Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Computational and Applied Mathematics 182 (2005) 51–66
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Convergence properties of nonmonotone spectral projected gradient methods☆

Changyu Wang[a,*], Qian Liu[a], Xinmin Yang[b]

[a] *Institute of Operations Research, Qufu Normal University, Qufu, Shandong 273165, China*
[b] *Department of Mathematics, Chongqing Normal University, Chongqing 400047, China*

Received 10 November 2003

Abstract

In a recent paper, a nonmonotone spectral projected gradient (SPG) method was introduced by Birgin et al. for the minimization of differentiable functions on closed convex sets and extensive presented results showed that this method was very efficient. In this paper, we give a more comprehensive theoretical analysis of the SPG method. In doing so, we remove various boundedness conditions that are assumed in existing results, such as boundedness from below of f, boundedness of x_k or existence of accumulation point of $\{x_k\}$. If $\nabla f(\cdot)$ is uniformly continuous, we establish the convergence theory of this method and prove that the SPG method forces the sequence of projected gradients to zero. Moreover, we show under appropriate conditions that the SPG method has some encouraging convergence properties, such as the global convergence of the sequence of iterates generated by this method and the finite termination, etc. Therefore, these results show that the SPG method is attractive in theory.

Keywords: Nonmonotone linear search; Convergence; Finite termination

1. Introduction

The problem of minimizing a continuously differentiable mapping $f : R^n \to R$ over a nonempty closed convex set $\Omega \subseteq R^n$,

$$\min\{f(x) : x \in \Omega\} \tag{1}$$

☆ This work is supported by the Natural Science Foundation of China (Grant no. 10171055).
* Corresponding author.
E-mail address: wcy0537@eyou.com (C. Wang).

0377-0427/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2004.10.018

has received considerable attention. Over the last few decades, there have been many different methods to solve problem (1). The simplest of these methods is the gradient projection method which was originally proposed by Goldstein [12] and Levitin and Polyak [17] and extended by Calamai and Moré [8]. This method possesses some advantages. Firstly, it is easy to implement (especially, for the optimization problem with simple bounds), uses little storage and readily exploits any sparsity or separable structure in $\nabla f(x)$ or Ω. Secondly, it is able to drop and add many constraints from the active set at each iteration. Hence, it has been developed for solving various cases of problem (1).

Given an inner product norm $\|\cdot\|$, the projection onto a nonempty closed convex set Ω is the mapping $P_\Omega : R^n \to \Omega$ defined by

$$P_\Omega(x) := \arg\min\{\|z - x\| : z \in \Omega\}.$$

The gradient projection algorithm is defined by

$$x_{k+1} = x_k(\lambda_k) = P_\Omega(x_k - \lambda_k \nabla f(x_k)),$$

where $\lambda_k > 0$ is the stepsize and $\nabla f(x)$ is the gradient of f.

Some convergence results of the gradient projection method were obtained (see, for example, [2,8,9,11–13,17,18,23–26] and references therein). In the algorithms of these papers, we noticed that the sequence $\{f(x_k)\}$ was monotonically decreasing. But for some functions the performance of methods with monotone strategies is poor. The numerical results imply that the methods with proper nonmonotone strategies are more efficient than the ones with monotone strategies (see [6,10,15,16,20]). Particularly, in some cases, the methods with nonmonotone line search can overcome the Maratos effect [6].

The spectral choice of steplength introduced by Barzilai and Borwein [1] is a technique for the choice of steplength. Numerical results show that this technique is very efficient to solve large-scale unconstrained optimization (see [3,20–22]).

Recently, combining the spectral choice of steplength with nonmonotone line search techniques, Birgin et al. [4] established nonmonotone spectral projected gradient (SPG) methods. In the paper [4], a mass of numerical experiments showed that the spectral choice of the steplength represented considerable progress in relation to constant choices and that the nonmonotone framework was useful. However, the proof of their convergence theorems (see [4]) contain minor errors. The reason is that in the proof of case 2 (see [4, Theorems 2.3 and 2.4, p. 1200, p. 1202]) they mistook the sequence $\{x_k\}$ for a subsequence $\{x_k\}_K$, which converged to an accumulation point $\bar{x}$. So the convergence results in the paper [4] could not be obtained. In fact, if we add some appropriate conditions, the convergence theorem will be proved.

In this paper, the authors study the convergence properties of SPG methods. We remove various boundedness conditions that are assumed in existing results, such as boundedness from below of f, boundedness of x_k or existence of accumulation point of $\{x_k\}$. If $\nabla f(\cdot)$ is uniformly continuous, we establish the convergence theory of this method and prove that the SPG method forces the sequence of projected gradients to zero. Moreover, we show under appropriate conditions that the SPG method has some encouraging convergence properties, such as the global convergence of the sequence of iterates generated by this method and the finite termination, which improve and generalize the corresponding results in the papers [7,19].

The paper is organized as follows. In the next section we introduce some concepts and lemmas which are used in the remainder of the paper. In Section 3, we present the nonmonotone spectral projected gradient algorithms. We then prove the convergence theorems in Section 4. In Section 5, we study the global convergence of the sequence $\{x_k\}$ of iterates generated by this method. Finally in Section 6, the finite convergence results under milder conditions are obtained.

2. Definitions and lemmas

We now review some definitions and lemmas that are used in subsequent sections.

We say that a point $x \in \Omega$ is a stationary point of problem (1) if it satisfies condition

$$\langle \nabla f(x), y - x \rangle \geqslant 0 \quad \forall y \in \Omega.$$

The set that consists of all the stationary points and the set that consists of all the global optimal solutions of problem (1), respectively, are denoted by Ω^+ and Ω^*. If the mapping $f(\cdot)$ is pseudo-convex on Ω, then $\Omega^* = \Omega^+$.

For a nonempty subset S of R^n, its polar cone is defined as

$$S^\circ = \{y \in R^n | \langle y, x \rangle \leqslant 0 \quad \forall x \in S\}.$$

The tangent cone of Ω at $x \in \Omega$ is given by

$$T_\Omega(x) = \{d \in R^n | \exists \tau_k \downarrow 0,\ d_k \to d \quad \forall k, x + \tau_k d_k \in \Omega\}.$$

The normal cone of Ω at x is defined as

$$N_\Omega(x) = T_\Omega(x)^\circ.$$

We call a mapping $\nabla_\Omega f : R^n \to R^n$ the projected gradient of $f(\cdot)$ with respect to the set Ω if

$$\nabla_\Omega f(x) = P_{T_{\Omega(x)}}(-\nabla f(x)) \quad \forall x \in \Omega.$$

By this definition, $x \in \Omega^+$ if and only if $\nabla_\Omega f(x) = 0$ or $-\nabla f(x) \in N_\Omega(x)$.

A mapping $\nabla f(\cdot)$ is Lipschitz continuous on Ω, if there exists a constant $L > 0$ such that for every x, $y \in \Omega$,

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L \|x - y\|.$$

Projection has been extensively studied and we here briefly recall some of its properties for our discussion.

Lemma 2.1 (*Calamai and Moré [8]*). *Let P_Ω be the projection onto Ω. Consider $x \in \Omega$, and define $x(\alpha) := P_\Omega(x - \alpha \nabla f(x))$, then*

(1) $\langle x(\alpha) - x + \alpha \nabla f(x), y - x(\alpha) \rangle \geqslant 0$, *for all* $y \in \Omega$ *and* $\alpha > 0$
(2) *for all* $\alpha > 0$,

$$\langle \nabla f(x), x - x(\alpha) \rangle \geqslant \frac{|x(\alpha) - x\|^2}{\alpha}.$$

Lemma 2.2 (*Calamai and Moré [8]*). *Let $\nabla_\Omega f(x)$ be the projected gradient of f at $x \in \Omega$. Then*

(1) $\min\{\langle \nabla f(x), v \rangle : \|v\| \leqslant 1\} = -\|\nabla_\Omega f(x)\|$;

(2) $\nabla_\Omega f(\cdot)$ *is lower semicontinuous on* Ω, *that is, if* $\lim_{k\to\infty} x_k = x$, *then*

$$\|\nabla_\Omega f(x)\| \leqslant \liminf_{k\to\infty} \|\nabla_\Omega f(x_k)\|.$$

Lemma 2.3 (*Wang and Xiu [25]*). *For each iterative point* $x_k \in \Omega,\ x_k = P_\Omega(x_{k-1} - \alpha_{k-1}\nabla f(x_{k-1}))$, *we have for any* $x \in \Omega$,

$$\frac{\langle \nabla f(x_k), x_k - x\rangle}{\|x_k - x\|} \leqslant \frac{\|x_k - x_{k-1}\|}{\alpha_{k-1}} + \|\nabla f(x_k) - \nabla f(x_{k-1})\|.$$

3. Convergence properties

In the paper [4], nonmonotone spectral projected gradient algorithms [4, Algorithms 2.1 and 2.2] to solve the constrained minimization problem (1) were introduced. Now we state their convergence theorems as follows.

Using the first part of the proof of the theorem in [15, p. 709], we can obtain the following lemma.

Lemma 3.1. *Let* $l(k)$ *be an integer such that* $k - \min\{k, M-1\} \leqslant l(k) \leqslant k$ *and*

$$f(x_{l(k)}) = \max_{0 \leqslant j \leqslant \min\{k, M-1\}} f(x_{k-j}).$$

If $\{x_k\}$ *is the sequence produced by Algorithm* 2.1 (*or Algorithm* 2.2), *then the sequence* $\{f(x_{l(k)})\}$ *is monotonically nonincreasing*.

Let $f_{\max} = \lim_{k\to\infty} f(x_{l(k)}),\ \ N = \{1, 2, \ldots\}$, we obtain the following result.

Theorem 3.1. *Let* $\{x_k\}$ *be the sequence generated by the Algorithm* 2.1, *if* $\nabla f(x)$ *is uniformly continuous on an open convex set containing* $\{x_k\}$, *then* $\lim_{k\to+\infty} f(x_k) = -\infty$ *or*

$$\lim_{k\to+\infty} \frac{\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle}{\lambda_k} = 0.$$

Proof. Note that if $\lim_{k\to+\infty} f(x_k) = -\infty$, then the result is obtained. Otherwise, we obtain that $\limsup_{k\to+\infty} f(x_k) > -\infty$. Thus Algorithm 2.1 implies that $f_{\max} > -\infty$.

Setting $K_i = \{l(k') + i - 1 \mid k' \in N\}$, where $i = 0, 1, 2, \ldots, M-1$. It follows from the definition of $f(x_{l(k')})$ that

$$\begin{aligned} l(k'+1) - l(k') &\leqslant l(k'+1) - l(l(k'+1) - 1), \\ &\leqslant l(k'+1) - \{l(k'+1) - 1 - \min\{l(k'+1) - 1, M-1\}\}, \\ &= 1 + \min\{l(k'+1) - 1, M-1\} \leqslant M. \end{aligned}$$

It is clear that $\bigcup_{i=0}^{M-1} - K_i = N$. It suffices to show that

$$\lim_{k\in \bigcup_{i=0}^{M-1} K_i, k\to\infty} \frac{\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle}{\lambda_k} = 0.$$

Let us first show that

$$\lim_{k\in K_0,k\to\infty}\frac{\langle\nabla f(x_k),x_k-x_k(\lambda_k)\rangle}{\lambda_k}=0.$$

Assume that there is an $\varepsilon_0>0$ and an infinite subsequence $\bar{K}_0\subset K_0$ such that

$$\frac{\langle\nabla f(x_k),x_k-x_k(\lambda_k)\rangle}{\lambda_k}\geqslant\varepsilon_0\quad\forall k\in\bar{K}_0. \tag{2}$$

We will prove that this assumption leads to a contradiction. First note that if $k\in\bar{K}_0$ then there exists k' such that $k=l(k')-1$, i.e., $k+1=l(k')$. By Algorithm 2.1 we have

$$f(x_{l(k')})\leqslant f(x_{l(k)})+\gamma\langle x_k(\lambda_k)-x_k\ \nabla f(x_k)\rangle\quad\forall k\in\bar{K}_0 \tag{3}$$

which, together with Lemma 2.1(2), shows that

$$0\leqslant\gamma\langle x_k-x_k(\lambda_k),\ \nabla f(x_k)\rangle\leqslant f(x_{l(k)})-f(x_{l(k')}).$$

Furthermore, taking limit in the above inequality as $k\in\bar{K}_0,\ k\to\infty$, we have

$$\lim_{k\in K_0,k\to\infty}\langle\nabla f(x_k),x_k-x_k(\lambda_k)\rangle=0. \tag{4}$$

From (2) and (4), we derive $\lim_{k\in\bar{K}_0,k\to\infty}\lambda_k=0$. So, by Algorithm 2.1, for all $k\in\bar{K}_0$ sufficiently large there exists $\sigma_1\leqslant\rho_k\leqslant\sigma_2,\ \psi_k=\lambda_k/\rho_k$ satisfying

$$f(x_k(\psi_k))>\max_{0\leqslant j\leqslant\min\{k,M-1\}}f(x_{k-j})+\gamma\psi_k\langle d_k,\ \nabla f(x_k)\rangle.$$

Hence,

$$\begin{aligned}f(x_k)-f(x_k(\psi_k))&\leqslant\max_{0\leqslant j\leqslant\min\{k,M-1\}}f(x_{k-j})-f(x_k(\psi_k)),\\&<\gamma\langle x_k-x_k(\psi_k),\ \nabla f(x_k)\rangle.\end{aligned} \tag{5}$$

Condition (5) shows that if

$$\rho_k(\lambda)=\frac{f(x_k)-f(x_k(\lambda))}{\langle x_k-x_k(\lambda),\nabla f(x_k)\rangle},$$

then

$$\begin{aligned}\rho_k(\psi_k)&=\frac{f(x_k)-f(x_k(\psi_k))}{\langle x_k-x_k(\psi_k),\nabla f(x_k)\rangle},\\&<\frac{\gamma\langle x_k-x_k(\psi_k),\nabla f(x_k)\rangle}{\langle x_k-x_k(\psi_k),\nabla f(x_k)\rangle}=\gamma.\end{aligned} \tag{6}$$

On the other hand, since $\langle\nabla f(x),x-x(\alpha)\rangle/\alpha$ is nonincreasing on $\alpha>0$ [26], Lemma 2.1(2) and (2) imply that if $k\in\bar{K}_0$, then

$$\begin{aligned}\langle\nabla f(x_k),x_k-x_k(\lambda_k)\rangle^2&\geqslant\lambda_k^2\frac{\langle\nabla f(x_k),x_k-x_k(\lambda_k)\rangle}{\lambda_k}\frac{\langle\nabla f(x_k),x_k-x_k(\psi_k)\rangle}{\psi_k}\\&\geqslant\varepsilon_0\rho_k^2\psi_k\langle\nabla f(x_k),x_k-x_k(\psi_k)\rangle\\&\geqslant\varepsilon_0\rho_k^2\|x_k-x_k(\psi_k)\|^2.\end{aligned}$$

Hence,

$$\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle \geqslant \sqrt{\varepsilon_0}\rho_k \|x_k - x_k(\psi_k)\|. \tag{7}$$

So the uniform continuity of $\nabla f(x)$ and (7) show that

$$\begin{aligned} |\rho_k(\psi_k) - 1| &= \left| \frac{f(x_k) - f(x_k(\psi_k)) - \langle x_k - x_k(\psi_k), \nabla f(x_k)\rangle}{\langle x_k - x_k(\psi_k), \nabla f(x_k)\rangle} \right| \\ &\leqslant \frac{o(\|x_k - x_k(\psi_k)\|)}{\sqrt{\varepsilon_0}\rho_k \|x_k - x_k(\psi_k)\|} \\ &\leqslant \frac{o(\|x_k - x_k(\psi_k)\|)}{\sqrt{\varepsilon_0}\sigma_1 \|x_k - x_k(\psi_k)\|}. \end{aligned}$$

This establishes that $\rho_k(\psi_k) > \gamma$ for all $k \in \bar{K}_0$ sufficiently large, which is the desired contradiction because (6) guarantees $\rho_k(\psi_k) < \gamma$. Hence,

$$\lim_{k \in K_0, k \to \infty} \frac{\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle}{\lambda_k} = 0.$$

Under the assumption that

$$\lim_{k \in K_{i-1}, k \to \infty} \frac{\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle}{\lambda_k} = 0, \tag{8}$$

we now establish

$$\lim_{k \in K_i, k \to \infty} \frac{\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle}{\lambda_k} = 0.$$

By using Lemma 2.1(2), we have

$$\begin{aligned} \frac{\|x_{k-1} - x_{k-1}(\lambda_{k-1})\|^2}{\alpha_{\max}^2} &\leqslant \frac{\|x_{k-1} - x_{k-1}(\lambda_{k-1})\|^2}{\lambda_{k-1}^2} \\ &\leqslant \frac{\langle \nabla f(x_{k-1}), x_{k-1} - x_{k-1}(\lambda_{k-1})\rangle}{\lambda_{k-1}}. \end{aligned} \tag{9}$$

Since $k \in K_i$ implies $k-1 \in K_{i-1}$, condition (8) and (9) show that

$$\lim_{k \in K_i, k \to \infty} \frac{\|x_{k-1} - x_{k-1}(\lambda_{k-1})\|}{\lambda_{k-1}} = 0, \tag{10}$$

$$\lim_{k \in K_i, k \to \infty} \|x_{k-1} - x_{k-1}(\lambda_{k-1})\| = 0. \tag{11}$$

Thus, by the uniform continuity of $\nabla f(x)$ and (11), we have

$$\lim_{k \in K_i, k \to \infty} \|\nabla f(x_{k-1}) - \nabla f(x_{k-1}(\lambda_{k-1}))\| = 0. \tag{12}$$

Moreover, Lemmas 2.1(2) and 2.3 show that

$$\begin{aligned}\frac{\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle}{\lambda_k} &= \frac{\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle}{\|x_k - x_k(\lambda_k)\|}\frac{\|x_k - x_k(\lambda_k)\|}{\lambda_k} \\ &\leqslant \frac{\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle^2}{\|x_k - x_k(\lambda_k)\|^2} \\ &\leqslant \left(\frac{\|x_{k-1} - x_k(\lambda_{k-1})\|}{\lambda_{k-1}} + \|\nabla f(x_{k-1}) - \nabla f(x_{k-1}(\lambda_{k-1}))\|\right)^2. \end{aligned} \tag{13}$$

Condition (13), together with (10) and (12), implies that

$$\lim_{k\in K_i, k\to\infty} \frac{\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle}{\lambda_k} = 0.$$

Thus, we obtain

$$\lim_{k\to\infty} \frac{\langle \nabla f(x_k), x_k - x_k(\lambda_k)\rangle}{\lambda_k} = 0.$$

This completes the proof of the theorem. □

By using Lemma 2.1(2) and Theorem 3.1, we have the following corollary.

Corollary 3.1. *Let* $\{x_k\}$ *be the sequence generated by the Algorithm* 2.1, *if* $\nabla f(x)$ *is uniformly continuous on an open convex set containing* $\{x_k\}$ *and* $f_{\max} > -\infty$, *then*

$$\lim_{k\to+\infty} \frac{\|x_k - x_k(\lambda_k)\|}{\lambda_k} = 0.$$

Corollary 3.2. *Let* $\{x_k\}$ *be the sequence generated by the Algorithm* 2.1, *if* $\nabla f(x)$ *is uniformly continuous on an open convex set containing* $\{x_k\}$ *and* $f_{\max} > -\infty$, *then*

$$\lim_{k\to+\infty} \|\nabla_\Omega f(x_k)\| = 0.$$

Proof. By using the definition of $T_\Omega(x_k)$ and Lemma 2.3, we have for any $d \in T_\Omega(x_k)$ and $\|d\| \leqslant 1$,

$$-\langle \nabla f(x_k), d\rangle \leqslant \frac{x_k - x_{k-1}}{\lambda_{k-1}} + \|\nabla f(x_k) - \nabla f(x_{k-1})\|. \tag{14}$$

Lemma 2.2(1), (14) implies that

$$\begin{aligned}\|\nabla_\Omega f(x_k)\| &= \max\{-\langle \nabla f(x_k), d\rangle | d \in T_\Omega(x_k), \|d\| \leqslant 1\} \\ &\leqslant \frac{x_k - x_{k-1}}{\lambda_{k-1}} + \|\nabla f(x_k) - \nabla f(x_{k-1})\|.\end{aligned}$$

Therefore, by Lemma 3.1, the uniform continuity of $\nabla f(\cdot)$ and the fact that λ_k is bounded, we have

$$\lim_{k\to+\infty} \|\nabla_\Omega f(x_k)\| = 0.$$

Corollary 3.3. *Let* $\{x_k\}$ *be the sequence generated by the Algorithm* 2.1, $\nabla f(x)$ *is uniformly continuous on an open convex set containing* $\{x_k\}$, *if* x^+ *is a cluster point of* $\{x_k\}$, *then* $x^+ \in \Omega^+$.

Proof. By using Lemmas 2.4 and 2.2(2), we can prove $\nabla_\Omega f(x^+) = 0$, that is, $x^+ \in \Omega^+$. □

Remark. Corollary 3.3 is the correct expression of Theorem 2.3 of the paper [4].

Theorem 3.2. *Let* $\{x_k\}$ *be the sequence generated by the Algorithm* 2.2, *if* $\nabla f(x)$ *is uniformly continuous on an open convex set containing* $\{x_k\}$, *then* $\lim_{k\to+\infty} f(x_k) = -\infty$ *or*

$$\lim_{k\to+\infty} \langle \nabla f(x_k), d_k\rangle = 0.$$

Proof. If $\lim_{k\to+\infty} f(x_k) = -\infty$, then the result is obtained. Otherwise, we obtain that $\limsup_{k\to+\infty} f(x_k) > -\infty$. Thus Algorithm 2.2 implies that $f_{\max} > -\infty$. Setting $K_i = \{l(k') + i - 1 | k' \in N\}$, where $i = 0, 1, 2, \ldots, M-1$. The proof proceeds as in Theorem 3.1. Thus it suffices to show that

$$\lim_{k\in \bigcup_{i=0}^{M-1} K_i, k\to\infty} \langle \nabla f(x_k), d_k\rangle = 0.$$

Let us first show that

$$\lim_{k\to K_0, k\to+\infty} \langle \nabla f(x_k), d_k\rangle = 0.$$

Suppose, on the contrary, that there is an $\varepsilon_0 > 0$ and an infinite subsequence $\bar{K}_0 \subset K_0$ such that

$$\langle \nabla f(x_k), d_k\rangle < -\varepsilon_0 \quad \forall k \in \bar{K}_0. \tag{15}$$

First note that if $k \in \bar{K}_0$ then there exists k' such that $k = l(k') - 1$, i.e. $k + 1 = l(k')$ and by Algorithm 2.2 we have

$$f(x_{l(k')}) \leqslant f(x_{l(k)}) + \gamma\lambda_k \langle \nabla f(x_k), d_k\rangle \quad \forall k \in K_0. \tag{16}$$

By (15) and (16), we have that

$$0 < \gamma\lambda_k\varepsilon_0 \leqslant -\gamma\lambda_k \langle \nabla f(x_k), d_k\rangle \leqslant f(x_{l(k)}) - f(x_{l(k')}).$$

Taking limit in the above inequality as $k \in \bar{K}_0$, $k \to +\infty$, Lemma 2.1 and $f_{\max} > -\infty$ show that

$$\lim_{k\in\bar{K}_0, k\to\infty} \lambda_k = 0, \qquad \lim_{k\in\bar{K}_0, k\to\infty} \gamma\lambda_k \frac{\|d_k\|^2}{\alpha_{\max}} \leqslant \lim_{k\in\bar{K}_0, k\to\infty} \gamma\lambda_k \langle \nabla f(x_k), d_k\rangle = 0.$$

Hence,

$$\lim_{k\in\bar{K}_0, k\to\infty} \|\lambda_k d_k\| = 0. \tag{17}$$

By Algorithm 2.2 and $\lambda_k \to 0 (k \in \bar{K}_0, k \to \infty)$, for all $k \in \bar{K}_0$ sufficiently large there exists $\sigma_1 \leqslant \rho_k \leqslant \sigma_2$, $\psi_k = \lambda_k/\rho_k$ satisfying

$$f(x_k + \psi_k d_k) > \max_{0\leqslant j\leqslant \min\{k, M-1\}} f(x_{k-j}) - \gamma\psi_k \langle \nabla f(x_k), d_k\rangle.$$

Therefore, we obtain

$$f(x_k) - f(x_k + \psi_k d_k) \leqslant \max_{0 \leqslant j \leqslant \min\{k, M-1\}} f(x_{k-j}) - f(x_k + \psi_k d_k), \\ < -\gamma\psi_k \langle \nabla f(x_k), d_k \rangle.$$

By using mean value theorem,

$$\langle \nabla f(x_k) - \nabla f(x_k + \theta_k \psi_k d_k), d_k \rangle < (1-\gamma)\langle \nabla f(x_k), d_k \rangle, \tag{18}$$

where $\theta_k \in (0, 1)$. By Lemma 2.1(2), the assumption shows that

$$-\langle \nabla f(x_k), d_k \rangle \geqslant \sqrt{\varepsilon_0} \frac{\|d_k\|}{\sqrt{\alpha_k}} \geqslant \sqrt{\frac{\varepsilon_0}{\alpha_{\max}}} \|d_k\|. \tag{19}$$

Thus, by (18) and (19), we have

$$(1-\gamma) < \frac{\|\nabla f(x_k) - \nabla f(x_k + \theta_k \psi_k d_k)\| \|d_k\|}{-\langle \nabla f(x_k), d_k \rangle}, \\ \leqslant \sqrt{\frac{\alpha_{\max}}{\varepsilon_0}} \|\nabla f(x_k) - \nabla f(x_k + \theta_k \psi_k d_k)\|. \tag{20}$$

Taking limit in (20) as $k \in K_0,\ k \to \infty$, the uniformly continuous of $\nabla f(x)$ and condition (17) show that $\gamma \geqslant 1$. An obvious contradiction now occurs. Hence,

$$\lim_{k \in K_0, k \to +\infty} \langle \nabla f(x_k), d_k \rangle = 0.$$

For the reminder of the proof, the argument used in the proof of Theorem 3.1 yield

$$\lim_{k \in K_i, k \to +\infty} \langle \nabla f(x_k), d_k \rangle = 0.$$

Thus, we obtain

$$\lim_{k \to \infty} \langle \nabla f(x_k), d_k \rangle = 0. \quad \square$$

Very similar to the proof of corollaries of Theorem 3.1, from Theorem 3.2 we obtain the following corollaries:

Corollary 3.4. *Let $\{x_k\}$ be the sequence generated by the Algorithm 2.2, if $\nabla f(x)$ is uniformly continuous on an open convex set containing $\{x_k\}$ and $f_{\max} > -\infty$, then*

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

Corollary 3.5. *Let $\{x_k\}$ be the sequence generated by the Algorithm 2.2, if $\nabla f(x)$ is uniformly continuous on an open convex set containing $\{x_k\}$ and $f_{\max} > -\infty$, then*

$$\lim_{k \to +\infty} \|\nabla_\Omega f(x_k)\| = 0.$$

Corollary 3.6. *Let $\{x_k\}$ be the sequence generated by the Algorithm 2.2, $\nabla f(x)$ is uniformly continuous on an open convex set containing $\{x_k\}$, if x^+ is a cluster point of $\{x_k\}$, then $x^+ \in \Omega^+$.*

4. Global convergence

An algorithm is called globally convergent, if there exists $\bar{x} \in \Omega^+$ such that the sequence $\{x_k\}$ produced by the algorithm satisfies $\lim_{k\to+\infty} x_k = \bar{x}$.

Under general conditions, various existing algorithms do not possess global convergence. Gonzaga [14] gave a counter example, in which if f is convex and strictly convex at all nonoptimal points, the steepest descent method with exact stepsize rule for solving unconstrained problem generates four distinct accumulation points. However, it is available to explore what are conditions for convergence of the method. Especially, when we estimate convergence rate of an algorithm, possessing global convergence is one of the preconditions. More recently, the authors [25] studied the global convergence of the gradient projected method with Armijo stepsize rule. In this paper, we prove that if f is generalized convex and $\nabla f(\cdot)$ is Lipschitz continuous on Ω, then the sequence $\{x_k\}$ produced by the nonmonotone spectral projected gradient method converges to a solution of problem (1). To prove this result, we first give two important lemmas.

Lemma 4.1. *Let $\{x_k\}$ be an infinite sequence generated by the Algorithm* 2.1. *If $\nabla f(x)$ is Lipschitz continuous on Ω and $\alpha_{\max} \leqslant 1/2L$, then we have for all $k \in N$ and any $x \in \Omega$,*

$$\|x_{k+1} - x\|^2 \leqslant \|x_k - x\|^2 + 2\alpha_{\max}\{f(x_k) - f(x_{k+1})\} + 2\lambda_k\langle\nabla f(x_k), x - x_k\rangle.$$

Proof. By using Lemma 2.1, $\alpha_{\max} \leqslant 1/2L$ and mean value theorem, we have for all $k \in N$ and any $x \in \Omega$,

$$\begin{aligned}
\|x_{k+1} - x\|^2 &= \|x_k - x\|^2 + 2\langle x_{k+1} - x_k, x_k - x\rangle + \|x_{k+1} - x_k\|^2 \\
&= \|x_k - x\|^2 + 2\langle x_{k+1} - x_k, x_{k+1} - x\rangle - \|x_{k+1} - x_k\|^2 \\
&\leqslant \|x_k - x\|^2 + 2\lambda_k\langle\nabla f(x_k), x - x_{k+1}\rangle - \|x_{k+1} - x_k\|^2 \\
&\leqslant \|x_k - x\|^2 + 2\alpha_{\max}\langle\nabla f(x_k), x_k - x_{k+1}\rangle + 2\lambda_k\langle\nabla f(x_k), x - x_k\rangle - \|x_{k+1} - x_k\|^2 \\
&\leqslant \|x_k - x\|^2 + 2\alpha_{\max}\{f(x_k) - f(x_{k+1})\} + 2\lambda_k\langle\nabla f(x_k), x - x_k\rangle \\
&\quad + 2\alpha_{\max}\|\nabla f(x_k) - \nabla f(\xi_k)\|\|x_k - x_{k+1}\| - \|x_{k+1} - x_k\|^2 \\
&\leqslant \|x_k - x\|^2 + 2\alpha_{\max}\{f(x_k) - f(x_{k+1})\} + (\theta_k - 1)\|x_k - x_{k+1}\|^2 \\
&\quad + 2\lambda_k\langle\nabla f(x_k), x - x_k\rangle \\
&\leqslant \|x_k - x\|^2 + 2\alpha_{\max}\{f(x_k) - f(x_{k+1})\} + 2\lambda_k\langle\nabla f(x_k), x - x_k\rangle,
\end{aligned}$$

where $\xi_k = x_k + \theta_k(x_{k+1} - x_k)$, $\theta_k \in (0, 1)$. The proof is completed. □

Lemma 4.2. *Let $\{x_k\}$ be an infinite sequence generated by the Algorithm* 2.2. *If $\nabla f(x)$ is Lipschitz continuous on Ω and $\alpha_{\max} \leqslant 1/2L$, then we have for all $k \in N$ and any $x \in \Omega$,*

$$\|x_{k+1} - x\|^2 \leqslant \|x_k - x\|^2 + 2\alpha_{\max}\{f(x_k) - f(x_{k+1})\} + 2\alpha_k\lambda_k\langle\nabla f(x_k), x - x_k\rangle.$$

The proof of this Lemma is similar to that of Lemma 4.1, so we omit the details.

Theorem 4.1. *Let f be pseudo-convex on Ω, $\nabla f(x)$ is Lipschitz continuous on Ω and $\{x_k\}$ be an infinite sequence generated by Algorithm* 2.1 (*or Algorithm* 2.2). *If* $\alpha_{\max} \leqslant 1/2L$, *then we have*

(1) $\Omega^* \neq \emptyset$ *if and only if* $\lim_{k\to+\infty} x_k = x^*$ *where* $x^* \in \Omega^*$.
(2) *Otherwise*, $\liminf_{k\to\infty} f(x_k) = \inf\{f(x)|x \in \Omega\}$.

Proof. (1) Assume that $\Omega^* \neq \emptyset$, then for any $x \in \Omega^*$, we have

$$f(x) < f(x_k) \quad \forall k \in N. \tag{21}$$

Using the pseudo-convexity of $f(x)$, we derive

$$\langle \nabla f(x_k), x - x_k \rangle < 0 \quad \forall k \in N$$

which, together with Lemma 4.1 (or Lemma 4.2), deduces that

$$\{\|x_k - x\|^2 + 2\alpha_{\max} f(x_k)\} \downarrow \tag{22}$$

Conditions (21) and (22) imply that $\{x_k\}$ is bounded. So, there exists at least one limit point x^* of $\{x_k\}$ and $K \subseteq N$ such that

$$\lim_{k\in K, k\to+\infty} x_k = x^*.$$

From the pseudo-convexity of f and Corollary 3.3 (Corollary 3.6), we have $x^* \in \Omega^*$. Taking $x = x^*$ in (22), we have

$$\{\|x_k - x^*\|^2 + 2\alpha_{\max} f(x_k)\} \downarrow .$$

So,

$$\begin{aligned}\lim_{k\to+\infty} \{\|x_k - x^*\|^2 + 2\alpha_{\max} f(x_k)\} &= \lim_{k\in K, k\to+\infty} \{\|x_k - x^*\|^2 + 2\alpha_{\max} f(x_k)\} \\ &= 2\alpha_{\max} f(x^*).\end{aligned} \tag{23}$$

Since the continuity of f, the boundedness of $\{x_k\}$ shows that $\{f(x_k)\}$ is bounded. Now there exists a subsequence $\{f(x_k)\}_{\bar{K}}$ of the sequence $\{f(x_k)\}$ satisfying

$$\lim_{k\in \bar{K}, k\to+\infty} f(x_k) = l.$$

$\bar{x}$ is a cluster point of $\{x_k\}_{\bar{K}}$. Again, by Corollary 3.3 (Corollary 3.6), we have $\bar{x} \in \Omega^*$. By the continuity of f, we have

$$l = \lim_{k\in \bar{K}, k\to\infty} f(x_k) = f(\bar{x}) = f(x^*).$$

Hence,

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

From (23), we obtain that

$$\lim_{k\to\infty} x_k = x^*.$$

We now prove (2). It suffices to prove (2) for the case where $\Omega^* = \emptyset$. In this case we have from (1),

$$\lim_{k\to\infty} \|x_k\| = +\infty. \tag{24}$$

Assume that $\lim\, \inf_{k\to\infty} f(x_k) > \inf\{f(x)|x \in \Omega\}$. Then there exists a point $\bar{x} \in \Omega$ such that

$$f(\bar{x}) < f(x_k) \quad \forall k \in N.$$

Similar to the previous proof, we derive that $\{x_k\}$ is bounded, which is a contradiction to (24). Therefore, (2) holds. This completes the proof. □

Theorem 4.2. *Let f be quasi-convex on Ω, $\nabla f(x)$ is Lipschitz continuous on Ω and $\{x_k\}$ be an infinite sequence generated by Algorithm* 2.1 (*or Algorithm* 2.2). *If $\alpha_{\max} \leqslant 1/2L$, then we have*

(1) $\Omega^+ \neq \emptyset$ *if and only if* $\lim_{k\to+\infty} x_k = x^+$ *where* $x^+ \in \Omega^+$.
(2) *Otherwise*, $\lim\, \inf_{k\to\infty} f(x_k) = \inf\{f(x)|x \in \Omega\}$.

Proof. The proof of this theorem is similar to that of Theorem 4.1, so we omit the details. □

Remark. All of the results in this section assume that $\alpha_{\max} \leqslant 1/2L$ and $\nabla f(x)$ is Lipschitz continuous on Ω. These are restrictive, which may be relative to the nonmonotone line search. However, if we set $M = 1$ in Algorithms 2.1 and 2.2, i.e. with monotone line search, the same results can be obtained without the assumptions of Lipschitz continuous of $\nabla f(x)$ and $\alpha_{\max} \leqslant 1/2L$. The proof of Lemmas 4.1 and 4.2 are similar to that of Lemma 3 in [25], so we omit the details.

5. Finite termination of algorithm

An algorithm is called finite convergent, if the sequence $\{x_k\}$ produced by the algorithm satisfies that there exists k_0 such that $x_k \in \Omega^+$ for all $k \geqslant k_0$. The finite termination of algorithm was originally studied by Burke and Ferris [7]. In Ref. [7], they introduced weak sharp condition on Ω and Ω^*, i.e., for every $x^* \in \Omega^*$,

$$-\nabla f(x^*) \in \operatorname{int} \bigcap_{x\in\Omega} [T_\Omega(x) \cap N_{\Omega^*}(x)]^\circ. \quad \text{(WS)}$$

If condition (WS) holds, they obtained a condition for the sequence $\{x_k\}$ generated by algorithms to converge finitely to an optimal solution of a problem of minimizing a differentiable convex function. Later, Marcotte and Zhu [19] gave a condition for the method of solving pseudo-monotone variational inequalities to terminate finitely, if condition (WS) holds.

In this section, for the general optimization problem (1) without convexity (convexity or pseudo-convexity) assumption which was required in Refs. [7,19], we prove under condition (WS) that if the sequence $\{x_k\}$ generated by the algorithm is bounded, then the Algorithm 2.1 (Algorithm 2.2) terminate finitely.

Theorem 5.1. *Suppose Ω^+ is a nonempty closed convex set in problem* (1) *and condition* (WS) *on Ω and Ω^+ holds. Let $\{x_k\}$ is bounded, then Algorithm* 2.1 (*Algorithm* 2.2) *terminate finitely*.

Proof. We only prove the result for Algorithm 2.1 (The proof for algorithm 2.2 is similar to that of Algorithm 2.1). Suppose, on the contrary, that there exists an infinite subsequence $\{x_k\}_{k\in K}(K \subseteq N)$ such that for all $k \in K$, $x_k \notin \Omega^+$, that is,

$$\|x_k - P_{\Omega^+}(x_k)\| > 0 \quad \forall k \in K. \tag{25}$$

Since $\{x_k\}$ is bounded, Corollary 3.3 implies that

$$\lim_{k\in K, k\to\infty} x_k = x^+, \tag{26}$$

where $x^+ \in \Omega^+$. By condition (WS) on Ω and Ω^+, we have

$$-\nabla f(x^+) \in \text{int} \bigcap_{x\in\Omega} [T_\Omega(x) \cap N_{\Omega^+}(x)]^\circ. \tag{27}$$

Using (27), there exists $\alpha > 0$ such that for any $x \in \Omega^+$,

$$-\nabla f(x^+) + \alpha B \in [T_\Omega(x) \cap N_{\Omega^+}(x)]^\circ, \tag{28}$$

where B is an unit sphere on R^n. By the definition of polar cone and (28), we derive that for any $x \in \Omega^+$ and any $d \in T_\Omega(x) \cap N_{\Omega^+}(x)$,

$$\left\langle -\nabla f(x^+) + \alpha \frac{d}{\|d\|}, d \right\rangle \leqslant 0,$$

i.e.

$$\alpha \|d\| \leqslant \langle \nabla f(x^+), d \rangle. \tag{29}$$

Setting $z_k = P_{\Omega^+}(x_k)$ and $d_k = x_k - z_k$, condition (25) implies that $d_k \neq 0$ and d_k is a feasible direction on Ω at z_k. So,

$$d_k \in T_\Omega(z_k).$$

By using projection properties, we have that for all $y \in \Omega^+$,

$$\langle d_k, y - z_k \rangle \leqslant 0.$$

Thus, by the convexity of Ω^+, we derive $d_k \in N_{\Omega^+}(z_k)$. Hence,

$$d_k = x_k - z_k \in T_\Omega(z_k) \cap N_{\Omega^+}(z_k).$$

Condition (29) implies that for all $k \in K$,

$$\alpha \leqslant \frac{\langle \nabla f(x^+), x_k - x_k \rangle}{\|x_k - z_k\|}. \tag{30}$$

By (30) and Lemma 2.3, we obtain that for all $k \in K$,

$$\begin{aligned}\alpha &\leqslant \frac{\langle \nabla f(x^+), x_k - z_k\rangle}{\|x_k - z_k\|}\\ &= \frac{\langle \nabla f(x_k), x_k - z_k\rangle}{\|x_k - z_k\|} + \frac{\langle \nabla f(x^+) - \nabla f(x_k), x_k - z_k\rangle}{\|x_k - z_k\|}\\ &\leqslant \frac{\|x_k - x_{k-1}\|}{\lambda_{k-1}} + \|\nabla f(x_k) - \nabla f(x_{k-1})\| + \|\nabla f(x_k) - \nabla f(x^+)\|. \end{aligned} \tag{31}$$

Moreover, from (26), the continuity of $\nabla f(\cdot)$ shows that

$$\lim_{k\in K, k\to\infty} \|\nabla f(x_k) - \nabla f(x^+)\| = 0. \tag{32}$$

Again, by Corollary 3.1 and the boundedness of λ_k, we have

$$\lim_{k\to\infty} \frac{\|x_k - x_{k-1}\|}{\lambda_{k-1}} = 0, \tag{33}$$

$$\lim_{k\in K, k\to\infty} \|\nabla f(x_k) - \nabla f(x_{k-1})\| = 0. \tag{34}$$

Therefore, taking limit in (31) as $k \in K$, $k \to \infty$, (32)–(34) imply that $\alpha \leqslant 0$. This is a contradiction. □

Corollary 5.1. *Suppose Ω^+ is a nonempty closed convex set in problem* (1) *and condition* (WS) *on Ω and Ω^+ holds. Let $f(\cdot)$ is pseudo-convex* (*or quasi-convex*). *If $\nabla f(\cdot)$ is Lipschitz continuous on Ω and $\alpha_{\max} = 1/2L$, then Algorithms* 2.1 *and* 2.2 *terminate finitely*.

Proof. By using Theorems 4.1 or 4.2, we have that the sequence $\{x_k\}$ generated by Algorithms 2.1 or 2.2 is convergent. Thus by Theorem 5.1 we obtain the desired result. □

If $f(\cdot)$ is convex, then the condition for the finite convergence of Algorithm 2.1 (or Algorithm 2.2) is obviously weaker than that of Theorem 5.1.

Theorem 5.2. *Suppose Ω^* is nonempty in problem* (1) *and condition* (WS) *on Ω and Ω^* holds. Let $f(\cdot)$ be convex on Ω and $\{x_k\}$ be an infinite sequence generated by Algorithms* 2.1 *or* 2.2. *If $\nabla f(\cdot)$ is uniformly continuous on an open convex set containing $\{x_k\}$, then the algorithm terminate finitely*.

Proof. We only prove the result for Algorithm 2.1. Since condition (WS) holds, using Corollary 2.7 in Ref. [7], we can show that there exists $\alpha > 0$ for any $x^* \in \Omega^*$ and $x \in \Omega$, such that

$$f(x) - f(x^*) \geqslant \alpha \operatorname{dist}(x, \Omega^*), \tag{35}$$

where $\operatorname{dist}(x, \Omega^*) = \|x - P_{\Omega^*}(x)\|$. Suppose, on the contrary, that there exists an infinite subsequence $\{x_k\}_{k\in K}(K \subseteq N)$ satisfying (25). From (25), (35) and Lemma 2.3, the convexity of $f(\cdot)$ shows that for all $k \in K$,

$$\begin{aligned} \alpha &\leqslant \frac{f(x_k) - f(P_{\Omega^*}(x_k))}{\|x_k - P_{\Omega^*}(x_k)\|} \\ &\leqslant \frac{\langle \nabla f(x_k), x_k - P_{\Omega^*}(x_k)\rangle}{\|x_k - P_{\Omega^*(x_k)}\|} \\ &\leqslant \frac{\|x_k - x_{k-1}\|}{\lambda_{k-1}} + \|\nabla f(x_k) - \nabla f(x_{k-1})\|. \end{aligned} \tag{36}$$

Taking limit in (36) as $k \in K, k \to \infty$, Corollary 3.1, the boundedness of $\{\lambda_k\}$ and the uniform continuity of $\nabla f(\cdot)$ imply that $\alpha \leqslant 0$, giving a contradiction. □

6. Final remarks

When we completed the paper and reported it at International Conference on Numerical Linear Algebra and Optimization (7–10 October, 2003, Guilin, China), Raydan told us that the proof of Theorems 2.3 and 2.4 in the paper [4] containing minor errors had been corrected in the paper [5]. In fact, it is easy to see that the main results in [5] are still special cases in our paper.

References

[1] J. Barzilai, J.M. Borwein, Two point step size gradient methods, IMA J. Numer. Anal. 8 (1988) 141–148.
[2] D.P. Bertsekas, Projected Newton method for optimization problems with simple constrains, SIAM J. Control Optim. 20 (1982) 221–246.
[3] E.G. Birgin, J.M. Martínez, A spectral conjugate gradient method for unconstrained optimization, Appl. Math. Optim. 43 (2001) 117–128.
[4] E.G. Birgin, J.M. Martínez, M. Raydan, Nonmonotone spectral projected gradient methods on convex set, SIAM J. Optim. 10 (2000) 1196–1211.
[5] E.G. Birgin, J.M. Martínez, M. Raydan, Inexact spectral projected gradient methods on convex sets, IMA J. Numer. Anal. 23 (2003) 539–559.
[6] J.F. Bonnans, E.R. Panier, A.L. Tits, J.L. Zhou, Avoiding the Maratos effect by means of a nonmonotone linear search II inequality constrained problems-feasible iterates, SIAM J. Numer. Anal. 29 (1992) 1187–1202.
[7] J.V. Burke, M.C. Ferris, Weak sharp minima in mathematical programming, SIAM J. Control. Optim. 31 (1993) 1340–1359.
[8] P.H. Calamai, J.J. Moré, Projected gradient methods for linearly constrained problems, Math. Programming 39 (1987) 93–116.
[9] Y.C. Cheng, On the gradient-projection method for solving the nonsymmetric linear complementarity problem, J. Optim. Theory Appl. 43 (1984) 527–541.
[10] N.Y. Deng, Y. Xiao, F.J. Zhou, A nonmonotonic trust region algorithm, J. Optim. Theory Appl. 76 (1993) 259–285.
[11] J.C. Dunn, On the convergence of projected gradient processes to singular attractors, J. Optim. Theory Appl. 55 (1987) 203–215.
[12] A.A. Goldstein, Convex programming in Hilbert space, Bull. Amer. Soc. 70 (1964) 709–710.
[13] A.A. Goldstein, On gradient projection, in: Proceedings of the 12th Annual Allerton Conference on Circuits and Systems, Allerton Park, IL, 1974, pp. 38–40.
[14] C.C. Gonzaga, Two facts on the convergence of the Cauchy algorithm, J. Optim. Theory Appl. 107 (2000) 591–600.
[15] L. Grippo, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's method, SIAM J. Numer. Anal. 23 (1986) 707–716.

[16] L. Grippo, F. Lampariello, S. Lucidi, A truncated Newton method with nonmonotone line search for unconstrainted optimization, J. Optim. Theory Appl. 60 (1989) 401–419.
[17] E.S. Leviton, B.T. Polyak, Constrained minimization problems, USSR. Comput. Math. Math. Phys. 6 (1966) 1–50.
[18] Z.Q. Luo, P. Tseng, On the linear convergence of descent methods for convex essentially smooth minimization, SIAM J. Control Optim. 30 (1992) 408–425.
[19] P. Marcotte, D. Zhu, Weak sharp solutions of variational inequalities, SIAM J. Optim. 9 (1998) 179–189.
[20] E.R. Panier, A.L. Tits, Avoiding the Maratos effect by means of a nonmonotone line search I. Constrained problems, SIAM J. Numer. Anal. 28 (1991) 1183–1196.
[21] M. Raydan, On the Barailai and Borwein choice of steplength for the gradient method, IMA J. Numer. Anal. 13 (1993) 321–326.
[22] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, SIAM J. Optim. 7 (1997) 26–33.
[23] A. Schwartz, E. Polak, Family of projected descent methods for optimization problems with simple bounds, J. Optim. Theory Appl. 92 (1997) 1–31.
[24] C.Y. Wang, B. Qu, Convergence of the gradient projection method with a new stepsize rule, OR Trans. 6 (1) (2002) 36–44.
[25] C.Y. Wang, N.H. Xiu, Convergence of the gradient projection method for generalized convex minimization, Comp. Optim. Appl. 16 (2000) 111–120.
[26] N.H. Xiu, C.Y. Wang, J.Z. Zhang, Convergence properties of projection and contraction methods for variational inequality problems, Appl. Math. Optim. 43 (2001) 147–168.