regularization parameter \( \mu \) is taken as
\[
\mu = 0.1\|A^T b\|_\infty.
\]
Notice that \( \mu \geq \|A^T b\|_\infty \), the unique minimum of (1.2) is the zero vector [28]. In
the experiments, we fix the matrix \( A \) size \( n = 4096 \) and \( m = \text{round}(0.1 \times n) \) (the
similar experiment data is obtained for \( m = \text{round}(0.2 \times n) \) and \( m = \text{round}(0.3 \times n) \))
and consider a range of degrees of sparseness: the number \( T \) of nonzero spikes
in \( x_{\text{ture}} \) ranges from 1 to 50. For each value of \( T \), we generate ten random data
sets. For each data set \( (x_{\text{ture}}, A, b) \), we first ran \( l1_ls \) and store the final value
of the objective function and then ran other algorithms until they reach the same
objective function value. Finally, we compute average MSE
\[
\text{MSE} = \frac{1}{n} \| \bar{x} - x_{\text{ture}} \|_2^2
\]
and average CPU time, over the ten runs.

We present our numerical results by using performance profiles as proposed in
[18]. A performance plot for the CPU time is presented in Figure 1 and a perfor-
mance plot for the MSE is presented in Figure 2. From Figures 1-2, we can observe
that our method is faster than other methods and obtain a better MSE.

5.2. Group-separable regularizer. In this subsection, we examine performance
using the group separable regularizer [38] for which
\[
\phi(x) = \frac{1}{2} \| Ax - b \|_2^2 \quad \text{and} \quad \psi(x) = \mu \sum_{i=1}^{n} ||x_{[i]}||_2,
\]
where \( x_{[1]}, x_{[2]}, \cdots, x_{[m]} \), are \( m \) disjoint subvectors of \( x \) and \( A \in R^{1024 \times 4096} \) was
obtained by the same way as that in subsection 5.1. The vector \( x_{\text{ture}} \) has 4096
components, divided into \( m = 64 \) groups of length \( l_i = 64 \). To generate \( x_{\text{ture}} \), we
randomly chose from one to eight groups and filled them with zero-mean Gaussian
random samples of unit variance, while all other groups are filled with zeros. The
target vector is \( b = Ax_{\text{ture}} + n \), where the noise \( n \) is a Gaussian noise with mean
zero and variance \( 10^{-4} \). The regularization parameter is chosen as suggested in [38]:
Figure 3. From top to bottom: original signal reconstruction via the minimization of (1.2) obtained by IST, FISTA, NBB and FPC_BB.

$$\mu = 0.3 \| A^T b \|_\infty.$$ We ran 10 test problems and gives the average CPU time needed by four methods in Table 2.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>FISTA</th>
<th>NBB</th>
<th>FPC_BB</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU time</td>
<td>1.9890</td>
<td>0.8015</td>
<td>0.5343</td>
</tr>
</tbody>
</table>

From Table 2, we can observe that NBB is faster than $\ell_1$ ls and is comparable to FPC_BB and FISTA.

5.3. **Image deblurring problem.** In this subsection, we present results for one image restoration problems referred to as Cameraman (see Figure 7). The images are $256 \times 256$ grayscale images; that is, $n = m = 256^2 = 65536$. The image restoration problem has the form (1.2), where $\mu = 0.00005$. Table 3 reports the average CPU time. The results in Table 3 and Figure 3 again indicate that NBB yields much better performance for the test problem.

Table 3. Statical data
6. Conclusions. In the paper, we introduce an algorithm framework for the more general problem of minimizing the sum \( f(x) + \psi(x) \), where \( f \) is smooth and \( \psi \) is convex, but possible nonsmooth. At each step, the search direction of the algorithm is obtained by solving an optimization problem involving a quadratic term with diagonal Hessian and Barzilai-Borwein steplength plus \( \psi(x) \). The method with the nonmonotone line search techniques is showed to be globally convergent. We also show that the method shares a sublinear global rate of convergence when \( f \) is convex and the convergence is R-linear when \( f \) is strongly convex. The numerical results presented in Section 5 demonstrate the effectiveness of the algorithm for solving sparse reconstruction problems of varying difficulties.

REFERENCES


Received May 2014; revised September 2014.

E-mail address: chengwanyou@sina.com
E-mail address: zixin@dgut.edu.cn
E-mail address: dhli@scnu.edu.cn