MATHEMATICAL COMMUNICATIONS
Math. Commun. **20**(2015), 161–173

Convergence of the steepest descent method with line searches and uniformly convex objective in reflexive Banach spaces*

Fernando Andrés Gallego[1], John Jairo Quintero[2] and Juan Carlos Riano[2,†]

[1] *Instituto de Matematicas, Universidade Federal do Rio de Janeiro, 21 945–970 Rio de Janeiro, Brazil*
[2] *PCM Computational Applications, Universidad Nacional de Colombia, Cra 27 # 64–60 Manizales, Colombia*

Received July 12, 2015; accepted August 4, 2015

Abstract. In this paper, we present some algorithms for unconstrained convex optimization problems. The development and analysis of these methods is carried out in a Banach space setting. We begin by introducing a general framework for achieving global convergence without Lipschitz conditions on the gradient, as usual in the current literature. This paper is an extension to Banach spaces to the analysis of the steepest descent method for convex optimization, most of them in less general spaces.

AMS subject classifications: 90C25, 49M29, 46N10, 46N40

Key words: uniformly convex functional, descent methods, step-size estimation, metric of gradient

1. Introduction

Optimization or minimization problems arise in many areas of modern science and technology. As examples, let us mention control theory, image processing and segmentation, drag reduction, shape optimization, etc. Here we are particularly interested in the global convergence of some optimization methods formulated in infinite dimensional spaces.

Optimization methods are iterative algorithms for finding (global or local) solutions of minimization problems. Usually, we are already satisfied if the method can be proved to converge to stationary points. These points may than satisfy the first order necessary optimal conditions.

The steepest descent method is one of the oldest and most widely used approaches in the literature when solving smooth optimization problems. However, classical results for an arbitrary objective function are not so robust because the convergence of the method depends on gradient-like Lipschitz conditions, see [14, 17] (finite

*The work of the first author was supported by DIMA-UNAL Colombia under contract No. ODS-58/2012.
†Corresponding author. *Email addresses:* fgallego@ufrj.br (F. A. Gallego), jjquintero@unal.edu.co (J. J. Quintero), jcrianoro@unal.edu.co (J. C. Riano)

dimension) and [4, 5, 12, 13, 15] (infinite dimension). We will analyze the convergence of these methods with a stepsize rule for a uniformly convex functional (see [8, 9, 18, 19]). In fact, not only the Lipschitz constant is not required, but it is also possible to prove the convergence of the full sequence under the only assumption of the existence of minimizers (besides convexity of the objective in Hilbert spaces, and uniform convexity in Banach spaces, as in this case). In the nonconvex case, even for a Lipschitz continuous objective, it is necessary to assume the existence of cluster points of the generated sequence, and the only attainable result is their stationarity, but not the convergence of the full sequence (even assuming the existence of the cluster points).

In Hilbert spaces, results of this type can be found in [10]. The convergence of gradient-like methods in Hilbert spaces in the convex but nonsmooth case, where the direction is taken as the opposite to a subgradient, rather than minus the gradient can be found in [1]; hence, no line search is possible because subgradient directions may fail to be descent ones, and therefore the steplengths must be given exogenously. The extension of results in the previous reference to reflexive Banach spaces was achieved in [2, 3]. The last three papers differ from this document, since they demand just convexity of the objective function instead of uniform convexity; nevertheless, they use exogenously given steplengths instead of line searches, due to the nonsmoothness of the objective function.

We refer to [11] for a precise historical background and motivation about the metrics of a functional gradient. For the historical convergence theorem of steepest descent algorithm in normed linear spaces, see also [7].

We will start with a common classical situation in which we want to minimize a Fréchet differentiable functional f defined on a reflexive Banach space. The assumption of uniform convexity guarantees the existence of a global minimizer and a global convergence towards it, allowing us to be concentrated on the central issues related to the infinite-dimensional setting. The minimization strategy will be based on a general method of the steepest descent type coupled with a line search. The advantage of this procedure consists of avoiding regularity assumptions such as the Lipschitz continuity of the first derivation. After making suitable assumptions regarding the well-posedness of the minimization problem, a steepest descent algorithm in the Banach space is built proving its convergence.

This paper is organized as follows. In Section 2, we start with a fairly general framework of convex optimization in Banach spaces, introducing some step-size rules. In Section 3, we present some assumptions regarding the functional to be minimized. Under these weak assumptions, we formulate an abstract algorithm in Section 4, introducing the notion of Riesz map and proving its convergence.

2. Convex optimization in Banach space

Let E be a reflexive Banach space and E' its dual. $\langle \cdot, \cdot \rangle$ will denote the duality coupling in $E \times E'$ and $E' \times E''$, respectively, i.e.; $\langle x, x' \rangle = x'(x)$, $\langle x', x'' \rangle = x''(x')$, for all $x \in E, x' \in E, x'' \in E''$.
Let $f : E \to R$ be a Fréchet differentiable functional. We consider the following

minimization problem: Find $u \in E$ such that

$$f(u) = \min_{v \in E} f(v). \tag{1}$$

The line search method for solveing (1) generates the following iteration:

$$u_{k+1} = u_k + \alpha_k s_k, \tag{2}$$

where $u_k \in E$ is the current iterative point, $s_k \in E$ is a search direction, and α_k is a positive step-size.

Definition 1. *Given $v \in E$, we call $s \in E$ an admissible descent direction for v if $\|s\| = 1$ and $\langle f'(v), s\rangle < 0$*

Once an admissible descent direction is determined, one has to find the minimum of the functional f along the search direction. In order to ensure that the line search in fact yields a smaller value of the functional than the initial one, we identify a possible range of step-sizes that guarantees this descent.

There are a lot of rules for choosing step-size α_k (see [12],[13],[14]), in this paper we take into account the following considerations:

1. *Minimization Rule*: At each iteration, α_k is selected so that

$$\phi_k(\alpha_k) = \min_{\alpha > 0} \phi_k(\alpha), \tag{3}$$

where $\phi(\alpha) = f(u_k + \alpha s_k)$.

2. *Approximate minimization rule*: At each iteration, α_k is selected so that

$$\alpha_k = \min\left\{\alpha : \langle f'\left(u_k + \alpha s_k\right), s_k\rangle = 0, \alpha > 0\right\}. \tag{4}$$

3. *Goldstein rule*: A fixed scalar $\sigma \in (0, \frac{1}{2})$ is selected, and α_k is chosen to satisfy

$$\sigma \leq \frac{f\left(u_k + \alpha_k s_k\right) - f(u_k)}{\alpha_k \left\langle f'(u_k), s_k\right\rangle} \leq 1 - \sigma. \tag{5}$$

4. *Wolfe rule*: α_k is chosen to satisfy simultaneously

$$f(u_k) - f(u_k + \alpha_k s_k) \geq -\sigma \alpha_k \langle f'(u_k), s_k\rangle \tag{6}$$

$$\langle f'(u_k + \alpha_k s_k), s_k\rangle \geq \beta \langle f'(u_k), s_k\rangle, \tag{7}$$

where σ and β are some scalars with $\sigma \in (0, \frac{1}{2})$ and $\beta \in (\sigma, 1)$.

5. *Strong Wolfe rule*: α_k is chosen to satisfy (6) and

$$|\langle f'(u_k + \alpha_k s_k), s_k\rangle| \leq -\beta \langle f'(u_k), s_k\rangle. \tag{8}$$

6. *Armijo rule*: Let us define scalars as follows: $\lambda_k = -\frac{\langle f'(u_k), s_k\rangle}{\|s_k\|^2}$, $\beta \in (0, 1)$, and $\sigma \in (0, \frac{1}{2})$. Then we define $\alpha_k = \beta^{m_k} \lambda_k$, where

$$m_k = \min_{m > 0}\left\{f(u_k) - f(u_k + \beta^m \lambda_k s_k) \geq -\sigma \beta^m \lambda_k \left\langle f'(u_k), s_k\right\rangle\right\}. \tag{9}$$

3. General assumptions

We start by collecting general definitions and assumptions on the functional f that will be held throughout the paper.

Definition 2 (see [18])**.** *Let $f : E \to \mathbb{R} \cup \{\infty\}$ be an l.s.c, convex function whose domain is not a singleton. f is uniformly convex at $x \in domf$ if there exists $\delta : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\}$ with $\delta(t) > 0$ for $t > 0$ such that*

$$f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)-\lambda(1-\lambda)\delta(\|y-x\|) \qquad \forall y \in domf, \lambda \in (0,1)$$

Assumption 1. *The functional f satisfies:*

1. *The Fréchet-derivative $f' : E \to E'$ is uniformly continuous on each bounded subset of E.*
2. *f is a uniformly convex functional (or p-elliptic functional), i.e., there exist $C > 0$ and $p > 1$ such that*

$$\|w - v\|_E^p \leq C\langle f'(w) - f'(v), w - v\rangle \tag{10}$$

for all $w, v \in E$.

The second item of Assumption 1 is equivalent to Definition 2 (see [18]). In the case of $E = \mathbb{R}^n$, the function $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle a, x\rangle$, $A = A^T$ satisfies the second item of Assumption 1 (with $p = 2$) if and only if the matrix A is positively defined. For the two-dimensional case, the surface representing f has the form of a paraboloid whose horizontal sections are ellipses. This explains the elliptic functional terminology, (i.e., a 2-elliptic functionals case $E = \mathbb{R}^n$). It forms a natural generalization of quadratic functionals with a symmetric, positively defined matrix (see [8]).

We recall some consequences of the above assumption:

Lemma 1. *Let Assumption 1 be satisfied. Then, the following statements hold.*

1. *For all $w, u \in E$ we have*

$$f(v) - f(u) \geq \langle f'(u), v - u\rangle + \frac{1}{Cp}\|u - v\|_E^p, \quad C > 0, p > 1. \tag{11}$$

2. *f is bounded from below and strictly convex.*
3. *Let $u_0 \in E$ be arbitrary. The set $L_{u_0}(f) := \{v \in E : f(v) \leq f(u_0)\}$ is bounded.*

The proof can be found in the appendix or see also [18].

Remark 1. *In particular, from Lemma 1 we see that our assumptions are slightly stronger than convexity, so that under Assumption 1 there exists a unique solution for the minimization problem (1).*

4. Convergent steepest descent algorithm

The gradient notion of a functional was given by Golomb and Tapia [11] as a rigorous generalization of arbitrary normed linear spaces as follows: Given a normed linear space E with dual E' and second dual E'', ζ denotes the natural or canonical embedding of E into E'', i.e ($\zeta(x)(y) = y(x), x \in E, y \in E'$). For all $x \in E$, let

$$J(x) := \{v \in E' : \|v\| = \|x\|, \langle v, x\rangle = \|x\|^2\}. \tag{12}$$

The set-valued map $J : E \to 2^{E'}$ is called the duality map of E. We denote the duality map of E' by J'.

Definition 3. *If $f : E \to R$ is Fréchet differentiable at $x \in E$, then by the metric gradient of f at x we mean*

$$\bigtriangledown f(x) = \zeta^{-1} J' f'(x) \subseteq E. \tag{13}$$

We call

$$\bigtriangledown f = \zeta^{-1} J' f' : E \to 2^E \tag{14}$$

the metric gradient of f.

Note that unlike the derivative, the metric gradient depends on the norm, not just on the topology of the space. The definition of the gradient of a functional defined on a normed linear space which in the case of Hilbert space reduces to

$$f'(x)(\eta) = (\bigtriangledown f(x), \eta), \tag{15}$$

where $(\cdot,\cdot)$ is an inner product. When E is not a Hilbert space the above relationship is no longer obvious. The main distinction between the metric gradient and the derivative is that $\bigtriangledown f(x) \in E$, while $f'(x) \in E'$.

We consider a special operator that, in the case of a Hilbert space, coincides with the Riesz representation; the name of this operator differs from author to author.

Definition 4. *Let E be a normed space and $R : E' \to 2^E$ the point-to-set mapping defined by*

$$R(v) := \{x \in E : \|x\|_E = 1 \quad \langle v, x\rangle = \|v\|_{E'}\}, \qquad \forall\, v \in E'.$$

The set-valued mapping R is called the Riesz map.

Note that if $x \in R(f)$, for $f \in E'$, then $\langle f, x\rangle = \sup\limits_{v \in E} \frac{\langle f,v\rangle}{\|v\|_E}$.

There exists a relation between the Riesz map and the metric gradient of f; in fact, is easy to see that

$$R(f'(u)) = \frac{\bigtriangledown f(u)}{\|f'(u)\|_{E'}} \quad \forall u \in E. \tag{16}$$

A relevant remark is about the existence of the Riesz map:

Proposition 1. *If E is a reflexive Banach space, then the Riesz map $R(f'(u_k)) \neq \phi$ for all k.*

Proof. By proposition 1 in [11], we know that the metric gradient of f is not an empty subset in E. By (16), we can conclude. □

We are now ready to introduce the steepest descent algorithm for solving the optimization problem (1).

Algorithm 1. *Let $u_0 \in E$ be given. Then, for $k = 0, 1, 2, ...,$ while $f'(u_k) \neq 0$,*

1. *choose the search direction as $-s_k \in R(f'(u_k))$,*
2. *determine the step-size α_k defined by the line search rule,*
3. *let $u_{k+1} := u_k + \alpha_k s_k$.*

Remark 2. *Note that $s_k \in -R(f'(u_k))$ is an admissible descent direction for u_k, because $\|s_k\| = 1$ and $\langle f'(u_k), s_k\rangle = -\|f'(u_k)\|_{E'} < 0$.*

In what follows, E is a Reflexive Banach space and u is the minimizer in (1). We start taking into account two results that help us to prove some properties of convergence.

Lemma 2. *Let $\{u_k\}$ be an infinite sequence generated by algorithm 1. If $\{f(u_k)\}$ is a decreasing sequence, $\{u_k\}$ is a Cauchy sequence and $\langle f'(u_{k+1}), s_k\rangle \geq 0$, then $\{u_k\}$ converges strongly for u in E.*

Proof. By definition 4, we have

$$\begin{aligned} \|f'(u_k)\|_{E'} &= -\langle f'(u_k), s_k\rangle = \langle f'(u_{k+1}) - f'(u_k), s_k\rangle - \langle f'(u_{k+1}), s_k\rangle \\ &\leq \|f'(u_{k+1}) - f'(u_k)\|_{E'}\|s_k\|_E - \langle f'(u_{k+1}), s_k\rangle \\ &\leq \|f'(u_{k+1}) - f'(u_k)\|_{E'}. \end{aligned} \tag{17}$$

As $\{f(u_k)\}$ is a decreasing sequence, $\{u_k\} \in L_{u_0}(f)$, thus

$$\|f'(u_{k+1}) - f'(u_k)\|_{E'} \to 0, \quad (k \to \infty)$$

since $L_{u_0}(f)$ is bounded by Lemma 1, $\{u_k\}$ is a Cauchy sequence and f' is uniformly continuous in a subset bounded of E by 1 of Assumption 1. By (17),

$$\|f'(u_k)\|_{E'} \to 0, \quad (k \to \infty).$$

Now using the second item of Assumption 1 and the fact that $f'(u) = 0$, we have

$$\|u_k - u\|_E^p \leq C\,\langle f'(u_k), u_k - u\rangle \leq C\|f'(u_k)\|_{E'}\|u_k - u\|_E$$

so $\|u_k - u\|_E \leq (C\|f'(u_k)\|_{E'})^{\frac{1}{p-1}}$ with $p > 1$. Therefore,

$$\lim_{k\to\infty} \|u_k - u\|_E = 0.$$

□

Lemma 3. *If there exist* $\sigma \in (0,1)$ *such that* $f(u_k) - f(u_{k+1}) \geq \sigma \alpha_k \|f'(u_k)\|_{E'}$, *then* $\{u_k\}$ *is a Cauchy sequence.*

Proof. By Lemma 1 and Definition 4, we have

$$\begin{aligned}
\frac{1}{Cp}\|u_{k+1}-u_k\|_E^p &\leq f(u_{k+1}) - f(u_k) - \langle f'(u_k), u_{k+1}-u_k\rangle \\
&= f(u_{k+1}) - f(u_k) - \alpha_k \langle f'(u_k), s_k\rangle \\
&= f(u_{k+1}) - f(u_k) + \alpha_k \|f'(u_k)\|_{E'} \\
&\leq \frac{1}{\sigma}\{f(u_k) - f(u_{k+1})\} - \{f(u_k) - f(u_{k+1})\} \\
&= \left(\frac{1}{\sigma} - 1\right)\{f(u_k) - f(u_{k+1})\}. \qquad (18)
\end{aligned}$$

Since $\{f(u_k)\}$ is a decreasing sequence and bounded below, as $\{f(u_k)\}$ is a Cauchy sequence, then by (18), $\{u_k\}$ is a Cauchy sequence. □

Theorem 1. *If we choose the step-size* α_k *by the Minimization Rule, then Algorithm* 1 *generates an infinite sequence* $\{u_k\}$ *that converges strongly to* u *in* E.

Proof. Let $\phi_k : R_+ \to R$ be defined by $\phi_k(\alpha) = f(u_k + \alpha s_k)$. By the Minimization rule, there exists α_k minimizer of ϕ_k. On the other hand, ϕ_k is a differential function, since f is Fréchet differentiable, with $\phi_k'(\alpha) = \langle f'(u_k + \alpha s_k), s_k\rangle$. Hence,

$$\langle f'(u_{k+1}), s_k\rangle = \langle f'(u_k + \alpha_k s_k), s_k\rangle = \phi_k'(\alpha_k) = 0. \qquad (19)$$

Now using Lemma 1 and (19), we have

$$\begin{aligned}
\frac{1}{Cp}\|u_{k+1}-u_k\|_E^p &\leq f(u_k) - f(u_{k+1}) - \langle f'(u_{k+1}), u_k - u_{k+1}\rangle \\
&= f(u_k) - f(u_{k+1}) + \langle f'(u_{k+1}), \alpha_k s_k\rangle \\
&= f(u_k) - f(u_{k+1}) + \alpha_k \langle f'(u_{k+1}), s_k\rangle \\
&= f(u_k) - f(u_{k+1}). \qquad (20)
\end{aligned}$$

By construction, $\{f(u_k)\}$ is a decreasing sequence and bounded below; thus $\{f(u_k)\}$ is a Cauchy sequence. Then by (20), $\{u_k\}$ is a Cauchy sequence. Now invoking Lemma 2, we have that $\{u_k\}$ converges strongly for u in E. □

Theorem 2. *If we choose the step-size* α_k *by the Approximate minimization rule, then Algorithm* 1 *generates an infinite sequence* $\{u_k\}$ *that converges strongly to* u *in* E.

Proof. By the Approximate minimization rule, α_k is chosen such that

$$\langle f'(u_{k+1}), s_k\rangle = \langle f'(u_k + \alpha_k s_k), s_k\rangle = 0.$$

Analogically, (20) implies that $\{u_k\}$ is a Cauchy sequence. Note that α_k is a critical point of $\phi_k(\alpha) = f(u_k + \alpha s_k)$, which is a convex function. Then α_k is a minimizer of ϕ_k, thus by construction the same as the minimization rule, $\{f(u_k)\}$ is a decreasing sequence. Now using Lemma 2, we have that $\{u_k\}$ converges for u in E. □

Theorem 3. *If we choose the step-size α_k by the Goldstein rule, then Algorithm* 1 *generates an infinite sequence $\{u_k\}$ that converges strongly to u in E.*

Proof. By the Goldstein rule, we have

$$\sigma \leq \frac{f(u_{k+1}) - f(u_k)}{\alpha_k(-\|f'(u_k)\|_{E'})} \leq 1-\sigma \Leftrightarrow \sigma \leq \frac{f(u_k) - f(u_{k+1})}{\alpha_k\|f'(u_k)\|_{E'}} \leq 1-\sigma. \quad (21)$$

In particular, $f(u_k) - f(u_{k+1}) \geq \sigma\alpha_k \|f'(u_k)\|_{E'}$, with $\sigma \in (0, \frac{1}{2})$, by Lemma 3, $\{u_k\}$ is a Cauchy sequence; hence,

$$\alpha_k = \alpha_k \|s_k\|_E = \|u_{k+1} - u_k\|_E \to 0 \quad (k \to \infty). \quad (22)$$

On the other hand, using the mean value theorem, there exists $c_k \in (0, \alpha_k)$, such that

$$f(u_{k+1}) - f(u_k) = \langle f'(u_k + c_k s_k), \alpha_k s_k \rangle$$

by (21),

$$-\langle f'(u_k + c_k s_k), s_k \rangle = \frac{f(u_k) - f(u_{k+1})}{\alpha_k} \leq (1-\sigma)\|f'(u_k)\|_{E'}. \quad (23)$$

By Definition 4 and (23), we have

$$\begin{aligned} \|f'(u_k)\|_{E'} &= -\langle f'(u_k), s_k\rangle \\ &= \langle f'(u_k + c_k s_k) - f'(u_k), s_k \rangle - \langle f'(u_k + c_k s_k), s_k\rangle \\ &\leq \|f'(u_k + c_k s_k) - f'(u_k)\|_{E'} + (1-\sigma)\|f'(u_k)\|_{E'}. \end{aligned}$$

Therefore,

$$\|f'(u_k)\|_{E'} \leq \frac{1}{\sigma}\|f'(u_k + c_k s_k) - f'(u_k)\|_{E'}. \quad (24)$$

By construction, $\{f(u_k)\}$ is a decreasing sequence, thus $\{u_k\}$ lies in $L_{u_0}(f)$, i.e,

$$f(u_k + \alpha_k s_k) = f(u_{k+1}) \leq f(u_0), \quad (25)$$

Let $\delta > 0$. We claim that

$$f(u_k + c_k s_k) \leq f(u_0) + \delta \quad (26)$$

for k large enough. In fact, by (22), there exist $N = N(\delta) > 0$ such that

$$|\alpha_k - c_k| < \delta \quad \forall k \geq N.$$

Hence

$$\|u_{k+1} - [u_k + c_k s_k]\|_E < \delta \quad \forall k \geq N,$$

by contininuity of f and (25), we obtain (26). Thus $\{u_k + c_k s_k\}$ lies in $L_{u_0+\delta}(f) := \{v \in E : f(v) \leq f(u_0) + \delta\}$. As the proof of 3 in Lemma 1 (see Appendix), we can see that it is a bounded set in E. Now by (22), we have that $c_k \to 0$ as $k \to \infty$. Then

$$\|f'(u_k + c_k s_k) - f'(u_k)\|_{E'} \to 0 \quad (k \to \infty)$$

since f' is uniformly continuous in a bounded subset of E. By (24), $\|f'(u_k)\|_{E'} \to 0$ as $k \to \infty$. Invoking the second item of Assumption 1 and the fact that $f'(u) = 0$, we have

$$\|u_k - u\|_E^p \leq C \langle f'(u_k), u_k - u\rangle \leq C\|f'(u_k)\|_{E'}\|u_k - u\|_E.$$

Therefore,

$$\|u_k - u\|_{E'} \leq (C\|f'(u_k)\|_{E'})^{\frac{1}{p-1}} \to 0 \quad (k \to \infty), \qquad p > 1.$$

□

Theorem 4. *If we choose the step-size α_k by the Wolfe rule, then Algorithm* 1 *generates an infinite sequence $\{u_k\}$ that converges strongly to u in E.*

Proof. The Wolfe rule takes the form:

$$f(u_k) - f(u_{k+1}) \geq -\sigma\alpha_k \langle f'(u_k), s_k\rangle = \sigma\alpha_k\|f'(u_k)\|_{E'} \tag{27}$$

and

$$-\langle f'(u_{k+1}), s_k\rangle \leq -\beta \langle f'(u_k), s_k\rangle = \beta\|f'(u_k)\|_{E'}, \tag{28}$$

where $\sigma \in (0, \frac{1}{2}), \beta \in (\sigma, 1)$. By Definition 4 and (28), we have

$$\begin{aligned}\|f'(u_k)\|_{E'} &= -\langle f'(u_k), s_k\rangle = \langle f'(u_{k+1}) - f'(u_k), s_k\rangle - \langle f'(u_{k+1}), s_k\rangle \\ &\leq \|f'(u_{k+1}) - f'(u_k)\|_{E'}\|s_k\|_E + \beta\|f'(u_k)\|_{E'}\end{aligned}$$

Hence,

$$\|f'(u_k)\|_{E'} \leq \frac{1}{1-\beta}\|f'(u_{k+1}) - f'(u_k)\|_{E'}. \tag{29}$$

By construction, $\{f(u_k)\}$ is a decreasing sequence, thus $\{u_k\} \in L_{u_0}(f)$ and by (27) and Lemma 3, $\{u_k\}$ is a Cauchy sequence, thus

$$\|f'(u_{k+1}) - f'(u_k)\|_{E'} \to 0 \quad (k \to \infty)$$

since $L_{u_0}(f)$ is bounded by Lemma 1 and f' is uniformly continuous in the subset bounded of E by the first item of Assumption 1. Thus, from (29) we get

$$\|f'(u_k)\|_{E'} \to 0 \quad (k \to \infty).$$

Now using Assumption 1 and the fact that $f'(u) = 0$, we have

$$\|u_k - u\|_E^p \leq C \langle f'(u_k), u_k - u\rangle \leq C\|f'(u_k)\|_{E'}\|u_k - u\|_E,$$

so $\|u_k - u\|_E \leq (C\|f'(u_k)\|_{E'})^{\frac{1}{p-1}}$ with $p > 1$. Therefore,

$$\lim_{k\to\infty} \|u_k - u\|_E = 0.$$

□

Theorem 5. *If we choose the step-size α_k by the Strong Wolfe rule, then Algorithm* 1 *generates an infinite sequence $\{u_k\}$ that converges strongly to u in E.*

Proof. The Strong Wolfe rule takes the form (27) and

$$|\langle f'(u_{k+1}), s_k\rangle| \leq -\beta \langle f'(u_k), s_k\rangle = \beta \|f'(u_k)\|_{E'}. \tag{30}$$

Note that by Lemma 3, $\{u_k\}$ is a Cauchy sequence. If $\langle f'(u_{k+1}), s_k\rangle \geq 0$, we can invoke Lemma 2. If $\langle f'(u_{k+1}), s_k\rangle < 0$, (30) implies that

$$\beta \|f'(u_k)\|_{E'} \geq -\langle f'(u_{k+1}), s_k\rangle \geq -\beta \|f'(u_k)\|_{E'}, \tag{31}$$

where $\sigma \in (0, \frac{1}{2}), \beta \in (\sigma, 1)$, which is the case of Theorem 4. □

The convergence of the descent method with the Armijo rule needs some extra requirement about the regularity of f':

Theorem 6. *If we choose the step-size α_k by the Armijo rule and suppose that f' is uniformly continuous in E, then Algorithm* 1 *generates an infinite sequence $\{u_k\}$ that converges strongly to u in E.*

Proof. First, note that the scalars generated by the Armijo rule take the form:

$$\begin{aligned}
\lambda_k &= -\langle f'(u_k), s_k\rangle = \|f'(u_k)\|_{E'} \\
\alpha_k &= \beta^{m_k} \lambda_k = \beta^{m_k} \|f'(u_k)\|_{E'}, \quad \beta \in (0,1) \\
m_k &= \min\left\{m \in Z_+ : f(u_k) - f(u_k + \beta^m \lambda_k s_k) \geq \sigma \beta^m \lambda_k^2\right\}, \quad \sigma \in (0, \frac{1}{2}).
\end{aligned} \tag{32}$$

Thus, (32) implies either

$$\alpha_k = \|f'(u_k)\|_{E'} \qquad if \quad m_k = 0 \tag{33}$$

or

$$f(u_k) - f\left(u_k + \frac{\alpha_k}{\beta} s_k\right) < \sigma \frac{\alpha_k}{\beta} \|f'(u_k)\|_{E'}. \tag{34}$$

Now by (34) and the mean value theorem, there exist $c_k \in (0, \alpha_k)$ such that

$$\left\langle f'\left(u_k + \frac{c_k}{\beta} s_k\right), \frac{\alpha_k}{\beta} s_k \right\rangle = f\left(u_k + \frac{\alpha_k}{\beta} s_k\right) - f(u_k).$$

Thus

$$-\left\langle f'\left(u_k + \frac{c_k}{\beta} s_k\right), s_k \right\rangle < \sigma \|f'(u_k)\| \tag{35}$$

by Definition 4 and (35), we have that

$$\begin{aligned}
\|f'(u_k)\|_{E'} &= -\langle f'(u_k), s_k\rangle \\
&= \left\langle f'\left(u_k + \frac{c_k}{\beta} s_k\right) - f'(u_k), s_k \right\rangle - \left\langle f'\left(u_k + \frac{c_k}{\beta} s_k\right), s_k \right\rangle \\
&\leq \left\| f'\left(u_k + \frac{c_k}{\beta} s_k\right) - f'(u_k) \right\|_{E'} + \sigma \|f'(u_k)\|_{E'}.
\end{aligned}$$

Thus

$$\|f'(u_k)\|_{E'} \leq \frac{1}{1-\sigma}\left\|f'\left(u_k + \frac{c_k}{\beta}s_k\right) - f'(u_k)\right\|_{E'} \qquad \sigma \in (0, \frac{1}{2}). \tag{36}$$

On the other hand, (32) implies that $f(u_k) - f(u_{k+1}) \geq \sigma\alpha_k\|f'(u_k)\|_{E'}$, by Lemma 3, we have that $\{u_k\}$ is a Cauchy sequence; i.e.,

$$\alpha_k = \alpha_k\|s_k\|_E = \|u_{k+1} - u_k\|_E \to 0 \quad (k \to \infty).$$

Hence, $\lim_{k\to\infty} c_k = 0$ since $c_k \in (0, \alpha_k)$. Then by (33) or (34) – (36) and the fact that f' is uniformly continuous in E, we have that

$$\|f'(u_k)\|_{E'} \to 0 \quad (k \to \infty).$$

Now using the second item of Assumption 1 and the fact that $f'(u) = 0$, we have

$$\|u_k - u\|_E^p \leq C\langle f'(u_k), u_k - u\rangle \leq C\|f'(u_k)\|_{E'}\|u_k - u\|_E$$

so $\|u_k - u\|_E \leq (C\|f'(u_k)\|_{E'})^{\frac{1}{p-1}}$ with $p > 1$. Hence,

$$\lim_{k\to\infty} \|u_k - u\|_E = 0.$$

□

5. Conclusions

To conclude, this result does not give any information about a nonrelfexive case, so we think that a full understanding of convergence properties for more general Banach spaces is relevant.

6. Appendix

Lemma 1. 1. By the fundamental theorem of Calculus and the second item in Assumption 1, we have:

$$\begin{aligned} f(v) - f(u) &= \int_0^1 \frac{d}{dt}f(u + t(v-u))dt = \int_0^1 \langle f'(u + t(v-u)), v - u\rangle\, dt \\ &= \langle f'(u), v-u\rangle + \int_0^1 \langle f'(u + t(v-u)) - f'(u), v-u\rangle\, dt \\ &\geq \langle f'(u), v-u\rangle + \frac{1}{C}\int_0^1 \frac{1}{t}\,\|t(u-v)\|_E^p\, dt \\ &= \langle f'(u), v-u\rangle + \frac{1}{C}\int_0^1 t^{p-1}\,\|u-v\|_E^p\, dt; \end{aligned}$$

hence

$$f(v) - f(u) \leq \langle f'(u), v-u\rangle + \frac{1}{Cp}\,\|u-v\|_E^p\,. \tag{37}$$

2. By the above item, we have that

$$f(v) - f(u) \geq \langle f'(u), v - u\rangle + \frac{1}{Cp}\|v - u\|_E^p > \langle f'(u), v - u\rangle$$

for all $v \neq u \in E$;then f is strictly convex. For boundedness, take $u = 0$. Then

$$f(v) \geq f(0) + \langle f'(0), v\rangle + \frac{1}{Cp}\|v\|_E^p \geq f(0) - \|f'(0)\|_{E'}\|v\|_E + \frac{1}{Cp}\|v\|_E^p.$$

Thus the functional f is coercive. Then there exists a unique solution of (1) (see Corollary 3.23 in [6]).

3. Take $u = 0$ in (37), and $v \in L_{u_0}$. Then

$$\frac{1}{Cp}\|v\|_E^p \leq f(v) - f(0) - \langle f'(0), v\rangle \leq |f(u_0)| + |f(0)| + \|f'(0)\|_{E'}\|v\|_E.$$

If $\|v\|_E > 1$, then, since $p > 1$, we have that

$$\|v\| \leq (Cp\,(|f(u_0)| + |f(0)| + \|f'(0)\|_E))^{\frac{1}{p-1}} = M.$$

Hence, $\|v\|_E \leq M$ for all $v \in L_{u_0}$.

□

Acknowledgement

The authors gratefully acknowledge financial support of the Dirección Nacional de Investigaciones of the Universidad Nacional de Colombia during the course of this research under project No. 15723 "Optimización con el Método de Búsqueda en Línea para Funciones con Dominio Convexo en Espacios de Banach".

References

[1] Ya. I. Alber, A. N. Iusem, M. V. Solodov, *On the projected subgradient method for nonsmooth convex optimization in Hilbert spaces*, Math. Programming **81**(1998), 23–37.

[2] Ya. I. Alber, A. N. Iusem, M. V. Solodov. *Minimization of nonsmooth convex functionals in Banach spaces*, J. Convex Anal. **4**(1997), 235–254.

[3] Ya. I. Alber, A. N. Iusem, *Extension of subgradient techniques for nonsmooth convex optimization in Banach spaces*, Set-Valued Analysis **9**(2001), 315–335.

[4] L. Armijo, *Minimization of function having Lipschitz continuous first partial derivatives*, Pac. J. Math. **16**(1966), 1–3.

[5] D. P. Bertsekas, *On the Goldstein-Levitin-Polyak Gradient Projection Method*, IEEE Trans. Automat. Control **21**(1976), 174–184.

[6] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, New York, 2010.

[7] R. H. Byrd, R. A. Tapia, *An extension of Curryâs theorem to steepest descent in normed linear spaces*, Math. Programming **9**(1975), 247–254.

[8] P. G. CIARLET, *Introduction to numerical linear algebra and optimization*, Cambridge University Press, New York, 1998.

[9] A. I. COHEN, *Stepsize analysis for descent methods*, J. Optim. Theory Appl. **33**(1981), 187–205.

[10] R. CORREA, C. LEMARECHAL, *Convergence of some algorithms for convex optimization*, Math. Programming **62**(1993), 261–275.

[11] M. GOLOMB, R. A. TAPIA, *The metric gradient in normed linear spaces*, Numer. Math. **20**(1972), 115–124.

[12] M. HINZE, R. PINNAU, M. ULBRICH, S. ULBRICH, *Optimization with PDE Constraints*, Springer, New York, 2009.

[13] D. G. LUENBERGER, *Optimization by Vector Space Methods*, John Wiley and Sons, Inc., 1969.

[14] J. NOCEDAL, J. W. STEPHEN, *Numerical Optimization*, Springer-Verlag, New York, 1999.

[15] J. P. PENOT, *On the convergence of descent algorithms*, Comput. Optim. Appl. **23**(2002), 279–284.

[16] B. T. POLJAK, *Gradient methods for solving equations and inequalities*, Z. Vyčisl. Mat. i Mat. Fiz. **4**(1964), 995–1005. (in Russian)

[17] Z.-J. SHI, *Convergence of line search methods for unconstrained optimization*. Appl. Math. Comput. **157**(2004), 393–405.

[18] C. ZALINESCU, *On uniformly convex functions*, J. Math. Anal. Appl. **95**(1983), 344–374.

[19] C. ZALINESCU, *Convex Analysis in General Vector Spaces*, World Scientific, New Jersey–London–Singapore–Hong Kong, 2002.