J Optim. Theory Appl. manuscript No.
(will be inserted by the editor)

Forward-backward Penalty Scheme for Constrained Convex Minimization without Inf-compactness

Nahla Noun · Juan Peypouquet

Communicated by Alfredo Iusem

Abstract In order to solve constrained minimization problems, Attouch et al. propose a forward-backward algorithm that involves an exterior penalization scheme in the forward step. They prove that every sequence generated by the algorithm converges weakly to a solution of the minimization problem if either the objective function or the penalization function corresponding to the feasible set is inf-compact. Unfortunately, this assumption leaves out problems that are not coercive, as well as several interesting applications in infinite-dimensional spaces. The purpose of this short article is to show this convergence result without the inf-compactness assumption.

Keywords constrained convex optimization · forward-backward algorithms · exterior penalization

The first author's PhD thesis is supported by UL − Azm & Saade Association, Lebanon. The second author is partly supported by FONDECYT Grant 11090023, Basal Project CMM Universidad de Chile, Núcleo Milenio Información y Coordinación en Redes ICM/FIC P10-024F, and Anillo Project ACT-1106: Analysis of Control Problems & Applications.

N. Noun
Institut de Mathématiques et Modélisation de Montpellier, UMR 5149 CNRS, Université Montpellier 2, place Eugène Bataillon, 34095 Montpellier cedex 5, France & Département et Laboratoire de Mathématiques, Faculté des Sciences 1 et Ecole Doctorale des Sciences et de Technologie, Hadath, Liban. E-mail: nahla.noun@math.univ-montp2.fr

J. Peypouquet (Corresponding Author)
Departamento de Matemática & AM2V, Universidad Técnica Federico Santa María, Avenida España 1680, Valparaíso, Chile. E-mail: juan.peypouquet@usm.cl

Mathematics Subject Classification (2000) 46N10 · 49M37 · 65K10 · 90C25

1 Introduction

In order to solve general constrained convex minimization problems, Attouch et al. propose in [1] a forward-backward algorithm that involves an exterior penalization scheme in the forward step. They prove that every sequence generated by that procedure converges weakly to a solution of the optimization problem if either the objective function or the penalization function is inf-compact (convergence being strong in the latter case). Unfortunately, the inf-compactness assumption leaves out several interesting applications in infinite-dimensional spaces, as well as finite-dimensional problems that are not coercive. The purpose of this short article is to prove the convergence result without the inf-compactness assumption.

A continuous-time version of the method was first introduced in [2] in the form of a differential inclusion for solving constrained variational inequalities determined by maximal monotone operators. In [3], the authors proposed two algorithms based on implicit discretizations of the differential inclusion in [2]. The forward-backward method we discuss here was studied in [1], where the most powerful convergence results correspond to the smooth case, where the penalization function involved in the forward step has Lipschitz-continuous gradient. In all the previously cited works, weak ergodic convergence holds for general maximal monotone operators, while weak convergence of the whole trajectory or sequence occurs in the subdifferential case (the constrained minimization problem). When both the objective function and the penalization function are smooth, a diagonal forward method of gradient type is proved to converge in [4]. The sequence of papers [1–4] contains several different applications and a more thorough bibliographical discussion. We shall not enter into the details here for the sake of brevity.

The paper is organized as follows: Section 2 contains the general setting, including the notation, the statement of the problem and the description of the algorithm, along with some examples. The main result and its proof are presented in Section 3, where we also provide a brief commentary on the hypotheses. Finally, in Section 4 we propose some directions to extend the results presented in this work.

2 General Setting

In this section, we describe the constrained minimization problem to be solved and the forward-backward penalty scheme proposed in [1], and mention some particular instances that we find interesting.

2.1 The Constrained Minimization Problem

Consider the constrained minimization problem

$$(\mathcal{P}) \qquad \text{Find } x^* \in \mathcal{S} = \operatorname{Argmin}\{\Phi(u) : u \in C\},$$

where C is a nonempty, closed and convex subset of a real Hilbert space $\mathcal{H}$ and Φ is a proper, lower-semicontinuous and convex function on $\mathcal{H}$ taking values in $\mathbb{R} \cup \{+\infty\}$. We shall assume in what follows that the set $\mathcal{S}$ is nonempty. Take a (everywhere finite) convex and differentiable function $\Psi : \mathcal{H} \to \mathbb{R}$ such that $C = \mathrm{Argmin}\Psi$ and restate $(\mathcal{P})$ as

$$\text{Find } x^* \in \mathcal{S} = \mathrm{Argmin}\{\Phi(u) : u \in \mathrm{Argmin}\Psi\}.$$

Here, we are interested in functions Ψ whose gradients are Lipschitz-continuous, and, for simplicity, we shall also assume that $\mathrm{Argmin}\Psi = \{x \in \mathcal{H} : \Psi(x) = 0\}$. Let us see some examples.

2.1.1 Sparse Control of a Linear Elliptic Equation

Let Ω be a bounded open subset of $\mathbb{R}^N$ with $\mathcal{C}^2$ boundary. Given $y_0 \in L^2(\Omega)$, consider the optimal control problem

$$\begin{cases} \text{Minimize } \ \|\alpha\|_{L^1(\Omega)} + \frac{1}{2}\|y - y_0\|^2_{L^2(\Omega)}, \\ \text{subject to } \ \alpha \in L^2(\Omega), \ y \in H^1_0(\Omega) \text{ and } -\Delta y = \alpha. \end{cases}$$

The nondifferentiable part involving the L^1-norm is known to induce sparsity of the optimal control α^*, while the quadratic term forces the optimal state y^* towards the reference signal y_0. Since every feasible point (y, α) verifies $\Delta y \in L^2(\Omega)$ and the boundary of Ω is $\mathcal{C}^2$, we must have $y \in H^2(\Omega)$ (see, for instance, Theorem 4 in Section 6.3 of [5]). Therefore, the solutions must belong to $\mathcal{H} = H^1_0(\Omega) \cap H^2(\Omega) \times L^2(\Omega)$, which is a Hilbert space when equipped with the inner product $\langle (y, \alpha), (z, \eta)\rangle_{\mathcal{H}} = \langle y, z\rangle_{H^2(\Omega)} + \langle \alpha, \eta\rangle_{L^2(\Omega)}$ and the corresponding norm. For $(y, \alpha) \in \mathcal{H}$ define

$$\Phi(y, \alpha) := \|\alpha\|_{L^1(\Omega)} + \frac{1}{2}\|y - y_0\|^2_{L^2(\Omega)} \quad \text{and} \quad \Psi(y, \alpha) := \frac{1}{2}\|\Delta y + \alpha\|^2_{L^2(\Omega)},$$

and rewrite the problem as

$$\text{Find } (y^*, \alpha^*) \in \mathrm{Argmin}\,\{\ \Phi(y, \alpha) \ : \ (y, \alpha) \in \mathrm{Argmin}\Psi\ \}\,.$$

Observe that neither Φ nor Ψ is inf-compact (i.e. their sublevel sets are not compact), but both Φ and Ψ are continuous, convex and everywhere defined[1]. The function Ψ is differentiable with $\nabla\Psi(y,\alpha) = (\Delta^*(\Delta y + \alpha), \Delta y + \alpha)$. Moreover, $\nabla\Psi$ is Lipschitz-continuous with constant $L_\Psi \leq 2N$.

2.1.2 A Convex Feasibility Problem

Consider the problem of finding a point in the intersection of two unbounded, closed and convex subsets C and D of $\mathbb{R}^N$ (just for simplicity). Assume that:

- i) There is, of course, an obvious choice of a smooth convex function Ψ such that $C = \text{Argmin}\Psi$. This is the case when C is given by a set of smooth convex inequalities: $C = \{x : g_j(x) \leq c_j,\ j = 1, \ldots, J\}$. Here, the function $\Psi(x) = \sum_{j=1}^{J} [g_j(x) - c_j]_+^2$ does the trick.
- ii) The set D is easy to project onto. For instance, if $D = \{x : Ax = b\}$, where b is a vector in $\mathbb{R}^M$ and A is a matrix of size $M \times N$ $(M < N)$ with linearly independent rows and such that $b \in R(A)$ (meaning that the system is well-posed, not redundant and has infinite solutions), then $\text{Proj}_D(x) = x - A^*(AA^*)^{-1}(Ax - b)$.

Under these conditions $C \cap D = \text{Argmin}\{\Phi(u) : u \in \text{Argmin}\Psi\}$ with $\Phi = \delta_D$, the indicator function of D. It seems reasonable to take advantage of this particular structure by using the function Ψ to approach C along with the projection onto D (see 2.2.2 below).

[1] Since Ω is bounded, $L^2(\Omega) \subset L^1(\Omega)$.

2.2 The Forward-Backward-Penalty Scheme

Following [1], in order to solve $(\mathcal{P})$, we introduce a forward-backward algorithm involving an exterior penalization scheme in the forward step:

$$(FBP) \qquad \begin{cases} x^1 \in \mathcal{H} \\ x^{n+1} = (I + \lambda_n \partial \Phi)^{-1} \left(x^n - \lambda_n \beta_n \nabla \Psi(x^n) \right), \quad \text{for } n \geq 1, \end{cases}$$

where (λ_n) and (β_n) are sequences of positive parameters. Some particular instances are the following:

2.2.1 (Multiscale) Forward-Backward Algorithm

If $\lambda_n \beta_n = \gamma$, then we have

$$x^{n+1} = (I + \lambda_n \partial \Phi)^{-1} \left(x^n - \gamma \nabla \Psi(x^n) \right), \quad \text{for } n \geq 1.$$

This can be interpreted as a multiscale Forward-Backward algorithm in the sense that the weights assigned to $\partial \Phi$ and $\nabla \Psi$ are of a different order. This induces a hierarchical selection of optimal points: the algorithm gives preference to the minimization of a primary criterion Ψ (the function with the higher weight) and then selects, among the minimizers of Ψ, one that minimizes the secondary criterion Φ.

2.2.2 The Gradient Projection Method

If D is a nonempty, closed and convex subset of $\mathcal{H}$ and $\Phi = \delta_D$, then we have $(I + \lambda_n \partial \Phi)^{-1} = \mathrm{Proj}_D$, independently of λ_n. In that case,

$$x^{n+1} = \mathrm{Proj}_D \left(x^n - \gamma_n \nabla \Psi(x^n) \right), \quad \text{for } n \geq 1,$$

where now $\gamma_n = \lambda_n \beta_n$ represents the stepsize for the gradient subiteration. This is the well-known *gradient projection method* introduced by Goldstein in [6]. As $n \to \infty$, the sequence (x^n) will converge weakly to a point in $C \cap D$, provided this set is nonempty.

3 The Convergence Result and its Proof

Let us begin this section by stating and quickly commenting the assumptions needed for the convergence result. The following hypotheses $(\mathbf{H}_1 - \mathbf{H}_4)$ are used in [1]:

$\mathbf{H}_1$: The operator $\partial\Phi + N_C$ is maximal monotone. This simply states that $\partial\Phi + N_C = \partial(\Phi + \delta_C)$ and so $\mathcal{S} = (\partial\Phi + N_C)^{-1}(0)$. Several qualification conditions are sufficient for this property. The interested reader may consult the introduction of [7] and the references therein.

$\mathbf{H}_2$: The sequence of stepsizes (λ_n) belongs to $\ell^2 \setminus \ell^1$.

$\mathbf{H}_3$: The gradient of Ψ is Lipschitz continuous with constant L_Ψ, and there exist positive numbers γ and Γ such that $0 < \gamma \leq \lambda_n \beta_n \leq \Gamma < 2/L_\Psi$ for all (sufficiently large) n. These are typical assumptions for the convergence of gradient methods (see, for instance, Section 1.2 in [8]).

$\mathbf{H}_4$: For each p in the range of N_C, $\sum_{n=1}^{\infty} \left[\Psi^*\left(\frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] < \infty$. This condition somehow relates the growth of the sequence (β_n) to the shape of Ψ around its minimizing set C. It was introduced in [3] and has its continuous counterpart in [2]. The reader is referred to [4] for further discussion.

In addition to the preceding hypotheses, we shall consider the following − very mild − condition, which appears in [3] and [4], and seems to be the key to dropping the inf-compactness assumption made in [1].

H$_5$: There exists $K > 0$ such that $\beta_{n+1} - \beta_n \leq K$ for all (sufficiently large) n. This prevents the sequence (β_n) from increasing too fast, which also helps to avoid numerical instabilities.

As mentioned in [4, Section 3.2], hypotheses $\mathbf{H}_2$ and $\mathbf{H}_3$ together imply $\mathbf{H}_4$ whenever Ψ is bounded from below by a multiple of the square of the distance to C. Then one can take, for instance, $\beta_n \sim n$ and $\lambda_n \sim \frac{1}{n}$ for hypothesis $\mathbf{H}_5$ to hold. Observe also that, in the context of the gradient projection method described above, if the space is infinite-dimensional and the sets C and D have nonempty interior, neither Φ nor Ψ are inf-compact. In fact, even the intersection of their level sets with sufficiently large closed balls fail to be compact.

The main result of this paper is the following:

Theorem 3.1 *Let (x^n) satisfy (FBP) and assume hypotheses $\mathbf{H}_1 - \mathbf{H}_5$ hold. Then (x^n) converges weakly as $n \to \infty$ to a point in $\mathcal{S}$.*

Before we proceed with the proof of Theorem 3.1, we state the following result, which collects several consequences of [1, Proposition 11] that will be useful in what follows:

Proposition 3.1 *Assume hypotheses $\mathbf{H}_1 - \mathbf{H}_4$ hold. Then we have the following:*

i) For each $u \in \mathcal{S}$, $\lim_{n\to\infty} \|x^n - u\|$ exists.

ii) The series $\sum_{n=1}^{\infty} \|x^{n+1} - x^n\|^2$, $\sum_{n=1}^{\infty} \Psi(x^n)$ and $\sum_{n=1}^{\infty} \|\nabla\Psi(x^n)\|^2$ are convergent.

In particular, $\lim_{n\to\infty} \Psi(x^n) = 0$ *and every weak cluster point of the sequence* (x^n) *lies in* C*.*

In the proof of Lemma 3.2 we make use of the following elementary fact concerning real sequences, whose obvious proof we omit:

Lemma 3.1 *If* (a_n) *and* (ρ_n) *are non-negative sequences such that* $(\rho_n) \in \ell^1$ *and* $a_{n+1} - a_n \leq \rho_n$ *for all* n*, then* $\lim_{n\to\infty} a_n$ *exists.*

The analysis relies on the study of the sequence $(\Omega_n(x^n))$, where Ω_n is the penalized function given by

$$\Omega_n = \Phi + \beta_n \Psi$$

for $n \geq 1$.

Lemma 3.2 *Assume hypotheses* $\mathbf{H}_1 - \mathbf{H}_5$ *hold. Then the sequence* $(\Omega_n(x^n))$ *converges as* $n \to \infty$*.*

Proof. Recall that

$$\frac{x^n - x^{n+1}}{\lambda_n} - \beta_n \nabla\Psi(x^n) \in \partial\Phi(x^{n+1}).$$

The subdifferential inequality for Φ gives

$$\Phi(x^n) \geq \Phi(x^{n+1}) + \left\langle \frac{x^n - x^{n+1}}{\lambda_n} - \beta_n \nabla\Psi(x^n), x^n - x^{n+1} \right\rangle$$

and so

$$\Phi(x^{n+1}) - \Phi(x^n) + \frac{1}{\lambda_n}\|x^{n+1} - x^n\|^2 \leq \beta_n \langle \nabla\Psi(x^n), x^n - x^{n+1} \rangle. \tag{1}$$

On the other hand, the Descent Lemma (see, for instance, Proposition A.24 in [8]) for Ψ gives

$$\Psi(x^{n+1}) \leq \Psi(x^n) + \langle \nabla\Psi(x^n), x^{n+1} - x^n \rangle + \frac{L_\Psi}{2}\|x^{n+1} - x^n\|^2, \qquad (2)$$

whence

$$\begin{aligned} &\beta_{n+1}\Psi(x^{n+1}) - \beta_n\Psi(x^n) - \frac{\beta_n L_\Psi}{2}\|x^{n+1} - x^n\|^2 \\ &\leq \beta_n\langle \nabla\Psi(x^n), x^{n+1} - x^n \rangle + (\beta_{n+1} - \beta_n)\Psi(x^{n+1}). \qquad (3) \end{aligned}$$

Adding (1) and (3) we obtain

$$\Omega_{n+1}(x^{n+1}) - \Omega_n(x^n) + \left[\frac{1}{\lambda_n} - \frac{\beta_n L_\Psi}{2}\right]\|x^{n+1} - x^n\|^2 \leq (\beta_{n+1} - \beta_n)\Psi(x^{n+1}).$$

Since $\lambda_n\beta_n < 2/L_\Psi$ and $\beta_{n+1} - \beta_n \leq K$, we have

$$\Omega_{n+1}(x^{n+1}) - \Omega_n(x^n) \leq K\Psi(x^{n+1}).$$

The sequence $(\Omega_n(x^n))$ is bounded from below because the sequence (x^n) is bounded. The right-hand side is summable by Part *ii)* in Proposition 3.1, whence Lemma 3.1 implies $\lim_{n\to\infty}\Omega_n(x^n)$ exists. □

The argument in the proof of Lemma 3.2 also provides the summability of the series $\sum_{n=1}^{\infty}\frac{\|x^{n+1}-x^n\|^2}{\lambda_n}$. This is stronger than the summability of $\sum_{n=1}^{\infty}\|x^{n+1} - x^n\|^2$, but does not imply that of $\sum_{n=1}^{\infty}\|x^{n+1} - x^n\|$, which would immediately give strong convergence of the sequence.

Lemma 3.3 *Assume hypotheses* $\mathbf{H}_1 - \mathbf{H}_5$ *hold. For each* $u \in C$ *one has*

$$\sum_{n=1}^{\infty}\lambda_n\left[\Omega_{n+1}(x^{n+1}) - \Phi(u)\right] < \infty \qquad (\textit{possibly } -\infty).$$

Proof. First observe that

$$\Omega_{n+1}(x^{n+1}) - \Phi(u)$$

$$= \Phi(x^{n+1}) + \beta_n \Psi(x^n) - \Phi(u) + (\beta_{n+1} - \beta_n)\Psi(x^{n+1}) + \beta_n(\Psi(x^{n+1}) - \Psi(x^n))$$

$$\leq \Phi(x^{n+1}) + \beta_n \Psi(x^n) - \Phi(u) + K\Psi(x^{n+1}) + \beta_n(\Psi(x^{n+1}) - \Psi(x^n)). \quad (4)$$

Using (2), we obtain

$$\begin{aligned} \beta_n(\Psi(x^{n+1}) - \Psi(x^n)) &\leq \beta_n \langle \nabla\Psi(x^n), x^{n+1} - x^n \rangle + \frac{\beta_n L_\Psi}{2}\|x^{n+1} - x^n\|^2 \\ &\leq \frac{\beta_n}{2}\|\nabla\Psi(x^n)\|^2 + \frac{\beta_n(L_\Psi + 1)}{2}\|x^{n+1} - x^n\|^2. \quad (5) \end{aligned}$$

Inequalities (4) and (5) give

$$\lambda_n \left[\Omega_{n+1}(x^{n+1}) - \Phi(u)\right] \leq \lambda_n \left[\Phi(x^{n+1}) + \beta_n \Psi(x^n) - \Phi(u)\right] + \varepsilon_n,$$

where

$$\varepsilon_n = \lambda_n K \Psi(x^{n+1}) + \frac{\lambda_n \beta_n}{2}\|\nabla\Psi(x^n)\|^2 + \frac{\lambda_n \beta_n (L_\Psi + 1)}{2}\|x^{n+1} - x^n\|^2.$$

Since the sequences (λ_n) and $(\lambda_n \beta_n)$ are bounded, Proposition 3.1 implies $(\varepsilon_n) \in \ell^1$. Therefore, it suffices to prove that

$$\sum_{n=1}^{\infty} \lambda_n \left[\Phi(x^{n+1}) + \beta_n \Psi(x^n) - \Phi(u)\right] < \infty. \quad (6)$$

The subdifferential inequality for Φ at points u and x^{n+1} gives

$$\Phi(u) \geq \Phi(x^{n+1}) + \left\langle \frac{x^n - x^{n+1}}{\lambda_n} - \beta_n \nabla\Psi(x^n), u - x^{n+1} \right\rangle. \quad (7)$$

On the other hand, since $\Psi(u) = 0$, the subdifferential inequality for Ψ at points u and x^n gives

$$\begin{aligned} 0 &\geq \Psi(x^n) + \langle \nabla\Psi(x^n), u - x^n\rangle \\ &= \Psi(x^n) + \langle \nabla\Psi(x^n), u - x^{n+1}\rangle + \langle \nabla\Psi(x^n), x^{n+1} - x^n\rangle. \end{aligned} \tag{8}$$

Combining (7) and (8) we obtain

$$2\lambda_n \left[\Phi(x^{n+1}) + \beta_n\Psi(x^n) - \Phi(u)\right]$$

$$\leq 2\langle x^n - x^{n+1}, x^{n+1} - u\rangle + 2\lambda_n\beta_n\langle \nabla\Psi(x^n), x^n - x^{n+1}\rangle.$$

However,

$$2\langle x^n - x^{n+1}, x^{n+1} - u\rangle = \|x^n - u\|^2 - \|x^{n+1} - u\|^2 - \|x^{n+1} - x^n\|^2$$

and

$$2\langle \lambda_n\beta_n\nabla\Psi(x^n), x^n - x^{n+1}\rangle \leq \frac{4}{L_\Psi^2}\|\nabla\Psi(x^n)\|^2 + \|x^{n+1} - x^n\|^2.$$

Thus

$$2\lambda_n \left[\Phi(x^{n+1}) + \beta_n\Psi(x^n) - \Phi(u)\right] \leq \|x^n - u\|^2 - \|x^{n+1} - u\|^2 + \frac{4}{L_\Psi^2}\|\nabla\Psi(x^n)\|^2.$$

We conclude that

$$2\sum_{n=1}^{m}\lambda_n \left[\Phi(x^{n+1}) + \beta_n\Psi(x^n) - \Phi(u)\right]$$

$$\leq \|x^1 - u\|^2 - \|x^{m+1} - u\|^2 + \frac{4}{L_\Psi^2}\sum_{n=1}^{m}\|\nabla\Psi(x^n)\|^2$$

for $m \geq 1$. In view of Proposition 3.1, this shows (6) and completes the proof.□

Proof of Theorem 3.1. In order to prove Theorem 3.1, we shall use a well-known result most commonly referred to as Opial's Lemma (see, for instance, [9, Lemma 4.1] for the simplified version used here, or also [10, Lemma 1] for its full form). In view of Part *i*) in Proposition 3.1, it suffices to verify that every weak cluster point of the sequence (x^n) must belong to $\mathcal{S}$. Since $(\lambda_n) \notin \ell^1$, Lemmas 3.2 and 3.3 imply $\lim_{n\to\infty} \Omega_n(x^n) \leq \Phi(u)$ whenever $u \in C$. Suppose a subsequence (x^{k_n}) of (x^n) converges weakly to some x^∞ as $n \to \infty$. Then $x^\infty \in C$ by Proposition 3.1. The weak lower-semicontinuity of Φ then gives

$$\Phi(x^\infty) \leq \liminf_{n\to\infty} \Phi(x^{k_n}) \leq \liminf_{n\to\infty} \Omega_{k_n}(x^{k_n}) = \lim_{n\to\infty} \Omega_n(x^n) \leq \Phi(u).$$

Therefore x^∞ minimizes Φ on C and so $x^\infty \in \mathcal{S}$. □

4 Concluding Remarks

In this model, the objective function is allowed to be nonsmooth, while the penalization function for the set of constraints is required to be smooth. A challenging problem arises when both the objective function and the penalization function have a smooth part and a nonsmooth part. Of course, one way to address this issue is to consider a fully nonsmooth setting, as in [3]. However, by doing so, the underlying smoothness of the problem is not exploited. One possible way to tackle this situation more efficiently is to perform a gradient step with respect to the smooth parts following the ideas in [4], and a proximal step with respect to the nonsmooth part, according to what is done in [3]. Some progress has been made in this direction, and this will be the matter of a forthcoming paper.

References

1. Attouch H., Czarnecki, M.-O., Peypouquet J.: Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities, SIAM J. Optim., 21, no. 4, 1251–1274 (2011)
2. Attouch H., Czarnecki M.-O.: Asymptotic behavior of coupled dynamical systems with multiscale aspects, J. Differential Equations, 248, no. 6, 1315–1344 (2010)
3. Attouch H., Czarnecki M.-O., Peypouquet J.: Prox-penalization and splitting methods for constrained variational problems, SIAM J. Optim., 21, no. 1, 149–173 (2011)
4. Peypouquet J.: Coupling the gradient method with a general exterior penalization scheme for convex minimization, J. Optim. Theory Appl., 153, no. 1, 123–138 (2012)
5. Evans L.-C.: Partial differential equations, Second edition, Graduate Studies in Mathematics 19, American Mathematical Society, Providence, RI (2010)
6. Goldstein A.-A.: Convex programming in Hilbert space., Bull. Amer. Math. Soc., 70, 709–710 (1964)
7. Ernst E., Théra M.: On the necessity of the Moreau-Rockafellar-Robinson qualification condition in Banach spaces, Math. Program., 117, no. 1-2, Ser. B, 149–161 (2009)
8. Bertsekas, D.: Nonlinear programming. Athena Scientific, Belmont MA (1999)
9. Peypouquet J., Sorin S.: Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time, J. Convex Anal. 17, no. 3-4, 1113–1163 (2010)
10. Brézis H., Browder F.-E.: Nonlinear ergodic theorems, Bull. Amer. Math. Soc., 82, no. 6, 959–961 (1976)