Available online at www.isr-publications.com/jnsa
J. Nonlinear Sci. Appl., 10 (2017), 5566–5575
Research Article

ISSN: 2008-1898

Journal of Nonlinear Sciences and Applications

Journal Homepage: www.tjnsa.com - www.isr-publications.com/jnsa

Perturbation resilience of proximal gradient algorithm for composite objectives

Yanni Guo[a], Wei Cui[a], Yansha Guo[b,*]

[a]College of Science, Civil Aviation University of China, Tianjin, 300300, P. R. China.

[b]School of Information Technology Engineering, Tianjin University of Technology and Education, Tianjin, 300222, P. R. China.

Communicated by Y. H. Yao

Abstract

In this paper, we study the perturbation resilience of a proximal gradient algorithm under the general Hilbert space setting. With the assumption that the error sequence is summable, we prove that the iterative sequence converges weakly to a solution of the composite optimization problem. We also show the bounded perturbation resilience of this iterative method and apply it to the lasso problem.

Keywords: Perturbation resilience, proximal gradient algorithm, composite optimization, lasso problem.
2010 MSC: 47H10, 47J25, 65J15.

1. Introduction

Perturbation resilience is concerned with the recent developed optimization scheme called "superiorization" [4], which provides a novel way to obtain solutions by carefully selecting perturbations in an active way. Contrary to exacted methods, superiorization approach does not necessarily find an optimal solution for a given objective function. Instead, it may try to find a point. This point will have a low cost function value and hence it will be superior to other points. This heuristic method is usually less time consuming that makes it applicable to some important practical problems such as medical image recovery [9, 16], computed tomography [20], intensity-modulated radiation therapy [8], and so on.

Nevertheless, the superiorized version produced by superiorization methodology is heavily based on the bounded perturbation resilience of the original iterative algorithm. Very recently, several articles investigated the perturbation resilience. [6] considered the perturbation resilience and convergence of dynamic string-averaging projection method for solving the convex feasibility problem. [10] studied the convex minimization problem:

$$\min_{x\in C} f(x),$$

where C is a nonempty closed convex subset of finite dimension space R^J and the objective function f is convex. The following projected scaled gradient (PSG) algorithm with errors:

*Corresponding author
Email addresses: ynguo@amss.ac.cn (Yanni Guo), 851241788@qq.com (Wei Cui), gysh337@126.com (Yansha Guo)

doi:10.22436/jnsa.010.10.36

Received 2017-07-17

$$x_{n+1} = P_C(x_n - \gamma_n D(x_n)\nabla f(x_n) + e(x_n)) \tag{1.1}$$

is introduced, and the exact PSG algorithm is proved to be bounded perturbation resilient. Later on, [22] generalized algorithm (1.1) to general Hilbert space setting and showed that (1.1) converges in a sublinear rate. The bounded perturbation resilience is also concluded. Equation (1.1) is in fact a kind of inexact iterative scheme, which was developed in many literatures over the past two decades for the reason that solving the problems is either expensive or impossible, see [11, 13, 19, 23, 24] and references therein for instance.

Motivated by [10] and [22], we consider the non-smooth composite optimization problem of the form

$$\min_{x\in H} \Phi(x), \quad \Phi(x) = f(x) + g(x) \tag{1.2}$$

in a real Hilbert space H, where f, g are proper lower semi-continuous convex functions and satisfy assumptions:

(A1) f is differentiable, whose gradient is Lipschitz continuous with constant L. That is,

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x-y\|, \quad x, y \in H.$$

g may not be differentiable.

(A2) $f + g = \Phi$ is coercive, namely,

$$\lim_{\|x\|\to+\infty} \Phi(x) = +\infty.$$

Hence, Φ has a minimizer over H, or $S := \operatorname{Argmin}(\Phi) \neq \emptyset$ ([3, Proposition 11.14]).

The proximal gradient (PG) algorithm, or forward-backward splitting method ([12, 17]), with the combinations of explicit (forward) gradient steps with respect to the smooth part and proximal (backward) part with respect to the non-smooth part, is an appealing approach for solving these types of non-smooth optimization problems because of their fast theoretical convergence rates and strong practical performance [15].

The classical proximal gradient algorithm generates a sequence $\{x_n\}_{n=0}^{\infty}$ starting from a given $x_0 \in H$ and recursively utilizes the rule

$$x_{n+1} = (I + \alpha\partial g)^{-1}(I - \alpha\nabla f)(x_n), \text{ for } n \geqslant 0, \tag{1.3}$$

where $\alpha > 0$ is the step size. It is known that if

$$S := \operatorname{Argmin}(\Phi) \neq \emptyset,$$

any sequence generated by algorithm (1.3) converges weakly to an element of S if

$$0 < \alpha < \frac{2}{L}$$

(see, for instance, [3, Theorem 25.8]).

Recently, a slightly more general PG algorithm, where α is replaced by α_n, was introduced by Xu [21]:

$$x_{n+1} = (I + \alpha_n\partial g)^{-1}(I - \alpha_n\nabla f)(x_n) \tag{1.4}$$

for convex optimization problem (1.2). The weak convergence of the algorithm is given as follows.

Theorem 1.1 ([21, Theorem 3.1])**.** *Let H be a real Hilbert space, $f, g \in \Gamma_0(H)$. Assume that (1.2) is solvable. Assume in addition that*

(1) *∇f is Lipschitz continuous on H, that is, there exists a number $L > 0$ such that*

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x-y\|, \quad \forall\, x, y \in H.$$

(2)
$$0 < \inf \alpha_n \leqslant \alpha_n \leqslant \sup \alpha_n < \frac{2}{L}.$$

Then the sequence $\{x_n\}_{n=0}^{\infty}$ *generated by algorithm* (1.4) *converges weakly to a solution of* (1.2). *No strong convergence is guaranteed if* $\dim H = \infty$.

In this paper, we shall prove that a more general PG algorithm is perturbation resilient in the general setting of Hilbert space. In addition, we will see that the bounded perturbation resilience can be changed into a special case of perturbation resilience. Given $x_0 \in H$ arbitrarily, the algorithm we shall focus on is

$$x_{n+1} = (I + \alpha_n \partial g)^{-1}(x_n - \alpha_n D(x_n)\nabla f(x_n) + e(x_n)) \tag{1.5}$$

with the following conditions hold.

(A3) For each $x \in H$, $D(x) : H \to H$ is a bounded linear operator.
(A4) $0 < \inf \alpha_n \leqslant \alpha_n \leqslant \sup \alpha_n < \frac{2}{L}$.
(A5) $e(x_n)$ denotes the computational error in computing the gradient $\nabla f(x_n)$ and satisfies

$$\sum_{n=0}^{\infty} \|e(x_n)\| < +\infty. \tag{1.6}$$

To ensure that $D(x_n)\nabla f(x_n)$ does not deviate too much from the gradient $\nabla f(x_n)$, we define

$$\theta^n := \nabla f(x_n) - D(x_n)\nabla f(x_n),$$

and assume that

(A6)
$$\sum_{n=0}^{\infty} \|\theta^n\| < +\infty. \tag{1.7}$$

We proceed as follows. In Section 2, we list some definitions and lemmas that will be useful in the next section. In Section 3, we prove that the sequence $\{x_n\}_{n=0}^{\infty}$ generated by (1.5) converges weakly to a solution of problem (1.2). In Section 4, we deal with the bounded perturbation resilience of (1.4) and the basic algorithm of (1.5). Finally, in Section 5, we apply the bounded perturbation resilience of (1.4) to the lasso problem.

2. Preliminaries

Let H be a real Hilbert space with inner product $< \cdot, \cdot >$ and the induced norm $\|\cdot\|$. Recall that a mapping $V : H \to H$ is non-expansive if

$$\|Vx - Vy\| \leqslant \|x - y\|, \ \forall x,\ y \in H.$$

A mapping $V : H \to H$ is α-averaged if
$$V = (1-\alpha)I + \alpha T,$$

where $\alpha \in (0, 1)$ and T is non-expansive. Let $\mathrm{Fix}(T)$ stand for the fixed point set of T.

Denote by

$$\Gamma_0(H) = \{f : H \to (-\infty, \infty] | \text{ f is proper lower semi-continuous, convex}\}.$$

We need the following definition.

Definition 2.1 ([14, Proximal Operator]). Let $g \in \Gamma_0(H)$. The proximal operator of g is defined by

$$\mathrm{prox}_g(x) := \arg\min_{y\in H}\left\{\frac{\|y-x\|^2}{2}+g(y)\right\}, \quad x \in H.$$

The proximal operator of g of order $\alpha > 0$ is defined as the proximal operator of αg, that is,

$$\mathrm{prox}_{\alpha g}(x) := \arg\min_{y\in H}\left\{\frac{\|y-x\|^2}{2\alpha}+g(y)\right\}, \quad x \in H.$$

The following Lemmas 2.2 and 2.3 describe the properties of the proximal operators.

Lemma 2.2 ([21]). *Let $g \in \Gamma_0(H)$, $\alpha > 0$, $\mu > 0$. Then*

(i) $\mathrm{prox}_{\alpha g}(x) = (I+\alpha\partial g)^{-1}(x)$;

(ii) $\mathrm{prox}_{\alpha g}(x) = \mathrm{prox}_{\mu g}\left(\frac{\mu}{\alpha} + (1-\frac{\mu}{\alpha})\mathrm{prox}_{\alpha g}x\right)$.

Lemma 2.3 ([3, Non-expansiveness]). *Let g be a convex function on H, and $\alpha > 0$. Then the proximal operator of g is non-expansive:*

$$\|(I+\alpha\partial g)^{-1}(x)-(I+\alpha\partial g)^{-1}(y)\| \leqslant \|x-y\|, \ \forall x,y \in H. \tag{2.1}$$

Lemma 2.4 ([18, Decent Lemma]). *Let $f : H \to \mathbb{R}$ be a differentiable function with L-Lipschitz continuous gradient ∇f. Then for any $x, y \in H$, we have*

$$f(x) \leqslant f(y)+ < \nabla f(y), x-y > +\frac{L}{2}\|x-y\|^2.$$

Lemma 2.5 ([21, Proposition 3.2]). *Let the functions $f, g \in \Gamma_0(H)$, $z \in H$ and $\alpha > 0$. Assume that f is differentiable on H. Then z is an element of the solution set S if and only if z solves the fixed point equation*

$$z = (I+\alpha\partial g)^{-1}(I-\alpha\nabla f)z.$$

Lemma 2.6 ([7, Lemma 2.1]). *Let $\{\alpha_n\}_{n=0}^{\infty} \subset \mathbb{R}_+$ be a sequence of nonnegative real numbers. If it holds that $0 \leqslant \alpha_{n+1} \leqslant \alpha_n + \delta_n$ for all $n \geqslant 0$, where $\delta_n \geqslant 0$ for all $n \geqslant 0$ and $\sum_{n=0}^{\infty}\|\delta_n\| < \infty$, then $\{\alpha_n\}_{n=0}^{\infty}$ converges.*

The following two lemmas play important roles in proving the weak convergence of $\{x_n\}_{n=0}^{\infty}$ generated by (1.5).

Lemma 2.7 ([16]). *Let H be a real Hilbert space, and $T : H \to H$ a non-expansive mapping with the fixed point set $\mathrm{Fix}(T) \neq \emptyset$. If $\{x_n\}_{n=0}^{\infty}$ is a sequence in H converging weakly to x, and if $\{(I-T)x_n\}$ converges strongly to 0, then $(I-T)x = 0$.*

Lemma 2.8 ([1, Opial]). *Let H be a Hilbert space, and $\{x_n\}_{n=0}^{\infty}$ be a sequence such that there exists a nonempty set $S \subset H$ satisfying*

(i) $\lim_{n\to\infty}\|x_n - p\|$, $\forall p \in S$ *exists;*

(ii) *if x_{n_j} converges weakly to a point x in H for a subsequence $n_j \to +\infty$, then $x \in S$.*

Then there exists $\bar{x} \in S$ such that $\{x_n\}_{n=0}^{\infty}$ converges weakly to $\bar{x}$ as $n \to \infty$.

3. Convergence analysis

The proof of the main convergence theorem (Theorem 3.5 below) is based on several propositions.

Proposition 3.1. *Let the functions $f, g \in \Gamma_0(H)$. Let $\{x_n\}_{n=0}^{\infty}$ be a sequence generated by algorithm (1.5). Assume that (A1)-(A6) hold. Then the sequence of function values $\{\Phi(x_n)\}_{n=0}^{\infty}$ converges.*

Proof. Note that (1.5) implies that

$$\frac{1}{\alpha_n}(x_n - x_{n+1}) - D(x_n)\nabla f(x_n) + \frac{1}{\alpha_n}e(x_n) \in \partial g(x_{n+1}).$$

We get from the convexity of g and Lemma 2.4,

$$g(x_n) \geqslant g(x_{n+1}) + < s, x_n - x_{n+1} >, \quad \forall s \in \partial g(x_{n+1}),$$
$$f(x_n) \geqslant f(x_{n+1}) - < \nabla f(x_n), x_{n+1} - x_n > -\frac{L}{2}\|x_{n+1} - x_n\|^2.$$

Summing the both inequalities, we conclude

$$\Phi(x_{n+1}) \leqslant \Phi(x_n) + < \nabla f(x_n) + s, x_{n+1} - x_n > + \frac{L}{2}\|x_{n+1} - x_n\|^2.$$

Choose

$$s = \frac{1}{\alpha_n}(x_n - x_{n+1}) - D(x_n)\nabla f(x_n) + \frac{1}{\alpha_n}e(x_n) \in \partial g(x_{n+1}),$$

and denote

$$\delta^n = \frac{1}{\alpha_n}e(x_n) + \theta^n. \tag{3.1}$$

We have

$$\begin{aligned}\Phi(x_{n+1}) &\leqslant \Phi(x_n) + (\frac{L}{2} - \frac{1}{\alpha_n})\|x_{n+1} - x_n\|^2 + \|\delta^n\| \cdot \|x_{n+1} - x_n\| \\ &\leqslant \Phi(x_n) - \eta_1\|x_{n+1} - x_n\|^2 + \|\delta^n\| \cdot \|x_{n+1} - x_n\|,\end{aligned}$$

where $\eta_1 = \frac{1}{\sup_n\{\alpha_n\}} - \frac{L}{2} > 0$ owing to assumption (A4). Or

$$\begin{aligned}\Phi(x_n) - \Phi(x_{n+1}) &\geqslant \eta_1\|x_{n+1} - x_n\|^2 - \|\delta^n\| \cdot \|x_{n+1} - x_n\| \\ &\geqslant \eta_1(\|x_{n+1} - x_n\| - \frac{1}{2\eta_1}\|\delta^n\|)^2 - \frac{1}{4\eta_1}\|\delta^n\|^2.\end{aligned} \tag{3.2}$$

Let $\underline{\Phi} = \inf_{x\in H}\Phi(x)$, then $\Phi(x) \geqslant \underline{\Phi}$ for all $x \in H$. Consequently,

$$0 \leqslant \Phi(x_{n+1}) - \underline{\Phi} \leqslant \Phi(x_n) - \underline{\Phi} + \frac{1}{4\eta_1}\|\delta^n\|^2. \tag{3.3}$$

So $\Phi(x_n) - \underline{\Phi}$ converges according to Lemma 2.5 and that $\sum_{n=1}^{\infty}\|\delta^n\|^2 < \infty$ (see (3.1), (A5)-(A6)). Hence the sequence $\{\Phi(x_n)\}_{n=0}^{\infty}$ also converges. □

Proposition 3.2. *Let* $\{x_n\}_{n=0}^{\infty}$ *be generated by algorithm* (1.5)*. It holds that*

$$\|x_{n+1} - x_n\| \leqslant \sqrt{\frac{2}{\eta_1}}|\Phi(x_{n+1}) - \Phi(x_n)|^{\frac{1}{2}} + \frac{1}{\eta_1}\|\delta_n\|.$$

Hence, it has $\lim_{n\to\infty}\|x_{n+1} - x_n\| \to 0$.

Proof. In view of (3.2), we observe

$$\begin{aligned}\|x_{n+1} - x_n\|^2 &= (\|x_{n+1} - x_n\| - \frac{1}{2\eta_1}\|\delta_n\| + \frac{1}{2\eta_1}\|\delta_n\|)^2 \\ &\leqslant 2\left[\left(\|x_{n+1} - x_n\| - \frac{\|\delta_n\|}{2\eta_1}\right)^2 + \frac{1}{4\eta_1^2}\|\delta_n\|^2\right] \leqslant \frac{2}{\eta_1}[\Phi(x_n) - \Phi(x_{n+1})] + \frac{\|\delta_n\|^2}{\eta_1^2}.\end{aligned}$$

Thereby,

$$\|x_{n+1} - x_n\| \leqslant \sqrt{\frac{2}{\eta_1}}|\Phi(x_n) - \Phi(x_{n+1})|^{1/2} + \frac{\|\delta_n\|}{\eta_1},$$

which follows from the inequality $\sqrt{a^2 + b^2} \leqslant a + b$ for all $a, b \geqslant 0$. Accordingly, we get

$$\lim_{n\to\infty}\|x_{n+1} - x_n\| \to 0 \tag{3.4}$$

by Proposition 3.1. □

We need the non-expansive property of the mapping $(I+\alpha\partial g)^{-1}(I-\alpha\nabla f)$.

Proposition 3.3. *Let $f, g \in \Gamma_0(H)$. For any $0<\alpha<\frac{2}{L}$, where L is the Lipschitz constant of ∇f, $(I+\alpha\partial g)^{-1}(I-\alpha\nabla f)$ is $\frac{\alpha L+2}{4}$-averaged. Hence, it is non-expansive.*

Proof. At first, ∇f is L-Lipschitzian indicates that ∇f is $\frac{1}{L}$-ism ([2]). Consequently, $I-\alpha\nabla f$ is $\frac{\alpha L}{2}$-averaged as $0<\alpha<\frac{2}{L}$ ([3, Proposition 4.33]). Besides, $(I+\alpha\partial g)^{-1}$ is $\frac{1}{2}$-averaged, the composite $(I+\alpha\partial g)^{-1}(I-\alpha\nabla f)$ is $\frac{\alpha L+2}{4}$-averaged, then it is non-expansive. □

Proposition 3.4. *Let $f, g \in \Gamma_0(H)$. Let $\{x_n\}_{n=0}^{\infty}$ be generated by algorithm (1.5), and assumptions (A1)-(A6) hold. Then there exists a subsequence $\{x_{n_k}\}_{k=0}^{\infty} \subseteq \{x_n\}_{n=0}^{\infty}$ converging weakly to $x^* \in S$.*

Proof. Applying (3.3) recursively, we obtain

$$\Phi(x_{n+1}) \leqslant \Phi(x_{n-1}) + \frac{1}{4\eta_1}(\|\delta^n\|^2+\|\delta^{n-1}\|^2) \leqslant \Phi(x_0) + \frac{1}{4\eta_1}\sum_{k=0}^{n}\|\delta^k\|^2.$$

Denote by $M := \frac{1}{4\eta_1}\sum_{k=0}^{\infty}\|\delta^k\|^2$ as $\sum_{k=0}^{\infty}\|\delta^k\|^2 < +\infty$. Clearly, $\{x_n\}_{n=0}^{\infty} \subset \{x \in H | \underline{\Phi} \leqslant \Phi(x_{n+1}) \leqslant \Phi(x_0)+M\}$, which is bounded owing to the coercivity of Φ and $\underline{\Phi} = \inf_{x\in H}\Phi(x) > -\infty$. The Bolzano Weierstrass theorem shows that there exists a subsequence $\{x_{n_k}\}_{k=0}^{\infty}$ which is weakly convergent.

Denote by x^* the weak convergence of $\{x_{n_k}\}_{k=0}^{\infty}$. Let α be the limitation of the subsequence $\{\alpha_{n_j}\}_{j=0}^{\infty}$ of $\{\alpha_n\}_{n=0}^{\infty}$ since $\{\alpha_n\}_{n=0}^{\infty}$ is bounded. Then $0<\alpha<\frac{2}{L}$ as assumption (A4) holds. Next, we will show that $x^* \in S$, which is equivalent to show that

$$\|x_{n_j} - (I+\alpha\partial g)^{-1}(I-\alpha\nabla f)(x_{n_j})\| \to 0, \quad \text{as} \quad j\to\infty$$

according to Lemma 2.5 as well as Lemma 2.7 and Proposition 3.3.

As a matter of fact, notice from Lemma 2.2 that, for any $\alpha, \alpha_{n_j} > 0$,

$$(I+\alpha\partial g)^{-1}(x) = (I+\alpha_{n_j}\partial g)^{-1}[\frac{\alpha_{n_j}}{\alpha}x + (1-\frac{\alpha_{n_j}}{\alpha})(I+\alpha\partial g)^{-1}(x)]. \tag{3.5}$$

We deduce from (3.5), (2.1), and Proposition 3.3, that

$$\begin{aligned}
&\|x_{n_j} - (I+\alpha\partial g)^{-1}(I-\alpha\nabla f)(x_{n_j})\| \\
&\quad \leqslant \|x_{n_j} - x_{n_{j+1}}\| + \|x_{n_{j+1}} - (I+\alpha\partial g)^{-1}(I-\alpha\nabla f)(x_{n_j})\| \\
&\quad = \|x_{n_j} - x_{n_{j+1}}\| + \|(I+\alpha_{n_j}\partial g)^{-1}[(I-\alpha_{n_j}D\nabla f)+e](x_{n_j}) - (I+\alpha\partial g)^{-1}(I-\alpha\nabla f)(x_{n_j})\| \\
&\quad = \|x_{n_j} - x_{n_{j+1}}\| + \|(I+\alpha_{n_j}\partial g)^{-1}[(I-\alpha_{n_j}D\nabla f)+e](x_{n_j}) \\
&\qquad - (I+\alpha_{n_j}\partial g)^{-1}[\frac{\alpha_{n_j}}{\alpha}I + (1-\frac{\alpha_{n_j}}{\alpha})(I+\alpha\partial g)^{-1}](I-\alpha\nabla f)(x_{n_j})\| \\
&\quad \leqslant \|x_{n_j} - x_{n_{j+1}}\| + \|[(I-\alpha_{n_j}D\nabla f)+e](x_{n_j}) \\
&\qquad - [\frac{\alpha_{n_j}}{\alpha}I + (1-\frac{\alpha_{n_j}}{\alpha})(I+\alpha\partial g)^{-1}](I-\alpha\nabla f)(x_{n_j})\| \\
&\quad = \|x_{n_j} - x_{n_{j+1}}\| + \|(1-\frac{\alpha_{n_j}}{\alpha})x_{n_j} + \frac{\alpha_{n_j}}{\alpha}\cdot\alpha\nabla f(x_{n_j}) - \alpha_{n_j}D\nabla f(x_{n_j}) + e(x_{n_j}) \\
&\qquad - (1-\frac{\alpha_{n_j}}{\alpha})(I+\alpha\partial g)^{-1}(I-\alpha\nabla f)(x_{n_j})\| \\
&\quad = \|x_{n_j} - x_{n_{j+1}}\| + \|(1-\frac{\alpha_{n_j}}{\alpha})x_{n_j} + \alpha_{n_j}(\nabla f - D\nabla f)(x_{n_j}) + e(x_{n_j}) \\
&\qquad - (1-\frac{\alpha_{n_j}}{\alpha})(I+\alpha\partial g)^{-1}(I-\alpha\nabla f)(x_{n_j})\| \\
&\quad \leqslant \|x_{n_j} - x_{n_{j+1}}\| + (1-\frac{\alpha_{n_j}}{\alpha})\|x_{n_j}\| + \alpha_{n_j}\|(\nabla f - D\nabla f)(x_{n_j})\| + \|e(x_{n_j})\|
\end{aligned}$$

$$
\begin{aligned}
&+ (1 - \frac{\alpha_{n_j}}{\alpha})\|(I + \alpha\partial g)^{-1}(I - \alpha\nabla f)(x_{n_j})\| \\
&\leqslant \|x_{n_j} - x_{n_{j+1}}\| + 2(1 - \frac{\alpha_{n_j}}{\alpha})\|x_{n_j}\| + \alpha_{n_j}\|\theta_{n_j}\| + \|e(x_{n_j})\|.
\end{aligned}
$$

Now since $\{\|x_{n_j}\|\}_{j=0}^{\infty}$ is bounded, we can use $\alpha_{n_j} \to \alpha$ $(j \to +\infty)$ and (1.6), (1.7), together with (3.4) to conclude that $\|x_{n_j} - (I + \alpha\partial g)^{-1}(I - \alpha\nabla f)(x_{n_j})\| \to 0$ as $j \to \infty$. The proof is completed. □

Theorem 3.5. *Let $f, g \in \Gamma_0(H)$. Let $\{x_n\}_{n=0}^{\infty}$ be generated by algorithm* (1.5)*. If the assumptions* (A1)-(A6) *hold, then $\{x_n\}_{n=0}^{\infty}$ converges weakly to $\bar{x} \in S$ as $n \to \infty$, where S is the solution set of problem* (1.2).

Proof. We will apply Lemma 2.8 to the sequence $\{x_n\}_{n=0}^{\infty}$ and the solution set S. Due to Proposition 3.4, it remains to prove that $\lim_{n\to\infty} \|x_n - z\|$ exists for each $z \in S$. To see this, we rewrite x_{n+1} as

$$
\begin{aligned}
x_{n+1} &= (I + \alpha_n\partial g)^{-1}[(I - \alpha_n D\nabla f)(x_n) + e(x_n)] \\
&= (I + \alpha_n\partial g)^{-1}(I - \alpha_n\nabla f)(x_n) + (I + \alpha_n\partial g)^{-1}[(I - \alpha_n D\nabla f)(x_n) + e(x_n)] \\
&\quad - (I + \alpha_n\partial g)^{-1}(I - \alpha_n\nabla f)(x_n) \\
&=: (I + \alpha_n\partial g)^{-1}(I - \alpha_n\nabla f)(x_n) + \tilde{e}(x_n),
\end{aligned}
$$

where

$$
\tilde{e}(x_n) = (I + \alpha_n\partial g)^{-1}[(I - \alpha_n D\nabla f)(x_n) + e(x_n)] - (I + \alpha_n\partial g)^{-1}(I - \alpha_n\nabla f)(x_n)
$$

such that

$$
\|\tilde{e}(x_n)\| \leqslant \|(I - \alpha_n D\nabla f)(x_n) + e(x_n) - (I - \alpha_n\nabla f)(x_n)\| \leqslant \alpha_n\|\theta_n\| + \|e(x_n)\|.
$$

Hence, we have $\sum_{n=0}^{\infty} \|\tilde{e}(x_n)\| < \infty$ in view of (1.6) and (1.7).

By observing that $\mathrm{Fix}[(I + \alpha_n\partial g)^{-1}(I - \alpha_n\nabla f)] = S$ for each n, we compute

$$
\begin{aligned}
\|x_{n+1} - z\| &= \|(I + \alpha_n\partial g)^{-1}[(I - \alpha_n D\nabla f) + e](x_n) - z\| \\
&\leqslant \|(I + \alpha_n\partial g)^{-1}(I - \alpha_n\nabla f)(x_n) - z\| + \|\tilde{e}(x_n)\| \\
&= \|(I + \alpha_n\partial g)^{-1}(I - \alpha_n\nabla f)(x_n) - (I + \alpha_n\partial g)^{-1}(I - \alpha_n\nabla f)z\| + \|\tilde{e}(x_n)\| \\
&\leqslant \|x_n - z\| + \|\tilde{e}(x_n)\|.
\end{aligned}
$$

Now by virtue of Lemma 2.6, and that $\sum_{n=0}^{\infty} \|\tilde{e}(x_n)\| < \infty$, we conclude that $\lim_{n\to\infty} \|x_n - z\|$ exists. □

4. Bounded perturbation resilience

In this section, we prove the bounded perturbation resilience of the proximal gradient method by transform the bounded perturbation into a special case of the summable perturbation studied in the previous section. The following definition can be found in [5], which was originally given with a finite-dimensional Euclidean space [4].

Definition 4.1 (Bounded Perturbation Resilience)**.** Let H be a real Hilbert space. Given a problem Φ, an algorithmic operator $A_\Phi : H \to H$ is said to be bounded perturbation resilient if the following holds. If the sequence $\{x_n\}_{n=0}^{\infty}$, generated by $x_{n+1} = A_\Phi x_n$, for all $n \in \mathbb{N}$, with $x_0 \in H$, converges to a solution of Φ, then any sequence $\{y_n\}_{n=0}^{\infty}$, starting from given $y_0 \in H$, generated by $y_{n+1} = A_\Phi(y_n + \beta_n v_n)$ also converges to a solution of Φ, where the vector sequence $\{v_n\}_{n=0}^{\infty}$ is bounded, and the scalars $\{\beta_n\}_{n=0}^{\infty}$ are such that $\beta_n \geqslant 0$, $n \in \mathbb{N}$, and $\sum_{n=0}^{\infty} \beta_n < \infty$.

If we treat the PG algorithm (1.5) with $e(x_n) \equiv 0$ $(n \in \mathbb{N})$ as the basic algorithm A_Φ, then the bounded perturbation of (1.5) is a sequence $\{x_n\}_{n=0}^{\infty}$ generated by

$$x_{n+1} = (I + \alpha_n \partial g)^{-1}(I - \alpha_n D\nabla f)(x_n + \beta_n \nu_n), \tag{4.1}$$

where $\{\beta_n\}_{n=0}^{\infty}$, $\{\nu_n\}_{n=0}^{\infty}$ satisfy the conditions in Definition 4.1, respectively. Denote by

$$\theta_n := (D\nabla f)(x_n + \beta_n \nu_n) - \nabla f(x_n + \beta_n \nu_n). \tag{4.2}$$

Theorem 4.2. *Let $f, g \in \Gamma_0(H)$, the assumptions* (A1)-(A4) *and* (A6) *hold with θ_n defined by* (4.2)*, and $\{\beta_n\}_{n=0}^{\infty}$, $\{\nu_n\}_{n=0}^{\infty}$ satisfy the conditions in Definition 4.1, respectively. Then any sequence $\{x_n\}_{n=0}^{\infty}$ generated by algorithm* (4.1) *converges weakly to a solution of* (1.2).

Proof. The idea of verifying the convergence of algorithm (4.1) is to build a relationship between (4.1) and (1.5). It is not hard to see that (4.1) can be rewritten as

$$x_{n+1} = (I + \alpha_n \partial g)^{-1}(x_n - \alpha_n \nabla f(x_n) + \hat{e}(x_n)),$$

where

$$\hat{e}(x_n) = \beta_n \nu_n + \alpha_n[\nabla f(x_n) - (D\nabla f)(x_n + \beta_n \nu_n)] = \beta_n \nu_n + \alpha_n[\nabla f(x_n) - \nabla f(x_n + \beta_n \nu_n) - \theta_n].$$

Consequently,

$$\begin{aligned}\|\hat{e}(x_n)\| &\leqslant \|\beta_n \nu_n\| + \alpha_n[\|\nabla f(x_n) - \nabla f(x_n + \beta_n \nu_n)\| + \|\theta_n\|] \\ &\leqslant \|\beta_n \nu_n\|[1 + \alpha_n L] + \alpha_n \|\theta_n\| \leqslant 3(\sup_n \|\nu_n\|)\beta_n + (\sup_n \alpha_n)\|\theta_n\| < \infty\end{aligned}$$

from the conditions imposed on $\{\nu_n\}_{n=0}^{\infty}$, $\{\beta_n\}_{n=0}^{\infty}$, and (A6). The conclusion follows from Theorem 3.5 if we take $D(x_n) = I$ in algorithm (1.5). $\square$

If we treat the PG algorithm (1.4) as the basic algorithm A_Φ, the bounded perturbation of (1.4) is a sequence $\{x_n\}_{n=0}^{\infty}$ generated by

$$x_{n+1} = (I + \alpha_n \partial g)^{-1}(I - \alpha_n \nabla f)(x_n + \beta_n \nu_n). \tag{4.3}$$

Theorem 4.3. *Let $f, g \in \Gamma_0(H)$, the assumptions* (A1)-(A2) *and* (A4) *hold, and $\{\beta_n\}_{n=0}^{\infty}$, $\{\nu_n\}_{n=0}^{\infty}$ satisfy the conditions in Definition 4.1, respectively. Then any sequence $\{x_n\}_{n=0}^{\infty}$ generated by algorithm* (4.3) *converges weakly to a critical point of* (1.2).

Proof. The conclusion is an obvious result of Theorem 4.2 if we choose $D(x_n + \beta_n \nu_n) = I$ for all $n \in \mathbb{N}$ in algorithm (4.1). $\square$

5. An application to lasso problem

Consider the lasso problem

$$\min_{x \in \mathbb{R}^J} \frac{1}{2}\|Ax - b\|_2^2 + \gamma\|x\|_1, \tag{5.1}$$

where A is an $m \times J$ (real) matrix, $b \in \mathbb{R}^m$, and $\gamma > 0$ is a regularization parameter. We take $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ and $g(x) = \gamma\|x\|_1$. Then $\nabla f(x) = A^T(Ax - b)$, which is Lipschitz continuous with constant $L = \|A\|_2^2$, and

$$(I + \alpha_n \partial g)^{-1}(x_n) = (I + \alpha_n \partial(\gamma\| \cdot \|_1))^{-1}(x_n) = ((1 + \alpha_n \gamma \partial | \cdot |)^{-1}(x_n^1), \cdots, (1 + \alpha_n \gamma \partial | \cdot |)^{-1}(x_n^J))^T,$$

where

$$(1 + \alpha_n \partial \gamma | \cdot |)^{-1}(x_n^k) = \mathrm{sgn}(x_n^k)\max\{|x_n^k| - \alpha_n \gamma, 0\}, \ k = 1, 2, \ \cdots, \ J.$$

We can solve the lasso problem (5.1) by applying Theorem 4.3.

Theorem 5.1. *Let the vector sequence* $\{v_n\}_{n=0}^{\infty}$ *be bounded in* $\mathbb{R}^J$, *and the nonnegative scalars* $\beta_n, n = 0, 1, \cdots$ *be such that* $\sum_{n=0}^{\infty} \beta_n < \infty$. *Then any sequence* $\{x_n\}_{n=0}^{\infty}$ *generated by the algorithm*

$$x_{n+1} = (I + \alpha_n \gamma \partial \| \cdot \|_1)^{-1}[(x_n + \beta_n v_n) - \alpha_n A^T A(x_n + \beta_n v_n) + \alpha_n A^T b)],$$

where α_n *satisfies* $0 < \inf \alpha_n \leqslant \alpha_n \leqslant \sup \alpha_n < \frac{2}{\|A\|_2^2}$, *converges weakly to a solution of* (5.1).

Proof. Observe that $\Phi(x) := \frac{1}{2}\|Ax - b\|_2^2 + \gamma\|x\|_1$ is a continuous convex function, and

$$\lim_{\|x\|_2 \to \infty} \Phi(x) = \infty.$$

Besides, the assumptions (A1)-(A2) and (A4) are satisfied. Consequently, the weak convergence of $\{x_n\}_{n=0}^{\infty}$ follows from Theorem 4.3. $\square$

Acknowledgment

The authors would like to acknowledge the National Natural Science Foundation of China (No. 11402294, 61503385) and Fundamental Research Funds for the Central Universities (No. 3122015L015) for financial support.

References

[1] F. Alvarez, H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Wellposedness in optimization and related topics, Gargnano, (1999), Set-Valued Anal., **9** (2001), 3–11. 2.8

[2] J. B. Baillon, G. Haddad, *Quelques proprits des oprateurs angle-borns et n-cycliquement monotones*, (French) Israel J. Math., **26** (1977), 137–150. 3

[3] H. H. Bauschke, P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, With a foreword by Hédy Attouch, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, New York, (2011). 1, 1, 2.3, 3

[4] Y. Censor, R. Davidi, G. T. Herman, *Perturbation resilience and superiorization of iterative algorithms*, Inverse Problems, **26** (2010), 12 pages. 1, 4

[5] Y. Censor, R. Davidi, G. T. Herman, R. W. Schulte, L. Tetruashvili, *Projected subgradient minimization versus superiorization*, J. Optim. Theory Appl., **160** (2014), 730–747. 4

[6] Y. Censor, A. J. Zaslavski, *Convergence and perturbation resilience of dynamic string-averaging projection methods*, Comput. Optim. Appl., **54** (2013), 65–76. 1

[7] Y. C. Cheng, *On the gradient-projection method for solving the nonsymmetric linear complementarity problem*, J. Optim. Theory Appl., **43** (1984), 527–541. 2.6

[8] R. Davidi, Y. Censor, R. W. Schulte, S. Geneser, L. Xing, *Feasibility-seeking and superiorization algorithms applied to inverse treatment planning in radiation therapy*, Infinite products of operators and their applications, Contemp. Math., Israel Math. Conf. Proc., Amer. Math. Soc., Providence, RI, **636** (2015), 83–92. 1

[9] E. S. Helou, M. V. W. Zibetti, E. X. Miqueles, *Superiorization of incremental optimization algorithms for statistical tomographic image reconstruction*, Inverse Problems, **33** (2017), 26 pages. 1

[10] W.-M. Jin, Y. Censor, M. Jiang, *Bounded perturbation resilience of projected scaled gradient methods*, Comput. Optim. Appl., **63** (2016), 365–392. 1, 1

[11] A. Kaplan, R. Tichatschke, *On inexact generalized proximal methods with a weakened error tolerance criterion*, Optimization, **53** (2004), 3–17. 1

[12] P.-L. Lions, B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., **16** (1979), 964–979. 1

[13] Z.-Q. Luo, P. Tseng, *Error bounds and convergence analysis of feasible descent methods: a general approach*, Degeneracy in optimization problems, Ann. Oper. Res., **46/47** (1993), 157–178. 1

[14] J. J. Moreau, *Proximité et dualité dans un espace hilbertien*, (French) Bull. Soc. Math. France, **93** (1965), 273–299. 2.1

[15] Y. Nesterov, *Gradient methods for minimizing composite objective function*, CORE Discussion Papers, **140** (2007), 125–161. 1

[16] T. Nikazad, R. Davidi, G. T. Herman, *Accelerated perturbation-resilient block-iterative projection methods with application to image reconstruction*, Inverse Problems, **28** (2012), 19 pages. 1, 2.7

[17] G. B. Passty, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, J. Math. Anal. Appl., **72** (1979), 383–390. 1

[18] B. T. Polyak, *Introduction to optimization*, Translated from the Russian. With a foreword by Dimitri P. Bertsekas, Translations Series in Mathematics and Engineering, Optimization Software, Inc., Publications Division, New York, (1987). 2.4

[19] S. A. Santos, R. C. M. Silva, *An inexact and nonmonotone proximal method for smooth unconstrained minimization*, J. Comput. Appl. Math., **269** (2014), 86–100. 1

[20] M. J. Schrapp, G. T. Herman, *Data fusion in X-ray computed tomography using a superiorization approach*, Rev. Sci. Instrum., **85** (2014), 9 pages. 1

[21] H.-K. Xu, *Properties and iterative methods for the lasso and its variants*, Chin. Ann. Math. Ser. B, **35** (2014), 501–518. 1, 1.1, 2.2, 2.5

[22] H.-K. Xu, *Bounded perturbation resilience and superiorization techniques for the projected scaled gradient method*, Inverse Problems, **33** (2017), 19 pages. 1

[23] Y.-H. Yao, N. Shahzad, *Strong convergence of a proximal point algorithm with general errors*, Optim. Lett., **6** (2012), 621–628. 1

[24] A. J. Zaslavski, *Convergence of a proximal point method in the presence of computational errors in Hilbert spaces*, SIAM J. Optim., **20** (2010), 2413–2421. 1