Journal of Nonlinear and Convex Analysis
Volume 14, Number 2, 2013, 245–251

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

STRONG CONVERGENCE OF THE GRADIENT-PROJECTION ALGORITHM IN HILBERT SPACES

HUANHUAN CUI AND FENGHUI WANG

ABSTRACT. In this paper we study a modified Gradient-Projection Algorithm recently introduced by Xu (Averaged mappings and the Gradient-Projection Algorithm, J. Optim. Theory Appl., 150 (2011), 360–378). The strong convergence of the algorithm is obtained under some weak conditions.

1. Introduction

Let $\mathcal{H}$ be a real Hilbert space, C a nonempty closed convex subset, and $f : \mathcal{H} \to \mathbb{R}$ a convex and differentiable function. Consider a convexly constrained minimization problem: Find $x^* \in C$ so that

$$f(x^*) = \min_{x \in C} f(x). \tag{1.1}$$

A classical way to solve such problem is the gradient projection algorithm(GPA): For any initial guess x_0, the GPA generates a sequence as

$$x_{n+1} = P_C(x_n - r\nabla f(x_n)),$$

where ∇f denotes the differential of f and $r > 0$ is known as the step of the algorithm. If we set $T = P_C(I - r\nabla f)$, then problem (1.1) is transformed into the fixed point problem:

$$\text{Find } x^* \in C \text{ so that } Tx^* = x^*;$$

accordingly the GPA has the form:

$$x_{n+1} = Tx_n.$$

If further ∇f is L-Lipschitz continuous, namely

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \forall x, y \in \mathcal{H},$$

and the step is chosen so that $0 < r < (2/L)$, then T is an averaged mapping. By the well-known Mann's theorem, the GPA converges weakly to a fixed point of T, i.e., a minimizer of prblem (1.1), whenever such an element exists.

However, Hundal's counterexample [6] reveals that the GPA can not converge strongly in general. To get the strongly convergent algorithm, Xu [10] recently introduced the following modified scheme: For any initial guess $x_0 \in C$, define

$$x_{n+1} = \theta_n h(x_n) + (1 - \theta_n)P_C(x_n - r_n\nabla f(x_n)), \tag{1.2}$$

2010 *Mathematics Subject Classification.* 47H09, 47H10.

Key words and phrases. Gradient projection algorithm, inverse-strongly monotone operator, averaged mappings.

This work is supported by Science Foundation of Department of Education, Henan (2011B110023, 12A110016) and Peiyu Foundation of Luoyang Normal University (2011-PYJJ-002).

where $(\theta_n) \subseteq [0,1], h : C \to C$ is a contractive mapping. Then algorithm (1.2) can be strongly convergent to a solution of problem (1.1) provided that

(i) $0 < \liminf_n r_n \leq \limsup_n r_n < (2/L)$,
(ii) $\lim_n \theta_n = 0, \sum \theta_n = \infty$,
(iii) $\sum |\theta_n - \theta_{n+1}| < \infty, \sum |r_n - r_{n+1}| < \infty$.

In what follows we shall prove that condition (i)-(ii) is sufficient to guarantee the strong convergence of the above algorithm.

2. Preliminary

Assume that C is a nonempty convex closed subset in $\mathcal{H}$. $\mathrm{Fix}(T)$ will denote the fixed point set of T, $\omega_w(x_n)$ the weak cluster points set of the sequence (x_n), $\to$ strong convergence, $\rightharpoonup$ weak convergence, and Ω the solution set of problem (1.1).

Definition 2.1. Let $T : C \to \mathcal{H}$ be a nonlinear mapping. Then

(1) T is said to be a ρ-contraction, if there exists $\rho \in (0,1)$ so that for $x, y \in C$,

$$\|Tx - Ty\| \leq \rho \|x - y\|.$$

(2) T is said to be a nonexpansive mapping, if for $x, y \in C$,

$$\|Tx - Ty\| \leq \|x - y\|.$$

(3) T is said to be an α-averaged mapping, if there is $\alpha \in (0,1)$ and a nonexpansive mapping S so that $T = (1-\alpha)I + \alpha S$.

(4) T is said to be a κ-inverse strongly monotone mapping, if there is $\kappa > 0$ so that

$$\langle Tx - Ty, x - y \rangle \geq \kappa \|Tx - Ty\|^2.$$

Lemma 2.2 (Demiclosedness Principle [5])**.** *Let C be a nonempty closed convex subset of $\mathcal{H}$ and $T : C \to \mathcal{H}$ a nonexpansive mapping with* $\mathrm{Fix}(T) \neq \emptyset$. *If (x_n) is a sequence in C such that $x_n \rightharpoonup x$ and $(I - T)x_n \to 0$, then $(I - T)x = 0$, i.e., $x \in \mathrm{Fix}(T)$.*

Averaged mappings have the following properties.

Lemma 2.3 (see [3, 10])**.** *(1) If $T : C \to \mathcal{H}$ is α-averaged, then for any $z \in \mathrm{Fix}(T)$ and for all $x \in C$,*

$$\|Tx - z\|^2 \leq \|x - z\|^2 - \frac{1-\alpha}{\alpha}\|Tx - x\|^2.$$

(2) Let $T_1 : \mathcal{H} \to \mathcal{H}$ and $T_2 : C \to \mathcal{H}$ be α_1 and α_2-averaged, respectively. Then T_1T_2 is $(\alpha_1 + \alpha_2 - \alpha_1\alpha_2)$-averaged.

For $x \in \mathcal{H}$, its projection $P_C x$ is defined as is the unique point in C with the property:

$$\|x - P_C x\| = \min_{y \in C} \|x - y\|.$$

Lemma 2.4. *Let P_C be the projection mapping. Then*

(i) *P_C is $(1/2)$-averaged and 1-inverse strongly monotone;*
(ii) *$y = P_C x$ if and only if $P_C x \in C, \langle x - y, z - y \rangle \leq 0,\ \forall z \in C$.*

Lemma 2.5. *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function with L-Lipschitz continuous differential ∇f. Let $0 < r < (2/L), T_r = P_C(I - r\nabla f)$. Then (1) $\mathrm{Fix}(T_r) = \Omega$; (2) T_r is $(2 + rL)/4$-averaged.*

Proof. (1) According to [4, Lemma 5.13], x^* solves problem (1.1) if and only if

$$\langle \nabla f(x^*), x^* - z\rangle \leq 0, \forall z \in C,$$

which is equivalent to

$$\langle x^* - (x^* - r\nabla f(x^*)), x^* - z\rangle \leq 0.$$

It then follows from Lemma 2.4 that $x^* = P_C(I - r\nabla f)x^*$.

(2) We note that

$$\begin{aligned}&\|(I - (2/L)\nabla f)x - (I - (2/L)\nabla f)y\|^2 = \|(x - y) - (2/L)(\nabla f(x) - \nabla f(y))\|^2\\ &\quad = \|x - y\|^2 + 4/L^2\|\nabla f(x) - \nabla f(y)\|^2 - 4/L\langle x - y, \nabla f(x) - \nabla f(y)\rangle.\end{aligned}$$

According to [1, Corollary 10], ∇f is $1/L$-inverse strongly monotone, if it is L-Lipschitz continuous, and thus $I - (2/L)\nabla f$ is nonexpansive. Note that

$$I - r\nabla f = \left(1 - \frac{r}{(2/L)}\right) I + \frac{r}{(2/L)}(I - (2/L)\nabla f).$$

This implies that $I - r\nabla f$ is $rL/2$-averaged. Since the projection mapping is $(1/2)$-averaged, Lemma 2.3 is therefore applicable. □

Lemma 2.6 (see [8])**.** *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function with L-Lipschitz continuous differential ∇f. Let $0 < r \leq r' \leq (2/L), T_r = P_C(I - r\nabla f)$. Then*

$$\|T_r x - x\| \leq \|T_{r'} x - x\|, \ x \in C.$$

Lemma 2.7 (see [9])**.** *Let (a_n) be a nonnegative real sequence satisfying*

$$a_{n+1} \leq (1 - \alpha_n)a_n + \alpha_n b_n,$$

where $(\alpha_n) \subset (0, 1)$ and (b_n) are real sequences. Then $a_n \to 0$ provided that

(i) $\sum \alpha_n = \infty, \lim_n \alpha_n = 0$;

(ii) $\overline{\lim}_n b_n \leq 0$ *or* $\sum \alpha_n |b_n| < \infty$.

3. Main result

We now prove our main theorem.

Theorem 3.1. *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function with L-Lipschitz continuous differential ∇f, $h : C \to C$ a ρ-contractive mapping. Assume that $(\theta_n) \subseteq [0, 1]$ and (r_n) satisfy*

$$0 < \liminf r_n \leq \limsup r_n < (2/L), \tag{3.1}$$

$$\lim_{n\to\infty} \theta_n = 0, \sum \theta_n = \infty. \tag{3.2}$$

Then the sequence generated by (1.2) *converges strongly to a minimizer x^* of problem* (1.1)*, which also solves*

$$\langle (I - h)x^*, y - x^*\rangle \geq 0, \forall y \in \Omega. \tag{3.3}$$

Proof. By Lemma 2.4, x^* solves (3.3) if and only if $x^* = P_\Omega h(x^*)$. It is easy to check the boundedness of the iterative algorithm. Now set

$$T_n := P_C(I - r_n \nabla f_n).$$

By Lemma 2.3, $\mathrm{Fix}(T_n) = S, T_n$ is an averaged mapping, and

$$\|T_n x_n - x^*\|^2 \leq \|x_n - x^*\|^2 - \frac{(2/L) - r_n}{(2/L) + r_n}\|x_n - T_n x_n\|^2. \tag{3.4}$$

On the other hand, by the subdifferential inequality

$$\|a + b\|^2 \leq \|a\|^2 + 2\langle b, a + b\rangle, \forall a, b \in \mathcal{H},$$

we have the following estimation:

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

Substituting (3.4) into the above,

$$\begin{aligned}
\|x_{n+1} - x^*\|^2 &\leq \theta_n\rho\|x_n - x^*\|^2 + (1 - \theta_n)\|x_n - x^*\|^2 \\
&\quad - (1 - \theta_n)\frac{(2/L) - r_n}{(2/L) + r_n}\|x_n - T_n x_n\|^2 + 2\theta_n\langle h(x^*) - x^*, x_{n+1} - x^*\rangle \\
&\leq [1 - (1 - \rho)\theta_n]\|x_n - x^*\|^2 \\
&\quad - (1 - \theta_n)\frac{(2/L) - r_n}{(2/L) + r_n}\|x_n - T_n x_n\|^2 + 2\theta_n\langle h(x^*) - x^*, x_{n+1} - x^*\rangle.
\end{aligned}$$

Since $\theta_n \to 0$, it is readily seen that

$$\liminf_{n\to\infty}(1 - \theta_n)\frac{(2/L) - r_n}{(2/L) + r_n} > 0.$$

Without loss of generality we may assume that there is $\sigma > 0$ so that for all $n \geq 0$

$$(1 - \theta_n)((2/L) - r_n)/((2/L) + r_n) \geq \sigma.$$

Let $s_n = \|x_n - x^*\|^2, \alpha_n = (1 - \rho)\theta_n$. Hence

$$s_{n+1} - s_n + \alpha_n s_n + \sigma\|x_n - T_n x_n\|^2 \leq \frac{2\alpha_n}{1 - \rho}\langle h(x^*) - x^*, x_{n+1} - x^*\rangle. \tag{3.5}$$

Finally we prove $s_n \to 0$ by considering two possible cases.

Case 1. (s_n) is eventually decreasing (i.e., there exists $N \geq 0$ such that (s_n) is decreasing for $n \geq N$). In this case, (s_n) must be convergent. According to (3.5),

$$\sigma\|x_n - T_n x_n\|^2 \leq M\alpha_n + (s_n - s_{n+1}), \tag{3.6}$$

where $M > 0$ is so large that

$$\frac{2}{1-\rho}\|h(x^*) - x^*\|\|x_{n+1} - x^*\| \leq M$$

for all $n \in \mathbb{N}$. Taking $n \to \infty$ in (3.6) yields $\|x_n - T_n x_n\| \to 0$. Since $\liminf r_n > 0$, we assume without loss of generality that there is $r > 0$ so that $r_n \geq r$ for all $n \in \mathbb{N}$. Set $T_r = P_C(I - rA)$. By Lemma 2.6,

$$\|x_n - T_r x_n\| \leq \|x_n - T_n x_n\| \to 0.$$

Using Demicolsedness Principle, $\omega_w(x_n) \subset \mathrm{Fix}(T_r) = \Omega$, and by Lemma 2.4,

$$\limsup_{n\to\infty}\langle h(x^*) - x^*, x_n - x^*\rangle = \max_{w\in\omega_w(x_n)}\langle h(x^*) - P_\Omega h(x^*), w - P_\Omega h(x^*)\rangle \leq 0.$$

Note that inequality (3.5) implies

$$s_{n+1} \leq (1-\alpha_n)s_n + \frac{2\alpha_n}{1-\rho}\langle h(x^*) - x^*, x_{n+1} - x^*\rangle.$$

We thus apply Lemma 2.7 to conclude $s_n \to 0$.

Case 2. (s_n) is not eventually decreasing. Hence, we can find a subsequence (s_{n_k}) so that $s_{n_k} \leq s_{n_k+1}$ for all $k \geq 0$. Now let us define

$$J_n := \{n_0 \leq k \leq n : s_k \leq s_{k+1}\}, \forall n > n_0.$$

Obviously J_n is nonempty and $J_n \subseteq J_{n+1}$. Let $\tau(n) := \max J_n, n > n_0$. Then $\tau(n) \to \infty$; otherwise (s_n) is nonincreasing. It is readily seen that $s_{\tau(n)} \leq s_{\tau(n)+1}$ for all $n > n_0$. Hence

$$s_n \leq s_{\tau(n)+1}, \ \forall n > n_0. \tag{3.7}$$

In fact, if $\tau(n) = n$, then (3.7) is trivial; otherwise, by the definition of $\tau(n)$

$$s_{\tau(n)+1} > s_{\tau(n)+2} > \cdots > s_n.$$

Hence (3.7) holds true. Since $s_{\tau(n)} \leq s_{\tau(n)+1}$ for $n > n_0$, we deduce from (3.5) that

$$\|x_{\tau(n)} - T_{\tau(n)}x_{\tau(n)}\|^2 \leq M\alpha_{\tau(n)} \to 0.$$

Similarly, we have $\omega_w(x_{\tau(n)}) \subset \Omega$. On the other hand, we have

$$\|x_{\tau(n)} - x_{\tau(n)+1}\| \leq \theta_{\tau(n)}\|h(x_{\tau(n)}) - x_{\tau(n)}\| + \|T_{\tau(n)}x_{\tau(n)} - x_{\tau(n)}\|.$$

Since $\|h(x_{\tau(n)}) - x_{\tau(n)}\|$ is bounded, letting $n \to \infty$ yields $\|x_{\tau(n)} - x_{\tau(n)+1}\| \to 0$. Hence

$$\begin{aligned}\limsup_{n\to\infty}\langle h(x^*) - x^*, x_{\tau(n)+1} - x^*\rangle &= \limsup_{n\to\infty}\langle h(x^*) - x^*, x_{\tau(n)} - x^*\rangle \\ &= \max_{w\in\omega_w(x_{\tau(n)})}\langle h(x^*) - P_\Omega h(x^*), w - P_\Omega h(x^*)\rangle \leq 0. \end{aligned} \tag{3.8}$$

Since $s_{\tau(n)} \leq s_{\tau(n)+1}$, we deduce from (3.5) that

$$s_{\tau(n)} \leq \frac{2}{1-\rho}\langle h(x^*) - x^*, x_{\tau(n)+1} - x^*\rangle, \quad n > n_0. \tag{3.9}$$

Combining (3.8) and (3.9), we have

$$\limsup_n s_{\tau(n)} \leq 0 \Rightarrow s_{\tau(n)} \to 0.$$

Consequently

$$\begin{aligned}\sqrt{s_{\tau(n)+1}} &\leq \|(x_{\tau(n)} - x^*) + (x_{\tau(n)+1} - x_{\tau(n)})\| \\ &\leq \sqrt{s_{\tau(n)}} + \|x_{\tau(n)+1} - x_{\tau(n)}\| \to 0.\end{aligned}$$

In view of (3.7), $s_n \to 0$, that is, $x_n \to x^*$. □

Remark 3.2. The construction of $(\tau(n))$ is motivated by an idea invented by Maingé [7].

We apply the above result to get the following.

Corollary 3.3. *Let condition* (3.1)-(3.2) *be satisfied. Given $u \in C$ and an initial guess $x_0 \in C$, let (x_n) be a sequence generated by*

$$x_{n+1} = \theta_n u + (1 - \theta_n) P_C(x_n - r_n \nabla f(x_n)).$$

Then (x_n) converges strongly to $x^* = P_\Omega u$, a minimizer of problem (1.1).

Acknowledgement

The authors would like to thank the reviewer for his/her comments that help improve the manuscript.

References

[1] J. B Baillon and G. Haddad, *Quelques proprietes des operateurs angle-bornes et ncycliquement monotones*, Israel J. Math. **26** (1977), 137–150.
[2] H .H Bauschke and P. L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces*, Math. Oper. Res. (2001) 26248–264.
[3] P.L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization **53** (2004), 475–504.
[4] H. W Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, 1996.
[5] K. Goebel and W. A. Kirk, *Topics on Metric Fixed Point Theory*, Cambridge University Press, Cambridge, 1990.
[6] H. Hundal, *An alternating projection that does not converge in norm*, Nonlinear Anal. **57** (2004) 35–61.
[7] P. Maingé, *A hybrid extragradient-viscosity method for monotone operators and fixed point problems*, SIAM J. Control Optim. **47** (2008), 1499–1515.
[8] F. Wang and H. K. Xu, *Strongly convergent iterative algorithms for solving a class of variational inequalities*, J. Nonlinear Convex Anal. **11** (2010), 407–421.
[9] H. K. Xu, *Iterative algorithms for nonlinear operators*, J. Lond. Math. Soc. **66** (2002), 240–256.
[10] H. K. Xu, *Averaged mappings and the gradient-projection algorithm*, J. Optim. Theory Appl. **150** (2011), 360–378.

Manuscript received September 11, 2011
revised March 14, 2012

Huanhuan Cui
Department of mathematics, Luoyang Normal University, Luoyang 471022 China
E-mail address: hhcui@live.cn

Fenghui Wang
Department of mathematics, Luoyang Normal University, Luoyang 471022 China
E-mail address: wfenghui@163.com