AIMS Mathematics, 9(6): 16335–16353.
DOI: 10.3934/math.2024791
Received: 05 March 2024
Revised: 23 April 2024
Accepted: 29 April 2024
Published: 09 May 2024

https://www.aimspress.com/journal/Math

Research article

Nonmonotone variable metric Barzilai-Borwein method for composite minimization problem

Xiao Guo[1], **Chuanpei Xu**[1], **Zhibin Zhu**[2] **and Benxin Zhang**[1,*]

[1] School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China

[2] School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin 541004, China

* **Correspondence:** Email: bxzhang@guet.edu.cn.

Abstract: In this study, we develop a nonmonotone variable metric Barzilai-Borwein method for minimizing the sum of a smooth function and a convex, possibly nondifferentiable, function. At each step, the descent direction is obtained by taking the difference between the minimizer of the scaling proximal function and the current iteration point. An adaptive nonmonotone line search is proposed for determining the step length along this direction. We also show that the limit point of the iterates sequence is a stationary point. Numerical results with parallel magnetic resonance imaging, Poisson, and Cauchy noise deblurring demonstrate the effectiveness of the new algorithm.

Keywords: variable metric; Barzilai-Borwein method; nonsmooth optimization; total variation
Mathematics Subject Classification: 65K10, 90-08, 90C30

1. Introduction

In this paper, we consider the following composite optimization problem:

$$\min_{x\in R^n} \phi(x) = f(x) + h(x), \tag{1.1}$$

where $f(x)$ is a differentiable function with a Lipschitz constant $L > 0$ for the gradient and $h(x)$ is a convex function that is possibly nonsmooth. Many models in sparse optimization can be formulated as problems (1.1), such as the l_1-regularized minimization in compressive sensing [1–3] and total variation regularization in image processing [4–6].

A special case of model (1.1) is the constrained problem over a convex set, when $h(x)$ is the indicator function of a convex set Ω, i.e.,

$$h(x) = l_\Omega(x)$$

and

$$l_{\Omega}(x) = \begin{cases} 0, & if \quad x \in \Omega, \\ +\infty, & if \quad x \notin \Omega. \end{cases}$$

For this case, a simple and famous algorithm is the gradient projection method [7]. To accelerate this projection algorithm, a scheme with line-search along the feasible direction is proposed in [8]. Using the Barzilai-Borwein (BB) stepsize [9] and nonmonotone line search strategy, Birgin et al. [10] developed the nonmonotone spectral projected gradient method, which is particularly appealing for large-scale problems. Dai and Fletcher [11] induced a new projected gradient method that alternately used the two BB steplengths. In [12], the authors proposed a modified spectral conjugate gradient projection algorithm and applied it to the total variation image restoration.

The other special case of model (1.1), which has attracted much interest in signal and image processing and machine learning, is the well-known l_2-l_1 problem, i.e.,

$$h(x) = \|\cdot\|_1.$$

One of the most popular methods is the iterative shrinkage thresholding algorithm [13]. Using Nesterov's technique, Beck and Teboulle [14] proposed the famous fast iterative shrinkage thresholding algorithm (FISTA) by forward-backward splitting. Hale et al. [2] developed the fixed-point continuation method for l_1-minimization. They also used BB stepsize and nonmonotone line search to enhance this algorithm in [15]. Another closely related method is the sparse reconstruction by separable approximation algorithm (SpaRSA) [16], which involves minimizing a non-smooth convex problem with separable structures. Hager et al. [17] showed that SpaRSA has a sublinear convergence rate for general convex functions or R-linear when the objective function is strongly convex. An improved version of SpaRSA based on a cyclic BB iteration and adaptive line search was also presented in [17]. Huang and Liu [18] proposed a BB-type method for minimizing composite function. A nonmonotone strategy was employed for determining the step length along the search direction, which was defined by the difference between the minimizer of the approximation objective function and the current iteration point. Recently, Cheng et al. [19] induced an algorithm framework for the more general type of (1.1). At each step, the search direction was obtained by solving an optimization problem. Xiao et al. [20] proposed a nonmonotone BB gradient algorithm for l_1-regularized nonsmooth minimization in compressive sensing. At each iteration, the search direction can be easily derived by minimizing a local approximal quadratic model. Tseng and Yun [21] gave a block coordinate gradient descent method. But numerical results showed the efficiency of this method only for small and medium-scale problems.

There are many variable metric forward-backward (FB) approaches for solving (1.1) in recent literatures. In this type of algorithm, the suitable symmetric positive definite scaling matrices changed at each iteration. A clever choice of them can lead to an improvement in convergence speed [22, 23]. Bonettini et al. [24] developed a variable metric inexact-line-search-based method for nonsmooth optimization. An Armijo-like rule was used to determine the stepsize along the descent direction. For the general nonconvex case, they proved that all limit points of the iterate sequence are stationary. In the last two decades, more variants of FB approaches have been proposed. These FB variants introduce a kind of extrapolation, or inertial, step exploiting the two previous iterations and can be traced back to two main typologies: FISTA-like methods [25] and heavy-ball ones [26]. Other recent

results can be found in [27, 28], where the analysis applies also to the nonconvex case, assuming that the Kurdyka-Lojasiewicz property holds.

In this paper, following the same lines as in [24] that exploiting some variable metric information and nonmonotone line search in [18, 19] can lead to fast image restorations, we propose a BB-type algorithm with adaptive nonmonotone line search and scaling technique. The nonmonotone strategy includes a convex combination of the maximum function value of some previous iterations and the current function value. At each step, the search direction of the algorithm is obtained by solving an optimization subproblem involving a quadratic term with a variable metric matrix and BB steplength plus $h(x)$. Then, we perform a nonmonotone line search along this direction. We prove that the method with the nomonotone line search techniques is globally convergent. Similar to [17, 18, 24], the new method does not need to know the value of the Lipschitz constant L. Then, we apply this method to total variation image restoration, such as parallel magnetic resonance imaging [29, 30], Poisson and Cauchy noise removal [24]. Numerical experiments show that our approach is competitive with several other known methods.

The rest of the paper is organized as follows: In Section 2, we present the new method for solving the model (1.1) . The convergence of the new algorithm is proved in Section 3. In Section 4, we show the results of the numerical experiments. Finally, the conclusion is given in Section 5.

Throughout this paper, we denote $\langle x, y\rangle = x^T y$ as the inner product of two vectors $x, y \in R^n$. $\|\cdot\|$ denotes the Euclidean norm. The scaled Euclidean norm associated with a symmetric positive definite matrix D is

$$\|x\|_D^2 = x^T D x.$$

Given

$$\mu_{\max} > \mu_{\min} > 0,$$

$M_{[\mu_{\min},\mu_{\max}]}$ denotes the set of all symmetric positive definite matrix with all eigenvalues contained in the interval $[\mu_{\min}, \mu_{\max}]$. For any

$$D \in M_{[\mu_{\min},\mu_{\max}]},$$

we have the result

$$\mu_{\min}\|x\|^2 \le \|x\|_D^2 \le \mu_{\max}\|x\|^2. \tag{1.2}$$

2. Proposed method

In this section, we first introduce the search direction and then introduce a new nonmonotone line search to determine the step length along this direction. At last, the new method is presented.

Our iteration step for solving (1.1) can be described as follows:

$$x_{k+1} = x_k + \lambda_k d_k, \tag{2.1}$$

where $\lambda_k \in (0, 1]$ is the step length and $d_k \in R^n$ is the search direction, which is given by

$$d_k := \arg\min_{v\in R^n}\{\nabla f(x_k)^T v + \frac{\alpha_k}{2} v^T D_k v + h(x_k + v)\}, \tag{2.2}$$

where $\alpha_k > 0$, and D_k is a symmetric positive definite matrix. Let

$$\tilde{x}_k = x_k + d_k,$$

from (2.2), we can get that $\tilde{x}_k$ is the minimizer of

$$\min_{z\in R^n} \ f(x_k) + \langle z - x_k, \nabla f(x_k)\rangle + \frac{\alpha_k}{2}\|z - x_k\|^2_{D_k} + h(z). \tag{2.3}$$

In fact, the problem (2.3) can be rewritten as the proximity operator style associated with the convex function $h(x)$, which is defined as

$$\tilde{x}_k = prox_h^{\alpha_k D_k}(y_k) = \arg\min_{z\in R^n} \ h(z) + \frac{\alpha_k}{2}\|z - y_k\|^2_{D_k}, \tag{2.4}$$

where

$$y_k = x_k - (\alpha_k D_k)^{-1}\nabla f(x_k).$$

So, (2.1) is equivalent to

$$x_{k+1} = x_k + \lambda_k(prox_h^{\alpha_k D_k}(x_k - (\alpha_k D_k)^{-1}\nabla f(x_k)) - x_k). \tag{2.5}$$

The setting of D_k and α_k will affect the performance of the method (2.5). A clever choice of them can lead to significant improvements in the practical convergence speed [31, 32]. Firstly, about the metric selection, we assume that

$$D_k \in M_{[\mu_{\min},\mu_{\max}]}.$$

Different problems can choose a different scaling matrix D_k. So, we will discuss how to choose D_k in the section on numerical experiments. Next, we will focus on how to choose the parameter α_k. Similar to [17–20], we choose α_k as follows:

$$\alpha_k = \frac{\langle s_{k-1}, y_{k-1}\rangle}{\langle s_{k-1}, s_{k-1}\rangle}, \tag{2.6}$$

where

$$s_{k-1} = x_k - x_{k-1}, \quad y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1}).$$

In fact, this step length was first proposed by Barzilai and Borwein for solving unconstrained problems [9]. Due to its efficiency, the BB approach has received considerable attention. As said in [9], a good property of α_k is that it satisfies the quasi-Newton condition

$$\alpha_k := \arg\min_{\alpha\in R} \ \|\alpha s_{k-1} - y_{k-1}\|^2.$$

We take

$$\alpha_k = \min\{\alpha_{\max}, \max\{\alpha_{\min}, \frac{\langle s_{k-1}, y_{k-1}\rangle}{\langle s_{k-1}, s_{k-1}\rangle}\}\}, \tag{2.7}$$

where $\alpha_{\max} > \alpha_{\min} > 0$ are fixed constants.

When the search direction is determined, a suitable stepsize along the decent direction should be found to determine the next iterative point. In the following, we will show how to determine the step length λ_k. We know that the performance of the BB algorithm benefits from the nonmonotone line search (see [10, 11]). The earliest nonmonotone line search technique, which differs from the traditional Armijo line search or the Wolfe-Powell line search, was given by Grippo et al. [33] for smooth, unconstrained optimization. In order to overcome some disadvantages of the nonmonotone line search, Amini et al. introduced a more relaxed nonmonotone strategy that included a convex

combination of the largest objective function in some recent past iteration and the current function value [34]. However, they are not directly applied to the nonsmooth problems. In [18], the authors proposed a line search for solving the nonsmooth problem (1.1) as follows:

$$\phi(x_k + \lambda d_k) \leq \phi(x_{l(k)}) - \frac{\gamma}{2}\lambda\alpha_k\|d_k\|^2, \tag{2.8}$$

where $\gamma \in (0, 1)$ is a constant and

$$\phi(x_{l(k)}) = \max_{0\leq j\leq \min\{k,M-1\}} \phi(x_{k-j}), \tag{2.9}$$

where M is a fixed integer and $l(k)$ is an integer such that

$$k - \min\{k, M - 1\} \leq l(k) \leq k.$$

Motivated by the idea from [18, 34], we modified the line search (2.8) and used the following acceptance criterion to determine the step length λ_k:

$$\phi(x_k + \lambda d_k) \leq R_k(\eta) - \frac{\gamma}{2}\lambda\alpha_k\|d_k\|^2_{D_k}, \tag{2.10}$$

where

$$R_k(\eta) = \eta\phi(x_{l(k)}) + (1 - \eta)\phi(x_k), \tag{2.11}$$

and $0 \leq \eta \leq 1$. Since

$$\phi(x_k) \leq \phi(x_{l(k)}),$$

we have

$$\phi(x_k) \leq R_k(\eta) \leq \phi(x_{l(k)}). \tag{2.12}$$

Given all the above derivations, we now describe the nonmonotone variable metric Barzilai-Borwein type algorithm as Algorithm 1 shows.

Algorithm 1 Variable metric BB method for solving (1.1).

Step 0: Given an initial point $x_0 \in R^n$, $k := 0$, $\eta > 0$,
M, $\alpha_{\min}$, $\alpha_{\max}$, $\mu_{\min}$, $\mu_{\max}$, $\gamma \in (0, 1)$.
Step 1: Stop if $\|d_k\| = 0$. Otherwise, continue.
Step 2: Compute d_k via (2.2),
where $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$, $D_k \in M_{[\mu_{\min},\mu_{\max}]}$.
Step 3: Compute λ_k via (2.10).
Step 4: Let $x_{k+1} = x_k + \lambda_k d_k$.
Step 5: Set $k := k + 1$ and go to Step 1.

Algorithm 1 is closely related to the method in [18]. However, compared (2.10) with (2.8), it can be easily found differences. Algorithm 1 use the adaptive nonmonotone line search and the variable metric matrix D_k to get the step length λ_k. If $\eta = 1$ and $D_k = I$, Algorithm 1 is reduced to the algorithms in [18]. Similar to the idea in [24], the variable metric technique is adopted to compute the descent direction, while the paper [24] used the modified Armijo line search to determine the step length λ_k.

3. Convergence analysis

In this section, we analyze the convergence of Algorithm 1.

First, we give some useful properties of d_k. In the nonsmooth case (1.1), d_k is a descent direction for $\phi(x)$ at x_k if $\phi'(x_k; d_k) < 0$, where $\phi'(x; d)$ is the direction derivative of $\phi(x)$ at x in the direction $d \in R^n$ and is defined as [35]

$$\phi'(x; d) = \lim_{t \to 0} \frac{\phi(x + td) - \phi(x)}{t}.$$

A feasible point x is said to be a stationary point of $\phi(x)$ if $\phi'(x; d) \geq 0$ for all d. Using the definition of $\phi'(x; d)$ and the property of stationary point, we show that the direction defined by (2.2) is descent if $d_k \neq 0$ and if $d_k = 0$, x_k is a stationary point of $\phi(x)$ in the following two lemmas. The proof is similar to the proofs of Lemmas 2.2 and 2.3 in [19]. So, we omit it here.

Lemma 3.1. *For any $x_k \in R^n$ and $d_k \in R^n$ determined by (2.2), we have*

$$\phi'(x_k; d_k) \leq \nabla f(x_k)^T d_k + h(x_k + d_k) - h(x_k) \leq -\frac{\alpha_k}{2} d_k D_k^T d_k.$$

Lemma 3.2. *$x_k \in R^n$ is a stationary point of (1.1) if and only if $d_k = 0$.*

Next, we show that Algorithm 1 is well defined.

Lemma 3.3. *Let $\{x_k\}$ be a sequence generated by Algorithm 1. If x_k is not a stationary point, then the line search (2.10) is satisfied whenever*

$$0 < \lambda \leq \min\{1, \frac{(1 - \gamma)\alpha_{\min}\mu_{\min}}{L}\}.$$

Moreover, if λ_k satisfies the inequality (2.10), then

$$\lambda_k \geq \min\{1, \frac{\rho(1 - \gamma)\alpha_{\min}\mu_{\min}}{L}\} := c, \tag{3.1}$$

where $\rho \in (0, 1)$.

Proof. By the inequality (2.12), the Lipschitz continuity of $\nabla f(x)$ with constant L, the convexity of $h(x)$, and Lemma 3.1, we can get

$$\begin{aligned} \phi(x_k + \lambda d_k) - R_k(\eta) &\leq \phi(x_k + \lambda d_k) - \phi(x_k) \\ &= f(x_k + \lambda d_k) - f(x_k) + h(x_k + \lambda d_k) - h(x_k) \\ &\leq \tfrac{\lambda^2 L}{2}\|d_k\|^2 + \lambda\langle \nabla f(x_k), d_k\rangle + \lambda(h(x_k + d_k) - h(x_k)) \\ &\leq \tfrac{\lambda^2 L}{2}\|d_k\|^2 - \tfrac{\lambda\alpha_k}{2}\|d_k\|_{D_k}^2. \end{aligned} \tag{3.2}$$

The use of inequality (2.10) yields

$$\frac{\lambda^2 L}{2}\|d_k\|^2 - \frac{\lambda\alpha_k}{2}\|d_k\|_{D_k}^2 \leq -\frac{\gamma}{2}\lambda\alpha_k\|d_k\|_{D_k}^2.$$

Using the assumption (1.2) and Lemma 3.2, we can have

$$\lambda \leq \frac{(1 - \gamma)\alpha_{\min}\mu_{\min}}{L}.$$

Now, we give the lower bound of λ_k. Assume that either $\lambda_k = 1$ or the line search (2.10) has failed at least once. Therefore, there is a constant a $\rho \in (0, 1)$ such that

$$\phi(x_k + \frac{\lambda_k}{\rho} d_k) > R_k(\eta) - \frac{\gamma}{2} \frac{\lambda_k}{\rho} \alpha_k \|d_k\|^2_{D_k}.$$

Combining the inequality (3.2) yields

$$\frac{\lambda_k}{\rho} L \|d_k\|^2 > (1 - \gamma)\alpha_k \|d_k\|^2_{D_k}.$$

So,

$$\lambda_k > \frac{\rho(1 - \gamma)\alpha_{\min}\mu_{\min}}{L}.$$

This completes the proof. □

The following theorem implies that every accumulation point of $\{x_k\}$ is a stationary point of (1.1). Moreover, if $f(x)$ is a convex function, Algorithm 1 converges to a global solution of (1.1).

Theorem 3.1. *Let the sequences $\{x_k\}$ and $\{d_k\}$ be generated by Algorithm 1, then*

$$\lim_{k\to\infty} \|d_k\| = 0 \tag{3.3}$$

and

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

where $\bar{\phi}$ is a constant.

Proof. Using the definition R_k, the inequalities (2.10) and (2.12), we have

$$\phi(x_{k+1}) \le R_k(\eta) - \frac{\gamma}{2} \lambda \alpha_k \|d_k\|^2_{D_k} \le \phi(x_{l(k)}) - \frac{\gamma}{2} \lambda \alpha_k \|d_k\|^2_{D_k}. \tag{3.4}$$

It indicates that

$$\phi(x_{k+1}) \le \phi(x_{l(k)}).$$

On the other hand, from (2.9), we get

$$\begin{aligned} \phi(x_{l(k+1)}) &= \max_{0\le j\le \min\{k+1,M\}} \phi(x_{k+1-j}) \\ &= \max\{\max_{1\le j\le \min\{k+1,M\}} \phi(x_{k+1-j}), \phi(x_{k+1})\} \\ &\le \max\{\phi(x_{l(k)}), \phi(x_{k+1})\} \\ &= \phi(x_{l(k)}). \end{aligned} \tag{3.5}$$

Then the sequence $\{\phi(x_{l(k)})\}$ is nonincreasing. Using the same assumption as the one in [19] that $\phi(x)$ is bounded below, there is a constant $\bar{\phi}$ such that

$$\lim_{k\to\infty} \phi(x_{l(k)}) = \bar{\phi}. \tag{3.6}$$

By applying (3.4) with k replaced by $l(k) - 1$, we have

$$\begin{aligned}\phi(x_{l(k)}) &= \phi(x_{l(k)-1} + \lambda_{l(k)-1} d_{l(k)-1}) \\ &\leq R_{l(k)-1} - \frac{\gamma}{2} c\alpha_{\min} \|d_{l(k)-1}\|^2_{D_{l(k)-1}} \\ &\leq \phi(x_{l(l(k)-1)}) - \frac{\gamma}{2} c\alpha_{\min}\mu_{\min} \|d_{l(k)-1}\|^2.\end{aligned}$$

Using (3.6), we obtain

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

By the continuity of $\phi(x)$, we deduce

$$\lim_{k\to\infty} \phi(x_{l(k)-1}) = \lim_{k\to\infty} \phi(x_{l(k)} - \lambda_{l(k)-1} d_{l(k)-1}) = \bar{\phi}.$$

Let $k = l(k) - j$, by induction and using the similar proof of Theorem 3.1 in [19] , we can show that the following limits are satisfied for $j = 1, 2, \cdots, M$,

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

For

$$1 \leq j = l(k) - k \leq M,$$

we have

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

This means that

$$\lim_{k\to\infty} \|x_k - x_{l(k)}\| = 0.$$

The facts with the continuity of $\phi(x)$ and (3.6) indicate that

$$\lim_{k\to\infty} \phi(x_k) = \lim_{k\to\infty} \phi(x_{l(k)}) = \bar{\phi}. \tag{3.7}$$

Taking the limits of (3.4) and (3.7), we obtain

$$\lim_{k\to\infty} \|d_k\| = 0.$$

From Lemma 3.2, we can also know that any limit point of $\{x_k\}$ is a stationary point of $\phi(x)$. Therefore, the proof is complete. □

4. Numerical experiments

In this section, we give some numerical experiments to illustrate the effectiveness of the proposed method. The first test problem of the form (1.1) is the total variation (TV) of magnetic resonance imaging (MRI). As a special type of (1.1), the parallel magnetic resonance imaging model [36, 37] can be written as follows:

$$\min_x \frac{1}{2}\|Ax - b\|^2 + \tau\|x\|_{TV},$$

where $x \in C^n$ and n is the total number of pixels in the image. $A \in C^{m\times n}$ is a sample matrix, and $b \in C^m$ is the vector of measured Fourier coefficients. $\|\cdot\|_{TV}$ is the total variation norm, which was first introduced in [5]. $\tau > 0$ is the regularization parameter. Obviously,

$$f(x) = \frac{1}{2}\|Ax - b\|^2$$

is Lipschitz continuous,

$$h(x) = \tau\|x\|_{TV}$$

is nonsmooth, and $\phi(x)$ is a convex function.

Our new method is named NMBBA, NVMBB, and NVMBBA, respectively, in different parameter settings. We compare it with the famous algorithms SpaRSA [16] and NMBB [18]. In all the experiments, we set

$$M = 5,\ \ \gamma = 0.001,\ \ \alpha_{\min} = 10^{-10},\ \ \alpha_{\max} = 10^{10},\ \ \mu_{\max} = \sqrt{1 + 1/k^2} = 1/\mu_{\min}.$$

In the first experiment, we use a Poisson mask with an acquisition rate 0.243. The image size is 256×256 (named Data1). The regularization parameter is set as $\tau = 0.0005$. In NMBBA , $\eta = 0.5$, $D_k = I$. In NVMBB,

$$\eta = 1, \quad D_k = 1/diag(x_k).$$

In NVMBBA,

$$\eta = 0.7, \quad D_k = 1/diag(x_k).$$

In the second experiment, we use radial mask with an acquisition rate 0.424. The image size is 512×512 (named Data2). The regularization parameter is set as $\tau = 0.00037$. In NMBBA, $\eta = 0.5$, $D_k = I$. In NVMBB,

$$\eta = 1, \quad D_k = 1/diag(x_k).$$

In NVMBBA,

$$\eta = 0.9, \quad D_k = 1/diag(x_k).$$

For the subproblem (2.4), we use the FISTA [14] method to solve the minimizing total variation proximal function. The inner iteration number is set to 5. The stopping criterion of the five algorithms is:

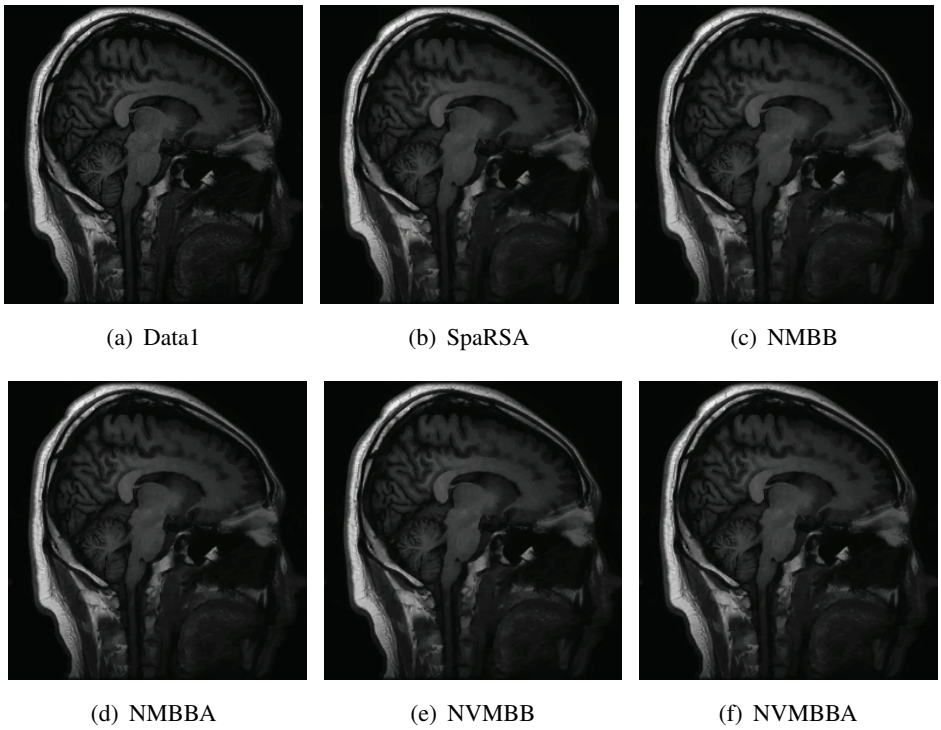
$$\frac{\|x_{k+1} - x_k\|}{\|x_{k+1}\|} \le 0.0001.$$

We use signal-to-noise ratio (SNR) as a means of judging performance, which is defined as follows:

$$SNR = 20\log_{10}\frac{\|x_k - \hat{x}\|}{\|x_k - x_0\|}(dB),$$

where x_0 is the original image, x_k is the recovery, and $\hat{x}$ is its mean value. All algorithms are implemented in the Matlab programming environment (Version R2011a). The experiments are performed on a laptop with 2 GB of RAM, a 2 GHz processor, and the Windows 7 operating system.

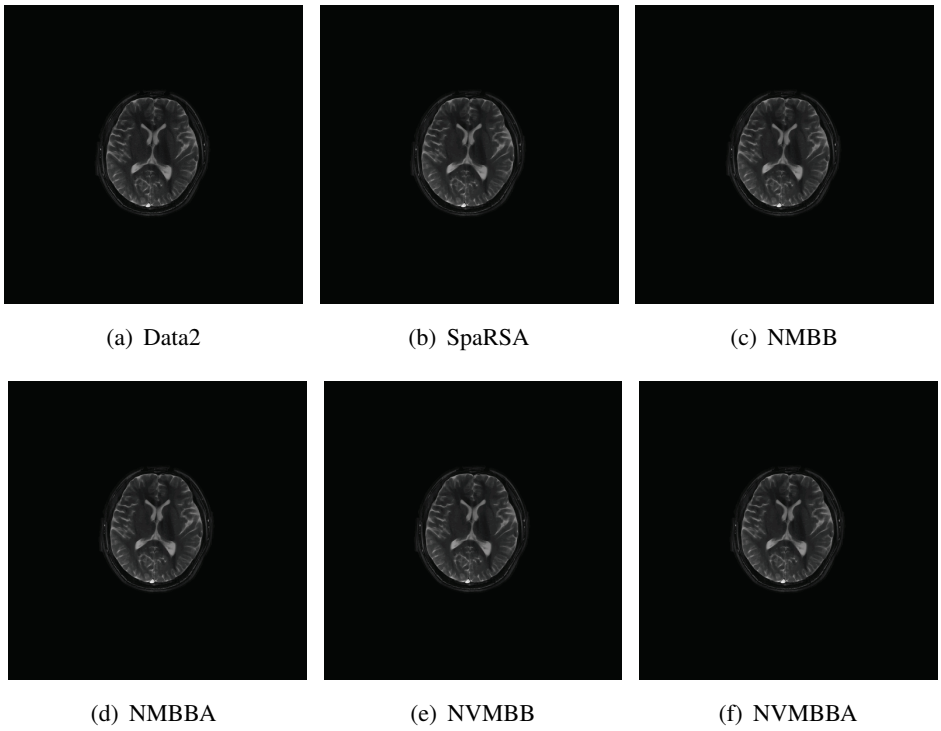
The reconstructed images are shown in Figures 1 and 2. It is seen that all algorithms adequately recovered the image from a small set of data samples. We can also see that NMBB appears to have higher accuracy than other methods for Data1 and NMBBA foe Data2.



(a) Data1 (b) SpaRSA (c) NMBB

(d) NMBBA (e) NVMBB (f) NVMBBA

Figure 1. Original image and reconstructed images of Data1 by different algorithms.



(a) Data2 (b) SpaRSA (c) NMBB

(d) NMBBA (e) NVMBB (f) NVMBBA

Figure 2. Original image and reconstructed images of Data2 by different algorithms.

This can also be seen in Table 1. From the table, we can see that the proposed method is very efficient in the sense that it reaches a similar SNR value in a shorter time (second) and with a lower iteration number (Iter) compared to the other methods.

Table 1. Performance comparison of different methods in MRI reconstruction.

	Data1			Data2		
Algorithms	Iter	Time(s)	SNR(dB)	Iter	Time(s)	SNR(dB)
SpaSRA	61	73.89	19.3562	28	139.93	28.1334
NMBB	40	38.79	19.6495	27	132.86	27.1879
NMBBA	30	28.59	19.4433	21	98.20	28.5226
NVMBB	22	21.74	19.4899	22	110.91	27.9799
NVMBBA	22	20.31	19.4899	22	107.48	27.9799

In Figure 3, we plot the objective function value versus CPU time. From the figure, we can see that, among these five algorithms, the decay rates of SpaRSA are much slower than those of the others.

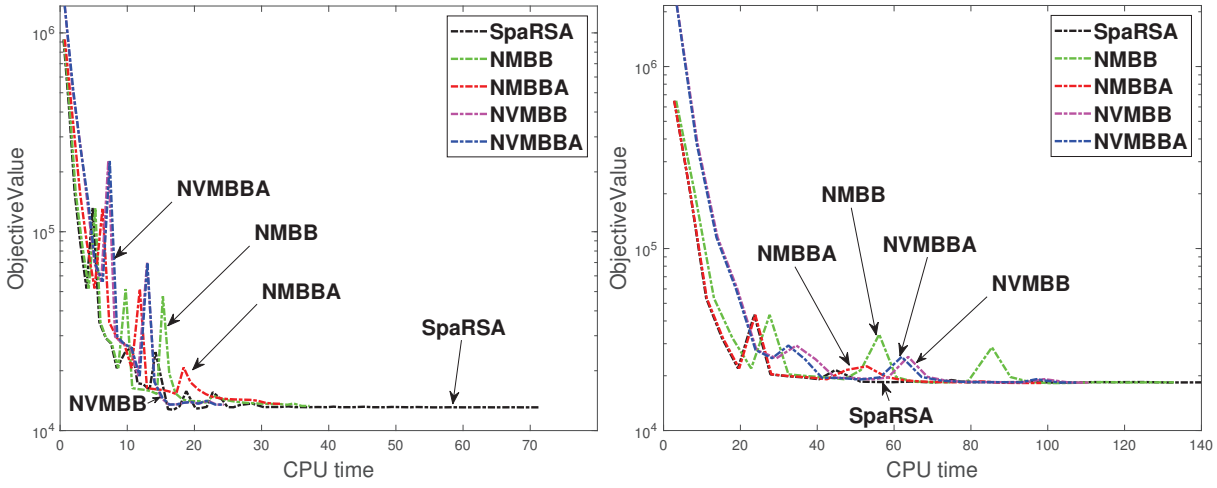


Figure 3. Objective function value versus CPU time for Data1 (a) and Data2 (b).

The reason may be that it has to solve the subproblem for each trial point on each line search, as said in [18]. NMBBA does not use the variable metric technique. This demonstrates that the convex combination of the maximum function value of some previous iteration and the current function value can improve the convergence speed. If we choose the suitable scaling matrix D_k for NVMBB and NVMBBA, they may converge quickly; for example, see their performance Data1. Otherwise, NVMBB and NVMBBA may be a little slower than NMBBA but also faster than SpaRSA and NMBB. So, a suitable combination of the stepsize α_k and the variable metric D_k is important for the practical problems discussed in [31].

The second test problem is TV Poisson noise image deconvolution. When the image is corrupted by Poisson- type noise, the Kullback-Leibler divergence is often used as the data-fitting term. Let b be

the observed blurry and noisy image, then the data discrepancy can be given as

$$f(x) = \langle b, ln\frac{b}{Hx + g\mathbf{1}}\rangle + Hx + g\mathbf{1} - b,$$

where

$$H \in R^{n\times n}$$

is a blur matrix, $\mathbf{1}$ is the vector of all ones, and g is a constant background term. To deal with this ill-posed problem, based on the prior information of the image, TV regularization with a nonnegative constraint is added to control the noise and artifacts, i.e.,

$$h(x) = \rho||x||_{TV} + l_{x\geq 0}(x),$$

where $\rho > 0$ is a regularization parameter. Hence, the Poisson noise image restoration problem is reformulated as a minimization problem

$$\min_{x\in R^n} \phi(x) = \langle b, ln\frac{b}{Hx + g\mathbf{1}}\rangle + Hx + g\mathbf{1} - b + \rho||x||_{TV} + l_{x\geq 0}(x).$$

So, Algorithm 1 can be used to solve the convex composite optimization problem.

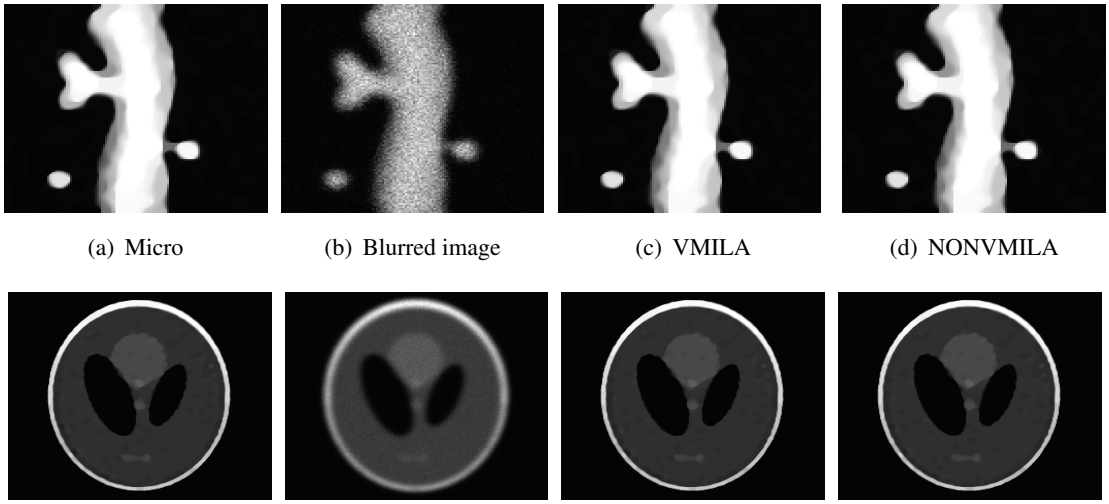
In this test, we compare it with the method named VMILA in [24]. In Algorithm 1, the variable metric technique is the same as VMILA, with $\eta = 1$. This is to say, we use the nonmonotone line search instead of the modified Armijo line search in VMILA to determine the step length λ_k. So, algorithm 1 is named NONVMILA. The two test images are Micro (128×128) and Phantom (256×256). For the micro image, $\rho = 0.09$ and $g = 0.5$. For the Phantom image, $\rho = 0.004$ and $g = 10$. We set $M = 3$ for Micro and $M = 9$ for Phantom, and the other parameters are the same as VMILA. We also used the FISTA [14] method to minimize the total variation of the proximal function. The inner iteration number is set to 1500. We refer the reader to [24] for more details. Both of the compared methods are stopped when the maximum number of iterations is more than 400. The peak-signal noise ratio (PSNR) is used to evaluate the quality of the restored image.

The numerical performance of image restoration is recorded in Table 2. We present the CPU times in seconds and the PSNR values of the two algorithms for various blurry observations.

Table 2. CPU time and PSNR values of two algorithms for Poisson noise image deblurring.

	VMILA		NONVMILA	
Images	Time(s)	PSNR	Time(s)	PSNR
Micro	4.4280	52.7215	3.8330	52.8909
Phantom	10.2243	24.6938	9.7457	25.9434

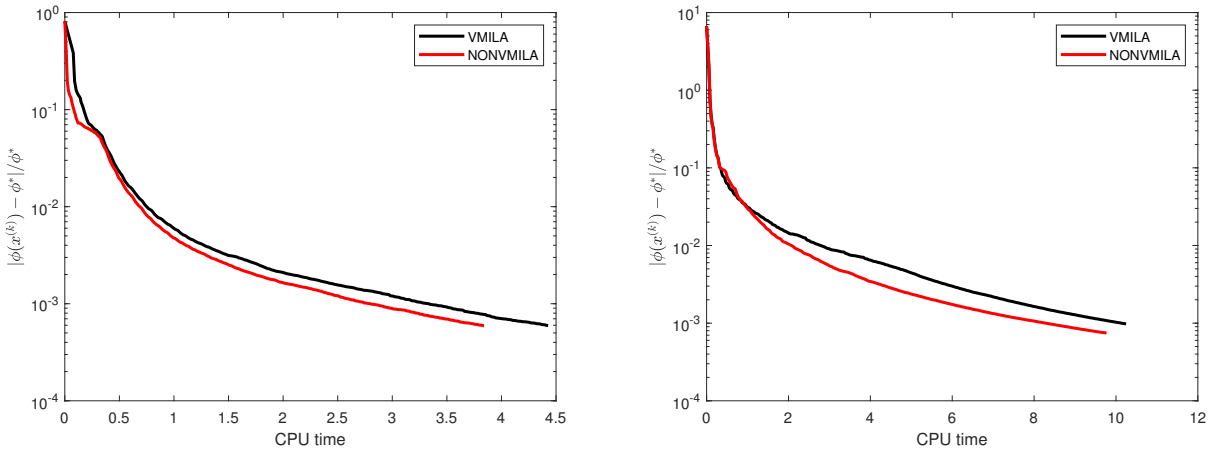
From the table, one can see that the two approaches can produce similar behaviors for the two images. Figure 4 displays the visual results of image deblurring. In terms of visual quality, the difference is not significant. This can also be seen from the PSNR values. Figure 5 reports the relative decrease of the objective function with respect to the optimal value ϕ^* (when $x = x_0$, x_0 is the original image) as a function of the computational time. We observe that NONVMILA requires less time.



(a) Micro (b) Blurred image (c) VMILA (d) NONVMILA

(e) Phantom (f) Blurred image (g) VMILA (h) NONVMILA

Figure 4. Original images ((a), (e)), blurred images ((b), (f)) with Poisson noise and images recovered by NMILA ((c) PSNR=52.7215, (g) PSNR=24.6938) and NONVMILA ((d) PSNR=52.8909, (h) PSNR=25.9434).



(a) (b)

Figure 5. Relative decrease of the objective functions with respect to the computational time for Mirco (a) and Phantom (b).

The last problem is Cauchy noise image deconvolution. When the image is corrupted by the Cauchy-type noise, the log function is used as the data fitting term. Let b be the observed blurred and noisy image, then the data discrepancy is

$$f(x) = \frac{1}{2}\log(\gamma^2 + (Hx - b)^2),$$

where

$$H \in R^{n\times n}$$

is a blur matrix with the 9×9 testing kernel and the variance 1, while has been set equal to 0.02. Based on the prior information of the image, TV regularization with a nonnegative constraint is added to control the noise and artifacts, i.e.,

$$h(x) = \rho\|x\|_{TV} + l_{x\geq 0}(x),$$

where $\rho > 0$ is a regularization parameter. Hence, the Cauchy noise image restoration problem is reformulated as

$$\min_{x\in R^n} \psi(x) = \frac{1}{2}\log(\gamma^2 + (Hx - b)^2) + \rho\|x\|_{TV} + l_{x\geq 0}(x).$$

So, Algorithm 1 can be used to solve this composite optimization problem.

In this test, we also compare it with the VMILA type in [38,39]. In Algorithm 1, the variable metric technique is the same as in VMILA, with $\eta = 0.9$, and is also named NONVMILA. We set $M = 5$ and the other parameters are the same as VMILA. The three test images are the cameraman (256×256), the house (256×256) and the bird (256×256). All methods are stopped when the maximum number of iterations is more than 5000 or

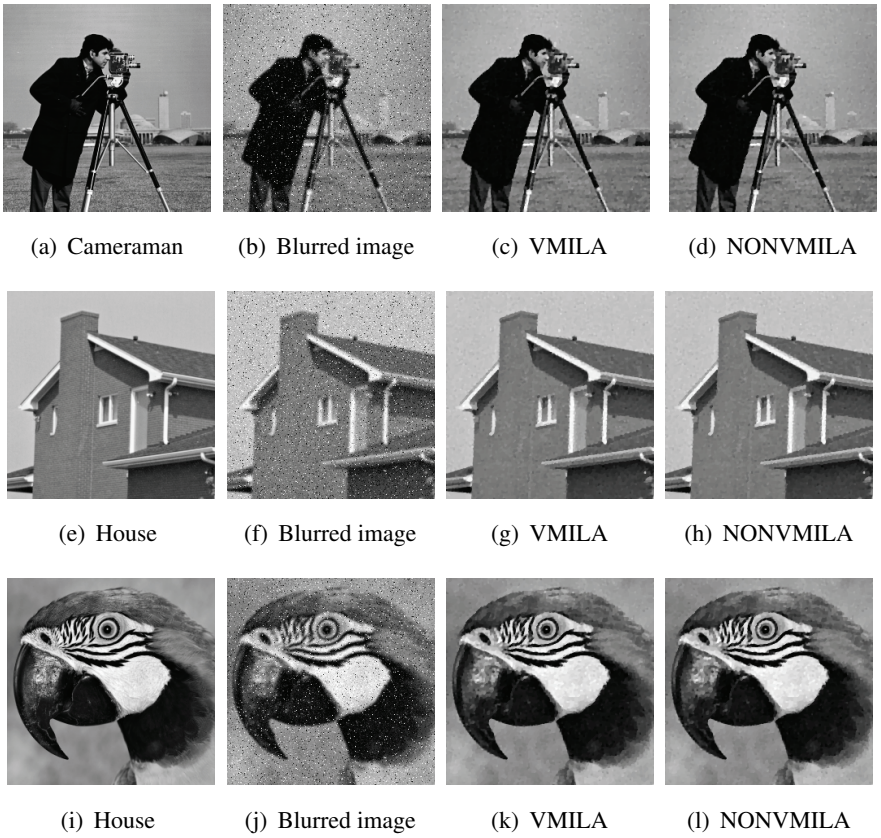
$$|\psi(x_k) - \psi(x_{k-10})| \leq 10^{-6} * \max\{1, |\psi(x_k)|\}.$$

The numerical performance of image restoration is reported in Table 3.

Table 3. CPU time and PSNR values of two algorithms for Cauchy noise image deblurring.

	VMILA		NONVMILA	
Images	Time(s)	PSNR	Time(s)	PSNR
Cameraman	105.2720	26.2798	48.6810	26.2892
House	87.3230	30.4242	47.5650	30.4463
Bird	123.6920	27.1690	55.8790	27.2572

We present the running time in seconds and PSNR values of the two algorithms. From the table, we observe that the two approaches can get similar behaviors for three images in terms of PSNR. We also find that NONVMILA requires less time than VMILA when the stopping criterion is satisfied. Figure 6 displays the visual results of the deblurred image. From Table 3 and Figure 6, it seems that our approach is very efficient for this tested problem.



(a) Cameraman (b) Blurred image (c) VMILA (d) NONVMILA

(e) House (f) Blurred image (g) VMILA (h) NONVMILA

(i) House (j) Blurred image (k) VMILA (l) NONVMILA

Figure 6. Original images ((a), (e), (i)), blurred images ((b), (f), (j)) with Poisson noise and images recovered by NMILA ((c) PSNR=26.2798, (g) PSNR=30.4242, (k) PSNR=27.1690) and NONVMILA ((d) PSNR=26.2892, (h) PSNR=30.4463, (l) PSNR=27.2572).

5. Conclusions

In this paper, we introduce a variable metric algorithm frame for composite optimization problems. With the nonmonotone line search techniques, the convergence of the method is discussed. With the growing use of nonconvex objective functions in applied fields like image processing and machine learning, the need for numerical methods in a fully nonconvex setting has increased significantly. So, future work will focus on a general nonconvex non-Lipschitz function, including theoretical and numerical analysis. Moreover, there is good potential to improve the convergence results using the conditions about the regularity of the cost function and the boundedness of the Hessian matrix [40]. Also, a multi-step inertial forward-backward splitting algorithm will be considered in the future [41, 42]. At last, it is interesting to develop a new scaling matrix and BB stepsize for solving the composite optimization problem.

Author contributions

Xiao Guo: writing-original draft, writing-review and editing, software. Chuanpei Xu: resources, writing-review and editing, supervision. Zhibin Zhu: methodology, writing-review and editing,

supervision. Benxin Zhang: resources, writing-review and editing.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

The authors wish to thank the editor and anonymous referees for their constructive suggestions and comments, which greatly improved the presentation of this paper. This work was supported in part by the National Natural Science Foundation of China under Grants 11901137 and 62171147, in part by the Guangxi Natural Science Foundation under Grant 2024GXNSFAA010512, in part by the Guangxi Key Laboratory of Automatic Detecting Technology and Instruments under Grant YQ22108, and in part by the Innovation Project of GUET Graduate Education under Grant 2022YCXS161. We would like to thank the authors of [24, 36, 37] for making their code and data free for academic use.

Conflict of interest

The authors declare no conflicts of interest.

References

1. E. J. Candes, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory*, **52** (2006), 489–509. https://doi.org/10.1109/TIT.2005.862083
2. E. T. Hale, W. Yin, Y. Zhang, Fixed-point continuation for l_1-minimization: methodology and convergence, *SIAM J. Optim.*, **19** (2008), 1107–1130. https://doi.org/10.1137/070698920
3. Y. Li, C. Li, W. Yang, W. Zhang, A new conjugate gradient method with a restart direction and its application in image restoration, *AIMS Math.*, **8** (2023), 28791–28807. https://doi.org/10.3934/math.20231475
4. A. Chambolle, An algorithm for total variation minimization and applications, *J. Math. Imag. Vis.*, **20** (2004), 89–97. https://doi.org/10.1023/B:JMIV.0000011325.36760.1e
5. L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, *Phys. D*, **60** (1992), 259–268. https://doi.org/10.1016/0167-2789(92)90242-F
6. N. Artsawang, Accelerated preconditioning Krasnosel'skii-Mann method for efficiently solving monotone inclusion problems, *AIMS Math.*, **8** (2023), 28398–28412. https://doi.org/10.3934/math.20231453
7. P. H. Calamai, J. J. Moré, Projeeted gradient methods for linearly constralned problems, *Math. Program.*, **39** (1987), 93–116. https://doi.org/10.1007/BF02592073
8. D. P. Bertsekas, *Nonlinear programming*, Athena scientific, Belmont, 1999.

9. J. Barzilai, J. M. Borwein, Two point step size gradient methods, *IMA J. Numer. Anal.*, **8** (1988), 141–148. https://doi.org/10.1093/imanum/8.1.141
10. E. G. Birgin, J. M. Martínez, M. Raydan, Nonmonotone spectral projected gradient methods on convex sets, *SIAM J. Optim.*, **10** (2000), 1196–1211. https://doi.org/10.1137/S1052623497330963
11. Y. H. Dai, R. Fletcher, Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming, *Numer. Math.*, **100** (2005), 21–47. https://doi.org/10.1007/s00211-004-0569-y
12. B. Zhang, Z. Zhu, S. Li, A modified spectral conjugate gradient projection algorithm for total variation image restoration, *Appl. Math. Lett.*, **27** (2014), 26–35. https://doi.org/10.1016/j.aml.2013.08.006
13. I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Commun. Pure Appl. Math.*, **57** (2004), 1413–1457. https://doi.org/10.1002/cpa.20042
14. A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imag. Sci.*, **2** (2009), 183–202. https://doi.org/10.1137/080716542
15. E. T. Hale, W. Yin, Y. Zhang, Fixed-point continuation applied to compressed sensing: implementation and numerical experiments, *J. Comput. Math.*, **28** (2010), 170–194. https://doi.org/10.4208/jcm.2009.10-m1007
16. S. J. Wright, R. D. Nowak, T. Figueiredo, Sparse reconstruction by separable approximation, *IEEE Trans. Signal Process.*, **57** (2009), 2479–2493. https://doi.org/10.1109/TSP.2009.2016892
17. W. W. Hager, D. T. Phan, H. Zhang, Gradient-based methods for sparse recovery, *SIAM J. Imag. Sci.*, **4** (2011), 146–165. https://doi.org/10.1137/090775063
18. Y. Huang, H. Liu, A Barzilai-Borwein type method for minimizing composite functions, *Numer. Algorithms*, **69** (2015), 819–838. https://doi.org/10.1007/s11075-014-9927-8
19. W. Cheng, Z. Chen, D. Li, Nonmonotone spectral gradient method for sparse recovery, *Inverse Probl. Imag.*, **9** (2015), 815–833. https://doi.org/10.3934/ipi.2015.9.815
20. Y. Xiao, S. Y. Wu, L. Qi, Nonmonotone Barzilai-Borwein gradient algorithm for l_1-regularized nonsmooth minimization in compressive sensing, *J. Sci. Comput.*, **61** (2014), 17–41. https://doi.org/10.1007/s10915-013-9815-8
21. P. Tseng, S. Yun, A coordinate gradient descent method for nonsmooth separable minimization, *Math. Program.*, **117** (2009), 387–423. https://doi.org/10.1007/s10107-007-0170-0
22. E. Chouzenoux, J. C. Pesquet, A. Repetti, Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function, *J. Optim. Theory Appl.*, **162** (2014), 107–132. https://doi.org/10.1007/s10957-013-0465-7
23. P. L. Combettes, B. C. Vû, Variable metric forward-backward splitting with applications to monotone inclusions in duality, *Optimization*, **63** (2014), 1289–1318. https://doi.org/10.1080/02331934.2012.733883
24. S. Bonettini, I. Loris, F. Porta, M. Prato, Variable metric inexact line-search-based methods for nonsmooth optimization, *SIAM J. Optim.*, **26** (2016), 891–921. https://doi.org/10.1137/15M1019325

25. S. Rebegoldi, L. Calatroni, Scaled, inexact and adaptive generalized FISTA for strongly convex optimization, *SIAM J. Optim.*, **32** (2022), 2428–2459. https://doi.org/10.1137/21M1391699

26. S.Bonettini, M. Prato, S. Rebegoldi, A new proximal heavy ball inexact line-search algorithm, *Comput. Optim. Appl.*, 2024. https://doi.org/10.1007/s10589-024-00565-9

27. , S. Bonettini, P. Ochs, M. Prato, S. Rebegoldi, An abstract convergence framework with application to inertial inexact forward-backward methods, *Comput. Optim. Appl.*, **84** (2023), 319–362. https://doi.org/10.1007/s10589-022-00441-4

28. H. Liu, T. Wang, Z. Liu, A nonmonotone accelerated proximal gradient method with variable stepsize strategy for nonsmooth and nonconvex minimization problems, *J. Glob. Optim.*, 2024. https://doi.org/10.1007/s10898-024-01366-4

29. Y. Chen, W. W. Hager, M. Yashtini, X. Ye, H. Zhang, Bregman operator splitting with variable stepsize for total variation image reconstruction, *Comput. Optim. Appl.*, **54** (2013), 317–342. https://doi.org/10.1007/s10589-012-9519-2

30. Y. Chen, W. W. Hager, F. Huang, D. T. Phan, X. Ye, W. Yin, Fast algorithms for image reconstruction with application to partially parallel MR imaging, *SIAM J. Imag. Sci.*, **5** (2012), 90–118. https://doi.org/10.1137/100792688

31. S. Bonettini, M. Prato, New convergence results for the scaled gradient projection method, *Inverse Probl.*, **31** (2015), 095008. https://doi.org/10.1088/0266-5611/31/9/095008

32. F. Porta, M. Prato, L. Zanni, A new steplength selection for scaled gradient methods with application to image deblurring, *J. Sci. Comput.*, **65** (2015), 895–919. https://doi.org/10.1007/s10915-015-9991-9

33. L. Grippo, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's method, *SIAM J. Numer. Anal.*, **23** (1986), 707–716. https://doi.org/10.1137/0723046

34. K. Amini, M. Ahookhosh, H. Nosratipour, An inexact line search approach using modified nonmonotone strategy for unconstrained optimization, *Numer. Algorithms*, **66** (2014), 49–78. https://doi.org/10.1007/s11075-013-9723-x

35. R. T. Rockafellar, *Convex analysis*, Princeton University Press, 2015.

36. Y. Ouyang, Y. Chen, G. Lan, J. E. Pasiliao, An accelerated linearized alternating direction method of multipliers, *SIAM J. Imag. Sci.*, **8** (2015), 644–681. https://doi.org/10.1137/14095697X

37. X. Ye, Y. Chen, F. Huang, Computational acceleration for MR image reconstruction in partially parallel imaging, *IEEE Trans. Med. Imag.*, **30** (2011), 1055–1063. https://doi.org/10.1109/TMI.2010.2073717

38. S. Bonettini, I. Loris, F. Porta, M. Prato, S. Rebegoldi, On the convergence of a linesearch based proximal-gradient method for nonconvex optimization, *Inverse Probl.*, **33** (2017), 055005. https://doi.org/10.1088/1361-6420/aa5bfd

39. S. Bonettini, M. Prato, S. Rebegoldi, Convergence of inexact forward-backward algorithms using the forward-backward envelope, *SIAM J. Optim.*, **30** (2020), 3069–3097. https://doi.org/10.1137/19M1254155

40. M. K. Riahi, I. A. Qattan, On the convergence rate of Fletcher-Reeves nonlinear conjugate gradient methods satisfying strong Wolfe conditions: application to parameter identification in problems governed by general dynamics, *Math. Methods Appl. Sci.*, **45** (2022), 3644–3664. https://doi.org/10.1002/mma.8009

41. X. Li, Q. L. Dong, A. Gibali, PMiCA-Parallel multi-step inertial contracting algorithm for solving common variational inclusions, *J. Nonlinear Funct. Anal.*, **2022** (2022), 7. https://doi.org//10.23952/jnfa.2022.7

42. L. O. Jolaoso, Y. Shehu, J. Yao, R. Xu, Double inertial parameters forward-backward splitting method: applications to compressed sensing, image processing, and SCAD penalty problems, *J. Nonlinear Var. Anal.*, **7** (2023), 627–646. https://doi.org/10.23952/jnva.7.2023.4.10

AIMS Press