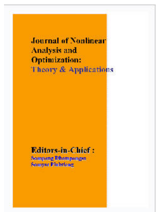
Journal of Nonlinear Analysis and Optimization
Vol. 1 No. 1 (2010), 35-43
ISSN : 1906-9685
http://www.sci.nu.ac.th/jnao



Remarks on the Gradient-Projection Algorithm

Meng Su[1], Hong-Kun Xu[2,*]

ABSTRACT: The gradient-projection algorithm (GPA) is a powerful method for solving constrained minimization problems in finite (and even infinite) dimensional Hilbert spaces. We consider GPA with variable stepsizes and show that if GPA generates a bounded sequence, then under certain assumptions, every accumulation point of the sequence is a solution of the minimization problem. We also look into the issue where the sequence of stepsizes is allowed to be the limiting case (e.g., approaching to zero).

KEYWORDS: Gradient-projection algorithm, constrained minimization, variational inequality, optimality condition, convex function, Lipschitz continuous gradient, Féjer-monotone.

1. **Introduction**

Consider the constrained minimization problem

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

where C is a nonempty closed convex set of $\mathbb{R}^n$, and the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. Assume that (1) is consistent (i.e., it has a solution) and we use Γ to denote its solution set.

It is well known that the (necessary) optimality condition for a point $x^* \in C$ to be a solution of (1) is that

(2) $$-\nabla f(x^*) \in N_C(x^*)$$

where $N_C(x^*)$ is the normal cone to C at x^*, namely,

$$N_C(x^*) = \{v \in \mathbb{R}^n : \langle v, x - x^*\rangle \le 0, \quad x \in C\}.$$

Condition (2) is equivalent to the following variational inequality (VI):

(3) $$x^* \in C, \quad \langle \nabla f(x^*), x - x^*\rangle \ge 0, \quad x \in C.$$

Let P_C be the nearest point projection from $\mathbb{R}^n$ onto C. We can then rewrite VI (3) equivalently to a fixed point equation

(4) $$x^* = P_C(I - \alpha\nabla f)x^*,$$

where $\alpha > 0$ is any (fixed) constant.

Corresponding author: Hong-Kun Xu (xuhk@math.nsysu.edu.tw).
Manuscript received July 11, 2010. Accepted July 31, 2010.
*Supported in part by NSC 97-2628-M-110-003-MY3 (Taiwan).

The gradient-projection algorithm (GPA, for short) is usually applied to solve the minimization problem (1). This algorithm generates a sequence $\{x^k\}$ through the recursion:

$$x^{k+1} = P_C(x^k - \alpha_k \nabla f(x^k)), \quad k = 0, 1, \cdots, \tag{5}$$

where the initial guess $x^0 \in C$ is chosen arbitrarily and $\{\alpha_k\}$ is a sequence of stepsizes which may be chosen in different ways.

GPA (5) has well been studied in the case of constant stepsizes $\alpha_k = \alpha$ for all k (see the books [4, 5], and the papers [1, 2, 3, 6, 7, 8]). A recent averaged mapping approach to GPA (5) can be found in [9].

A fundamental convergence result for GPA (5) is the following one which can be found in literature (cf. [5, Theorem 6.1] or [4, Theorem 1, Section 7.2] with constant stepsize).

Theorem 1.1. *Let $\{x^k\}$ be the sequence generated by GPA (5). Assume*

(i) *f is continuously differentiable and its gradient is Lipschitz continuous:*

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad x, y \in \mathbb{R}^n, \tag{6}$$

where $L \ge 0$ is a constant;

(ii) *the set $C_0 := \{x \in C : f(x) \le f(x^0)\}$ is bounded;*

(iii) *the sequence $\{\alpha_k\}$ satisfies the condition:*

$$0 < \liminf_{k\to\infty} \alpha_k \le \limsup_{k\to\infty} \alpha_k < \frac{2}{L}. \tag{7}$$

Then every accumulation point x^* of $\{x^k\}$ satisfies the optimality condition (2). Moreover, if f is also convex, then $\{x^k\}$ converges to a solution of the minimization (1).

In this paper we deal with the gradient-projection algorithm with variable stepsize. We show that if GPA (5) generates a bounded sequence $\{x^k\}$ (which is guaranteed by the boundedness of the set C_0), then under condition (7), every accumulation point of $\{x^k\}$ is a solution to the minimization problem (1). If the objective f is, in addition, convex, then $\{x^k\}$ indeed converges to a solution of (1). We also deal with the saturated situation where we can allow the sequence $\{\alpha_k\}$ to close zero (see the precise descriptions in Section 3).

2. Preliminaries

Let H be the real Euclidean n-space $\mathbb{R}^n$ and C be a nonempty closed convex subset of H.

Definition 2.1. The (metric or nearest-point or orthogonal) projection from H onto C is a mapping that assigns, to each point $x \in H$, a unique point in C, denoted $P_C x$, with the property:

$$\|x - P_C x\| = \min\{\|x - y\| : y \in C\}. \tag{8}$$

The following characterizes the relation $z = P_C x$.

Proposition 2.2. *Given $z \in C$ and $x \in H$. Then $z = P_C x$ if and only if there holds the relation:*

$$\langle x - z, y - z\rangle \le 0, \quad y \in C.$$

Consequently, P_C is firmly nonexpansive, that is,

$$\langle P_C x - P_C y, x - y, x - y\rangle \ge \|P_C x - P_C y\|^2, \quad x, y \in H.$$

In particular, P_C is nonexpansive:

$$\|P_C x - P_C y\| \le \|x - y\|, \quad x, y \in H.$$

We next recall the optimality condition for the minimization problem (1).

Lemma 2.3. (Optimality Condition.) *A necessary condition of optimality for a point $x^* \in C$ to be a solution of the minimization problem (1) is that x^* solves the variational inequality:*

$$x^* \in C, \quad \langle \nabla f(x^*), x - x^* \rangle \geq 0, \quad x \in C. \tag{9}$$

Equivalently, $x^* \in C$ solves the fixed point equation

$$x^* = P_C(x^* - \alpha \nabla f(x^*))$$

for every constant $\alpha > 0$.

If, in addition, f is convex, then the optimality condition (9) is also sufficient.

Lemma 2.4. *(cf.* [4, Lemma 2, Section 1.4]*) Suppose a continuously differentiable convex function $f : \mathbb{R}^n \to \mathbb{R}$ has a Lipschitz continuous gradient:*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad x, y \in \mathbb{R}^n. \tag{10}$$

Then ∇f satisfies the so-called inversely strong (also known as co-coercive) monotonicity:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2, \quad x, y \in \mathbb{R}^n. \tag{11}$$

3. The Gradient-Projection Algorithm

Recall that the gradient-projection method (GPM) for solving the minimization problem (1) generates a sequence $\{x^k\}$ according to the recursive process

$$x^{k+1} = P_C(x^k - \alpha_k \nabla f(x^k)), \quad k = 0, 1, 2, \cdots, \tag{12}$$

where $\{\alpha_k\}$ is a sequence of stepsizes with $\alpha_k \geq 0$ for each $k \geq 0$. The convergence of GPM (12) depends on the choice of the stepsize sequence $\{\alpha_k\}$ (and also on the behavior of the gradient ∇f). The purpose of this section is to discuss the convergence of GPM (12) under different choices of the stepsize sequence $\{\alpha_k\}$. (Note that in both books [5, 4], constant stepsize $\alpha_k \equiv \alpha$ is considered; here we deal with variable stepsize.)

3.1. Variable Stepsize

Theorem 3.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function such that the gradient ∇f satisfies the Lipschitz continuity condition (10). For a given initial guess $x^0 \in C$, let $\{x^k\}$ be the sequence generated by GPA (12). Assume*

- $(A1)$ *$\{\alpha_k\}$ satisfies condition (7).*
- $(A2)$ *$\{x^k\}$ is bounded. (This is guaranteed by the boundedness assumption (ii) of the set $C_0 = \{x \in C : f(x) \leq f(x^0)\}$ in Theorem 1.1, as assumed in* [5]*).*

Then we have

- (i) *Every accumulation point of $\{x^k\}$ is a solution of the minimization problem (1).*
- (ii) $\lim_{k\to\infty} f(x^k) = f_{\min} := \min\{f(x) : x \in C\}$.
- (iii) *If $\{x^k\}$ is Féjer-monotone with respect to the solution set Γ of the minimization problem (1), that is,*

$$\|x^{k+1} - x^*\| \leq \|x^k - x^*\|, \quad k \geq 0, \; x^* \in \Gamma, \tag{13}$$

then the sequence $\{x^k\}$ converges to a solution of (1).

Proof. We have, using (6),

$$f(x^{k+1}) - f(x^k) = \int_0^1 \langle \nabla f(x^k + t(x^{k+1} - x^k)), x^{k+1} - x^k \rangle dt$$
$$= \int_0^1 \langle \nabla f(x^k + t(x^{k+1} - x^k)) - \nabla f(x^k), x^{k+1} - x^k \rangle dt + \langle \nabla f(x^k), x^{k+1} - x^k \rangle$$
$$\le \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2. \tag{14}$$

On the other hand, since x^{k+1} is the projection of $x^k - \alpha_k \nabla f(x^k)$ onto C, it follows from Proposition 2.2 that

$$\langle x^k - \alpha_k \nabla f(x^k) - x^{k+1}, x^k - x^{k+1} \rangle \le 0.$$

This implies that

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle \le -\frac{1}{\alpha_k}\|x^{k+1} - x^k\|^2. \tag{15}$$

Substituting (15) into (14), we get

$$f(x^{k+1}) \le f(x^k) - \left(\frac{1}{\alpha_k} - \frac{L}{2}\right)\|x^{k+1} - x^k\|^2. \tag{16}$$

Let $\underline{\alpha} = \liminf_{k\to\infty} \alpha_k$ and $\overline{\alpha} = \limsup_{k\to\infty} \alpha_k$. Then, by assumption (7), we see that there is an integer $N \ge 1$ such that

$$0 < \underline{\alpha} \le \alpha_k \le \overline{\alpha} < \frac{2}{L} \tag{17}$$

for all $k \ge N$. With no loss of generality, we may assume that (17) holds true for all $k \ge 0$. It then turns out from (16) that

$$f(x^{k+1}) \le f(x^k) - \frac{2 - \overline{\alpha}L}{2\overline{\alpha}}\|x^{k+1} - x^k\|^2. \tag{18}$$

The sequence $\{f(x^k)\}$ is therefore decreasing and $\lim_{k\to\infty} f(x^k)$ exists. Moreover,

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0. \tag{19}$$

We next prove that every accumulation point of $\{x^k\}$ is a solution of the minimization problem (1). Note that since $\{x^k\}$ is assumed to be bounded, the set of accumulation points of $\{x^k\}$ is nonempty. Let $\hat{x}$ be an accumulation point of $\{x^k\}$ and let $\{x^{k_j}\}$ be a subsequence of $\{x^k\}$ converging to $\hat{x}$. Due to (19), we also have that $\{x^{k_j+1}\}$ converges to $\hat{x}$.

We may also assume that the subsequence $\{\alpha_{k_j}\}$ is convergent to some number $\hat{\alpha}$ (say). According to (17), we get

$$0 < \underline{\alpha} \le \hat{\alpha} \le \overline{\alpha} < \frac{2}{L}.$$

Taking the limit as $j \to \infty$ in the relation

$$x^{k_j+1} = P_C(x^{k_j} - \alpha_{k_j}\nabla f(x^{k_j}))$$

yields $\hat{x} = P_C(\hat{x} - \hat{\alpha}\nabla f(\hat{x}))$ which exactly says that $\hat{x}$ solves the minimization problem (1) and hence, $\lim_{k\to\infty} f(x^k) = f(\hat{x}) = f_{\min}$.

Finally we show that the entire sequence $\{x^k\}$ converges if the Féjer monotonicity condition (13) holds. Let x^* be an accumulation point of $\{x^k\}$. Due to (13), the full limit as $k \to \infty$ of the sequence $\{\|x^k - x^*\|\}$, $\lim_{k\to\infty} \|x^k - x^*\|$, exists. However, a subsequence of it converges to zero. We therefore conclude that $\lim_{k\to\infty} \|x^k - x^*\| = 0$ and the full sequence $\{x^k\}$ converges to x^*. □

If f is also convex, then we can remove the boundedness assumption on $\{x^k\}$ in Theorem 3.1.

Theorem 3.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable convex function such that the gradient ∇f satisfies the Lipschitz continuity condition (10). Assume the minimization problem (1) is consistent. For a given initial guess $x^0 \in C$, let $\{x^k\}$ be the sequence generated by GPA (12) where we assume $\{\alpha_k\}$ satisfies condition (7). Then $\{x^k\}$ converges to a solution of (1).*

Proof. We only need to prove the Féjer-monotonicity (13) holds when f is convex. To see this, we take $x^* \in \Gamma$. Observing $x^* = P_C(I - \alpha\nabla f)x^*$ for every $\alpha > 0$, and using the nonexpansivity of P_C and Lemma 2.4, we derive that

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 &= \|P_C(I - \alpha_k\nabla f)x^k - P_C(I - \alpha_k\nabla f)x^*\|^2 \\
&\le \|(I - \alpha_k\nabla f)x^k - (I - \alpha_k\nabla f)x^*\|^2 \\
&= \|(x^k - x^*) - \alpha_k(\nabla f(x^k) - \nabla f(x^*))\|^2 \\
&= \|x^k - x^*\|^2 - 2\alpha_k\langle x^k - x^*, \nabla f(x^k) - \nabla f(x^*)\rangle \\
&\quad + \alpha_k{}^2\|\nabla f(x^k) - \nabla f(x^*)\|^2 \\
&\le \|x^k - x^*\|^2 - \alpha_k\left(\frac{2}{L} - \alpha_k\right)\|\nabla f(x^k) - \nabla f(x^*)\|^2 \\
&\le \|x^k - x^*\|^2 - \frac{\underline{\alpha}(2 - \overline{\alpha})}{L}\|\nabla f(x^k) - \nabla f(x^*)\|^2.
\end{aligned}$$

It is now immediately clear that $\|x^{k+1} - x^*\| \le \|x^k - x^*\|$, namely, $\{x^k\}$ satisfies the Féjer-monotonicity condition (13). Consequently, $\{x^k\}$ is bounded, and moreover converges to a point in Γ. □

3.2. **Two-Slope Test**

The two-slope test for the unconstrained minimization problem

(20) $$\min\{f(x) : x \in \mathbb{R}^n\}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, determines the stepsize sequence $\{\alpha_k\}$ in such a way that

(21) $$f(x^k) - b\alpha_k\|\nabla f(x^k)\|^2 \le f(x^k - \alpha_k\nabla f(x^k)) \le f(x^k) - a\alpha_k\|\nabla f(x^k)\|^2,$$

where $0 < a < b < 1$ are two fixed constants.

The following theorem is known [5, Theorem 5.3].

Theorem 3.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and its gradient ∇f is Lipschitz with constant L. Let $\{x^k\}$ be the sequence generated by the steepest-descent method:*

(22) $$x^{k+1} = x^k - \alpha_k\nabla f(x^k), \quad k = 0, 1, \cdots,$$

where the stepsize sequence $\{\alpha_k\}$ is determined by the two-slope test (21). Assume the set

$$M_0 := \{x \in \mathbb{R}^n : f(x) \le f(x^0)\}$$

is bounded. Then $\{x^k\}$ is bounded and every accumulation point x^* of $\{x^k\}$ satisfies the optimality condition $\nabla f(x^*) = 0$ for the unconstrained minimization (20).

Below we extend this two-slope test from unconstrained to constrained minimization.

The two-slope test for the constrained minimization problem (1) determines the stepsize for the $(k+1)$th iterate x^{k+1} in such a way that

(23) $$f(x^k) - b\langle\nabla f(x^k), x^k - x^k(\alpha_k)\rangle \le f(x^k(\alpha_k)) \le f(x^k) - a\langle\nabla f(x^k), x^k - x^k(\alpha_k)\rangle,$$

where $0 < a < b < 1$ are two fixed constants, and where

$$x^k(\alpha) = P_C(x^k - \alpha\nabla f(x^k)), \quad \alpha \ge 0.$$

It is easily found that if C is the entire space $\mathbb{R}^n$ (i.e., the constrained minimization (1) is reduced to the unconstrained minimization (20)), then (23) is reduced to (21).

We have the following convergence result which extends Theorem 3.3 to the case of constrained minimization.

Theorem 3.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function such that its gradient ∇f satisfies the Lipschitz continuity condition (10). Let $\{x^k\}$ be the sequence generated by GPM (12), where the sequence of stepsizes, $\{\alpha_k\}$, satisfies the two-slope test (23) and $\liminf_{k\to\infty} \alpha_k > 0$. Assume, in addition, that the set*

$$C_0 := \{x \in C : f(x) \le f(x_0)\}$$

is bounded. Then $\{x^k\}$ is bounded and every accumulation point $\bar{x}$ of $\{x^k\}$ is a stationary point of the constrained minimization problem (1); namely, $\bar{x}$ satisfies the variational inequality

$$\bar{x} \in C, \quad \langle \nabla f(\bar{x}), x - \bar{x} \rangle \ge 0, \quad x \in C. \tag{24}$$

In particular, if f is also convex, then $\bar{x}$ is a solution of (1).

Proof. Noticing $x^{k+1} = x^k(\alpha_k)$, we get immediately from (23) that $f(x^{k+1}) \le f(x^k)$ for all $k \ge 0$. This implies that $\{x^k\} \subset C_0$; in particular, $\{x^k\}$ is bounded. Hence

$$\lim_{k\to\infty} f(x^k) \quad \text{exists.}$$

Also (23) implies that

$$\langle \nabla f(x^k), x^k - x^k(\alpha_k) \rangle \ge 0. \tag{25}$$

Since there also holds (again from (23))

$$a\langle \nabla f(x^k), x^k - x^k(\alpha_k) \rangle \le f(x^k) - f(x^{k+1}),$$

we get that

$$\lim_{k\to\infty} \langle \nabla f(x^k), x^k - x^{k+1} \rangle = 0. \tag{26}$$

Now since x^{k+1} is the projection of $x^k - \alpha_k \nabla f(x^k)$ onto C, we have by Proposition 2.2

$$\langle (x^k - \alpha_k \nabla f(x^k)) - x^{k+1}, y - x^{k+1} \rangle \le 0, \quad y \in C.$$

It turns out that

$$\langle x^k - x^{k+1}, y - x^{k+1} \rangle \le \alpha_k \langle \nabla f(x^k)), y - x^{k+1} \rangle, \quad y \in C.$$

Setting $y := x^k \in C$ we get

$$\|x^k - x^{k+1}\|^2 \le \alpha_k \langle \nabla f(x^k)), x^k - x^{k+1} \rangle. \tag{27}$$

Combining (26) and (27), we obtain

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0. \tag{28}$$

Now let $\bar{x}$ be an accumulation point of $\{x^k\}$ and let $\{x^{k_j}\}$ be a subsequence of $\{x^k\}$ converging to $\bar{x}$. With no loss of generality, we may assume that $\alpha_{k_j} \to \bar{\alpha} > 0$ (for $\liminf_{k\to\infty} \alpha_k > 0$). Now taking the limit as $j \to \infty$ in the relation

$$x^{k_j+1} = P_C(x^{k_j} - \alpha_{k_j} \nabla f(x^{k_j}))$$

gives that

$$\bar{x} = P_C(\bar{x} - \bar{\alpha} \nabla f(\bar{x})).$$

This equivalently says that $\bar{x}$ satisfies VI (24) and is a stationary point of the minimization problem (1).

When f is convex, the optimality condition is also sufficient and $\bar{x}$ is therefore a solution of (1). □

3.3. **Strongly Monotone Gradient**

Assume that the objective function f is continuously differentiable such that its gradient is Lipschitzian and strongly monotone. Namely, there exist constants $\beta > 0$ and $L \geq 0$ satisfying, for all $x, y \in C$,

$$\langle \nabla f(x) - \nabla f(x), x - y \rangle \geq \beta \|x - y\|^2 \tag{29}$$

and

$$\|\nabla f(x) - \nabla f(x)\| \leq L\|x - y\|. \tag{30}$$

We then have that the mapping

$$T \equiv T_\gamma := P_C(I - \gamma \nabla f)$$

is a contraction provided $0 < \gamma < 2\beta/L^2$. As a matter of fact, we have

$$\begin{aligned} \|Tx - Ty\|^2 &= \|P_C(I - \gamma\nabla f)x - P_C(I - \gamma\nabla f)y\|^2 \\ &\leq \|(I - \gamma\nabla f)x - (I - \gamma\nabla f)y\|^2 \\ &= \|(x - y) - \gamma(\nabla f(x) - \nabla f(y))\|^2 \\ &= \|x - y\|^2 - 2\gamma\langle \nabla f(x) - \nabla f(y), x - y\rangle + \gamma^2\|\nabla f(x) - \nabla f(y)\|^2 \\ &\leq \|x - y\|^2 - 2\gamma\beta\|x - y\|^2 + \gamma^2 L^2\|x - y\|^2 \\ &= (1 - \gamma(2\beta - \gamma L^2))\|x - y\|^2. \end{aligned}$$

This shows that T is a contraction with constant $\sqrt{1 - \gamma(2\beta - \gamma L^2)}$.

Therefore, for such a choice of γ, we can apply Banach's contraction principle to get that for each $x^0 \in C$, the sequence $\{T_\gamma^k x^0\}$ converges to the unique fixed point of T_γ (or the unique solution of the minimization (1)).

We however look at the case where the stepsizes $\{\gamma_k\}$ are variable such that

$$0 < \liminf_{k\to\infty} \gamma_k < \limsup_{k\to\infty} \gamma_k < \frac{2\beta}{L^2}. \tag{31}$$

We have the following convergence result.

Theorem 3.5. *Let $x^0 \in C$ and define a sequence $\{x^k\}$ by the iterative algorithm:*

$$x^{k+1} = P_C(x^k - \gamma_k \nabla f(x^k)), \tag{32}$$

where the sequence $\{\gamma_k\}$ is selected according to the selection rule (31). Then $\{x^k\}$ converges to the unique solution x^* of the minimization (1).

Proof. By (31), there exist some natural number N and positive constants a and b such that

$$0 < a \leq b < 2\beta/L^2 \quad \text{and} \quad a \leq \gamma_k \leq b \ (k \geq N).$$

Set

$$h = \max_{a\leq\gamma\leq b} \sqrt{1 - \gamma(2\beta - \gamma L^2)}.$$

Then $0 \leq h < 1$ and it is easy to see that

$$0 \leq \sqrt{1 - \gamma_k(2\eta - \gamma_k L^2)} \leq h$$

for all $k \geq N$.

Denote by x^* the unique solution of the minimization (1). We now compute, for all $k \geq N$,

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &= \|P_C(I - \gamma_k \nabla f)x^k - P_C(I - \gamma_k \nabla f)x^*\|^2 \\ &\leq \|(I - \gamma_k \nabla f)x^k - (I - \gamma_k \nabla f)x^*\|^2 \\ &= \|(x^k - x^*) - \gamma_k(\nabla f(x^k) - \nabla f(x^*))\|^2 \\ &= \|x^k - x^*\|^2 + \gamma_k^2\|\nabla f(x^k) - \nabla f(x^*)\|^2 \\ &\qquad - 2\gamma_k\langle x^k - x^*, \nabla f(x^k) - \nabla f(x^*)\rangle \\ &\leq [1 - \gamma_k(2\eta - \gamma_k L^2)]\|x^k - x^*\|^2 \\ &\leq h^2\|x^k - x^*\|^2.\end{aligned}$$

Consequently

$$\|x^{k+1} - x^*\| \leq h\|x^k - x^*\| \leq \cdots \leq h^{k-N+1}\|x^N - x^*\|.$$

Therefore, we get $\lim_{k\to\infty} \|x^k - x^*\| = 0$. □

Theorem 3.5 asserts that if the parameter sequence $\{\gamma_k\}$ is bounded away below from zero and above from $2\beta/L^2$, then the sequence $\{x^k\}$ generated by GPA (32) converges to the unique solution of the minimization (1). The result below shows that we can allow $\{\gamma_k\}$ to close either zero or $2\beta/L^2$ and still keep the convergence of the sequence $\{x^k\}$.

Theorem 3.6. *Assume that the sequence $\{\gamma_k\}$ satisfies the condition*

$$0 < \gamma_k < \frac{2\eta}{L^2} \quad \textit{for all } k \textit{ and} \quad \sum_{k=0}^{\infty} \gamma_k \left(\frac{2\beta}{L^2} - \gamma_k\right) = \infty. \tag{33}$$

Then the sequence $\{x^k\}$ generated by GPA (32) converges to the unique solution x^* of the minimization (1).

Proof. The first part of condition (33) assures that that the mapping $P_C(I - \gamma_k \nabla f) : C \to C$ is a contraction with the coefficient $\sqrt{1 - \gamma_k(2\beta - \gamma_k L^2)}$ for all $k \geq 0$. Observing $x^* = P_C(I - \gamma_k \nabla f)x^*$ for all $k \geq 0$, we have

$$\begin{aligned}\|x^{k+1} - x^*\| &= \|P_C(I - \gamma_k \nabla f)x_n - P_C(I - \gamma_k \nabla f)x^*\| \\ &\leq \sqrt{1 - \gamma_k(2\beta - \gamma_k L^2)}\|x^k - x^*\| \\ &\leq \left(1 - \frac{1}{2}\gamma_k(2\beta - \gamma_k L^2)\right)\|x^k - x^*\| \qquad (34) \\ &\leq \|x^k - x^*\|.\end{aligned}$$

In particular, $\{x^k\}$ is bounded and $\lim_{k\to\infty} \|x^k - x^*\|$ exists. Also it follows from (34) that

$$\frac{L^2}{2}\gamma_k \left(\frac{2\beta}{L^2} - \gamma_k\right)\|x^k - x^*\| \leq \|x^k - x^*\| - \|x^{k+1} - x^*\|. \tag{35}$$

Put $r = \lim_{k\to\infty} \|x^k - x^*\|$. If $r > 0$, then by (35), we get (noticing $\|x^k - x^*\| \geq r$ for all k)

$$\frac{rL^2}{2}\gamma_k \left(\frac{2\beta}{L^2} - \gamma_k\right) \leq \|x^k - x^*\| - \|x^{k+1} - x^*\|.$$

Consequently,

$$\sum_{k=0}^{\infty} \gamma_k \left(\frac{2\beta}{L^2} - \gamma_k\right) < \infty$$

which contradicts the assumption (33). So we must have $r = 0$ and $x^k \to x^*$ as $k \to \infty$. □

References

[1] E.M. Gafni and D.P. Bertsekas, Two-metric projection methods for constained optimization, *SIAM J. Control Optim* 22 (1984), 936-964.
[2] P.H. Calamai and J.J. Moré, Projected gradient methods for linearly constained problems, *Mathematical Programming* 39 (1987), 93-116.
[3] E.S. Levitin and B.T. Polyak, Constrained minimization problems , *USSR Computationnal Mathematics and Mathematical Phsics* 6 (1966), 1-50.
[4] B.T. Polyak, "Introduction to Optimization," *Optimization Software*, New York, 1987.
[5] A. Ruszczynski, "Nonlinear optimization," *Princeton University Press*, New Jersey, 2006.
[6] C.Y. Wang and N.H. Xiu, Convergence of gradient projection methods for generalize convex minimization , *Computational Optim. Appl.* 16 (2000), 111-120.
[7] N.H. Xiu, C.Y. Wang, C.Y. and J.Z. Zhang, Convergence properties of projection and contraction methods for variational inequality problems, *Applied Math. Opt.* 43 (2001), 147-168.
[8] N. Xiu, D. Wang and L. Kong, A note on the gradient projection method with exact stepsize rule, *Journal of Computational Mathematics* 25 (2007), 221-230.
[9] H.K. Xu, Averaged mappings and the gradient-projection algorithm. (Preprint.)

[1]PENN STATE UNIVERSITY AT ERIE, THE BEHREND COLLEGE, 4205 COLLEGE DRIVE, ERIE, PA 16563-0203, U.S.A.
Email address: mengsu@psu.edu

[2]DEPARTMENT OF APPLIED MATHEMATICS, NATIONAL SUN YAT-SEN UNIVERSITY, KAOHSIUNG 80424, TAIWAN.
Email address: xuhk@math.nsysu.edu.tw