Annals of Mathematical Sciences and Applications
Volume 8, Number 2, 347–363, 2023

Refinements of some convergence results of the gradient-projection algorithm

Hong-Kun Xu

Dedicated to Professor Anthony To-Ming Lau on the occasion of his 80th birthday

In this note we refine some of the results of [18] on the gradient-projection algorithm in the infinite-dimensional Hilbert space setting by weakening the conditions imposed on the choices of the parameters in [18, Theorems 4.2, 4.3 and 5.2]. In addition, we also show that the relaxed gradient-projection algorithm has a sublinear rate of convergence.

AMS 2000 subject classifications: Primary 90C25, 47H26; secondary 47H09, 49J40.
Keywords and phrases: Gradient-projection, fixed point algorithm, averaged mapping, viscosity approximation method.

1. Introduction

The gradient-projection algorithm (GPA) is one of the most popular methods for solving a constrained minimization problem of the form

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

where C is a nonempty closed convex subset of a Hilbert space H and $f : H \to \mathbb{R}$ is a continuously differentiable convex function. We will use S to denote the set of solutions of (1) and always assume that $S \neq \emptyset$ throughout the rest of this paper.

The GPA is an iteration process that generates a sequence (x_n) by the recursive procedure

$$x_{n+1} = P_C(x_n - \gamma_n \nabla f(x_n)), \quad n \geq 0, \tag{2}$$

where $x_0 \in C$ is an initial guess, $\gamma_n > 0$ is a stepsize, and P_C is the metric projection from H onto C. The convergence (either weak or strong) of

GPA (2) depends on the gradient ∇f and the stepsizes (γ_n). The following convergence result is known.

Theorem 1.1. *[5] Assume further that the gradient ∇f of f satisfies the Lipschitz continuity condition (in this case, f is said to be L-smooth):*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad x, y \in H. \tag{3}$$

Assume also that the stepsize sequence (γ_n) satisfies the condition:

$$0 < \liminf_{n\to\infty} \gamma_n \leq \limsup_{n\to\infty} \gamma_n < \frac{2}{L}. \tag{4}$$

Then the sequence (x_n) generated by GPA (2) converges weakly to a point in S.

An averaged mapping approach to GPA (2) is provided in [18]. In connection with Mann's iteration method [8, 11], the following result was proved in [18].

Theorem 1.2. *[18, Theorems 4.2 and 4.3] Assume that f is convex and L-smooth (i.e., (3) holds). Let a sequence (x_n) be generated by the relaxed gradient-projection algorithm (RGPA):*

$$x_{n+1} = (1 - \alpha_n)x_n + \alpha_n P_C(x_n - \gamma_n \nabla f(x_n)), \quad n = 0, 1, 2, \cdots. \tag{5}$$

Assume that $\{\gamma_n\}$ and $\{\alpha_n\}$ satisfy the condition (4) and the following condition

$$0 < \alpha_n < \frac{4}{2 + \gamma_n L}, \quad 0 < \liminf_{n\to\infty} \alpha_n \leq \limsup_{n\to\infty} \alpha_n < \frac{4}{2 + L \cdot \limsup_{n\to\infty} \gamma_n}. \tag{6}$$

Then the sequence (x_n) converges weakly to a point in S.

If, in addition, the stepsizes $\gamma_n \equiv \gamma \in (0, 2/L)$ for all $n \geq 0$, that is, the algorithm (5) is reduced to

$$x_{n+1} = (1 - \alpha_n)x_n + \alpha_n P_C(x_n - \gamma \nabla f(x_n)), \quad n = 0, 1, 2, \cdots, \tag{7}$$

and the sequence $\{\alpha_n\}$ satisfies the condition

$$\sum_{n=1}^{\infty} \alpha_n \left(\frac{4}{2 + \gamma L} - \alpha_n \right) = \infty, \tag{8}$$

then (x_n) converges weakly to a point in S.

It is known that GPA (2), in an infinite-dimensional setting, has weak convergence only, in general. In order to get strong convergence, a technique, known as viscosity approximation method (VAM), is needed. This method was first introduced by Attouch [1] to convex optimization theory and later extended by Moudafi [9] and Xu [17] to nonexpansive mappings in Hilbert and Banach spaces, respectively. Applying VAM to GPA (2) leads to the algorithm below:

$$x_{n+1} = \theta_n h(x_n) + (1-\theta_n)P_C(x_n - \gamma_n \nabla f(x_n)), \quad n = 0, 1, 2, \cdots, \tag{9}$$

where $\theta_n \in [0, 1]$ for all $n \geq 0$, and $h : C \to C$ is a ρ-contraction with $\rho \in [0, 1)$, i.e.,

$$\|h(x) - h(y)\| \leq \rho\|x - y\| \quad \text{for all } x, y \in C.$$

Note also that VAM is indeed an extension of Halpern's iteration method [2, 4, 6, 7, 13, 14, 15].

Theorem 1.3. *[18, Theorem 5.2] Assume that f is convex and L-smooth (i.e., (3) holds) and let (x_n) be generated by VAM (9). Assume $\{\gamma_n\}$ satisfies the condition (4) and, in addition, the following conditions are satisfied:*

(i) $\theta_n \to 0$;
(ii) $\sum_{n=0}^{\infty} \theta_n = \infty$;
(iii) $\sum_{n=0}^{\infty} |\theta_{n+1} - \theta_n| < \infty$;
(iv) $\sum_{n=0}^{\infty} |\gamma_{n+1} - \gamma_n| < \infty$.

Then (x_n) converges in norm to a point $x^* \in S$ which is the unique solution of the variational inequality (VI)

$$x^* \in S, \quad \langle (I-h)x^*, x - x^*\rangle \geq 0, \quad x \in S. \tag{10}$$

Equivalently, x^* is the unique fixed point of the contraction $P_S h$, i.e., $x^* = (P_S h)x^*$.

The purpose of this note is to refine Theorems 1.2 and 1.3 by weakening the conditions imposed on the parameters (α_n), (θ_n), and (γ_n). More precisely, we will prove the weak convergence of the algorithm RGPA (11) under standard conditions, that is, (γ_n) satisfies (4) and (α_n) satisfies the condition: $\alpha_n \geq \underline{\alpha} > 0$ for all $n \geq 0$. Moreover, we will weaken the conditions in Theorem 1.3 by completely removing the condition (iii) satisfied by (θ_n) and also condition (iv) will be weakened to the condition $\gamma_{n+1} - \gamma_n \to 0$. In addition, we also show that the relaxed gradient-projection algorithm has a sublinear rate of convergence.

2. Preliminaries

Suppose H is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$, respectively, and C is a nonempty closed convex subset of H. A mapping $T : C \to C$ is said to be nonexpansive if $\|Tx - Ty\| \le \|x - y\|$ for all $x, y \in C$. A mapping $V : H \to H$ is said to be α-averaged (α-AV) if for some $\alpha \in (0, 1)$ and another nonexpansive mapping $T : H \to H$ one has

$$V = (1 - \alpha)I + \alpha T.$$

A typical example of an averaged mapping is the (metric) projection $P_C : H \to C$ defined by

$$P_C x = \arg\min\{\|x - y\|^2 : y \in C\}, \quad x \in H.$$

Some useful properties of projections are listed below.

Proposition 2.1. *Given $x \in H$ and $z \in C$.*

(i) $z = P_C x$ *if and only if* $\langle x - z, y - z \rangle \le 0$ *for all* $y \in C$.
(ii) $\langle x - y, P_C x - P_C y \rangle \ge \|P_C x - P_C y\|^2$ *for all* $x, y \in H$. *In particular,* P_C *is* $\frac{1}{2}$*-AV (also known as firmly nonexpansive).*
(iii) $\|x - P_C x\|^2 \le \|x - y\|^2 - \|y - P_C x\|^2$ *for all* $x \in H$ *and* $y \in C$.

The following result is immediately clear, but nevertheless useful.

Lemma 2.1. *If a mapping $V : H \to H$ is α-AV for some $\alpha \in (0, 1)$, then, for each $\alpha' \in [\alpha, 1)$, V is α'-AV.*

Proof. Since V is α-AV, $V = (1 - \alpha)I + \alpha T$, where T is nonexpansive. Now for $1 > \alpha' > \alpha$, set $T' := (1 - \alpha/\alpha')I + (\alpha/\alpha')T$ which is nonexpansive. It is easily seen that $V = (1 - \alpha')I + \alpha' T'$. Consequently, V is α'-AV. □

Lemma 2.2. *Suppose $f : H \to \mathbb{R}$ is convex and L-smooth with $L \ge 0$ (i.e., ∇f is L-Lipschitz). Then, for each $0 < \gamma < \frac{2}{L}$, the mapping $P_C(I - \gamma \nabla f)$ is β-AV with $\beta = \frac{2+\gamma L}{4}$. In other words, there exists a nonexpansive mapping $T : H \to H$ such that*

$$P_C(I - \gamma \nabla f) = (1 - \beta)I + \beta T = \frac{2 - \gamma L}{4} I + \frac{2 + \gamma L}{4} T.$$

Proof. Details of proof can be found in the proof of [18, Theorem 4.1]. □

Lemma 2.3 ([5]). *Suppose $f : H \to \mathbb{R}$ is convex and L-smooth with $L \geq 0$ (i.e., ∇f is L-Lipschitz). Then, for each $x, y \in H$, we have*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2.$$

Lemma 2.4 (Opial's lemma [10]). *Let K be a nonempty subset of a real Hilbert space H. Let $\{x_n\}$ be a bounded sequence in H satisfying the properties:*

(i) $\lim_{n\to\infty} \|x_n - x\|$ *exists for each $x \in K$;*
(ii) $\omega_w(x_n) \subset K$.

Then $\{x_n\}$ is weakly convergent to a point in K.

Here $\omega_w(x_n)$ denotes the set of all accumulation points in the weak topology of the sequence (x_n).

Lemma 2.5. *[12] Suppose (β_n) is a sequence of real numbers in $[0, 1]$ such that*

$$0 < \liminf_{n\to\infty} \beta_n \leq \limsup_{n\to\infty} \beta_n < 1.$$

Let (x_n) and (y_n) be bounded sequences in a Banach space such that $x_{n+1} = (1 - \beta_n)x_n + \beta_n y_n$ for all $n \geq 1$. Suppose, in addition,

$$\limsup_{n\to\infty}(\|y_{n+1} - y_n\| - \|x_{n+1} - x_n\|) \leq 0.$$

Then $\lim_{n\to\infty} \|y_n - x_n\| = 0$. Consequently, one also has $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$.

Lemma 2.6. *[16] Assume $\{a_n\}$ is a sequence of nonnegative real numbers such that*

$$a_{n+1} \leq (1 - \gamma_n)a_n + \gamma_n \delta_n, \quad n \geq 0,$$

where $\{\gamma_n\}$ is a sequence in (0,1) and $\{\delta_n\}$ a sequence in $\mathbb{R}$ satisfying the conditions:

(i) $\sum_{n=1}^{\infty} \gamma_n = \infty$,
(ii) $\limsup_{n\to\infty} \delta_n \leq 0$.

Then $\lim_{n\to\infty} a_n = 0$.

Lemma 2.7 (Demiclosedness Principle; cf. [3]). *Let C be a closed convex subset of a Hilbert space H and let $T : C \to C$ be a nonexpansive mapping with $Fix(T) \neq \emptyset$. If $\{x_n\}$ is a sequence in C such that $x_n \to x$ $x_n - Tx_n \to 0$ strongly, then $x - Tx = 0$, i.e., $x \in Fix(T)$.*

3. Main results

In this section we shall refine Theorems 1.2 and 1.3 by weakening the conditions imposed on the parameters (α_n), (θ_n), and (γ_n). In addition, we also show that the relaxed gradient-projection algorithm has a sublinear rate of convergence.

Theorem 3.1. *Assume that f is convex and L-smooth (i.e., (3) holds). Let a sequence (x_n) be generated by the relaxed gradient-projection algorithm (RGPA):*

$$x_{n+1} = (1-\alpha_n)x_n + \alpha_n P_C(x_n - \gamma_n \nabla f(x_n)), \quad n = 0, 1, 2, \cdots . \tag{11}$$

Assume in addition that

(i) $\{\gamma_n\}$ satisfies (4), i.e., $0 < \liminf_{n\to\infty} \gamma_n \le \limsup_{n\to\infty} \gamma_n < \frac{2}{L}$.
(ii) $(\alpha_n) \subset (0, 1]$ and $\liminf_{n\to\infty} \alpha_n > 0$.

Then (x_n) converges weakly to a point in S.

Proof. Observe from (4) that we may assume that $0 < a \le \gamma_n \le b < 2/L$ for some constants $0 < a \le b < 1$ and all $n \ge 0$. Also by Lemma 2.2, for each n, $P_C(I-\gamma_n\nabla f)$ is β_n-AV with $\beta_n = \frac{2+\gamma_n L}{4}$. Set $\beta = \frac{2+bL}{4}$; then $\beta_n \le \beta < 1$ for all n. By Lemma 2.1, we further get that $P_C(I-\gamma_n\nabla f)$ is β-AV. Therefore, we can write

$$V_n := P_C(I-\gamma_n\nabla f) = (1-\beta)I + \beta T_n = \frac{2-bL}{4} I + \frac{2+bL}{4} T_n, \tag{12}$$

where $T_n : H \to H$ is nonexpansive. We can rewrite x_{n+1} as

$$\begin{aligned} x_{n+1} &= (1-\alpha_n)x_n + \alpha_n V_n x_n \\ &= (1-\alpha_n')x_n + \alpha_n' T_n x_n, \end{aligned}$$

where $\alpha_n' = \alpha_n\beta$. It is easy to find from condition (ii) that

$$0 < \liminf_{n\to\infty} \alpha_n' \le \limsup_{n\to\infty} \alpha_n' \le \beta < 1.$$

Consequently, there exist $0 < \alpha_* \le \alpha^* < 1$ such that

$$\alpha_* \le \alpha_n' \le \alpha^* \tag{13}$$

for all n (large enough). Now take an $x^* \in S$ to get

$$\|x_{n+1} - x^*\|^2 = \|(1-\alpha_n')(x_n - x^*) + \alpha_n'(T_n x_n - x^*)\|^2$$

$$= (1-\alpha_n')\|x_n - x^*\|^2 + \alpha_n'\|T_nx_n - x^*\|^2$$
$$- \alpha_n'(1-\alpha_n')\|x_n - T_nx_n\|^2$$
$$\le \|x_n - x^*\|^2 - \alpha_*(1-\alpha^*)\|x_n - T_nx_n\|^2. \tag{14}$$

It turns out that $\|x_{n+1} - x^*\| \le \|x_n - x^*\|$, hence,

$$\lim_{n\to\infty} \|x_n - x^*\| \quad \text{for each } x^* \in S. \tag{15}$$

Since we also have $\alpha_*(1-\alpha^*)\|x_n - T_nx_n\|^2 \le \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2$. This immediately implies that $\lim_{n\to\infty} \|x_n - T_nx_n\|^2 = 0$ which in turns implies that $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$.

Next we show

$$\omega_w(x_n) \subset S. \tag{16}$$

To see this, we take $x' \in \omega_w(x_n)$ and assume $x_{n'} \to x'$ weakly for some subsequence $(x_{n'})$ of (x_n). With no loss of generality, we may assume $\gamma_{n'} \to \gamma' \in (0, 2/L)$. Set $V' = P_C(I - \gamma'\nabla f)$. Notice that V' is nonexpansive and $Fix(V') = S$. It turns out that

$$\begin{aligned}
&\|x_{n'} - V'x_{n'}\| \\
&\le \|x_{n'} - V_{n'}x_{n'}\| + \|V_{n'}x_{n'} - V'x_{n'}\| \\
&\le \|x_{n'} - x_{n'+1}\| + \|x_{n'+1} - V_{n'}x_{n'}\| \\
&\quad + \|P_C(I - \gamma_{n'}\nabla f)x_{n'} - P_C(I - \gamma'\nabla f)x_{n'}\| \\
&\le \|x_{n'} - x_{n'+1}\| + \theta_{n'}\|h(x_{n'}) - V_{n'}x_{n'}\| + |\gamma_{n'} - \gamma'|\|\nabla f(x_{n'})\| \\
&\le \|x_{n'} - x_{n'+1}\| + 2M(\theta_{n'} + |\gamma_{n'} - \gamma'|) \to 0.
\end{aligned}$$

Thus, the demiclosedness principle of nonexpansive mappings (i.e., Lemma 2.7) asserts that $x' \in Fix(V') = S$; hence, $\omega_w(x_n) \subset S$.

By virtue of (15) and (16), Lemma 2.4 is applicable to the sequence (x_n) and the set S. Consequently, (x_n) converges weakly to a point of S. □

The following is a straightforward consequence of Theorem 3.1.

Corollary 3.1. *Assume that f is convex and L-smooth (i.e., (3) holds). Let a sequence (x_n) be generated by the gradient-projection algorithm (GPA):*

$$x_{n+1} = P_C(x_n - \gamma_n\nabla f(x_n)), \quad n = 0, 1, 2, \cdots.$$

Assume $\{\gamma_n\}$ satisfies (4), i.e., $0 < \liminf_{n\to\infty} \gamma_n \le \limsup_{n\to\infty} \gamma_n < \frac{2}{L}$. Then (x_n) converges weakly to a point in S.

Proof. In Theorem 3.1, take $\alpha_n = 1$ for all $n \geq 0$. □

Remark 3.1. *We observe that the choices of the parameter sequences* (γ_n) *and* (α_n) *in Theorem 3.1 are decoupled; moreover, the choice of* (α_n) *is irrelevant to the Lipschitz constant* L *of* ∇f*, as opposed to the condition* (6) *of Theorem 3.1.*

Theorem 3.2. *Assume that* f *is convex and* L*-smooth (i.e.,* (3) *holds) and let* (x_n) *be generated by VAM* (9)*. Assume the following conditions are satisfied:*

(a) $0 < \liminf_{n\to\infty} \gamma_n \leq \limsup_{n\to\infty} \gamma_n < \frac{2}{L}$*, i.e., (4) holds.*
(b) $\gamma_{n+1} - \gamma_n \to 0$.
(c) $\theta_n \to 0$.
(d) $\sum_{n=0}^{\infty} \theta_n = \infty$.

Then (x_n) *converges in norm to the unique solution* x^* *of VI* (10)*. Equivalently,* x^* *is the unique fixed point of the contraction* $P_S h$*, i.e.,* $x^* = (P_S h)x^*$*.*

Proof. First we show that (x_n) is bounded. As a matter of fact, since $P_C(I - \gamma_n \nabla f)$ is nonexpansive with S as its fixed point set, it follows from (9) that, for $\bar{x} \in S$,

$$\begin{aligned}
&\|x_{n+1} - \bar{x}\| \\
&= \|\theta_n[h(x_n) - h(\bar{x}) + h(\bar{x}) - \bar{x}] + (1-\theta_n)[P_C(I - \gamma_n \nabla f)x_n - \bar{x}]\| \\
&\leq \theta_n(\rho\|x_n - \bar{x}\| + \|h(\bar{x}) - \bar{x}\|) + (1-\theta_n)\|x_n - \bar{x}\| \\
&= (1 - (1-\rho)\theta_n)\|x_n - \bar{x}\| + \theta_n\|h(\bar{x}) - \bar{x}\| \\
&\leq \max\left\{\|x_n - \bar{x}\|, (1-\rho)^{-1}\|h(\bar{x}) - \bar{x}\|\right\}.
\end{aligned}$$

Hence, an induction argument shows that

$$\|x_n - \bar{x}\| \leq \max\left\{\|x_0 - \bar{x}\|, \frac{1}{1-\rho}\|h(\bar{x}) - \bar{x}\|\right\}, \quad n \geq 0.$$

In particular, (x_n) is bounded. Let M be a constant such that

$$M \geq \max\{\|x_n\|, \|h(x_n)\|, \|\nabla f(x_n)\|, \|T_m x_n\|, \|P_C(x_n - \gamma_n \nabla f(x_n))\|\}$$

for all $m, n \geq 0$.

Observe that (12) remains valid, due to condition (a). We can therefore rewrite the algorithm (9) as

$$x_{n+1} = \theta_n h(x_n) + (1-\beta)(1-\theta_n)x_n + \beta(1-\theta_n)T_n x_n$$

$$= (1-(\beta+(1-\beta)\theta_n))x_n + \theta_n h(x_n) + \beta(1-\theta_n)T_n x_n$$
$$= (1-\tau_n)x_n + \tau_n y_n,$$

where $\tau_n = \beta + (1-\beta)\theta_n$ and

$$y_n = \frac{\theta_n h(x_n) + \beta(1-\theta_n)}{\tau_n} T_n x_n = \frac{\theta_n}{\tau_n} h(x_n) + \frac{\beta(1-\theta_n)}{\tau_n} T_n x_n.$$

It follows that

$$\begin{aligned} y_{n+1} - y_n &= \frac{\theta_{n+1}}{\tau_{n+1}} h(x_{n+1}) \\ &\quad + \frac{\beta(1-\theta_{n+1})}{\tau_{n+1}} T_{n+1}x_{n+1} - \left(\frac{\theta_n}{\tau_n} h(x_n) + \frac{\beta(1-\theta_n)}{\tau_n} T_n x_n\right) \\ &= \frac{\theta_{n+1}}{\tau_{n+1}} h(x_{n+1}) - \frac{\theta_n}{\tau_n} h(x_n) \\ &\quad + \frac{\beta(1-\theta_{n+1})}{\tau_{n+1}} (T_{n+1}x_{n+1} - T_{n+1}x_n) \\ &\quad + \frac{\beta(1-\theta_{n+1})}{\tau_{n+1}} (T_{n+1}x_n - T_n x_n) \\ &\quad + \left(\frac{\beta(1-\theta_{n+1})}{\tau_{n+1}} - \frac{\beta(1-\theta_n)}{\tau_n}\right) T_n x_n. \end{aligned} \tag{17}$$

Again using (12), we get

$$P_C(I - \gamma_{n+1}\nabla f)x_n = (1-\beta)x_n + \beta T_{n+1}x_n,$$
$$P_C(I - \gamma_n \nabla f)x_n = (1-\beta)x_n + \beta T_n x_n.$$

Hence,

$$\begin{aligned} \|T_{n+1}x_n - T_n x_n\| &= \|P_C(I-\gamma_{n+1}\nabla f)x_n - P_C(I-\gamma_n\nabla f)x_n\|/\beta \\ &\le |\gamma_{n+1}-\gamma_n|\|\nabla f(x_n)\|/\beta \le (M/\beta)|\gamma_{n+1}-\gamma_n| \to 0. \end{aligned} \tag{18}$$

Combining (17) and (18) we obtain

$$\begin{aligned} &\|y_{n+1} - y_n\| - \|x_{n+1} - x_n\| \\ &\quad \le M\left(\frac{\theta_{n+1}}{\tau_{n+1}} + \frac{\theta_n}{\tau_n}\right) + \left(\frac{\beta(1-\theta_{n+1})}{\tau_{n+1}} - 1\right)\|x_{n+1} - x_n\| \\ &\qquad + M\left\{\frac{(1-\theta_{n+1})}{\tau_{n+1}}|\gamma_{n+1}-\gamma_n| + \left|\frac{\beta(1-\theta_{n+1})}{\tau_{n+1}} - \frac{\beta(1-\theta_n)}{\tau_n}\right|\right\}. \end{aligned} \tag{19}$$

Now noticing the facts that $\theta_n \to 0$ and $\tau_n \to \beta \in (0,1)$, we immediately get from (19) and the assumption $|\gamma_{n+1} - \gamma_n| \to 0$ that

$$\limsup_{n\to\infty}(\|y_{n+1} - y_n\| - \|x_{n+1} - x_n\|) \le 0.$$

Consequently, by Lemma 2.5 we obtain $\lim_{n\to\infty} \|y_n - x_n\| = 0$ and also $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$.

Next, we prove $\omega_w(x_n) \subset S$. Let $\hat{x} \in \omega_w(x_n)$ and assume $x_{n_j} \rightharpoonup \hat{x}$ for some subsequence (x_{n_j}) of (x_n). With no loss of generality, we may assume $\gamma_{n_j} \to \gamma \in (0, 2/L)$, due to condition (a), i.e., (4). Set $V_n = P_C(I - \gamma_n \nabla f)$ and $V = P_C(I - \gamma\nabla f)$. Notice that V is nonexpansive and $Fix(V) = S$. It turns out that

$$\begin{aligned}
\|x_{n_j} - Vx_{n_j}\| &\le \|x_{n_j} - V_{n_j}x_{n_j}\| + \|V_{n_j}x_{n_j} - Vx_{n_j}\| \\
&\le \|x_{n_j} - x_{n_j+1}\| + \|x_{n_j+1} - V_{n_j}x_{n_j}\| \\
&\quad + \|P_C(I - \gamma_{n_j}\nabla f)x_{n_j} - P_C(I - \gamma\nabla f)x_{n_j}\| \\
&\le \|x_{n_j} - x_{n_j+1}\| + \theta_{n_j}\|h(x_{n_j}) - V_{n_j}x_{n_j}\| + |\gamma_{n_j} - \gamma|\|\nabla f(x_{n_j})\| \\
&\le \|x_{n_j} - x_{n_j+1}\| + 2M(\theta_{n_j} + |\gamma_{n_j} - \gamma|) \to 0.
\end{aligned}$$

Thus, the demiclosedness principle (Lemma 2.7) of nonexpansive mappings asserts that $\hat{x} \in Fix(V) = S$; hence, $\omega_w(x_n) \subset S$.

Now let x^* be the unique solution of VI (10) and we show $x_n \to x^*$ in norm. To see this, we claim

$$\limsup_{n\to\infty}\langle h(x^*) - x^*, x_n - x^*\rangle \le 0. \tag{20}$$

As a matter of fact, we can take a subsequence (x_{n_j}) of (x_n) such that

$$\limsup_{n\to\infty}\langle h(x^*) - x^*, x_n - x^*\rangle = \lim_{j\to\infty}\langle h(x^*) - x^*, x_{n_j} - x^*\rangle.$$

With no loss of generality, we may further assume $x_{n_j} \to \hat{x}$ weakly; then $\hat{x} \in S$ as proved above. Now since x^* solves VI (10), it is immediately clear that

$$\limsup_{n\to\infty}\langle h(x^*) - x^*, x_n - x^*\rangle = \langle h(x^*) - x^*, \hat{x} - x^*\rangle \le 0$$

and (20) is proven.

We are finally in a position to prove $x_n \to x^*$ in norm. Recalling that $V_n = P_C(I - \gamma_n\nabla f)$ and $x_{n+1} = \theta_n h(x_n) + (1 - \theta_n)V_n x_n$, we have

$$\|x_{n+1} - x^*\|^2 = \|\theta_n(h(x_n) - x^*) + (1 - \theta_n)(V_n x_n - x^*)\|^2$$

$$
\begin{aligned}
&= \|\theta_n(h(x_n) - h(x^*)) + (1-\theta_n)(V_n x_n - x^*) + \theta_n(h(x^*) - x^*)\|^2 \\
&\le \|\theta_n(h(x_n) - h(x^*)) + (1-\theta_n)(V_n x_n - x^*)\|^2 \\
&\quad + 2\theta_n\langle h(x^*) - x^*, x_{n+1} - x^*\rangle \\
&\le \theta_n\|h(x_n) - h(x^*)\|^2 + (1-\theta_n)\|V_n x_n - x^*\|^2 \\
&\quad + 2\theta_n\langle h(x^*) - x^*, x_{n+1} - x^*\rangle \\
&\le (1-(1-\rho^2)\theta_n)\|x_n - x^*\|^2 + 2\theta_n\langle h(x^*) - x^*, x_{n+1} - x^*\rangle. \qquad (21)
\end{aligned}
$$

Due to condition (d) and (20), Lemma 2.6 is applicable to (21) to get $\|x_n - x^*\| \to 0$. This completes the proof. □

It is interesting to know if condition (b) can be removed.

When the contraction h is taken to be constant, we get the following result.

Corollary 3.2. *Assume that f is convex and L-smooth (i.e., (3) holds) and let (x_n) be generated by the following Halpern iteration method:*

$$(22) \qquad x_{n+1} = \theta_n u + (1-\theta_n)P_C(x_n - \gamma_n\nabla f(x_n)), \quad n = 0, 1, 2, \cdots,$$

where $u \in C$ is a fixed point in C (commonly referred to as anchor). Assume the condition (a)-(c) of Theorem 3.2 are satisfied. Then (x_n) converges in norm to $P_S u$.

Remark 3.2. *The conditions of Theorem 3.2 are much weaker than the conditions of Theorem 1.3. For instance, condition (iii) of Theorem 1.3 is completely removed and condition (iv) is weakened to condition (b) (i.e., $\gamma_{n+1} - \gamma_n \to 0$) of Theorem 3.2.*

Note also that conditions (c) and (d) are standard and necessary conditions for Halpern's iteration method to converge in norm [4].

Finally we discuss the convergence rate of RGPA (11). Below we show that RGPA (11) has a sublinear rate of convergence.

Theorem 3.3. *The RGPA (11) has at least a sublinear rate of convergence. More precisely, we have the estimate:*

$$(23) \qquad f(x_n) - f(x^*) \le \frac{1}{n}\left(\frac{f(x_0) - f(x^*)}{\underline{\alpha}} + \tau\|x_0 - x^*\|^2\right), \quad n \ge 1,$$

where $x^* \in S$ and $\tau = \tau(\underline{\alpha}, \overline{\alpha}, \underline{\gamma}, \overline{\gamma}, L) > 0$ is a constant. Here L is the Lipschitz constant of ∇f, $\underline{\alpha} := \inf_{n\ge0}\alpha_n > 0$, $\overline{\alpha} := \sup_{n\ge0}\alpha_n \le 1$, $\underline{\gamma} := \inf_{n\ge0}\gamma_n > 0$, and $\overline{\gamma} := \sup_{n\ge0}\gamma_n < \frac{2}{L}$.

Proof. By (11), we get

$$\frac{x_{n+1}-x_n}{\alpha_n}+x_n=P_C(x_n-\gamma_n\nabla f(x_n)).$$

It turns out from Proposition 2.1(i) that

$$-\left\langle \gamma_n\nabla f(x_n)+\frac{x_{n+1}-x_n}{\alpha_n}, y-\frac{x_{n+1}-x_n}{\alpha_n}-x_n\right\rangle\le 0\quad \forall y\in C. \tag{24}$$

In particular, replacing the y in (24) with x_n yields

$$\langle\nabla f(x_n), x_{n+1}-x_n\rangle\le-\frac{\|x_{n+1}-x_n\|^2}{\alpha_n\gamma_n}. \tag{25}$$

On the other hand, taking $y:=x^*\in S$ in (24), we obtain

$$\begin{aligned}
&-\langle\nabla f(x_n), x^*-x_n\rangle+\frac{1}{\alpha_n}\langle\nabla f(x_n), x_{n+1}-x_n\rangle\\
&\quad-\frac{1}{\alpha_n\gamma_n}\langle x_{n+1}-x_n, x^*-x_n\rangle+\frac{\|x_{n+1}-x_n\|^2}{\alpha_n^2\gamma_n}\le 0.
\end{aligned} \tag{26}$$

We have from (26) and Lemma 2.3

$$\begin{aligned}
f(x^*)-f(x_{n+1})&\ge f(x_n)+\langle\nabla f(x_n), x^*-x_n\rangle\\
&\qquad-[f(x_n)+\langle\nabla f(x_n), x_{n+1}-x_n\rangle+\frac{L}{2}\|x_{n+1}-x_n\|^2]\\
&=\langle\nabla f(x_n), x^*-x_{n+1}\rangle-\frac{L}{2}\|x_{n+1}-x_n\|^2\\
&=\langle\nabla f(x_n), x^*-x_n\rangle+\langle\nabla f(x_n), x_n-x_{n+1}\rangle-\frac{L}{2}\|x_{n+1}-x_n\|^2\\
&\ge\frac{1}{\alpha_n}\langle\nabla f(x_n), x_{n+1}-x_n\rangle-\frac{1}{\alpha_n\gamma_n}\langle x_{n+1}-x_n, x^*-x_n\rangle\\
&\quad+\frac{\|x_{n+1}-x_n\|^2}{\alpha_n^2\gamma_n}+\langle\nabla f(x_n), x_n-x_{n+1}\rangle-\frac{L}{2}\|x_{n+1}-x_n\|^2\\
&=\left(\frac{1}{\alpha_n}-1\right)\langle\nabla f(x_n), x_{n+1}-x_n\rangle-\frac{1}{\alpha_n\gamma_n}\langle x_{n+1}-x_n, x^*-x_n\rangle\\
&\quad-\left(\frac{L}{2}-\frac{1}{\alpha_n^2\gamma}\right)\|x_{n+1}-x_n\|^2.
\end{aligned} \tag{27}$$

Since

$$\langle x_{n+1} - x_n, x^* - x_n \rangle = -\frac{1}{2}(\|x_{n+1} - x^*\|^2 - \|x_{n+1} - x_n\|^2 - \|x_n - x^*|^2),$$

we obtain from (27) that

$$\begin{aligned} f(x^*) - f(x_{n+1}) \geq & \left(\frac{1}{\alpha_n} - 1\right) \langle \nabla f(x_n), x_{n+1} - x_n \rangle \\ & + \frac{1}{2\alpha_n \gamma_n}(\|x_{n+1} - x^*\|^2 - \|x_{n+1} - x_n\|^2 - \|x_n - x^*|^2) \\ & - \left(\frac{L}{2} - \frac{1}{\alpha_n^2 \gamma_n}\right) \|x_{n+1} - x_n\|^2. \end{aligned}$$

Consequently,

$$\begin{aligned} f(x_{n+1}) - f(x^*) \leq & \left(\frac{1}{\alpha_n} - 1\right) \langle \nabla f(x_n), x_n - x_{n+1} \rangle \\ & + \frac{1}{2\alpha_n \gamma_n}(\|x_n - x^*\|^2 - \|x_{n+1} - x^*|^2) \\ & + \left(\frac{L}{2} - \frac{1}{\alpha_n^2 \gamma_n} + \frac{1}{2\alpha_n \gamma_n}\right) \|x_{n+1} - x_n\|^2. \end{aligned} \tag{28}$$

Substituting into (28) the inequality

$$\langle \nabla f(x_n), x_n - x_{n+1} \rangle \leq f(x_n) - f(x_{n+1}) + \frac{L}{2}\|x_{n+1} - x_n\|^2$$

which holds by Lemma 2.3, we further arrive at

$$\begin{aligned} f(x_{n+1}) - f(x^*) \leq & \left(\frac{1}{\alpha_n} - 1\right) [f(x_n) - f(x_{n+1}) + \frac{L}{2}\|x_{n+1} - x_n\|^2] \\ & + \frac{1}{2\alpha_n \gamma_n}(\|x_n - x^*\|^2 - \|x_{n+1} - x^*|^2) \\ & + \left(\frac{L}{2} - \frac{1}{\alpha_n^2 \gamma_n} + \frac{1}{2\alpha_n \gamma_n}\right) \|x_{n+1} - x_n\|^2. \end{aligned}$$

Hence by multiplying both sides by α_n, we get

$$\begin{aligned} \alpha_n[f(x_{n+1}) - f(x^*)] \leq & (1 - \alpha_n)[f(x_n) - f(x_{n+1}) + \frac{L}{2}\|x_{n+1} - x_n\|^2] \\ & + \frac{1}{2\gamma_n}(\|x_n - x^*\|^2 - \|x_{n+1} - x^*|^2) \end{aligned}$$

$$+\left(\frac{\alpha_n L}{2}-\frac{1}{\alpha_n\gamma_n}+\frac{1}{2\gamma_n}\right)\|x_{n+1}-x_n\|^2$$
$$=(1-\alpha_n)[f(x_n)-f(x_{n+1})]+\frac{1}{2\gamma_n}(\|x_n-x^*\|^2-\|x_{n+1}-x^*|^2)$$
$$+\left(\frac{L}{2}-\frac{1}{\alpha_n\gamma_n}+\frac{1}{2\gamma_n}\right)\|x_{n+1}-x_n\|^2.$$

It follows that

$$f(x_{n+1})-f(x_n)+\alpha_n[f(x_n)-f(x^*)]$$
$$\le\frac{1}{2\gamma_n}(\|x_n-x^*\|^2-\|x_{n+1}-x^*|^2)+\left(\frac{L}{2}-\frac{1}{\alpha_n\gamma_n}+\frac{1}{2\gamma_n}\right)\|x_{n+1}-x_n\|^2.$$

Setting $d_1 := 1/(2\overline{\lambda})$ and $d_2 := (L/2)+(1/2\underline{\gamma})$, we get, for each $j\ge 0$,

$$\begin{aligned}&f(x_{j+1})-f(x_j)+\underline{\alpha}[f(x_j)-f(x^*)]\\ &\le d_1(\|x_j-x^*\|^2-\|x_{j+1}-x^*|^2)+d_2\|x_{j+1}-x_j\|^2.\end{aligned}\tag{29}$$

Summing up from $j=0$ to $j=n-1$ and then divided by n yields

$$\begin{aligned}&\underline{\alpha}\left(\frac{1}{n}\sum_{j=0}^{n-1}f(x_j)-f(x^*)\right)+\frac{1}{n}\left(f(x_n)-f(x_0)\right)\\ &\le\frac{d_1}{n}(\|x_0-x^*\|^2-\|x_n-x^*|^2)+\frac{d_2}{n}\sum_{j=0}^{n-1}\|x_{j+1}-x_j\|^2.\end{aligned}\tag{30}$$

Since, by Lemma 2.3, $f(x_{j+1})\le f(x_j)+\langle\nabla f(x_j),x_{j+1}-x_j\rangle+\frac{L}{2}\|x_{j+1}-x_j\|^2$, we have by (25),

$$\begin{aligned}f(x_j)-f(x_{j+1})&\ge\langle\nabla f(x_j),x_j-x_{j+1}\rangle-\frac{L}{2}\|x_{j+1}-x_j\|^2\\ &\ge\left(\frac{1}{\alpha_j\gamma_j}-\frac{L}{2}\right)\|x_{j+1}-x_j\|^2\ge 0\end{aligned}$$

since $\alpha_j\gamma_j\le\gamma_j<2/L$ for all j. That is, $\{f(x_j)\}$ is nonincreasing. As a result, we find that (30) implies

$$\underline{\alpha}[f(x_n)-f(x^*)]+\frac{1}{n}\left(f(x_n)-f(x_0)\right)$$

$$(31) \qquad \leq \frac{d_1}{n}\|x_0 - x^*\|^2 + \frac{d_2}{n}\sum_{j=0}^{n-1}\|x_{j+1} - x_j\|^2.$$

By (14), we get

$$\begin{aligned}\|x_{n+1} - x_n\|^2 = (\alpha_n')^2\|T_n x_n - x_n\|^2 \leq \beta^2\|T_n x_n - x_n\|^2 \\ \leq \frac{\beta}{\underline{\alpha}(1-\overline{\alpha}\beta)}(\|x_n - x^*\|^2 - x_{n+1} - x^*\|^2)\end{aligned}$$

with $\beta = \frac{2+\overline{\gamma}L}{4} < 1$ (as $\overline{\gamma} < 2/L$). Hence,

$$\frac{1}{n}\sum_{j=0}^{n-1}\|x_{j+1} - x_j\|^2 \leq \frac{\beta}{\underline{\alpha}(1-\overline{\alpha}\beta)n}(\|x_0 - x^*\|^2 - x_n - x^*\|^2).$$

Substituting this into (31), we obtain

$$\begin{aligned}f(x_n) - f(x^*) &\leq \frac{1}{n}\left\{\frac{f(x_0) - f(x_n)}{\underline{\alpha}} + \frac{\|x_0 - x^*\|^2}{\underline{\alpha}}\left(d_1 + \frac{d_2\beta}{\underline{\alpha}(1-\overline{\alpha}\beta)}\right)\right\} \\ &\leq \frac{1}{n}\left\{\frac{f(x_0) - f(x^*)}{\underline{\alpha}} + \|x_0 - x^*\|^2\left(\frac{d_1}{\underline{\alpha}} + \frac{d_2\beta}{\underline{\alpha}^2(1-\overline{\alpha}\beta)}\right)\right\}.\end{aligned}$$

Consequently, (23) holds with $\tau = (d_1/\underline{\alpha}) + d_2\beta/(\underline{\alpha}^2(1-\overline{\alpha}\beta))$. The proof is complete. □

Acknowledgements

The author was grateful to the referee for his/her suggestion to include a convergence rate discussion for the RGPA (11), which resulted in Theorem 3.3 and strengthened the presentation of this manuscript.

References

[1] H. Attouch, Viscosity approximation methods for minimization problems, *SIAM J. Optim.* **6** (1996), no. 3, 769–806. MR1402205

[2] C. E. Chidume and C. O. Chidume, Iterative approximation of fixed points of nonexpansive mappings, *J. Math. Anal. Appl.* **318** (2006), 288–295. MR2210888

[3] K. Geobel and W. A. Kirk, *Topics in Metric Fixed Point Theory*, Cambridge Studies in Advanced Mathematics, vol. 28, Cambridge University Press, 1990. MR1074005

[4] B. R. Halpern, Fixed points of nonexpanding maps, *Bull. Amer. Math. Soc.* **73** (1967) 957–961. MR0218938

[5] E. S. Levitin and B. T. Polyak, Constrained minimization methods, *Zh. vychisl. Mat. mat. Fiz.* **6** (1966). 787–823, 1966. MR0211590

[6] P.-L. Lions, Approximation de points fixes de contractions, *C. R. Acad. Sci. Paris Ser. A-B* **284** (1977), A1357–A1359. MR0470770

[7] G. Lopez, V. Martin and H. K. Xu, Halpern's iteration for nonexpansive mappings, in "Nonlinear Analysis and Optimization I: Nonlinear Analysis," *Contemporary Mathematics* **513** (2010), 211–230. MR2668248

[8] W. R. Mann, Mean value methods in iteration, *Proc. Amer. Math. Soc.* **4** (1953), 506–510. MR0054846

[9] A. Moudafi, Viscosity approximation methods for fixed-points problems, *J. Math. Anal. Appl.* **241** (2000), 46–55. MR1738332

[10] Z. Opial, Weak convergence of the sequence of successive approximations of nonexpansive mappings, *Bull. Amer. Math. Soc.* **73** (1967), 595–597. MR0211301

[11] S. Reich, Weak convergence theorems for nonexpansive mappings in Banach spaces, *J. Math. Anal. Appl.* **67** (1979), 274–276. MR0528688

[12] T. Suzuki, Strong convergence theorems for infinite families of nonexpansive mappings in general Banach spaces, *Fixed Point Theory Appl.* **2005** (2005), 103–123. MR2172156

[13] T. Suzuki, A sufficient and necessary condition for Halpern-type strong convergence to fixed points of nonexpansive mappings, *Proc. Amer. Math. Soc.* **135** (2007), 99–106. MR2280199

[14] T. Suzuki, Reich's problem concerning Halpern's convergence, *Arch. Math.* **92** (2009), 602–613. MR2516166

[15] R. Wittmann, Approximation of fixed points of nonexpansive mappings, *Arch. Math.* **58** (1992), 486–491. MR1156581

[16] H. K. Xu, Iterative algorithms for nonlinear operators, *J. London Math. Soc.* **66** (2002), 240–256. MR1911872

[17] H. K. Xu, Viscosity approximation methods for nonexpansive mappings, *J. Math. Anal. Appl.* **298** (2004), 279–291. MR2086546

[18] H. K. Xu, Averaged mappings and the gradient-projection algorithm, *J. Optim. Theory Appl.* **150** (2011), 360–378. MR2818926

Hong-Kun Xu
School of Science
Hangzhou Dianzi University
Hangzhou
310018
China
E-mail address: xuhk@hdu.edu.cn

Received May 31, 2023