Extension of Subgradient Techniques for Nonsmooth Optimization in Banach Spaces

Ya. I. Alber
Department of Mathematics
The Technion-Israel Institute of Technology
32000 Haifa, Israel *
A. N. Iusem
Instituto de Matemática Pura e Aplicada
Estrada Dona Castorina 110,
Rio de Janeiro, RJ, CEP 22460-320, Brazil †

Abstract.

In this paper we continue the study of the subgradient method for nonsmooth convex constrained minimization problems in a uniformly convex and uniformly smooth Banach space. We consider the case when the stepsizes satisfy $\sum_{k=1}^{\infty} \alpha_k = \infty$, $\lim_{k\to\infty} \alpha_k = 0$.

Key words: *subgradient method, nonsmooth convex functional, subgradient, duality mapping, generalized projection, J-Lipschitz continuous functional, weak convergence, estimates of convergence rate.*

1991 AMS Subject Classification: 90C25, 90C30, 49J40.

1 Introduction

We study the following optimization problem:

$$\min f(x), \tag{1.1}$$

$$s.t.\ \ x \in \Omega, \tag{1.2}$$

*This work was supported in part by a joint grant from the Center for Absorption in Science of the Ministry of Immigrant Absorption and the Committee for Planning and Budgeting of the Council for Higher Education under the framework of the KAMEA Program.

†This research was supported by CNPq Grant no. 301280/86.

where B is a Banach space, $f : B \to \mathbf{R}$ is a convex function (possibly nonsmooth), and Ω is a nonempty, closed and convex subset of B.

In order to introduce our algorithm, we need some notation and basic properties of Banach spaces, which we discuss next.

Let $\langle \cdot, \cdot \rangle : B^* \times B \to \mathbf{R}$ be the standard duality product, and $\|\cdot\|,\ \|\cdot\|_*$ the norms in B and B^*, respectively. We denote by $\partial f(x)$ the subdifferential of $f(x)$ at $x \in B$, that is

$$\partial f(x) = \{u \in B^* : f(y) - f(x) \geq \langle u, y - x \rangle,\ \text{for all } y \in B\}.$$

The elements of $\partial f(x)$ are called subgradients of f at x.

Let $\delta_B(\epsilon)$ and $\rho_B(\tau)$ be the modulus of convexity and the modulus of smoothness of the Banach space B, defined as

$$\delta_B(\epsilon) := \inf\{1 - \|(x+y)/2\| \mid \|x\| = 1,\ \|y\| = 1,\ \|x - y\| = \epsilon\} \tag{1.3}$$

and

$$\rho_B(\tau) := \sup\{\|(x+y)/2\| + \|(x-y)/2\| - 1 \mid \|x\| = 1,\ \|y\| = \tau\}. \tag{1.4}$$

B is uniformly convex if and only if $\delta_B(\epsilon) > 0$ for all $\epsilon > 0$, and uniformly smooth if and only if $\lim_{\tau \to 0} \tau^{-1} \rho_B(\tau) = 0$ (see, for instance, [1, 14, 17]).

We assume that B is a real uniformly convex and uniformly smooth (reflexive) Banach space, and we observe that in this case B^* is also a reflexive, uniformly convex and uniformly smooth Banach space (see [11]). We also assume throughout the following condition on f:

(P1) $\partial f : B \to \mathcal{P}(B^*)$ is bounded on bounded sets.

We mention that (P1) implies that f itself is bounded on bounded sets, as it follows easily from the definition of ∂f. Under these assumptions, it has been proved in Proposition 2.2.6 of [9] that f is locally Lipschitz, and henceforth continuous.

Let $J : B \to B^*$ be the normalized duality mapping determined by the following relations:

$$\langle Jx, x \rangle = \|Jx\|_* \|x\| = \|x\|^2. \tag{1.5}$$

Similarly, J^* will denote the normalized duality mapping from B^* to B. In a Hilbert space H, J is the identity operator, i.e. $J = I_H$. Under our assumptions on the Banach space B (and, more generally, in any reflexive strictly convex Banach space with a strictly convex dual), it holds that $JJ^* = I_{B^*}$ and $J^*J = I_B$.

The following function $V : B^* \times B \to \mathbf{R}$ is central in the subsequent analysis:

$$V(\varphi, x) := \frac{1}{2}(\|\varphi\|_*^2 - 2\langle \varphi, x \rangle + \|x\|^2), \quad x \in B,\ \varphi \in B^*. \tag{1.6}$$

In a Hilbert space H, $V(\varphi, x) = 2^{-1}\|\varphi - x\|^2$, where $x \in H,\ \varphi \in H$. Other properties of this function will be presented in Section 3.

Given a closed and convex set $\Omega \subseteq B$, we use V as given by (1.6) in order to define the generalized projection operator $\pi_\Omega : B^* \to \Omega$:

$$\pi_\Omega(\varphi) := \arg\min_{x \in \Omega} V(\varphi, x). \tag{1.7}$$

In a Hilbert space H, the generalized projection reduces to the standard metric projection, that is $\pi_\Omega = P_\Omega$. In the case when $\Omega = B$, it holds that $\pi_\Omega = J^*$.

In this paper we study the following algorithm for problem (1.1)–(1.2):

$$x^1 \in B,$$
$$x^{k+1} = \pi_\Omega \left[Jx^k - \alpha_k \frac{u^k}{\|u^k\|_*} \right], \tag{1.8}$$

where u^k is an arbitrary element of $\partial f(x^k)$ and the stepsizes $\{\alpha_k\} \subset \mathbf{R}_{++}$ are chosen according to

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \quad \alpha_k > 0, \quad \lim_{k\to\infty} \alpha_k = 0, \tag{1.9}$$

J and π_Ω are the duality mapping and the generalized projection operator as defined by (1.5) and (1.7), respectively.

We mention next some previously established results on related algorithms. In general, strong convergence of the sequence $\{x^k\}$, generated by (1.8), (1.9), cannot be proved (even for the smooth unconstrained case) unless B is finite dimensional. In a Hilbert space, convergence of the sequence of functional values $\{f(x^k)\}$, in the sense that

$$\liminf_{k\to\infty} f(x^k) = \inf_{x\in\Omega} f(x),$$

was obtained in [18] for the algorithm

$$x^{k+1} = P_\Omega \left[x^k - \alpha_k \frac{u^k}{\|u^k\|_*} \right], \tag{1.10}$$

where u^k is an arbitrary subgradient of f at x^k and P_Ω is the metric projection operator. For the smooth and unconstrained case in Hilbert spaces, weak convergence of the sequence $\{x^k\}$ generated by the gradient method

$$x^{k+1} = x^k - \alpha_k f'(x^k), \tag{1.11}$$

with stepsizes satisfying $0 \le \bar{\alpha} < \alpha_k \le \bar{\bar{\alpha}}$, to a minimizer of f, was established in [2] for the case when $f \in C^{1,\mu}$ (i.e., the gradient of f satisfies the Lipschitz-Hölder condition with $0 < \mu \le 1$, and $\bar{\bar{\alpha}}$ is related to μ); see also [12, 15]. As a consequence, strong convergence holds in finite dimensional spaces. Analogous results are presented in [3] for the constrained problem (1.1)–(1.2) and the projection method

$$x^{k+1} = P_\Omega(x^k - \alpha_k f'(x^k)),$$

similar to (1.11). The case of $f \in C^{1,0}$ (the gradient of f is continuous) in finite dimensional spaces was considered in [4] where strong convergence is proved. A similar convergence result in the nonsmooth case was noted in the literature long ago, but a complete proof appeared for the first time in [10]. The nonsmooth case

in infinite dimensional spaces was investigated recently in [6, 7, 5]. In [6], it was shown that the method (1.10), with $\max\{1, ||u^k||_*\}$ substituting for $||u^k||_*$, is weakly convergent and stable if $\sum_{k=1}^{\infty} \alpha_k = \infty$, $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$. The same results were obtained in [7] for Banach spaces and the method (1.8) under the conditions $\delta_B(\epsilon) > C\epsilon^2$ for some $C > 0$, and $\sum_{k=1}^{\infty} \rho_{B^*}(\alpha_k) < \infty$. Finally, in [5] weak convergence and stability of the average approximations to a minimizer were studied in Banach spaces, and for the more general problem of monotone variational inequalities, rather than convex optimization.

Next, we summarize the main results in this paper. In Theorem 3.1, we assume that either $\delta_B(\epsilon) > C\epsilon^2$ for some $C > 0$ or the set Ω is bounded, but we do not assume existence of solutions of the problem (1.1), (1.2). Then we prove that $\liminf_{k\to\infty} f(x^k) = \inf_{x\in\Omega} f(x)$ for the method (1.8)-(1.9). In Theorems 3.4 and 3.5 we obtain estimates of the convergence rate of the functional values on some subsequence $\{x^{\ell_k}\}$ of $\{x^k\}$, under the additional assumption that the set S of solutions of problem (1.1), (1.2) is nonempty. Unfortunately, we were not able to obtain convergence rate results for the whole sequence, which would be obviously more meaningful, and we were forced to leave this item as an open problem, deserving further research.

Stronger results (including weak convergence of the whole sequence $\{x^k\}$ to a solution x^*) are proved in Theorem 3.9 under the following additional joint assumption on f and Ω:
(P2) The intersection of Ω with any level set of f is either empty or bounded.
Property (P2) clearly holds either when f is coercive or when Ω is bounded.

Convergence of the whole sequence of functional values $\{f(x^k)\}$ to $\inf_{x\in\Omega} f(x)$ and weak convergence of the whole sequence $\{x^k\}$ to some $\bar{x} \in S$ are established in Theorem 3.15 for the case when f is J-Lipschitz continuous (see Definition 3.11) and $\Omega = B$.

2 Some recurrence inequalities

We present in this section some preliminary results, in the form of recurrence numerical inequalities, which are needed in our convergence analysis.

Lemma 2.1 *Let $\{\alpha_k\}$ and $\{\beta_k\}$ be sequences of nonnegative real numbers. Assume that*

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \quad \sum_{k=1}^{\infty} \alpha_k \beta_k < \infty.$$

Then:
(i) there exists an infinite subsequence $\{\beta_{\ell_k}\} \subset \{\beta_k\}$ such that $\lim_{k\to 0} \beta_{\ell_k} = 0$,
(ii) if $\lim_{k\to\infty} \alpha_k = 0$, and there exists $\theta > 0$ such that

$$|\beta_{k+1} - \beta_k| \leq \theta\alpha_k, \tag{2.1}$$

for all $k \geq 1$, then $\lim_{k\to\infty} \beta_k = 0$.

Proof: See [6], Proposition 2.

Lemma 2.2 *Let* $\{\mu_k\}$, $\{\alpha_k\}$, $\{\beta_k\}$, *and* $\{\gamma_k\}$ *be sequences of nonnegative real numbers satisfying the recurrence inequality*

$$\mu_{k+1} \leq \mu_k - \alpha_k \beta_k + \gamma_k. \tag{2.2}$$

Assume that

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \quad \sum_{k=1}^{\infty} \gamma_k < \infty.$$

Then:
(i) there exists an infinite subsequence $\{\beta_{\ell_k}\} \subset \{\beta_k\}$ *such that*

$$\beta_{\ell_k} \leq \frac{1}{\sum_{j=1}^{\ell_k} \alpha_j}, \tag{2.3}$$

and, consequently, $\lim_{k\to 0} \beta_{\ell_k} = 0$,
(ii) if $\lim_{k\to\infty} \alpha_k = 0$, *and there exists* $\theta > 0$ *such that (2.1) holds, then* $\lim_{k\to\infty} \beta_k = 0$,
(iii) there exists $\lim_{k\to\infty} \mu_k = \mu \geq 0$.

Proof. i) If the result does not hold, then there exists $M > 0$ such that for all $k \geq M$ it holds that

$$\beta_k > \frac{1}{\sum_{j=1}^{k} \alpha_j}.$$

Therefore, by (2.2)

$$\mu_{k+1} \leq \mu_k - \frac{\alpha_k}{\sum_{j=1}^{k} \alpha_j} + \gamma_k, \ \ \forall k \geq M.$$

As a consequence,

$$\mu_k \leq \mu_M - \sum_{s=M}^{k-1} \mathcal{A}_s + \sum_{s=M}^{k-1} \gamma_s, \ \forall k > M,$$

where

$$\mathcal{A}_s = \frac{\alpha_s}{\sum_{j=1}^{s} \alpha_j},$$

which contradicts nonnegativity of $\{\mu_k\}$, because $\sum_{k=1}^{\infty} \gamma_k < \infty$, while $\sum_{s=1}^{\infty} \mathcal{A}_s = \infty$, as a consequence of the fact that $\sum_{k=1}^{\infty} \alpha_k = \infty$ and Abel-Dini's criterion for divergent series (see e.g. [16]).
ii) The result follows from Lemma 2.1. Indeed, applying successively the inequality (2.2), we obtain

$$\sum_{k=1}^{n} \alpha_k \beta_k \leq \mu_1 - \mu_n + \sum_{k=1}^{n} \gamma_k \leq \mu_1 + \sum_{k=1}^{n} \gamma_k$$

for all $n \geq 1$. Thus,

$$\sum_{k=1}^{\infty} \alpha_k \beta_k < \infty.$$

This means that all the assumptions of Lemma 2.1 hold, and therefore $\lim_{k\to\infty} \beta_k = 0$.
iii) See [6], Proposition 1. ■

Lemma 2.3 *Let $\{\mu_k\}$, $\{\alpha_k\}$, $\{\beta_k\}$, and $\{\gamma_k\}$ be sequences of nonnegative real numbers satisfying (2.2). Assume that*

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \quad \lim_{k\to\infty} \alpha_k = 0, \quad \sum_{k=1}^{\infty} \gamma_k = \infty, \quad \lim_{k\to\infty} \frac{\gamma_k}{\alpha_k} = 0.$$

Then for any $c_0 > 1$, there exists an infinite subsequence $\{\beta_{\ell_k}\} \subset \{\beta_k\}$ such that

$$\beta_{\ell_k} \leq \frac{c_0 \gamma_{\ell_k}}{\alpha_{\ell_k}}, \tag{2.4}$$

and, consequently, $\lim_{k\to 0} \beta_{\ell_k} = 0$.

Proof. We use again the argument of Lemma 2.2. If the result does not hold, then there exists $M > 0$ such that

$$\beta_k > \frac{c_0 \gamma_k}{\alpha_k}$$

for all $k \geq M$. As a consequence,

$$\mu_{k+1} \leq \mu_k - (c_0 - 1)\gamma_k,$$

and therefore,

$$\mu_k \leq \mu_M - (c_0 - 1) \sum_{s=M}^{k-1} \gamma_s$$

for all $k > M$, which contradicts nonnegativity of $\{\mu_k\}$, because $\sum_{k=1}^{\infty} \gamma_k = \infty$. ■

Lemma 2.4 *Let $\{\mu_k\}$, $\{\alpha_k\}$, $\{\beta_k\}$, and $\{\gamma_k\}$ be sequences of nonnegative real numbers satisfying the inequality (2.2). Assume that $\sum_{k=1}^{\infty} \alpha_k = \infty$ and that there exist $\bar{\alpha} > 0$, $C_2 > 0$ such that $\alpha_k \leq \bar{\alpha}$, $\gamma_k/\alpha_k \leq C_2$. Then:*
(i) for any $C_1 > 0$ there exists a subsequence $\{\ell_k\} \in \mathbf{N}$ such that

$$\beta_{\ell_k} \leq C_1 + \frac{\gamma_{\ell_k}}{\alpha_{\ell_k}}, \tag{2.5}$$

(ii) if $\{\mu_{\ell_k}\}$ is bounded, where $\{\ell_k\}$ is the set of subindices of the subsequence in (i), then the whole sequence $\{\mu_k\}$ is also bounded.

Proof. i) Define

$$I_1 = \left\{ k \in \mathbf{N} : \ \beta_k \le C_1 + \frac{\gamma_k}{\alpha_k} \right\}, \quad I_2 = \mathbf{N} \setminus I_1.$$

If I_1 is finite then there exists M such that

$$\mu_{k+1} \le \mu_k - C_1 \alpha_k$$

for all $k > M$, which cannot be true because $\mu_k \ge 0$ and $\sum_{k=1}^{\infty} \alpha_k = \infty$. Thus, I_1 is an infinite set and (2.5) holds for all $\ell_k \in I_1$. Furthermore, for all $\ell_k \in I_1$,

$$\beta_{\ell_k} \le C_1 + \frac{\gamma_{\ell_k}}{\alpha_{\ell_k}} \le C_1 + C_2,$$

i.e., $\{\beta_{\ell_k}\}$ is bounded. We need this last fact below.
ii) Consider the subsequence $\{\beta_{\ell_k}\}$ with $\{\ell_k\} = I_1$. If I_1 coincides with I, i.e., $\ell_{k+1} = \ell_k + 1$, then $\beta_k \le C_1 + C_2$ for all $k \in \mathbf{N}$, and (ii) holds. Otherwise, we proceed as follows. Consider an interval $[\ell_k, \ell_{k+1}]$. Then $\ell_{k+1} > \ell_k$, $\ell_k \in I_1$, $\ell_{k+1} \in I_1$, and

$$\{\ell_k + 1, \ell_k + 2, ..., \ell_{k+1} - 1\} \in I_2.$$

$\{\mu_{\ell_k}\}$ is bounded by assumption, say, $\mu_{\ell_k} \le C_3$. Then,

$$\mu_{\ell_k+1} \le \mu_{\ell_k} - \alpha_{\ell_k}\beta_{\ell_k} + \gamma_{\ell_k} \le \mu_{\ell_k} + \gamma_{\ell_k} \le C_3 + C_4,$$

where C_4 is an upper bound of $\{\gamma_k\}$, which is finite because $\alpha_k \le \bar{\alpha}$ and $\gamma_k \le C_2\alpha_k \le \bar{\alpha}C_2$. Therefore,

$$\mu_{\ell_k+2} \le \mu_{\ell_k+1} - \alpha_{\ell_k+1}\beta_{\ell_k+1} + \gamma_{\ell_k+1} \le \mu_{\ell_k+1} - C_1\alpha_{\ell_k+1} \le \mu_{\ell_k+1} \le C_3 + C_4,$$

because $\ell_k + 1 \in I_2$, i.e., $\beta_{\ell_k+1} > C_1 + \gamma_{\ell_k+1}/\alpha_{\ell_k+1}$. The same estimate is valid for $\ell_k + 2, ..., \ell_{k+1} - 1$. Thus, $\mu_k \le C_3 + C_4$ for all $k \in \mathbf{N}$. ■

Lemma 2.5 *Suppose that* $\{\mu_k\}$, $\{\alpha_k\}$, $\{\beta_k\}$, *and* $\{\gamma_k\}$ *are sequences of nonnegative real numbers satisfying (2.2). Assume that there exists* $\bar{\alpha} > 0$ *such that* $\alpha_k \le \bar{\alpha}$ *for all* k *and that*

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \quad \lim_{k\to\infty} \frac{\gamma_k}{\alpha_k} = 0.$$

Then:
(i) there exists a sequence of indices $\{\ell_k\} \subset \mathbf{N}$ *such that*

$$\beta_{\ell_k} \le \frac{1}{\sum_{j=1}^{\ell_k} \alpha_j} + \frac{\gamma_{\ell_k}}{\alpha_{\ell_k}} \tag{2.6}$$

and, consequently, $\lim_{k\to\infty} \beta_{\ell_k} = 0$,
(ii) if $\lim_{k\to\infty} \mu_{\ell_k} = 0$ *then* $\lim_{k\to\infty} \mu_k = 0$,
(iii) if $\{\beta_k\}$ *converges then* $\lim_{k\to\infty} \beta_k = 0$.

Proof. (i) is proved with the same argument as in the proof of Lemma 2.2, assuming that there exists M such that

$$\beta_k > \frac{1}{\sum_{j=1}^{k} \alpha_j} + \frac{\gamma_k}{\alpha_k} \tag{2.7}$$

for all $k \geq M$, and obtaining a contradiction. (ii) is proved in a similar way, with the proofline of Lemma 2.4. (iii) is obvious. ■

Remark 2.6 *Suppose that β_k of Lemma 2.5 is of the form $\beta_k = \psi(\mu_k)$, where $\psi : \mathbf{R}_+ \to \mathbf{R}$ is a continuous function such that $\psi(t) > 0$ for all $t > 0$ and $\psi(0) = 0$. Then the inequality (2.2) takes the form*

$$\mu_{k+1} \leq \mu_k - \alpha_k \psi(\mu_k) + \gamma_k,$$

with

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \quad \alpha_k \leq \bar{\alpha}, \quad \lim_{k\to\infty} \frac{\gamma_k}{\alpha_k} = 0.$$

In this case, Lemma 2.5 asserts that $\lim_{k\to\infty} \mu_k = 0$, and it coincides with Lemma 1 from [8]. Indeed, we get from (2.6)

$$\psi(\mu_{\ell_k}) \leq \frac{1}{\sum_{k=1}^{\ell_k} \alpha_j} + \frac{\gamma_{\ell_k}}{\alpha_{\ell_k}},$$

which means that $\lim_{k\to\infty} \psi(\mu_{\ell_k}) = 0$, implying $\lim_{k\to\infty} \mu_{\ell_k} = 0$, and therefore, by Lemma 2.5 (ii), $\lim_{k\to\infty} \mu_k = 0$.

3 Subgradient methods: convergence and estimates of the convergence rate

We present first some known properties of V and π_Ω, defined by (1.6) and (1.7) respectively, which are needed in our convergence analysis (see [1]).
(P3) $V(\phi, x)$ is differentiable for each fixed $x \in B$ and $\nabla_\varphi V(\varphi, x) = J^*\varphi - x$.
(P4) $V(\cdot, x)$ is convex for each fixed $x \in B$. Therefore, for all $y \in B$ and all $z \in B$

$$V(Jy, x) \leq V(Jz, x) + \langle Jy - Jz, y - x\rangle.$$

(P5) For all $x \in B$, $\varphi \in B^*$

$$\frac{1}{2}(\|\varphi\|_* - \|x\|)^2 \leq V(\varphi, x) \leq \frac{1}{2}(\|\varphi\|_* + \|x\|)^2.$$

In particular, V is nonnegative and finite, and $V(Jy, x) = 0$, with $x, y \in B$ if and only if $x = y$;
(P6) The operator π_Ω satisfies

$$V(J\pi_\Omega(\phi), \xi) \leq V(\phi, \xi) - V(\phi, \pi_\Omega(\phi)), \quad \forall \xi \in \Omega.$$

Consequently, we say that π_Ω is *the conditionally nonexpansive* operator relative to the functional V in the Banach space B, in the sense that $V(J\pi_\Omega(\phi), \xi) \leq V(\phi, \xi)$ for all $\xi \in \Omega$.

Theorem 3.1 *Let B be a uniformly convex and uniformly smooth Banach space and Ω a convex and closed subset of B. Consider the sequence $\{x^k\}$ generated by (1.8), (1.9). If:*
1) either the modulus of convexity $\delta_B(\epsilon) \geq C\epsilon^2$ for some $C > 0$ or
2) Ω is bounded,
then it holds that

$$\liminf_{k\to\infty} f(x^k) = \inf_{x\in\Omega} f(x).$$

Proof. We discuss first the case of assumption 1. It is well known (see [14]) that if $\delta_B(\epsilon) \geq C\epsilon^p$ for some $C > 0$ with $1 < p < \infty$, then there exists a constant $D > 0$ such that $\rho_{B^*}(\tau) \leq D\tau^q$, and $p^{-1}+q^{-1} = 1$. In our case, $p = 2$, and therefore $\rho_{B^*}(\tau) \leq D\tau^2$.

Take any $\bar{x} \in \Omega$, and let $\{x^k\}$, $\{u^k\}$ be the sequences defined by the algorithm, $V_k = V(Jx^k, \bar{x})$, with V as in (1.6), and

$$\tilde{\beta}_k = \frac{\langle u^k, x^k - \bar{x}\rangle}{\|u^k\|_*}. \tag{3.1}$$

First of all, we shall show that

$$V_{k+1} \leq V_k - \alpha_k\tilde{\beta}_k + 32D{\alpha_k}^2, \tag{3.2}$$

for all $k \geq 0$. Denote

$$\phi^k = Jx^k - \alpha_k\frac{u^k}{\|u^k\|_*}. \tag{3.3}$$

With this definition, we have

$$x^{k+1} = \pi_\Omega(\phi^k), \tag{3.4}$$

and

$$\|\phi^k - Jx^k\|_* = \alpha_k. \tag{3.5}$$

Properties (P3) and (P4) above imply that

$$V(\phi^k, \bar{x}) \leq V(Jx^k, \bar{x}) + \langle\phi^k - Jx^k, J^*\phi^k - \bar{x}\rangle. \tag{3.6}$$

By (P6) and (3.4), we have

$$V(Jx^{k+1}, \bar{x}) \leq V(\phi^k, \bar{x}).$$

Hence,

$$V_{k+1} \leq V_k + \langle\phi^k - Jx^k, J^*\phi^k - \bar{x}\rangle. \tag{3.7}$$

We proceed to estimate the second term in the right hand side of (3.7). Define $R : \mathbf{R}\times\mathbf{R} \to \mathbf{R}$ as

$$R(s,t) = \sqrt{(s^2 + t^2)/2}. \tag{3.8}$$

The following estimate has been established in [1], p. 38:

$$\langle Jy - Jx, y - x\rangle \leq 2R^2(\|x\|, \|y\|)\rho_B\Big(4\|y - x\|/R(\|x\|, \|y\|)\Big)$$

for all x, $y \in B$. It follows easily that for all $\phi, \psi \in B^*$, the following inequality holds:

$$\langle \phi - \psi, J^*\phi - J^*\psi\rangle \leq 2R^2(\|\phi\|_*, \|\psi\|_*)\rho_{B^*}\Big(4\|\phi - \psi\|_*/R(\|\phi\|_*, \|\psi\|_*)\Big). \tag{3.9}$$

By the definition (3.1) of $\tilde{\beta}_k$ and the definition (3.3) of ϕ^k, we have

$$\begin{aligned}
&\langle \phi^k - Jx^k, J^*\phi^k - \bar{x}\rangle \\
&= \langle \phi^k - Jx^k, x^k - \bar{x}\rangle + \langle \phi^k - Jx^k, J^*\phi^k - J^*Jx^k\rangle \\
&= -\alpha_k\tilde{\beta}_k + \langle \phi^k - Jx^k, J^*\phi^k - J^*Jx^k\rangle.
\end{aligned} \tag{3.10}$$

¿From (3.9) we get

$$\langle \phi^k - Jx^k, J^*\phi^k - J^*Jx^k\rangle \leq 2T_k^2\rho_{B^*}(4\|\phi^k - Jx^k\|_*/T_k),$$

with $T_k = R(\|x^k\|, \|\phi^k\|)$, which implies, in view of (3.1), (3.3) and (3.5),

$$\langle \phi^k - Jx^k, J^*\phi^k - \bar{x}\rangle \leq -\alpha_k\tilde{\beta}_k + 2T_k^2\rho_{B^*}(4\alpha_k/T_k). \tag{3.11}$$

Thus, it follows from (3.11) and the property of $\rho_{B^*}(\tau)$ that

$$\langle \phi^k - Jx^k, J^*\phi^k - \bar{x}\rangle \leq -\alpha_k\tilde{\beta}_k + 32D\alpha_k^2. \tag{3.12}$$

Combining (3.7) and (3.12) we obtain (3.2).

Suppose now that the assertion of the theorem is not true. Then there exists $M > 0$ and $\bar{f} > f^*$ such that

$$f(x^k) \geq \bar{f}, \ \ \forall k \geq M.$$

Furthermore, there exists a point $\bar{x} \in \Omega$ such that $f(\bar{x}) < \bar{f}$ and it follows from the continuity of f (which is itself a consequence of (P1), as discussed above) that there exists $d > 0$ such that $f(x) < \bar{f}$ for all x satisfying

$$\|x - \bar{x}\| \leq d.$$

Let

$$y^k = \bar{x} + d\frac{J^*u^k}{\|u^k\|_*}.$$

Clearly, $\|y^k - \bar{x}\| = d$, and so

$$f(y^k) < \bar{f} \leq f(x^k), \ \ \forall k \geq M.$$

Thus, since $\langle \phi, J^*\phi \rangle = \|\phi\|_*^2$ for all $\phi \in B^*$, we have

$$0 > f(y^k) - f(x^k) \geq \langle u^k, y^k - x^k \rangle =$$
$$\langle u^k, y^k - \bar{x} \rangle + \langle u^k, \bar{x} - x^k \rangle = d\|u^k\|_* + \langle u^k, \bar{x} - x^k \rangle.$$

This implies the inequality

$$\tilde{\beta}_k \geq d > 0 \tag{3.13}$$

for all $k > M$. Now we apply (3.2), obtaining:

$$V_{k+1} \leq V_k - \alpha_k d + 32D{\alpha_k}^2. \tag{3.14}$$

Comparing (2.2) with (3.14) with $\mu_k = V_k$, we see that $\gamma_k = 32D{\alpha_k}^2$, and therefore

$$\lim_{k\to\infty} \frac{\gamma_k}{\alpha_k} = 0.$$

Then the fact that $d > 0$ contradicts Lemma 2.5.

Now we consider the case of assumption 2, for which we will use the following estimate, obtained in [1], p. 38: if $\|x\| \leq \eta$ and $\|y\| \leq \eta$ for some $\eta > 0$, then

$$\langle Jx - Jy, x - y \rangle \leq 2L\eta^2 \rho_{B^*}(4\|Jx - Jy\|_*/\eta), \tag{3.15}$$

for all $x, y \in B$, where L is Figiel's constant, which satisfies $1 < L < 3.18$ (see [14, 1]). We mention that the upper bound on L was improved from 3.18 to 1.7 in [19]. Note that, in view of (1.8), $\{x^k\} \subset \Omega$, so that $\{x^k\}$ is bounded under assumption 2. It follows from (3.3) and the boundedness of $\{\alpha_k\}$ that $\{\phi^k\}$ is bounded, and henceforth there exists some $\bar{\eta} > 0$ such that $\|\phi^k\|_* \leq \bar{\eta}$, $\|Jx^k\|_* \leq \bar{\eta}$ for all k. Now, using (3.10) and (3.15) with $\bar{\eta}$ instead of η, as well as (3.5), we have

$$\langle \phi^k - Jx^k, J^*\phi^k - \bar{x} \rangle \leq -\alpha_k\tilde{\beta}_k + 2L\bar{\eta}^2\rho_{B^*}(4\alpha_k/\bar{\eta})$$
$$= -\alpha_k\tilde{\beta}_k + C_1\rho_{B^*}(C_2\alpha_k), \tag{3.16}$$

where

$$C_1 = 2L\bar{\eta}^2, \;\; C_2 = 4\bar{\eta}^{-1}. \tag{3.17}$$

Thus, by (3.7),

$$V_{k+1} \leq V_k - \alpha_k\tilde{\beta}_k + C_1\rho_{B^*}(C_2\alpha_k). \tag{3.18}$$

Assume again that the result does not hold. Then the inequality (3.14) becomes

$$V_{k+1} \leq V_k - \alpha_k d + C_1\rho_{B^*}(C_2\alpha_k), \tag{3.19}$$

On the other hand, we have,

$$\lim_{k\to\infty} \frac{\rho_{B^*}(C_2\alpha_k)}{\alpha_k} = 0, \tag{3.20}$$

because B^* is uniformly smooth. (3.19) and (3.20) contradict Lemma 2.5, and therefore the theorem is proved. ■

Remark 3.2 *More generally, Theorem 3.1 holds when the sequence $\{x^k\}$ is known beforehand to be bounded (cf. [7], Theorem 4.2).*

Corollary 3.3 *Under the hypotheses of Theorem 3.1, if $\{f(x^k)\}$ has a limit as $k \to \infty$ then all weak accumulation points of $\{x^k\}$, if any, are solutions of problem (1.1)-(1.2).*

Proof. Let $\bar{x}$ be a weak accumulation point of $\{x^k\}$. Since Ω is closed and convex, it is weakly closed, and thus $\bar{x} \in \Omega$, because $x^k \in \Omega$ for all k. Let $\{x^{\ell_k}\}$ be a subsequence of $\{x^k\}$ such that $\{x^{\ell_k}\}$ is weakly convergent to $\bar{x}$. Since f is convex, it is weakly lower semicontinuous. Let $f^* = \inf_{x\in\Omega} f(x)$. By Theorem 3.1,

$$f^* \le f(\bar{x}) \le \operatorname{liminf}_{k\to\infty} f(x^{\ell_k}) = \lim_{k\to\infty} f(x^k) = \operatorname{liminf}_{k\to\infty} f(x^k) = f^*.$$

Thus, $f(\bar{x}) = f^*$, i.e., $\bar{x}$ solves the problem. ■

The next two theorems give estimates for the convergence rate of some subsequence $\{x^{\ell_k}\}$ of $\{x^k\}$.

Theorem 3.4 *Let B be a uniformly convex and uniformly smooth Banach space, B^* its dual space. Assume that there exists a constant $C > 0$ such that modulus of convexity $\delta_B(\epsilon) \ge C\epsilon^2$. Suppose that the solution set S of problem (1.1), (1.2) is nonempty and that the subdifferential ∂f of f is bounded on Ω. Take any $c_0 > 1$, and consider the sequence $\{x^k\}$, generated by (1.8), (1.9). It holds that $\operatorname{liminf}_{k\to\infty} f(x^k) = f^* := \inf_{x\in\Omega} f(x)$ and there exist a subsequence $\{x^{\ell_k}\} \subset \{x^k\}$ and a constant $C_3 > 0$ such that either*

$$\beta_{\ell_k} = f(x^{\ell_k}) - f^* \le 32C_3c_0D\alpha_{\ell_k}, \quad \text{if} \quad \sum_{k=1}^{\infty} \alpha_k^2 = \infty, \tag{3.21}$$

or

$$\beta_{\ell_k} = f(x^{\ell_k}) - f^* \le C_3\Big(\sum_{s=1}^{\ell_k} \alpha_s\Big)^{-1}, \quad \text{if} \quad \sum_{k=1}^{\infty} \alpha_k^2 < \infty. \tag{3.22}$$

Proof. Since $\delta_B(\epsilon) \ge C\epsilon^2$, there exists a constant $D > 0$ such that modulus of smoothness $\rho_{B^*}(\tau) \le D\tau^2$. Since the subdifferential ∂f is bounded on Ω, there exist a constant $C_3 > 0$ such that $\|u\|_* \le C_3$ for all $u \in \partial f(x)$ and all $x \in \Omega$.

The statement on $\operatorname{liminf}_{k\to\infty} f(x^k)$ follows from Theorem 3.1. Take $x^* \in S$, assumed to be nonempty. Since $x^* \in \Omega$, we have, by (3.1)-(3.2) with $\bar{x} = x^*$ and $C_1 = 32D$,

$$V_{k+1} \le V_k - \alpha_k \frac{\langle u^k, x^k - x^*\rangle}{\|u^k\|_*} + 32D{\alpha_k}^2. \tag{3.23}$$

Applying to (3.23) the subgradient inequality for convex functions

$$\langle u^k, x^k - x^*\rangle \ge f(x^k) - f^*,$$

we get

$$V_{k+1} \leq V_k - C_3^{-1}\alpha_k(f(x^k) - f^*) + 32D{\alpha_k}^2. \tag{3.24}$$

Defining

$$\mu_k = V_k, \;\; \beta_k = f(x^k) - f^*,$$

we can rewrite (3.24) as

$$\mu_{k+1} \leq \mu_k - C_3^{-1}\alpha_k\beta_k + \gamma_k, \tag{3.25}$$

with $\gamma_k = 32D\alpha_k^2$, and

$$\lim_{k\to\infty} \frac{\gamma_k}{\alpha_k} = 0.$$

Then (3.21) and (3.22) follow from Lemma 2.3 and Lemma 2.2, respectively. ■

Similarly, we establish now the following result.

Theorem 3.5 *Let B be a uniformly convex and uniformly smooth Banach space and Ω be a convex, closed and bounded subset of B. Suppose that the solution set S of problem (1.1)-(1.2) is nonempty. Take any $c_0 > 1$. Consider the sequence $\{x^k\}$, generated by (1.8), (1.9). It holds that $\liminf_{k\to\infty} f(x^k) = f^* := \inf_{x\in\Omega} f(x)$, and there exist $C_3 > 0$ and a subsequence $\{x^{\ell_k}\} \subset \{x^k\}$, such that either*

$$\beta_{\ell_k} = f(x^{\ell_k}) - f^* \leq \frac{C_1C_3c_0\rho_{B^*}(C_2\alpha_{\ell_k})}{\alpha_{\ell_k}}, \;\; \text{if} \;\; \sum_{k=1}^{\infty} \rho_{B^*}(\alpha_k) = \infty, \tag{3.26}$$

where C_1 and C_2 are defined by (3.17), or

$$\beta_{\ell_k} = f(x^{\ell_k}) - f^* \leq \Big(\sum_{s=1}^{\ell_k} \alpha_s\Big)^{-1}, \;\; \text{if} \;\; \sum_{k=1}^{\infty} \rho_{B^*}(\alpha_k) < \infty. \tag{3.27}$$

Proof. The statement on $\liminf_{k\to\infty} f(x^k)$ follows from Theorem 3.1. If Ω is bounded then we can assume, without loss of generality, that $\|x^k\| \leq R$ for some $R > 0$ and all k, so that C_1 and C_2 are well defined. Furthermore, (P1) implies $\|u^k\|_* \leq C_3$ for some $C_3 > 0$ and all k. In this case, we obtain, for $\mu_k = V_k$ and $\beta_k = f(x^k) - f^*$, the inequality (cf. (3.25))

$$\mu_{k+1} \leq \mu_k - C_3^{-1}\alpha_k\beta_k + \gamma_k,$$

with $\gamma_k = C_1\rho_{B^*}(C_2\alpha_k)$. The proofs of (3.26) and (3.27) follow the pattern of the proof of Lemmas 2.3 and 2.2 with minor changes. In both cases, we assume that the result does not hold, in which case the inequalities in (3.26) and (3.27) are reversed for all k larger than a given M, and a contradiction arises, using the fact that $\lim_{k\to\infty} \gamma_k/\alpha_k = 0$, by uniform smoothness of B^*. ■

In the special case of Hilbert spaces and strongly convex objective function with Lipschitz continuous derivatives, it is possible to establish an upper bound ξ_k for $f(x^k) - f^*$ such that $\sum_{k=1}^{\infty} \xi_k < \infty$ (see e.g. [20], p.70). In all other cases, known upper bounds for $f(x^k) - f^*$ are not summable (see, for instance, [2]). Taking into account (3.26) and (3.27), the next remark shows that this is also the case under the assumptions considered in this paper if $\ell_k \equiv k$.

Remark 3.6 *If* $\sum_{k=1}^{\infty}\alpha_k=\infty$ *and* $\alpha_k\le C$ *then also* $\sum_{k=1}^{\infty}(\rho_{B^*}(\alpha_k)/\alpha_k)=\infty$ *and* $\sum_{k=1}^{\infty}(\sum_{s=1}^{k}\alpha_s)^{-1}=\infty$.

Indeed, it is known that for any Banach space B and any Hilbert space H it holds that

$$\rho_B(\tau)\ge\rho_H(\tau)$$

and

$$\rho_H(\tau)=\frac{\tau^2}{(1+\tau^2)^{1/2}+1},$$

(see e.g. [11]). Therefore,

$$\rho_H(\alpha_k)\ge C_1\alpha_k^2,\ \ C_1=\frac{1}{(1+C^2)^{1/2}+1}.$$

Then

$$\frac{\rho_{B^*}(\alpha_k)}{\alpha_k}\ge C_1\alpha_k,$$

establishing the first result. The second one follows from Abel-Dini's criterion for divergent series.

In order to prove the next theorem we need the two following lemmas.

Lemma 3.7 *Take* $f:B\to\mathbf{R}$ *convex and let* x^* *be a minimizer of* f. *Take sequences* $\{y^k\}\in B,\{v^k\}\in B^*$ *such that* $0\ne v^k\in\partial f(y^k)$ *for all* $k\ge1$. *Define the sequence* $\{\sigma_k\}$ *as*

$$\sigma_k=\frac{\langle v^k,y^k-x^*\rangle}{\|v^k\|_*}. \tag{3.28}$$

If $\lim_{k\to\infty}f(y^k)=\infty$ *then* $\lim_{k\to\infty}\sigma_k=\infty$.

Proof. Suppose that $\lim_{k\to\infty}f(y^k)=\infty$ but $\{\sigma_k\}$ has a bounded subsequence. Define

$$\hat y^k=x^*+\frac{\langle v^k,y^k-x^*\rangle\ J^*v^k}{\|v^k\|_*^2}. \tag{3.29}$$

It is easy to check that

$$\langle v^k,\hat y^k-y^k\rangle=0. \tag{3.30}$$

Indeed,

$$\langle v^k,\hat y^k-y^k\rangle=\langle v^k,x^*-y^k\rangle+\frac{\langle v^k,y^k-x^*\rangle\langle v^k,J^*v^k\rangle}{\|v^k\|_*^2}=0$$

because $\langle v^k,J^*v^k\rangle=\|v^k\|_*^2$. Further, $\langle v^k,y^k-x^*\rangle\ge f(y^k)-f(x^*)$ yields $\sigma_k\ge0$.

By definition of $\partial f(y^k)$, we have

$$f(\hat y^k)\ge f(y^k)+\langle v^k,\hat y^k-y^k\rangle.$$

Then, in view of (3.30), we conclude that, for all $k \geq 0$,

$$f(\hat{y}^k) \geq f(y^k). \tag{3.31}$$

By (3.29), since $\|J^* v^k\| = \|v^k\|_*$ and $\{\sigma_k\}$ is eventually nonnegative, we get

$$\|\hat{y}^k - x^*\| = \sigma_k. \tag{3.32}$$

If $\{\sigma_k\}$ has a bounded subsequence $\{\sigma_{j_k}\}$, then $\{\hat{y}^{j_k}\}$ is bounded by (3.32), and by (3.31)

$$\lim_{k\to\infty} f(\hat{y}^{j_k}) \geq \lim_{k\to\infty} f(y^k) = \infty,$$

in contradiction with (P1), which clearly implies that f is Lipschitz continuous on bounded sets. The lemma is proved. ■

Lemma 3.8 *Suppose that condition (P2) is satisfied. Take $\{y^k\} \subset \Omega$. If the sequence $\{\sigma_k\}$ defined by (3.28) is bounded then $\{y^k\}$ is also bounded.*

Proof. If the result does not hold, then there exists an unbounded subsequence $\{y^{\ell_k}\}$ of $\{y^k\}$. By (P2), we get $\lim_{k\to\infty} f(y^{\ell_k}) = \infty$, because otherwise f would have an unbounded and nonempty level set. By Lemma 3.7, $\{\sigma_{\ell_k}\}$ is unbounded, in contradiction with our assumption of boundedness of $\{\sigma_k\}$. ■

Theorem 3.9 *Let B be a uniformly convex and uniformly smooth Banach space with modulus of convexity $\delta_B(\epsilon) \geq C\epsilon^2$ for some $C > 0$. Consider the sequence $\{x^k\}$ generated by (1.8), (1.9). Suppose that the set S of minimizers of f is nonempty and that (P1) and (P2) are satisfied. Then, it holds that*
(i) the sequence $\{x^k\}$ is bounded,
(ii) there exists $C_2 > 0$ such that $\|x^{k+1} - x^k\| \leq C_2\alpha_k$, and consequently $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$,
(iii) for any $c_0 > 1$ there exists a subsequence $\{x^{\ell_k}\} \in \{x^k\}$ such that either (3.21) or (3.22) are satisfied,
(iv) for any $x^* \in S$,

$$\lim_{\ell_k\to\infty} \langle u^{\ell_k}, x^{\ell_k} - x^* \rangle \to 0, \tag{3.33}$$

where $\{x^{\ell_k}\}$ is the subsequence whose existence is ensured by (iii),
(v) there exists $C_4 > 0$ such that $|f(x^{k+1}) - f(x^k)| \leq C_4\alpha_k$, and consequently $\lim_{k\to\infty} |f(x^{k+1}) - f(x^k)| = 0$,
(vi) $\{x^k\}$ has weak accumulation points.
(vii) if $\{f(x^k)\}$ is convergent then all weak accumulation points belong to S and, in particular, if S is a singleton then $\{x^k\}$ is weakly convergent to the unique solution of the problem,
(viii) under the asumption of (vii), if the sequence $\{V(Jx^k, x^*)\}$ is convergent for some $x^* \in S$ and the duality mapping J is sequentially weakly continuous (on some bounded set containing $\{x^k\}$), then $\{x^k\}$ is weakly convergent to a point $\tilde{x} \in S$.

Proof. (i) We return to (3.2) and (3.1) with $\bar{x} = x^* \in S$ again. Using Lemma 2.5(i) we conclude that there exists a subsequence $\{\tilde{\beta}_{\ell_k}\} \subset \{\tilde{\beta}_k\}$ such that $\lim_{k\to\infty} \tilde{\beta}_{\ell_k} = 0$ and the estimate

$$\tilde{\beta}_{\ell_k} = \frac{\langle u^{\ell_k}, x^{\ell_k} - x^* \rangle}{\|u^{\ell_k}\|_*} \leq 32D\alpha_{\ell_k} + \frac{1}{\sum_{j=1}^{\ell_k} \alpha_j} \tag{3.34}$$

is satisfied. It is clear that $\{\tilde{\beta}_{\ell_k}\}$ is bounded and $\tilde{\beta}_{\ell_k} \geq 0$. Thus, we may apply Lemma 3.8 with $\sigma_k = \tilde{\beta}_{\ell_k}$. Since $x^k \in \Omega$ for all $k > 0$ by (1.8), we conclude that $\{x^{\ell_k}\}$ is bounded, in view of the definition of $\tilde{\beta}_{\ell_k}$. Now, it is clear that the sequence $\{V_{\ell_k}\}$, with $V_k = V(Jx^k, x^*)$ is also bounded, because

$$V(Jx^{\ell_k}, x^*) \leq 2^{-1}(\|x^{\ell_k}\| + \|x^*\|)^2,$$

by (P5). In this situation, we conclude from Lemma 2.4(ii) that the whole sequence $\{V_k\}$ is bounded. This implies boundedness of the whole sequence $\{x^k\}$, because

$$V(Jx^k, x^*) \geq 2^{-1}(\|x^k\| - \|x^*\|)^2,$$

again by (P5). Thus the result holds.
(ii) Use (i) to find $C_0 > 0$ such that $\|x^k\| \leq C_0$ for all k. The inclusion $\{x^k\} \subset \Omega$ implies $x^k = \pi_\Omega(Jx^k)$. Then from the inequality

$$\|\pi_\Omega(\phi_1) - \pi_\Omega(\phi_2)\| \leq 2Rg_B^{-1}(4LR^{-1}\|\phi_1 - \phi_2\|_*),$$

established in [1], p.33, where

$$g_B(\epsilon) = \frac{\delta_B(\epsilon)}{\epsilon},$$

and R is any real number satisfying $\max\{\|\phi_1\|_*,\ \|\phi_2\|_*\} \leq R$, we get, using (3.3)-(3.5),

$$\|x^{k+1} - x^k\| = \|\pi_\Omega(\phi^k) - \pi_\Omega(Jx^k)\| \leq 2Rg_B^{-1}(4LR^{-1}\alpha_k), \tag{3.35}$$

with $R = C_0 + \bar{\alpha}$, because

$$\max_k \{\|\phi^k\|_*, \|Jx^k\|_*\} = \max_k \{\|\phi^k\|_*, \|x^k\|\} \leq C_0 + \bar{\alpha}.$$

Recall that $\delta_B(\epsilon) \geq C\epsilon^2$ and hence $g_B(\epsilon) \geq C\epsilon$. Therefore,

$$g_B^{-1}(\epsilon) \leq \frac{\epsilon}{C} \tag{3.36}$$

and we conclude from (3.35) that

$$\|x^{k+1} - x^k\| \leq C_2\alpha_k, \tag{3.37}$$

with $C_2 = 8L/C$.

(iii) Let us return again to (3.2). By (P1) and (i), $\{u^k\}$ is bounded, say $\|u^k\|_* \leq C_3$. Then, the proof of Theorem 3.4 yields the conclusion.
(iv) (3.33) follows from (3.34) and the boundedness of $\{u^k\}$ (cf. the proof of (iii)).
(v) By the convexity of f and item (iii), we get

$$f(x^{k+1}) - f(x^k) \leq \langle u^{k+1}, x^{k+1} - x^k \rangle \leq C_3 \|x^{k+1} - x^k\|.$$

On the other hand,

$$f(x^{k+1}) - f(x^k) \geq \langle u^k, x^{k+1} - x^k \rangle \geq -C_3 \|x^{k+1} - x^k\|.$$

Using (3.37), we have

$$|f(x^{k+1}) - f(x^k)| \leq C_3 \|x^{k+1} - x^k\| \leq 8LC^{-1}C_3\alpha_k.$$

(vi) and (vii) The results follow from boundedness of the sequence $\{x^k\}$, weak lower semicontinuity of convex functions on closed convex sets and Corollary 3.3.
(viii) See the proof of Lemma 2.8 in [7]. ■

Remark 3.10 *In a Hilbert space H, $\pi_\Omega = P_\Omega$, where P_Ω is the metric projection operator and J is the identity operator. Recall that J is sequentially weakly continuous in H. Since*

$$\frac{\epsilon^2}{8} \leq \delta_H(\varepsilon) \leq \frac{\epsilon^2}{4},$$

all our results above hold for the sequence $\{x^k\}$ generated by the subgradient method (1.10), (1.9). This particular case was studied in [6].

Let us make a comment on item (vii) of Theorem 3.9. We will say that the sequence $\{x^k\}$ is $f-$quasi-Fejér (cf. [13]) if

$$f(x^{k+1}) \leq f(x^k) + \gamma_k \text{ and } \sum_{k=1}^{\infty} \gamma_k < \infty. \tag{3.38}$$

It is obvious that if the sequence $\{x^k\}$ generated by (1.8), (1.9) is $f-$quasi-Fejér then $\{f(x^k\})$ has a limit. Furthermore (3.21) and (3.22) imply $\lim_{k\to\infty} f(x^k) = f^*$.

For example, assume that f is a convex and differentiable function in a Hilbert space, and that its gradient $f'(x)$ satisfies the Hölder condition

$$\|f'(x) - f'(y)\| \leq h\|x - y\|^\mu,$$

for some $0 < \mu \leq 1$ and some $h > 0$. In this case,

$$f(x^{k+1}) \leq f(x^k) + \langle f'(x^k), x^{k+1} - x^k \rangle + \langle f'(x^{k+1}) - f'(x^k), x^{k+1} - x^k \rangle,$$

and we have for (1.8), (1.9) with $\Omega = H$, the inequality

$$f(x^{k+1}) \leq f(x^k) + h\|x^{k+1} - x^k\|^{1+\mu}.$$

Under the assumptions of Theorem 3.9, if

$$\sum_{k=1}^{\infty} \alpha_k^{1+\mu} < \infty,$$

then (see (3.37))

$$\sum_{k=1}^{\infty} \|x^{k+1} - x^k\|^{1+\mu} \leq C_2^{1+\mu} \sum_{k=1}^{\infty} \alpha_k^{1+\mu} < \infty,$$

and $\lim_{k\to\infty} f(x^k) = f^*$. We remark that, in general, f in (3.38) need not be differentiable.

In the following theorem, we omit the condition $\delta_B(\epsilon) \geq C\epsilon^2$, assuming just that B is a uniformly convex and uniformly smooth Banach space, but at the same time we impose a stronger requirement on the stepsizes α_k.

Definition 3.11 *Let B be a uniformly convex and uniformly smooth Banach space and K a subset of B. The function $f : B \to \mathbf{R}$ is said to be J-Lipschitz continuous on K if there exists $L_K > 0$ such that*

$$|f(x) - f(y)| \leq L_K \|Jx - Jy\|_*, \ \ \forall x, y \in K.$$

Example 3.12 If $\delta_B(\epsilon) \geq C\epsilon^2$ for some $C > 0$, f is convex and (P1) holds, then f is J-Lipschitz continuous on any bounded set $K \subset B$. Indeed, for all $x, y \in K$ and all $u \in \partial f(x)$, as in the proof of Theorem 3.9(v), we have

$$|f(x) - f(y)| \leq \|u\|_* \|x - y\|.$$

If $\|x\| \leq C_1$ for all $x \in K$ then $\|u\|_* \leq C_2$. Also, using the inequality (7.8) from [1], one can obtain (see (3.36))

$$\|x - y\| \leq 2C_1 g_B^{-1}(4LC_1^{-1}\|Jx - Jy\|_*) \leq 8LC^{-1}\|Jx - Jy\|_*.$$

Hence,

$$|f(x) - f(y)| \leq C_K \|Jx - Jy\|_*,$$

where

$$C_K = 8C^{-1}C_2 L.$$

Example 3.13 The function $f(x) = \|x\|^s, \ s > 1$, is J-Lipschitz continuous on each bounded set K. Indeed, for any $a, b > 0$ and $r > 1$ it holds that

$$a^r - b^r \leq r a^{r-1}(a - b).$$

Therefore,

$$\begin{aligned} |\|x\|^s - \|y\|^s| &\leq s \max_{x,y\in K}\{\|x\|^{s-1}, \|y\|^{s-1}\}|\|x\| - \|y\|| \\ &\leq \bar{C}|\|x\| - \|y\|| \leq \bar{C}\|Jx - Jy\|_*, \end{aligned}$$

because

$$\|Jx - Jy\|_* \geq |\|Jx\|_* - \|Jy\|_*| = |\|x\| - \|y\||.$$

Example 3.14 Fix $\xi \in B$ and let

$$f(x) = V(Jx, \xi) = \frac{1}{2}(\|Jx\|_*^2 - 2\langle Jx, \xi\rangle + \|\xi\|^2),$$

i.e., V is as in (1.6). Since

$$grad_\phi V(Jx, \xi) = x - \xi,$$

we have

$$|V(Jx, \xi) - V(Jy, \xi)| \leq \ max\{\|x - \xi\|, \|y - \xi\|\}\|Jx - Jy\|_*$$

for any $x, y \in B$. Thus, it is clear that f is J-Lipschitz continuous on any bounded subset of B.

We consider next the following method for unconstrained minimization:

$$Jx^{k+1} = Jx^k - \alpha_k \frac{u^k}{\|u^k\|_*}, \quad u^k \in \partial f(x^k) \quad x^{k+1} = J^* Jx^k. \tag{3.39}$$

Theorem 3.15 *Let B be a uniformly convex and uniformly smooth Banach space, and $f : B \to \mathbf{R}$ a convex function, J-Lipschitz continuous on all bounded sets. Suppose that the set S of minimizers of f is nonempty and (P1) is satisfied. Consider the sequence $\{x^k\}$ generated by (3.39), where*

$$\sum_{k=1}^{\infty} \alpha_k = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \rho_{B^*}(\alpha_k) < \infty. \tag{3.40}$$

Then,
(i) the sequence $\{x^k\}$ is bounded,
(ii) $\liminf_{k\to\infty} f(x^k) = f^*$, and moreover, there exists a subsequence $\{x^{\ell_k}\} \subset \{x^k\}$ such that $\beta_{\ell_k} = f(x^{\ell_k}) - f^* \leq (\sum_{k=1}^{\ell_k} \alpha_k)^{-1}$,
(iii) there exists a weak accumulation point $\bar{x}$ of $\{x^k\}$ such that $\bar{x} \in S$,
(iv) $\lim_{k\to\infty} f(x^k) = f^*$, and all weak accumulation points of $\{x^k\}$ belong to S,
(v) if either S is a singleton or the duality mapping J is sequentially weakly continuous (on some bounded set containing $\{x^k\}$), then $\{x^k\}$ converges weakly to a point $\bar{x} \in S$.

The proof is similar to that of Theorem 3.1 in [7]. Indeed, (3.40) implies boundedness of $\{x^k\}$, and, using (P1), one can derive the inequality (cf. (3.18))

$$V_{k+1} \leq V_k - C\alpha_k\beta_k + C_1\rho_{B^*}(C_2\alpha_k), \tag{3.41}$$

where $\beta_k = f(x^k) - f^*$ and C, C_1, C_2 are positive constants. Then Theorem 3.1 and Theorem 3.5 establish (ii) (see also Remark 3.2). Weak lower semicontinuity of convex functions on closed convex sets guarantees (iii). Furthermore, we have, from Definition 3.11 and from (3.39),

$$|\beta_{k+1} - \beta_k| = |f(x^{k+1}) - f(x^k)| \leq L_K\|Jx^{k+1} - Jx^k\| = L_K\alpha_k, \ L_K > 0.$$

Now Lemma 2.2(ii) allows us to conclude that $\lim_{k\to\infty} \beta_k = 0$, so that all weak accumulation points of $\{x^k\}$ belong to S, establishing (iv). Item (v) follows as in the proof of Lemma 2.8 of [7]. ■

Acknowledgements. We thank K. Kiwiel for some very enlightening discussions on a previous version of this paper, and also an anonymous referee, whose comments and suggestions allowed us to attain a considerable improvement over that version.

References

[1] Alber, Ya.I.: Metric and generalized projection operators in Banach spaces: properties and applications. In *Theory and Applications of Nonlinear Operators of Monotone and Accretive Type* (A. Kartsatos, editor). Marcel Dekker, New York (1996) 15-50.

[2] Alber, Ya.I.: The problem of the minimization of smooth functionals by gradient methods, *USSR Computational Mathematics and Mathematical Physcs* **11** (1971), 263-272.

[3] Alber, Ya.I.: Minimization of functionals of class $C^{1,\mu}$ on bounded sets, *Economica i Matematicheskie Metody* **16** (1980), 185-190.

[4] Alber, Ya.I.: Recurrence relations and variational inequalities, *Soviet Matematicheskie Doklady* **27** (1983), 511-517.

[5] Alber, Ya.I.: On average convergence of the iterative projection methods (to be published).

[6] Alber, Ya.I., Iusem, A.N., Solodov, M.V.: On the projection subgradient method for nonsmooth convex optimization in a Hilbert space, *Mathematical Programming* **81** (1998), 23-35.

[7] Alber, Ya.I., Iusem, A.N., Solodov, M.V.: Minimization of nonsmooth convex functionals in Banach spaces, *Journal of Convex Analysis* **4** (1997), 235-254.

[8] Alber, Ya.I., Reich, S.: An iterative method for solving a class of nonlinear operator equations in Banach spaces, *Panamerican Mathematical Journal* **4** (1994), 39-54.

[9] Clarke, F.H. *Optimization and Nonsmooth Analysis*. SIAM Publications, Philadelphia (1983).

[10] Correa, R., Lemaréchal, L.: Convergence of some algorithms for convex minimization, *Mathematical Programming* **62** (1993), 261-275.

[11] Diestel, J.: *The Geometry of Banach Spaces*, Lecture Notes in Mathematics **485**, Springer, Berlin (1975).

[12] Dunn, J.C.: Global and asymptotic convergence rates estimates for a class of projected gradient processes, *SIAM Journal on Control and Optimization* **19** (1981), 368-400.

[13] Ermoliev, Yu.: Methods for solving nonlinear extremal problems, *Cybernetics* **2** (1966), 1-17.

[14] Figiel, T.: On the moduli of convexity and smoothness, *Studia Mathematica* **56** (1976), 121-155.

[15] Golshtein, E., Tretyakov, N.: *Modified Lagrange Functions*, Moscow (1989).

[16] K. Knopp, *Theory and Applications of Infinite Series*, Dover, New York (1990).

[17] Lindenstrauss, J., Tzafriry, L.: *Classical Banach Spaces II*, Springer, Berlin (1979).

[18] Polyak, B.T.: A general method of solving extremum problems, *Soviet Matematicheskie Doklady* **8** (1967), 593–597.

[19] Serb, I.: Estimates for the modulus of smoothness and convexity of a Banach space, *Mathematica* **34** (1992), 61-70.

[20] Vasil'ev, F.P.: *Methods of Solving Extremal Problems*, Nauka, Moskow (1981).