Journal of Convex Analysis
Volume 4 (1997), No. 2, 235–255

Minimization of Nonsmooth Convex Functionals in Banach Spaces*

Ya. I. Alber
Department of Mathematics, The Technion – Israel Institute of Technology, 32000 Haifa, Israel.
e-mail: alberya@tx.technion.ac.il

A. N. Iusem
Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ, CEP 22460-320, Brazil.
e-mail: iusp@impa.br

M. V. Solodov
Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ, CEP 22460-320, Brazil.
e-mail: solodov@impa.br

Received April 29, 1996

We develop a unified framework for convergence analysis of subgradient and subgradient projection methods for minimization of nonsmooth convex functionals in Banach spaces. The important novel features of our analysis are that we neither assume that the functional is uniformly or strongly convex, nor use regularization techniques. Moreover, no boundedness assumptions are made on the level sets of the functional or the feasible set of the problem. In fact, the solution set can be unbounded. Under very mild assumptions, we prove that the sequence of iterates is bounded and it has at least one weak accumulation point which is a minimizer. Moreover, all weak accumulation points of the sequence of Cesàro averages of the iterates are solutions of the minimization problem. Under certain additional assumptions (which are satisfied for several important instances of Banach spaces), we are able to exhibit weak convergence of the whole sequence of iterates to one of the solutions of the optimization problem. To our knowledge, this is the first result of this kind for general nonsmooth convex minimization in Banach spaces.

Keywords: Banach spaces, nonsmooth optimization, subgradient methods, metric projection, generalized projection, weak convergence.

1991 Mathematics Subject Classification: 90C25, 90C30, 49J40, 46M37.

1. Introduction

We consider the problem

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

* Research of the second author is supported in part by CNPq grant 301280/86; research of the third author is supported in part by CNPq grant 300734/95-6

ISSN 0944-6532 / \$ 2.50

of minimizing a convex functional $f : B \to \Re$ over a convex closed subset C of a uniformly smooth and uniformly convex Banach space B. We assume that the solution set S of (1.1) is nonempty.

Let B^* be the dual space to B, and let $J : B \to B^*$ be the normalized duality mapping determined by the following relations :

$$\langle Jx, x\rangle = \|Jx\|_{B^*}\|x\|_B = \|x\|_B^2, \tag{1.2}$$

where $\langle\cdot,\cdot\rangle$ denotes the duality pairing of B^* and B, and $\|\cdot\|_B$, $\|\cdot\|_{B^*}$ denote norms in B and B^*, respectively (when it is clear from the context, we will use $\|\cdot\|$ to denote the norm in B). 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 $J^*J = I_B$. By $\rho_B(\cdot)$ and $\delta_B(\cdot)$ we will denote the moduli of smoothness and convexity, respectively, of a Banach space B [10, 13, 14]. Duality mappings and moduli of smoothness and convexity will play an important role in this paper, and we will elaborate more on their properties in Section 2.

Suppose first that $C = B$, i.e. the problem (1.1) is unconstrained. Let $\partial f(x)$ be the subdifferential of $f(\cdot)$ at $x \in B$, that is

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

We shall consider the following iterative method for solving (1.1): having x^i, a current approximation to a solution, compute

$$x^{i+1} = J^*y^i, \quad \text{where} \quad y^i = Jx^i - \alpha_i\frac{u^i}{\|u^i\|_{B^*}}, \quad u^i \in \partial f(x^i), \tag{1.3}$$

with the stepsizes $\{\alpha_i\}$ chosen according to the rule

$$\sum_{i=0}^{\infty}\alpha_i = \infty, \quad \sum_{i=0}^{\infty}\rho_{B^*}(\alpha_i) < \infty. \tag{1.4}$$

Note that by the properties of $\rho_{B^*}(\cdot)$ (see Section 2), it follows that there exist choices of $\{\alpha_i\}$ such that the two conditions in (1.4) are satisfied. It also follows that $\lim_{i\to\infty}\alpha_i = 0$. In the case of an unconstrained problem in a Hilbert space H (i.e. $C = B = H$), it is relatively easy to show that, if the solution set is nonempty, then the sequence of iterates $\{x^i\}$ generated by the subgradient method is bounded, and it is a "minimizing" sequence in the sense that $\liminf_{i\to\infty} f(x^i) = \inf_{x\in H} f(x)$. Results along those lines in Euclidean and Hilbert spaces have been obtained in [15, 8, 9]. It then follows that the sequence $\{x^i\}$ has a weak accumulation point which is a minimizer. The proof of convergence of the whole sequence $\{x^i\}$ to a weak limit in a Hilbert space (i.e. uniqueness of a weak accumulation point) is considerably more involved and was obtained only recently [2]. It can be claimed that the results reported in [2] made the analysis of subgradient methods in a Hilbert space complete. As we shall see, things become more difficult in a Banach space (even for unconstrained problems, and even more so for problems with constraints). In a Banach space, a method related to (1.3)–(1.4) was studied in [4, 7] under the considerably stronger assumption of $f(\cdot)$ being uniformly convex (or, in the case of operator equations, the operator being uniformly monotone). In that case, the problem solution is unique, and strong convergence of the iterates is established in [4, 7]. We emphasize that, in the setting of this paper, the solution of (1.1) needs not be unique. Moreover, the solution set S may even be unbounded.

In the constrained case (i.e. when $C \neq B$), it is clear that iteration (1.3) should be modified by applying some kind of projection operator which forces the sequence to stay in the feasible set C. However, the metric projection operator $P_C : B \to C$, defined as

$$P_C[x] := \arg\min_{y \in C} \|y - x\|, \tag{1.5}$$

seems to be unsuitable for the task. The reason is that, in a general Banach space, the metric projection operator does not appear to have certain "nice" properties, such as nonexpansiveness and monotonicity (see the discussion in [1]). In this paper, we employ the *generalized* projection operator introduced in [1]. The following function $V : B^* \times B \to \Re$ is central in the subsequent analysis:

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

In a Hilbert space H, $V(\varphi, x) = 2^{-1}\|\varphi - x\|^2_H$ where $x \in H$, $\varphi \in H$. Other properties of this function are summarized in Section 2. We define the generalized projection operator $\pi_C : B^* \to C$ as follows:

$$\pi_C[\varphi] := \arg\min_{x \in C} V(\varphi, x). \tag{1.7}$$

In a Hilbert space H, the generalized projection reduces to the standard metric projection, that is $\pi_C = P_C$. In the case when $C = B$, it holds that $\pi_C = J^*$. It turns out that this projection operator has properties (these are summarized in Section 2) sufficient for convergence of the following iterative algorithm: having x^i, a current approximation to the solution of (1.1), compute

$$x^{i+1} = \pi_C\left[Jx^i - \alpha_i \frac{u^i}{\|u^i\|_{B^*}}\right], \quad u^i \in \partial f(x^i), \tag{1.8}$$

where the stepsizes $\{\alpha_i\}$ are chosen according to (1.4). This scheme is similar to the one considered in [1] for uniformly monotone variational inequalities.

We will also consider a modified method with a metric projection in Banach space. This algorithm is of the following form: having x^i, a current approximation to the solution of (1.1), compute

$$x^{i+1} = J^* y^i, \quad y^i = Jx^i - \alpha_i\left(\frac{\bar{u}^i}{\|\bar{u}^i\|_{B^*}} + 2\frac{J(x^i - \bar{x}^i)}{\|x^i - \bar{x}^i\|}\right), \tag{1.9}$$

where

$$\bar{x}^i = P_C[x^i], \quad \bar{u}^i \in \partial f(\bar{x}^i)$$

with the the stepsizes $\{\alpha_i\}$ chosen according to (1.4). A similar algorithm was proposed in [5] to solve uniformly monotone variational inequalities. We note that in the case studied here, the subdifferential operator $\partial f(\cdot)$ is merely monotone. Another related technique is considered in [3], where no uniform convexity assumption is made but a regularization term is added at every iteration.

2. Preliminary results

This section contains some definitions and preliminary results that have to do with the properties and geometry of Banach spaces [14, 10, 11, 13, 4, 6].

Definition 2.1. A Banach space B is said to be uniformly convex if for any $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in B$, $y \in B$ satisfying $\|x\| = 1$, $\|y\| = 1$, $\|x - y\| = \varepsilon$, it holds that $\|(x + y)/2\| \leq 1 - \delta$.

The function

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

is called the modulus of convexity of B. Banach space B is uniformly convex if and only if $\delta_B(\varepsilon) > 0$ for all $\varepsilon > 0$.

Example 2.2. The Banach spaces l^p, L^p and the Sobolev spaces W^p_m are uniformly convex for all $p \in (1, \infty)$. Let B be any of those spaces. Then

For $p \in [2, \infty)$, $\quad \delta_B(\varepsilon) \geq p^{-1}2^{-p}\varepsilon^p$.
For $p \in (1, 2]$, $\quad \delta_B(\varepsilon) \geq (p - 1)8^{-1}\varepsilon^2$.

Definition 2.3. A Banach space B is said to be uniformly smooth if for any $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in B$, $y \in B$ satisfying $\|x\| = 1$, $\|y\| \leq \delta$, it holds that $\|(x + y)/2\| + \|(x - y)/2\| - 1 \leq \varepsilon\|y\|$.

The function

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

is called the modulus of smoothness of B. Banach space B is uniformly smooth if and only if $\lim_{\tau\to 0} \tau^{-1}\rho_B(\tau) = 0$.

Example 2.4. The Banach spaces l^p, L^p and Sobolev spaces W^p_m are uniformly smooth for all $p \in (1, \infty)$. Let B be any of those spaces. Then

For $p \in [2, \infty)$, $\quad \rho_B(\tau) \leq (p - 1)\tau^2$.
For $p \in (1, 2]$, $\quad \rho_B(\tau) \leq p^{-1}\tau^p$.

It is known that for any Hilbert space H and any uniformly smooth and uniformly convex Banach space B,

$$\rho_B(\tau) \geq \rho_H(\tau) \quad \text{and} \quad \delta_B(\varepsilon) \leq \delta_H(\varepsilon).$$

Furthermore, the moduli of smoothness and convexity of any Hilbert space H are given by (see [10])

$$\rho_H(\tau) = (1 + \tau^2)^{1/2} - 1 \quad \text{and} \quad \delta_H(\varepsilon) = 1 - (1 - (\varepsilon/2)^2)^{1/2}.$$

Therefore, for any uniformly smooth and uniformly convex Banach space B,

$$\rho_B(\tau) \geq \frac{\tau^2}{1+(1+\tau^2)^{1/2}}. \tag{2.1}$$

These facts will be used in the sequel.
We next give examples of the duality mapping $J(\cdot)$ in the uniformly smooth and uniformly convex Banach spaces l^p, L^P and W^p_m, $\quad p \in (1,\infty)$.

Example 2.5. (See [1])

For l^p, $Jx = \|x\|^{2-p}_{l^p} y \in l^q$, where $x = \{x_1, x_2, \dots\}$ and $y = \{x_1|x_1|^{p-2}, x_2|x_2|^{p-2}, \dots\}$ with $1/p + 1/q = 1$.
For L^p, $Jx = \|x\|^{2-p}_{L^p}|x|^{p-2}x \in L^q$, where $1/p+1/q=1$.
For W^p_m, $Jx = \|x\|^{2-p}_{W^p_m}\sum(-1)^{|t|}D^t(|D^t x|^{p-2}D^t x) \in W^q_{-m}$, where $1/p+1/q=1$.

The following two lemmas are proved in [4, 6, 1].

Lemma 2.6. *Let B be a uniformly smooth and uniformly convex Banach space. The following properties hold:*

1. *$J(\cdot)$ is a uniformly monotone operator on any bounded set. More precisely, for all $x \in B$, $y \in B$*

$$\langle Jx - Jy, x-y\rangle \geq C_1^2(\|x\|,\|y\|)\delta_B\left(\frac{\|x-y\|}{2C_1(\|x\|,\|y\|)}\right),$$

where

$$C_1(\|x\|,\|y\|) = (\|x\|^2/2 + \|y\|^2/2)^{1/2}.$$

If $\|x\| \leq C_1$ and $\|y\| \leq C_1$, where C_1 is constant, then

$$\langle Jx - Jy, x-y\rangle \geq \frac{C_1^2}{2L}\delta_B\left(\frac{\|x-y\|}{2C_1}\right),$$

$1 < L < 3.18$ *is the Figiel constant* [11].

2. *For all $x \in B$, $y \in B$*

$$\langle Jx - Jy, x-y\rangle \leq 8\|Jx-Jy\|^2_{B^*} + C_2(\|x\|,\|y\|)\rho_{B^*}(\|Jx-Jy\|_{B^*}),$$

where

$$C_2(\|x\|,\|y\|) \leq 4\max\{2L; \|x\|+\|y\|\}.$$

3. *For all $x \in B$, $y \in B$, such that $\|x\| \leq C_1$ and $\|y\| \leq C_1$,*

$$\|x-y\| \leq 2C_1 g_B^{-1}(4C_1^{-1}L\|Jx-Jy\|_{B^*}),$$

where $g_B(\varepsilon) = \delta_B(\varepsilon)/\varepsilon$.

Note that if x and y are contained in some bounded set, then C_1 and C_2 are absolute constants.

Lemma 2.7 summarizes the properties of $V(\cdot,\cdot)$ defined by (1.6), and $\pi_C[\cdot]$ defined by (1.7) which will be used later.

Lemma 2.7. *The following properties hold:*

1. *$V(\cdot,x)$ is differentiable for each $x \in B$ and $\nabla_\varphi V(\varphi,x) = J^*\varphi - x$;*
2. *$V(\cdot,x)$ is convex for each $x \in B$. Therefore, for all $x \in B$, $y \in B$, $z \in B$*
$$V(Jy,x) \le V(Jz,x) + \langle Jy - Jz, y - x\rangle;$$
3. *For all $x \in B$, $\varphi \in B^*$*
$$\frac{1}{2}(\|\varphi\|_{B^*} - \|x\|)^2 \le V(\varphi,x) \le \frac{1}{2}(\|\varphi\|_{B^*} + \|x\|)^2.$$
In particular, $V(\varphi,x)$ is nonnegative and finite, and $V(Jy,x) = 0$ for some $y \in B$, $x \in B$ if and only if $y = x$.
4. *For every $x \in B$, $y \in B$, such that $\|x\| \le C_1$ and $\|y\| \le C_1$,*
$$2L^{-1}C_1^2\delta_B\left(\frac{\|x-y\|}{4C_1}\right) \le V(Jx,y) \le 4L^{-1}C_1^2\rho_B\left(\frac{4\|x-y\|}{C_1}\right).$$
5. *The generalized projection operator $\pi_C[\cdot]$, given by (1.7), is well-defined.*
6. *The following property of absolutely best approximation holds: for any $\varphi \in B^*$ and any $x \in C$*
$$V(J\pi_C[\varphi],x) \le V(\varphi,x) - V(\varphi,\pi_C[\varphi]).$$
In particular, this implies the property of "conditional" nonexpansiveness:
$$V(J\pi_C[\varphi],x) \le V(\varphi,x).$$
7. *For any $\varphi \in B^*$ and any $x \in C$*
$$\langle \varphi - J\pi_C[\varphi], \pi_C[\varphi] - x\rangle \ge 0.$$
8. *The operator $\pi_C[\cdot]$ is J-fixed in every point $x \in C$, i.e.*
$$\pi_C[Jx] = x.$$
9. *Let $R > 0$ and $T := \{\varphi \in B^* \mid \|\pi_C[\varphi]\| \le R\}$. Then for any $\varphi_1 \in T$ and $\varphi_2 \in T$,*
$$\|\pi_C[\varphi_1] - \pi_C[\varphi_2]\| \le 2Rg_B^{-1}\left(\frac{4L}{R}\|\varphi_1 - \varphi_2\|_{B^*}\right).$$

The following lemma is of independent interest. It can be thought of as a generalization of the result on quasi-Fejér convergent sequences in a Hilbert space [2, Proposition 1]. In particular, the generalization is two-fold. First of all, quasi-Fejér relation (2.2) is written with respect to the functional $V(\cdot,\cdot)$ defined by (1.6). This functional arises naturally in the analysis in Banach spaces and, in a Hilbert space, it reduces to the standard distance function induced by the norm. The other generalization consists of allowing an additional term β_i in inequality (2.2) below ($\beta_i = 0$ was considered in [2]). Lemma 2.8 is the basis for our convergence analysis.

Lemma 2.8. *Let T be a nonempty subset of a Banach space B. Let $\{x^i\}$ be a sequence in B such that for every $\tilde{x} \in T$ there exist two sequences of nonnegative real numbers $\{\beta_i\}$ and $\{\gamma_i\}$ satisfying $\sum_{i=0}^{\infty} \beta_i < \infty$, $\sum_{i=0}^{\infty} \gamma_i < \infty$ and*

$$V(Jx^{i+1}, \tilde{x}) \leq (1+\beta_i)V(Jx^i, \tilde{x}) + \gamma_i. \tag{2.2}$$

Then the following properties hold:

1. *The sequence $\{V(Jx^i, \tilde{x})\}$ is bounded for all $\tilde{x} \in T$;*
2. *The sequence $\{x^i\}$ is bounded;*
3. *The sequence $\{V(Jx^i, \tilde{x})\}$ converges to a nonnegative limit for all $\tilde{x} \in T$;*
4. *If the set of weak accumulation points W of $\{x^i\}$ belongs to T, and if, in addition, either T is a singleton, or $J(\cdot)$ is a sequentially weakly continuous operator (on some bounded set containing $\{x^i\}$), then $\{x^i\}$ is weakly convergent to a point in T.*

Proof. The fact that the sequence $\{V(Jx^i, \tilde{x})\}$ converges to a nonnegative limit follows from a well known lemma on numerical sequences [16, Lemma 2, p.44]. Thus $\{V(Jx^i, \tilde{x})\}$ is also bounded. Boundedness of the sequence $\{x^i\}$ now follows from Lemma 2.7 (3). So the first three assertions are established.

Existence of weak accumulation points of the sequence $\{x^i\}$ follows from Part (2). Let $W \neq \emptyset$ be the set of weak accumulation points of $\{x^i\}$. It is assumed that $W \subset T$. If T is a singleton, i.e. it consists of just one point, it immediately follows that $\{x^i\}$ converges weakly to this point. In that case, the claim is established. Otherwise suppose that $J(\cdot)$ is sequentially weakly continuous, i.e. from the fact that $\{x^k\}$ weakly converges to x it follows that $\{J(x^k)\}$ weakly converges to Jx. Let $\hat{x} \in W$, $\tilde{x} \in W$ and $\{x^{i_n}\}$, $\{x^{i_m}\}$ be two subsequences of $\{x^i\}$ weakly converging to $\hat{x}$ and $\tilde{x}$ respectively. Then we have

$$2V(Jx^i, \hat{x}) = \|x^i\|^2 - 2\langle Jx^i, \hat{x}\rangle + \|\hat{x}\|^2$$

and

$$2V(Jx^i, \tilde{x}) = \|x^i\|^2 - 2\langle Jx^i, \tilde{x}\rangle + \|\tilde{x}\|^2.$$

Since, by Part (3), both $\{V(Jx^i, \hat{x})\}$ and $\{V(Jx^i, \tilde{x})\}$ converge, it follows that

$$\begin{aligned} v &= 2\lim_{i\to\infty}(V(Jx^i, \hat{x}) - V(Jx^i, \tilde{x})) \\ &= 2\lim_{i\to\infty}\langle Jx^i, \tilde{x} - \hat{x}\rangle + \|\hat{x}\|^2 - \|\tilde{x}\|^2. \end{aligned}$$

Taking the limit as $n \to \infty$ along the subsequence $\{x^{i_n}\}$, we obtain

$$v = 2\langle J\hat{x}, \tilde{x} - \hat{x}\rangle + \|\hat{x}\|^2 - \|\tilde{x}\|^2,$$

while taking the limit as $m \to \infty$ along the subsequence $\{x^{i_m}\}$, we have

$$v = 2\langle J\tilde{x}, \tilde{x} - \hat{x}\rangle + \|\hat{x}\|^2 - \|\tilde{x}\|^2.$$

Subtracting the last two relations yields

$$\begin{aligned} 0 &= 2\langle J\tilde{x} - J\hat{x}, \tilde{x} - \hat{x}\rangle \\ &\geq \frac{C_1^2}{2L}\delta_B\left(\frac{\|\tilde{x} - \hat{x}\|}{2C_1}\right), \end{aligned}$$

where the inequality follows from Lemma 2.6 (1). It immediately follows that $\tilde{x} = \hat{x}$. This completes the proof. □

Remark 2.9. We refer the reader to [1] for examples of Banach spaces with sequentially weakly continuous duality mapping $J(\cdot)$ and a discussion of related issues.

3. Unconstrained minimization

In this section, we consider the iterative method given by (1.3), (1.4), for solving (1.1) in the unconstrained case when $C = B$.

Throughout the rest of the paper, we make the following standing assumptions: the solution set S is nonempty, and the subdifferential operator $\partial f(\cdot)$ is bounded on bounded sets. These assumptions are not restrictive. In fact, a sufficient (and also necessary) condition for $\partial f(\cdot)$ to be bounded on bounded sets is boundedness of $|f(\cdot)|$ on bounded sets. This issue is discussed in more detail in [2].

Our main result of this section is the following.

Theorem 3.1. *Let B be a uniformly smooth and uniformly convex Banach space. Let $f : B \to \Re$ be a convex functional. For any sequence of iterates $\{x^i\}$ generated by* (1.3), (1.4) *the following statements hold:*

1. *$\{x^i\}$ is bounded;*
2. $\liminf_{i\to\infty} f(x^i) = \min_{x\in B} f(x) =: f^*$. *Moreover, there exists a subsequence $\{x^{i_k}\}$ of $\{x^i\}$ such that $f(x^{i_k}) - f^* \leq \left(\sum_{j=0}^{i_k} \alpha_j\right)^{-1}$;*
3. *There exists $\bar{x}$, a weak accumulation point of $\{x^i\}$, such that $\bar{x} \in S$.*
4. *If, in addition, the modulus of convexity of the Banach space B is such that $\delta_B(\varepsilon) \geq D_1\varepsilon^2$ for some constant $D_1 > 0$, then $\lim_{i\to\infty} f(x^i) = f^*$ and all weak accumulation points of $\{x^i\}$ belong to the solution set S.*
5. *If $\delta_B(\varepsilon) \geq D_1\varepsilon^2$ and either S is a singleton or the duality mapping $J(\cdot)$ is sequentially weakly continuous (on some bounded set containing $\{x^i\}$), then $\{x^i\}$ converges weakly to a point $\bar{x} \in S$.*

Proof. First of all, by (1.3), we have that

$$Jx^{i+1} - Jx^i = JJ^*y^i - Jx^i = y^i - Jx^i = -\alpha_i\frac{u^i}{\|u^i\|_{B^*}},$$

and

$$\|Jx^{i+1} - Jx^i\|_{B^*} = \alpha_i.$$

Take any $\bar{x} \in S$, and let $f^* := f(\bar{x})$. By Lemma 2.7 (2), we obtain

$$\begin{aligned} V(Jx^{i+1}, \bar{x}) &\le V(Jx^i, \bar{x}) + \langle Jx^{i+1} - Jx^i, x^{i+1} - \bar{x} \rangle \\ &= V(Jx^i, \bar{x}) + \langle Jx^{i+1} - Jx^i, x^i - \bar{x} \rangle + \langle Jx^{i+1} - Jx^i, x^{i+1} - x^i \rangle. \end{aligned} \tag{3.1}$$

By convexity of $f(\cdot)$, we have

$$\begin{aligned} \langle Jx^{i+1} - Jx^i, x^i - \bar{x} \rangle &= -\frac{\alpha_i}{\|u^i\|_{B^*}} \langle u^i, x^i - \bar{x} \rangle \\ &\le -\frac{\alpha_i}{\|u^i\|_{B^*}} (f(x^i) - f^*) \le 0. \end{aligned} \tag{3.2}$$

We now turn our attention to the last term in (3.1). By Lemma 2.6 (2),

$$\begin{aligned} \langle Jx^{i+1} - Jx^i, x^{i+1} - x^i \rangle &\le 8\|Jx^{i+1} - Jx^i\|_{B^*}^2 + C_2\left(\|x^{i+1}\|, \|x^i\|\right) \rho_{B^*}\left(\|Jx^{i+1} - Jx^i\|_{B^*}\right) \\ &= 8\alpha_i^2 + C_2\left(\|x^{i+1}\|, \|x^i\|\right) \rho_{B^*}(\alpha_i), \end{aligned} \tag{3.3}$$

where

$$C_2(\|x^{i+1}\|, \|x^i\|) \le 4 \max\{2L; \|x^{i+1}\| + \|x^i\|\}.$$

Note that

$$\|x^{i+1}\| = \|Jx^{i+1}\|_{B^*} = \|y^i\|_{B^*} = \|Jx^i - \alpha_i \frac{u^i}{\|u^i\|_{B^*}}\|_{B^*} \le \|x^i\| + \alpha_i,$$

where the last step follows from the triangular inequality. Thus

$$C_2(\|x^{i+1}\|, \|x^i\|) \le 4 \max\{2L; 2\|x^i\| + \alpha_i\}.$$

Combining the last relation with (3.1)–(3.3), we obtain

$$V(Jx^{i+1}, \bar{x}) \le V(Jx^i, \bar{x}) + 8\alpha_i^2 + 4\max\{2L; 2\|x^i\| + \alpha_i\}\rho_{B^*}(\alpha_i). \tag{3.4}$$

For every i, we consider two possible cases:

1. $2\|x^i\| + \alpha_i < 2L$, or
2. $2\|x^i\| + \alpha_i \ge 2L$.

In the first case (3.4) yields

$$V(Jx^{i+1}, \bar{x}) \le V(Jx^i, \bar{x}) + 8\alpha_i^2 + 8L\rho_{B^*}(\alpha_i). \tag{3.5}$$

By (2.1),

$$\rho_{B^*}(\alpha_i) \ge \frac{\alpha_i^2}{1 + (1 + \alpha_i^2)^{1/2}}.$$

Because $\{\alpha_i\}$ is bounded above, say by some $\bar{\alpha} > 0$, it follows that

$$\rho_{B^*}(\alpha_i) \ge \frac{\alpha_i^2}{D_2},$$

where $D_2 := 1 + (1+\bar{\alpha})^{1/2}$. Hence, by (3.5),

$$V(Jx^{i+1}, \bar{x}) \le V(Jx^i, \bar{x}) + 8(D_2 + L)\rho_{B^*}(\alpha_i). \tag{3.6}$$

In the second case, (3.4) yields

$$\begin{aligned} V(Jx^{i+1}, \bar{x}) &\le V(Jx^i, \bar{x}) + 8\alpha_i^2 + 8\|x^i\|\rho_{B^*}(\alpha_i) + 4\alpha_i\rho_{B^*}(\alpha_i) \\ &= V(Jx^i, \bar{x}) + 8\alpha_i^2 + 4(\alpha_i + 2\|\bar{x}\|)\rho_{B^*}(\alpha_i) + 8(\|x^i\| - \|\bar{x}\|)\rho_{B^*}(\alpha_i). \end{aligned} \tag{3.7}$$

Taking into account that for any real numbers a and b, $2ab \le a^2 + b^2$, we further obtain

$$\begin{aligned} 2(\|x^i\| - \|\bar{x}\|)\rho_{B^*}(\alpha_i) &= 2(\|x^i\| - \|\bar{x}\|)\sqrt{\rho_{B^*}(\alpha_i)}\sqrt{\rho_{B^*}(\alpha_i)} \\ &\le (\|x^i\| - \|\bar{x}\|)^2\rho_{B^*}(\alpha_i) + \rho_{B^*}(\alpha_i) \\ &\le (2V(Jx^i, \bar{x}) + 1)\rho_{B^*}(\alpha_i), \end{aligned}$$

where the last inequality follows from Lemma 2.7 (3). Combining the latter relation with (3.7), we obtain

$$V(Jx^{i+1}, \bar{x}) \le (1 + 8\rho_{B^*}(\alpha_i))V(Jx^i, \bar{x}) + 8\alpha_i^2 + 4(\alpha_i + 2\|\bar{x}\| + 1)\rho_{B^*}(\alpha_i). \tag{3.8}$$

Because $\rho_{B^*}(\alpha_i) \ge \alpha_i^2/D_2$, it follows that

$$V(Jx^{i+1}, \bar{x}) \le (1 + 8\rho_{B^*}(\alpha_i))V(Jx^i, \bar{x}) + 4(2D_2 + \bar{\alpha} + 2\|\bar{x}\| + 1)\rho_{B^*}(\alpha_i). \tag{3.9}$$

From (3.6) and (3.9) it follows that, for all i,

$$V(Jx^{i+1}, \bar{x}) \le (1 + 8\rho_{B^*}(\alpha_i))V(Jx^i, \bar{x}) + D_3\rho_{B^*}(\alpha_i), \tag{3.10}$$

where $D_3 := \max\{8(D_2 + L); 4(2D_2 + \bar{\alpha} + 2\|\bar{x}\| + 1)\}$. We are now in position to apply Lemma 2.8 (with $T = S$, $\beta_i = 8\rho_{B^*}(\alpha_i)$ and $\gamma_i = D_3\rho_{B^*}(\alpha_i)$). In particular, it follows that $\{x^i\}$ is bounded, say, by C_1.

We now prove the second assertion of the theorem. By (3.1), (3.2) and (3.10), we obtain

$$V(Jx^{i+1}, \bar{x}) \le (1 + 8\rho_{B^*}(\alpha_i))V(Jx^i, \bar{x}) - \frac{\alpha_i}{\|u^i\|_{B^*}}(f(x^i) - f^*) + D_3\rho_{B^*}(\alpha_i).$$

Under our assumptions, from boundedness of $\{x^i\}$ it follows that there exists some positive constant D_4 such that $\|u^i\|_{B^*} \le D_4^{-1}$. Therefore

$$V(Jx^{i+1}, \bar{x}) \le (1 + 8\rho_{B^*}(\alpha_i))V(Jx^i, \bar{x}) - D_4\alpha_i(f(x^i) - f^*) + D_3\rho_{B^*}(\alpha_i). \tag{3.11}$$

Let $s_i := \left(\sum_{j=0}^{i-1} \alpha_j\right)^{-1}$. For every i, we consider two possible cases.

1. $f(x^i) - f^* \le s_i$, or
2. $f(x^i) - f^* > s_i$.

We will show that there exists an infinite sequence of indices for which the first case takes place. Suppose the opposite. Then for i sufficiently large, say $i \geq i_0$, the second inequality above is satisfied. For any $i \geq i_0$, by (3.11) and boundedness of $\{V(Jx^i, \bar{x})\}$, we obtain

$$\begin{aligned} V(Jx^{i+1}, \bar{x}) &\leq V(Jx^i, \bar{x}) - D_4 \alpha_i s_i + D_5 \rho_{B^*}(\alpha_i) \\ &\leq V(Jx^0, \bar{x}) - D_4 \sum_{j=i_0}^{i} \alpha_j s_j + D_5 \sum_{j=i_0}^{i} \rho_{B^*}(\alpha_j), \end{aligned}$$

where $D_5 := D_3 + D_6$ with D_6 satisfying $8V(Jx^i, \bar{x}) \leq D_6$ for all i. By (1.4) and the Abel-Dini criterion for divergent series, it follows that the first series above diverges; on the other hand, the second series above is finite by (1.4). But then the latter inequality implies that $\{V(Jx^i, \bar{x})\}$ is unbounded below, which contradicts Lemma 2.8 (1). Hence there exists an infinite subsequence of indices $\{i_k\}$ for which the first condition holds. This condition, combined with the fact that $\lim_{i\to\infty} s_i = 0$ (because of (1.4)), completes the proof of the second assertion of the theorem.

Let $\{x^{i_k}\}$ be a subsequence of $\{x^i\}$ such that $\lim_{k\to\infty} f(x^{i_k}) = \liminf_{i\to\infty} f(x^i) = f^*$. Since $\{x^{i_k}\}$ is bounded (because so is $\{x^i\}$), it follows that it has some weak accumulation point $\tilde{x}$. Let $\{x^{i_{k_m}}\}$ be a subsequence of $\{x^{i_k}\}$ whose weak limit is $\tilde{x}$. By weak lower semicontinuity of convex functionals [12], it follows that

$$\begin{aligned} f(\tilde{x}) &\leq \liminf_{m\to\infty} f(x^{i_{k_m}}) \\ &= \lim_{k\to\infty} f(x^{i_k}) \\ &= f^*. \end{aligned}$$

Therefore $\tilde{x} \in S$.

We now prove the fourth assertion of the theorem. Denote $\lambda_i := f(x^i) - f^* \geq 0$. By (3.11) we have that

$$D_4 \alpha_i \lambda_i \leq V(Jx^i, \bar{x}) - V(Jx^{i+1}, \bar{x}) + D_5 \rho_{B^*}(\alpha_i),$$

where D_5 was defined above. Summing the latter relation from 0 to j, we obtain

$$\begin{aligned} D_4 \sum_{i=0}^{j} \alpha_i \lambda_i &\leq \sum_{i=0}^{j} (V(Jx^i, \bar{x}) - V(Jx^{i+1}, \bar{x})) + D_5 \sum_{i=0}^{j} \rho_{B^*}(\alpha_i) \\ &= V(Jx^0, \bar{x}) - V(Jx^{j+1}, \bar{x}) + D_5 \sum_{i=0}^{j} \rho_{B^*}(\alpha_i). \end{aligned}$$

Letting $j \to \infty$, from (1.4) and boundedness of $\{V(Jx^i, \bar{x})\}$ it follows that

$$\sum_{i=0}^{\infty} \alpha_i \lambda_i < \infty. \tag{3.12}$$

Suppose that $\{\lambda_i\}$ does not converge to zero. Then there exist some $\tau > 0$ and some subsequence $\{\lambda_{i_m}\}$ such that $\lambda_{i_m} \geq \tau$ for all m. For each m, by $k(m)$ denote the smallest integer greater than i_m such that $\lambda_{k(m)} \leq \tau/2$. Note that this is well defined because we have already established that there exists a subsequence $\{\lambda_{i_k}\}$ such that $\lim_{k\to\infty} \lambda_{i_k} = 0$. By definition of $k(m)$ we have that $\lambda_j > \tau/2$ for all $j \in \{i_m, \ldots, k(m)-1\}$. By (3.12) we obtain that

$$\begin{aligned} \infty &> \sum_{i=0}^{\infty} \alpha_i \lambda_i \\ &\geq \sum_{m=0}^{\infty} \sum_{j=i_m}^{k(m)-1} \alpha_j \lambda_j \\ &\geq \frac{\tau}{2} \sum_{m=0}^{\infty} \sum_{j=i_m}^{k(m)-1} \alpha_j . \end{aligned}$$

It follows that

$$\lim_{m\to\infty} \left(\sum_{j=i_m}^{k(m)-1} \alpha_j \right) = 0. \tag{3.13}$$

By convexity of $f(\cdot)$, the Cauchy-Schwartz inequality and boundedness of $\{u^i\}$, we have

$$\begin{aligned} \lambda_i - \lambda_{i+1} &= f(x^i) - f(x^{i+1}) \\ &\leq \langle u^i, x^i - x^{i+1} \rangle \\ &\leq \|u^i\|_{B^*} \|x^{i+1} - x^i\| \\ &\leq D_4^{-1} \|x^{i+1} - x^i\|. \end{aligned} \tag{3.14}$$

By Lemma 2.6 (3),

$$\|x^{i+1} - x^i\| \leq C_1 g_B^{-1} (4 C_1^{-1} L \|Jx^{i+1} - Jx^i\|_{B^*}).$$

Therefore

$$\|x^{i+1} - x^i\| \ \leq \ C_1 g_B^{-1} (4 C_1^{-1} L \alpha_i).$$

Assume now that the modulus of convexity of the Banach space B satisfies $\delta_B(\varepsilon) \geq D_1 \varepsilon^2$ for some $D_1 > 0$. Recall that $g_B(\varepsilon) = \delta_B(\varepsilon)/\varepsilon$ and $g_B(0) = 0$. Under our assumption, it follows that $g_B(\varepsilon) \geq D_1 \varepsilon$ and hence $g_B^{-1}(\varepsilon) \leq D_1^{-1} \varepsilon$. Therefore

$$\|x^{i+1} - x^i\| \leq 4 D_1^{-1} L \alpha_i. \tag{3.15}$$

By the choice of indices i_m and $k(m)$, and by (3.14) and (3.15) we further obtain

$$\begin{aligned}\frac{\tau}{2} &\le \lambda_{i_m} - \lambda_{k(m)} \\ &= \sum_{j=i_m}^{k(m)-1} (\lambda_j - \lambda_{j+1}) \\ &\le D_4^{-1} \sum_{j=i_m}^{k(m)-1} \|x^{j+1} - x^j\| \\ &\le 4D_1^{-1} L D_4^{-1} \sum_{j=i_m}^{k(m)-1} \alpha_j,\end{aligned}$$

which contradicts (3.13). Hence $\lim_{i\to\infty} \lambda_i = 0$, that is $\lim_{i\to\infty} f(x^i) = f^*$. Let $\tilde{x}$ be any weak accumulation point of $\{x^i\}$ and $\{x^{i_k}\}$ be any subsequence which weakly converges to $\tilde{x}$. By weak lower semicontinuity of convex functionals, we have

$$\begin{aligned}f(\tilde{x}) &\le \liminf_{k\to\infty} f(x^{i_k}) \\ &= \lim_{i\to\infty} f(x^i) \\ &= f^*.\end{aligned}$$

It follows that the set of weak accumulation points of $\{x^i\}$ is contained in S, and the fourth assertion is established.
Now the last assertion follows from Lemma 2.8 (4). The proof is complete. □

Remark 3.2. Similarly to [2], we could consider the use of inexact subgradients in (1.3). In particular, we could choose u^i from an appropriate set of ε-subgradients defined by

$$\partial_\varepsilon f(x) = \{u \in B^* \mid f(y) - f(x) \ge \langle u, y - x\rangle - \varepsilon \text{ for all } y \in B\},$$

where $\varepsilon \ge 0$. If $u^i \in \partial_{\varepsilon_i} f(x^i)$ and $\{\varepsilon_i\}$ are chosen so that

$$\varepsilon_i \le \mu\alpha_i$$

for some nonnegative constant μ, then Theorem 3.1 is still valid.

Remark 3.3. Similarly to [2], we could treat the case when the solution set S of (1.1) is empty. In particular, it still holds that $\liminf_{i\to\infty} f(x^i) = \inf_{x\in C} f(x)$. It then follows that if S is empty, then the sequence $\{x^i\}$ must be unbounded.

Remark 3.4. The assumption $\delta_B(\varepsilon) \ge D_1\varepsilon^2$ which is needed in parts 4 and 5 of the theorem, is satisfied, for example, by l^p, L^p and W_m^p for $p \in (1, 2]$ (see Example 2.2). Examples of Banach spaces with sequentially weakly continuous $J(\cdot)$ can be found in [1].

We conclude this section with an interesting result which establishes strong convergence of the sequence of projections of $\{x^i\}$ onto the solution set S.

Theorem 3.5. *Let B be a uniformly smooth and uniformly convex Banach space, and $f : B \to \Re$ be a convex functional. Let $\{x^i\}$ be a sequence generated by (1.3), (1.4) and let $v^i := \pi_S[Jx^i]$. It follows that the sequence $\{v^i\}$ converges strongly to a point $v \in S$.*

Proof. Since $v^i \in S$, from the analysis in Theorem 3.1, we have that

$$V(Jx^{i+1}, v^i) \leq V(Jx^i, v^i) + D\rho_{B^*}(\alpha_i),$$

for some $D > 0$. Hence, for any $m \geq 1$,

$$\begin{aligned} V(Jx^{i+m}, v^i) &\leq V(Jx^i, v^i) + D \sum_{j=i}^{i+m-1} \rho_{B^*}(\alpha_j) \\ &\leq V(Jx^i, v^i) + s_i, \end{aligned} \tag{3.16}$$

where $s_i := D\sum_{j=i}^{\infty} \rho_{B^*}(\alpha_j)$. By Lemma 2.7 (6),

$$V(Jx^{i+m}, v^{i+m}) \leq V(Jx^{i+m}, v^i) - V(Jv^{i+m}, v^i).$$

Combining the latter relation with (3.16), we obtain

$$\begin{aligned} V(Jx^{i+m}, v^{i+m}) &\leq V(Jx^i, v^i) - V(Jv^{i+m}, v^i) + s_i \qquad (3.17) \\ &\leq V(Jx^i, v^i) + s_i. \end{aligned}$$

Because m is arbitrary and $\lim_{i\to\infty} s_i = 0$, it now follows that the sequence $\{V(Jx^i, v^i)\}$ converges. By Lemma 2.7 (3) it then follows that the sequence $\{v^i\}$ is bounded. By (3.17),

$$V(Jv^{i+m}, v^i) \leq V(Jx^i, v^i) - V(Jx^{i+m}, v^{i+m}) + s_i.$$

Since $\{V(Jx^i, v^i)\}$ converges and $\{s_i\}$ tends to zero, it follows that

$$\lim_{i\to\infty} V(Jv^{i+m}, v^i) = 0.$$

By Lemma 2.7 (4), we conlcude that

$$\lim_{i\to\infty} \delta_B\left(\frac{\|v^{i+m} - v^i\|}{2C_1}\right) = 0,$$

where, by boundedness of $\{v^i\}$, C_1 is a constant. Consequently

$$\lim_{i\to\infty} \|v^{i+m} - v^i\| = 0.$$

Since m was arbitrary, $\{v^i\} \subset S$ and S is a closed set, it follows that $\{v^i\}$ converges strongly to some $v \in S$. □

4. Constrained minimization

In this section, we consider methods for solving optimization problems with constraints. In particular, two methods are presented: one makes use of the generalized projection operator discussed above, and the other utilizes standard metric projection but modifies the subgradient direction.

4.1. Method with generalized projection

We first consider the algorithm generated according to (1.8), (1.4), for solving constrained optimization problems in Banach spaces. This method employs generalized projection operator (1.7) instead of the "standard" metric projection. This novel feature allows us to prove convergence properties which are not known for the metric projection scheme.

Theorem 4.1. *Let B be a uniformly smooth and uniformly convex Banach space. Let $f : B \to \Re$ be a convex functional. For any sequence generated by* (1.8), (1.4) *all the conclusions of* Theorem 3.1 *are valid.*

Proof. Denote $\varphi^i := Jx^i - \alpha_i u^i/\|u^i\|_{B^*}$. With this definition we have $x^{i+1} = \pi_C[\varphi^i]$ and $\|\varphi^i - Jx^i\|_{B^*} = \alpha_i$. Take any $\bar{x} \in S$. By Lemma 2.7 (2), we obtain

$$V(\varphi^i, \bar{x}) \le V(Jx^i, \bar{x}) + 2\langle \varphi^i - Jx^i, J^*\varphi^i - \bar{x}\rangle.$$

By Lemma 2.7 (6), $V(Jx^{i+1}, \bar{x}) \le V(\varphi^i, \bar{x})$. Hence

$$V(Jx^{i+1}, \bar{x}) \le V(Jx^i, \bar{x}) + 2\langle \varphi^i - Jx^i, J^*\varphi^i - \bar{x}\rangle. \tag{4.1}$$

We further obtain

$$\begin{aligned}
\langle \varphi^i - Jx^i, J^*\varphi^i - \bar{x}\rangle &= \langle \varphi^i - Jx^i, x^i - \bar{x}\rangle + \langle \varphi^i - Jx^i, J^*\varphi^i - x^i\rangle \\
&= -\frac{\alpha_i}{\|u^i\|_{B^*}}\langle u^i, x^i - \bar{x}\rangle + \langle \varphi^i - Jx^i, J^*\varphi^i - J^*Jx^i\rangle \\
&\le -\frac{\alpha_i}{\|u^i\|_{B^*}}(f(x^i) - f^*) + 8\|\varphi^i - Jx^i\|_{B^*}^2 \\
&\quad + C_2(\|J^*\varphi^i\|, \|J^*Jx^i\|)\rho_{B^*}(\|\varphi^i - Jx^i\|_{B^*}) \\
&= -\frac{\alpha_i}{\|u^i\|_{B^*}}(f(x^i) - f^*) + 8\alpha_i^2 + C_2(\|\varphi^i\|_{B^*}, \|x^i\|)\rho_{B^*}(\alpha_i),
\end{aligned}$$

where convexity of $f(\cdot)$ and Lemma 2.6 (2) are used in the inequality. Combining the last relation with (4.1), we obtain

$$V(Jx^{i+1}, \bar{x}) \le V(Jx^i, \bar{x}) - 2\frac{\alpha_i}{\|u^i\|_{B^*}}(f(x^i) - f^*) + 16\alpha_i^2 + 2C_2(\|\varphi^i\|_{B^*}, \|x^i\|)\rho_{B^*}(\alpha_i). \tag{4.2}$$

Now the first three assertions of the theorem are established exactly the same way as in the proof of Theorem 3.1.

In the proof of the last two assertions of the theorem, we need to bound the quantity $\|x^{i+1} - x^i\|$. Here this has to be done differently from what we did in Theorem 3.1. In particular, we have

$$\begin{aligned}\|x^{i+1}-x^i\| &= \|\pi_C[\varphi^i]-\pi_C[Jx^i]\| \\ &\le 2Rg_B^{-1}\left(\frac{4L}{R}\|\varphi^i-Jx^i\|_{B^*}\right) \\ &= 2Rg_B^{-1}\left(\frac{4L}{R}\alpha_i\right),\end{aligned}$$

where the first equality follows from Lemma 2.7 (8), and the inequality follows from Lemma 2.7 (9).
The rest of the proof follows the pattern of Theorem 3.1. □

We note that if the sequence $\{x^i\}$ is assumed to be bounded (for example, this is always true in the case when the feasible region C is bounded or a certain level set of $f(\cdot)$ is bounded), then the stepsize rule (1.4) can be relaxed by eliminating the second condition. In particular, we have the following result.

Theorem 4.2. *Let B be a uniformly smooth and uniformly convex Banach space and $f: B\to\Re$ be a convex functional. Let $\{x^i\}$ be any sequence of iterates generated by* (1.8) *with the stepsizes $\{\alpha_i\}$ satisfying*

$$\sum_{i=0}^{\infty}\alpha_i=\infty,\qquad \lim_{i\to\infty}\alpha_i=0.$$

If the sequence $\{x^i\}$ is bounded then $\liminf_{i\to\infty} f(x^i)=f^*$ *and there exists $\bar{x}$, a weak accumulation point of $\{x^i\}$, such that $\bar{x}\in S$.*

Proof. We consider two possible cases. First, if it holds that $\sum_{i=0}^{\infty}\rho_{B^*}(\alpha_i)<\infty$ then the result follows from Theorem 4.1.
Let $\sum_{i=0}^{\infty}\rho_{B^*}(\alpha_i)=\infty$ and suppose that the result is not true. Then there exists $\tau>0$ such that $f(x^i)-f^*\ge\tau$ for all i. Let $\tilde{x}\in S$. By the analysis in Theorem 3.1 and Theorem 4.1, we have for any $j\ge 1$

$$\begin{aligned}D_4\tau\sum_{i=0}^{j}\alpha_i &\le D_4\sum_{i=0}^{j}\alpha_i(f(x^i)-f^*) \\ &\le V(Jx^0,\tilde{x})-V(Jx^{j+1},\tilde{x})+D_5\sum_{i=0}^{j}\rho_{B^*}(\alpha_i) \\ &\le D+D_5\sum_{i=0}^{j}\rho_{B^*}(\alpha_i)\end{aligned}$$

for some $D>0$, where boundedness of $\{x^i\}$ (and hence also of $\{V(Jx^i,\tilde{x})\}$) was used in the last inequality. It follows that

$$\sum_{i=0}^{j}\left(D_4\tau-D_5\alpha_i^{-1}\rho_{B^*}(\alpha_i)\right)\alpha_i\le D.$$

Since $\lim_{i\to\infty} \alpha_i^{-1}\rho_{B^*}(\alpha_i) = 0$ by uniform smoothness of B^*, we have $D_5\alpha_i^{-1}\rho_{B^*}(\alpha_i) \leq D_4\tau/2$ for i large enough, say $i \geq i_0$. Then, for $j > i_0$,

$$\frac{D_4\tau}{2}\sum_{i=i_0}^{j}\alpha_i \leq D.$$

Letting $j \to \infty$ in the latter relation gives a contradiction. Hence the hypothesis is invalid, and therefore $\liminf_{i\to\infty} f(x^i) = f^*$. □

The next result establishes some interesting properties of Cesàro averages of the sequence $\{x^i\}$. We recall that Cesàro averages $\{z^m\}$ of $\{x^i\}$ are defined by the following formula

$$z^m := (\sum_{i=0}^{m}\alpha_i)^{-1}\sum_{i=0}^{m}\alpha_i x^i.$$

Cesàro averages were studied in [17] for methods of solving operator equations.

Theorem 4.3. *Let B be a uniformly smooth and uniformly convex Banach space and $f : B \to \Re$ be a convex functional. Let $\{x^i\}$ be any sequence of iterates generated by* (1.8), (1.4) *and let $\{z^m\}$ be the sequence of its Cesàro averages. It follows that $\lim_{m\to\infty} f(z^m) = f^*$ and all weak accumulation points of $\{z^m\}$ belong to the solution set S. In particular, if $S = \{\bar{x}\}$ then $\{z^m\}$ converges weakly to $\bar{x}$.*

Proof. Let $\bar{x} \in S$ and $f^* = f(\bar{x})$. Denote $\lambda_i := f(x^i) - f^*$. By (4.2), we have that

$$D_4\alpha_i\lambda_i \leq V(Jx^i,\bar{x}) - V(Jx^{i+1},\bar{x}) + D_5\rho_{B^*}(\alpha_i),$$

where D_4 and D_5 are determined similarly to Theorem 3.1. Summing the latter relation from 0 to m, we obtain

$$\begin{aligned} D_4\sum_{i=0}^{m}\alpha_i\lambda_i &\leq \sum_{i=0}^{m}(V(Jx^i,\bar{x}) - V(Jx^{i+1},\bar{x})) + D_5\sum_{i=0}^{m}\rho_{B^*}(\alpha_i) \\ &= V(Jx^0,\bar{x}) - V(Jx^{m+1},\bar{x}) + D_5\sum_{i=0}^{m}\rho_{B^*}(\alpha_i). \end{aligned}$$

Let $s_m := (\sum_{i=0}^{m}\alpha_i)^{-1}$. Then

$$0 \leq D_4 s_m\sum_{i=0}^{m}\alpha_i\lambda_i \leq s_m(V(Jx^0,\bar{x}) - V(Jx^{m+1},\bar{x})) + D_5 s_m\sum_{i=0}^{m}\rho_{B^*}(\alpha_i).$$

By boundedness of $\{V(Jx^i,\bar{x})\}$ and (1.4), it follows that the right-hand-side, and hence also the left-hand-side, in the relation above tend to zero as $m \to \infty$. By the convexity of $f(\cdot)$, we further obtain

$$\begin{aligned} s_m\sum_{i=0}^{m}\alpha_i\lambda_i &= \frac{\sum_{i=0}^{m}\alpha_i(f(x^i) - f^*)}{\sum_{i=0}^{m}\alpha_i} \\ &= \frac{\sum_{i=0}^{m}\alpha_i f(x^i)}{\sum_{i=0}^{m}\alpha_i} - f^* \\ &\geq f(z^m) - f^*, \end{aligned}$$

where z^m are the Cesàro averages. It follows that $\lim_{m\to\infty} f(z^m) = f^*$. Furthermore, the sequence $\{z^m\}$ is bounded because so is $\{x^i\}$. Let z be any weak accumulation point of $\{z^m\}$ and $\{z^{m_k}\}$ be any subsequence which weakly converges to z. By weak lower semicontinuity of convex functionals, we have

$$\begin{aligned} f(z) &\le \liminf_{k\to\infty} f(z^{m_k}) \\ &= \lim_{m\to\infty} f(z^m) \\ &= f^*. \end{aligned}$$

It follows that the set of weak accumulation points of $\{z^m\}$ is contained in S. □

4.2. Method with metric projection

We next consider the interesting algorithm given by (1.9), (1.4). This method does make use of the metric projection, however the iterates generated need not be feasible (i.e. contained in the feasible set C).

Theorem 4.4. *Let B be a uniformly smooth and uniformly convex Banach space. Let $f : B \to \Re$ be a convex functional. For any sequence generated by* (1.9), (1.4) *all the conclusions of* Theorem 3.1 *are valid.*

Proof. First note that, by (1.9),

$$Jx^{i+1} - Jx^i = JJ^*y^i - Jx^i = y^i - Jx^i = -\alpha_i \frac{\bar{u}^i}{\|\bar{u}^i\|_{B^*}} - 2\alpha_i \frac{J(x^i - \bar{x}^i)}{\|x^i - \bar{x}^i\|},$$

and, by the triangular inequality,

$$\|Jx^{i+1} - Jx^i\|_{B^*} = \alpha_i \Big\| \frac{\bar{u}^i}{\|\bar{u}^i\|_{B^*}} + 2\frac{J(x^i - \bar{x}^i)}{\|x^i - \bar{x}^i\|} \Big\|_{B^*} \le 3\alpha_i.$$

Take any $\tilde{x} \in S$. By Lemma 2.7 (2),

$$\begin{aligned} V(Jx^{i+1}, \tilde{x}) &\le V(Jx^i, \tilde{x}) + \langle Jx^{i+1} - Jx^i, x^{i+1} - \tilde{x} \rangle \\ &= V(Jx^i, \tilde{x}) + \langle Jx^{i+1} - Jx^i, x^i - \bar{x}^i \rangle + \langle Jx^{i+1} - Jx^i, \bar{x}^i - \tilde{x} \rangle \\ &\quad + \langle Jx^{i+1} - Jx^i, x^{i+1} - x^i \rangle. \end{aligned} \tag{4.3}$$

For the second term in the latter relation, we have

$$\begin{aligned} \langle Jx^{i+1} - Jx^i, x^i - \bar{x}^i \rangle &= -\frac{\alpha_i}{\|\bar{u}^i\|_{B^*}} \langle \bar{u}^i, x^i - \bar{x}^i \rangle - 2\frac{\alpha_i}{\|x^i - \bar{x}^i\|} \langle J(x^i - \bar{x}^i), x^i - \bar{x}^i \rangle \\ &\le \frac{\alpha_i}{\|\bar{u}^i\|_{B^*}} \|x^i - \bar{x}^i\| \|\bar{u}^i\|_{B^*} - 2\frac{\alpha_i}{\|x^i - \bar{x}^i\|} \|x^i - \bar{x}^i\|^2 \\ &= -\alpha_i \|x^i - \bar{x}^i\| \le 0. \end{aligned} \tag{4.4}$$

For the third term in (4.3), we obtain

$$\begin{aligned} \langle Jx^{i+1} - Jx^i, \bar{x}^i - \tilde{x} \rangle &= -\frac{\alpha_i}{\|\bar{u}^i\|_{B^*}} \langle \bar{u}^i, \bar{x}^i - \tilde{x} \rangle - 2\frac{\alpha_i}{\|x^i - \bar{x}^i\|} \langle J(x^i - \bar{x}^i), \bar{x}^i - \tilde{x} \rangle \\ &\le -\frac{\alpha_i}{\|\bar{u}^i\|_{B^*}} (f(\bar{x}^i) - f^*), \end{aligned} \tag{4.5}$$

where the inequality follows from convexity of $f(\cdot)$ and the properties of the metric projection operator $P_C[\cdot]$ (in particular, $\bar{x}^i = P_C[x^i] = \arg\min_{x\in C} \|x^i - x\|^2$ implies that for all $x \in C$, $\langle J(x^i - \bar{x}^i), \bar{x}^i - x\rangle \geq 0$). Now consider the last term in (4.3). We have

$$\begin{aligned}\langle Jx^{i+1} - Jx^i, x^{i+1} - x^i\rangle &\leq 8\|Jx^{i+1} - Jx^i\|^2_{B^*} + C_2(\|x^{i+1}\|, \|x^i\|)\rho_{B^*}(\|Jx^{i+1} - Jx^i\|_{B^*}) \\ &\leq 72\alpha_i^2 + C_2(\|x^{i+1}\|, \|x^i\|)\rho_{B^*}(3\alpha_i), \end{aligned} \tag{4.6}$$

where the first inequality follows from Lemma 2.6 (2). Combining (4.3)–(4.6) yields

$$\begin{aligned} &V(Jx^{i+1}, \tilde{x}) \leq \\ &V(Jx^i, \tilde{x}) - \alpha_i\left(\frac{f(\bar{x}^i) - f^*}{\|\bar{u}^i\|_{B^*}} + \|x^i - \bar{x}^i\|\right) + 72\alpha_i^2 + C_2(\|x^{i+1}\|, \|x^i\|)\rho_{B^*}(3\alpha_i). \end{aligned} \tag{4.7}$$

Using the same arguments as in the proof of Theorem 3.1, we conclude that $\{x^i\}$ is bounded. Therefore, for some constant $D > 0$, $\|u^i\| \leq D$ and $\|\bar{u}^i\| \leq D$ for all i. By convexity of $f(\cdot)$, we have

$$\begin{aligned} f(x^i) - f^* &\leq f(\bar{x}^i) - f^* - \langle u^i, \bar{x}^i - x^i\rangle \\ &\leq f(\bar{x}^i) - f^* + \|u^i\|_{B^*}\|x^i - \bar{x}^i\| \\ &\leq f(\bar{x}^i) - f^* + D\|x^i - \bar{x}^i\|. \end{aligned}$$

Combining the latter relation with (4.7), and taking into account that $\|\bar{u}^i\| \leq D$, we obtain

$$V(Jx^{i+1}, \tilde{x}) \leq V(Jx^i, \tilde{x}) - \frac{\alpha_i}{D}(f(x^i) - f^*) + 72\alpha_i^2 + C_2(\|x^{i+1}\|, \|x^i\|)\rho_{B^*}(3\alpha_i).$$

The rest of the proof follows the pattern of Theorem 3.1. □

5. Concluding remarks

A unified framework for convergence analysis of subgradient projection methods for nonsmooth convex minimization in a Banach space was presented. The important distinctive features of our analysis are the following:

1. No strong or uniform monotonicity assumptions are made with respect to the subdifferential operator;
2. No boundedness assumptions are made with respect to the solution set of the problem or the level sets of the objective function;
3. No iterative regularization techniques are used to solve the problem.

Under very mild assumptions, we proved that the sequence of iterates is bounded and it has at least one weak accumulation point which is a minimizer. Moreover, all weak accumulation points of the sequence of Cesàro averages of the iterates are solutions of the minimization problem. Under certain additional assumptions (which are satisfied for several important instances of Banach spaces), we are able to exhibit weak convergence of the whole sequence of iterates to one of the solutions of the optimization problem.

Finally, we would like to state the following remaining open questions concerning the convergence properties of subgradient methods in a general Banach space.

1. Is it possible to prove weak convergence of the iterates (i.e. the uniqueness of a weak accumulation point) without making additional assumptions on the Banach space, like the ones we made in this paper (see parts 4 and 5 of Theorem 3.1) ?
2. Is it possible to prove weak convergence of the Cesàro averages in a general Banach space ?
3. Under which conditions (other than boundedness of $\{x^i\}$) the requirement that $\sum_{i=0}^{\infty} \rho_{B^*}(\alpha_i) < \infty$ can be relaxed ?

Acknowledgements. The first author thanks the Institute for Pure and Applied Mathematics (IMPA) at Rio de Janeiro, Brazil, where he was a visiting professor while this paper was written.

References

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

[2] Ya. I. Alber, A. N. Iusem, and M. V. Solodov: On the projected subgradient methods for nonsmooth convex optimization in a Hilbert space, 1996. To be published.

[3] Ya. I. Alber, A. Kartsatos, and E. Litsyn: Iterative solution of unstable variational inequalities on approximately given sets, Abstract and Applied Analysis, 1(1) (1996) 45–65.

[4] Ya. I. Alber, A. I Notik: Geometric properties of Banach spaces and approximate methods for solving nonlinear operator equations, Soviet Math. Dokl., 29 (1984) 611–615.

[5] Ya. I. Alber, A. I Notik: On minimization of functionals and solution of variational inequalities in Banach spaces, Soviet Math. Dokl., 34(2) (1987) 296–300.

[6] Ya. I. Alber, A. I Notik: Parallelogram inequalities in Banach spaces and some properties of the duality mapping, Ukranian Math. Journal, 40 (1988) 650–652.

[7] Ya. I. Alber, A. I Notik: On the iterative method for variational inequalities with nonsmooth unbounded operators in Banach space, Journal of Mathematical Analysis and Applications, 188(3) (1994) 928–939.

[8] A. A. Auslender: Optimisation Méthodes Numériques, Masson, Paris, 1976.

[9] R. Correa, C. Lemaréchal: Convergence of some algorithms for convex minimization, Mathematical Programming, 62 (1993) 261–275.

[10] J. Diestel: The geometry of Banach spaces, Lecture Notes in Mathematics 485, Springer-Verlag, Berlin, 1975.

[11] T. Figiel: On the muduli of convexity and smoothness, Studia Mathematica, 56 (1976) 121–155.

[12] J. R. Giles: Convex analysis with applications to differentiation of convex functions, Research notes in mathematics 58, Pitman, Boston, 1982.

[13] J. Lindenstrauss, L. Tzafriri: Classical Banach spaces II, Springer-Verlag, Berlin, 1979.

[14] V. D. Milman: Geometric theory of Banach spaces. Part II: Geometry of the unit sphere, Russian Mathematics Surveys, 26(6) (1971) 79–163.

[15] B. T. Polyak: A general method of solving extremum problems, Soviet Math. Dokl, 8 (1967) 593–597.

[16] B. T. Polyak: Introduction to Optimization, Optimization Software, Inc., Publications Division, New York, 1987.

[17] I. P. Ryazantseva: Convergence of mean iterates for nonlinear monotone equations in Banach space, Vestnik Moscovskogo Universiteta, Series XV, 4 (1987) 61–63. (In Russian).

HIER :

Leere Seite
256