



# Convergence rate analysis for fixed-point iterations of generalized averaged nonexpansive operators

Yizun Lin and Yuesheng Xu

**Abstract.** We estimate convergence rates for fixed-point iterations of a class of nonlinear operators which are partially motivated by convex optimization problems. We introduce the notion of the generalized averaged nonexpansive (GAN) operator with a positive exponent, and provide convergence rate analysis of the fixed-point iteration of the GAN operator. The proposed generalized averaged nonexpansiveness is weaker than averaged nonexpansiveness while stronger than nonexpansiveness. We show that the fixed-point iteration of a GAN operator with a positive exponent converges to its fixed-point and estimate the local convergence rate (the convergence rate in terms of the distance between consecutive iterates) depending on the range of the exponent. We prove that the fixed-point iteration of a GAN operator with a positive exponent strictly smaller than 1 can achieve an exponential global convergence rate (the convergence rate in terms of the distance between an iterate and the solution). Furthermore, we establish the global convergence rate of the fixed-point iteration of a GAN operator, depending on both the exponent of generalized averaged nonexpansiveness and the exponent of the Hölder regularity, if the GAN operator is also Hölder regular. We then apply the established theory to three types of convex optimization problems that appear often in data science to design fixed-point iterative algorithms for solving these optimization problems and to analyze their convergence properties.

**Mathematics Subject Classification.** 47J26, 47H09, 65K05, 90C25.

**Keywords.** Convex optimization, fixed-point iteration, generalized averaged nonexpansive, convergence rate.

## 1. Introduction

We consider in this paper the convergence rate analysis of fixed-point algorithms. Fixed-point type algorithms have been popular in solving nondifferentiable convex or nonconvex optimization problems such as image processing [20, 30, 35, 37, 38, 50], medical imaging [29, 34, 47, 56], machine learning [18, 32, 33, 43], and compressed sensing [25, 57]. Existing fixed-point type algorithms for optimization include the gradient descent algorithm [10, 48], the proximal point algorithm [13, 46], the proximal gradient algorithm [9, 42], the forward-backward splitting algorithm [19, 54] and the fixed-point proximity algorithm [30, 34, 37, 38].

Traditionally, fixed-point algorithms were often developed by constructing contractive operators or averaged nonexpansive operators [1, 3, 6, 37, 44]. Such constructions bring advantages for fixed-point algorithms, since they make the corresponding algorithmic convergence analysis straightforward and provides robust and monotonic convergence. That is, as the fixed-point iteration proceeds, the distance between the iterate and the true solution is monotonically decreasing. In addition, fixed-point algorithms are comparatively simple and easy to implement. Most optimization problems in real-world applications may be reformulated as fixed-point equations of averaged nonexpansive operators but usually not contractive operators. It is also known [22] that the local convergence rate (the convergence rate in terms of the distance between consecutive iterates) of the fixed-point iteration of an averaged nonexpansive operator is  $o(k^{-\frac{1}{2}})$ , where  $k$  denotes the step of the iteration. However, certain optimization problems cannot be reformulated as fixed-point problems of averaged nonexpansive operators but operators with weaker conditions. For analysis of such fixed-point iterations, the theory of the averaged nonexpansive operator is not applicable. Therefore, there is a need to extend the existing results.

We are interested in understanding the following two issues: Is there a class of operators, satisfying a condition weaker than the averaged nonexpansiveness, whose fixed-point iterations still converge? Is there a subclass of the averaged nonexpansive operators whose fixed-point iterations have convergence rates higher than order  $o(k^{-\frac{1}{2}})$ ? For the first issue, some classes of operators were proposed, such as demicontractive operators [26, 36] and quasi-firmly nonexpansive operators [51, 52]. However, these classes of operators do not ensure the closedness of the composition operation, which makes them not applicable to a large range of real-world optimization problems. In addition, their fixed-point iterations do not have a convergence rate higher than that the averaged nonexpansive operators have.

To address these two issues, we introduce the notion of the generalized averaged nonexpansive (GAN) operator with a positive exponent  $\gamma$ , establish the convergence property of the fixed-point iterations of GAN operators and prove their convergence rates higher than the known result for a range of the exponent  $\gamma$ . Specifically, this notion generalizes the averaged nonexpansive operators in two aspects. First, the generalized averaged nonexpansiveness

with exponent  $\gamma$  of an operator for  $\gamma > 2$  is weaker than the averaged nonexpansiveness which corresponds to  $\gamma = 2$ , but it still guarantees convergence of its fixed-point iterations. Second, the exponent  $\gamma$  allows us to refine the local convergence rates of the resulting fixed-point iterations, leading to a local convergence rate higher than that the averaged nonexpansive operator has.

We organize this paper in six sections. In Sect. 2, we describe fixed-point formulations for three convex optimization models, which motivate the class of operators investigated in this paper. We introduce in Sect. 3 the notion of the GAN operator and study its connection with nonexpansive, averaged nonexpansive and contractive operators. Several basic properties of GAN operators are also provided. Section 4 is devoted to local and global convergence rate analysis of fixed-point iterations of GAN operators. In Sect. 5, we employ the convergence rate results developed in Sect. 4 to analyze the convergence rate of the fixed-point algorithms for three convex optimization models described in Sect. 2. Section 6 offers a conclusion.

## 2. Fixed-point formulations for optimization

Solutions of optimization problems are often formulated as fixed-points of nonlinear operators. Such formulations have great advantages for algorithm development and convergence analysis. We describe in this section fixed-point formulations for convex optimization problems.

By  $\Gamma_0(\mathbb{R}^n)$  we denote the class of all proper lower semicontinuous convex functions from  $\mathbb{R}^n$  to  $\mathbb{R} \cup \{+\infty\}$ . We assume that  $\Psi \in \Gamma_0(\mathbb{R}^n)$  and consider the convex optimization problem

$$\operatorname{argmin}_{x \in \mathbb{R}^n} \Psi(x). \quad (2.1)$$

Throughout this paper, we assume that the objective function  $\Psi$  has at least one minimizer without further mentioning. Solutions of problem (2.1) may be reformulated as fixed-points of certain operators, depending on the smoothness of the objective function  $\Psi$ . To this end, we first recall the notions of the proximity operator and subdifferential of a convex function. Let  $H \in \mathbb{R}^{n \times n}$  be a symmetric positive definite matrix. For  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^n$ , we define the  $H$ -weighted inner product by  $\langle x, y \rangle_H := x^\top H y$  and the corresponding  $H$ -weighted norm by  $\|x\|_H := \langle x, x \rangle_H^{\frac{1}{2}}$ . Then the  $\ell_2$  inner product and  $\ell_2$  norm are given by  $\langle x, y \rangle_2 := \langle x, y \rangle_I$  and  $\|x\|_2 := \|x\|_I$  respectively, where  $I \in \mathbb{R}^{n \times n}$  denotes the identity matrix. For  $\psi \in \Gamma_0(\mathbb{R}^n)$ , the proximity operator of  $\psi$  at  $x \in \mathbb{R}^n$  is defined by

$$\operatorname{prox}_\psi(x) := \operatorname{argmin}_{u \in \mathbb{R}^n} \left\{ \frac{1}{2} \|u - x\|_2^2 + \psi(u) \right\},$$

and the subdifferential of  $\psi$  at  $x \in \mathbb{R}^n$  is defined by

$$\partial\psi(x) := \{y \in \mathbb{R}^n : \psi(z) \geq \psi(x) + \langle y, z - x \rangle_2 \text{ for all } z \in \mathbb{R}^n\}.$$

Below, we list examples of the operators derived from problem (2.1) for different types of objective functions. In the following three cases, we assume

that function  $f \in \Gamma_0(\mathbb{R}^n)$  is differentiable with an  $L$ -Lipschitz continuous gradient with respect to  $\|\cdot\|_2$ . We let  $\mathbb{R}_+$  denote the set of all positive real numbers throughout the paper.

**Case 1.**  $\Psi := f$ . In this case, a minimizer of (2.1) is identified as a fixed-point of operator

$$T_1 := \mathcal{I} - \beta \nabla f, \quad \text{where } \beta \in \mathbb{R}_+. \quad (2.2)$$

We will call  $T_1$  a gradient descent operator. This type of optimization problems has important applications in machine learning (e.g. smoothed SVM, ridge regression) [55] and medical imaging [2, 23].

**Case 2.**  $\Psi := f + g$ , where  $g \in \Gamma_0(\mathbb{R}^n)$  may not be differentiable, but its proximity operator has a closed form or can be easily computed. By using Fermat's rule (Theorem 16.3 of [6]) and a relation between the subdifferential and the proximity operator (Proposition 2.6 of [37]), a minimizer of (2.1) is identified as a fixed-point of operator

$$T_2 := \text{prox}_{\beta g} \circ (\mathcal{I} - \beta \nabla f), \quad \text{where } \beta \in \mathbb{R}_+. \quad (2.3)$$

Obviously,  $T_2 = \text{prox}_{\beta g} \circ T_1$ . Optimization models of this type are raised from machine learning (e.g.  $\ell_1$ -SVM, LASSO regression) [33], compressed sensing [25] and image processing [7, 24].

**Case 3.**  $\Psi = f + g \circ B + h$ , where  $g \in \Gamma_0(\mathbb{R}^m)$ ,  $h \in \Gamma_0(\mathbb{R}^n)$ , and  $B \in \mathbb{R}^{m \times n}$  is a matrix. We also assume that the proximity operators of both  $g$  and  $h$  have closed forms or can be easily computed. Let  $g^*$  denote the conjugate function of  $g$ , that is,

$$g^*(z) := \sup_{y \in \mathbb{R}^m} \{ \langle z, y \rangle_2 - g(y) \}, \quad \text{for } z \in \mathbb{R}^m.$$

By using Fermat's rule, the chain rule of subdifferential, a relation between the subdifferential and the proximity operator, and introducing a dual variable, a minimizer of (2.1) in this case can be identified as a fixed-point of a more complicated operator. Specifically, we let  $v := \begin{pmatrix} x \\ y \end{pmatrix}$ , for  $x \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^m$ , and introduce  $r : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$  by  $r(v) := f(x)$ ,  $\tilde{T} : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^{n+m}$  by  $\tilde{T}(v) := \begin{pmatrix} \text{prox}_{\beta h}(x) \\ \text{prox}_{\eta g^*}(y) \end{pmatrix}$ , where  $\beta$  and  $\eta$  are two positive parameters. Let

$$E := \begin{pmatrix} I_n & -\beta B^\top \\ \eta B & I_m \end{pmatrix}, \quad G := \begin{pmatrix} I_n & -\beta B^\top \\ -\eta B & I_m \end{pmatrix}, \quad W := \begin{pmatrix} \frac{1}{\beta} I_n & -B^\top \\ -B & \frac{1}{\eta} I_m \end{pmatrix}.$$

We then define the operators

$$T_G : u \rightarrow \left\{ v : (u, v) \text{ satisfies that } v = \tilde{T}((E - G)v + Gu) \right\}$$

and

$$T_3 := T_G \circ (\mathcal{I} - W^{-1} \nabla r), \quad \text{where } \beta, \eta \in \mathbb{R}_+. \quad (2.4)$$

It can be verified that if  $v \in \mathbb{R}^{n+m}$  is a fixed-point of  $T_3$ , then the corresponding  $x \in \mathbb{R}^n$  is a minimizer of (2.1). One can refer to [31, 34] for more details of the derivation of operator  $T_3$ . The model in this case has applications

in image processing [15, 17, 49], machine learning [53] and medical imaging [28, 29, 31, 34].

### 3. Generalized averaged nonexpansive operators

In this section, we introduce the notion of the generalized averaged nonexpansive (GAN) operator and study its connection with the nonexpansive, averaged nonexpansive and contractive operators. Several basic properties of GAN operators are also provided.

We first describe the definition of GAN operator. Let  $\mathcal{I}$  denote the identity operator and  $\|\cdot\|$  a norm on  $\mathbb{R}^n$ .

**Definition 3.1.** An operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is said to be generalized averaged nonexpansive if there exist  $\gamma, \mu \in \mathbb{R}_+$  such that

$$\|Tx - Ty\|^\gamma + \mu\|(\mathcal{I} - T)x - (\mathcal{I} - T)y\|^\gamma \leq \|x - y\|^\gamma, \quad \text{for all } x, y \in \mathbb{R}^n. \quad (3.1)$$

Specifically, we say that  $T$  is  $\mu$ -generalized averaged nonexpansive ( $\mu$ -GAN) with exponent  $\gamma$  with respect to  $\|\cdot\|$ .

The norm  $\|\cdot\|$  used in Definition 3.1 can be any norm including the norm induced by an inner product, weighted inner product and the  $\ell_1$  norm. According to Definition 3.1, for  $\mu_1 > \mu_2 > 0$ , if  $T$  is  $\mu_1$ -GAN with exponent  $\gamma \in \mathbb{R}_+$ , then it is also  $\mu_2$ -GAN with exponent  $\gamma$ .

Let  $\text{Fix}(T)$  denote the set of all fixed-points of operator  $T$  and

$$\Lambda := \{T : \mathbb{R}^n \rightarrow \mathbb{R}^n \mid \text{Fix}(T) \neq \emptyset\}.$$

Throughout this paper, we will assume that  $T \in \Lambda$  without further mentioning. It follows from Definition 3.1 that if  $T$  is GAN, then

$$\|Tx - \hat{x}\|^\gamma + \mu\|Tx - x\|^\gamma \leq \|x - \hat{x}\|^\gamma, \quad \text{for all } x \in \mathbb{R}^n, \hat{x} \in \text{Fix}(T). \quad (3.2)$$

We next discuss connections of the GAN operators with the nonexpansive, averaged nonexpansive, firmly nonexpansive and contractive operators. For notational simplicity, throughout the remaining part of this paper, we use  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$  to represent a weighted inner product and the corresponding weighted norm with respect to a symmetric positive definite matrix, respectively, unless there is a need to specify the weight matrix. An operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is called nonexpansive if

$$\|Tx - Ty\| \leq \|x - y\|, \quad \text{for all } x, y \in \mathbb{R}^n,$$

and is called firmly nonexpansive if

$$\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y \rangle, \quad \text{for all } x, y \in \mathbb{R}^n.$$

If there exists a nonexpansive operator  $\mathcal{N} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $\alpha \in (0, 1)$  such that  $T = (1 - \alpha)\mathcal{I} + \alpha\mathcal{N}$ , we say that  $T$  is  $\alpha$ -averaged nonexpansive [3, 27, 45].

If there exists  $\rho \in (0, 1)$  such that

$$\|Tx - Ty\| \leq \rho\|x - y\|, \quad \text{for all } x, y \in \mathbb{R}^n, \quad (3.3)$$

we say that  $T$  is contractive ( $\rho$ -contractive). From Definition 3.1, we can immediately see that GAN operators are nonexpansive.

To see the connection of the generalized averaged nonexpansiveness with the averaged nonexpansiveness, we recall a known result (Proposition 4.35 of [6]).

**Proposition 3.2.** *Let  $\alpha \in (0, 1)$ . Operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $\alpha$ -averaged nonexpansive if and only if*

$$\|Tx - Ty\|^2 + \frac{1 - \alpha}{\alpha} \|(\mathcal{I} - T)x - (\mathcal{I} - T)y\|^2 \leq \|x - y\|^2, \quad \text{for all } x, y \in \mathbb{R}^n.$$

Proposition 3.2 implies that the  $\alpha$ -averaged nonexpansiveness is equivalent to the  $\frac{1-\alpha}{\alpha}$ -generalized averaged nonexpansiveness with exponent 2. In particular, the firm nonexpansiveness is equivalent to the 1-generalized averaged nonexpansiveness with exponent 2, since it is also equivalent to the  $\frac{1}{2}$ -averaged nonexpansiveness (see Remark 4.34 of [6]). We will show later in this section that for any given  $\gamma \in \mathbb{R}_+$ , a contractive operator must be GAN with exponent  $\gamma$ . The generalization from averaged nonexpansiveness to generalized averaged nonexpansiveness will lead to higher order convergence rate for fixed-point algorithm defined by a GAN operator with an exponent smaller than 2. We will discuss this point in a later section.

We now study the relation among the GAN operators with different exponents and the relation among the generalized averaged nonexpansiveness, contractivity and FP-contractivity (which we will define later). We first present a technical lemma whose proof is straightforward and thus is omitted.

**Lemma 3.3.** *Let  $a, b$  and  $c$  be three nonnegative real numbers,  $\gamma, \gamma' \in \mathbb{R}_+$ . If  $\gamma' > \gamma$  and  $a^\gamma + b^\gamma \leq c^\gamma$ , then  $a^{\gamma'} + b^{\gamma'} \leq c^{\gamma'}$ .*

We establish the inclusion relation of the class of GAN operators with different exponents in the following proposition.

**Proposition 3.4.** *Suppose that  $0 < \gamma_1 < \gamma_2$ . If operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is GAN with exponent  $\gamma_1$ , then  $T$  is GAN with exponent  $\gamma_2$ . In particular, if  $T$  is  $\mu$ -GAN with exponent  $\gamma_1$  for  $\mu \geq 1$ , then  $T$  is  $\mu$ -GAN with exponent  $\gamma_2$ .*

*Proof.* Since  $T$  is GAN with exponent  $\gamma_1$ , there exists  $\mu \in \mathbb{R}_+$  such that

$$\|Tx - Ty\|^{\gamma_1} + \mu \|(\mathcal{I} - T)x - (\mathcal{I} - T)y\|^{\gamma_1} \leq \|x - y\|^{\gamma_1}, \quad \text{for all } x, y \in \mathbb{R}^n.$$

Applying Lemma 3.3 with  $a := \|Tx - Ty\|$ ,  $b := \mu^{\frac{1}{\gamma_1}} \|(\mathcal{I} - T)x - (\mathcal{I} - T)y\|$ ,  $c := \|x - y\|$ ,  $\gamma := \gamma_1$  and  $\gamma' := \gamma_2$ , we obtain that

$$\|Tx - Ty\|^{\gamma_2} + \mu^{\frac{\gamma_2}{\gamma_1}} \|(\mathcal{I} - T)x - (\mathcal{I} - T)y\|^{\gamma_2} \leq \|x - y\|^{\gamma_2}, \quad (3.4)$$

which implies that  $T$  is GAN with exponent  $\gamma_2$ . In particular, if  $\mu \geq 1$ , then it follows from (3.4) and the fact  $\frac{\gamma_2}{\gamma_1} \geq 1$  that

$$\|Tx - Ty\|^{\gamma_2} + \mu \|(\mathcal{I} - T)x - (\mathcal{I} - T)y\|^{\gamma_2} \leq \|x - y\|^{\gamma_2},$$

which implies that  $T$  is  $\mu$ -GAN with exponent  $\gamma_2$ .  $\square$

We next show that contractivity implies generalized averaged nonexpansiveness.

**Proposition 3.5.** *If an operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $\rho$ -contractive for some  $\rho \in (0, 1)$ , then it is  $\hat{\rho}$ -GAN with exponent  $\gamma$ , where  $\gamma \in \mathbb{R}_+$  is an arbitrarily fixed number and  $\hat{\rho} := \frac{1-\rho^\gamma}{(1+\rho)^\gamma}$ .*

*Proof.* Since  $T$  is  $\rho$ -contractive, for any fixed  $\gamma \in \mathbb{R}_+$ , we have that

$$\|Tx - Ty\|^\gamma \leq \rho^\gamma \|x - y\|^\gamma, \quad \text{for all } x, y \in \mathbb{R}^n. \quad (3.5)$$

We choose  $\hat{\rho} := \frac{1-\rho^\gamma}{(1+\rho)^\gamma}$  and observe that  $\rho^\gamma = 1 - \hat{\rho}(1+\rho)^\gamma$ . We thus obtain that

$$\rho^\gamma \|x - y\|^\gamma = \|x - y\|^\gamma - \hat{\rho}(\|x - y\| + \rho\|x - y\|)^\gamma.$$

By the definition (3.3) of the contractive operator and the triangle inequality, we find that

$$\|x - y\| + \rho\|x - y\| \geq \|x - y\| + \|Tx - Ty\| \geq \|(\mathcal{I} - T)x - (\mathcal{I} - T)y\|.$$

Substituting this inequality into the right hand side of the above equation and then combining with (3.5), we get that

$$\|Tx - Ty\|^\gamma \leq \|x - y\|^\gamma - \hat{\rho}\|(\mathcal{I} - T)x - (\mathcal{I} - T)y\|^\gamma,$$

which proves the desired result.  $\square$

Proposition 3.5 provides the inclusion of the class of contractive operators in the class of GAN operators with exponent  $\gamma$  for any  $\gamma \in \mathbb{R}_+$ . Moreover, the class of contractive operators is a proper subset of the class of GAN operators (see, Example 3.9 to be presented later).

We next investigate the inclusion relation of the class of FP-contractive operators and the class of GAN operators with exponent  $\gamma \in (0, 1)$ . We now define the FP-contractive operator. For  $T \in \Lambda$ , if there exists  $\rho \in (0, 1)$  such that

$$\|Tx - \hat{x}\| \leq \rho\|x - \hat{x}\|, \quad \text{for all } x \in \mathbb{R}^n \setminus \text{Fix}(T), \hat{x} \in \text{Fix}(T), \quad (3.6)$$

then we say that  $T$  is  $\rho$ -contractive with respect to its fixed-point set (or FP- $\rho$ -contractive). From the definition of the FP-contractivity, contractive operators must be FP-contractive. However, an FP-contractive operator may not be contractive. For example, the identity operator  $\mathcal{I}$  is FP-contractive but not contractive. In addition, the fixed-point of a FP-contractive operator may not be unique.

We need a technical lemma on the monotonicity of the function  $\psi$  defined for  $\gamma \in \mathbb{R}_+$  by

$$\psi(\alpha) := \frac{1 - \alpha^\gamma}{(1 - \alpha)^\gamma}, \quad \alpha \in [0, 1). \quad (3.7)$$

The following lemma can be easily proved by considering the signs of  $\psi'$  and using L'Hospital's Rule.

**Lemma 3.6.** *Let  $\psi : [0, 1) \rightarrow \mathbb{R}$  be defined by (3.7). If  $\gamma < 1$ , then  $\psi$  is strictly decreasing on  $(0, 1)$  and  $\lim_{\alpha \rightarrow 1^-} \psi(\alpha) = 0$ . If  $\gamma > 1$ , then  $\psi$  is strictly increasing on  $(0, 1)$  and  $\lim_{\alpha \rightarrow 1^-} \psi(\alpha) = +\infty$ .*

**Proposition 3.7.** *If  $T \in \Lambda$  is GAN with exponent  $\gamma$  for some  $\gamma \in (0, 1)$ , then it is FP- $\rho$ -contractive for some  $\rho \in (0, 1)$ .*

*Proof.* We prove this proposition by contradiction. Assume to the contrary that  $T$  is not FP- $\rho$ -contractive for any  $\rho \in (0, 1)$ . That is, for any  $\rho \in (0, 1)$ , there exist  $x \in \mathbb{R}^n \setminus \text{Fix}(T)$  and  $\hat{x} \in \text{Fix}(T)$  such that  $\|Tx - \hat{x}\| > \rho\|x - \hat{x}\|$ . We next prove this implies that  $T$  is not GAN with exponent  $\gamma$  for any  $\gamma \in (0, 1)$ , that is, for any  $\gamma \in (0, 1)$  and any  $\mu \in \mathbb{R}_+$ , there exist  $x \in \mathbb{R}^n \setminus \text{Fix}(T)$  and  $\hat{x} \in \text{Fix}(T)$  such that

$$\|Tx - \hat{x}\|^\gamma + \mu\|Tx - x\|^\gamma > \|x - \hat{x}\|^\gamma. \quad (3.8)$$

By Lemma 3.6, for any  $\gamma \in (0, 1)$ ,  $\psi$  defined by (3.7) is continuous and strictly decreasing on  $(0, 1)$ , and  $\lim_{\alpha \rightarrow 1^-} \psi(\alpha) = 0$ . This ensures that for any  $\mu > 0$ , there exists  $\rho_{\mu, \gamma} \in (0, 1)$  such that  $\mu > \psi(\rho_{\mu, \gamma})$ . By the contradiction hypothesis, for this  $\rho_{\mu, \gamma}$ , there exist some  $x \in \mathbb{R}^n \setminus \text{Fix}(T)$  and  $\hat{x} \in \text{Fix}(T)$  such that  $\|Tx - \hat{x}\| > \rho_{\mu, \gamma}\|x - \hat{x}\|$ . We next prove that (3.8) holds.

Since  $x \in \mathbb{R}^n \setminus \text{Fix}(T)$ , we know that  $\|Tx - x\| > 0$ . If  $\|Tx - \hat{x}\| \geq \|x - \hat{x}\|$ , it is clear that (3.8) holds. If  $\|Tx - \hat{x}\| < \|x - \hat{x}\|$ , we choose  $\rho'_{\mu, \gamma} := \|Tx - \hat{x}\|/\|x - \hat{x}\|$ . We then observe that  $\rho'_{\mu, \gamma} \in (\rho_{\mu, \gamma}, 1)$  and satisfies

$$\|Tx - \hat{x}\| = \rho'_{\mu, \gamma}\|x - \hat{x}\|, \quad (3.9)$$

and  $\mu > \psi(\rho'_{\mu, \gamma})$ , due to the decreasing monotonicity of  $\psi$  on  $(0, 1)$ . This inequality together with  $\|Tx - x\| > 0$  and the definition of  $\psi$  implies that

$$\mu\|Tx - x\|^\gamma > \psi(\rho'_{\mu, \gamma})\|Tx - x\|^\gamma = \frac{1 - \rho'^{\gamma}_{\mu, \gamma}}{(1 - \rho'^{\gamma}_{\mu, \gamma})^\gamma} \|Tx - x\|^\gamma.$$

This inequality combined with the triangle inequality  $\|x - \hat{x}\| \leq \|Tx - x\| + \|Tx - \hat{x}\|$  and (3.9) yields that

$$\mu\|Tx - x\|^\gamma > \frac{1 - \rho'^{\gamma}_{\mu, \gamma}}{(1 - \rho'^{\gamma}_{\mu, \gamma})^\gamma} (\|x - \hat{x}\| - \|Tx - \hat{x}\|)^\gamma = (1 - \rho'^{\gamma}_{\mu, \gamma}) \|x - \hat{x}\|^\gamma.$$

Combining the above inequality and (3.9) leads to

$$\|Tx - \hat{x}\|^\gamma + \mu\|Tx - x\|^\gamma > (\rho'_{\mu, \gamma}\|x - \hat{x}\|)^\gamma + (1 - \rho'^{\gamma}_{\mu, \gamma}) \|x - \hat{x}\|^\gamma = \|x - \hat{x}\|^\gamma.$$

This is (3.8), a contradiction to the generalized averaged nonexpansiveness of  $T$  with exponent  $\gamma$  for some  $\gamma \in (0, 1)$ . Therefore, we conclude that  $T$  is FP- $\rho$ -contractive for some  $\rho \in (0, 1)$ .  $\square$

We next demonstrate by a one-dimensional example that the class of contractive operators is a proper subset of the class of GAN operators with exponent 1. To this end, we first establish a technical lemma. We mention here that a one-dimensional operator  $T : \mathbb{R} \rightarrow \mathbb{R}$  is said to be monotonically increasing if  $Tx \geq Ty$  for any  $x, y \in \mathbb{R}$  satisfying that  $x > y$ .

**Lemma 3.8.** *If an operator  $T : \mathbb{R} \rightarrow \mathbb{R}$  is nonexpansive and monotonically increasing, then it is GAN with exponent 1.*



*Proof.* It suffices to prove that for all  $t_1, t_2 \in \mathbb{R}$ ,

$$|T(t_1) - T(t_2)| + |(t_1 - t_2) - (T(t_1) - T(t_2))| \leq |t_1 - t_2|. \quad (3.10)$$

If  $t_1 = t_2$ , (3.10) clearly holds. Without loss of generality, we prove that (3.10) holds for the case  $t_1 > t_2$ . In this case, we know that  $T(t_1) \geq T(t_2)$  since  $T$  is monotonically increasing. Furthermore, the nonexpansiveness of  $T$  implies that  $T(t_1) - T(t_2) \leq t_1 - t_2$ . Therefore,

$$|T(t_1) - T(t_2)| + |(t_1 - t_2) - (T(t_1) - T(t_2))| = |t_1 - t_2|,$$

which completes the proof.  $\square$

*Example 3.9.* Let  $T := \text{prox}_{\lambda|\cdot|}$ , where  $\lambda \in \mathbb{R}_+$ . Then  $T$  is GAN with exponent 1, but it is not GAN with exponent  $\gamma$  for any  $\gamma \in (0, 1)$  and nor contractive.

*Proof.* Note that  $T$  is firmly nonexpansive [21], and hence it is nonexpansive. It follows from Example 2.3 in [37] that

$$\text{prox}_{\lambda|\cdot|}(t) = \begin{cases} t - \lambda, & \text{if } t > \lambda, \\ 0, & \text{if } -\lambda \leq t \leq \lambda, \\ t + \lambda, & \text{if } t < -\lambda, \end{cases} \quad (3.11)$$

which is monotonically increasing. Then we conclude from Lemma 3.8 that  $T$  is GAN with exponent 1.

We next show that  $T$  is not GAN with exponent  $\gamma$  for all  $\gamma \in (0, 1)$ . Suppose that there exists some  $\gamma \in (0, 1)$  such that  $T$  is GAN with exponent  $\gamma$ . Since  $\text{Fix}(T) = \{0\}$ , there exists  $\mu \in \mathbb{R}_+$  such that

$$|Tx - 0|^\gamma + \mu |Tx - x|^\gamma \leq |x - 0|^\gamma, \quad \text{for all } x \in \mathbb{R},$$

that is,

$$\mu |Tx - x|^\gamma \leq |x|^\gamma - |Tx|^\gamma, \quad \text{for all } x \in \mathbb{R}. \quad (3.12)$$

Since  $Tx = x - \lambda$  for  $x > \lambda$ , we have  $|Tx - x|^\gamma = \lambda^\gamma$ . Then (3.12) implies that

$$\mu \leq \lambda^{-\gamma} [x^\gamma - (x - \lambda)^\gamma], \quad \text{for all } x \in (\lambda, +\infty). \quad (3.13)$$

Let  $\varphi(x) := x^\gamma - (x - \lambda)^\gamma$ ,  $x \in (\lambda, +\infty)$ . It is obvious that  $\varphi$  is continuous on  $(\lambda, +\infty)$ . Moreover, by letting  $x = \frac{1}{t}$  and using L'Hospital's rule, we have that

$$\lim_{x \rightarrow \infty} \varphi(x) = \lim_{t \rightarrow 0} \frac{1 - (1 - \lambda t)^\gamma}{t^\gamma} = \lim_{t \rightarrow 0} \frac{\lambda(1 - \lambda t)^{\gamma-1}}{t^{\gamma-1}} = \lambda \lim_{t \rightarrow 0} \left( \frac{1}{t} - \lambda \right)^{\gamma-1} = 0$$

for  $\gamma \in (0, 1)$ . This implies that for any  $\mu > 0$ , there exists sufficiently large  $x \in \mathbb{R}_+$  such that  $\mu > \lambda^{-\gamma} \varphi(x)$ , which contradicts (3.13). Thus  $T$  is not GAN with exponent  $\gamma$ . According to Proposition 3.5, we know that  $T$  is not contractive either.  $\square$

We mention here that the projection operator onto a closed convex set  $E \subset \mathbb{R}^n$  defined by

$$P_{E, \|\cdot\|_2}(x) := \underset{y \in E}{\operatorname{argmin}} \|x - y\|_2, \quad x \in \mathbb{R}^n$$

may not be GAN with exponent 1 with respect to the  $\ell_2$  norm. In addition, neither the proximity operator of the  $\ell_1$  norm nor the proximity operator of the  $\ell_2$  norm is GAN with exponent 1 with respect to the  $\ell_2$  norm, when the dimension is greater than or equal to 2. The  $\ell_1$  norm is defined by  $\|x\|_1 := \sum_{i=1}^n |x_i|$ , for  $x \in \mathbb{R}^n$ .

*Example 3.10.* Let  $E := \{p \in \mathbb{R}^2 : p_1 \in [0, 1], p_2 = 0\}$ , operator  $\mathcal{T}_1 := P_{E, \|\cdot\|_2}$ , and define operator  $\mathcal{T}_2 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  by  $\mathcal{T}_2 := \text{prox}_{\lambda \|\cdot\|_1}$  and  $\mathcal{T}_3 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  by  $\mathcal{T}_3 := \text{prox}_{\lambda \|\cdot\|_2}$ , where  $\lambda \in \mathbb{R}_+$ . Then none of  $\mathcal{T}_1$ ,  $\mathcal{T}_2$  and  $\mathcal{T}_3$  is GAN with exponent 1 with respect to  $\|\cdot\|_2$ .

*Proof.* By the definition of generalized averaged nonexpansiveness with exponent 1, to prove the desired result, it suffices to show that for any  $\mu \in \mathbb{R}_+$ , there exist  $x^{(i)}$  and  $y^{(i)}$  such that  $\|(\mathcal{I} - \mathcal{T}_i)x^{(i)} - (\mathcal{I} - \mathcal{T}_i)y^{(i)}\|_2 \neq 0$  and

$$p_i < \mu, \quad (3.14)$$

where

$$p_i := \frac{\|x^{(i)} - y^{(i)}\|_2 - \|\mathcal{T}_i x^{(i)} - \mathcal{T}_i y^{(i)}\|_2}{\|(\mathcal{I} - \mathcal{T}_i)x^{(i)} - (\mathcal{I} - \mathcal{T}_i)y^{(i)}\|_2}, \quad i = 1, 2, 3. \quad (3.15)$$

We first consider the case  $i = 1$ . For any  $\mu \in \mathbb{R}_+$ , it is clear that

$$\frac{\sqrt{4\mu^2 + 1} - 1}{2\mu} < \mu. \quad (3.16)$$

Let  $x^{(1)} = (0, 0)^\top$  and  $y^{(1)} = (1, 2\mu)^\top$ . Substituting  $x^{(1)}$  and  $y^{(1)}$  into the definition of  $p_1$  in (3.15) and then using (3.16) yield that  $p_1 < \mu$ .

We next consider the case  $i = 2$ . We know from Example 2.4 of [37] that

$$\mathcal{T}_2 x = (\text{prox}_{\lambda|\cdot|}(x_1), \text{prox}_{\lambda|\cdot|}(x_2))^\top, \quad x \in \mathbb{R}^2,$$

where the closed form of  $\text{prox}_{\lambda|\cdot|}$  is given by (3.11). For any  $\mu \in \mathbb{R}_+$ , we let  $x^{(2)} = (\frac{\lambda}{2}, \lambda)^\top$  and  $y^{(2)} = (\lambda, \lambda(\frac{1}{4\mu} + 1))^\top$ . Substituting  $x^{(2)}$  and  $y^{(2)}$  into the definition of  $p_2$  in (3.15) and then using (3.16) yield that

$$p_2 = \frac{\sqrt{\frac{1}{4}\lambda^2 + \frac{\lambda^2}{16\mu^2}} - \frac{\lambda}{4\mu}}{\frac{1}{2}\lambda} = \frac{\sqrt{4\mu^2 + 1} - 1}{2\mu} < \mu.$$

Last, we consider the case  $i = 3$ . From Example 2.5 of [37], we know that

$$\mathcal{T}_3(x) = \text{prox}_{\lambda|\cdot|}(\|x\|_2) \frac{x}{\|x\|_2}, \quad x \in \mathbb{R}^2.$$

For any  $\mu \in \mathbb{R}_+$ , we let  $r := \frac{\lambda}{2} \min\{\mu^2, 1\}$ ,  $x^{(3)} = (\sqrt{\lambda^2 - (\lambda - r)^2}, \lambda - r)^\top$  and  $y^{(3)} = (0, 2\lambda)^\top$ . Then  $\|x^{(3)}\|_2 = \lambda$  and  $\|y^{(3)}\|_2 = 2\lambda$ , and hence  $\mathcal{T}_3 x^{(3)} = (0, 0)^\top$ ,  $\mathcal{T}_3 y^{(3)} = (0, \lambda)^\top$ . Substituting  $x^{(3)}$  and  $y^{(3)}$  into the definition of  $p_3$

in (3.15) and simplifying, we get that  $p_3 = \frac{\sqrt{\lambda^2 + 4\lambda r} - \lambda}{\sqrt{2\lambda r}}$ . Then rationalizing the numerator and using the fact  $\sqrt{\lambda^2 + 4\lambda r} > \lambda$  give that

$$p_3 = \frac{4\lambda r}{\sqrt{2\lambda r}(\sqrt{\lambda^2 + 4\lambda r} + \lambda)} < \sqrt{\frac{2}{\lambda}} \sqrt{r}. \quad (3.17)$$

By the definition of  $r$ , we know that  $\sqrt{r} \leq \sqrt{\frac{\lambda}{2}} \mu$ , which together with (3.17) yields that  $p_3 < \mu$ .

We conclude that (3.14) holds for  $i = 1, 2, 3$ , which implies the desired result.  $\square$

We next provide a theorem showing that there exists a class of GAN operators with exponent 1 for a high-dimensional case. An example satisfies this theorem will be given later in Corollary 5.6.

**Theorem 3.11.** *Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a firmly nonexpansive operator. If there exists  $\alpha \in (0, 1]$  such that*

$$\|Tx - Ty\| \geq \alpha \|x - y\|, \quad \text{for all } x, y \in \mathbb{R}^n, \quad (3.18)$$

*then for  $\beta \in (0, 2)$ ,  $\mathcal{I} - \beta T$  is GAN with exponent 1.*

*Proof.* To prove the desired result, it suffices to show that there exists  $\mu \in \mathbb{R}_+$  such that for all  $x, y \in \mathbb{R}^n$ ,

$$\|(x - y) - \beta(Tx - Ty)\| + \mu \|\beta(Tx - Ty)\| \leq \|x - y\|. \quad (3.19)$$

Let  $w := x - y$  and  $v := Tx - Ty$ . Then (3.19) can be rewritten by

$$\|w - \beta v\| \leq \|w\| - \mu \|\beta v\|. \quad (3.20)$$

To prove (3.20), it suffices to show that

$$\mu \|\beta v\| \leq \|w\| \quad \text{and} \quad \|w - \beta v\|^2 \leq (\|w\| - \mu \|\beta v\|)^2, \quad (3.21)$$

where the second inequality is equivalent to

$$(1 - \mu^2)\beta \|v\|^2 + 2\mu \|w\| \|v\| \leq 2\langle w, v \rangle. \quad (3.22)$$

To prove the first inequality in (3.21) and inequality (3.22), we let

$$\mu := \min \left\{ \frac{1}{\beta}, \alpha \left( 1 - \frac{\beta}{2} \right) \right\}.$$

Then  $\mu \in (0, 1)$ . The firm nonexpansiveness of  $T$  implies that it is nonexpansive. Since  $\mu \leq \frac{1}{\beta}$ , the first inequality in (3.21) follows from the nonexpansiveness of  $T$  immediately. We next show the validity of (3.22). To this end, we show that

$$\frac{2\mu}{2 - (1 - \mu^2)\beta} \in (0, \alpha). \quad (3.23)$$

Since  $\mu \in (0, 1)$  and  $\beta \in (0, 2)$ , it is obvious that  $2 - (1 - \mu^2)\beta > 2 - \beta > 0$ , which together with the fact  $\mu \leq \alpha \left( 1 - \frac{\beta}{2} \right)$  yields that  $\frac{2\mu}{2 - (1 - \mu^2)\beta} < \frac{2\mu}{2 - \beta} \leq \alpha$ . Hence (3.23) holds. Combining (3.23) and (3.18), we obtain that

$$\|v\| \geq \frac{2\mu}{2 - (1 - \mu^2)\beta} \|w\|.$$

Multiplying  $[2 - (1 - \mu^2)\beta]\|v\|$  on both sides of the above inequality leads to

$$(1 - \mu^2)\beta\|v\|^2 + 2\mu\|w\|\|v\| \leq 2\|v\|^2. \quad (3.24)$$

The firm nonexpansiveness of  $T$  gives that  $\|v\|^2 \leq \langle w, v \rangle$ , which together with (3.24) implies (3.22). This completes the proof.  $\square$

Note that the identity operator is a trivial GAN operator with exponent  $\gamma$  for any  $\gamma \in \mathbb{R}_+$ . In the following proposition, we identify ranges of  $\mu$  for the non-triviality of GAN operators for different ranges of  $\gamma$ . To simplify the notation, throughout the remaining part of this paper, we define

$$\Omega_\mu^\gamma := \{T \in \Lambda : T \text{ is } \mu\text{-GAN with exponent } \gamma\}. \quad (3.25)$$

**Proposition 3.12.** *Let  $\gamma, \mu \in \mathbb{R}_+$ .*

(i) *For any  $\gamma \leq 1$ ,*

$$\Omega_\mu^\gamma \setminus \{\mathcal{I}\} \neq \emptyset \text{ if and only if } \mu \leq 1.$$

(ii) *For any  $\gamma > 1$  and  $\mu \in \mathbb{R}_+$ ,  $\Omega_\mu^\gamma \setminus \{\mathcal{I}\} \neq \emptyset$ .*

*Proof.* We first establish (i). Suppose that  $\mu \leq 1$  and show that  $\Omega_\mu^\gamma \setminus \{\mathcal{I}\} \neq \emptyset$ . It suffices to find some operator  $T \neq \mathcal{I}$  such that  $T \in \Omega_\mu^\gamma$  for any  $\gamma \in (0, 1]$ . To this end, we define  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  by  $T(x) := z$  for all  $x \in \mathbb{R}^n$ , where  $z \in \mathbb{R}^n$  is a constant vector. It is obvious that  $\text{Fix}(T) = \{z\}$ . Since  $\mu \leq 1$ , for any  $\gamma \in \mathbb{R}_+$  and for all  $x, y \in \mathbb{R}^n$ , we have that

$$\|Tx - Ty\|^\gamma + \mu\|(\mathcal{I} - T)x - (\mathcal{I} - T)y\|^\gamma = \mu\|x - y\|^\gamma \leq \|x - y\|^\gamma.$$

Hence  $T \in \Omega_\mu^\gamma \setminus \{\mathcal{I}\}$  for any  $\gamma \in (0, 1]$ .

Conversely, for any  $\gamma \in (0, 1]$ , if  $\Omega_\mu^\gamma \setminus \{\mathcal{I}\} \neq \emptyset$ , then there exists  $T \in \Omega_\mu^\gamma$  such that for some  $x \in \mathbb{R}^n$ ,  $Tx \neq x$ . Since  $T \in \Omega_\mu^\gamma$ , for any given  $\hat{x} \in \text{Fix}(T)$ , we have that

$$\|Tx - \hat{x}\|^\gamma + \mu\|Tx - x\|^\gamma \leq \|x - \hat{x}\|^\gamma. \quad (3.26)$$

We next prove that the validity of (3.26) implies  $\mu \leq 1$ . By (3.26) and the fact that  $\|Tx - x\| > 0$ , we know that  $\|Tx - \hat{x}\| < \|x - \hat{x}\|$ . Let  $\alpha := \frac{\|Tx - \hat{x}\|}{\|x - \hat{x}\|}$ . Then  $\alpha \in [0, 1)$  and

$$\|Tx - \hat{x}\| = \alpha\|x - \hat{x}\|. \quad (3.27)$$

Hence

$$\|x - \hat{x}\|^\gamma - \|Tx - \hat{x}\|^\gamma = (1 - \alpha^\gamma)\|x - \hat{x}\|^\gamma \quad (3.28)$$

and, by the triangle inequality,

$$\|Tx - x\|^\gamma \geq (\|x - \hat{x}\| - \|Tx - \hat{x}\|)^\gamma = (1 - \alpha)^\gamma\|x - \hat{x}\|^\gamma. \quad (3.29)$$

By combining (3.26), (3.28) and (3.29), we obtain that

$$\mu \leq \frac{\|x - \hat{x}\|^\gamma - \|Tx - \hat{x}\|^\gamma}{\|Tx - x\|^\gamma} \leq \frac{1 - \alpha^\gamma}{(1 - \alpha)^\gamma} = \psi(\alpha), \quad \alpha \in [0, 1), \quad (3.30)$$

where  $\psi$  is defined by (3.7). We know from Lemma 3.6 that for any  $\gamma \in (0, 1)$ ,

$$\max_{\alpha \in [0, 1)} \{\psi(\alpha)\} = 1. \quad (3.31)$$

Obviously, (3.31) also holds for  $\gamma = 1$ . That is, (3.31) holds for any  $\gamma \in (0, 1]$ , which together with (3.30) implies that  $\mu \leq 1$ .

Now we prove (ii). For any  $\gamma > 1$ , if  $\mu \leq 1$ , we have shown in (i) that there exists a constant operator  $T$  such that  $T \in \Omega_\mu^\gamma$ . If  $\mu > 1$ , we let  $\psi$  be defined by (3.7). Since  $\gamma > 1$ , by Lemma 3.6, there exists  $\alpha \in (0, 1)$  such that  $\psi(\alpha) \geq \mu$ , that is,

$$\alpha^\gamma + \mu(1 - \alpha)^\gamma \leq 1. \quad (3.32)$$

We next verify that  $\alpha\mathcal{I} \in \Omega_\mu^\gamma$ . By using (3.32), for all  $x, y \in \mathbb{R}^n$ , we have that

$$\begin{aligned} \|\alpha x - \alpha y\|^\gamma + \mu\|(1 - \alpha)x - (1 - \alpha)y\|^\gamma &= (\alpha^\gamma + \mu(1 - \alpha)^\gamma) \|x - y\|^\gamma \\ &\leq \|x - y\|^\gamma. \end{aligned}$$

Thus,  $\alpha\mathcal{I} \in \Omega_\mu^\gamma \setminus \{\mathcal{I}\}$ .  $\square$

We know that the classes of contractive operators, nonexpansive operators and averaged nonexpansive operators are closed under the composition operation as well as the convex combination (see [5, 6, 8, 14, 16]). These properties are important for their applications in convex optimization. In the next proposition, we prove the closedness of the class of GAN operators with exponent  $\gamma \geq 1$  under the composition operation as well as the convex combination. To establish these results, we recall Example 8.23 of [6] as a lemma.

**Lemma 3.13.** *If  $\gamma \in [1, +\infty)$ , then  $\|\cdot\|^\gamma$  is convex and*

$$\|x + y\|^\gamma \leq 2^{\gamma-1} (\|x\|^\gamma + \|y\|^\gamma), \quad \text{for all } x, y \in \mathbb{R}^n.$$

**Proposition 3.14.** *Let  $\gamma \in [1, +\infty)$  and  $\mu_1, \mu_2 \in \mathbb{R}_+$ ,  $\alpha \in (0, 1)$ . If  $\mathcal{T}_1 : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $\mu_1$ -GAN with exponent  $\gamma$  and  $\mathcal{T}_2 : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $\mu_2$ -GAN with exponent  $\gamma$ , then the following statements hold:*

- (i)  $\mathcal{T}_1 \circ \mathcal{T}_2$  is  $\mu$ -GAN with exponent  $\gamma$ , where  $\mu := 2^{1-\gamma} \min\{\mu_1, \mu_2\}$ .
- (ii)  $(1 - \alpha)\mathcal{T}_1 + \alpha\mathcal{T}_2$  is  $\mu'$ -GAN with exponent  $\gamma$ , where  $\mu' := \min\{\mu_1, \mu_2\}$ .

*Proof.* We first prove item (i). For any  $x, y \in \mathbb{R}^n$ , we define

$$p := (\mathcal{I} - \mathcal{T}_2)x - (\mathcal{I} - \mathcal{T}_2)y, \quad q := (\mathcal{I} - \mathcal{T}_1)(\mathcal{T}_2x) - (\mathcal{I} - \mathcal{T}_1)(\mathcal{T}_2y).$$

Then

$$p + q = (\mathcal{I} - \mathcal{T}_1 \circ \mathcal{T}_2)x - (\mathcal{I} - \mathcal{T}_1 \circ \mathcal{T}_2)y. \quad (3.33)$$

From Lemma 3.13, we know that

$$\|p + q\|^\gamma \leq 2^{\gamma-1} (\|p\|^\gamma + \|q\|^\gamma). \quad (3.34)$$

Let  $\mu := 2^{1-\gamma} \min\{\mu_1, \mu_2\}$ . It follows from (3.33) and (3.34) that

$$\mu\|(\mathcal{I} - \mathcal{T}_1 \circ \mathcal{T}_2)x - (\mathcal{I} - \mathcal{T}_1 \circ \mathcal{T}_2)y\|^\gamma \leq \mu_2\|p\|^\gamma + \mu_1\|q\|^\gamma. \quad (3.35)$$

The  $\mu_1$ -generalized averaged nonexpansiveness of  $\mathcal{T}_1$  and  $\mu_2$ -generalized averaged nonexpansiveness of  $\mathcal{T}_2$  (with exponent  $\gamma$ ) give that

$$\mu_1\|q\|^\gamma \leq \|\mathcal{T}_2x - \mathcal{T}_2y\|^\gamma - \|(\mathcal{T}_1 \circ \mathcal{T}_2)x - (\mathcal{T}_1 \circ \mathcal{T}_2)y\|^\gamma \quad (3.36)$$

and

$$\mu_2 \|p\|^\gamma \leq \|x - y\|^\gamma - \|\mathcal{T}_2 x - \mathcal{T}_2 y\|^\gamma, \quad (3.37)$$

respectively. Adding inequalities (3.36) and (3.37) together yields

$$\mu_2 \|p\|^\gamma + \mu_1 \|q\|^\gamma \leq \|x - y\|^\gamma - \|(\mathcal{T}_1 \circ \mathcal{T}_2)x - (\mathcal{T}_1 \circ \mathcal{T}_2)y\|^\gamma. \quad (3.38)$$

Now combining (3.35) and (3.38), we conclude that

$$\mu \|(\mathcal{I} - \mathcal{T}_1 \circ \mathcal{T}_2)x - (\mathcal{I} - \mathcal{T}_1 \circ \mathcal{T}_2)y\|^\gamma \leq \|x - y\|^\gamma - \|(\mathcal{T}_1 \circ \mathcal{T}_2)x - (\mathcal{T}_1 \circ \mathcal{T}_2)y\|^\gamma,$$

that is,  $\mathcal{T}_1 \circ \mathcal{T}_2$  is  $\mu$ -GAN with exponent  $\gamma$ , which proves item (i).

We next prove item (ii). For  $\alpha \in (0, 1)$  and  $x, y \in \mathbb{R}^n$ , we define  $\mathcal{T} := (1 - \alpha)\mathcal{T}_1 + \alpha\mathcal{T}_2$ ,  $w := x - y$ ,  $v := \mathcal{T}x - \mathcal{T}y$ ,  $v^{(1)} := \mathcal{T}_1 x - \mathcal{T}_1 y$  and  $v^{(2)} := \mathcal{T}_2 x - \mathcal{T}_2 y$ . To prove item (ii), it suffices to show that

$$\|v\|^\gamma + \mu' \|w - v\|^\gamma \leq \|w\|^\gamma, \quad (3.39)$$

where  $\mu' := \min\{\mu_1, \mu_2\}$ . It is easy to see that  $w - v = (1 - \alpha)(w - v^{(1)}) + \alpha(w - v^{(2)})$ , which together with the convexity of  $\|\cdot\|^\gamma$  (see Lemma 3.13) and the definition of  $\mu'$  yields that

$$\mu' \|w - v\|^\gamma \leq (1 - \alpha)\mu_1 \|w - v^{(1)}\|^\gamma + \alpha\mu_2 \|w - v^{(2)}\|^\gamma. \quad (3.40)$$

The generalized averaged nonexpansiveness of  $\mathcal{T}_1$  and  $\mathcal{T}_2$  yields that

$$\mu_1 \|w - v^{(1)}\|^\gamma \leq \|w\|^\gamma - \|v^{(1)}\|^\gamma \quad \text{and} \quad \mu_2 \|w - v^{(2)}\|^\gamma \leq \|w\|^\gamma - \|v^{(2)}\|^\gamma. \quad (3.41)$$

Combining (3.40) and (3.41), then using the convexity of  $\|\cdot\|^\gamma$ , we obtain that

$$\begin{aligned} \mu' \|w - v\|^\gamma &\leq \|w\|^\gamma - \left[ (1 - \alpha)\|v^{(1)}\|^\gamma + \alpha\|v^{(2)}\|^\gamma \right] \\ &\leq \|w\|^\gamma - \|(1 - \alpha)v^{(1)} + \alpha v^{(2)}\|^\gamma = \|w\|^\gamma - \|v\|^\gamma, \end{aligned}$$

that is, (3.39) holds. This completes the proof.  $\square$

Before closing this section, we illustrate certain geometric properties of nonexpansive, firmly nonexpansive, averaged nonexpansive and contractive operators, and the proposed GAN operators with different exponents. Such geometric properties are useful in guiding us for the convergence analysis of the Picard sequences of these operators. For  $x \in \mathbb{R}^n$ ,  $r \in \mathbb{R}_+$ , we define the ball with center  $x$  and radius  $r$  by

$$B(x, r) := \{y \in \mathbb{R}^n \mid \|y - x\| \leq r\}.$$

Let  $T \in \Lambda$  and  $\hat{x}$  be an arbitrary fixed-point of  $T$ . According to the definitions of nonexpansive, contractive, firmly nonexpansive and averaged nonexpansive operators, for any given  $x \in \mathbb{R}^n$ , ranges of  $Tx$  can be illustrated by balls with distinct centers and radii if  $T$  is one of these types of operators. We show the ranges of  $Tx$  for different types of  $T$  in Fig. 1. The case that  $T$  is GAN with exponent 3 is also included.

Figure 1 illustrates that the nonexpansiveness of operator  $T$  is not sufficient to guarantee convergence of its fixed-point iteration. For example, when  $T$  is nonexpansive without additional hypothesis, for a given point  $x$ ,

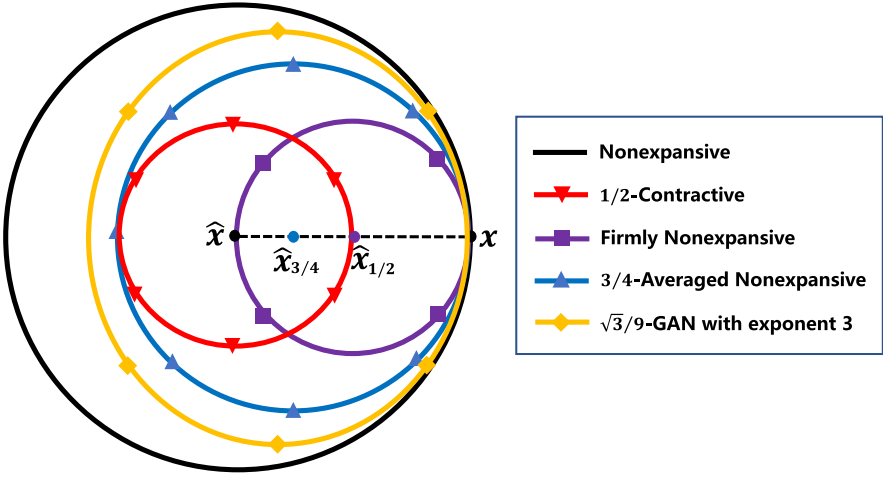


FIGURE 1. The range of  $Tx$  for a given  $x \in \mathbb{R}^2$  when  $T$  is nonexpansive, contractive, firmly nonexpansive, averaged nonexpansive or GAN with exponent 3 with respect to  $\|\cdot\|_2$ : inner region of the circles including the boundaries

$Tx$  may stay on the boundary of  $B(\hat{x}, \|x - \hat{x}\|)$  and not be equal to  $x$ , that is,  $\|Tx - \hat{x}\| = \|x - \hat{x}\|$  but  $Tx \neq x$ . For the same reason, for any positive integer  $k$ ,  $T^k x$  may always stay on the boundary, which implies that the fixed-point iteration of  $T$  does not converge. If  $T$  is contractive, the range of  $T^k x$  will shrink as  $k$  increases, which leads to the convergence of  $T^k x$  to  $\hat{x}$ . For the case when  $T$  is averaged nonexpansive, the range of  $Tx$  is an inscribed ball of  $B(\hat{x}, \|x - \hat{x}\|)$  with tangent point  $x$ . One can refer to Theorem 2.2 of [14] for convergence of the fixed-point iteration of  $T$  for this case.

In Fig. 2, we show the range of  $Tx$  for the cases that  $T$  is  $\frac{1}{2}$ -GAN with exponent  $\frac{1}{2}$ , 1, 2 or 3. When the exponent  $\gamma$  is equal to 2, the range of  $Tx$  is a ball the same as in the case of averaged nonexpansiveness. It is of the egg shape for exponent 3 and the water-drop shape for exponent 1. We will show that the convergence rate of the fixed-point iteration of  $T$  improves as  $\gamma$  decreases. Especially, when  $\gamma < 1$ , the range of  $T$  has some kind of contractive property (the point  $x$  is included), which is called as the FP-contractive property.

#### 4. Convergence rate analysis for fixed-point iterations of GAN operators

In this section, we establish convergence of the fixed-point iteration of a GAN operator. The local convergence rate (the convergence rate of the distance between two consecutive iterates) and global convergence rate (the convergence rate in terms of the distance between an iterate and a fixed-point) are provided.

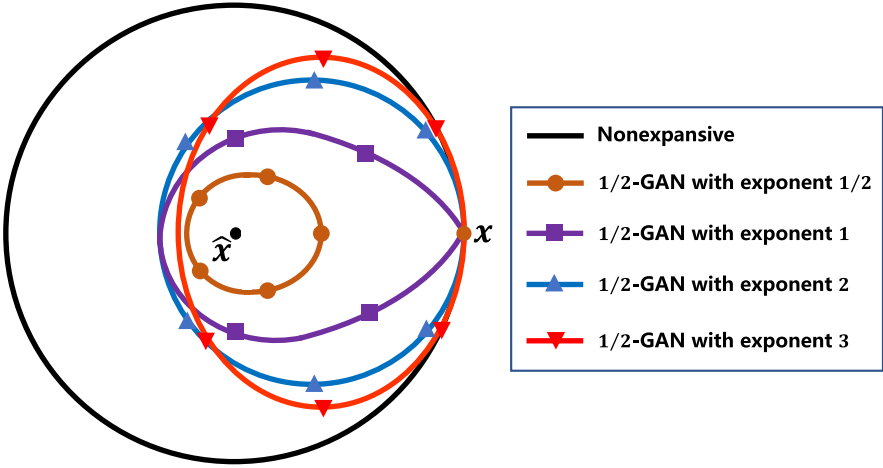


FIGURE 2. The range of  $Tx$  for a given  $x \in \mathbb{R}^2$  when  $T$  is nonexpansive or  $\frac{1}{2}$ -GAN with exponent  $\frac{1}{2}$ , 1, 2 and 3 with respect to  $\|\cdot\|_2$ : inner region of the closed curves including the boundaries

#### 4.1. Local convergence rate analysis

We will show in this subsection that the local convergence rate of the fixed-point iteration of a GAN operator with exponent  $\gamma$  is  $o(k^{-\frac{1}{\gamma}})$ . The smaller the exponent  $\gamma$  a GAN operator has, the higher local convergence rate its fixed-point iteration results.

We first describe the fixed-point iteration of an operator. By  $\mathbb{N}_0$  and  $\mathbb{N}_+$  we denote the set of all nonnegative integers and the set of all positive integers, respectively. Given an initial vector  $x^0 \in \mathbb{R}^n$ , the fixed-point iteration of  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is given by

$$x^{k+1} = Tx^k, \quad k \in \mathbb{N}_0.$$

We call the sequence  $\{x^k\}$  generated by the fixed-point iteration of  $T$  a Picard sequence of operator  $T$ .

We begin with stating the main theorem of this section. For sequences  $\{a_k\} \subset \mathbb{R}_+ \cup \{0\}$  and  $\{b_k\} \subset \mathbb{R}_+$ , both tending to zero, if  $\lim_{k \rightarrow \infty} \frac{a_k}{b_k} = 0$ , we write  $a_k = o(b_k)$ . If there exist constants  $c > 0$  and  $K \in \mathbb{N}_0$  such that  $a_k \leq cb_k$  for all  $k \geq K$ , we write  $a_k = O(b_k)$ .

**Theorem 4.1.** *If  $T \in \Lambda$  is GAN with exponent  $\gamma \in \mathbb{R}_+$ , then for any initial vector  $x^0 \in \mathbb{R}^n$ , the Picard sequence  $\{x^k\}$  of operator  $T$  converges to some  $x^* \in \text{Fix}(T)$ , and*

$$\|x^{k+1} - x^k\| = o\left(k^{-\frac{1}{\gamma}}\right). \quad (4.1)$$

We now proceed to prove Theorem 4.1. To this end, we recall Theorem 1 of [41] as a lemma.



**Lemma 4.2.** *Let  $T \in \Lambda$  be a nonexpansive operator. For any initial vector  $x^0 \in \mathbb{R}^n$ , if the Picard sequence  $\{x^k\}$  of  $T$  satisfies  $\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0$ , then  $\{x^k\}$  converges to a fixed-point of  $T$ .*

We also need Lemma 3 of [22], which we state below.

**Lemma 4.3.** *Suppose that  $\{a_k\}$  and  $\{b_k\}$  are two nonnegative sequences in  $\mathbb{R}$ . If  $\sum_{k=0}^{\infty} a_k b_k < +\infty$ ,  $\{b_k\}$  is monotonically decreasing, and there exists  $\varepsilon > 0$  such that  $a_k \geq \varepsilon$  for all  $k \in \mathbb{N}_0$ , then  $b_k = o(\frac{1}{k})$ .*

*Proof of Theorem 4.1.* We first show convergence of the sequence  $\{x^k\}$ . Since  $T$  is GAN with exponent  $\gamma$ , we know that it is nonexpansive and there exists  $\mu \in \mathbb{R}_+$  such that for any  $\hat{x} \in \text{Fix}(T)$ ,

$$\|x^{k+1} - \hat{x}\|^\gamma + \mu \|x^{k+1} - x^k\|^\gamma \leq \|x^k - \hat{x}\|^\gamma. \quad (4.2)$$

For any positive integer  $K$ , summing both sides of the inequality (4.2) for  $k = 0, 1, \dots, K$  yields that

$$\sum_{k=0}^K \mu \|x^{k+1} - x^k\|^\gamma \leq \|x^0 - \hat{x}\|^\gamma - \|x^{K+1} - \hat{x}\|^\gamma \leq \|x^0 - \hat{x}\|^\gamma. \quad (4.3)$$

Inequality (4.3) ensures that series

$$\sum_{k=0}^{\infty} \mu \|x^{k+1} - x^k\|^\gamma < +\infty. \quad (4.4)$$

Result (4.4) implies that  $\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0$ . By Lemma 4.2, we conclude that  $\{x^k\}_{k \in \mathbb{N}_0}$  converges to some  $x^* \in \text{Fix}(T)$ .

We next employ Lemma 4.3 to show that (4.1) holds. Applying Lemma 4.3 to the sequences  $a_k := \mu$  and  $b_k := \|x^{k+1} - x^k\|^\gamma$ ,  $k \in \mathbb{N}_0$ , it suffices to show that  $\{\|x^{k+1} - x^k\|\}$  is monotonically decreasing. This follows from the nonexpansiveness of  $T$  since it implies that

$$\|x^{k+2} - x^{k+1}\| = \|Tx^{k+1} - Tx^k\| \leq \|x^{k+1} - x^k\|$$

for all  $k \in \mathbb{N}_0$ . Therefore, by Lemma 4.3, (4.1) holds.  $\square$

Since an averaged nonexpansive operator is GAN with exponent  $\gamma = 2$ , Theorem 4.1 covers the well-known result that the local convergence rate of the fixed-point iteration of an averaged nonexpansive operator is  $o(k^{-\frac{1}{2}})$ , see [22]. Moreover, it ensures that the local convergence rate of the fixed-point iteration of a GAN operator with exponent  $\gamma < 2$  is higher than that of an averaged nonexpansive operator.

## 4.2. Global convergence rate analysis

We consider in this subsection the global convergence rate of the fixed-point iteration of a GAN operator and investigate the relation between the local convergence rate and the global convergence rate. We will show that the fixed-point iteration of a GAN operator with exponent  $\gamma \in (0, 1)$  can achieve an exponential global convergence rate. Moreover, if a GAN operator is also Hölder regular, the global convergence rate of its fixed-point iteration will depend on both the exponent of generalized averaged nonexpansiveness and

the exponent of Hölder regularity. The definition of Hölder regularity will be given later in this subsection.

We first establish a relation between the local convergence rate and the global convergence rate.

**Theorem 4.4.** *If  $T \in \Omega_\mu^1$  for  $\mu \in (0, 1]$ , then for any initial vector  $x^0 \in \mathbb{R}^n$ , the Picard sequence  $\{x^k\}$  of  $T$  converges to some  $x^* \in \text{Fix}(T)$ , and there holds the equivalence relation for all positive integers  $k$ ,*

$$\mu \sum_{j=k}^{\infty} \|x^{j+1} - x^j\| \leq \|x^k - x^*\| \leq \sum_{j=k}^{\infty} \|x^{j+1} - x^j\|. \quad (4.5)$$

*Proof.* Convergence of  $\{x^k\}$  to some  $x^* \in \text{Fix}(T)$  follows from Theorem 4.1. It remains to establish the equivalence relation (4.5). Since  $T \in \Omega_\mu^1$ , by the definition of generalized averaged nonexpansiveness, we have that

$$\mu \|x^{j+1} - x^j\| \leq \|x^j - x^*\| - \|x^{j+1} - x^*\|, \quad j \in \mathbb{N}_0.$$

For any  $N > k$ , summing the above inequality for  $j = k, k+1, \dots, N$  yields that

$$\mu \sum_{j=k}^N \|x^{j+1} - x^j\| \leq \|x^k - x^*\| - \|x^{N+1} - x^*\| \leq \|x^k - x^*\|. \quad (4.6)$$

In the inequality above, we let  $N \rightarrow \infty$  and get the left inequality of (4.5).

To establish the right inequality of (4.5), for any  $N > k$ , we repeatedly use the triangle inequality and obtain that

$$\|x^k - x^*\| \leq \sum_{j=k}^N \|x^j - x^{j+1}\| + \|x^{N+1} - x^*\|. \quad (4.7)$$

Inequality (4.6) implies that  $\sum_{j=k}^{\infty} \|x^{j+1} - x^j\| < +\infty$ . Moreover, the first part of this theorem ensures that  $\lim_{N \rightarrow \infty} \|x^{N+1} - x^*\| = 0$ . Hence, letting  $N \rightarrow \infty$  in inequality (4.7) yields the right inequality of (4.5).  $\square$

Theorem 4.4 indicates that when the operator  $T$  is GAN with exponent 1, the global convergence rate of its Picard sequence is equivalent to the convergence rate of  $\sum_{j=k}^{\infty} \|x^{j+1} - x^j\|$ . We next show how Theorem 4.4 provides a way to estimate the global convergence rate. We first show a technical result.

**Proposition 4.5.** *If  $\{a_k\} \subset \mathbb{R}$  is a nonnegative sequence with  $a_k = o(k^{-\alpha})$ , then  $\sum_{j=k}^{\infty} a_j = o(k^{-(\alpha-1)})$ .*

*Proof.* Since  $a_k = o(k^{-\alpha})$ , for any  $\varepsilon > 0$ , there is  $K \in \mathbb{N}_0$  such that  $a_j < \frac{\varepsilon}{j^\alpha}$  for all  $j \geq K$ . Summing this inequality for  $j = k, k+1, \dots$ , with  $k \geq K$ , we obtain that

$$\sum_{j=k}^{\infty} a_j < \varepsilon \sum_{j=k}^{\infty} \frac{1}{j^\alpha} \leq \varepsilon \int_{k-1}^{\infty} \frac{1}{t^\alpha} dt = \frac{\varepsilon}{(\alpha-1)(k-1)^{\alpha-1}}.$$

This establishes the desired estimate.  $\square$

Theorem 4.4 together with Propositions 3.4 and 4.5 leads to the following theorem.

**Theorem 4.6.** *If  $T \in \Lambda$  is GAN with exponent  $\gamma \in (0, 1)$ , then for any initial vector  $x^0 \in \mathbb{R}^n$ , the Picard sequence  $\{x^k\}$  of  $T$  converges to some  $x^* \in \text{Fix}(T)$ , and  $\|x^k - x^*\| = o\left(k^{-\frac{1-\gamma}{\gamma}}\right)$ .*

*Proof.* It follows from Theorem 4.1 that  $\{x^k\}$  converges to some  $x^* \in \text{Fix}(T)$  and  $\|x^{k+1} - x^k\| = o\left(k^{-\frac{1}{\gamma}}\right)$ . Applying Proposition 4.5 with  $a_k := \|x^{k+1} - x^k\|$ , we obtain that

$$\sum_{j=k}^{\infty} \|x^{j+1} - x^j\| = o\left(k^{-\frac{1-\gamma}{\gamma}}\right).$$

Moreover, by Proposition 3.4, we see that  $T$  is GAN with exponent 1. Thus, the desired result of this theorem follows from Theorem 4.4.  $\square$

In fact, according to Proposition 3.7, we know that GAN operator with exponent  $\gamma \in (0, 1)$  is FP- $\rho$ -contractive for some  $\rho \in (0, 1)$ , which leads to higher order global convergence rate of its Picard sequence than the result shown in Theorem 4.6. To this end, we first show that the Picard sequence of a FP-contractive operator has an exponential global convergence rate.

**Theorem 4.7.** *If operator  $T \in \Lambda$  is FP-contractive, then for any initial vector  $x^0 \in \mathbb{R}^n$ , the Picard sequence  $\{x^k\}$  of  $T$  either converges to some  $x^* \in \text{Fix}(T)$  within a finite number of iterations or there exists  $\rho \in (0, 1)$  such that*

$$\|x^k - x^*\| \leq \rho^k \|x^0 - x^*\|, \quad \text{for all } k \in \mathbb{N}_0. \quad (4.8)$$

*Proof.* If there exists an integer  $K \in \mathbb{N}_0$  such that  $x^K \in \text{Fix}(T)$ , then  $x^k = x^K$  for all  $k > K$ , and hence  $\lim_{k \rightarrow \infty} x^k = x^K$ . Otherwise,  $x^k \notin \text{Fix}(T)$  for all  $k \in \mathbb{N}_0$ . In this case, by the definition of the FP-contractive operator, there exist  $x^* \in \text{Fix}(T)$  and  $\rho \in (0, 1)$  such that

$$\|x^{k+1} - x^*\| \leq \rho \|x^k - x^*\|, \quad \text{for all } k \in \mathbb{N}_0.$$

Repeatedly using this inequality, we obtain the desired estimate (4.8).  $\square$

The next corollary improves the global convergence rate given in Theorem 4.6.

**Corollary 4.8.** *If operator  $T \in \Lambda$  is GAN with exponent  $\gamma \in (0, 1)$ , then for any initial vector  $x^0 \in \mathbb{R}^n$ , the Picard sequence  $\{x^k\}$  of  $T$  either converges to some  $x^* \in \text{Fix}(T)$  within finite iterations or there exists some  $\rho \in (0, 1)$  such that estimate (4.8) holds.*

*Proof.* By Proposition 3.7, a GAN operator  $T \in \Lambda$  with exponent  $\gamma \in (0, 1)$  is FP-contractive. Then the desired result of this corollary follows from Theorem 4.7.  $\square$

To obtain global convergence rates for the case with the exponent  $\gamma \geq 1$ , we need an additional condition that establishes a relation between the local convergence rate and the global convergence rate. In view of this, we recall the definition of Hölder regular operators introduced in [11]. For a set  $E \subset \mathbb{R}^n$  and  $x \in \mathbb{R}^n$ , we define

$$d(x, E) := \inf_{y \in E} \|x - y\|.$$

**Definition 4.9.** Let  $T \in \Lambda$ . We say that  $T$  is a Hölder regular (HR) operator with exponent  $\gamma$ , if there exist  $\gamma \in \mathbb{R}_+$  and  $\mu \in \mathbb{R}_+$  such that

$$d(x, \text{Fix}(T)) \leq \mu \|x - Tx\|^\gamma, \quad \text{for all } x \in \mathbb{R}^n.$$

We verify below that for any  $\rho \in (0, 1)$ , a FP- $\rho$ -contractive operator  $T \in \Lambda$  is HR with exponent 1. By the FP-contractivity of  $T$  and the triangle inequality, for all  $x \in \mathbb{R}^n$  and  $\hat{x} \in \text{Fix}(T)$ , we have that

$$\|Tx - x^*\| \leq \rho \|x - x^*\| \quad \text{and} \quad \|x - x^*\| \leq \|Tx - x^*\| + \|x - Tx\|,$$

which imply that  $\|x - x^*\| \leq \frac{1}{1-\rho} \|x - Tx\|$ , and hence

$$d(x, \text{Fix}(T)) \leq \frac{1}{1-\rho} \|x - Tx\|, \quad \text{for all } x \in \mathbb{R}^n.$$

Thus,  $T$  is HR with exponent 1. We shall show in the next section that the gradient descent operator is also HR with exponent 1 under appropriate assumptions.

Now we state the main result on the global convergence rate of the fixed-point iteration of GAN operators.

**Theorem 4.10.** *If  $T \in \Lambda$  is GAN with exponent  $\gamma_1 \in \mathbb{R}_+$  and HR with exponent  $\gamma_2 \in \mathbb{R}_+$ , then for any initial vector  $x^0 \in \mathbb{R}^n$ , the Picard sequence  $\{x^k\}$  of  $T$  converges to some  $x^* \in \text{Fix}(T)$ , and there exists  $\rho \in (0, 1)$  such that*

$$\|x^k - x^*\| = \begin{cases} O\left(k^{-\frac{\gamma_2}{\gamma_1(1-\gamma_2)}}\right), & 0 < \gamma_2 < 1, \\ O(\rho^k), & \gamma_2 \geq 1. \end{cases} \quad (4.9)$$

To prove Theorem 4.10, we recall Lemma 4.1 of [12].

**Lemma 4.11.** *Suppose that  $\{a_k\}$  and  $\{b_k\}$  be two sequences of nonnegative numbers. For  $p > 0$ , if there exists  $K \in \mathbb{N}_0$  such that*

$$a_{k+1} \leq a_k(1 - b_k a_k^p), \quad \text{for all } k \geq K,$$

*then*

$$a_k \leq \left( a_K^{-p} + p \sum_{j=K}^{k-1} b_j \right)^{-\frac{1}{p}}, \quad \text{for all } k > K.$$

For a closed convex set  $E \subset \mathbb{R}^n$ , we define  $P_E(x) := \underset{y \in E}{\operatorname{argmin}} \{\|x - y\|\}$ .

Note that  $\text{Fix}(T)$  is closed and convex if  $T \in \Lambda$  is nonexpansive. Hence,  $P_{\text{Fix}(T)}(x)$  is well-defined, which will be used in the proof of the next proposition.

**Proposition 4.12.** *Suppose that  $T \in \Lambda$  is nonexpansive. For the Picard sequence  $\{x^k\}$  of  $T$  with a given initial vector  $x^0 \in \mathbb{R}^n$ , let  $d_k := d(x^k, \text{Fix}(T))$ ,  $k \in \mathbb{N}_0$ . If there exist  $\gamma > 0$ ,  $\mu > 0$ ,  $\vartheta \geq 1$  and  $K \in \mathbb{N}_0$  such that*

$$d_{k+1}^\gamma \leq d_k^\gamma - \mu d_k^{\gamma\vartheta}, \quad \text{for all } k \geq K, \quad (4.10)$$

*then  $\{x^k\}$  converges to some  $x^* \in \text{Fix}(T)$ . Moreover, there exist  $C \in \mathbb{R}_+$  and  $\rho \in [0, 1)$  such that for  $k > K$ ,*

$$\|x^k - x^*\| \leq \begin{cases} Ck^{-\frac{1}{\gamma(\vartheta-1)}}, & \vartheta > 1, \\ C\rho^{k-K}, & \vartheta = 1. \end{cases}$$

*Proof.* Let  $a_k = d_k^\gamma$  and  $p = \vartheta - 1 \geq 0$ . Then (4.10) becomes

$$a_{k+1} \leq a_k(1 - \mu a_k^p), \quad \text{for all } k \geq K. \quad (4.11)$$

We consider two cases based on the value of  $\vartheta$ .

*Case 1:*  $\vartheta > 1$ . We first show that  $\{x^k\}$  converges to some  $x^* \in \text{Fix}(T)$ . It follows from Lemma 4.11 with  $b_k := \mu$  that

$$a_k \leq (a_K^{-p} + p\mu(k-K))^{-\frac{1}{\vartheta-1}}, \quad \text{for all } k > K.$$

Hence, there exists  $C_1 > 0$  such that for  $k > K$ ,

$$d_k = a_k^{\frac{1}{\gamma}} \leq C_1 k^{-\frac{1}{\gamma(\vartheta-1)}} \rightarrow 0, \quad \text{as } k \rightarrow \infty.$$

By the nonexpansiveness of  $T$ , we know that  $\{\|x^k - \hat{x}\|\}$  is monotonically decreasing for any  $\hat{x} \in \text{Fix}(T)$ . Then

$$\begin{aligned} \|x^{k+1} - x^k\| &\leq \|x^{k+1} - P_{\text{Fix}(T)}(x^k)\| + \|x^k - P_{\text{Fix}(T)}(x^k)\| \\ &\leq 2\|x^k - P_{\text{Fix}(T)}(x^k)\| = 2d_k \rightarrow 0. \end{aligned}$$

We conclude from Lemma 4.2 that  $\{x^k\}$  converges to some  $x^* \in \text{Fix}(T)$ . Using the monotonicity of  $\{\|x^k - \hat{x}\|\}$  for any  $\hat{x} \in \text{Fix}(T)$  again, we have that

$$\|x^m - P_{\text{Fix}(T)}(x^k)\| \leq \|x^{m-1} - P_{\text{Fix}(T)}(x^k)\| \leq \dots \leq \|x^k - P_{\text{Fix}(T)}(x^k)\| = d_k$$

for all  $m > k$ ,  $k \in \mathbb{N}_0$ . Letting  $m$  tend to infinity, the above inequality becomes

$$\|x^* - P_{\text{Fix}(T)}(x^k)\| \leq d_k, \quad \text{for all } k \in \mathbb{N}_0,$$

which together with the triangle inequality implies for all  $k > K$  that

$$\begin{aligned} \|x^k - x^*\| &\leq \|x^k - P_{\text{Fix}(T)}(x^k)\| + \|x^* - P_{\text{Fix}(T)}(x^k)\| \\ &\leq 2d_k \leq 2C_1 k^{-\frac{1}{\gamma(\vartheta-1)}}. \end{aligned}$$

*Case 2:* Suppose that  $\vartheta = 1$ . Then (4.10) becomes  $d_{k+1}^\gamma \leq (1 - \mu)d_k^\gamma$  for all  $k \geq K$ . This implies that  $\mu \in (0, 1]$  and for  $k > K$ ,  $d_k \leq d_K(1 - \mu)^{\frac{k-K}{\gamma}} \rightarrow 0$ . By the same argument as Case 1, there exists some  $x^* \in \text{Fix}(T)$  such that

$$\|x^k - x^*\| \leq 2d_k \leq 2d_K(1 - \mu)^{\frac{k-K}{\gamma}}.$$

Therefore, the proof is completed by setting  $C := \max\{2C_1, 2d_K\}$  and  $\rho := (1 - \mu)^{\frac{1}{\gamma}}$ .  $\square$

Note that the result in Proposition 3.1 of [11] is a special case of the above proposition with  $\gamma = 2$ . The generalization for any  $\gamma \in \mathbb{R}_+$  is necessary for the global convergence rate analysis of the fixed-iteration of GAN operator. We next employ Proposition 4.12 to prove Theorem 4.10.

*Proof of Theorem 4.10.* Since  $T$  is GAN with exponent  $\gamma_1$ , by Theorem 4.1, we know that  $\{x^k\}$  converges to some  $x^* \in \text{Fix}(T)$  and  $\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0$ . Moreover, there exists  $\mu_1 \in \mathbb{R}_+$  such that for all  $k \in \mathbb{N}_0$ ,

$$\|x^{k+1} - P_{\text{Fix}(T)}(x^k)\|^{\gamma_1} \leq \|x^k - P_{\text{Fix}(T)}(x^k)\|^{\gamma_1} - \mu_1 \|x^{k+1} - x^k\|^{\gamma_1}. \quad (4.12)$$

Let  $d_k := d(x^k, \text{Fix}(T))$ ,  $k \in \mathbb{N}_0$ . By the definition of  $d_{k+1}$  and (4.12), we obtain that

$$\begin{aligned} d_{k+1}^{\gamma_1} &\leq \|x^{k+1} - P_{\text{Fix}(T)}(x^k)\|^{\gamma_1} \\ &\leq d_k^{\gamma_1} - \mu_1 \|x^{k+1} - x^k\|^{\gamma_1}, \quad \text{for all } k \in \mathbb{N}_0. \end{aligned} \quad (4.13)$$

It follows from the Hölder regularity of  $T$  that there exists  $\mu_2 \in \mathbb{R}_+$  such that

$$d_k \leq \mu_2 \|x^{k+1} - x^k\|^{\gamma_2}, \quad \text{for all } k \in \mathbb{N}_0. \quad (4.14)$$

Since  $\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0$ , there exists  $K$  such that  $\|x^{k+1} - x^k\| < 1$  for all  $k \geq K$ , which together with (4.14) implies that for  $\gamma_2 \geq 1$ ,

$$d_k \leq \mu_2 \|x^{k+1} - x^k\|, \quad \text{for all } k \geq K. \quad (4.15)$$

Now combining (4.13) with (4.14) for  $0 < \gamma_2 < 1$  gives that

$$d_{k+1}^{\gamma_1} \leq d_k^{\gamma_1} - \mu_1 \mu_2^{-\frac{\gamma_1}{\gamma_2}} d_k^{\frac{\gamma_1}{\gamma_2}}, \quad \text{for all } k \in \mathbb{N}_0. \quad (4.16)$$

Combining (4.13) with (4.15) for  $\gamma_2 \geq 1$  gives that

$$d_{k+1}^{\gamma_1} \leq d_k^{\gamma_1} - \mu_1 \mu_2^{-\gamma_1} d_k^{\gamma_1}, \quad \text{for all } k \geq K. \quad (4.17)$$

Then we conclude from Proposition 4.12 that there exist  $C \in \mathbb{R}_+$  and  $\rho \in [0, 1)$  such that for all  $k > K$ ,

$$\|x^k - x^*\| \leq \begin{cases} C k^{-\frac{\gamma_2}{\gamma_1(1-\gamma_2)}}, & 0 < \gamma_2 < 1, \\ C \rho^{k-K}, & \gamma_2 \geq 1. \end{cases} \quad (4.18)$$

For  $0 < \gamma_2 < 1$ ,  $\|x^k - x^*\| = O\left(k^{-\frac{\gamma_2}{\gamma_1(1-\gamma_2)}}\right)$  follows from (4.18) immediately. We next consider the case  $\gamma_2 \geq 1$ . In this case, if  $\rho = 0$ , then it is obvious that  $\|x^k - x^*\| = O(\rho^k)$  holds according to (4.18). If  $\rho \in (0, 1)$ , then for all  $k > K$ ,  $\|x^k - x^*\| \leq C' \rho^k$ , where  $C' = C \rho^{-K}$ . Thus  $\|x^k - x^*\| = O(\rho^k)$ .  $\square$

Theorem 4.10 extends the result given in [11] where operators that are averaged nonexpansive (GAN with exponent  $\gamma_1 = 2$ ) and HR with exponent  $\gamma_2 \in (0, 1]$  were considered.

We close this section by listing convergence rates of the fixed-point iterations of GAN operators with different exponents in Table 1.

TABLE 1. Convergence rates of the fixed-point iterations of GAN operators

Case	Conditions	Convergence rate
1	GAN with exponent $\gamma \in [1, +\infty)$	Local: $o\left(k^{-\frac{1}{\gamma}}\right)$
2	GAN with exponent $\gamma \in (0, 1)$	Global: exponential
3	GAN with exponent $\gamma_1 \in [1, +\infty)$ & HR with exponent $\gamma_2 \in (0, 1)$	Global: $O\left(k^{-\frac{\gamma_2}{\gamma_1(1-\gamma_2)}}\right)$
4	GAN with exponent $\gamma_1 \in [1, +\infty)$ & HR with exponent $\gamma_2 \in [1, +\infty)$	Global: exponential

## 5. Convergence rate analysis for optimization

In this section, we first describe the fixed-point algorithms for the convex optimization problems described in Sect. 2, and then employ the results in Sect. 4 to analyze their convergence rates. The GAN operators provide a unified framework for developing fixed-point iterative schemes for convex optimization problems and analyzing their convergence and convergence rates.

By the definition (2.2) of operator  $T_1$ , its fixed-point iteration is the gradient descent algorithm given as follows:

$$x^{k+1} = x^k - \beta \nabla f(x^k), \quad \text{where } \beta \in \mathbb{R}_+. \quad (5.1)$$

The fixed-point iteration (2.3) of  $T_2$  is given by

$$x^{k+1} = \text{prox}_{\beta g}(x^k - \beta \nabla f(x^k)), \quad \text{where } \beta \in \mathbb{R}_+. \quad (5.2)$$

We next derive the fixed-point iteration of  $T_3$  defined by (2.4). Note that  $W = R^{-1}G$ , where  $R := \begin{pmatrix} \beta I_n & \\ & \eta I_m \end{pmatrix}$ . We can verify that the fixed-point iteration  $v^{k+1} = T_3(v^k)$  is equivalent to

$$v^{k+1} = \tilde{T}((E - G)v^{k+1} + (G - R\nabla r)(v^k)),$$

that is,

$$\begin{cases} x^{k+1} = \text{prox}_{\beta h}(x^k - \beta(\nabla f(x^k) + B^\top y^k)), \\ y^{k+1} = \text{prox}_{\eta g^*}(y^k + \eta B(2x^{k+1} - x^k)). \end{cases}$$

By using the well-known Moreau decomposition [34, 39]

$$\mathcal{I} = \text{prox}_{\eta g^*} + (\eta \mathcal{I}) \circ \text{prox}_{\frac{1}{\eta} g} \circ (\eta^{-1} \mathcal{I}),$$

we have the following fixed-point iteration of  $T_3$ ,

$$\begin{cases} x^{k+1} = \text{prox}_{\beta h}(x^k - \beta(\nabla f(x^k) + B^\top y^k)), \\ y^{k+1} = \eta \left( \mathcal{I} - \text{prox}_{\frac{1}{\eta} g} \right) \left( \frac{1}{\eta} y^k + B(2x^{k+1} - x^k) \right), \end{cases} \quad \text{where } \beta, \eta \in \mathbb{R}_+. \quad (5.3)$$

We next show the generalized averaged nonexpansiveness with exponent 2 of  $T_1$ ,  $T_2$  and  $T_3$ , which offers  $o(k^{-\frac{1}{2}})$  local convergence rate for algorithms (5.1), (5.2) and (5.3). We then provide higher order convergence rates for the fixed-point algorithms (5.1) and (5.2) under additional assumptions.

**Proposition 5.1.** *Let  $T_1$ ,  $T_2$  and  $T_3$  be defined by (2.2), (2.3) and (2.4), respectively. If  $\beta < \frac{2}{L}$  for  $T_1$ ,  $T_2$  and  $T_3$ , and  $\mu < \frac{2(2-\beta L)}{4\beta\|B\|_2^2 + L(2-\beta L)}$  for  $T_3$ , then  $T_1$  and  $T_2$  are GAN with exponent 2 with respect to  $\|\cdot\|_2$ ,  $T_3$  is GAN with exponent 2 with respect to  $\|\cdot\|_W$ .*

*Proof.* We first show the generalized averaged nonexpansiveness of  $T_1$  and  $T_2$ . It follows from the proof of Theorem 26.14 of [6] that  $T_1$  and  $T_2$  are both averaged nonexpansive with respect to  $\|\cdot\|_2$ . Hence they are both GAN with exponent 2 with respect to  $\|\cdot\|_2$  by Proposition 3.2.

We now turn to considering operator  $T_3$ . It follows from Lemma 7 of [34] that  $T_3$  is averaged nonexpansive with respect to  $\|\cdot\|_W$  if the minimum eigenvalue of  $W$  is greater than  $\frac{L}{2}$ , that is,  $W - \frac{L}{2}I$  is positive definite. Let  $\tilde{B} := \frac{1}{\sqrt{(\frac{1}{\beta} - \frac{L}{2})(\frac{1}{\mu} - \frac{L}{2})}}B$ . According to Lemma 6.2 of [30],  $W - \frac{L}{2}I$  is positive definite if and only if  $\|\tilde{B}\|_2 < 1$ , that is,

$$\left(\frac{1}{\beta} - \frac{L}{2}\right) \left(\frac{1}{\mu} - \frac{L}{2}\right) > \|B\|_2^2. \quad (5.4)$$

Since  $\beta \in (0, \frac{2}{L})$ ,  $\eta \in (0, \frac{2(2-\beta L)}{4\beta\|B\|_2^2 + L(2-\beta L)})$ , it is easy to verify that (5.4) holds, which implies that  $T_3$  is averaged nonexpansive with respect to  $\|\cdot\|_W$ , and hence it is GAN with exponent 2 with respect to  $\|\cdot\|_W$ .  $\square$

**Proposition 5.2.** *Suppose that  $\beta < \frac{2}{L}$  and  $\mu < \frac{2(2-\beta L)}{4\beta\|B\|_2^2 + L(2-\beta L)}$ . Then for arbitrary initial vectors  $x^0 \in \mathbb{R}^n$  and  $y^0 \in \mathbb{R}^m$ , the following statements hold:*

- (i) *Sequence  $\{x^k\}$  generated by Algorithm (5.1) converges to a minimizer of the objective function  $f$ .*
- (ii) *Sequence  $\{x^k\}$  generated by Algorithm (5.2) converges to a minimizer of the objective function  $f + g$ .*
- (iii) *Sequences  $\{x^k\}$  generated by Algorithm (5.3) with  $\{y^k\}$  converges to a minimizer of the objective function  $f + g \circ B + h$ .*
- (iv) *The local convergence rate of  $\{x^k\}$  in any of the above three cases is  $o(k^{-\frac{1}{2}})$ .*

*Proof.* By Proposition 5.1, we know that  $T_1$  and  $T_2$  are both GAN with exponent 2 with respect to  $\|\cdot\|_2$ ,  $T_3$  is GAN with exponent 2 with respect to  $\|\cdot\|_W$ . Then we conclude from Theorem 4.1 and the equivalence of all norms on  $\mathbb{R}^n$  that the fixed-point iterations of  $T_1$ ,  $T_2$  and  $T_3$  (or Algorithm (5.1), (5.2) and (5.3)) converge to their fixed-points and (iv) holds. Let  $v^* := \begin{pmatrix} x^* \\ y^* \end{pmatrix}$  be the fixed-point of  $T_3$  that the fixed-point iteration of  $T_3$  converges to, where  $x^* \in \mathbb{R}^n$ ,  $y^* \in \mathbb{R}^m$ . The proof is completed by noticing that the fixed-points



of  $T_1$  and  $T_2$  are minimizers of  $f$  and  $f + g$  respectively, and  $x^*$  is a minimizer of  $f + g \circ B + h$ .  $\square$

We comment that we have recovered the local convergence rate  $o\left(k^{-\frac{1}{2}}\right)$  of Algorithm (5.3) previously obtained in [31], by showing that  $T_3$  is the generalized averaged nonexpansiveness with exponent 2.

Based on the convergence rate analysis in previous sections, we are able to obtain further convergence rate results for the fixed-point algorithms (5.1) and (5.2). We first consider the one-dimensional case for Algorithm (5.1).

**Proposition 5.3.** *Suppose that function  $f \in \Gamma_0(\mathbb{R})$  is differentiable with an  $L$ -Lipschitz continuous derivative, where  $L \in \mathbb{R}_+$ . Then for  $\beta \in (0, \frac{2}{L})$ , the following hold:*

- (i)  $T_1$  is GAN with exponent 1.
- (ii) For any initial vector  $x^0 \in \mathbb{R}^n$ , the sequence  $\{x^k\}$  generated by Algorithm (5.1) converges to a minimizer of  $f$  with an  $o\left(\frac{1}{k}\right)$  local convergence rate.

*Proof.* We first prove (i). By the definition of generalized averaged nonexpansiveness with exponent 1 and the definition of  $T_1$ , it suffices to show that there exists  $\mu \in \mathbb{R}_+$  such that for all  $x, y \in \mathbb{R}$ ,

$$|(x - y) - \beta(f'(x) - f'(y))| + \mu|\beta(f'(x) - f'(y))| \leq |x - y|. \quad (5.5)$$

Let  $w := x - y$ ,  $v := \beta(f'(x) - f'(y))$  and  $\mu := \min\left\{\frac{1}{2}, \frac{2}{\beta L} - 1\right\}$ . Then  $\mu \in (0, 1)$  and  $L \leq \frac{2}{\beta(1+\mu)}$ . It follows from the  $L$ -Lipschitz continuity of  $f'$  that

$$|v| \leq \beta L|w| \leq \frac{2}{1+\mu}|w|. \quad (5.6)$$

The convexity of  $f$  implies that  $f'$  is monotonically increasing, and hence  $wv \geq 0$ . Multiplying  $(1 - \mu^2)|v|$  on both sides of (5.6), we obtain that

$$(1 - \mu^2)v^2 \leq 2(1 - \mu)wv,$$

which implies that

$$v^2 - 2wv + w^2 \leq w^2 - 2\mu wv + \mu^2 v^2,$$

that is,

$$(w - v)^2 \leq (|w| - \mu|v|)^2. \quad (5.7)$$

By (5.6) and the fact that  $\mu \in (0, 1)$ , it is easy to see that  $\mu|v| \leq |w|$ . Hence (5.7) is equivalent to  $|w - v| \leq |w| - \mu|v|$ , that is, (5.5) holds, and hence  $T_1$  is GAN with exponent 1.

Now we employ (i) and Theorem 4.1 to prove (ii). The convergence of  $\{x^k\}$  to a minimizer of  $f$  has been shown in Proposition 5.2 (i). Since  $T_1$  is GAN with exponent 1, the  $o\left(\frac{1}{k}\right)$  local convergence rate of its fixed-point iteration follows from Theorem 4.1 immediately.  $\square$

In Proposition 5.3, we have shown the generalized averaged nonexpansiveness with exponent 1 of  $T_1$  and the convergence rate of Algorithm (5.1) in one-dimensional case. We next consider the higher-dimensional case.

In fact, we are able to show that  $T_1$  is both GAN with exponent 1 and HR with exponent 1 under appropriate assumptions, which leads to an exponential global convergence rate for Algorithm (5.1) by Theorem 4.10. To establish this result, we recall the Baillon–Haddad theorem [4].

**Lemma 5.4.** *Suppose that  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$  is a differentiable convex function. Then  $\nabla\psi$  is  $L$ -Lipschitz with respect to  $\|\cdot\|$  for some  $L \in \mathbb{R}_+$  if and only if*

$$\|\nabla\psi(x) - \nabla\psi(y)\|^2 \leq L\langle x - y, \nabla\psi(x) - \nabla\psi(y) \rangle, \quad \text{for all } x, y \in \mathbb{R}^n.$$

**Theorem 5.5.** *Let  $f \in \Gamma_0(\mathbb{R}^n)$  be differentiable. If there exist  $L_1 \geq L_2 > 0$  such that*

$$L_2\|x - y\| \leq \|\nabla f(x) - \nabla f(y)\| \leq L_1\|x - y\|, \quad \text{for all } x, y \in \mathbb{R}^n, \quad (5.8)$$

*then for  $\beta \in (0, \frac{2}{L_1})$ , the following hold:*

- (i)  $T_1$  is both GAN with exponent 1 and HR with exponent 1.
- (ii) For any initial vector  $x^0 \in \mathbb{R}^n$ , the sequence  $\{x^k\}$  generated by Algorithm (5.1) converges to a minimizer of  $f$  with an exponential global convergence rate.

*Proof.* We first prove the generalized averaged nonexpansiveness of  $T_1$  by employing Theorem 3.11. Let  $T := \frac{1}{L_1}\nabla f$ . It follows from the second inequality of (5.8) and Lemma 5.4 that

$$\|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty \rangle, \quad \text{for all } x, y \in \mathbb{R}^n,$$

that is,  $T$  is firmly nonexpansive. By the first inequality of (5.8), we have

$$\|Tx - Ty\| \geq \frac{L_2}{L_1}\|x - y\|,$$

where  $\frac{L_2}{L_1} \in (0, 1]$ . Since  $T_1 = \mathcal{I} - \beta L_1 T$  and  $\beta L_1 \in (0, 2)$ , the generalized averaged nonexpansiveness with exponent 1 of  $T_1$  follows from Theorem 3.11 immediately.

We next show the Hölder regularity of  $T_1$ . Let  $\mu = \frac{1}{\beta L_2}$ . Since  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable, by Fermat's lemma [58], we know that  $\nabla f(\hat{x}) = \mathbf{0}$  for any  $\hat{x} \in \text{Fix}(T_1)$ . Using the first inequality of (5.8), for any  $x \in \mathbb{R}^n$ ,  $\hat{x} \in \text{Fix}(T_1)$ ,

$$\|x - \hat{x}\| \leq \frac{1}{L_2}\|\nabla f(x) - \nabla f(\hat{x})\| = \mu\|\beta\nabla f(x)\| = \mu\|x - T_1x\|,$$

which implies that  $d(x, \text{Fix}(T_1)) \leq \mu\|x - T_1x\|$ . Thus,  $T_1$  is HR with exponent 1.

Now we employ (i) and Theorem 4.10 to prove (ii). The convergence of  $\{x^k\}$  to a minimizer of  $f$  has been shown in Proposition 5.2 (i). Since  $T_1$  is both GAN with exponent 1 and HR with exponent 1, (ii) follows from Theorem 4.10 immediately.  $\square$

It is known (Theorem 2.1.15 of [40]) that if the objective function  $f$  is differentiable with a Lipschitz continuous gradient and strongly convex, then item (ii) of Theorem 5.5 holds. In fact, according to inequality (2.1.17) of Theorem 2.1.9 of [40] and the Cauchy–Schwarz inequality, for such a function  $f$ , the strong convexity of  $f$  implies the first inequality of (5.8) for some  $L_2 \in \mathbb{R}_+$ .

We next provide an example whose objective function satisfies (5.8).

**Corollary 5.6.** *Suppose function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is defined by  $f(x) := \frac{1}{2}\|Ax - b\|_2^2$ , where  $A \in \mathbb{R}^{m \times n}$  is a full column rank matrix,  $b \in \mathbb{R}^m$ . Then for any initial vector  $x^0 \in \mathbb{R}^n$ , the sequence  $\{x^k\}$  generated by Algorithm (5.1) converges to the minimizer of  $f$  with an exponential global convergence rate for  $\beta \in (0, \frac{2}{L})$ , where  $L$  is the maximum eigenvalue of  $A^\top A$ .*

*Proof.* It is easy to verify that  $f \in \Gamma_0(\mathbb{R}^n)$  and it is differentiable. The fact that  $A$  has full column rank implies the positive definiteness of the Hessian matrix  $H := A^\top A$  of  $f$ . Hence  $f$  is strictly convex and has a unique minimizer. According to Theorem 5.5, to prove this corollary, it suffices to show that there exist  $L_1 \geq L_2 > 0$  such that (5.8) holds. By the definition of  $f$ ,

$$\|\nabla f(x) - \nabla f(y)\|_2^2 = z^\top H^\top H z,$$

where  $z := x - y$ . Of course,  $H^\top H \in \mathbb{R}^{n \times n}$  is positive definite. Let  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  be the  $n$  eigenvalues of  $H^\top H$ . Then  $H^\top H - \lambda_1 I$  and  $\lambda_n I - H^\top H$  are both positive semi-definite, which implies that

$$\lambda_1 \|z\|_2^2 \leq \|H z\|_2^2 \leq \lambda_n \|z\|_2^2,$$

that is, (5.8) holds by setting  $L_2 = \sqrt{\lambda_1}$  and  $L_1 = \sqrt{\lambda_n}$ . Therefore, the desired result of this corollary follows from Theorem 5.5 (ii) immediately.  $\square$

To close this section, we present a local convergence rate for Algorithm (5.2). Note that when the  $\ell_2$  norm in the definition of generalized averaged nonexpansiveness is replaced by the  $\ell_1$  norm (generalized averaged nonexpansiveness with respect to  $\ell_1$  norm), Proposition 3.14 and Theorem 4.1 still hold with the  $\ell_2$  norms in them is replaced by the  $\ell_1$  norms. Moreover, we have the following theorem.

**Theorem 5.7.** *Suppose that for  $i \in \mathbb{N}_n$ ,  $f_i \in \Gamma_0(\mathbb{R})$  is differentiable with an  $L_i$ -Lipschitz continuous derivative, for some  $L_i \in \mathbb{R}_+$ . If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is given by*

$$f(x) := f_1(x_1) + f_2(x_2) + \dots + f_n(x_n),$$

*$g := \lambda \|\cdot\|_1$ , for  $\lambda \in [0, +\infty)$  and  $\beta \in \left(0, \frac{2}{\max_{i \in \mathbb{N}_n} \{L_i\}}\right)$ , then the following hold:*

- (i) *Operator  $T_2$  is GAN with exponent 1 with respect to  $\|\cdot\|_1$ .*
- (ii) *For any initial vector  $x^0 \in \mathbb{R}^n$ , the sequence  $\{x^k\}$  generated by Algorithm (5.2) converges to a minimizer of  $f + g$  with a local convergence rate  $o(\frac{1}{k})$  with respect to  $\|\cdot\|_1$ .*

*Proof.* We first prove (i). By Example 3.9 and Proposition 5.3 (i), we know that both  $\text{prox}_{\beta\lambda|\cdot|}$  and  $\mathcal{I} - \beta f'_i$  are GAN with exponent 1. This implies that both  $\text{prox}_{\beta g}$  and  $\mathcal{I} - \beta \nabla f$  are GAN with exponent 1 with respect to the  $\ell_1$  norm. Then, by the  $\ell_1$  norm version of Proposition 3.14,  $T_2$  is GAN with exponent 1 with respect to the  $\ell_1$  norm.

Now we conclude from (i) and the  $\ell_1$  norm version of Theorem 4.1 that the fixed-point iteration of  $T_2$  converges to a minimizer of  $f + g$  with the convergence rate  $o\left(\frac{1}{k}\right)$  in terms of  $\|x^{k+1} - x^k\|_1$ , which completes the proof of (ii).  $\square$

Theorem 5.7 establishes the local convergence rate  $o\left(\frac{1}{k}\right)$  with respect to  $\|\cdot\|_1$  for Algorithm (5.2) by employing the generalized averaged nonexpansiveness with exponent 1 with respect to the  $\ell_1$  norm. The same local convergence rate with respect to an inner product norm for Algorithm (5.2) has been shown in Theorem 3 of [22].

## 6. Conclusions

We have introduced the notion of the generalized averaged nonexpansive (GAN) operator, which allows us to study convergence and convergence rates of fixed-point iterations of GAN operators not covered by the existing theory of the averaged nonexpansive operators. The introduced notion provides a unified approach for analyzing the convergence and convergence rates of convex optimization algorithms. The convergence rate results of optimization algorithms obtained from this approach cover existing understanding and lead to new findings.

## Acknowledgements

Yizun Lin was supported in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515110541, by the Fundamental Research Funds for the Central Universities of China under Grant 21620352, by the Opening Project of Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University under Grant 2021006, and by National Natural Science Foundation of China under Grant 62176103. Yuesheng Xu was supported in part by US National Science Foundation under grants DMS-1912958 and DMS-2208386, and by National Natural Science Foundation of China under grant 11771464. All correspondence should be addressed to Y. Xu.

**Data availability statement** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## References

- [1] Agarwal, R.P., Meehan, M., O'regan, D.: Fixed Point Theory and Applications, vol. 141. Cambridge University Press, Cambridge (2001)
- [2] Ahn, S., Fessler, J.A.: Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans. Med. Imaging* **22**(5), 613–626 (2003)
- [3] Baillon, J.-B., Bruck, R.E., Reich, S.: On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces. *Houst. J. Math.* **4**(1), 1–9 (1978)
- [4] Baillon, J.-B., Haddad, G.: Quelques propriétés des opérateurs angle-bornés et  $\epsilon$ -cycliquement monotones. *Isr. J. Math.* **26**(2), 137–150 (1977)
- [5] Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. *SIAM Rev.* **38**(3), 367–426 (1996)
- [6] Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Space*, 2nd edn. Springer, New York (2017)
- [7] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009)
- [8] Bërdëllima, A., Steidl, G.: On  $\alpha$ -firmly nonexpansive operators in  $r$ -uniformly convex spaces. *Results Math.* **76**, 172 (2021)
- [9] Bertsekas, D.P.: On the Goldstein–Levitin–Polyak gradient projection method. *IEEE Trans. Autom. Control* **21**(2), 174–184 (1976)
- [10] Bertsekas, D.P.: *Convex Optimization Algorithms*. Athena Scientific, Belmont (2015)
- [11] Borwein, J.M., Li, G., Tam, M.K.: Convergence rate analysis for averaged fixed point iterations in common fixed point problems. *SIAM J. Optim.* **27**(1), 1–33 (2017)
- [12] Borwein, J.M., Li, G., Yao, L.: Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets. *SIAM J. Optim.* **24**(1), 498–527 (2014)
- [13] Bruck, R.E., Reich, S.: Nonexpansive projections and resolvents of accretive operators in Banach spaces. *Houst. J. Math.* **3**(4), 459–470 (1977)
- [14] Byrne, C.: A unified treatment of some iterative algorithms in signal processing and image reconstruction. *Inverse Probl.* **20**(1), 103–120 (2004)
- [15] Cai, J.-F., Chan, R.H., Shen, Z.: A framelet-based image inpainting algorithm. *Appl. Comput. Harmon. Anal.* **24**(2), 131–149 (2008)

- [16] Cegielski, A.: Iterative Methods for Fixed Point Problems in Hilbert Spaces. Lecture Notes in Mathematics, vol. 2057. Springer, Heidelberg (2012)
- [17] Chan, R.H., Chan, T.F., Shen, L., Shen, Z.: Wavelet algorithms for high-resolution image reconstruction. *SIAM J. Sci. Comput.* **24**(4), 1408–1432 (2003)
- [18] Chen, B., Wang, J., Zhao, H., Zheng, N., Príncipe, J.C.: Convergence of a fixed-point algorithm under maximum correntropy criterion. *IEEE Signal Process. Lett.* **22**(10), 1723–1727 (2015)
- [19] Chen, G.H.-G., Rockafellar, R.T.: Convergence rates in forward-backward splitting. *SIAM J. Optim.* **7**(2), 421–444 (1997)
- [20] Chen, P., Huang, J., Zhang, X.: A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Probl.* **29**(2), 025011 (2013)
- [21] Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **4**(4), 1168–1200 (2005)
- [22] Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. In: *Splitting Methods in Communication, Imaging, Science, and Engineering*, pp. 115–163. Springer, New York (2016)
- [23] Fessler, J.A.: Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Trans. Med. Imaging* **13**(2), 290–300 (1994)
- [24] Figueiredo, M.A.T., Nowak, R.D.: An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.* **12**(8), 906–916 (2003)
- [25] Figueiredo, M.A.T., Nowak, R.D., Wright, S.J.: Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J. Select. Top. Signal Process.* **1**(4), 586–597 (2007)
- [26] Hicks, T.L., Kubicek, J.D.: On the Mann iteration process in a Hilbert space. *J. Math. Anal. Appl.* **59**(3), 498–504 (1977)
- [27] Kazimierz, G., Reich, S.: Uniform Convexity, Hyperbolic Geometry, and Non-expansive Mappings. Marcel Dekker, New York (1984)
- [28] Komodakis, N., Pesquet, J.-C.: Playing with duality: an overview of recent primal-dual approaches for solving large-scale optimization problems. *IEEE Signal Process. Mag.* **32**(6), 31–54 (2015)
- [29] Krol, A., Li, S., Shen, L., Xu, Y.: Preconditioned alternating projection algorithms for maximum a posteriori ECT reconstruction. *Inverse Probl.* **28**(11), 115005 (2012)
- [30] Li, Q., Shen, L., Xu, Y., Zhang, N.: Multi-step fixed-point proximity algorithms for solving a class of optimization problems arising from image processing. *Adv. Comput. Math.* **41**(2), 387–422 (2015)
- [31] Li, Q., Zhang, N.: Fast proximity-gradient algorithms for structured convex optimization problems. *Appl. Comput. Harmon. Anal.* **41**(2), 491–517 (2016)
- [32] Li, Z., Song, G., Xu, Y.: A fixed-point proximity approach to solving the support vector regression with the group lasso regularization. *Int. J. Numer. Anal. Model.* **15**, 154–169 (2018)
- [33] Li, Z., Song, G., Xu, Y.: A two-step fixed-point proximity algorithm for a class of non-differentiable optimization models in machine learning. *J. Sci. Comput.* **81**(2), 923–940 (2019)

- [34] Lin, Y., Schmidtlein, C.R., Li, Q., Li, S., Xu, Y.: A Krasnoselskii–Mann algorithm with an improved EM preconditioner for PET image reconstruction. *IEEE Trans. Med. Imaging* **38**(9), 2114–2126 (2019)
- [35] Lu, J., Shen, L., Xu, C., Xu, Y.: Multiplicative noise removal in imaging: an exp-model and its fixed-point proximity algorithm. *Appl. Comput. Harmon. Anal.* **41**(2), 518–539 (2016)
- [36] Mărușter, L., Mărușter, Ș: Strong convergence of the Mann iteration for  $\alpha$ -demicontractive mappings. *Math. Comput. Model.* **54**(9–10), 2486–2492 (2011)
- [37] Micchelli, C.A., Shen, L., Xu, Y.: Proximity algorithms for image models: denoising. *Inverse Probl.* **27**(4), 045009 (2011)
- [38] Micchelli, C.A., Shen, L., Xu, Y., Zeng, X.: Proximity algorithms for the L1/TV image denoising model. *Adv. Comput. Math.* **38**(2), 401–426 (2013)
- [39] Moreau, J.J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes rendus hebdomadaires des séances de l’Académie des sciences* **255**, 2897–2899 (1962)
- [40] Nesterov, Y.: *Introductory Lectures on Convex Optimization: A Basic Course*, vol. 87. Springer Science & Business Media, New York (2003)
- [41] Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Am. Math. Soc.* **73**(4), 591–597 (1967)
- [42] Parikh, N., Boyd, S.: Proximal algorithms. *Found. Trends Optim.* **1**(3), 127–239 (2014)
- [43] Polson, N.G., Scott, J.G., Willard, B.T.: Proximal algorithms in statistics and machine learning. *Stat. Sci.* **30**(4), 559–581 (2015)
- [44] Reich, S.: Weak convergence theorems for nonexpansive mappings in Banach spaces. *J. Math. Anal. Appl.* **67**(2), 274–276 (1979)
- [45] Reich, S.: Averaged mappings in the Hilbert ball. *J. Math. Anal. Appl.* **109**(1), 199–206 (1985)
- [46] Rockafellar, R.T.: Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**(5), 877–898 (1976)
- [47] Schmidtlein, C.R., Lin, Y., Li, S., Krol, A., Beattie, B.J., Humm, J.L., Xu, Y.: Relaxed ordered subset preconditioned alternating projection algorithm for PET reconstruction with automated penalty weight selection. *Med. Phys.* **44**(8), 4083–4097 (2017)
- [48] Ruder, S.: An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747* (2016)
- [49] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **60**(1–4), 259–268 (1992)
- [50] Shen, L., Xu, Y., Zeng, X.: Wavelet inpainting with the  $\ell_0$  sparse regularization. *Appl. Comput. Harmon. Anal.* **41**(1), 26–53 (2016)
- [51] Song, Y., Chai, X.: Halpern iteration for firmly type nonexpansive mappings. *Nonlinear Anal. Theory Methods Appl.* **71**(10), 4500–4506 (2009)
- [52] Song, Y., Li, Q.: Successive approximations for quasi-firmly type nonexpansive mappings. *Math. Commun.* **16**(1), 251–264 (2011)
- [53] Sra, S., Nowozin, S., Wright, S.J.: *Optimization for Machine Learning*. The MIT Press, Cambridge (2012)
- [54] Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.* **38**(2), 431–446 (2000)

- [55] Zhang, T.: Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the Twenty-First International Conference on Machine Learning*, p. 116 (2004)
- [56] Zheng, W., Li, S., Krol, A., Schmidtlein, C.R., Zeng, X., Xu, Y.: Sparsity promoting regularization for effective noise suppression in SPECT image reconstruction. *Inverse Probl.* **35**(11), 115011 (2019)
- [57] Zhu, Y., Wu, J., Yu, G.: A fast proximal point algorithm for  $\ell_1$ -minimization problem in compressed sensing. *Appl. Math. Comput.* **270**, 777–784 (2015)
- [58] Zorich, V.A.: *Mathematical Analysis I*, 2nd edn. Springer, Berlin (2015)

Yizun Lin

Department of Mathematics, College of Information Science and Technology  
Jinan University  
Guangzhou 510632  
China  
e-mail: linyizun@jnu.edu.cn

Yuesheng Xu

Department of Mathematics and Statistics  
Old Dominion University  
Norfolk VA 23529  
USA  
e-mail: y1xu@odu.edu

Accepted: July 27, 2022.