

Article

# Bounded Perturbation Resilience and Superiorization of Proximal Scaled Gradient Algorithm with Multi-Parameters

Yanni Guo \* and Xiaozhi Zhao

College of Science, Civil Aviation University of China, Tianjin 300300, China; 18535478198@163.com

\* Correspondence: ynguo@amss.ac.cn

Received: 17 April 2019; Accepted: 9 June 2019; Published: 12 June 2019



**Abstract:** In this paper, a multi-parameter proximal scaled gradient algorithm with outer perturbations is presented in real Hilbert space. The strong convergence of the generated sequence is proved. The bounded perturbation resilience and the superiorized version of the original algorithm are also discussed. The validity and the comparison with the use or not of superiorization of the proposed algorithms were illustrated by solving the  $l^1 - l^2$  problem.

**Keywords:** strong convergence; proximal scaled gradient algorithm; multi-parameter; superiorization; convex minimization problem

**MSC:** 47J25; 49J53; 49M37; 65J15; 90C25

## 1. Introduction

The superiorization method, which was introduced by Censor in 2010 [1], can solve a broad class of nonlinear constrained optimal problems that result from many practical problems such as computed tomography [2], medical image recovery [3,4], convex feasibility problems [5,6], inverse problems of radiation therapy [7] and so on, which generates an automatic procedure based on the fact that the basic algorithm has the property of bounded perturbation resilience so that it is expected to get lower values of the objective function. In recent years, some researchers have focused on finding more applications of superiorization methodology while some other researchers have investigated the bounded perturbation resilience of algorithms—see, for examples, [8–17].

In this paper, we study the bounded perturbation resilience property and the corresponding superiorization of a proximal scaled gradient algorithm with multi-parameters for solving the following non-smooth composite optimization problem of the form

$$\min_{x \in H} [f(x) + g(x)] =: \min_{x \in H} \Phi(x), \quad (1)$$

where  $H$  is a real Hilbert space endowed with an inner product  $\langle \cdot, \cdot \rangle$  and the induced norm  $\| \cdot \|$ .  $f, g \in \Gamma_0(H)$  with  $\Gamma_0(H)$  defined by

$$\Gamma_0(H) := \{f : H \rightarrow (-\infty, +\infty] \mid f \text{ is proper lower semicontinuous convex}\}.$$

In addition,  $f$  has  $L$ -Lipschitz continuous gradient  $\nabla f$  on  $H$  with  $L > 0$ .

The proximal gradient method is one of the popular iterative methods used for solving problem (1), which has received a lot of attention in the recent past due to its fast theoretical convergence rates and strong practical performance. Given an initial value  $x_0 \in H$ , the proximal gradient method generates the following sequence  $\{x_n\}$ :

$$x_{n+1} = prox_{\gamma g}(I - \gamma \nabla f)(x_n), \quad \forall n \geq 0, \tag{2}$$

where  $\gamma > 0$  is the step size and  $prox_{\gamma g}$  is the proximal operator of  $g$  of order  $\gamma$  (please refer to Definition 2, Section 2). Then, the generated sequence  $\{x_n\}$  converges weakly to a solution of problem (1) if the solution set  $S := Argmin[f(x) + g(x)] \neq \emptyset$  and  $0 < \gamma < \frac{2}{L}$  (see, for instance, [18], Theorem 25.8).

Xu [19] raised the following more general proximal gradient algorithm:

$$x_{n+1} = prox_{\gamma_n g}(I - \gamma_n \nabla f)(x_n). \tag{3}$$

The weak convergence of the generated sequence  $\{x_n\}$  was obtained. If  $\dim H = \infty$ , the strong convergence can not be guaranteed.

The scaled method was proposed by Strand [20] for increasing the rate of convergence of some algorithm. In a finite dimensional space, the selection of scaling matrices depends on the particular problem [21,22]. Jin, Censor and Jiang [13] introduced the following projected scaled gradient (PSG) algorithm:

$$x_{n+1} = P_C(I - \gamma_n D \nabla f)(x_n), \quad \forall n \geq 0, \tag{4}$$

where  $D(x_n)$  is a diagonal scaling matrix for each  $x_n$ ,  $P_C : \mathbb{R}^n \rightarrow C(\subset \mathbb{R}^n)$  is defined as  $P_C(x) = \inf\{\|x - y\|, \forall y \in C\}$  for solving the following convex minimization problems:

$$\text{minimize } J(x) \text{ subject to } x \in C, \tag{5}$$

where  $C \subset \mathbb{R}^n$  is a nonempty, closed and convex set, the objective function  $J : C \rightarrow \mathbb{R}$  is convex. With the assumption that

$$\sum_{n=0}^{\infty} \|\nabla f(x_n) - D(x_n) \nabla f(x_n)\| < \infty \tag{6}$$

and other conditions, the convergence of the PSG method in the presence of bounded perturbations was proved.

Motivated by [13], Guo, Cui and Guo [23] discussed the proximal gradient algorithm with perturbations:

$$x_{n+1} = prox_{\lambda_n g}(I - \lambda_n D \nabla f + e)x_n. \tag{7}$$

They proved that the generated sequence  $\{x_n\}$  converges weakly to the solution of problem (1). After that, Guo and Cui [15] applied the convex combination of contraction operator and proximal gradient operator to obtain the strong convergence of the generated sequence and discussed the bounded perturbation resilience of the exact algorithm.

In this paper, we will study the following proximal scaled gradient algorithm with multi-parameters:

$$x_{n+1} := t_n h(x_n) + \gamma_n x_n + \lambda_n prox_{\alpha_n g}(x_n - \alpha_n D(x_n) \nabla f(x_n)) + e_n, \quad n \geq 0, \tag{8}$$

which is a further generalization of the above algorithms. We will discuss the strong convergence of (8) and the bounded perturbation resilience of its exact algorithm just like the algorithms named above. In addition, we also study the superiorized version of the exact algorithm of (8).

The rest is organized as follows. In the next section, we introduce some basic concepts and lemmas. In Section 3, we discuss the strong convergence results of the exact and non-exact algorithms. In Section 4, we provide two numerical examples for illustrating the performances of the iterations. Finally, we summarize the main points of this paper in Section 5.

## 2. Preliminaries

Let  $H$  be a real Hilbert space endowed by an inner product  $\langle \cdot, \cdot \rangle$  and the induced norm  $\| \cdot \|$ . Let  $\{x_n\}$  be a sequence in  $H$ .  $z \in H$  is said to be a weak cluster point of  $\{x_n\}$  if there exists a subsequence  $\{x_{n_j}\}$  of  $\{x_n\}$  that converges weakly to it. The set of all weak cluster points of  $\{x_n\}$  is denoted by  $\omega_w(x_n)$ . Let  $T : H \rightarrow H$  be a nonlinear operator. Set  $Fix(T) := \{x \in H : Tx = x\}$ .

The following definitions are needed in proving our main results.

**Definition 1.** ([24], Proposition 2.1)

(i)  $T$  is non-expansive if

$$\|Tx - Ty\| \leq \|x - y\|, \quad \forall x, y \in H. \tag{9}$$

(ii)  $T$  is  $L$ -Lipschitz continuous with  $L \geq 0$ , if

$$\|Tx - Ty\| \leq L\|x - y\|, \quad \forall x, y \in H. \tag{10}$$

We call  $T$  a contractive mapping if  $0 \leq L < 1$ .

(iii)  $T$  is firmly non-expansive if

$$\|Tx - Ty\|^2 \leq \|x - y\|^2 - \|(I - T)x - (I - T)y\|^2. \tag{11}$$

(iv)  $T$  is  $\alpha$ -averaged if there exists a non-expansive operator  $S : H \rightarrow H$  and  $\alpha \in (0, 1)$ , such that

$$T = (1 - \alpha)I + \alpha S. \tag{12}$$

In particular, a firmly non-expansive mapping is  $\frac{1}{2}$ -averaged.

(v)  $T$  is  $v$ -inverse strongly monotone ( $v$ -ism) with  $v > 0$ , if

$$\langle Tx - Ty, x - y \rangle \geq v\|Tx - Ty\|^2, \quad \forall x, y \in H. \tag{13}$$

**Definition 2.** ([25], Proximal Operator) Let  $g \in \Gamma_0(H)$ . The proximal operator of  $g$  is defined by

$$prox_g(x) := \arg \min_{y \in H} \left\{ \frac{\|y - x\|^2}{2} + g(y) \right\}, \quad x \in H. \tag{14}$$

The above definition is well defined since  $\frac{\|y - x\|^2}{2} + g(y)$  has only one minimizer on  $H$  for each  $x \in H$  and for given  $g \in \Gamma_0(H)$  (see [18], Proposition 12.15).

The proximal operator of  $g$  of order  $\alpha > 0$  is defined as

$$prox_{\alpha g}(x) := \arg \min_{y \in H} \left\{ \frac{\|y - x\|^2}{2\alpha} + g(y) \right\}, \quad x \in H. \tag{15}$$

The following Lemmas 1–3 describe the properties of proximal operators.

**Lemma 1.** ([19,26], Lemma 2.4, Lemma 3.3) Let  $g \in \Gamma_0(H)$ , and  $\alpha > 0, \mu > 0$ . Then,

$$prox_{\alpha g}(x) = prox_{\mu g} \left( \frac{\mu}{\alpha}x + \left(1 - \frac{\mu}{\alpha}\right)prox_{\alpha g}x \right). \tag{16}$$

Moreover, if  $\alpha < \mu$ , we also have

$$\|x - prox_{\alpha g}(I - \alpha \nabla f)x\| \leq 2\|x - prox_{\mu g}(I - \mu \nabla f)x\|, \quad \forall x \in H. \tag{17}$$

**Lemma 2.** ([18] (Non-expansiveness) Let  $g \in \Gamma_0(H)$  and  $\alpha > 0$ . Then, the proximal operator  $prox_{\alpha g}$  is  $\frac{1}{2}$ -averaged. We obtain the non-expansiveness of the proximal operator

$$\|prox_{\alpha g}(x) - prox_{\alpha g}(y)\| \leq \|x - y\|, \quad \forall x, y \in H. \tag{18}$$

**Lemma 3.** ([19], Propostion 3.2) Let  $f, g \in \Gamma_0(H)$ ,  $f$  be differentiable and  $z \in H$ ,  $\alpha > 0$ . Then,  $z$  is a solution to (1) if and only if  $z$  is the fixed point of the following equation:

$$z = prox_{\alpha g}(I - \alpha \nabla f)z. \tag{19}$$

The following lemmas play important roles in proving the strong convergence result.

**Lemma 4.** ([18], Corollary 4.18) Let  $T : H \rightarrow H$  be a non-expansive mapping with  $Fix(T) \neq \emptyset$ . If  $\{x_n\}$  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 5.** ([27], Lemma 2.5) Assume that  $\{s_n\}$  is a sequence of nonnegative real numbers satisfying

$$s_{n+1} \leq (1 - \gamma_n)s_n + \gamma_n\delta_n + \beta_n, \quad n \geq 0, \tag{20}$$

where  $\{\gamma_n\} \subset [0, 1]$ ,  $\{\delta_n\} \subset \mathbb{R}$  such that

- (i)  $\sum_{n=0}^{\infty} \gamma_n = \infty$ ;
- (ii)  $\limsup_{n \rightarrow \infty} \delta_n \leq 0$ ;
- (iii)  $\sum_{n=0}^{\infty} \beta_n < \infty$ .

Then,  $\lim_{n \rightarrow \infty} s_n = 0$ .

**Lemma 6.** ([28], Lemma 2.4) Let  $x, y \in H$  and  $\alpha, \beta \in \mathbb{R}$ . Then,

- (i)  $\|x + y\|^2 \leq \|x\|^2 + 2\langle y, x + y \rangle$ ;
- (ii)  $\|\alpha x + \beta y\|^2 = \alpha(\alpha + \beta)\|x\|^2 + \beta(\alpha + \beta)\|y\|^2 - \alpha\beta\|x - y\|^2$ ;
- (iii)  $\|\alpha x + (1 - \alpha)y\|^2 = \alpha\|x\|^2 + (1 - \alpha)\|y\|^2 - \alpha(1 - \alpha)\|x - y\|^2$ .

**Lemma 7.** ([23], Proposition 3.3) Let  $f, g \in \Gamma_0(H)$ . For any  $0 < \alpha < \frac{2}{L}$ , where  $L$  is the Lipschitz constant of  $\nabla f$ .  $prox_{\alpha g}(I - \alpha \nabla f)$  is  $\frac{\alpha L + 2}{4}$ -averaged. Hence, it is non-expansive.

### 3. The Convergence Analysis and the Superiorized Version

In this section, we prove that the sequence  $\{v_n\}$  generated by the exact form of (8) converges strongly to a solution of problem (1) at first. Then, we discuss the strong convergence of algorithm (8). Finally, we investigate the bounded perturbation resilience of the exact iteration by viewing it as a special case of algorithm (8). The superiorized version is also presented at the end of this section.

#### 3.1. The Exact Form of Algorithm (8)

Given the errors  $e_n \equiv 0$ ,  $n \geq 0$  in (8), we get the exact version of (8):

$$\begin{aligned} v_{n+1} &= t_n h(v_n) + \gamma_n v_n + \lambda_n prox_{\alpha_n g}(v_n - \alpha_n D(v_n) \nabla f(v_n)) \\ &=: t_n h(v_n) + \gamma_n v_n + \lambda_n prox_{\alpha_n g}(I - \alpha_n D \nabla f)v_n, \end{aligned} \tag{21}$$

where  $\{t_n\}, \{\gamma_n\}, \{\lambda_n\} \subset [0, 1]$  such that  $0 < \inf_n \gamma_n$  and  $t_n + \gamma_n + \lambda_n = 1$  for all  $n \geq 0$ .  $h : H \rightarrow H$  is a  $\rho$ -contraction for some  $\rho \in [0, 1)$ .  $f, g \in \Gamma_0(H)$ ,  $f$  has the Lipschitz continuous gradient  $\nabla f$  with Lipschitz constant  $L > 0$ .  $D(x) : H \rightarrow H$  is a linear bounded operator for each  $x \in H$  with an upper bound  $N_x$  and satisfies

$$\sum_{n=0}^{\infty} \|\nabla f(v_n) - D(v_n) \nabla f(v_n)\| =: \sum_{n=0}^{\infty} \|\theta(v_n)\| < \infty. \tag{22}$$

Provided that  $\{t_n\}$ ,  $\{\lambda_n\}$  and  $\{\alpha_n\}$  satisfy some additional conditions, we get the following strong convergence result of algorithm (21).

**Theorem 1.** Suppose that the solution set  $S$  of (1) is not empty, (22) and the following conditions hold:

- (i)  $0 < \alpha := \inf_n \alpha_n \leq \alpha_n < \frac{2}{L}$  for all  $n$ ;
- (ii)  $\lim_{n \rightarrow \infty} t_n = 0$  and  $\sum_{n=0}^{\infty} t_n = \infty$ ;
- (iii)  $\lim_{n \rightarrow \infty} \frac{t_n}{\lambda_n} = 0$ .

The sequence  $\{v_n\}$  generated by algorithm (21) converges strongly to a point  $z \in S$ , where  $z$  is the unique solution of the following variational inequality problem:

$$\langle (I - h)z, v - z \rangle \geq 0, v \in S. \tag{23}$$

**Proof of Theorem 1.** We will complete the proof by three steps.

**Step 1.**  $\{v_n\}$  is a bounded sequence in  $H$ .

Let  $z \in S$ , then we have  $z = \text{prox}_{\alpha_n g}(I - \alpha_n \nabla f)z$  by Lemma 3. In view of Lemmas 2 and 7, we also get that  $\text{prox}_{\alpha_n g}$  and  $\text{prox}_{\alpha_n g}(I - \alpha_n \nabla f)$  are non-expansive for  $n \geq 0$ . Now, let us calculate

$$\begin{aligned} & \|v_{n+1} - z\| \\ &= \|t_n(h(v_n) - z) + \gamma_n(v_n - z) + \lambda_n(\text{prox}_{\alpha_n g}(I - \alpha_n D\nabla f)v_n - z)\| \\ &\leq t_n\|h(v_n) - h(z) + h(z) - z\| + \gamma_n\|v_n - z\| \\ &\quad + \lambda_n\|\text{prox}_{\alpha_n g}(I - \alpha_n D\nabla f)v_n - \text{prox}_{\alpha_n g}(I - \alpha_n \nabla f)z\| \\ &\leq t_n\rho\|v_n - z\| + t_n\|h(z) - z\| + \gamma_n\|v_n - z\| \\ &\quad + \lambda_n\|\text{prox}_{\alpha_n g}(I - \alpha_n D\nabla f)v_n - \text{prox}_{\alpha_n g}(I - \alpha_n \nabla f)v_n \\ &\quad + \text{prox}_{\alpha_n g}(I - \alpha_n \nabla f)v_n - \text{prox}_{\alpha_n g}(I - \alpha_n \nabla f)z\| \\ &\leq (1 - t_n(1 - \rho))\|v_n - z\| + t_n\|h(z) - z\| + \lambda_n\alpha_n\|\theta(v_n)\| \\ &= (1 - t_n(1 - \rho))\|v_n - z\| + t_n(1 - \rho)\frac{1}{(1 - \rho)}\|h(z) - z\| \\ &\quad + (1 - t_n - \gamma_n)\alpha_n\|\theta(v_n)\| \\ &\leq (1 - t_n(1 - \rho))\|v_n - z\| + t_n(1 - \rho)\frac{1}{(1 - \rho)}\|h(z) - z\| \\ &\quad + (1 - t_n + t_n\rho)\alpha_n\|\theta(v_n)\| \\ &= (1 - t_n(1 - \rho))(\|v_n - z\| + \alpha_n\|\theta(v_n)\|) + t_n(1 - \rho)\frac{1}{(1 - \rho)}\|h(z) - z\| \\ &\leq \max\{\|v_n - z\| + \alpha_n\|\theta(v_n)\|, \frac{\|h(z) - z\|}{1 - \rho}\}. \end{aligned} \tag{24}$$

An induction argument shows that

$$\begin{aligned} \|v_{n+1} - z\| &\leq \max\{\|v_0 - z\| + \sum_{k=0}^n \alpha_k\|\theta(v_k)\|, \frac{\|h(z) - z\|}{1 - \rho}\} \\ &\leq \max\{\|v_0 - z\| + M, \frac{\|h(z) - z\|}{1 - \rho}\}, \end{aligned} \tag{25}$$

where  $M := \sum_{n=0}^{\infty} \alpha_n\|\theta(v_n)\| < \infty$  as  $\{\alpha_n\}$  is bounded and  $\sum_{n=0}^{\infty} \|\theta(v_n)\| < \infty$ . Hence,  $\{v_n\}$  is bounded. Consequently,  $\{h(v_n)\}$  is bounded since  $h$  is a  $\rho$ -contraction.

**Step 2.** There exists a subsequence  $\{v_{n_j}\} \subset \{v_n\}$  such that  $\omega_w(v_{n_j}) \subset S$ .

We denote by

$$\begin{aligned} D_n &:= \text{prox}_{\alpha_n g}(I - \alpha_n D\nabla f), \\ T_n &:= \text{prox}_{\alpha_n g}(I - \alpha_n \nabla f) \end{aligned} \tag{26}$$

for brevity.

Using the notation  $\rho_n := \frac{\lambda_n}{1-t_n}$ , one has  $\rho_n \in (0, 1)$ ,

$$\begin{aligned} u_n &:= \frac{1}{1-t_n}(\gamma_n v_n + \lambda_n \text{prox}_{\alpha_n g}(I - \alpha_n D \nabla f)v_n) \\ &= (1 - \rho_n)v_n + \rho_n D_n v_n, \\ v_{n+1} &:= t_n h(v_n) + (1 - t_n)u_n. \end{aligned} \tag{27}$$

Given  $z \in S$ , we consider by utilizing Lemma 6 (iii)

$$\begin{aligned} &\|v_{n+1} - z\|^2 \\ &= \|t_n(h(v_n) - z) + (1 - t_n)(u_n - z)\|^2 \\ &= \|t_n(h(v_n) - h(z) + h(z) - z) + (1 - t_n)(u_n - z)\|^2 \\ &\leq \|t_n(h(v_n) - h(z)) + (1 - t_n)(u_n - z)\|^2 + 2t_n \langle h(z) - z, v_{n+1} - z \rangle \\ &\leq t_n \rho^2 \|v_n - z\|^2 + (1 - t_n)\|u_n - z\|^2 + 2t_n \langle h(z) - z, v_{n+1} - z \rangle. \end{aligned} \tag{28}$$

Meanwhile, we derive

$$\begin{aligned} &\|u_n - z\|^2 \\ &= \|(1 - \rho_n)v_n + \rho_n D_n v_n - z\|^2 \\ &= \|v_n - z + \rho_n(D_n v_n - v_n)\|^2 \\ &= \|v_n - z\|^2 + \rho_n^2 \|D_n v_n - v_n\|^2 - 2\rho_n \langle v_n - z, v_n - D_n v_n \rangle \\ &= \|v_n - z\|^2 + \rho_n^2 \|D_n v_n - v_n\|^2 - \rho_n(\|v_n - z\|^2 + \|D_n v_n - v_n\|^2 - \|D_n v_n - z\|^2) \\ &= (1 - \rho_n)\|v_n - z\|^2 - \rho_n(1 - \rho_n)\|D_n v_n - v_n\|^2 + \rho_n \|D_n v_n - z\|^2. \end{aligned} \tag{29}$$

Notice that

$$\begin{aligned} &\|D_n v_n - v_n\|^2 \\ &= \|(D_n - T_n)v_n + T_n v_n - v_n\|^2 \\ &\leq \alpha_n^2 \|\theta(v_n)\|^2 + \|T_n v_n - v_n\|^2 + 2\alpha_n \|\theta(v_n)\| \|T_n v_n - v_n\|, \end{aligned} \tag{30}$$

and that

$$\begin{aligned} &\|D_n v_n - z\|^2 \\ &= \|D_n v_n - T_n v_n + T_n v_n - z\|^2 \\ &\leq \alpha_n^2 \|\theta(v_n)\|^2 + \|v_n - z\|^2 + 2\alpha_n \|\theta(v_n)\| \|T_n v_n - z\| \\ &\leq \alpha_n^2 \|\theta(v_n)\|^2 + \|v_n - z\|^2 + 2\alpha_n \|\theta(v_n)\| \|v_n - z\| \end{aligned} \tag{31}$$

since  $T_n$  is non-expansive and  $z = T_n z$  (see Lemma 3). We then obtain by substituting (30) and (31) into (29)

$$\begin{aligned} &\|u_n - z\|^2 \\ &= (1 - \rho_n)\|v_n - z\|^2 - \rho_n(1 - \rho_n)\|D_n v_n - v_n\|^2 + \rho_n \|D_n v_n - z\|^2 \\ &\leq (1 - \rho_n)\|v_n - z\|^2 \\ &\quad - \rho_n(1 - \rho_n) \left( \alpha_n^2 \|\theta(v_n)\|^2 + \|T_n v_n - v_n\|^2 + 2\alpha_n \|\theta(v_n)\| \|T_n v_n - v_n\| \right) \\ &\quad + \rho_n \left( \alpha_n^2 \|\theta(v_n)\|^2 + \|v_n - z\|^2 + 2\alpha_n \|\theta(v_n)\| \|v_n - z\| \right) \\ &= \|v_n - z\|^2 - \rho_n(1 - \rho_n)\|T_n v_n - v_n\|^2 + \rho_n^2 \alpha_n^2 \|\theta(v_n)\|^2 \\ &\quad - 2\alpha_n \rho_n(1 - \rho_n)\|\theta(v_n)\| \|T_n v_n - v_n\| + 2\rho_n \alpha_n \|\theta(v_n)\| \|v_n - z\|. \end{aligned} \tag{32}$$

Combining (28) and (32), we get

$$\begin{aligned}
 & \|v_{n+1} - z\|^2 \\
 & \leq t_n \rho^2 \|v_n - z\|^2 + (1 - t_n) \|u_n - z\|^2 + 2t_n \langle h(z) - z, v_{n+1} - z \rangle \\
 & \leq t_n \rho^2 \|v_n - z\|^2 + (1 - t_n) \left[ \|v_n - z\|^2 - \rho_n (1 - \rho_n) \|T_n v_n - v_n\|^2 + \rho_n^2 \alpha_n^2 \|\theta(v_n)\|^2 \right. \\
 & \quad \left. - 2\alpha_n \rho_n (1 - \rho_n) \|\theta(v_n)\| \|T_n v_n - v_n\| + 2\rho_n \alpha_n \|\theta(v_n)\| \|v_n - z\| \right] \\
 & \quad + 2t_n \langle h(z) - z, v_{n+1} - z \rangle \\
 & = (t_n \rho^2 + (1 - t_n)) \|v_n - z\|^2 - \rho_n (1 - \rho_n) (1 - t_n) \|T_n v_n - v_n\|^2 \\
 & \quad + \rho_n^2 (1 - t_n) \alpha_n^2 \|\theta(v_n)\|^2 - 2\alpha_n \lambda_n (1 - \rho_n) \|\theta(v_n)\| \|T_n v_n - v_n\| \\
 & \quad + 2\lambda_n \alpha_n \|\theta(v_n)\| \|v_n - z\| + 2t_n \langle h(z) - z, v_{n+1} - z \rangle \\
 & \leq (1 - t_n (1 - \rho^2)) \|v_n - z\|^2 - \lambda_n (1 - \rho_n) \|T_n v_n - v_n\|^2 \\
 & \quad + \rho_n^2 (1 - t_n) \alpha_n^2 \|\theta(v_n)\|^2 + 2\lambda_n \alpha_n \|\theta(v_n)\| \|v_n - z\| + 2t_n \langle h(z) - z, v_{n+1} - z \rangle \\
 & = (1 - \bar{t}_n) \|v_n - z\|^2 + \rho_n \lambda_n \alpha_n^2 \|\theta(v_n)\|^2 + 2\lambda_n \alpha_n \|\theta(v_n)\| \|v_n - z\| \\
 & \quad + \bar{t}_n \left[ \frac{2}{1 - \rho^2} \langle h(z) - z, v_{n+1} - z \rangle - \frac{\lambda_n \gamma_n}{t_n (1 - t_n) (1 - \rho^2)} \|T_n v_n - v_n\|^2 \right] \\
 & \leq (1 - \bar{t}_n) \|v_n - z\|^2 + \bar{t}_n \zeta_n + U_n,
 \end{aligned} \tag{33}$$

where  $\bar{t}_n := t_n (1 - \rho^2)$ ,

$$\begin{aligned}
 \zeta_n & := \frac{2}{1 - \rho^2} \langle h(z) - z, v_{n+1} - z \rangle - \frac{\lambda_n \gamma_n}{t_n (1 - t_n) (1 - \rho^2)} \|T_n v_n - v_n\|^2, \\
 U_n & := \rho_n \lambda_n \alpha_n^2 \|\theta(v_n)\|^2 + 2\lambda_n \alpha_n \|\theta(v_n)\| \|v_n - z\|.
 \end{aligned} \tag{34}$$

Obviously, we have

$$|\zeta_n| \leq \frac{1}{1 - \rho^2} \|h(z) - z\| \|v_n - z\| < \infty, \tag{35}$$

which implies that  $\limsup_{n \rightarrow \infty} \zeta_n$  is a finite number. Thus, there exists a subsequence  $\{\zeta_{n_j}\} \subset \{\zeta_n\}$  such that  $\limsup_{n \rightarrow \infty} \zeta_n = \lim_{j \rightarrow \infty} \zeta_{n_j}$ . In addition, without loss of generality, we may assume that  $v_{n_j}$  converges weakly to some  $v^* \in H$  as  $j \rightarrow \infty$  since  $\{v_n\}$  is bounded. Notice that

$$\begin{aligned}
 \|v_{n_j+1} - v_{n_j}\| & = \|t_{n_j} (h(v_{n_j}) - v_{n_j}) + \lambda_{n_j} (D_{n_j} v_{n_j} - v_{n_j})\| \\
 & \leq t_{n_j} \|h(v_{n_j}) - v_{n_j}\| + \lambda_{n_j} \|D_{n_j} v_{n_j} - T_{n_j} v_{n_j} + T_{n_j} v_{n_j} - v_{n_j}\| \\
 & \leq t_{n_j} [\|h(v_{n_j})\| + \|v_{n_j}\|] + \lambda_{n_j} \alpha_{n_j} \|\theta(v_{n_j})\| + \lambda_{n_j} \|T_{n_j} v_{n_j} - v_{n_j}\| \\
 & \rightarrow 0, \text{ as } j \rightarrow \infty.
 \end{aligned} \tag{36}$$

We conclude that  $\{v_{n_j+1}\}$  also converges weakly to  $v^*$ . As a result,  $\lim_{n \rightarrow \infty} \langle h(z) - z, v_{n_j+1} - z \rangle$  exists. Hence, we have

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} \zeta_n & = \lim_{j \rightarrow \infty} \zeta_{n_j} \\
 & = \lim_{j \rightarrow \infty} \left[ \frac{1}{1 - \rho^2} \langle h(z) - z, v_{n_j+1} - z \rangle - \frac{\lambda_{n_j} \gamma_{n_j}}{t_{n_j} (1 - t_{n_j}) (1 - \rho^2)} \|T_{n_j} v_{n_j} - v_{n_j}\|^2 \right] \\
 & = \frac{1}{1 - \rho^2} \langle h(z) - z, v^* - z \rangle - \lim_{j \rightarrow \infty} \frac{\lambda_{n_j} \gamma_{n_j}}{t_{n_j} (1 - t_{n_j}) (1 - \rho^2)} \|T_{n_j} v_{n_j} - v_{n_j}\|^2.
 \end{aligned} \tag{37}$$

In light of the fact that  $\{\frac{\gamma_{n_j}}{1 - t_{n_j}}\}$  is bounded and  $\inf_n \frac{\gamma_{n_j}}{1 - t_{n_j}} > 0$ , the sequence  $\{\frac{\lambda_{n_j}}{t_{n_j}} \|T_{n_j} v_{n_j} - v_{n_j}\|^2\}$  is bounded. Then, condition (iii) implies

$$\lim_{j \rightarrow \infty} \|T_{n_j} v_{n_j} - v_{n_j}\| = 0. \tag{38}$$

Set  $T := \text{prox}_{\alpha g}(I - \alpha \nabla f)$  with  $\alpha = \inf_n \alpha_n > 0$  and apply Lemma 1. We get

$$\lim_{j \rightarrow \infty} \|Tv_{n_j} - v_{n_j}\| \leq 2 \lim_{j \rightarrow \infty} \|T_{n_j}v_{n_j} - v_{n_j}\| = 0. \tag{39}$$

Then, Lemma 4 guarantees that  $\omega_w(v_{n_j}) \subset S$ .

**Step 3.**  $\{v_n\}$  converges strongly to  $z \in S$ .

Let us show that (33) satisfies the conditions of Lemma 5. From Step 2, it has  $v_{n_j+1} \rightarrow v^* (\in S)$  as  $j \rightarrow \infty$ . Therefore,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \zeta_n &\leq \frac{2}{1-\rho^2} \limsup_{n \rightarrow \infty} \langle h(z) - z, v_{n+1} - z \rangle \\ &= \frac{2}{1-\rho^2} \lim_{j \rightarrow \infty} \langle h(z) - z, v_{n_j+1} - z \rangle \\ &= \frac{2}{1-\rho^2} \langle h(z) - z, v^* - z \rangle \\ &\leq 0. \end{aligned} \tag{40}$$

In addition, it is obvious that  $\sum_{n=0}^{\infty} U_n < \infty$  (see (34)) since  $\{\rho_n\}, \{\lambda_n\}, \{\alpha_n\}, \{\|v_n - z\|\}$  are bounded sequences and  $\sum_{n=0}^{\infty} \|\theta(v_n)\| < \infty$ . Finally, we apply Lemma 5 to (33) to conclude that  $\|v_n - z\| \rightarrow 0$ , as  $n \rightarrow \infty$ . This ends the proof.  $\square$

### 3.2. The Strong Convergence of Algorithm (8)

**Theorem 2.** Suppose that the solution set  $S$  of (1) is not empty, (22) and the following conditions hold:

- (i)  $0 < \alpha := \inf_n \alpha_n \leq \alpha_n < \frac{2}{L}$  for all  $n$ ;
- (ii)  $\lim_{n \rightarrow \infty} t_n = 0$  and  $\sum_{n=0}^{\infty} t_n = \infty$ ;
- (iii)  $\lim_{n \rightarrow \infty} \frac{t_n}{\lambda_n} = 0$ ;
- (iv)  $\sum_{n=0}^{\infty} \|e_n\| < \infty$ .

Then, the sequence  $\{x_n\}$  generated by algorithms (8) converges strongly to a point  $z \in S$ .

**Proof of Theorem 2.** Let  $\{x_n\}, \{v_n\}$  be generated by (8) and (21), respectively. Then,  $\{v_n\}$  converges strongly to a solution of problem (1) according to Theorem 1. Thus, we only need to prove that  $\|x_n - v_n\| \rightarrow 0$  as  $n \rightarrow \infty$ .

We denote by  $D_n := \text{prox}_{\alpha_n g}(I - \alpha_n D \nabla f)$  and  $T_n := \text{prox}_{\alpha_n g}(I - \alpha_n \nabla f)$ , respectively.  $T_n$  is non-expansive according to Lemma 7. Then, we have

$$\begin{aligned} &\|x_{n+1} - v_{n+1}\| \\ &= \|t_n(h(x_n) - h(v_n)) + \gamma_n(x_n - v_n) + \lambda_n[\text{prox}_{\alpha_n g}(I - \alpha_n D \nabla f)x_n \\ &\quad - \text{prox}_{\alpha_n g}(I - \alpha_n D \nabla f)v_n] + e_n\| \\ &= \|t_n(h(x_n) - h(v_n)) + \gamma_n(x_n - v_n) + \lambda_n((D_n - T_n)(x_n) - (D_n - T_n)(v_n)) \\ &\quad + T_n(x_n) - T_n(v_n)) + e_n\| \\ &\leq t_n\|h(x_n) - h(v_n)\| + \gamma_n\|x_n - v_n\| + \lambda_n(\|(D_n - T_n)(x_n)\| + \|(D_n - T_n)(v_n)\| \\ &\quad + \|T_n(x_n) - T_n(v_n)\|) + \|e_n\| \\ &\leq t_n\rho\|x_n - v_n\| + \gamma_n\|x_n - v_n\| + \lambda_n[\alpha_n(\|\theta(x_n)\| + \|\theta(v_n)\|) + \|x_n - v_n\|] + \|e_n\| \\ &\leq (1 - t_n(1 - \rho))\|x_n - v_n\| + \lambda_n\alpha_n(\|\theta(x_n)\| + \|\theta(v_n)\|) + \|e_n\|. \end{aligned} \tag{41}$$

Applying Lemma 5 to inequality (41), we get  $\|x_n - v_n\| \rightarrow 0$  as  $n \rightarrow \infty$ . We then have completed the proof owing to Theorem 1.  $\square$

### 3.3. Bounded Perturbation Resilience

This subsection is devoted to verifying the bounded perturbation resilience property of algorithm (21) and showing the superiorized version of it.

Given a problem  $\Psi$ , let  $A : H \rightarrow H$  be a basic algorithm operator.

**Definition 3.** [9] An algorithmic operator  $A$  is bounded perturbation resilient if the sequence  $\{v_n\}$ , generated by  $v_{n+1} = Av_n$  with  $v_0 \in H$ , converges to a solution to  $\Psi$ ; then, any sequence  $\{x_n\}$  generated by  $x_{n+1} = A(x_n + \beta_n y_n)$  with any  $x_0 \in H$ , also converges to a solution of  $\Psi$ , where  $\{y_n\}_{n=0}^\infty \subset H$  is bounded, and  $\{\beta_n\}_{n=0}^\infty \subset \mathbb{R}$  are such that  $\beta_n \geq 0$  for all  $n \geq 0$  and  $\sum_{n=0}^\infty \beta_n < \infty$ .

If we take algorithm (21) as the basic algorithm  $A$ , the following iteration is the bounded perturbation of it:

$$x_{n+1} := t_n h(x_n + \beta_n y_n) + \gamma_n (x_n + \beta_n y_n) + \lambda_n \text{prox}_{\alpha_n g}(I - \alpha_n D \nabla f)(x_n + \beta_n y_n), \quad n \geq 0. \tag{42}$$

We have the following result.

**Theorem 3.** Let  $H$  be a real Hilbert space, and  $h : H \rightarrow H$  a  $\rho$ -contractive operator with  $\rho \in (0, 1)$ ,  $f, g \in \Gamma_0(H)$ . Assume that the solution set  $S$  to (1) is nonempty and that  $f$  has Lipschitz continuous gradient  $\nabla f$  on  $H$  with the Lipschitz constant  $L > 0$ .  $\{\beta_n\}, \{y_n\}$  satisfy the conditions in Definition 3,  $\{t_n\}, \{\gamma_n\}, \{\lambda_n\}$  and  $\{\alpha_n\}$  satisfy the conditions in Theorem 1, respectively. Then, any sequence  $\{x_n\}$  generated by (42) converges strongly to a point  $x^*$  in  $S$ . Thus, algorithm (21) is bounded perturbation resilient.

**Proof of Theorem 3.** We can rewrite algorithm (42) as

$$x_{n+1} = t_n h(x_n) + \gamma_n x_n + \lambda_n \text{prox}_{\alpha_n g}(I - \alpha_n D \nabla f)x_n + \tilde{e}_n \tag{43}$$

with

$$\begin{aligned} \tilde{e}_n := & t_n (h(x_n + \beta_n y_n) - h(x_n)) + \gamma_n \beta_n y_n \\ & + \lambda_n [\text{prox}_{\alpha_n g}(I - \alpha_n D \nabla f)(x_n + \beta_n y_n) - \text{prox}_{\alpha_n g}(I - \alpha_n D \nabla f)x_n], \end{aligned} \tag{44}$$

which is obviously the same form as (8) if we certify that  $\sum_{n=0}^\infty \|\tilde{e}_n\| < \infty$ . In fact, we have

$$\begin{aligned} \|\tilde{e}_n\| &= \|t_n (h(x_n + \beta_n y_n) - h(x_n)) + \gamma_n \beta_n y_n \\ &\quad + \lambda_n [\text{prox}_{\alpha_n g}(I - \alpha_n D \nabla f)(x_n + \beta_n y_n) - \text{prox}_{\alpha_n g}(I - \alpha_n D \nabla f)x_n]\| \\ &= \|t_n (h(x_n + \beta_n y_n) - h(x_n)) + \gamma_n \beta_n y_n \\ &\quad + \lambda_n [(D_n - T_n)(x_n + \beta_n y_n) - (D_n - T_n)x_n + T_n(x_n + \beta_n y_n) - T_n x_n]\| \\ &\leq t_n \rho \beta_n \|y_n\| + \gamma_n \beta_n \|y_n\| \\ &\quad + \lambda_n [\alpha_n (\|\theta(x_n + \beta_n y_n)\| + \|\theta(x_n)\|) + \beta_n \|y_n\|] \\ &= (1 - t_n(1 - \rho))\beta_n \|y_n\| + \lambda_n \alpha_n (\|\theta(x_n + \beta_n y_n)\| + \|\theta(x_n)\|), \end{aligned} \tag{45}$$

where  $D_n$  and  $T_n$  are defined by (26), respectively. Then, it is easy to conclude that

$$\sum_{n=0}^\infty \|\tilde{e}_n\| < \infty \tag{46}$$

since  $\{\beta_n\}, \{\|\theta(x_n + \beta_n y_n)\|\}, \{\|\theta(x_n)\|\}$  are all summable. Hence, Theorem 2 guarantees that any sequence  $\{x_n\}$  generated by (42) converges strongly to a solution of (1). That is to say, algorithm (21) is bounded perturbation resilient.  $\square$

The superiorized version is equipped with an optimization criterion, which is usually a function  $\phi : H \rightarrow \mathbb{R}$ , with the convention that, for  $x \in H$ ,  $\phi(x)$  being smaller is considered superior. To ensure this, it needs a new concept, named nonascending direction for  $\phi$  at  $x$ . A vector  $v \in H$  is called nonascending for  $\phi$  at  $x \in H$  if  $\|v\| \leq 1$ , and there exists a constant  $\delta > 0$  such that, for all  $\lambda \in [0, \delta]$ ,  $\phi(x + \lambda v) \leq \phi(x)$ . Such  $v$  at least exists one, namely, zero vector. The superiorization method then provides us with an automatic way of turning the original iterative algorithm for solving problem (1) into an algorithm, for which the value of the objective function at each iteration is not larger than that under the original iterative algorithm. At the same time, the value of  $\phi$  is smaller than it is under

the original algorithm. Superiorization does this by assuming that there are a summable sequence  $\{\beta_k\}$  of positive real numbers and a bounded vector sequence  $\{v_k\} \subset H$  (Each  $\{v_k\}$  is a nonascending direction for  $\phi$  at some  $x \in H$ , and  $\beta_k v_k$ , together with the original iterative point, generates a new iterative point), and further by depending on a  $I$  steering steps aimed at reducing the values of  $\phi$  at these iterative points. In addition, it makes use of a logical variable called loop. In this paper, we choose the optimization criterion function as the objective function in problem (1). Then, the superiorized version of (21) is as specified below:

#### 4. Numerical Experiments

In this section, we solve the  $l^1 - l^2$  norm problem by two numerical examples to illustrate the performance of the proposed iterations. The concerned algorithms are Algorithm 1 (MPGAS), the bounded perturbation algorithm (42) (MPGAB) and basic algorithm (21) (MPGA). All of these experiments were done on a quad core Intel i7-8550U CPU @1.8 GHz with 16 GB DDR4 memory.

---

**Algorithm 1: Superiorized Version of (21)**

---

```

1: Given  $x_0$ 
2: set  $k = 0, l = -1$ 
3: set  $x_k = x_0$ 
4: set Error=Constant
5: while Error >  $\epsilon$ 
6:   set  $n = 0$ 
7:   set  $x_k^n = x_k$ 
8:   while  $n \leq I$ 
9:     set  $y_n$  to be a nonascending vector for  $\phi$  at  $x_k^n$ 
10:    set loop=true
11:    while loop
12:       $l = l + 1$ 
13:      set  $\beta_n = c^l$ 
14:      set  $z_n = x_k^n + \beta_n y_n$ 
15:      if  $\phi(z_n) \leq \phi(x_k^n)$  and  $\Phi(z_n) \leq \Phi(x_k^n)$ 
16:        set  $n = n + 1$ 
17:        set  $x_k^n = z_n$ 
18:        set loop=false
19:      end if
20:    end while
21:  end while
22: set  $x_{k+1} = t_k * h(x_k^l) + \gamma_k * x_k^l + (1 - t_k - \gamma_k) * prox_{\alpha_k g}(I - \alpha_k D \nabla f)(x_k^l)$ 
23: set Error= $\|x_{k+1} - x_k\|$ 
24: set  $k = k + 1$ .

```

---

##### 4.1. The $l^1 - l^2$ Norm Problem

Let  $b_k, 1 \leq k \leq N$  be an orthogonal basis of  $\mathbb{R}^N, \mu_k, 1 \leq k \leq N$  be strictly positive real numbers, let  $A \in \mathbb{R}^{M \times N} \setminus \{0\}$ , and  $d \in \mathbb{R}^M$ . The  $l^1 - l^2$  problem has the following form:

$$\min_{x \in \mathbb{R}^N} \sum_{k=1}^N \mu_k (|x^T b_k|) + \frac{1}{2} \|Ax - d\|_2^2. \tag{47}$$

In signal recovery problems,  $d$  is the observed signal and the original signal  $x$  is known to have a sparse representation.

We take  $f(x) = \frac{1}{2} \|Ax - b\|_2^2, g(x) = \sum_{k=1}^N \mu_k (|x^T b_k|)$ . Then,  $\nabla f(x) = A^T(Ax - b)$  with Lipschitz constant  $L = \|A^T A\|$ , where  $A^T$  refers to the transpose of  $A$ . The above  $l^1 - l^2$  problem is a special case to problem (1).

In this case, the  $k^{th}$  component of  $prox_{\alpha_n g}(x)$  is

$$[prox_{\alpha_n g}(x)]_k = \begin{cases} x^k + \alpha_n \mu_k, & x^k < -\alpha_n \mu_k, \\ 0, & x^k \in [-\alpha_n \mu_k, \alpha_n \mu_k], \\ x^k - \alpha_n \mu_k, & x^k > \alpha_n \mu_k, \end{cases} \tag{48}$$

where  $x = (x^1, x^2, \dots, x^N)^T \in \mathbb{R}^N$ . Then, given  $v_0, x_0 \in \mathbb{R}^N$ , arbitrarily, sequences  $\{v_n\}$ , generated by (21), and  $\{x_n\}$ , generated by (42), can be rewritten as

$$v_{n+1} = t_n h(v_n) + \gamma_n v_n + \lambda_n prox_{\alpha_n g}(v_n - \alpha_n D(v_n) A^T (A(v_n) - b)); \tag{49}$$

$$x_{n+1} = t_n h(x_n + \beta_n y_n) + \gamma_n (x_n + \beta_n y_n) + \lambda_n prox_{\alpha_n g}(x_n + \beta_n y_n - \alpha_n D(x_n + \beta_n y_n) A^T (A(x_n + \beta_n y_n) - b)), \tag{50}$$

respectively.

#### 4.2. Numerical Examples

**Example 1.** Let  $H = \mathbb{R}^2, \mu_1 = \mu_2 = 1,$

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad d = \begin{pmatrix} 1 \\ 2 \end{pmatrix}. \tag{51}$$

A straightforward calculation shows that the solution set  $S = \{(0, 0.6)^T\}$  to (47) and the minimum value of the objective function for (47) is 1.6. We solve this problem with the algorithms proposed in this paper. The numerical results can be found in Table 1.

Suppose that the contraction  $h(x) = \frac{1}{3}x$ , and the diagonal scaling matrix  $D(x_n) = \text{diag}\{d_{ii}(x_n)\} = \text{diag}\{(1 + \frac{1}{n^2})\}$ . We choose  $t_n = \frac{1}{3n}, \gamma_n = 0.01 + \frac{1}{3k}, \lambda_n = 1 - t_n - \gamma_n,$  and the step size sequence  $\alpha_n = \frac{n}{L(n+1)}$ . For algorithm (21) with bounded perturbations, we choose the bounded sequence  $\{y_n\}$  as

$$y_n = \begin{cases} -\frac{x_n}{\|x_n\|_1}, & \text{if } x_n \neq 0, \\ 0, & \text{if } x_n = 0, \end{cases} \tag{52}$$

the summable nonnegative real sequence  $\{\beta_n\}$  as  $\beta_n = c^n$  for some  $c \in (0, 1)$ . For the superiorized version of (21), we take the function  $\phi$  as the objective function in problem (47), that is

$$\phi(x) = \sum_{k=1}^2 (|x^T b_k|) + \frac{1}{2} \|Ax - d\|_2^2. \tag{53}$$

The iteration numbers (“Iter”), the values of  $x_n$  (“ $x_n$ ”), the values of the objective function (“Obj”) are reported in Table 1 when the stopping criterion

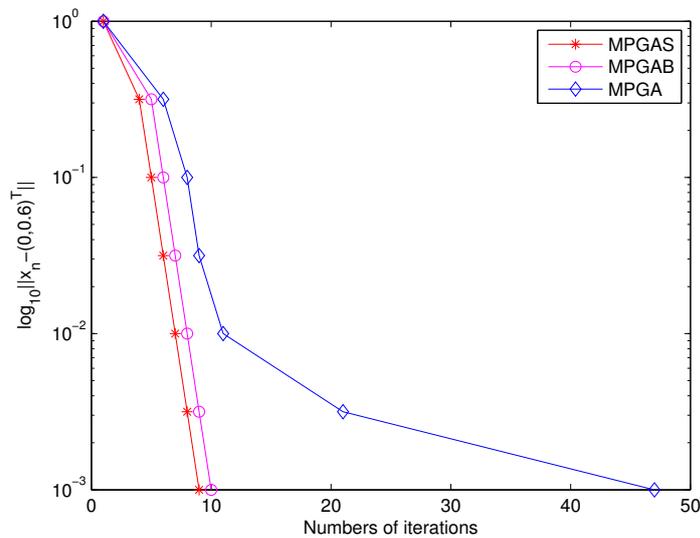
$$Err := \|x_n - (0, 0.6)^T\|_2 < 10^{-3} \tag{54}$$

is reached.

**Table 1.** Results for Example 1.

Methods	$x_0 = (0, 0)^T$			$x_0 = 2 * \text{rand}(2, 1)$		
	Iter	$x_n$	Obj	Iter	$x_n$	Obj
MPGAS	9	$[0.0000, 0.6000]^T$	1.600	9	$[0.0000, 0.6000]^T$	1.600
MPGAB	9	$[0.0000, 0.6000]^T$	1.600	9	$[0.0000, 0.6001]^T$	1.600
MPGA	47	$[0.0000, 0.599044]^T$	1.600	47	$[0.0000, 0.599044]^T$	1.600

The following Figure 1 is for  $x_0 = 2 * rand(2, 1)$ .



**Figure 1.** The number of iterations with Algorithm 1(MPGAS), algorithm (42)(MPGAB) and algorithm (21)(MPGA).

From Table 1 and Figure 1 above, we see that the superiorized version and the bounded perturbation algorithm of (21) arrived at the minimum and the unique minimum point by nine iterations while the original algorithm (21) took 47 iterations to attain the same minimum with the zero initial value. Similar results were also obtained with the initial value  $x_0$  from uniform distribution.

We now discuss a general case of problem (47) by the above-mentioned algorithms.

**Example 2.** Let the system matrix  $A \in \mathbb{R}^{50 \times 200}$  be stimulated by standard Gaussian distribution,  $\mu_k = 1, k = 1, 2, \dots, 200$ . Let the vector  $d \in \mathbb{R}^{50}$  be generated from a uniform distribution in the interval  $[-2, 2]$ . Solve the optimal problem (47) with the above-mentioned algorithms.

We take the parameters in the algorithms as follows:

1. Algorithm parameters:

The contraction  $h(x) = x/3$ . The diagonal scaling matrix

$$D(x_n) = \text{diag}\{d_{ii}(x_n)\} = \text{diag}\left\{1 - \frac{1}{n^2}\right\}. \tag{55}$$

2. Algorithm parameters for the superiorized version:  
 $t_n = \frac{1}{3n}, \gamma_n = 0.01 + \frac{1}{2n}$ , then  $\lambda_n = 1 - t_n - \gamma_n$ . The step size sequence  $\alpha_n = \frac{n}{L(n+1)}$ .

The summable nonnegative real sequence  $\{\beta_n\}$ :  $\beta_n = c^n$  for some  $c \in (0, 1)$  and  $I = 10$ . We set  $\phi$  as the objective function in problem (47).

The iteration numbers “Iter”, the computing time in seconds (“CPU(s)”), the error’s values (“Err”) are reported in Table 2 with a random initial guess  $x_0 = 2 * rand(200, 1)$  when the stopping criterion

$$\text{Err} := \|x_{n+1} - x_n\|_2 < \varepsilon \tag{56}$$

is reached, where  $\varepsilon$  is a given small positive constant.

**Table 2.** Results for Example 2 with  $x_0 = 2 * rand(200, 1)$ .

Methods	$Err < 10^{-4}$			$Err < 10^{-6}$		
	Iter	CPU(s)	Obj	Iter	CPU(s)	Obj
MPGAS	367	4.265625	27.127	1652	3.343750	22.790
MPGAB	1154	5.812500	28.563	2354	7.593750	22.793
MPGA	2084	9.781250	32.386	2354	8.046875	22.793

We find from Table 2 that there is no increase in the execution time of the computer by running the superiorized version, MPGAS, of original algorithm (21). In contrast, compared to the algorithms MPGAB and MPGA, MPGAS even reduces the operation time to get a smaller objective function value under the same stop criterions and initial value  $x_0$ .

## 5. Conclusions

In this paper, we have proposed a proximal scaled gradient algorithm with multi-parameters and studied the strong convergence of it in a real Hilbert space for solving a composite optimization problem. We have investigated the bounded perturbation resilience and the superiorized version of it as well. The validity of the proposed algorithm and the comparison of the original iteration, the bounded perturbation form and the superiorized version of it were illustrated by numerical examples. The results and numerical examples in this paper are a new attempt or application of a newly developed superiorization method. It shows that this method works well to some degree for the proposed algorithm.

**Author Contributions:** All authors contributed equally and significantly to this paper. Conceptualization, Y.G.; Data curation, Y.G. and X.Z.; Formal analysis, Y.G. and X.Z.

**Funding:** This research was funded by the Fundamental Research Funds for the Central Universities (Grant No. 3122018L004) and China Scholarship Council (Grant No. 201807315013).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Censor, Y.; Davidi, R.; Herman, G.T. Perturbation resilience and superiorization of iterative algorithms. *Inverse Probl.* **2010**, *26*, 65008. [[CrossRef](#)] [[PubMed](#)]
2. Davidi, R.; Schulte, R.W.; Censor, Y.; Xing, L. Fast superiorization using a dual perturbation scheme for proton computed tomography. *Trans. Am. Nucl. Soc.* **2012**, *106*, 73–76.
3. Davidi, R.; Herman, G.T.; Censor, Y. Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections. *Int. Trans. Oper. Res.* **2009**, *16*, 505–524. [[CrossRef](#)] [[PubMed](#)]
4. Nikazad, T.; Davidi, R.; Herman, G.T. Accelerated perturbation-resilient block-iterative projection methods with application to image reconstruction. *Inverse Probl.* **2012**, *28*, 035005. [[CrossRef](#)] [[PubMed](#)]
5. Censor, Y.; Chen, W.; Combettes, P.L.; Davidi, R.; Herman, G.T. On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints. *Comput. Optim. Appl.* **2012**, *51*, 1065–1088. [[CrossRef](#)]
6. Censor, Y.; Zaslavski, A.J. Strict Fejér monotonicity by superiorization of feasibility-seeking projection methods. *J. Optim. Theory Appl.* **2015**, *165*, 172–187. [[CrossRef](#)]
7. Davidi, R.; Censor, Y.; Schulte, R.W.; Geneser, S.; Xing, L. Feasibility-seeking and superiorization algorithm applied to inverse treatment planning in radiation therapy. *Contemp. Math.* **2015**, *636*, 83–92.
8. Censor, Y.; Zaslavski, A.J. Convergence and perturbation resilience of dynamic string averaging projection methods. *Comput. Optim. Appl.* **2013**, *54*, 65–76. [[CrossRef](#)]
9. Censor, Y.; Davidi, R.; Herman, G.T.; Schulte, R.W.; Tetruashvili, L. Projected subgradient minimization versus superiorization. *J. Optim. Theory Appl.* **2014**, *160*, 730–747. [[CrossRef](#)]

10. Dong, Q.L.; Lu, Y.Y.; Yang, J. The extragradient algorithm with inertial effects for solving the variational inequality. *Optimization* **2016**, *65*, 2217–2226. [[CrossRef](#)]
11. Garduño, E.; Herman, G. Superiorization of the ML-EM algorithm. *IEEE Trans. Nucl. Sci.* **2014**, *61*, 162–172.
12. He, H.; Xu, H.K. Perturbation resilience and superiorization methodology of averaged mappings. *Inverse Probl.* **2017**, *33*, 040301. [[CrossRef](#)]
13. Jin, W.; Censor, Y.; Jiang, M. Bounded perturbation resilience of projected scaled gradient methods. *J. Comput. Optim. Appl.* **2016**, *63*, 365–392. [[CrossRef](#)]
14. Schrapp, M.J.; Herman, G.T. Data fusion in X-ray computed tomography using a superiorization approach. *Rev. Sci. Instrum.* **2014**, *85*, 055302. [[CrossRef](#)] [[PubMed](#)]
15. Guo, Y.N.; Cui, W. Strong convergence and bounded perturbation resilience of a modified proximal gradient algorithm. *J. Inequal. Appl.* **2018**, *2018*, 103. [[CrossRef](#)] [[PubMed](#)]
16. Zhu, J.H.; Penfold, S. Total variation superiorization in dualenergy CT reconstruction for proton therapy treatment planning. *Inverse Probl.* **2017**, *33*, 044013. [[CrossRef](#)]
17. Zibetti, M.V.W.; Lin, C.A.; Herman, G.T. Total variation superiorized conjugate gradient method for image reconstruction. *Inverse Probl.* **2018**, *34*, 034001. [[CrossRef](#)]
18. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Space*; Dilcher, K., Taylor, K., Eds.; Springer: New York, NY, USA, 2011.
19. Xu, H.K. Properties and Iterative Methods for the Lasso and Its Variants. *Chin. Ann. Math.* **2014**, *35*, 501–518. [[CrossRef](#)]
20. Strand, O.N. Theory and methods related to the singular-function expansion and Landweber iteration for integral equations of the first kind. *SIAM J. Numer. Anal.* **1974**, *11*, 798–825. [[CrossRef](#)]
21. Piana, M.; Bertero, M. Projected Landweber method and preconditioning. *Inverse Probl.* **1997**, *13*, 441–463. [[CrossRef](#)]
22. Neto, E.S.; Helou, D.; Álvaro, R. Convergence results for scaled gradient algorithms in positron emission tomography. *Inverse Probl.* **2005**, *21*, 1905–1914. [[CrossRef](#)]
23. Guo, Y.N.; Cui, W.; Guo, Y.S. Perturbation resilience of proximal gradient algorithm for composite objectives. *J. Nonlinear Sci. Appl.* **2017**, *10*, 5566–5575. [[CrossRef](#)]
24. Xu, H.K. Iterative methods for the split feasibility problem in infinite-dimensional Hilbert space. *Inverse Probl.* **2010**, *26*, 105018. [[CrossRef](#)]
25. Moreau, J.J. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. Fr.* **1965**, *93*, 273–299. [[CrossRef](#)]
26. Marino, G.; Xu, H.K. Convergence of generalized proximal point algorithm. *Commun. Pure Appl. Anal.* **2004**, *3*, 791–808.
27. Xu, H.K. Iterative algorithms for nonlinear operators. *J. Lond. Math. Soc.* **2002**, *66*, 240–256. [[CrossRef](#)]
28. Xu, H.K. Error sensitivity for strongly convergent modifications of the proximal point algorithm. *J. Optim. Theory Appl.* **2015**, *168*, 901–916.

