Fixed Point Theory, 26(2025), No. 2, 497-516
DOI: 10.24193/fpt-ro.2025.2.10
http://www.math.ubbcluj.ro/~nodeacj/sfptcj.html

STRONG CONVERGENCE OF PROJECTED SUBGRADIENT METHODS IN INFINITE-DIMENSIONAL HILBERT SPACES

YUNSHU FU* AND HONG-KUN XU**

*School of Science, Hangzhou Dianzi University, Hangzhou 310018, China
E-mail: 844876873@qq.com

**School of Science, Hangzhou Dianzi University, Hangzhou, 310018, China
E-mail: xuhk@hdu.edu.cn (Corresponding author)

Abstract. Subgradient methods, introduced by Shor and developed by Albert, Iusem, Nesterov, Polyak, Solodov, and many others, are used to solve nondifferentiable optimization problems. In this paper we discuss weak and strong convergence of projected subgradient methods in an infinite-dimensional Hilbert space. We apply the viscosity approximation method to the projected subgradient method to obtain strongly convergent subgradient algorithms. In addition, we develop the forcing strong convergence technique and the CQ algorithm to solve nondifferentiable convex optimization problems.

Key Words and Phrases: Projection, ε-subgradient, nondifferentiable convex optimization, variational inequality, strong convergence, infinite-dimensional.

2020 Mathematics Subject Classification: 90C25, 47J25, 47H26, 47H10.

1. Introduction

Consider a constrained convex optimization problem of the form

$$\min_{x\in C} \varphi(x), \tag{1.1}$$

where C is a nonempty closed convex subset of a Hilbert space H, and $\varphi : H \to \mathbb{R}$ is a continuous, convex function. We use S to denote the set of solutions of (1.1) and assume $S \neq \emptyset$. A sufficient condition to guarantee the existence of a solution of (1.1) is either C is, in addition, bounded, or φ is coercive (i.e., $\varphi(x) \to \infty$ as $\|x\| \to \infty$).

If φ is continuously differentiable, the minimization problem (1.1) can be solved by the projection-gradient algorithm (PGA) which generates a sequence (x_n) by the iterative algorithm:

$$x_{n+1} = P_C(x_n - \alpha_n \nabla\varphi(x_n)), \quad n \geq 0, \tag{1.2}$$

**Supported in part by Natural Science Foundation of China grant number U1811461 and by Australian Research Council/Discovery Project DP200100124.

where the initial guess $x_0 \in C$, and (α_n) is the stepsize sequence, $\nabla\varphi$ is the gradient of φ, and P_C is the metric projection onto C; that is,

$$P_C x = \arg\min_{z\in C} \|z - x\|^2, \quad x \in H.$$

The PGA (1.2) has extensively been studied. The following convergence result is well known [8, 17] (see also [25] for an averaged mapping approach).

Theorem 1.1. *Assume the following two conditions are satisfied:*

(C1) *The gradient of φ, $\nabla\varphi$, satisfies the L-Lipschitz continuity condition (φ is said to be L-smooth):*

$$\|\nabla\varphi(x) - \nabla\varphi(y)\| \le L\|x - y\|, \quad x, y \in C. \tag{1.3}$$

(C2) *The stepszes $\{\alpha_n\}$ satisfy the condition:*

$$0 < \liminf_{n\to\infty} \alpha_n \le \limsup_{n\to\infty} \alpha_n < \frac{2}{L}, \tag{1.4}$$

Then the sequence $\{x_n\}$ generated by PGA (1.2) converges weakly to a point $x^* \in S$. [No strong convergence holds in general (see [25] *for a counterexample).]*

Condition (C2) shows that in the case of an L-smooth objective function φ, constant stepsizes (i.e., $\alpha_n = \alpha \in (0, \frac{2}{L})$) work for solving (1.1).

However, in the case of a nonsmooth objective function, the situation is different: constant stepsizes do not work anymore. This was first observed by Shor [20]. He considered an unconstrained nondifferentiable convex optimization in the finite-dimensional space $H = \mathbb{R}^N$:

$$\min_{x\in\mathbb{R}^N} \varphi(x) \tag{1.5}$$

where φ is a continuous convex function on $\mathbb{R}^N$. The steepest descent method generates a sequence (x_k) by

$$x_{k+1} = x_k - \alpha u_k, \quad u_k \in \partial\varphi(x_k) \tag{1.6}$$

where $x_0 \in \mathbb{R}^N$ is the initial guess, α is the constant stepsize, and $\partial\varphi$ is the subdifferential of φ. Shor used the following example to show that the steepest descent method (1.6) is unsuitable for minimizing the nondifferentiable convex minimization problem (1.5).

Example 1. [20, 26] Define a function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ by

$$\varphi(x) = \begin{cases} 5\sqrt{9x^2 + 16y^2}, & \text{if } x > |y|, \\ 9x + 16|y|, & \text{if } 0 < x \le |y|, \\ 9x + 16|y| - x^9, & \text{if } x < 0, \end{cases}$$

for $(x, y) \in \mathbb{R}^2$. Then the following hold:

(i) φ is convex and differentiable outside the half-line $\{(x, y) \in \mathbb{R}^2 : x \le 0,\ y = 0\}$.

(ii) φ attains its minimal value at the point $(-1, 0)$.

(iii) If the initial point (x_0, y_0) is taken from the domain $\{(x, y) \in \mathbb{R}^2 : x > |y| > (\frac{9}{16})^2|x|\}$, then the steepest descent method with exact line search converges to the point $(0, 0)$ which is non-optimal.

We find the following simpler example to show that the subgradient method (1.6) fails to converge to an optimal solution when constant stepsizes are utilized.

Example 2. Consider $\min_{x\in\mathbb{R}} \varphi(x) := |x|$. The steepest descent method with constant stepsize $\alpha > 0$ is

$$x_{n+1} = x_n - \alpha u_n, \quad u_n \in \partial\varphi(x_n).$$

Choose the initial guess $x_0 = 1$ and constant stepsize $\alpha \in (\frac{1}{2}, 1)$. Since

$$\partial|x| = \begin{cases} 1, & \text{if } x > 0, \\ [-1,1], & \text{if } x = 0, \\ -1, & \text{if } x < 0, \end{cases}$$

it is easy to find that

$$x_{2n-1} = 1 - \alpha > 0, \quad x_{2n} = 1 - 2\alpha < 0$$

for all $n \geq 1$. Consequently, the steepest descent method with constant stepsize no longer converges in general for nondifferentiable optimization.

Shor thus pointed out that the way of selection of stepsizes makes a big difference between differentiable (smooth) and nondifferentiable (nonsmooth) convex optimization problems, constant stepsizes are no longer suitable for nondifferentiable objective functions, and instead, diminishing stepsizes have to be implemented. He suggested the following subgradient method for solving (1.5):

$$x_{k+1} = x_k - \frac{\gamma_k}{\|u_k\|} u_k, \quad u_k \in \partial\varphi(x_k) \tag{1.7}$$

where $\gamma_k > 0$ and $0 \neq u_k \in \partial\varphi(x_k)$ (note that if $u_k = 0$, then x_k is an optimal solution).

Alber, et al [1] developed Shor's theory to the framework of a general Hilbert space (possibly infinite-dimensional) for the constrained nondifferentiable convex optimization problem (1.1).

The projected subgradient algorithm (PSA) of Alber, et al [1] generates a sequence (x_n) by the following iteration process:

$$x_{n+1} = P_C\left(x_n - \frac{\alpha_n}{\eta_n} u_n\right), \quad n = 0, 1, \cdots, \tag{1.8}$$

where the initial guess $x_0 \in C$, $u_n \in \partial_{\varepsilon_n}\varphi(x_n)$ (ε_n-subdifferential of φ), $\eta_n = \max\{1, \|u_n\|\}$, and (α_n) and (ε_n) are sequences of nonnegative real numbers such that

$$\sum_{n=0}^{\infty} \alpha_n = \infty, \quad \sum_{n=0}^{\infty} \alpha_n^2 < \infty, \tag{1.9}$$

$$\varepsilon_n \leq \mu\alpha_n \quad \text{for all } n \text{ and some constant } \mu \geq 0. \tag{1.10}$$

Regarding the PSA (1.8), the following has been proved.

Theorem 1.2. [1] *Suppose that Assumption A (see Section 3) holds and let (x_n) be generated by PSA (1.8). Assume, in addition, the conditions (1.9) and (1.10) are satisfied. Then (x_n) converges weakly to a point in S, i.e., a solution of* (1.1).

In this paper we are aimed to develop the projected subgradient method for solving the nondifferentiable convex optimization problem (1.1) in the infinite-dimensional Hilbert space framework, focusing on strongly convergent algorithms which are developed from modifications of PSA (1.8). More precisely, we will provide a simpler proof of Theorem 1.2 in Section 3 after some preliminaries in Section 2. In Section 4, we shall apply the viscosity approximation method to obtain two strongly convergent projected subgradient algorithms, followed by a result on the forcing strong convergence technique applied to PSA (1.8). Our final result is a strongly convergent subgradient CQ algorithm presented in Section 6. A summary will be given in Section 7.

2. Preliminaries

Let H be a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$, respectively. Given a nonempty closed convex subset $C \subset H$. We use P_C to denote the metric (or nearest point) projection from H onto C, that is, for each $x \in H$, $P_C x$ is the only point in C with the property

$$\|x - P_C x\| \le \|x - z\|$$

for all $z \in C$. In notation, we write $P_C x = \arg\min_{z\in C} \|x - z\|$.

The following properties of projections are pertinent to our argument.

Proposition 2.1. *Projections satisfy the properties:*

(i) $\|x - P_C x\|^2 \le \|x - z\|^2 - \|z - P_C x\|^2$ *for* $x \in H$ *and* $z \in C$.

(ii) $\|P_C x - P_C y\|^2 \le \langle x - y, P_C x - P_C y\rangle$ *for* $x, y \in H$. *It turns out that both mappings* P_C *and* $2P_C - I$ *are nonexpansive (here* I *is the identity mappping). Recall that a mapping* $T : H \to H$ *is said to be nonexpansive if*

$$\|Tx - Ty\| \le \|x - y\| \tag{2.1}$$

for all $x, y \in H$.

Proposition 2.2. [5] *(The demiclosedness principle for nonexpansive mappings.) Suppose* $T : C \to C$ *is a nonexpansive mapping. Then* $I - T$ *is demiclosed in the sense:*

$$\forall (x_n) \subset C : \ x_n \rightharpoonup x, \ x_n - Tx_n \to y \ \Rightarrow \ (I - T)x = y.$$

Here we always use the notation:

- $x_n \rightharpoonup x$ means that (x_n) converges to x weakly.
- $x_n \to x$ means that (x_n) converges to x strongly (i.e., in norm).

We need the concept of subdifferential and ε-subdifferential of convex functions.

Definition 2.3. [17] Let $\varphi : H \to \mathbb{R}$ be a convex function and let $\varepsilon \ge 0$ be given. The ε-subdifferential of φ at a point $x \in H$ is defined as

$$\partial_\varepsilon \varphi(x) := \{\xi \in H : \varphi(y) \ge \varphi(x) + \langle \xi, y - x\rangle - \varepsilon \quad \forall y \in H\}.$$

Recall that $\partial\varphi(x) := \partial_0\varphi(x)$ is the subdifferential of φ at x; that is,

$$\partial\varphi(x) = \{\xi \in H : \varphi(y) \ge \varphi(x) + \langle \xi, y - x\rangle \quad \forall y \in H\}.$$

If $\partial_\varepsilon\varphi(x) \ne \emptyset$, then we say that φ is ε-subdifferentiable at x.

Lemma 2.4. *In a Hilbert space H, the identity below holds:*

$$2\langle w-v, v-u\rangle = \|w-u\|^2 - \|w-v\|^2 - \|v-u\|^2, \quad u,v,w \in H.$$

Lemma 2.5. *(cf.* [17]*) Let (a_n) be a sequence of nonnegative real numbers such that*

$$a_{n+1} \le a_n + \delta_n, \quad n \ge 1,$$

where (δ_k) is a summable sequence of nonnegative real numbers (i.e., $\sum_{n=1}^{\infty} \delta_n < \infty$). Then $\lim_{n\to\infty} a_n$ exists.

Lemma 2.6. *(*[23, Lemma 2.5], [9, Lemma 2.2]*) Assume (s_n) is a sequence of nonnegative real numbers satisfying the condition:*

$$s_{n+1} \le (1-\lambda_n)s_n + \lambda_n\beta_n + \delta_n, \quad n \ge 0, \tag{2.2}$$

where (λ_n) and (δ_n) are sequences in (0,1) and (β_n) is a sequence in $\mathbb{R}$. Assume

(i) $\sum_{n=1}^{\infty} \lambda_n = \infty$,
(ii) $\limsup_{n\to\infty} \beta_n \le 0$, *and*
(iii) $\sum_{n=1}^{\infty} \delta_n < \infty$.

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

Lemma 2.7. [1, Proposition 2]*,* [16, Lemma 3.4] *(see also* [10, Lemma 2.1]*) Let (α_n) and (β_n) be sequences of nonnegative real numbers. Suppose the following conditions are satisfied:*

(i) $\sum_{n=1}^{\infty} \alpha_n = \infty$;
(ii) $\sum_{n=1}^{\infty} \alpha_n\beta_n < \infty$;
(iii) $\beta_{n+1} - \beta_n \le c\alpha_n$ *for all* $n \ge 1$ *and some constant* $c > 0$.

Then (β_n) converges to zero.

Lemma 2.8. [5] *Let K be a nonempty subset of a real Hilbert space H. Let $\{x_n\}$ be a sequence in H satisfying the properties:*

(i) $\lim_{n\to\infty} \|x_n - x\|$ *exists for each* $x \in K$; *and*
(ii) $\omega_w(x_n) \subset K$, *where* $\omega_w(x_n) = \{\xi \in H : \text{there exists a subsequence } x_{n_k} \rightharpoonup \xi\}$ *is the ω-weak limit point set of* $\{x_n\}$.

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

Lemma 2.9. [11, Lemma 1.5] *Let K be a nonempty closed convex subset of a Hilbert space H. Let $\{x_n\}$ be a sequence in H and $u \in H$. Suppose the conditions below are satisfied:*

- $\omega_w(x_n) \subset K$,
- $\|x_n - u\| \le \|u - P_K u\|$ *for all* n.

Then $x_n \to P_K u$.

We include a necessary and sufficient optimality condition for (1.1) as follows.

Proposition 2.10. *A point $x^* \in C$ solves the nondifferentiable convex optimization problem (1.1) if and only if x^* satisfies the optimality condition (or variational inequality):*

$$\text{There is } s^* \in \partial\varphi(x^*) \text{ such that } \langle s^*, y - x^*\rangle \ge 0 \quad \forall y \in C. \tag{2.3}$$

3. Weak convergence of projected subgradient method

Subgradient methods have been studied by many researchers, see, e.g., [3, 4, 6, 7, 14, 15, 18] and the references cited therein. We begin by rewriting the projected subgradient algorithm (1.8) as the form:

$$x_{n+1} = P_C(x_n - t_n u_n), \quad n = 0, 1, 2, \cdots \tag{3.1}$$

where $t_n = \alpha_n/\eta_n$ for $n \geq 0$.

The inequality (3.2) below will play a fundamental role in our argument of proving convergence of PSA (1.8). We therefore refer it to as the basic inequality.

Lemma 3.1. *Let (x_n) be defined by the PSA* (3.1)*. Then the following inequality, known as the basic inequality, holds:*

$$\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 - 2t_n[\varphi(x_n) - \varphi(x)] + t_n(t_n\|u_n\|^2 + 2\varepsilon_n) \tag{3.2}$$

for all $x \in C$. In particular,

$$\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 - 2t_n\beta_n + t_n(t_n\|u_n\|^2 + 2\varepsilon_n), \tag{3.3}$$

where $x^* \in S$ and $\beta_n = \varphi(x_n) - \varphi^*$ with $\varphi^* = \inf_C \varphi$.

Proof. For $x \in C$, we have, using the nonexpansivity of the projection P_C,

$$\begin{aligned} \|x_{n+1} - x\|^2 &= \|P_C(x_n - t_n u_n) - x\|^2 \\ &\leq \|x_n - x - t_n u_n\|^2 \\ &= \|x_n - x\|^2 - 2t_n\langle u_n, x_n - x\rangle + t_n^2\|u_n\|^2. \end{aligned} \tag{3.4}$$

Since $u_n \in \partial_{\varepsilon_n}\varphi(x_n)$, we get $\langle u_n, x_n - x\rangle \geq \varphi(x_n) - \varphi(x) - \varepsilon_n$. Substituting this into (3.4) and rearranging the terms yields the desired inequality (3.2). □

It is evident that the projection-gradient method (1.2) is no longer straightforwardly applicable to a nondifferentiable objective function φ. In this case, Albert, et al [1] introduced the projected subgradient method which generates a sequence (x_n) according to the following rule:

$$x_{n+1} = P_C\left(x_n - \frac{\alpha_n}{\eta_n}u_n\right), \quad u_n \in \partial_{\varepsilon_n}\varphi(x_n), \tag{3.5}$$

where $x_0 \in C$, $\eta_n = \max\{1, \|u_n\|\}$ and $\varepsilon_n \geq 0$ for each n.

To discuss convergence of the algorithm (3.5) in an infinite-dimensional setting, Albert, et al [1] used the following assumption.

Assumption A: For each fixed $\varepsilon \geq 0$, $\partial_\varepsilon\varphi$ is bounded on bounded sets, i.e., $\bigcup_{x\in K}\partial_\varepsilon\varphi(x)$ is bounded for each bounded subset K of H. [Note: Assumption A is always satisfied in a finite-dimensional Hilbert space.]

The weak convergence of PSA (1.8) was proved in [1]. However, here we give a simpler proof.

Theorem 3.2. [1, Theorem 1] *Let Assumption A be satisfied. Let (x_n) be generated by the projected subgradient method* (3.5)*. Assume that the conditions below are satisfied:*

(i) $\sum_{n=0}^{\infty}\alpha_n = \infty$ *and* $\sum_{n=0}^{\infty}\alpha_n^2 < \infty$.

(ii) (ε_n) *is nonincreasing and* $\varepsilon_n \leq \mu\alpha_n$ *for some constant* $\mu \geq 0$ *and all* n.

Then (x_n) *is weakly convergent to a solution of the nonsmooth optimization problem* (1.1).

Proof. Apply the basic inequality (3.3) of Lemma 3.1 to obtain

$$\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 - 2t_n\beta_n + t_n(t_n\|u_n\|^2 + 2\varepsilon_n), \tag{3.6}$$

where $x^* \in S$, $t_n = \frac{\alpha_n}{\eta_n}$, and $\beta_n = \varphi(x_n) - \varphi(x^*)$. Since $\eta_n \geq \|u_n\|$ and $\varepsilon_n \leq \mu\alpha_n$, it turns out from (3.6) that

$$\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 - \frac{2\alpha_n\beta_n}{\eta_n} + (2\mu+1)\alpha_n^2 \leq \|x_n - x^*\|^2 + (2\mu+1)\alpha_n^2. \tag{3.7}$$

Since $\sum_{n=1}^{\infty} \alpha_n^2 < \infty$ and $\eta_n \geq 1$ for all n, we obtain

- $\lim_{n\to\infty} \|x_n - x^*\|$ exists for each $x^* \in S$;
- $\sum_{n=0}^{\infty} \alpha_n\beta_n < \infty$.

Observing $\|x_{n+1} - x_n\| = \|P_C(x_n - \frac{\alpha_n}{\eta_n}u_n) - x_n\| \leq \frac{\alpha_n}{\eta_n}\|u_n\| \leq \alpha_n$, the fact $u_{n+1} \in \partial_{\varepsilon_{n+1}}\varphi(x_{n+1})$ which implies

$$\varphi(x_n) \geq \varphi(x_{n+1}) + \langle x_n - x_{n+1}, u_{n+1}\rangle - \varepsilon_{n+1}$$

we get

$$\begin{aligned}\beta_{n+1} - \beta_n &= \varphi(x_{n+1}) - \varphi(x_n)\\ &\leq \langle x_{n+1} - x_n, u_{n+1}\rangle + \varepsilon_{n+1}\\ &\leq \|x_n - x_{n+1}\|\|u_{n+1}\| + \varepsilon_n\\ &\leq (M+\mu)\alpha_n.\end{aligned}$$

Here M is a constant such that $M \geq \|u_n\|$ for all n. From the above discussions, we find that Lemma 2.7 is applicable and we obtain that $\beta_n \to 0$, i.e., $\varphi(x_n) \to \varphi(x^*) = \inf_C \varphi$. Then the lower semicontinuity of φ ensures that every weak cluster point of $\{x_n\}$ is a minimum point of φ (i.e., a point in S):

- $x_{n_k} \rightharpoonup \hat{x} \;\Rightarrow\; \hat{x} \in S$.

It turns out that Lemma 2.8 is also applicable with K replaced with S. Therefore, the sequence $\{x_n\}$ converges weakly to a point in S. □

4. Strong convergence of projected subgradient methods

Since strong convergence of the projected subgradient method (3.5) is not valid in general, regularization (or viscosity) techniques ([2, 12, 24]) are needed to guarantee strong convergent modifications of the algorithm (3.5). Maingé [10] introduced such a regularized strongly convergent modification of the projected subgradient method (3.5). His algorithm ([10, Eq. (10), page 902]) is restated as follows: Initializing with $x_0 \in C$ and given $a \in H$, the $(n+1)$th iterate x_{n+1} is defined by the recursion process:

$$x_{n+1} = P_C\left(x_n - \frac{\lambda_n}{\eta_n}d_n\right), \quad u_n \in \partial_{\varepsilon_n}\varphi(x_n), \quad d_n = u_n + \alpha_n(x_n - a), \tag{4.1}$$

where $\eta_n = \max\{\mu, \|d_n\|\}$ with $\mu > 0$.

In order to see the viscosity nature of the algorithm (4.1), we rewrite it in the form

$$x_{n+1} = P_C\left(\left(1-\frac{\alpha_n\lambda_n}{\eta_n}\right)x_n + \frac{\alpha_n\lambda_n}{\eta_n}a - \frac{\lambda_n}{\eta_n}u_n\right). \tag{4.2}$$

Since the map $x \mapsto P_C((1-\beta)x+w)$ is a contraction for any fixed $\beta \in (0,1]$ and $w \in H$, we find that the algorithm (4.1) is of viscosity nature [12], making possible the strong convergence of (4.1) which is the main result of Maingé [10].

Theorem 4.1. [10, Theorem 3.1] *Assume Assumption (A) in Theorem 3.2 and the following conditions:*

(i) $\sum_{n=0}^{\infty}\lambda_n = \infty$ *and* $\sum_{n=0}^{\infty}\lambda_n^2 < \infty$.
(ii) $\lim_{n\to\infty}\alpha_n = 0$, $\sum_{n=0}^{\infty}\alpha_n = \infty$, *and* $\sum_{n=0}^{\infty}\alpha_n\lambda_n = \infty$.
(iii) (ε_n) *is nonincreasing and* $\varepsilon_n \le \mu\lambda_n$ *for some constant* $\mu \ge 0$ *and all* n.

Then (x_n) *is strongly convergent to an optimal solution of the optimization problem* (1.5).

The proof provided by Maingé [10] depends heavily on the following technical lemma.

Lemma 4.2. [10, Lemma 3.1] *Let* (Γ_n) *be a sequence of real numbers that does not decrease at infinity, in the sense that there exists a subsequence* (Γ_{n_j}) *of* (Γ_n) *such that*

$$\Gamma_{n_j} < \Gamma_{n_j+1} \quad \text{for all } j \ge 0.$$

Also consider the sequence of integers $(\tau(n))_{n\ge n_0}$ *defined by*

$$\tau(n) = \max\{k \le n \mid \Gamma_k < \Gamma_{k+1}\}.$$

Then $(\tau(n))_{n\ge n_0}$ *is a nondecreasing sequence verifying*

$$\lim_{n\to\infty}\tau(n) = \infty,$$

and, for all $n \ge n_0$, *the following two estimates hold:*

$$\Gamma_{\tau(n)} \le \Gamma_{\tau(n)+1}, \quad \Gamma_n \le \Gamma_{\tau(n)+1}.$$

In the rest of this section we will extend Theorem 4.1 to more general regularization, i.e., the viscosity approximation methods (see Theorems 4.3 and 4.5). In addition, we will slightly relax the condition (iii) for the choice of the perturbation parameters (ε_n). Our proof is more straightforward and elementary (without employing the technical Lemma 4.2).

4.1. **Viscosity Approximation Method.** The viscosity approximation method (VAM) for optimization was introduced by Attouch [2]. Let H be a real Hilbert space and let $f : H \to \mathbb{R}\cup\{\infty\}$ be a proper, lower-semicontinuous, convex function. Consider the minimization problem

$$\min_{x\in H} f(x). \tag{4.3}$$

Given $z \in H$ and $\varepsilon > 0$. Let u_ε be the unique solution of the (regularized) problem:

$$\varepsilon u_\varepsilon + \partial f(u_\varepsilon) \ni z \quad \text{or} \quad u_\varepsilon = J_{1/\varepsilon}^{\partial f}\left(\frac{1}{\varepsilon}z\right). \tag{4.4}$$

Here $J_\lambda^{\partial f} = (I + \lambda\partial f)^{-1}$ is the resolvent of ∂f. Attouch [2] proved that $\{u_\varepsilon\}$ remains bounded as $\varepsilon \to 0$ if and only if $S_z := (\partial f)^{-1}z = \{v \in H : z \in \partial f(v)\} \neq \emptyset$; in this case, $\{u_\varepsilon\}$ converges, in norm, to the unique point $\hat{u}$ of minimal norm of S_z, that is, $\hat{u} = P_{S_z}(0)$. In particular, if $z = 0$ and the set S of optimal solutions of (4.3) is nonempty, then the solution z_ε of (4.4) with $z = 0$ converges in norm as $\varepsilon \to 0$ to the minimal norm element of S.

Attouch's viscosity technique was extended to nonxpansive mappings by Moudafi [12] (Hilbert spaces) and Xu [24] (Banach spaces).

A viscosity approximation method (VAM) for the projected subgradient method generates a sequence $\{x_n\}$ as follows:

$$x_{n+1} = P_C\left(\left(1 - \frac{\alpha_n\lambda_n}{\eta_n}\right)x_n + \frac{\alpha_n\lambda_n}{\eta_n}h(x_n) - \frac{\lambda_n}{\eta_n}u_n\right), \tag{4.5}$$

where the starting point $x_0 \in C$, $u_n \in \partial_{\varepsilon_n}\varphi(x_n)$, and $h : H \to H$ is a ρ-contraction (i.e., $\|h(x) - h(y)\| \le \rho\|x - y\|$ for all $x, y \in H$ and some $\rho \in [0, 1)$). Clearly, when $h(x) \equiv a$ is constant, VAM (4.5) is reduced to Maingé's algorithm (4.2).

We next analyze the strong convergence property of VAM (4.5), prior to which, however, for convenience we make the following convention.

Convention: The objective function φ of the nonsmooth optimization problem (1.1) is assumed to satisfy Assumption A, as defined in Section 3, in all the convergence theorems below. Assumption A is always assumed in an infinite-dimensional framework, though it is always satisfied in the finite-dimensional framework.

Theorem 4.3. *Assume the following conditions are satisfied:*

(i) $\sum_{n=0}^\infty \lambda_n = \infty$ *and* $\sum_{n=0}^\infty \lambda_n^2 < \infty$.
(ii) $\lim_{n\to\infty} \alpha_n = 0$, $\sum_{n=0}^\infty \alpha_n = \infty$, *and* $\sum_{n=0}^\infty \alpha_n\lambda_n = \infty$.
(iii) (ε_n) *is such that* $\sum_{n=0}^\infty \varepsilon_n\lambda_n < \infty$.
(iv) $\eta_n = \max\{\eta, \|d_n\|\}$, *where* $\eta > 0$ *is a constant and* $d_n = u_n + \alpha_n(x_n - h(x_n))$.

Then (x_n) *is strongly convergent to the solution* q *of the nonsmooth optimization problem* (1.1) *that is the unique fixed point of the contraction* $P_S f$*; equivalently, the unique solution of the variational inequality (VI):*

$$\langle q - h(q), x - q\rangle \ge 0, \quad x \in S. \tag{4.6}$$

Proof. First we prove the boundedness of (x_n). Notice that we can rewrite the algorithm (4.5) as

$$x_{n+1} = P_C\left(x_n - \frac{\lambda_n}{\eta_n}d_n\right), \quad d_n = u_n + \alpha_n(x_n - h(x_n)). \tag{4.7}$$

Now let q be the unique solution of VI (4.6) so that $q = P_S(h(q))$. We have

$$\begin{aligned}\|x_{n+1}-q\|^2 &= \|P_C(x_n - \frac{\lambda_n}{\eta_n}d_n) - q\|^2 \le \|x_n - q - \frac{\lambda_n}{\eta_n}d_n\|^2\\ &= \|x_n-q\|^2 - \frac{2\lambda_n}{\eta_n}\langle x_n - q, d_n\rangle + \frac{\lambda_n^2}{\eta_n^2}\|d_n\|^2\\ &= \|x_n-q\|^2 - \frac{2\lambda_n}{\eta_n}\langle x_n - q, u_n\rangle\\ &\quad - \frac{2\lambda_n\alpha_n}{\eta_n}\langle x_n - q, x_n - h(x_n)\rangle + \frac{\lambda_n^2}{\eta_n^2}\|d_n\|^2. \end{aligned} \tag{4.8}$$

Now since $u_n \in \partial_{\varepsilon_n}\varphi(x_n)$, we have $\varphi^* = \varphi(q) \ge \varphi(x_n) + \langle u_n, q - x_n\rangle - \varepsilon_n$. This implies that

$$\langle x_n - q, u_n\rangle \ge \beta_n - \varepsilon_n \tag{4.9}$$

where $\beta_n = \varphi(x_n) - \varphi^*$. Substituting (4.9) into (4.8) and noting $\eta_n = \max\{\eta, \|d_n\|\}$, we obtain

$$\|x_{n+1}-q\|^2 + \frac{2\lambda_n\beta_n}{\eta_n} \le \|x_n-q\|^2 - \frac{2\lambda_n\alpha_n}{\eta_n}\langle x_n - q, x_n - h(x_n)\rangle + \frac{2}{\eta}\varepsilon_n\lambda_n + \lambda_n^2. \tag{4.10}$$

Observing

$$\langle x_n - q, x_n - h(x_n)\rangle = \|x_n - q\|^2 + \langle x_n - q, q - h(x_n)\rangle, \tag{4.11}$$

we may rewrite (4.10) equivalently as

$$\|x_{n+1}-q\|^2 + \frac{2\lambda_n\beta_n}{\eta_n} \le \left(1 - \frac{2\lambda_n\alpha_n}{\eta_n}\right)\|x_n-q\|^2 + \frac{2\lambda_n\alpha_n}{\eta_n}\langle x_n - q, h(x_n) - q\rangle + \delta_n, \tag{4.12}$$

where $\delta_n = \frac{2}{\eta}\varepsilon_n\lambda_n + \lambda_n^2$. Note that $\sum_{n=0}^{\infty}\delta_n < \infty$ due to the conditions (i) and (iii). Noticing

$$\begin{aligned}\langle x_n - q, h(x_n) - q\rangle &= \langle x_n - q, h(x_n) - h(q)\rangle + \langle x_n - q, h(q) - q\rangle\\ &\le \rho\|x_n - q\|^2 + \langle x_n - q, h(q) - q\rangle,\end{aligned}$$

we further obtain from (4.12)

$$\begin{aligned}\|x_{n+1}-q\|^2 + \frac{2\lambda_n\beta_n}{\eta_n} &\le \left(1 - (1-\rho)\frac{2\lambda_n\alpha_n}{\eta_n}\right)\|x_n-q\|^2\\ &\quad + \frac{2\lambda_n\alpha_n}{\eta_n}\langle x_n - q, h(q) - q\rangle + \delta_n.\end{aligned} \tag{4.13}$$

Consequently, we get, using the Cauchy-Schwartz inequality $\langle x_n - q, h(q) - q\rangle \le \|x_n - q\|\|h(q) - q\|$,

$$\begin{aligned}\|x_{n+1}-q\|^2 &\le \left(1 - (1-\rho)\frac{2\lambda_n\alpha_n}{\eta_n}\right)\|x_n-q\|^2\\ &\quad + (1-\rho)\frac{2\lambda_n\alpha_n}{\eta_n}\|x_n - q\| \cdot \frac{\|f(q) - q\|}{1-\rho} + \delta_n.\end{aligned}$$

It turns out that

$$\|x_{n+1} - q\|^2 \le \max\left\{\|x_n - q\|^2, \|x_n - q\| \cdot \frac{\|h(q) - q\|}{1-\rho}\right\} + \delta_n$$
$$\le \max\left\{\|x_n - q\|^2, \frac{\|h(q) - q\|^2}{(1-\rho)^2}\right\} + \delta_n.$$

By induction, it is easy to prove that, for all $n \ge 0$,

$$\|x_n - q\|^2 \le \max\left\{\|x_0 - q\|^2, \frac{\|h(q) - q\|^2}{(1-\rho)^2}\right\} + \sum_{i=0}^{n-1} \delta_i.$$

In particular, (x_n) is bounded (hence, so is (u_n) by Assumption A), since

$$\sum_{n=0}^{\infty} \delta_n < \infty.$$

Next we rewrite (4.13) as

$$s_{n+1} \le (1-\mu_n)s_n + \mu_n \sigma_n + \delta_n, \tag{4.14}$$

where

$$s_n = \|x_n - q\|^2, \quad \mu_n = (1-\rho)\frac{2\lambda_n \alpha_n}{\eta_n}, \quad \sigma_n = \frac{1}{1-\rho}\left(\langle x_n - q, h(q) - q\rangle - \frac{\beta_n}{\alpha_n}\right).$$

Notice that we have $\mu_n \to 0$ and $\sum_{n=0}^{\infty} \mu_n = \infty$. It is also clear that (σ_n) is bounded from above since (x_n) is bounded. Hence, $\limsup_{n\to\infty} \sigma_n$ exists (and is finite). Let (x_{n_k}) be a subsequence of (x_n) such that

$$\limsup_{n\to\infty} \sigma_n = \lim_{k\to\infty} \sigma_{n_k} = \lim_{k\to\infty} \frac{1}{1-\rho}\left(\langle x_{n_k} - q, h(q) - q\rangle - \frac{\beta_{n_k}}{\alpha_{n_k}}\right). \tag{4.15}$$

Again since the sequence (x_n) is bounded, we may further assume that $x_{n_k} \to x'$ weakly as $k \to \infty$. It then follows from (4.15) that $\lim_{k\to\infty} \frac{\beta_{n_k}}{\alpha_{n_k}}$ exists; hence, $\left(\frac{\beta_{n_k}}{\alpha_{n_k}}\right)$ is bounded. This implies that (for $\alpha_n \to 0$)

$$\lim_{k\to\infty} \beta_{n_k} = \lim_{k\to\infty} \alpha_{n_k}\left(\frac{\beta_{n_k}}{\alpha_{n_k}}\right) = 0.$$

Namely, $\varphi(x_{n_k}) \to \varphi^*$. Consequently, $\varphi(x') = \varphi^*$ by weak lower semicontinuity of φ. Thus, $x' \in S$. Now we have

$$\limsup_{n\to\infty} \sigma_n = \lim_{k\to\infty} \frac{1}{1-\rho}\left(\langle x_{n_k} - q, h(q) - q\rangle - \frac{\beta_{n_k}}{\alpha_{n_k}}\right)$$
$$\le \lim_{k\to\infty} \frac{1}{1-\rho}\langle x_{n_k} - q, h(q) - q\rangle$$
$$= \frac{1}{1-\rho}\langle x' - q, h(q) - q\rangle \le 0$$

since q is the projection of $h(q)$ onto S and $x' \in S$.

Finally, applying Lemma 2.6 to (4.14) we obtain $\|x_n - q\|^2 \to 0$; namely, $x_n \to q$ in norm. The proof is complete. □

Remark 4.4. The above proof does not require that (ε_n) be nonincreasing; also the assumption $\varepsilon_n \le \mu\lambda_n$ (for all n) is replaced with the weaker condition $\sum_{n=1}^{\infty} \varepsilon_n\lambda_n < \infty$ (notice that the condition $\sum_{n=1}^{\infty} \lambda_n^2 < \infty$ is coupled).

An example of choice of the parameters (λ_n), (α_n), and (ε_n) is as follows:

$$\lambda_n = \frac{1}{(n+1)^{\lambda}}, \quad \alpha_n = \frac{1}{(n+1)^{\alpha}}, \quad \varepsilon_n = \frac{1}{(n+1)^{\varepsilon}},$$

where $\lambda > 0$, $\alpha > 0$, $\varepsilon > 0$.

To satisfy the conditions (i)-(iii) in Maingé's theorem (i.e., Theorem 4.1), the parameters λ, α, ε must satisfy the conditions

$$\frac{1}{2} < \lambda \le 1,\ 0 < \alpha \le 1,\ 0 < \lambda + \alpha \le 1,\ \varepsilon \ge \lambda. \tag{4.16}$$

While to satisfy the conditions (i)-(iii) of our Theorem 4.3, the parameters λ, α, ε need satisfy the conditions which are strictly weaker than (4.16):

$$\frac{1}{2} < \lambda \le 1,\ 0 < \alpha \le 1,\ 0 < \lambda + \alpha \le 1,\ \varepsilon + \lambda > 1. \tag{4.17}$$

Indeed, it is easy to find that $\varepsilon \ge \lambda \ \Rightarrow\ \varepsilon + \lambda \ge 2\lambda > 1$, but evidently, not vice versa.

Alternatively, we introduce another VAM of the following form:

$$x_{n+1} = \gamma_n h(x_n) + (1-\gamma_n) P_C\left(x_n - \frac{\alpha_n}{\eta_n} u_n\right), \quad u_n \in \partial_{\varepsilon_n}\varphi(x_n), \tag{4.18}$$

where $\gamma_n \in (0,1)$ for all n, h is a ρ-contraction on C for some $\rho \in [0,1)$, and $\eta_n = \max\{\eta, \|u_n\|\}$ with $\eta > 0$.

Theorem 4.5. *Suppose the following conditions are satisfied:*

(i) $\lim_{n\to\infty} \gamma_n = 0$, $\sum_{n=0}^{\infty} \gamma_n = \infty$,
(ii) $\sum_{n=0}^{\infty} \alpha_n^2 < \infty$,
(iii) $\lim_{n\to\infty} \frac{\gamma_n}{\alpha_n} = 0$,
(iv) $\sum_{n=0}^{\infty} \varepsilon_n \alpha_n < \infty$.

Then the sequence $\{x_n\}$ generated by VAM (4.18) *converges in norm to the solution q of the nonsmooth optimization problem* (1.1) *which is also the unique solution of VI* (4.6).

Proof. Let $q \in S$ be the unique solution of VI (4.6), i.e., $q = P_S(h(q))$.

We first show that $\{x_n\}$ is bounded. As a matter of fact, we have by the convexity of $\|\cdot\|^2$ and the basic inequality (3.3),

$$\begin{aligned}\|x_{n+1} - q\|^2 &= \|\gamma_n(h(x_n) - q) + (1-\gamma_n)(P_C\left(x_n - \frac{\alpha_n}{\eta_n}u_n\right) - q)\|^2 \\ &\le \gamma_n\|h(x_n) - q\|^2 + (1-\gamma_n)\|P_C\left(x_n - \frac{\alpha_n}{\eta_n}u_n\right) - q\|^2 \\ &\le \gamma_n\|h(x_n) - q\|^2 + (1-\gamma_n)\left(\|x_n - q\|^2 + (\frac{\alpha_n}{\eta_n}\|u_n\|)^2 + 2\frac{\alpha_n}{\eta_n}\varepsilon_n\right).\end{aligned} \tag{4.19}$$

By the Cauchy-Schwartz inequality we obtain

$$\begin{aligned}\|h(x_n)-q\|^2 &\le (\|h(x_n)-h(q)\|+\|h(q)-q\|)^2\\ &\le (\rho\|x_n-q\|+\|h(q)-q\|)^2\\ &= \rho^2\|x_n-q\|^2+2\rho\|h(q)-q\|\|x_n-q\|+\|h(q)-q\|^2\\ &\le \rho^2\|x_n-q\|^2+\rho\{(1-\rho)\|x_n-q\|^2\\ &\quad +\frac{1}{1-\rho}\|h(q)-q\|^2\}+\|h(q)-q\|^2\\ &= \rho\|x_n-q\|^2+\frac{1}{1-\rho}\|h(q)-q\|^2. \end{aligned}\tag{4.20}$$

Since $\|u_n\|/\eta_n \le 1$ and $\eta_n \ge \eta$, substituting (4.20) into (4.19), we obtain

$$\|x_{n+1}-q\|^2 \le (1-(1-\rho)\gamma_n)\|x_n-q\|^2+\frac{\gamma_n}{1-\rho}\|h(q)-q\|^2+\alpha_n^2+\frac{2}{\eta}\alpha_n\varepsilon_n.$$

Setting $\delta_n=\alpha_n^2+\frac{2}{\eta}\alpha_n\varepsilon_n$ which satisfies $\sum_{n=1}^{\infty}\delta_n<\infty$ (due to conditions (ii) and (iv)), we further obtain

$$\|x_{n+1}-q\|^2 \le \max\{\|x_n-q\|^2,\frac{1}{(1-\rho)^2}\|h(q)-q\|^2\}+\delta_n.$$

By induction, we get

$$\|x_n-q\|^2 \le \max\{\|x_0-q\|^2,\frac{1}{(1-\rho)^2}\|h(q)-q\|^2\}+\sum_{i=0}^{n-1}\delta_i$$

for all $n\ge 0$. Hence, $\{x_n\}$ is bounded. Let M be such that $M\ge\|h(x_n)-x_n\|$ for all $n\ge 0$. It follows from (4.18) (also noting $\eta_n\ge\|u_n\|$) that $\|x_{n+1}-x_n\|\le M\gamma_n+\alpha_n\to 0$. Namely, $\{x_n\}$ is asymptotically regular.

Again by the basic inequality (3.3), we derive that

$$\begin{aligned}\|x_{n+1}-q\|^2 &= \|(1-\gamma_n)(P_C\left(x_n-\frac{\alpha_n}{\eta_n}u_n\right)-q)+\gamma_n(h(x_n)-q)\|^2\\ &\le (1-\gamma_n)^2\|P_C\left(x_n-\frac{\alpha_n}{\eta_n}u_n\right)-q\|^2+2\gamma_n\langle h(x_n)-q,x_{n+1}-q\rangle\\ &\le (1-\gamma_n)^2\left(\|x_n-q\|^2-\frac{2\alpha_n\beta_n}{\eta_n}+(\frac{\alpha_n}{\eta_n}\|u_n\|)^2+\frac{2\alpha_n\varepsilon_n}{\eta_n}\right)\\ &\quad +2\gamma_n(\langle h(x_n)-q,x_{n+1}-q\rangle. \end{aligned}\tag{4.21}$$

Since

$$\begin{aligned}\langle h(x_n)-q,x_{n+1}-q\rangle &= \langle h(x_n)-h(q),x_{n+1}-q\rangle+\langle h(q)-q,x_{n+1}-q\rangle\\ &\le \rho\|x_n-q\|\cdot\|x_{n+1}-q\|+\langle h(q)-q,x_{n+1}-q\rangle\\ &\le (\rho/2)(\|x_n-q\|^2+\|x_{n+1}-q\|^2)+\langle h(q)-q,x_{n+1}-q\rangle \end{aligned}$$

and observing $(\eta_n/\|u_n\|) \le 1$ and $\eta_n \ge 1$, we get from (4.21)

$$\|x_{n+1}-q\|^2 \le (1-\gamma_n)^2\|x_n-q\|^2 - \frac{2\alpha_n\beta_n(1-\gamma_n)}{\eta_n} + \delta_n + \gamma_n[\rho(\|x_n-q\|^2+\|x_{n+1}-q\|^2) + 2\langle h(q)-q, x_{n+1}-q\rangle].$$

It turns out that

$$\|x_{n+1}-q\|^2 \le \frac{(1-\gamma_n)^2+\rho\gamma_n}{1-\rho\gamma_n}\|x_n-q\|^2 - \frac{2\alpha_n\beta_n(1-\gamma_n)}{\eta_n(1-\rho\gamma_n)} + \frac{\delta_n}{1-\rho\gamma_n} + \frac{2\gamma_n}{1-\rho\gamma_n}\langle h(q)-q, x_{n+1}-q\rangle. \tag{4.22}$$

Setting $s_n = \|x_n-q\|^2$,

$$\tau_n = \frac{2(1-\rho)\gamma_n-\gamma_n^2}{1-\rho\gamma_n}, \quad \mu_n = \frac{2}{2(1-\rho)-\gamma_n}\langle h(q)-q, x_{n+1}-q\rangle - \frac{2\alpha_n\beta_n(1-\gamma_n)}{\eta_n\gamma_n(2(1-\rho)-\gamma_n)},$$

and $\hat{\delta}_n = \frac{\delta_n}{1-\rho\gamma_n}$, we can rewrite (4.22) as

$$s_{n+1} \le (1-\tau_n)s_n + \tau_n\mu_n + \hat{\delta}_n. \tag{4.23}$$

Since $\frac{\tau_n}{\gamma_n} \to 2(1-\rho) > 0$, we have $\tau_n \approx \gamma_n$; thus $\tau_n \to 0$ and $\sum_n \tau_n = \infty$. We also have $\sum_{n=1}^{\infty}\hat{\delta}_n < \infty$. In order to apply Lemma 2.6, it suffices to prove that $\limsup_{n\to\infty}\mu_n \le 0$. It is evident that $\{\mu_n\}$ is bounded from above by $\sup_n\{[2/(2(1-\rho)-\gamma_n)]\|h(q)-q\|\cdot\|x_n-q\|\}$. [We may assume that $\sup_{n\ge 0}\gamma_n < 2(1-\rho)$ since $\gamma_n \to 0$.] Thus, $\limsup_{n\to\infty}\mu_n$ exists and is finite. Take a subsequence $\{\mu_{n_k}\}$ of $\{\mu_n\}$ such that $x_{n_k} \rightharpoonup \hat{x}$ and (noticing $\gamma_n \to 0$)

$$\begin{aligned}\limsup_{n\to\infty}\mu_n &= \lim_{k\to\infty}\mu_{n_k}\\ &= \lim_{k\to\infty}\left(\frac{2}{2(1-\rho)-\gamma_{n_k}}\langle h(q)-q, x_{n_k}-q\rangle - \frac{2\alpha_{n_k}\beta_{n_k}(1-\gamma_{n_k})}{\eta_{n_k}\gamma_{n_k}(2(1-\rho)-\gamma_{n_k})}\right)\\ &= \lim_{k\to\infty}\left(\frac{1}{1-\rho}\langle h(q)-q, \hat{x}-q\rangle - \frac{2\alpha_{n_k}\beta_{n_k}}{2(1-\rho)\eta_{n_k}\gamma_{n_k}}\right).\end{aligned} \tag{4.24}$$

This implies that $\lim_{k\to\infty}\frac{\alpha_{n_k}\beta_{n_k}}{\eta_{n_k}\gamma_{n_k}}$ exists. In particular, $\{\frac{\alpha_{n_k}\beta_{n_k}}{\gamma_{n_k}}\}$ is bounded since $\{\eta_n\}$ is bounded. Suppose $\frac{\alpha_{n_k}\beta_{n_k}}{\gamma_{n_k}} \le M$ for some constant $M \ge 0$. It turns out from condition (iii) that

$$\beta_{n_k} \le M\frac{\gamma_{n_k}}{\alpha_{n_k}} \to 0.$$

Namely, $\varphi(x_{n_k}) \to \varphi^*$. The weak lower semicontinuity of φ implies $\hat{x} \in S$, and we have from (4.24)

$$\limsup_{n\to\infty}\mu_n = \lim_{k\to\infty}\mu_{n_k} \le \frac{1}{1-\rho}\langle h(q)-q, \hat{x}-q\rangle \le 0.$$

Consequently, Lemma 2.6 is applicable to (4.23) and we obtain $s_n \to 0$. Namely, $x_n \to q$ in norm. □

Remark 4.6. The following choices of $\{\alpha_n\}$, $\{\gamma_n\}$, $\{\varepsilon_n\}$ satisfy the conditions (i)-(iv) of Theorem 4.5:

$$\alpha_n = \frac{1}{(n+1)^\alpha}, \quad \gamma_n = \frac{1}{(n+1)^\gamma}, \quad \varepsilon_n = \frac{1}{(n+1)^\varepsilon}; \quad \frac{1}{2} < \alpha < \gamma \leq 1, \ \varepsilon > 1-\alpha.$$

5. Forcing strong convergence

In [21], Solodov and Svaiter introduced a method, referred to as forcing strong convergence, for a modification of the proximal point algorithm of Rockafellar [19] in order to have strong convergence (see [22] for another strongly convergent modification). This method was extended by Bello Cruz and Iusem [4] to the case of nonsmooth convex objectives, using exact subgradients. The basic idea is to add additional projections onto appropriately constructed closed convex subsets. We here make a further extension by making use of ε-subgradients.

Bello Cruz and Iusem's algorithm for the convex nondifferentiable minimization problem (1.1) reads as follows.

Initializing $x_0 \in C$ and when $x_n \in C$ and $u_n \in \partial\varphi(x_n)$ are given, we set $\beta_n := \varphi(x_n) - \varphi^* \geq 0$ with $\varphi^* := \inf_C \varphi$ and define two closed half-spaces H_n and W_n by

$$H_n := \{x \in H : \langle x - x_n, u_n\rangle + \beta_n \leq 0\}, \tag{5.1}$$

$$W_n := \{x \in H : \langle x - x_n, x_0 - x_n\rangle \leq 0\}. \tag{5.2}$$

The $(n+1)$th iterate x_{n+1} is defined as the projection of x_0 to $H_n \cap W_n \cap C$:

$$x_{n+1} = P_{H_n \cap W_n \cap C} x_0. \tag{5.3}$$

Using ε-subdifferential, we can extend Bello Cruz and Iusem's algorithm (5.1)-(5.3) as follows. Initializing $x_0 \in C$ and when $x_n \in C$ and $u_n \in \partial_{\varepsilon_n}\varphi(x_n)$ are constructed, we set $\beta_n = \varphi(x_n) - \varphi^*$ and define x_{n+1} in the following way:

$$\begin{cases} H_n = \{x \in H : \langle x - x_n, u_n\rangle \leq \varepsilon_n - \beta_n\}, & (5.4a) \\ W_n = \{x \in H : \langle x - x_n, x_0 - x_n\rangle \leq 0\}, & (5.4b) \\ x_{n+1} = P_{H_n \cap W_n \cap C} x_0. & (5.4c) \end{cases}$$

Lemma 5.1. *Suppose $S \neq \emptyset$. Then we have*

$$H_n \cap W_n \cap C \supset S \tag{5.5}$$

for all $n \geq 0$. Consequently, the algorithm (5.4) *is well-defined.*

Proof. We prove (5.5) by induction. First observe $S \subset C$. Now for $n = 0$, we have $W_0 = H$ and by the ε-subdifferential inequality, we have, for each $x^* \in S$,

$$\varphi^* = \varphi(x^*) \geq \varphi(x_0) + \langle u_0, x^* - x_0\rangle - \varepsilon_0.$$

Namely, $0 \geq \beta_0 + \langle u_0, x^* - x_0\rangle - \varepsilon_0$. It turns out that $x^* \in H_0$; hence $S \subset H_0$. This proves (5.5) for the case of $n = 0$.

Next assume that (5.5) holds for $n > 0$. Thus, $H_n \cap W_n \cap C$ is nonempty and x_{n+1} is well-defined; moreover, we have

$$\langle x_0 - x_{n+1}, z - x_{n+1}\rangle \leq 0, \quad z \in H_n \cap W_n \cap C. \tag{5.6}$$

In particular, (5.6) holds for all $x^* \in S$; consequently, $S \subset W_{n+1}$. Applying again the ε-subdifferential inequality, we deduce that, for $x^* \in S$,

$$\varphi^* = \varphi(x^*) \geq \varphi(x_{n+1}) + \langle x^* - x_{n+1}, u_{n+1}\rangle - \varepsilon_{n+1}.$$

This can be rewritten as

$$\langle x^* - x_{n+1}, u_{n+1}\rangle \leq \varepsilon_{n+1} - \beta_{n+1}.$$

It turns out that $x^* \in H_{n+1}$ and we have proved that $x^* \in H_{n+1} \cap W_{n+1} \cap C$. That is, (5.5) holds for $n+1$. This completes the proof of (5.5) for all $n \geq 0$. □

Lemma 5.2. *Suppose S is nonempty and let (x_n) be generated by the algorithm* (5.4). *Then we have*

(i) *(x_n) is bounded;*
(ii) *$\|x_0 - x_n\|$ is increasing;*
(iii) $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$.

Proof. By (5.3), $\|x_{n+1} - x_0\| \leq \|x^* - x_0\|$ for all n and $x^* \in S$. This proves (i).

Since $x_{n+1} \in W_n$, it follows from (5.4b) that

$$\langle x_{n+1} - x_n, x_n - x_0\rangle \geq 0. \tag{5.7}$$

Applying Lemma 2.4 yields

$$\|x_{n+1} - x_0\|^2 - \|x_{n+1} - x_n\|^2 - \|x_n - x_0\|^2 \geq 0.$$

This immediately implies (ii). We also have

$$\|x_{n+1} - x_n\|^2 \leq \|x_{n+1} - x_0\|^2 - \|x_n - x_0\|^2$$

and (iii) follows; indeed, we have $\sum_{n=0}^{\infty} \|x_{n+1} - x_n\|^2 < \infty$. □

Theorem 5.3. *Suppose S is nonempty and $\varepsilon_n \to 0$. Then the sequence (x_n) be generated by the algorithm* (5.4). *converges in norm to $P_S x_0$.*

Proof. Since $x_{n+1} \in H_n$ we get

$$\langle x_{n+1} - x_n, u_n\rangle + \beta_n \leq \varepsilon_n. \tag{5.8}$$

Now since $\|x_{n+1} - x_n\| \to 0$ and (u_n) is bounded, it follows from (5.8) that $\beta_n \to 0$, namely, $\varphi(x_n) \to \varphi^*$. We therefore have by weak lower semicontinuity of φ

$$\omega_w(x_k) \subset S.$$

On the other hand, by (5.4b), we find that $x_n = P_{W_n} x_0$. Since $S \subset W_n$, we obtain that

$$\|x_0 - x_n\| \leq \|x_0 - P_S x_0\|$$

for all n. We can now apply Lemma 2.9 to obtain $x_n \to P_S x_0$ in norm. □

6. Projected subgradient CQ algorithm

The CQ method was first introduced in [13] (see also [11]) for the Krasnoselskii-Mann method of nonexpasive mappings in a Hilbert space as follows. Let $T : C \to C$ be a nonexpansive mapping such that $Fix(T) \neq \emptyset$. Initializing $x_0 \in C$, the CQ method defines (x_n) as follows:

$$\begin{cases} y_n = \alpha_n x_n + (1 - a_n)Tx_n, \\ C_n = \{z \in C : \|y_n - z\| \le \|x_n - z\|\}, \\ Q_n = \{z \in C : \langle x_n - z, x_n - x_0\rangle \le 0\}, \\ x_{n+1} = P_{C_n \cap Q_n} x_0. \end{cases} \tag{6.1}$$

It is proved [13] that if $\{\alpha_n\} \subset [0, \alpha]$ for some $\alpha \in (0, 1)$, then the sequence $\{x_n\}$ generated by the CQ algorithm (6.1) converges in norm to $P_{Fix(T)}x_0$.

Below we shall make an adaptation of (6.1) for the nondifferentiable minimization problem (1.1). Our algorithm, referred to as projected subgradient CQ algorithm, generates a sequence $\{x_n\}$ of iterates as follows.

$$\begin{cases} y_n = P_C(x_n - (\lambda_n/\eta_n)u_n), \quad u_n \in \partial_{\varepsilon_n}\varphi(x_n), \quad \eta_n = \max\{\eta, \|u_n\|\} & \text{(6.2a)} \\ C_n = \{z \in H : \|y_n - z\|^2 \le \|x_n - z\|^2 - 2(\lambda_n/\eta_n)\beta_n + \delta_n\}, & \text{(6.2b)} \\ Q_n = \{z \in H : \langle x_n - z, x_n - x_0\rangle \le 0\}, & \text{(6.2c)} \\ x_{n+1} = P_{C_n \cap Q_n \cap C} x_0. & \text{(6.2d)} \end{cases}$$

where $\beta_n = \varphi(x_n) - \varphi^*$ and $\delta_n = \lambda_n^2 + (2/\eta)\lambda_n\varepsilon_n$.

Using Lemma 2.4, we see that Q_n is equivalently defined by

$$Q_n = \{z \in C : \|z - x_n\|^2 + \|x_n - x_0\|^2 \le \|z - x_0\|^2\}. \tag{6.3}$$

The convergence of the algorithm (6.2) is presented below.

Theorem 6.1. *Assume the conditions:*

(i) $\sum_{n=0}^{\infty} \lambda_n = \infty$ *and* $\sum_{n=0}^{\infty} \lambda_n^2 < \infty$,
(ii) $\varepsilon_n \to 0$ *and* $\sum_{n=0}^{\infty} \lambda_n\varepsilon_n < \infty$

Let (x_n) *be generated by the projected subgradient CQ algorithm* (6.2)*. We have*

(a) (x_n) *is a minimizing sequence in the sense:*

$$\liminf_{n\to\infty} \varphi(x_n) = \varphi^* := \inf_{x\in C} \varphi(x). \tag{6.4}$$

(b) (x_n) *has a subsequence convergent in norm to* $P_S x_0$.

If, in addition, $(1/\lambda_n)\|x_{n+1} - x_n\|^2 \to 0$*, then* $\varphi(x_n) \to \varphi^*$ *and* $x_n \to P_S x_0$ *in norm.*

Proof. By the basic inequality (3.3), we have, for each $x^* \in S$,

$$\begin{aligned} \|y_n - x^*\|^2 &= \|P_C(x_n - (\lambda_n/\eta_n)u_n) - x^*\|^2 \\ &\le \|x_n - x^*\|^2 - 2(\lambda_n/\eta_n)\beta_n + [(\lambda_n/\eta_n)\|u_n\|]^2 + 2(\lambda_n/\eta_n)\varepsilon_n. \end{aligned}$$

Noticing $\eta_n = \max\{\eta, \|u_n\|\} \ge \|u_n\|$, we obtain

$$\|y_n - x^*\|^2 \le \|x_n - x^*\|^2 - 2(\lambda_n/\eta_n)\beta_n + \lambda_n^2 + \frac{2}{\eta}\lambda_n\varepsilon_n. \tag{6.5}$$

This shows that $x^* \in C_n$ and $S \subset C_n$ for all $n \ge 0$.

We now prove that $S \subset C_n \cap Q_n \cap C$ for all $n \geq 0$. This is trivial when $n = 0$ since $Q_0 = H$. Assume $S \subset C_n \cap Q_n \cap C$ for some $n > 0$. By (6.2d), x_{n+1} is the projection of x_0 onto $C_n \cap Q_n \cap C$. It turns out that

$$\langle x_{n+1} - x_0, x_{n+1} - z \rangle \leq 0, \quad z \in C_n \cap Q_n \cap C. \tag{6.6}$$

This inequality holds particularly for $z \in S$, which exactly means that $z \in Q_{n+1}$; i.e., $S \subset Q_{n+1}$. Consequently, the sequence $\{x_n\}$ is well-defined for all $n \geq 0$.
Also the definition of Q_n implies that $x_n = P_{Q_n} x_0$. Since $S \subset Q_n$, we get $\|x_n - x_0\| \leq \|q - x_0\|$ for $q \in S$. It turns out that (x_n) is bounded, and

$$\|x_n - x_0\| \leq \|q - x_0\| \quad \text{with } q = P_S x_0. \tag{6.7}$$

Moreover, since $x_{n+1} \in Q_n$, we get from (6.3)

$$\|x_{n+1} - x_n\|^2 + \|x_n - x_0\|^2 \leq \|x_{n+1} - x_0\|^2. \tag{6.8}$$

It follows that $\|x_n - x_0\| \leq \|x_{n+1} - x_0\|$ and thus, $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$.

On the other hand, by (6.2a), we get

$$\|y_n - x_n\| \leq \frac{\lambda_n}{\eta_n} \|u_n\| \leq \lambda_n \to 0.$$

Now since $x_{n+1} \in C_n$, we have from (6.2b)

$$\|y_n - x_{n+1}\|^2 \leq \|x_n - x_{n+1}\|^2 - \frac{\lambda_n}{\eta_n}(2\beta_n) + \delta_n \to 0.$$

It follows that

$$\frac{2\lambda_n}{\eta_n}\beta_n \leq \|x_n - x_{n+1}\|^2 + \delta_n \tag{6.9}$$

and (noticing that $\{\eta_n\}$ is bounded)

$$\sum_{n=1}^{\infty} \lambda_n \beta_n < \infty. \tag{6.10}$$

As $\sum_{n=1}^{\infty} \lambda_n = \infty$, we get from (6.10) that $\liminf_{n\to\infty} \beta_n = 0$. Let (x_{n_i}) be a subsequence of (x_n) such that $\lim_{i\to\infty} \beta_{n_i} = 0$. With no loss of generality, we may assume that $x_{n_i} \rightharpoonup \hat{x}$. By the w-lsc of φ, it turns out that $\varphi(\hat{x}) \leq \liminf_{i\to\infty} \varphi(x_{n_i}) = \varphi^*$. Hence, $\hat{x} \in S$. Applying Lemma 2.9 to (6.7) (with respect to the subsequence $\{x_{n_i}\}$) we obtain $x_{n_i} \to P_S x_0$ in norm.

Finally, from (6.9) we get

$$\beta_n \leq \frac{M}{2}\left(\frac{\|x_n - x_{n+1}\|^2}{\lambda_n} + \frac{\delta_n}{\lambda_n}\right), \tag{6.11}$$

where $M \geq \eta_n$ for all $n \geq 0$. Since $\delta_n/\lambda_n \leq \lambda_n + (2/\eta)\varepsilon_n \to 0$, it follows from (6.11) that $\beta_n \to 0$ under the condition $\|x_n - x_{n+1}\|^2/\lambda_n \to 0$. This implies that $\varphi(x_n) \to \varphi^*$; hence, $\omega_w(x_n) \subset S$ and the equation (6.7) makes Lemma 2.9 applicable to the full sequence $\{x_n\}$, which yields that $x_n \to P_S x_0$. This completes the proof. $\square$

7. Conclusions

We have studied the projected subgradient method (PSM) in the setting of an infinite-dimensional Hilbert space. We have provided a simpler proof of the weak convergence theorem, due to of Albert, et al [1]. We have applied the viscosity approximation method (VAM) of Attouch [2] and Moudafi [12] to obtain two regularized PSMs (Theorems 4.3 and 4.5) which are strongly convergent to a solution of the nondifferentiable convex optimization problem (1.1). We have also extended the forcing strong convergence technique of Solodov and Svaiter [21] (see also [4]) to obtain a strongly convergent subgradient algorithm in Theorem 5.3. Finally, we have extended the CQ algorithm of Nakajo and Takahashi [13] to obtain the projected subgradient CQ algorithm (6.2) and its strong convergence in Theorem 6.1.

Acknowledgments. The authors were grateful to the referees for their helpful comments which improved the presentation of this manuscript.

References

[1] Y.I. Albert, A.N. Iusem, M.V. Solodov, *On the projected subgradient method for nonsmooth convex optimization in a Hilbert space*, Math. Program., **81**(1998), 23-35.

[2] H. Attouch, *Viscosity solutions of minimization problems*, SIAM J. Optim., **6**(1996), no. 3, 769-806.

[3] A. Beck, M. Teboulle, *Mirror descent and nonnlinear projected subgradient methods for convex optimization*, Operations Research Letters, **31**(2003), 167-175.

[4] J.Y. Bello Cruz, A.N. Iusem, *A strongly convergent method for nonsmooth convex minimization in Hilbert spaces*, Numerical Functional Analysis and Optimization, **32**(2011), no. 10, 1009-1018.

[5] K. Goebel, W.A. Kirk, *Topics in Metric Fixed Point Theory*, Cambridge University Press, New York, 1990.

[6] D.V. Hieu, *Projected subgradient algorithms on systems of equilibrium problems*, Optim. Lett., **12**(2018), 551-566.

[7] K. Hishinuma, H. Iiduka, *Fixed point quasiconvex subgradient method*, European Journal of Operational Research, **282**(2020), no. 2, 428-437.

[8] E.S. Levitin, B.T. Polyak, *Constrained minimization methods*, Zh. Vychisl. Mat. Mat. Fiz., **6**(1966), 787-823.

[9] G. Lopez, V. Martin, H.K. Xu, *Perturbation techniques for nonexpansive mappings with applications*, Nonlinear Analysis: Real World Applications, **10**(2009), 2369-2383.

[10] P.-E. Maingé, *Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization*, Set-Valued Anal., **16**(2008), 899-912.

[11] C. Martinez-Yanes, H.K. Xu, *Strong convergence of the CQ method for fixed point iteration processes*, Nonlinear Anal., **64**(2006), 2400-2411.

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

[13] K. Nakajo, W. Takahashi, *Strong convergence theorems for nonexpansive mappings and nonexpansive semigroups*, J. Math. Anal. Appl., **279**(2003), 372-379.

[14] A. Nedić, A. Ozdaglar, *Subgradient methods for saddle-point problems*, J. Optim. Theory Appl., **142**(2009), 205-228.

[15] Yu. Nesterov, *Subgradient methods for huge-scale optimization problems*, Math. Program., Ser. A, **146**(2013), 275-297.

[16] B. Peng, H.K. Xu, *A cyclic coordinate-update fixed point algorithm*, Carpathian J. Math., **35**(2019), No. 3, 365-370.

[17] B.T. Polyak, *Introduction to Optimization*, Optimization Software, New York, 1987.

[18] S.M. Robinson, *Linear convergence of epsilon-subgradient descent methods for a class of convex functions*, Math. Prog. Ser. A, **86**(1999), 41-50.
[19] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, **14**(1976), 877-898.
[20] N.Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer Series in Computational Mathematics, Springer, 1985.
[21] M.V. Solodov, B.F. Svaiter, *Forcing strong convergence of proximal point iterations in a Hilbert space*, Mathematical Programming, Ser. A, **87**(2000), 189-202.
[22] Y. Wang, F. Wang, H.K. Xu, *Error sensitivity for strongly convergent modifications of the proximal point algorithm*, J. Optim. Theory Appl., **168**(2016), 901-916.
[23] H.K. Xu, *Iterative algorithms for nonlinear operators*, J. Lond. Math. Soc., **66**(2002), 240-256.
[24] H.K. Xu, *Viscosity approximation methods for nonexpansive mappings*, J. Math. Anal. Appl., **298**(2004), 279-291.
[25] H.K. Xu, *Averaged mappings and the gradient-projection algorithm*, J. Optim. Theory Appl., **150**(2011), 360-378.
[26] J. Zowe, *Nondifferentiable Optimization – A Motivation and a Short Introduction into the Subgradient and Bundle Concept*, in: Computational Mathematical Programming (ed. K. Schittkowski), NATO ASI Series, Computer and System Sciences, Vol. 15, Springer-Verlag, Heidelberg, 1985.

Received: March 24, 2023; Accepted: April 18, 2024.