



HAL
open science

Prox-Penalization and Splitting Methods for Constrained Variational Problems

Hedy Attouch, Marc-Olivier Czarnecki, Juan Peypouquet

► **To cite this version:**

Hedy Attouch, Marc-Olivier Czarnecki, Juan Peypouquet. Prox-Penalization and Splitting Methods for Constrained Variational Problems. *SIAM Journal on Optimization*, 2011, 21 (1), pp.149-173. 10.1137/100789464 . hal-00803589

HAL Id: hal-00803589

<https://hal.science/hal-00803589>

Submitted on 22 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PROX-PENALIZATION AND SPLITTING METHODS FOR CONSTRAINED VARIATIONAL PROBLEMS

HÉDY ATTOUCH, MARC-OLIVIER CZARNECKI & JUAN PEYPOUQUET

ABSTRACT. This paper is concerned with the study of a class of prox-penalization methods for solving variational inequalities of the form

$$Ax + N_C(x) \ni 0$$

where \mathcal{H} is a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximal monotone operator and N_C is the outward normal cone to a closed convex set $C \subset \mathcal{H}$. Given $\Psi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ which acts as a penalization function with respect to the constraint $x \in C$, and a penalization parameter β_n , we consider a diagonal proximal algorithm of the form

$$x_n = \left(I + \lambda_n (A + \beta_n \partial \Psi) \right)^{-1} x_{n-1},$$

and an algorithm which alternates proximal steps with respect to A and penalization steps with respect to C and reads as

$$x_n = (I + \lambda_n \beta_n \partial \Psi)^{-1} (I + \lambda_n A)^{-1} x_{n-1}.$$

We obtain weak ergodic convergence for a general maximal monotone operator A , and weak convergence of the whole sequence $\{x_n\}$ when A is the subdifferential of a proper lower-semicontinuous convex function. Mixing with Passty's idea, we can extend the ergodic convergence theorem, so obtaining the convergence of a prox-penalization splitting algorithm for constrained variational inequalities governed by the sum of several maximal monotone operators. Our results are applied to an optimal control problem where the state variable and the control are coupled by an elliptic equation. We also establish robustness and stability results that account for numerical approximation errors.

INTRODUCTION

Let \mathcal{H} be a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a general maximal monotone operator, and C a closed convex set in \mathcal{H} . We denote by N_C the outward normal cone to C . This paper is concerned with the study of a class of prox-penalization and splitting algorithms for solving variational inequalities of the form

$$(1) \quad Ax + N_C(x) \ni 0,$$

which combine proximal steps with respect to A and penalization steps with respect to C . We begin by describing two model situations that motivate our study:

1. Sum of maximal monotone operators. Let \mathcal{X} be a real Hilbert space and set $\mathcal{H} = \mathcal{X} \times \mathcal{X}$. Define $A : \mathcal{H} \rightrightarrows \mathcal{H}$ by $A(x_1, x_2) = (A_1 x_1, A_2 x_2)$ where A_1 and A_2 are maximal monotone operators on \mathcal{X} . If $C = \{(x_1, x_2) \in \mathcal{X} \times \mathcal{X} : x_1 = x_2\}$ the inclusion (1) reduces to

$$(2) \quad A_1 x + A_2 x \ni 0.$$

Date: March, 15, 2010.

1991 Mathematics Subject Classification. 37N40, 46N10, 49M30, 65K05, 65K10, 90B50, 90C25.

Key words and phrases. Nonautonomous gradient-like systems; monotone inclusions; asymptotic behaviour; hierarchical convex minimization; splitting methods; optimal control.

With the support of the French ANR grant ANR-08-BLAN-0294-03. J. Peyrouquet was partly supported by FONDECYT grant 11090023 and Basal Proyect, CMM, Universidad de Chile.

In the line of Trotter-Kato formula, we would like to solve this problem by using splitting methods which only require to compute resolvents (proximal steps) with respect to A_1 and A_2 , separately. A valuable guideline is a theorem from [24, Passty] which states that any sequence $\{x_n\}$ generated by the algorithm

$$(3) \quad x_n = (I + \lambda_n A_2)^{-1} (I + \lambda_n A_1)^{-1} x_{n-1}$$

converges weakly in average to some x satisfying (2) provided $\{\lambda_n\} \in \ell^2(\mathbb{N}) \setminus \ell^1(\mathbb{N})$. \square

2. Structured convex minimization. Coupled variational problems where the coupling occurs in the constraint play a central role in decision and engineering sciences. Consider the minimization problem

$$(4) \quad \min \{f_1(x_1) + f_2(x_2) : L_1 x_1 = L_2 x_2, (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2\},$$

where \mathcal{X}_1 , \mathcal{X}_2 and \mathcal{Z} are real Hilbert spaces and each L_i is bounded linear (or affine) operator from \mathcal{X}_i to \mathcal{Z} . This type of structured variational problem appears in optimal control of linear systems, in the study of domain decomposition methods for PDE's, transport, imaging and signal processing. Considering infinite dimensional spaces is crucial for these types of applications. Problem (4) falls in our setting by taking $A(x_1, x_2) = (\partial f_1(x_1), \partial f_2(x_2))$ and $C = \{(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 : L_1 x_1 = L_2 x_2\}$. Splitting algorithms attached to such coupled variational problems have a rich interpretation in terms of best response dynamics for potential games, see [4, Attouch, Bolte, Redont and Soubeyran]. \square

In order to address these problems in a unified way, we use the links between algorithms and continuous dissipative dynamical systems, and their asymptotic analysis by Liapunov methods. As we shall see, our algorithms can be derived by time discretization of the continuous nonautonomous dynamical system

$$(5) \quad \dot{x}(t) + Ax(t) + \beta(t)\partial\Psi(x(t)) \ni 0,$$

which has been recently introduced in [5, Attouch and Czarnecki], and whose trajectories (under certain conditions on the function $\beta(\cdot)$) asymptotically reach equilibria given by (1). In system (5) above, the function $\Psi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ acts as an external penalization function with respect to the constraint $x \in C$. The corresponding penalization parameter $\beta(t)$ tends to $+\infty$ as $t \rightarrow +\infty$.

Observe that an implicit discretization of the differential inclusion (5) gives

$$(6) \quad \frac{1}{\lambda_n} (x_n - x_{n-1}) + Ax_n + \beta_n \partial\Psi(x_n) \ni 0,$$

where λ_n and β_n are sequences of positive parameters. Inclusion (6) can be rewritten as

$$(7) \quad x_n = \left(I + \lambda_n (A + \beta_n \partial\Psi) \right)^{-1} x_{n-1},$$

giving a diagonal proximal point algorithm. On the other hand, since the resolvent of the sum of two maximal monotone operators may be hard to compute, we propose also an alternating method:

$$(8) \quad x_n = (I + \lambda_n \beta_n \partial\Psi)^{-1} (I + \lambda_n A)^{-1} x_{n-1},$$

which combines proximal steps corresponding to the operator A and the set C .

The implicit scheme described by (8) makes sense for any maximal monotone operator A and any $\Psi \in \Gamma_0(\mathcal{H})$. Under more restrictive assumptions on Ψ one may also consider a mixed

explicit-implicit algorithm of the form

$$\frac{1}{\lambda_n}(x_n - x_{n-1}) + Ax_n + \beta_n \partial \Psi(x_{n-1}) \ni 0,$$

which can be rewritten as

$$(9) \quad x_n = (I + \lambda_n A)^{-1}(x_{n-1} - \lambda_n \beta_n w_n),$$

for $w_n \in \partial \Psi(x_{n-1})$. Explicit schemes (in general) have the advantage of being easier to compute, which ensures enhanced applicability to real-life problems. However, they tend to be less stable than the implicit ones. Assuming Ψ to satisfy some additional regularity properties, it is reasonable to expect that algorithm (9) still enjoys good asymptotic convergence properties. This interesting subject requires further studies, which go beyond the scope of this paper.

Diagonal algorithms of the form

$$\frac{1}{\lambda_n}(x_n - x_{n-1}) + A_n(x_n) \ni 0$$

for general families of maximal monotone operators A_n have been studied [17, Kato], [18, Kobayasi, Kobayashi and Oharu], [2, Alvarez and Peypouquet]; especially in terms of their relationship with continuous-time trajectories solving a differential inclusion of the form

$$0 \in \dot{x}(t) + A(t)x(t).$$

In the special case where $A_n = \partial f_n$, asymptotic properties of algorithms such as (6) are proved in [7, Auslender, Crouzeix and Fedit] in the framework of exterior penalization. Further study was carried out in [1, Alart and Lemaire] under variational convergence assumptions and in [8, Bahraoui and Lemaire] and [21, Lemaire] using convergence of the subdifferentials in terms of the Hausdorff excess function. Convergence in value is studied in [25, Peypouquet]. For general penalization schemes in convex programming the reader can consult [13, Cominetti and Courdurier] and the references therein.

There are natural links between prox-penalization and proximal methods involving asymptotically vanishing terms (viscosity methods). They both involve multiscale aspects and lead to hierarchical minimization. Regarding their continuous versions, passing from one to the other relies on time rescaling, see [5]. For Tikhonov regularization see [20, Lehdili and Moudafi] and [14, Cominetti, Peypouquet and Sorin]. See [11, Cabot] for some further related results and references.

As we shall see, the use of a penalization-like scheme for general maximal monotone operators is an effective tool for finding solutions of constrained variational inequalities.

In classical penalization, most results rely on the smoothness and other special features of the penalization function. Having in view a large range of applications, we shall not assume any particular structure or regularity on the penalization function Ψ . Instead, we just suppose that Ψ is convex, lower-semicontinuous and $C = \operatorname{argmin} \Psi \neq \emptyset$. If $A = \partial \Phi$ for some proper lower-semicontinuous convex function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ and some qualification condition holds, the inclusion (1) is equivalent to

$$x \in \operatorname{argmin}\{\Phi(z) : z \in \operatorname{argmin} \Psi\}.$$

Therefore, our results can also be considered from a multiscale or hierarchical point of view.

Our main results can be summarized as follows: Under certain hypotheses on the sequences λ_n and β_n , and assuming a geometric condition involving the Fenchel conjugate of Ψ (which we shall state explicitly later on) we are able to prove the following results which can be classified into the following categories **A**, **B**, **C**:

Let $\{x_n\}$ be a sequence satisfying either (7) or (8) up to a numerical error ε_n (see Section 5 for precise details) and let $\{z_n\}$ be the sequence of weighted averages

$$(10) \quad z_n = \frac{1}{\tau_n} \sum_{k=1}^n \lambda_k x_k, \quad \text{where} \quad \tau_n = \sum_{k=1}^n \lambda_k.$$

- A.** The sequence $\{z_n\}$ converges weakly to a solution of (1) (Theorems 2.3 and 3.3).
- B.** If A is strongly monotone then $\{x_n\}$ converges strongly (Theorems 2.4 and 3.4).
- C.** If $A = \partial\Phi$ for some proper lower-semicontinuous convex function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ then $\{x_n\}$ converges weakly (Theorems 2.6 and 3.7).

This is the same type of asymptotic behavior as in the well-known proximal point algorithm with variable time step λ_n , see [22, Lions] and [10, Brézis and Lions]. See also [26, Peypouquet and Sorin] for a complete survey on the topic.

The paper is organized as follows: In section 1 we recall some basic facts about convex analysis and monotone operators, we state and discuss on the standing assumptions, and present some results from [23, Opial] and [24] that are useful for proving weak convergence of a sequence in a Hilbert space without *a priori* knowledge of the limit. Sections 2 and 3 contain our main results of type **A**, **B** and **C** for the algorithms given by (7) and (8), respectively. By mixing our techniques with Passty's idea [24], we obtain the convergence of a splitting algorithm for constrained variational inequalities governed by the sum of M maximal monotone operators. In section 4 we mention some applications of our results, with an illustrating example in optimal control of linear systems. We also show a simple numerical experience. Finally, in section 5 we provide some robustness and stability results concerning the dependence on the initial conditions and the convergence of the algorithms when the iterates are computed inexactly.

1. PRELIMINARIES

1.1. Some facts of convex analysis and monotone operators. Let \mathcal{H} be a real Hilbert space and let $\Gamma_0(\mathcal{H})$ denote the set of all proper lower-semicontinuous convex functions $F : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. Given $F \in \Gamma_0(\mathcal{H})$ and $x \in \mathcal{H}$, the *subdifferential* of F at x is the set

$$\partial F(x) = \{x^* \in \mathcal{H} : F(y) \geq F(x) + \langle x^*, y - x \rangle \text{ for all } y \in \mathcal{H}\}.$$

The *Fenchel conjugate* of F is the function $F^* \in \Gamma_0(\mathcal{H})$ defined by

$$F^*(x^*) = \sup_{y \in \mathcal{H}} \{\langle x^*, y \rangle - F(y)\}.$$

For $x, x^* \in \mathcal{H}$ one has $F(x) + F^*(x^*) \geq \langle x^*, x \rangle$ with equality if, and only if, $x^* \in \partial F(x)$.

Given a nonempty closed convex set $C \subset \mathcal{H}$, its *indicator function* is defined as $\delta_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise. The *support function* of C at a point x^* is $\sigma_C(x^*) = \sup_{c \in C} \langle x^*, c \rangle$. The *normal cone* to C at x is

$$N_C(x) = \{x^* \in \mathcal{H} : \langle x^*, c - x \rangle \leq 0 \text{ for all } c \in C\}$$

if $x \in C$ and \emptyset otherwise. We denote by $R(N_C)$ the range of N_C . Observe that $\delta_C^* = \sigma_C$ and $\partial\delta_C = N_C$. Notice also that $x^* \in N_C(x)$ if, and only if, $\sigma_C(x^*) = \langle x^*, x \rangle$.

A *monotone operator* is a set-valued mapping $A : \mathcal{H} \rightrightarrows \mathcal{H}$ such that

$$\langle x^* - y^*, x - y \rangle \geq 0$$

whenever $x^* \in Ax$ and $y^* \in Ay$. It is *maximal monotone* if its graph is not properly contained in the graph of any other monotone operator. The subdifferential of a function in $\Gamma_0(\mathcal{H})$ is maximal monotone. For any maximal monotone operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and for any $\lambda > 0$, the operator $I + \lambda A$ is surjective by Minty's Theorem (see [9, Brézis] or [26]). The operator $(I + \lambda A)^{-1}$ is a contraction that is everywhere defined. It is called the *resolvent* of A of index λ .

Finally, an operator A is *strongly monotone* with parameter $\alpha > 0$ if

$$\langle x^* - y^*, x - y \rangle \geq \alpha \|x - y\|^2$$

whenever $x^* \in Ax$ and $y^* \in Ay$. Observe that the set of zeroes of a strongly monotone operator is nonempty and contains a single element.

1.2. Standing assumptions. Let A be a maximal monotone operator on \mathcal{H} and $\Psi \in \Gamma_0(\mathcal{H})$ with $C = \operatorname{argmin} \Psi \neq \emptyset$. Without any loss of generality we assume $\min \Psi = 0$ (Ψ enters into the algorithm only via its subdifferential). Define the monotone operator $T_{A,C} = A + N_C$.

For our main results we shall make the following assumptions:

$$(\mathbf{H}) \quad \left\{ \begin{array}{l} (H_1) \quad \text{The solution set } \mathcal{S} = T_{A,C}^{-1}0 \text{ is nonempty.} \\ (H_2) \quad \text{The operator } T_{A,C} \text{ is maximal monotone.} \\ (H_3) \quad \sum_{n=1}^{\infty} \lambda_n = +\infty. \\ (H_4) \quad \text{For each } p \in R(N_C), \sum_{n=1}^{\infty} \lambda_n \beta_n \left[\Psi^* \left(\frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] < +\infty. \end{array} \right.$$

Some comments are in order:

(H_1): This simply states that problem (1) has a solution.

(H_2): The maximal monotonicity of $T_{A,C}$ gives the following characterization of the solution set: a point $z \in \mathcal{H}$ belongs to \mathcal{S} if, and only if,

$$\langle w, z - u \rangle \leq 0 \quad \text{for all } u \in C \cap \operatorname{dom}(A) \text{ and all } w \in T_{A,C}u.$$

If $A = \partial\Phi$ for some $\Phi \in \Gamma_0(\mathcal{H})$, the maximal monotonicity of $T_{A,C}$ simply states that $\partial\Phi + N_C = \partial(\Phi + \delta_C)$. Thus $z \in \mathcal{S}$ if, and only if, $0 \in \partial(\Phi + \delta_C)(z)$, which is equivalent to

$$z \in \operatorname{argmin}\{\Phi(x) : x \in C\}.$$

This holds if Φ and δ_C satisfy some qualification condition, such as the Moreau-Rockafellar or Attouch-Brézis.

(H_3): Since λ_n has a natural interpretation as a time step $\lambda_n = t_n - t_{n-1}$ in the discretization of (5), it is natural to assume $\sum_{n=1}^{\infty} \lambda_n = +\infty$ in order to preserve the asymptotic convergence properties of the continuous dynamics (see [5]). In other words, (H_3) is the discrete-time analogue of $t \rightarrow +\infty$.

(H_4): This is a discrete version of the following condition

$$(\mathcal{H}_1) \quad \text{For each } p \in R(N_C), \quad \int_0^{\infty} \beta(t) \left[\Psi^* \left(\frac{p}{\beta(t)} \right) - \sigma_C \left(\frac{p}{\beta(t)} \right) \right] dt < +\infty,$$

which was introduced in [5]. The analysis carried out in [5] remains valid in our discrete setting: First, all the terms in the sum are nonnegative. Indeed, since $\Psi(x) \leq \delta_C(x)$ for all $x \in \mathcal{H}$, one always has the reverse inequality for their Fenchel conjugates, namely $\Psi^*(p) - \sigma_C(p) \geq 0$ for all $p \in \mathcal{H}$.

In the special case where $\Psi(x) = \frac{1}{2} \operatorname{dist}(x, C)^2$, we have $\Psi^*(p) - \sigma_C(p) = \frac{1}{2} \|p\|^2$ for all $p \in \mathcal{H}$ and so

$$(11) \quad (H_4) \quad \iff \quad \sum_{n=1}^{\infty} \frac{\lambda_n}{\beta_n} < +\infty.$$

Suppose now that $\Psi(\cdot) \geq \frac{\theta}{2} \text{dist}(\cdot, C)^2$ for some $\theta > 0$. Then $\Psi^*(p) - \sigma_C(p) \leq \frac{1}{2\theta} \|p\|^2$ for all $p \in R(N_C)$ and so

$$\sum_{n=1}^{\infty} \frac{\lambda_n}{\beta_n} < +\infty \quad \Longrightarrow \quad (H_4).$$

Moreover, if the sequence $\{\beta_n\}$ is chosen so that $\limsup_{n \rightarrow \infty} \lambda_n \beta_n < +\infty$ and $\liminf_{n \rightarrow \infty} \lambda_n \beta_n > 0$ then

$$\sum_{n=1}^{\infty} \frac{\lambda_n}{\beta_n} < +\infty \quad \Longleftrightarrow \quad \sum_{n=1}^{\infty} \lambda_n^2 < +\infty.$$

The particular case $\Psi = 0$ corresponds to $C = \mathcal{H}$, which is the unconstrained case. In this situation $R(N_C) = \{0\}$, and since $\Psi^*(0) = \sigma_C(0) = 0$, condition (H_4) is trivially satisfied.

For the weak convergence of the sequence $\{x_n\}$ itself in the subdifferential case (Theorems 2.6 and 3.7) we shall assume an exponential-type growth condition on the sequence of parameters $\{\beta_n\}$, namely

(G) There is a constant $K \in \mathbb{R}$ such that $\frac{\beta_n - \beta_{n-1}}{\lambda_{n-1} \beta_{n-1}} \leq K$ for all sufficiently large n .

This is a discrete version of condition (\mathcal{H}_2) in [5]:

$$(\mathcal{H}_2) \quad \dot{\beta}(t) \leq K\beta(t).$$

1.3. A tool for proving weak convergence. The following lemma gathers results from [23, 24] (see also [26]). It is a simple but very useful tool for proving weak convergence in Hilbert spaces. What is interesting about this method is that one does not need to know the limit beforehand, but only a set to which this limit is expected to belong. Let $\{x_n\}$ be any sequence in \mathcal{H} and define $\{z_n\}$ as in (10) (recall that, by (H_3) , $\tau_n = \sum_{k=1}^n \lambda_k \rightarrow +\infty$ as $n \rightarrow +\infty$).

Lemma 1.1 (Opial-Passty). *Let F be a nonempty subset of \mathcal{H} and let $\lim_{n \rightarrow \infty} \|x_n - x\|$ exist for each $x \in F$. If every weak cluster point of $\{x_n\}$ (resp. $\{z_n\}$) lies in F , then $\{x_n\}$ (resp. $\{z_n\}$) converges weakly to a point in F as $n \rightarrow +\infty$.*

Proof. Since this result is less known in its ergodic form, let us prove it in this setting. Thus we want to prove weak convergence of the sequence $\{z_n\}$. Clearly the sequence $\{z_n\}$ is bounded. The space being reflexive, it suffices to prove that $\{z_n\}$ has only one weak cluster point as $n \rightarrow \infty$. Suppose otherwise that $z_{k_n} \rightharpoonup z$ and $z_{k'_n} \rightharpoonup z'$. Since

$$2\langle x_n, z - z' \rangle = \|x_n - z'\|^2 - \|x_n - z\|^2 - \|z'\|^2 + \|z\|^2,$$

we deduce the existence of $\lim_{n \rightarrow \infty} \langle x_n, z - z' \rangle$. But then $\lim_{n \rightarrow \infty} \langle z_n, z - z' \rangle$ exists as well, which implies that $\langle z, z - z' \rangle = \langle z', z - z' \rangle$. We conclude that $z = z'$. \blacksquare

2. PROX-PENALIZATION ALGORITHM

In this section we study the prox-penalization algorithm given by (7), namely

$$x_n = (I + \lambda_n(A + \beta_n \partial \Psi))^{-1} x_{n-1}.$$

Our results remain true if we allow an error ε_n in the computation of x_n . For the sake of clarity, we present the results in this section with $\varepsilon_n = 0$ and refer the reader to Section 5 for the general setting.

In order to guarantee the well-posedness of (7), all along this section we make the following standing qualification assumption (Q_n):

(Q_n) for each $n \in \mathbb{N}$, the monotone operator $A + \beta_n \partial \Psi$ is maximal monotone.

One can consult [9], [6, Attouch, Riahi and Théra] and the references therein, for general conditions ensuring that the sum of two maximal monotone operators is still maximal monotone. Observe that if $A = \partial \Phi$ this is true, for instance, under Moreau-Rockafellar or Attouch-Brézis qualification condition.

For any initial data $x_0 \in \mathcal{H}$, this procedure generates a unique trajectory $\{x_n\}$. The preceding equality can be equivalently written as

$$x_n + \lambda_n A x_n + \lambda_n \beta_n \partial \Psi(x_n) \ni x_{n-1},$$

which means that there exist $v_n^1 \in A x_n$ and $v_n^2 \in \partial \Psi(x_n)$ such that

$$(12) \quad x_n + \lambda_n v_n^1 + \lambda_n \beta_n v_n^2 = x_{n-1}.$$

Let us denote by $\{x_n\}$ an arbitrary sequence generated by algorithm (7) (corresponding to an arbitrary choice of the initial data $x_0 \in \mathcal{H}$). Let us recall that $T_{A,C} = A + N_C$. Consequently, for each $u \in D(A) \cap C$ and $w \in T_{A,C}u$ there exists $p \in N_C(u)$ with $w - p \in Au$.

Lemma 2.1. *Take $u \in D(A) \cap C$, $w \in T_{A,C}u$ and let $p \in N_C(u)$ be such that $w - p \in Au$. Then, for each $n \geq 1$, the following inequality holds*

$$\begin{aligned} \|x_n - u\|^2 - \|x_{n-1} - u\|^2 + \|x_{n-1} - x_n\|^2 + \lambda_n \beta_n \Psi(x_n) \\ \leq 2\lambda_n \langle u - x_n, w \rangle + \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right]. \end{aligned}$$

Proof. First observe that

$$\begin{aligned} \|x_{n-1} - u\|^2 &= \|x_n - u\|^2 + \|x_{n-1} - x_n\|^2 + 2 \langle x_n - u, x_{n-1} - x_n \rangle \\ &= \|x_n - u\|^2 + \|x_{n-1} - x_n\|^2 + 2 \langle x_n - u, \lambda_n v_n^1 + \lambda_n \beta_n v_n^2 \rangle \end{aligned}$$

where $v_n^1 \in A x_n$ and $v_n^2 \in \partial \Psi(x_n)$ are given by (12). The monotonicity of A gives

$$(13) \quad \langle (w - p) - v_n^1, u - x_n \rangle \geq 0,$$

while the subdifferential inequality for Ψ yields

$$0 = \Psi(u) \geq \Psi(x_n) + \langle u - x_n, v_n^2 \rangle.$$

Hence

$$\langle u - x_n, \lambda_n v_n^1 + \lambda_n \beta_n v_n^2 \rangle \leq \lambda_n \langle u - x_n, w - p \rangle - \lambda_n \beta_n \Psi(x_n)$$

and so

$$2 \langle u - x_n, \lambda_n v_n^1 + \lambda_n \beta_n v_n^2 \rangle + \lambda_n \beta_n \Psi(x_n) \leq 2\lambda_n \langle u - x_n, w - p \rangle - \lambda_n \beta_n \Psi(x_n).$$

Thus, if we set

$$E_n(u) = \|x_n - u\|^2 - \|x_{n-1} - u\|^2 + \|x_{n-1} - x_n\|^2 + \lambda_n \beta_n \Psi(x_n)$$

we see that

$$\begin{aligned} E_n(u) &= 2 \langle u - x_n, \lambda_n v_n^1 + \lambda_n \beta_n v_n^2 \rangle + \lambda_n \beta_n \Psi(x_n) \\ &\leq 2\lambda_n \langle u - x_n, w - p \rangle - \lambda_n \beta_n \Psi(x_n). \end{aligned}$$

But the right-hand side satisfies

$$\begin{aligned} 2\lambda_n \langle u - x_n, w - p \rangle - \lambda_n \beta_n \Psi(x_n) &= \lambda_n \beta_n \left[\left\langle \frac{2p}{\beta_n}, x_n \right\rangle - \Psi(x_n) - \left\langle \frac{2p}{\beta_n}, u \right\rangle \right] + 2\lambda_n \langle u - x_n, w \rangle \\ &\leq \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right] + 2\lambda_n \langle u - x_n, w \rangle, \end{aligned}$$

where Ψ^* is the Fenchel conjugate of Ψ . Finally

$$E_n(u) \leq \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right] + 2\lambda_n \langle u - x_n, w \rangle,$$

which is the desired inequality. ■

This immediately gives the following:

Corollary 2.2. *Under hypotheses (H_1) and (H_4) we have:*

- i) $\lim_{n \rightarrow \infty} \|x_n - u\|$ exists for each $u \in \mathcal{S}$.
- ii) $\sum_n \|x_{n-1} - x_n\|^2 < +\infty$.
- iii) $\sum_n \lambda_n \beta_n \Psi(x_n) < +\infty$.
- iv) *If $\liminf_{n \rightarrow \infty} \lambda_n \beta_n > 0$ then $\lim_{n \rightarrow \infty} \Psi(x_n) = 0$ and every weak cluster point of the sequence $\{x_n\}$ lies in C .*

Proof. Since $\mathcal{S} \neq \emptyset$ we can take $u \in \mathcal{S}$, $w = 0$ and $p \in N_C(u) \cap (-Au)$, so that Lemma 2.1 yields for each $n \geq 1$

$$\|x_n - u\|^2 - \|x_{n-1} - u\|^2 + \|x_{n-1} - x_n\|^2 + \lambda_n \beta_n \Psi(x_n) \leq \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right].$$

Hypothesis (H_4) immediately gives i), ii) and iii). Part iv) follows from iii) and the weak lower-semicontinuity of Ψ . ■

2.1. Ergodic convergence. We can now properly state and prove the weak ergodic convergence of the sequence $\{x_n\}$ given by (7) (result of type **A** in the introduction). Recall from (10) that $\tau_n = \sum_{k=1}^n \lambda_k$, $z_n = \frac{1}{\tau_n} \sum_{k=1}^n \lambda_k x_k$, and that, by (H_3) , $\tau_n \rightarrow +\infty$ as $n \rightarrow +\infty$.

Theorem 2.3 (Type **A**). *Under hypothesis (\mathbf{H}) the sequence $\{z_n\}$ converges weakly to a point in \mathcal{S} .*

Proof. By virtue of Lemma 1.1 and Corollary 2.2 i) it suffices to prove that each cluster point of the sequence $\{z_n\}$ lies in \mathcal{S} . Take $[u, w] \in T_{A,C}$. By Lemma 2.1, for each $1 \leq k \leq n$

$$\|x_k - u\|^2 - \|x_{k-1} - u\|^2 \leq 2\lambda_k \langle u - x_k, w \rangle + \lambda_k \beta_k \left[\Psi^* \left(\frac{2p}{\beta_k} \right) - \sigma_C \left(\frac{2p}{\beta_k} \right) \right]$$

where positive terms on the left hand side have been omitted (they have no significant contribution since they asymptotically vanish). Summing up these inequalities, $1 \leq k \leq n$, and dividing by $2\tau_n$ we obtain

$$-\frac{\|x_0 - u\|^2}{2\tau_n} \leq \langle u - z_n, w \rangle + \frac{1}{2\tau_n} \sum_{k=1}^n \lambda_k \beta_k \left[\Psi^* \left(\frac{2p}{\beta_k} \right) - \sigma_C \left(\frac{2p}{\beta_k} \right) \right].$$

Passing to the limit we deduce that every weak cluster point z of the sequence $\{z_n\}$ satisfies

$$0 \leq \langle u - z, w \rangle \quad \text{for each } [u, w] \in T_{A,C}.$$

By maximal monotonicity of $T_{A,C}$, (assumption (H_2)), this implies $z \in \mathcal{S}$. ■

2.2. Strong convergence for strongly monotone operators. Recall that A is strongly monotone with parameter $\alpha > 0$ if

$$\langle x^* - y^*, x - y \rangle \geq \alpha \|x - y\|^2$$

whenever $x^* \in Ax$ and $y^* \in Ay$. As a distinctive feature, the set of zeroes of a strongly monotone operator is nonempty, and it is equal to a singleton. We now prove the strong convergence of the sequence $\{x_n\}$ defined by (7) (result of type **B**) when A is strongly monotone.

Theorem 2.4 (Type **B**). *Under hypothesis (H), if the operator A is strongly monotone then the sequence $\{x_n\}$ converges strongly to the unique $u \in \mathcal{S}$.*

Proof. Recall from (12) that there exist $v_n^1 \in Ax_n$ and $v_n^2 \in \partial\Psi(x_n)$ such that

$$x_n + \lambda_n v_n^1 + \lambda_n \beta_n v_n^2 = x_{n-1}.$$

Let A be strongly monotone with parameter α and let u be the unique element in \mathcal{S} . Inequality (13) becomes

$$\langle (w - p) - v_n^1, u - x_n \rangle \geq \alpha \|x_n - u\|^2.$$

We follow the arguments in the proof of Lemma 2.1 to obtain

$$2\alpha \|x_n - u\|^2 + 2\langle u - x_n, \lambda_n v_n^1 + \lambda_n \beta_n v_n^2 \rangle + \lambda_n \beta_n \Psi(x_n) \leq 2\lambda_n \langle u - x_n, w - p \rangle - \lambda_n \beta_n \Psi(x_n).$$

Hence

$$(1 + 2\alpha\lambda_n) \|x_n - u\|^2 - \|x_{n-1} - u\|^2 \leq \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right].$$

Summation gives

$$2\alpha \sum_{n=1}^{\infty} \lambda_n \|x_n - u\|^2 \leq \|x_0 - u\|^2 + \sum_{n=1}^{\infty} \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right] < \infty.$$

Since $\sum_{n=1}^{\infty} \lambda_n = +\infty$ and $\lim_{n \rightarrow \infty} \|x_n - u\|$ exists, we must have $\lim_{n \rightarrow \infty} \|x_n - u\| = 0$. ■

2.3. Weak convergence for subdifferentials. Let $A = \partial\Phi$ be the subdifferential operator associated to some $\Phi \in \Gamma_0(\mathcal{H})$. For each $n \in \mathbb{N}$ let us define $\Omega_n \in \Gamma_0(\mathcal{H})$ by

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

Since the operator $\partial\Phi + \beta_n \partial\Psi$ has been assumed to be maximal monotone, the algorithm can be equivalently written as

$$\begin{aligned} x_n &= (I + \lambda_n \partial\Omega_n)^{-1} x_{n-1} \\ &= \operatorname{argmin}_{\xi \in \mathcal{H}} \left\{ \Phi(\xi) + \beta_n \Psi(\xi) + \frac{1}{2\lambda_n} \|\xi - x_{n-1}\|^2 \right\}. \end{aligned}$$

We are going to prove that the sequence $\{x_n\}$ defined by (7) converges weakly to a point in \mathcal{S} . We shall use the following auxiliary result:

Lemma 2.5. *Assume hypothesis (H) holds. Then for each $u \in \mathcal{S}$*

$$\sum_{n=1}^{\infty} \lambda_n [\Phi(x_n) - \Phi(u) + \beta_n \Psi(x_n)] < +\infty \quad (\text{possibly } -\infty).$$

Proof. The subdifferential inequality gives

$$(14) \quad \Omega_n(u) \geq \Omega_n(x_n) + \frac{1}{\lambda_n} \langle x_{n-1} - x_n, u - x_n \rangle$$

for each $u \in \mathcal{H}$. If $u \in \mathcal{S}$ one has $\Omega_n(u) = \Phi(u)$, hence

$$(15) \quad 2\lambda_n \left[\Phi(x_n) - \Phi(u) + \beta_n \Psi(x_n) \right] \leq \|x_{n-1} - u\|^2 - \|x_n - u\|^2 - \|x_{n-1} - x_n\|^2.$$

By summing these inequalities with respect to $n = 1, 2, \dots$, we obtain

$$\sum_{n=1}^{\infty} \lambda_n [\Phi(x_n) - \Phi(u) + \beta_n \Psi(x_n)] \leq \frac{1}{2} \|x_0 - u\|^2 < \infty,$$

which gives the result. \blacksquare

Now we are in position to prove the following result of type **C**:

Theorem 2.6 (Type **C**). *Let hypotheses **(H)** hold. Assume moreover that one of the following conditions holds*

- i) $\liminf_{n \rightarrow \infty} \lambda_n \beta_n > 0$ and $\liminf_{n \rightarrow \infty} \lambda_n > 0$;
- ii) Hypothesis **(G)** holds and $\liminf_{n \rightarrow \infty} \lambda_n \beta_n > 0$; or
- iii) Hypothesis **(G)** holds and $\lim_{n \rightarrow \infty} \beta_n = \infty$.

Then the sequence $\{x_n\}$ converges weakly to some $x^* \in \mathcal{S}$.

Proof. By Lemma 1.1 and part i) of Corollary 2.2 it suffices to prove that every weak cluster point of the sequence $\{x_n\}$ lies in \mathcal{S} . In view of the weak lower-semicontinuity of Φ and Ψ , we just need to verify that $\lim_{n \rightarrow \infty} \Psi(x_n) = 0$ and $\limsup_{n \rightarrow \infty} \Phi(x_n) \leq \Phi(u)$ for all $u \in \mathcal{S}$.

Assume condition i) holds. The fact that $\lim_{n \rightarrow \infty} \Psi(x_n) = 0$ follows from part iv) in Corollary 2.2. Next, Lemma 2.5 and the fact that $\sum \lambda_n \beta_n \Psi(x_n) < +\infty$ together imply

$$\sum \lambda_n [\Phi(x_n) - \Phi(u)] < +\infty$$

and since $\liminf_{n \rightarrow \infty} \lambda_n > 0$ we conclude that $\limsup_{n \rightarrow \infty} \Phi(x_n) \leq \Phi(u)$ for each $u \in \mathcal{S}$.

In the settings ii) and iii), let us suppose **(G)** holds. As before, we just need to verify that $\lim_{n \rightarrow \infty} \Psi(x_n) = 0$ and $\limsup_{n \rightarrow \infty} \Phi(x_n) \leq \Phi(u)$ for all $u \in \mathcal{S}$. By the definition of the algorithm one has $\Omega_n(x_n) \leq \Omega_n(x_{n-1})$. Setting $u = x_{n-1}$ in (14) and using **(G)** we obtain

$$\Omega_n(x_n) \leq \Omega_n(x_{n-1}) \leq \Omega_{n-1}(x_{n-1}) + K \lambda_{n-1} \beta_{n-1} \Psi(x_{n-1}).$$

Since $\Omega_n(x_n)$ is bounded from below, part iii) in Corollary 2.2 ensures the existence of $\lim_{n \rightarrow \infty} \Omega_n(x_n)$. Lemma 2.5 implies $\lim_{n \rightarrow \infty} \Omega_n(x_n) \leq \Phi(u)$ for any $u \in \mathcal{S}$ and so

$$\limsup_{n \rightarrow \infty} \Phi(x_n) \leq \Phi(u).$$

The fact that $\lim_{n \rightarrow \infty} \Psi(x_n) = 0$ follows from part iv) in Corollary 2.2 if $\liminf_{n \rightarrow \infty} \lambda_n \beta_n > 0$, and from equality $\beta_n \Psi(x_n) = \Omega_n(x_n) - \Phi(x_n)$ if $\lim_{n \rightarrow \infty} \beta_n = \infty$. \blacksquare

3. SPLITTING PROX-PENALIZATION ALGORITHM

Let $\{\beta_n\}$ and $\{\lambda_n\}$ be sequences of positive numbers. In this section we study the alternating algorithm given by

$$(16) \quad \begin{cases} y_n &= (I + \lambda_n A)^{-1} x_{n-1} \\ x_n &= (I + \lambda_n \beta_n \partial \Psi)^{-1} y_n \end{cases}$$

and give the corresponding convergence results. As we did in the preceding section, we shall study the algorithm in its exact form (16) and refer the reader to Section 5 for the general setting which accounts for computational errors. By contrast with the preceding section, where we needed assumption (Q_n) in order the algorithm to be well defined, here the algorithm is well defined without any further assumptions. For any initial data $x_0 \in \mathcal{H}$, algorithm (16) generates a unique sequence $\{x_n\}$.

The following estimation, closely related to Lemma 2.1, will be useful throughout this discussion:

Lemma 3.1. *For $u \in D(A) \cap C$ take $w \in T_{A,C}u$ so that $w = v + p$ for some $v \in Au$ and $p \in N_C(u)$, by definition. For each $n \geq 1$ the following inequality holds:*

$$\begin{aligned} \|x_n - u\|^2 - \|x_{n-1} - u\|^2 + \|x_{n-1} - y_n\|^2 + \frac{1}{2} \|x_n - y_n\|^2 + \lambda_n \beta_n \Psi(x_n) \\ \leq 2\lambda_n \langle w, u - x_n \rangle + \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right] + 2\lambda_n^2 \|v\|^2. \end{aligned}$$

Proof. We have $x_{n-1} - y_n \in \lambda_n A y_n$ and $y_n - x_n \in \lambda_n \beta_n \partial \Psi(x_n)$. The monotonicity of A implies

$$(17) \quad \langle x_{n-1} - y_n, y_n - u \rangle \geq \lambda_n \langle v, y_n - u \rangle,$$

which can be rewritten as

$$(18) \quad \|x_{n-1} - u\|^2 - \|y_n - u\|^2 - \|x_{n-1} - y_n\|^2 \geq 2\lambda_n \langle v, y_n - u \rangle.$$

On the other hand, the subdifferential inequality gives

$$0 = \Psi(u) \geq \Psi(x_n) + \left\langle \frac{y_n - x_n}{\lambda_n \beta_n}, u - x_n \right\rangle.$$

Thus $\langle y_n - x_n, x_n - u \rangle \geq \lambda_n \beta_n \Psi(x_n)$, which is equivalent to

$$(19) \quad \|y_n - u\|^2 - \|x_n - u\|^2 - \|x_n - y_n\|^2 \geq 2\lambda_n \beta_n \Psi(x_n).$$

Adding inequalities (18) and (19) we deduce that

$$\|x_{n-1} - u\|^2 - \|x_n - u\|^2 \geq \|x_{n-1} - y_n\|^2 + \|x_n - y_n\|^2 + 2\lambda_n \langle v, y_n - u \rangle + 2\lambda_n \beta_n \Psi(x_n).$$

But

$$2\lambda_n \langle v, y_n - u \rangle = 2\lambda_n \langle v, x_n - u \rangle + 2\langle \lambda_n v, y_n - x_n \rangle \geq 2\lambda_n \langle v, x_n - u \rangle - 2\lambda_n^2 \|v\|^2 - \frac{1}{2} \|y_n - x_n\|^2.$$

Replacing in the previous inequality we obtain

$$\|x_{n-1} - u\|^2 - \|x_n - u\|^2 \geq \|x_{n-1} - y_n\|^2 + \frac{1}{2} \|x_n - y_n\|^2 + 2\lambda_n \langle v, x_n - u \rangle + 2\lambda_n \beta_n \Psi(x_n) - 2\lambda_n^2 \|v\|^2$$

thus

$$\begin{aligned} \|x_n - u\|^2 - \|x_{n-1} - u\|^2 + \|x_{n-1} - y_n\|^2 + \frac{1}{2} \|x_n - y_n\|^2 + \lambda_n \beta_n \Psi(x_n) \\ \leq 2\lambda_n \langle v, u - x_n \rangle - \lambda_n \beta_n \Psi(x_n) + 2\lambda_n^2 \|v\|^2. \end{aligned}$$

Finally recall that $v = w - p$ so that, setting $D_n = 2\lambda_n \langle v, u - x_n \rangle - \lambda_n \beta_n \Psi(x_n)$,

$$\begin{aligned} D_n &= 2\lambda_n \langle w, u - x_n \rangle + \lambda_n \beta_n \left[\left\langle \frac{2p}{\beta_n}, x_n \right\rangle - \Psi(x_n) - \left\langle \frac{2p}{\beta_n}, u \right\rangle \right] \\ &\leq 2\lambda_n \langle w, u - x_n \rangle + \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right], \end{aligned}$$

which completes the proof. \blacksquare

An immediate consequence of Lemma 3.1 is the following:

Corollary 3.2. *Let hypotheses (H_1) and (H_4) hold. If $\sum \lambda_n^2 < \infty$ then*

- i) *For each $u \in \mathcal{S}$, $\lim_{n \rightarrow \infty} \|x_n - u\|$ exists.*
- ii) *The series $\sum \|x_{n-1} - y_n\|^2$, $\sum \|x_n - y_n\|^2$ and $\sum \lambda_n \beta_n \Psi(x_n)$ are convergent.*

In particular, $\lim_{n \rightarrow \infty} \|x_{n-1} - y_n\| = \lim_{n \rightarrow \infty} \|x_n - y_n\| = \lim_{n \rightarrow \infty} \|x_n - x_{n-1}\| = 0$.

Proof. For $u \in \mathcal{S}$ we can take $w = 0$ in Lemma 3.1 and conclude as in Corollary 2.2. As a byproduct one obtains

$$\sum_{n=1}^{\infty} \|x_{n-1} - y_n\|^2 + \frac{1}{2} \sum_{n=1}^{\infty} \|x_n - y_n\|^2 + \sum_{n=1}^{\infty} \lambda_n \beta_n \Psi(x_n) \leq \|x_0 - u\|^2 + L,$$

where

$$(20) \quad L = \sum_{n=1}^{\infty} \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right] + 2\|v\|^2 \sum_{n=1}^{\infty} \lambda_n^2$$

is finite. \blacksquare

Note that, as a difference with Corollary 2.2, we need to assume here that $\sum \lambda_n^2 < \infty$.

3.1. Ergodic convergence. Keeping the notations of the preceding section let us set $z_n = \frac{1}{\tau_n} \sum_{k=1}^n \lambda_k x_k$, where $\tau_n = \sum_{k=1}^n \lambda_k$. For the alternating algorithm given by (8) we need an additional hypothesis on the step sizes in order to guarantee its stability. The following gives the weak ergodic convergence of the sequence $\{x_n\}$ (result of type **A**):

Theorem 3.3 (Type A). *Under hypothesis **(H)**, if $\sum \lambda_n^2 < \infty$ then the sequence $\{z_n\}$ converges weakly to a point in \mathcal{S} .*

Proof. By Lemma 1.1 and Corollary 3.2, it suffices to prove that every weak cluster point of the sequence $\{z_n\}$ lies in \mathcal{S} . With the notation introduced in Lemma 3.1, if $u \in D(A) \cap C$ we have

$$\|x_n - u\|^2 - \|x_{n-1} - u\|^2 \leq \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right] + 2\lambda_n^2 \|v\|^2 + 2\lambda_n \langle w, u - x_n \rangle.$$

Summing up for $n = 1, \dots, m$, neglecting the positive term on the left-hand side and dividing by $2\tau_m$ we obtain

$$-\frac{\|x_0 - u\|^2}{2\tau_m} \leq \frac{L}{2\tau_m} + \langle w, u - z_m \rangle,$$

where L is given by (20). Therefore, if z_{m_k} converges weakly to z , then $0 \leq \langle w, u - z \rangle$. Since this is true for each $w \in T_{A,C}u$, we conclude from the maximality of $T_{A,C}$ that $z \in \mathcal{S}$. \blacksquare

3.2. Strong convergence for strongly monotone operators. When A is strongly monotone the sequence $\{x_n\}$ defined by (7) converges strongly to the unique $u \in \mathcal{S}$ (result of type **B**).

Theorem 3.4 (Type **B**). *Under hypothesis **(H)**, if A is strongly monotone and $\sum \lambda_n^2 < \infty$ then the sequence $\{x_n\}$ converges strongly to the unique $u \in \mathcal{S}$.*

Proof. Let A be strongly monotone with parameter α and let $\mathcal{S} = \{u\}$. Inequality (18) becomes

$$\|x_{n-1} - u\|^2 - \|y_n - u\|^2 - \|x_{n-1} - y_n\|^2 \geq 2\lambda_n \langle v, y_n - u \rangle + 2\alpha \|y_n - u\|^2.$$

Following the steps in the proof of Lemma 3.1 we obtain

$$2\alpha \lambda_n \|y_n - u\|^2 \leq \|x_{n-1} - u\|^2 - \|x_n - u\|^2 + \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right] + 2\lambda_n^2 \|p\|^2,$$

where $p \in (-Au) \cap N_C(u)$. Whence

$$2\alpha \sum_{n=1}^{\infty} \lambda_n \|y_n - u\|^2 \leq \|x_0 - u\|^2 + 2L,$$

where L is given by (20) with $v = -p$. Inequality (19) gives $\|x_n - u\|^2 \leq \|y_n - u\|^2$ and so

$$\sum_{n=1}^{\infty} \lambda_n \|x_n - u\|^2 < \infty.$$

Since $\lim_{n \rightarrow \infty} \|x_n - u\|$ exists and $\sum_{n=1}^{\infty} \lambda_n = +\infty$, the sequence $\{x_n\}$ must converge strongly to u . ■

3.3. Weak convergence for subdifferentials. Let $A = \partial\Phi$ for $\Phi \in \Gamma_0(\mathcal{H})$. Recall that, by hypothesis (H_2) , $\mathcal{S} = \operatorname{argmin}\{\Phi(x) \mid x \in \operatorname{argmin} \Psi\}$. We shall prove that the sequences $\{x_n\}$ and $\{y_n\}$ defined above converge weakly to an element of \mathcal{S} . We need some preliminary results. First define the energy-like function

$$E_n = \Phi(y_n) + \beta_n \Psi(x_n) + \frac{\|x_n - y_n\|^2}{2\lambda_n}.$$

Notice the dissymmetry in the roles of x_n and y_n as respective arguments of Φ and Ψ . In order to establish the weak convergence of the sequence $\{x_n\}$ we shall use two auxiliary results, which we now prove:

Lemma 3.5. *Under hypothesis **(H)**, for each $u \in \mathcal{S}$ one has $\sum_{n=1}^{\infty} \lambda_n [E_n - \Phi(u)] < +\infty$ (possibly $-\infty$).*

Proof. From the subdifferential inequality and the properties of the inner product we have

$$\|x_{n-1} - u\|^2 - \|y_n - u\|^2 - \|x_{n-1} - y_n\|^2 \geq 2\lambda_n [\Phi(y_n) - \Phi(u)].$$

Now adding this to (19) we obtain

$$2\lambda_n \beta_n \Psi(x_n) + 2\lambda_n [\Phi(y_n) - \Phi(u)] + \|x_n - y_n\|^2 \leq \|x_{n-1} - u\|^2 - \|x_n - u\|^2.$$

This completes the proof. ■

Lemma 3.6. *Let hypotheses **(H)** and **(G)** hold. Assume also that $\sum \lambda_n^2 < +\infty$ and that the sequence $\frac{1}{\lambda_n} - \frac{1}{\lambda_{n-1}}$ is bounded from above. Then $\lim_{n \rightarrow \infty} E_n$ exists and does not exceed the value $\Phi(u)$ for $u \in \mathcal{S}$.*

Proof. From the subdifferential inequality we obtain

$$\begin{aligned} 2\lambda_n\Phi(y_n) &\leq 2\lambda_n\Phi(y_{n-1}) + \|x_{n-1} - y_{n-1}\|^2 - \|x_{n-1} - y_n\|^2 - \|y_n - y_{n-1}\|^2 \\ 2\lambda_n\beta_n\Psi(x_n) &\leq 2\lambda_n\beta_n\Psi(x_{n-1}) + \|x_{n-1} - y_n\|^2 - \|x_n - y_n\|^2 - \|x_n - x_{n-1}\|^2. \end{aligned}$$

Thus

$$\Phi(y_n) + \beta_n\Psi(x_n) \leq \Phi(y_{n-1}) + \beta_n\Psi(x_{n-1}) + \frac{1}{2\lambda_n} [\|x_{n-1} - y_{n-1}\|^2 - \|x_n - y_n\|^2].$$

Now observe that

$$\begin{aligned} \Phi(y_{n-1}) + \beta_n\Psi(x_{n-1}) &\leq \Phi(y_{n-1}) + \beta_{n-1}\Psi(x_{n-1}) + (\beta_n - \beta_{n-1})\Psi(x_{n-1}) \\ &\leq \Phi(y_{n-1}) + \beta_{n-1}\Psi(x_{n-1}) + K\lambda_{n-1}\beta_{n-1}\Psi(x_{n-1}) \end{aligned}$$

by hypothesis **(G)**. On the other hand,

$$\frac{1}{2\lambda_n} [\|x_{n-1} - y_{n-1}\|^2 - \|x_n - y_n\|^2] = b_{n-1} - b_n + \left(\frac{1}{\lambda_n} - \frac{1}{\lambda_{n-1}} \right) \frac{\|x_{n-1} - y_{n-1}\|^2}{2},$$

where $b_n = \frac{1}{2\lambda_n}\|x_n - y_n\|^2$. Hence

$$E_n - E_{n-1} \leq K\lambda_{n-1}\beta_{n-1}\Psi(x_{n-1}) + \left(\frac{1}{\lambda_n} - \frac{1}{\lambda_{n-1}} \right) \frac{\|x_{n-1} - y_{n-1}\|^2}{2}.$$

Finally notice that E_n is bounded from below. Indeed, the sequence $\{y_n\}$ is bounded by Corollary 3.2. Since Φ is convex, $\Phi(y_n)$ is bounded from below and so is E_n . As a consequence, $\lim_{n \rightarrow \infty} E_n$ exists because the positive parts of the terms on the right-hand side of the previous inequality are summable. Lemma 3.5 then implies $\lim_{n \rightarrow \infty} E_n \leq \Phi(u)$ for $u \in \mathcal{S}$. \blacksquare

The hypotheses on the sequence $\{\lambda_n\}$ are satisfied, for instance, if $\lambda_n = \frac{1}{n}$. We now prove that the sequence $\{x_n\}$ converges weakly to a point in \mathcal{S} (result of type **C**).

Theorem 3.7 (Type **C**). *Let hypotheses **(H)** and **(G)** hold. Assume also that $\sum \lambda_n^2 < +\infty$ and that the sequence $\frac{1}{\lambda_n} - \frac{1}{\lambda_{n-1}}$ is bounded from above. Moreover, suppose that either $\liminf_{n \rightarrow \infty} \lambda_n\beta_n > 0$ or that $\lim_{n \rightarrow \infty} \beta_n = +\infty$. Then the sequence $\{x_n\}$ converges weakly to an element of \mathcal{S} .*

Proof. As before, by Lemma 1.1 and Corollary 3.2 it suffices to prove that every weak cluster point of the sequence $\{x_n\}$ lies in \mathcal{S} . Now Lemma 3.5 gives

$$\limsup_{n \rightarrow \infty} \Phi(y_n) \leq \lim_{n \rightarrow \infty} E_n \leq \Phi(u)$$

for all $u \in \mathcal{S}$. This shows that every weak cluster point \bar{y} of the sequence $\{y_n\}$ satisfies $\Phi(\bar{y}) \leq \Phi(u)$ for each $u \in \mathcal{S}$. By Corollary 3.2, $\lim_{n \rightarrow \infty} \|x_n - y_n\| = 0$ so these two sequences have the same cluster points. In order to prove that $\lim_{n \rightarrow \infty} \psi(x_n) = 0$, we use the argument in the proof of Theorem 2.6: if $\liminf_{n \rightarrow \infty} \lambda_n\beta_n > 0$ it follows from part ii) in Corollary 3.2, whereas if $\lim_{n \rightarrow \infty} \beta_n = +\infty$, it follows from the convergence of E_n . This completes the proof. \blacksquare

3.4. The case of M maximal monotone operators. By mixing the techniques developed in the preceding sections with Passty's idea, we are able to generalize the result of Theorem 3.3 to the case of M maximal monotone operators ($M \in \mathbb{N}$). The main result of this section, Theorem 3.9, includes Passty's result (by taking $\Psi = 0$ and $M = 2$) and our Theorem 3.3 (by taking $M = 1$).

Let us give M maximal monotone operators acting on \mathcal{H} , A_1, A_2, \dots, A_M . We are interested in computing a zero of the operator $T_{A,C} = \sum_{m=1}^M A_m + N_C$. In this setting, assumption (H_2) is

naturally replaced by the maximal monotonicity of $T_{A,C} = \sum_{m=1}^M A_m + N_C$.

Given an arbitrary $x_0 \in \mathcal{H}$, let us consider the sequence $\{x_n\}$ generated by the following algorithm:

Given x_{n-1} compute x_n as follows: set $y_n^0 = x_{n-1}$ and find

$$(21) \quad \begin{cases} y_n^m &= (I + \lambda_n A_m)^{-1} y_n^{m-1} & \text{for } m = 1, \dots, M \\ x_n &= (I + \lambda_n \beta_n \partial \Psi)^{-1} y_n^M. \end{cases}$$

For $u \in D(A) \cap C$ take $w \in T_{A,C}u$ and $p \in N_C(u)$ so that $w = p + \sum_{m=1}^M v_m$ where $v_m \in A_m u$ for $m = 1, \dots, M$ by definition.

Lemma 3.8. *With the notation introduced above, the following inequality holds for all $n \geq 1$:*

$$\|x_n - u\|^2 - \|x_{n-1} - u\|^2 \leq 2\lambda_n \langle w, u - x_n \rangle + 2\lambda_n \beta_n \left[\Psi^* \left(\frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] + M(M+1)\lambda_n^2 \sum_{m=1}^M \|v_m\|^2.$$

Proof. For each $m = 1, \dots, M$ one has $y_n^{m-1} - y_n^m \in \lambda_n A_m y_n^m$. The monotonicity of A_m gives

$$(22) \quad \|y_n^{m-1} - u\|^2 - \|y_n^m - u\|^2 - \|y_n^{m-1} - y_n^m\|^2 \geq 2\lambda_n \langle v_m, y_n^m - u \rangle.$$

On the other hand, since $y_n^M - x_n \in \lambda_n \beta_n \Psi(x_n)$, the subdifferential inequality yields

$$(23) \quad \|y_n^M - u\|^2 - \|x_n - u\|^2 - \|x_n - y_n^M\|^2 \geq 2\lambda_n \beta_n \Psi(x_n).$$

Summing up inequalities (22) and adding the result to (23) we obtain

$$\|x_{n-1} - u\|^2 - \|x_n - u\|^2 - \sum_{m=1}^M \|y_n^{m-1} - y_n^m\|^2 - \|x_n - y_n^M\|^2 \geq 2\lambda_n \sum_{m=1}^M \langle v_m, y_n^m - u \rangle + 2\lambda_n \beta_n \Psi(x_n).$$

Since

$$\sum_{m=1}^M \langle v_m, y_n^m - u \rangle = \sum_{m=1}^M \left[\langle v_m, x_n - u \rangle + \langle v_m, y_n^m - x_n \rangle \right] = \langle w - p, x_n - u \rangle + \sum_{m=1}^M \langle v_m, y_n^m - x_n \rangle,$$

we deduce that

$$\|x_n - u\|^2 - \|x_{n-1} - u\|^2 + \|x_n - y_n^M\|^2 + \sum_{m=1}^M \|y_n^{m-1} - y_n^m\|^2$$

$$\begin{aligned}
&\leq 2\lambda_n \left[\langle w - p, u - x_n \rangle - \beta_n \Psi(x_n) \right] + 2\lambda_n \sum_{m=1}^M \langle v_m, x_n - y_n^m \rangle \\
&\leq 2\lambda_n \langle w, u - x_n \rangle + 2\lambda_n \beta_n \left[\left\langle \frac{p}{\beta_n}, x_n \right\rangle - \Psi(x_n) - \left\langle \frac{p}{\beta_n}, u \right\rangle \right] + 2\lambda_n \sum_{m=1}^M \langle v_m, x_n - y_n^m \rangle \\
&\leq 2\lambda_n \langle w, u - x_n \rangle + 2\lambda_n \beta_n \left[\Psi^* \left(\frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] + 2\lambda_n \sum_{m=1}^M \langle v_m, x_n - y_n^m \rangle.
\end{aligned}$$

The proof will be complete if we verify that

$$2\lambda_n \sum_{m=1}^M \langle v_m, x_n - y_n^m \rangle \leq M(M+1)\lambda_n^2 \sum_{m=1}^M \|v_m\|^2 + \|x_n - y_n^M\|^2 + \sum_{m=1}^M \|y_n^{m-1} - y_n^m\|^2.$$

First observe that

$$2\lambda_n \langle v_m, x_n - y_n^m \rangle \leq M(M+1)\lambda_n^2 \|v_m\|^2 + \frac{1}{M(M+1)} \|x_n - y_n^m\|^2.$$

Therefore, we only need to show that

$$\sum_{m=1}^M \|x_n - y_n^m\|^2 \leq M(M+1) \left[\|x_n - y_n^M\|^2 + \sum_{m=1}^M \|y_n^{m-1} - y_n^m\|^2 \right].$$

Indeed,

$$\|x_n - y_n^m\| \leq \|x_n - y_n^M\| + \sum_{k=m+1}^M \|y_n^k - y_n^{k-1}\| \leq \|x_n - y_n^M\| + \sum_{k=1}^M \|y_n^k - y_n^{k-1}\|$$

and so

$$\begin{aligned}
\sum_{m=1}^M \|x_n - y_n^m\|^2 &\leq M \left[\|x_n - y_n^M\| + \sum_{k=1}^M \|y_n^k - y_n^{k-1}\| \right]^2 \\
&\leq M(M+1) \left[\|x_n - y_n^M\|^2 + \sum_{k=1}^M \|y_n^k - y_n^{k-1}\|^2 \right]
\end{aligned}$$

as required. \blacksquare

This immediately implies the convergence of the sequence $\{\|x_n - u\|\}$ for $u \in \mathcal{S}$ under the hypotheses of Corollary 3.2. We are in position to prove the ergodic convergence of the sequence $\{x_n\}$, namely

Theorem 3.9 (Type **A**). *Let $\{x_n\}$ be defined by algorithm (21). Assume hypothesis **(H)** holds and $\sum \lambda_n^2 < +\infty$. Then the sequence $\{z_n\}$ given by $z_n = \frac{1}{\tau_n} \sum_{k=1}^n \lambda_k x_k$, where $\tau_n = \sum_{k=1}^n \lambda_k$, converges weakly to a point in \mathcal{S} .*

Proof. As in the proof of Theorem 3.3, it suffices to show that every weak cluster point of the sequence $\{z_n\}$ lies in \mathcal{S} . Summing up the inequalities in Lemma 3.8 obtained for $n = 1, \dots, N$, then dividing by $2\tau_N$ and letting $N \rightarrow +\infty$, one finally obtains that every weak cluster point z of $\{z_n\}$ satisfies $0 \leq \langle w, u - z \rangle$. Whence $z \in \mathcal{S}$ by maximal monotonicity of $T_{A,C}$. \blacksquare

Observe that this procedure uses the resolvents successively in order to find a point in the set

$$\mathcal{S} = [A_1 + \cdots + A_M + N_C]^{-1} 0.$$

A special case of Theorem 3.9 is obtained by taking $M = 2$, $\Psi = 0$ and $C = H$, namely

Corollary 3.10 (Passty [24]). *Let A_1 and A_2 be two maximal monotone operators such that their sum $A_1 + A_2$ is maximal monotone. Suppose $S = (A_1 + A_2)^{-1}(0) \neq \emptyset$. Let us assume that $\sum_{n=1}^{\infty} \lambda_n = +\infty$ and $\sum_{n=1}^{\infty} \lambda_n^2 < +\infty$. Then any sequence $\{x_n\}$ generated by the algorithm*

$$(24) \quad x_n = (I + \lambda_n A_2)^{-1} (I + \lambda_n A_1)^{-1} x_{n-1}$$

converges weakly in average to some $x \in S$.

Let us remark that hypothesis (H_4) is trivially satisfied, and there is no assumption on β_n .

In the next section we describe an algorithm that provides a point in \mathcal{S} but uses the resolvents in parallel and then computes a barycenter.

4. EXAMPLES

4.1. Prox-projection. Take $\Psi = \delta_C$, where C is a closed convex subset of \mathcal{H} . Then the algorithm described by (16) becomes

$$(25) \quad \begin{cases} y_n &= (I + \lambda_n A)^{-1} x_{n-1} \\ x_n &= P_C y_n, \end{cases}$$

where P_C denotes the projection onto the set C . This is a prox-projection algorithm. In that case $\Psi^* = \sigma_C$ and hypothesis (H_4) is automatically satisfied. Thus weak ergodic convergence holds under the sole assumption $\sum_{n=1}^{\infty} \lambda_n = +\infty$. Weak convergence of the whole sequence $\{x_n\}$ holds for example with $\lambda_n = \frac{1}{n}$. Let us consider the two following special cases of particular interest:

Let $A = \partial\Phi$ be the subdifferential of $\Phi \in \Gamma_0(\mathcal{H})$, and let D be a closed convex subset of \mathcal{H} .

- (1) If $\Phi = \delta_D$, we recover the classical alternating projection method to find points in $C \cap D$ whenever this set is nonempty ($\mathcal{S} \neq \emptyset$). Hypothesis (H) is satisfied trivially because the resolvents do not depend on the parameters λ_n and β_n .
- (2) If $\Phi(x) = \frac{1}{2} \text{dist}(x, D)^2$ then \mathcal{S} is reduced to the point in C which is closest to D . Let us explicit algorithm (25) in that case. We need to compute $(I + \lambda_n A)^{-1} x$ with $A = \nabla\Phi$. Let us notice that $\nabla\Phi$ is the Yosida approximation of index 1 of $\partial\phi$ with $\phi = \delta_D$, namely $\nabla\Phi = (\partial\phi)_1$. By using the resolvent equation $((\partial\phi)_1)_\lambda = (\partial\phi)_{1+\lambda}$ (see [9] Proposition 2.6) we obtain

$$\begin{aligned} (I + \lambda_n A)^{-1} x &= x - \lambda_n (\partial\phi)_{1+\lambda_n}(x) \\ &= x - \frac{\lambda_n}{1+\lambda_n} (x - P_D x) . \\ &= \frac{1}{1+\lambda_n} (x + \lambda_n P_D x) . \end{aligned}$$

Thus the algorithm reads as

$$\begin{cases} y_n &= \frac{1}{1+\lambda_n} (x_{n-1} + \lambda_n P_D x_{n-1}) \\ x_n &= P_C y_n. \end{cases}$$

4.2. Barycenter. Let A_1 and A_2 be maximal monotone operators in a Hilbert space H . Set $\mathcal{H} = H \times H$ and define the (maximal monotone) operator A on \mathcal{H} by $A(x^1, x^2) = (A_1x^1, A_2x^2)$. Let $\Psi(x^1, x^2) = \frac{1}{2}\|x^1 - x^2\|^2$ so that $\nabla\Psi(x^1, x^2) = (x^1 - x^2, x^2 - x^1)$. The algorithm described by (16) gives

$$\begin{cases} y_n^1 + \lambda_n A_1 y_n^1 & \ni x_{n-1}^1 \\ y_n^2 + \lambda_n A_2 y_n^2 & \ni x_{n-1}^2 \end{cases}$$

and

$$\begin{cases} x_n^1 + \lambda_n \beta_n (x_n^1 - x_n^2) & = y_n^1 \\ x_n^2 + \lambda_n \beta_n (x_n^2 - x_n^1) & = y_n^2. \end{cases}$$

Solving for (x_n^1, x_n^2) one obtains

$$(26) \quad \begin{cases} x_n^1 & = (1 - \alpha_n)y_n^1 + \alpha_n y_n^2 \\ x_n^2 & = \alpha_n y_n^1 + (1 - \alpha_n)y_n^2, \end{cases}$$

where $\alpha_n = \frac{\lambda_n \beta_n}{1 + 2\lambda_n \beta_n} \in (0, \frac{1}{2})$. The second step amounts to computing two barycenters for the points found in the first step. For this type of algorithm, the interested reader can consult [19, Lehdili and Lemaire] and the references therein. Observe that condition (H) is satisfied if $\lambda_n \in \ell^2 \setminus \ell^1$ and $\sum \frac{\lambda_n}{\beta_n} < \infty$, so ergodic convergence is granted under these assumptions. In particular, one can take $\beta_n = \frac{1}{\lambda_n}$ and (26) becomes

$$\begin{cases} x_n^1 & = \frac{2}{3}y_n^1 + \frac{1}{3}y_n^2 \\ x_n^2 & = \frac{1}{3}y_n^1 + \frac{2}{3}y_n^2. \end{cases}$$

The limit (\bar{x}^1, \bar{x}^2) satisfies

$$(A_1 \bar{x}^1, A_2 \bar{x}^2) + N_C(\bar{x}^1, \bar{x}^2) \ni 0,$$

where $C = \{(x, y) : x = y\}$. In particular, if $A = \partial\Phi$ with $\Phi(x^1, x^2) = f(x^1) + g(x^2)$ then

$$(\bar{x}^1, \bar{x}^2) \in \operatorname{argmin}\{f(x) + g(y) : x = y\}$$

and so

$$\bar{x}^1 \in \operatorname{argmin}\{f(x) + g(x)\}.$$

One obtains weak convergence (not only ergodic), for instance, if $\lambda_n = \frac{1}{n}$ and $\beta_n = n$.

This procedure can be easily generalized to M variables. Let A_1, \dots, A_M be maximal monotone operators in a Hilbert space H . Set $\mathcal{H} = H^M$ and denote $\vec{x} = (x^1, \dots, x^M)$. Define the operator A on \mathcal{H} by $A(\vec{x}) = (A_1x^1, \dots, A_Mx^M)$ and set $\Psi(\vec{x}) = \frac{1}{2} \sum_{i < j} \|x^i - x^j\|^2$ so that

$$\nabla\Psi(\vec{x}) = M(\vec{x})' - \left[\sum_{k=1}^M x_k \right] \mathbf{1}_{M \times 1},$$

where $\mathbf{1}_{i \times j}$ denotes the matrix of size $i \times j$ whose entries all equal 1. The algorithm described by (16) gives

$$y_n^m + \lambda_n A_m y_n^m \ni x_{n-1}^m \quad \text{for } m = 1, \dots, M$$

and then

$$x_n^m + \lambda_n \beta_n M x_n^m - \lambda_n \beta_n \sum_{k=1}^M x_k = y_n^m \quad \text{for } m = 1, \dots, M.$$

The latter system of equations can be written in matricial form as

$$M_n \vec{x}_n = \vec{y}_n \quad \text{with} \quad M_n = (1 + \lambda_n \beta_n M)I - \lambda_n \beta_n \mathbf{1}_{M \times M}.$$

Simple computations show that M_n is invertible and the (i, j) -th entry of M_n^{-1} is

$$(M_n^{-1})_{i,j} = \begin{cases} \frac{1+\lambda_n\beta_n}{1+\lambda_n\beta_n M} & \text{if } i = j \\ \frac{\lambda_n\beta_n}{1+\lambda_n\beta_n M} & \text{if } i \neq j. \end{cases}$$

As a consequence, M_n^{-1} is symmetric and doubly stochastic, whence

$$\vec{x}_n = M_n^{-1} \vec{y}_n$$

represents the computation of each component x_n^m of \vec{x}_n as a barycenter of the vectors y_n^1, \dots, y_n^M according to the weights given in M_n^{-1} .

4.3. Optimal control of linear systems. Consider an optimal control problem:

$$(27) \quad \min\{\Phi(y, u) : Ay = Bu\},$$

where $A : Y \rightarrow Z$ and $B : U \rightarrow Z$ are linear operators (possibly unbounded), Y, U, Z are Hilbert spaces and $\Phi : Y \times U \rightarrow \mathbf{R} \cup \{+\infty\}$ is a proper, lower-semicontinuous convex function representing the cost to be minimized. A classical approach introduced by J.-L. Lions consists in taking the state equation as a constraint:

$$C = \{(y, u) \in Y \times U : Ay = Bu\}.$$

A natural way to deal with this type of constraint is to use the penalization function

$$\Psi(y, u) = \frac{1}{2} \|Ay - Bu\|_Z^2.$$

Set $\mathcal{H} = Y \times U$ being equipped with the Hilbert product structure. Algorithm (7) takes the form

$$(y_n, u_n) = \underset{(y, u) \in Y \times U}{\operatorname{argmin}} \left\{ \Phi(y, u) + \frac{\beta_n}{2} \|Ay - Bu\|^2 + \frac{1}{2\lambda_n} \left[\|y - y_{n-1}\|_Y^2 + \|u - u_{n-1}\|_U^2 \right] \right\}.$$

In view of (11), in order to fulfill our key hypothesis **(H)** it suffices to verify that Ψ is lower semicontinuous and satisfies

$$(28) \quad \Psi(y, u) \geq \frac{\theta}{2} \operatorname{dist}((y, u), C)^2$$

for some $\theta > 0$ (the distance is taken in $\mathcal{H} = Y \times U$). Then one can take, for instance $\{\lambda_n\} \in \ell^2 \setminus \ell^1$ and $\beta_n = \frac{1}{\lambda_n}$. Assuming Φ to be continuous, the qualification condition (H_2) is satisfied. Therefore, assuming that the set \mathcal{S} of solutions of problem (27) is nonempty we obtain the weak convergence of the sequence (y_n, u_n) to some $(y^*, u^*) \in \mathcal{S}$ as well as the convergence of the values $\Phi(y_n, u_n)$ to the optimal value of the problem.

As an example consider the optimal control of the following elliptic boundary-value problem: Let Ω be an open bounded subset of \mathbf{R}^N . Set $Y = H_0^1(\Omega)$, $U = Z = L^2(\Omega)$ and let

$$C = \{(y, u) \in H_0^1(\Omega) \times L^2(\Omega) : -\Delta y = u\}.$$

This is a closed convex subset of $H_0^1(\Omega) \times L^2(\Omega)$. Let us define $\Psi : H_0^1(\Omega) \times L^2(\Omega) \rightarrow \mathbf{R} \cup \{+\infty\}$

$$\Psi(y, u) = \begin{cases} \|\Delta y + u\|_{L^2(\Omega)}^2 & \text{if } \Delta y \in L^2(\Omega) \\ +\infty & \text{otherwise.} \end{cases}$$

Note that, by Agmon-Douglis-Nirenberg regularity result for elliptic equations, when Ω is sufficiently smooth,

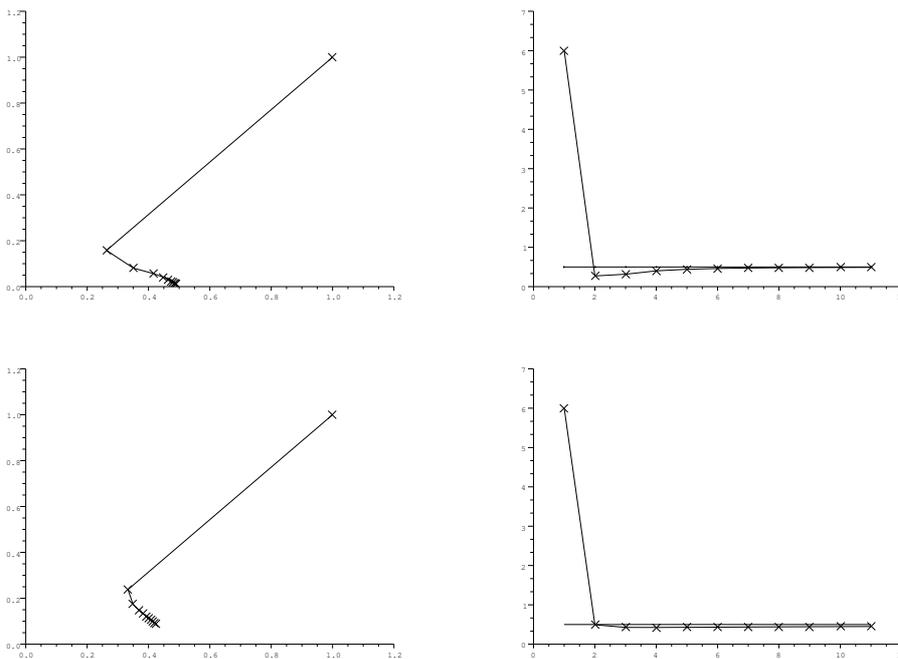
$$y \in H_0^1(\Omega) \text{ and } \Delta y \in L^2(\Omega) \iff y \in H_0^1(\Omega) \cap H^2(\Omega).$$

One can easily verify that Ψ is convex and lower semicontinuous on $H_0^1(\Omega) \times L^2(\Omega)$. We claim that (28) holds for some $\theta > 0$. Indeed, for any $(y, u) \in H_0^1(\Omega) \times L^2(\Omega)$ such that $\Psi(y, u) < +\infty$, i.e. $\Delta y \in L^2(\Omega)$ (otherwise the inequality is trivially satisfied)

$$\begin{aligned} \text{dist}((y, u), C)^2 &\leq \|(y, u) - (-\Delta^{-1}u, u)\|_{H_0^1 \times L^2}^2 = \|y + \Delta^{-1}u\|_{H_0^1}^2 \\ &\leq c\|\Delta y + u\|_{L^2}^2 = c\Psi(y, u) \end{aligned}$$

where c is the operator norm of $(-\Delta)^{-1} : L^2(\Omega) \rightarrow H_0^1(\Omega)$, which can be evaluated using the Poincaré inequality. Finally it suffices to set $\theta = 2/c$ to verify (28). Let us mention that in [16], Kaplan and Tichatschke have been studying numerically a penalization method for optimal control problems like the one above.

4.4. A simple numerical illustration. Take $\mathcal{H} = \mathbb{R} \times \mathbb{R}$. We perform a numerical simulation to find the point in the straight line $2u + v = 1$ that minimizes the function $\Phi(u, v) = 2(u^2 + uv + v^2)$. We define $\Psi(u, v) = \frac{1}{2}|2u + v - 1|^2$. All the hypothesis of theorems 2.6 and 3.7 are satisfied, for instance, with $\lambda_n = \frac{1}{n}$ and $\beta_n = n^2$. The solution set is $\mathcal{S} = \{(\frac{1}{2}, 0)\}$ and the optimal value $\Phi(\frac{1}{2}, 0) = \frac{1}{2}$. We run 10 iterations of algorithm (7) with initial point $(1, 1)$. Figures 1 and 2 show the evolution of the iterates (u_n, v_n) and the values $\Phi(u_n, v_n)$, respectively. We obtain $(0.49, 0.01)$ and 0.49 . The same is done for algorithm (8) to get $(0.42, 0.09)$ and 0.45 . The evolution is shown in Figures 3 and 4.



5. STABILITY AND ROBUSTNESS

5.1. Sensitivity with respect to initial data. We now derive some stability properties of the algorithms described in the preceding sections with respect to perturbations of the initial data.

First, let us consider two trajectories $\{x_n\}$ and $\{\hat{x}_n\}$ emanating respectively from x_0 and \hat{x}_0 following the algorithm given by (7):

$$\begin{aligned} x_n &= \left(I + \lambda_n(A + \beta_n \partial \Psi) \right)^{-1} x_{n-1} \\ \hat{x}_n &= \left(I + \lambda_n(A + \beta_n \partial \Psi) \right)^{-1} \hat{x}_{n-1}. \end{aligned}$$

As a resolvent, the operator $(I + \lambda_n(A + \beta_n \partial \Psi))^{-1}$ is a contraction. Hence,

$$(29) \quad \|x_n - \hat{x}_n\| \leq \|x_{n-1} - \hat{x}_{n-1}\| \leq \dots \leq \|x_0 - \hat{x}_0\|.$$

This dissipation property is characteristic of proximal schemes for monotone operators.

Now recall that $z_n = \frac{1}{\tau_n} \sum_{k=1}^n \lambda_k x_k$, where $\tau_n = \sum_{k=1}^n \lambda_k$. By the triangle inequality we have

$$\|z_n - \hat{z}_n\| \leq \frac{1}{\tau_n} \sum_{k=1}^n \lambda_k \|x_k - \hat{x}_k\| \leq \frac{1}{\tau_n} \sum_{k=1}^n \lambda_k \|x_0 - \hat{x}_0\| = \|x_0 - \hat{x}_0\|.$$

Let x_∞ and \hat{x}_∞ denote the weak limits, as $n \rightarrow \infty$, of the sequences $\{z_n\}$ and $\{\hat{z}_n\}$, respectively. Their existence is guaranteed by Theorem 2.3 and coincide with the limits of the sequences $\{x_n\}$ and $\{\hat{x}_n\}$ whenever the latter exist. By the weak lower-semicontinuity of the norm we obtain

$$\|x_\infty - \hat{x}_\infty\| \leq \|x_0 - \hat{x}_0\|.$$

Finally, define $L : \mathcal{H} \rightarrow \mathcal{H}$ in the following way: for $x \in \mathcal{H}$ compute the sequence $\{x_n\}$ using (7) and $x_0 = x$. Then set $L(x) = \text{w-} \lim_{n \rightarrow \infty} z_n$.

Proposition 5.1. *The function L is nonexpansive.*

In a similar way, if $\{x_n\}$ and $\{\hat{x}_n\}$ are produced using (8), then

$$\begin{aligned} \|x_n - \hat{x}_n\| &= \left\| (I + \lambda_n \beta_n \partial \Psi)^{-1} y_n - (I + \lambda_n \beta_n \partial \Psi)^{-1} \hat{y}_n \right\| \\ &\leq \|y_n - \hat{y}_n\| \\ &= \left\| (I + \lambda_n A)^{-1} x_{n-1} - (I + \lambda_n A)^{-1} \hat{x}_{n-1} \right\| \\ &\leq \|x_{n-1} - \hat{x}_{n-1}\|, \end{aligned}$$

so that the dissipation property (29) holds as well.

If we define $M : \mathcal{H} \rightarrow \mathcal{H}$ by $M(x) = \text{w-} \lim_{n \rightarrow \infty} z_n$, where x_n satisfies (8) we have

Proposition 5.2. *The function M is nonexpansive.*

5.2. Inexact computation of the iterates. Let us assume that we can compute the iterates following the rule (7) only approximately. More precisely, assume the sequence $\{x_n\}$ satisfies

$$(30) \quad \left\| x_n - \left(I + \lambda_n(A + \beta_n \partial \Psi) \right)^{-1} x_{n-1} \right\| \leq \varepsilon_n.$$

We shall prove that if the errors are summable, the convergence properties of the algorithm remain unaltered. To accomplish this, for $n \in \mathbf{N}$ and $x \in \mathcal{H}$ define $U(n, n)x = x$ and

$$U(N, n)x = \prod_{k=n+1}^N \left(I + \lambda_k(A + \beta_k \partial \Psi) \right)^{-1} x$$

for $N \geq n$. Here the product denotes the composition of resolvents. The family of operators $\{U(N, n)\}_{N \geq n}$ is a contracting evolution system, as defined in [2, 3]. That is, it satisfies

- i) $U(n, n)x = x$.

ii) $U(M, N)U(N, n) = U(M, n)$ for $M \geq N \geq n$.

iii) $\|U(N, n)x - U(N, n)y\| \leq \|x - y\|$.

The last property follows from (29).

On the other hand, the sequence $\{x_n\}$ satisfies

$$\begin{aligned} \|x_N - U(N, n)x_n\| &= \|x_N - U(N, N-1)U(N-1, n)x_n\| \\ &\leq \|x_N - U(N, N-1)x_{N-1}\| \\ &\quad + \|U(N, N-1)x_{N-1} - U(N, N-1)U(N-1, n)x_n\| \\ &\leq \varepsilon_N + \|x_{N-1} - U(N-1, n)x_n\|. \end{aligned}$$

By induction one easily shows that

$$\|x_N - U(N, n)x_n\| \leq \sum_{k=n+1}^N \varepsilon_k.$$

If $\sum \varepsilon_k < \infty$ then

$$\lim_{n \rightarrow \infty} \left[\sup_{N \geq n} \|x_N - U(N, n)x_n\| \right] \leq \lim_{n \rightarrow \infty} \left[\sum_{k=n+1}^{\infty} \varepsilon_k \right] = 0,$$

so that $\{x_n\}$ is an almost-orbit of the evolution system U . By Propositions 9 and 12 in [3, Alvarez and Peypouquet]¹, the almost-orbits have the same asymptotic behavior and so we have

Proposition 5.3. *The conclusions of Theorems 2.3, 2.4 and 2.6 remain true under the same hypotheses if x_n satisfies (30) with $\sum \varepsilon_n < \infty$.*

In an analogous way one can consider errors in the computation of the sequence generated by (8):

$$(31) \quad \begin{aligned} \|y_n - (I + \lambda_n A)^{-1}x_{n-1}\| &\leq \varepsilon_n \\ \|x_n - (I + \lambda_n \beta_n \partial \Psi)^{-1}y_n\| &\leq \delta_n. \end{aligned}$$

Following the arguments presented above the reader may easily check the following:

Proposition 5.4. *The conclusions of Theorems 3.3, 3.4 and 3.7 remain true under the same hypotheses if x_n satisfies (31) with $\sum \varepsilon_n < \infty$ and $\sum \delta_n < \infty$.*

Proposition 5.5. *The same is true for Theorem 3.9 if x_n satisfies*

$$\begin{aligned} \|y_n^m - (I + \lambda_n A_m)^{-1}y_n^{m-1}\| &\leq \varepsilon_{m,n} \quad \text{for } m = 1, \dots, M \\ \|x_n - (I + \lambda_n \beta_n \partial \Psi)^{-1}y_n^M\| &\leq \delta_n \end{aligned}$$

with $\sum_{m=1}^M \sum_{n=1}^{\infty} \varepsilon_{m,n} < \infty$ and $\sum_{n=1}^{\infty} \delta_n < \infty$.

REFERENCES

- [1] Alart P. and Lemaire B., Penalization in non-classical convex programming via variational convergence, *Math. Program.*, **51** (1991), 307-331.
- [2] Alvarez F. and Peypouquet J., Asymptotic equivalence and Kobayashi-type estimates for nonautonomous monotone operators in Banach spaces, *Discrete and Continuous Dynamical Systems*, **25** (2009), no 4, 1109-1128.
- [3] Alvarez F. and Peypouquet J., Asymptotic almost-equivalence of abstract evolution systems, arXiv:0904.2157, *paper under review*.

¹The result for the convergence of the whole sequence $\{x_n\}$ can also be found in [2].

- [4] Attouch H., Bolte J., Redont P. and Soubeyran A., Alternating proximal algorithms for weakly coupled convex minimization problems, Applications to dynamical games and PDE's, *J. Convex Anal.*, **15** (2008), no. 3, 485-506.
- [5] Attouch H. and Czarnecki M.-O., Asymptotic behavior of coupled dynamical systems with multiscale aspects, *J. Differential Equations*, **248** (2010), no. 6, 1315-1344.
- [6] Attouch H., Riahi H. and Théra M., Somme ponctuelle d'opérateurs maximaux monotones, *Serdica Math. J.*, **22** (1996), no. 3, 165-190.
- [7] Auslender A., Crouzeix, J.-P. and Fedit P., Penalty-proximal methods in convex programming, *J. Optim. Theory Appl.*, **55** (1987), no. 1, 1-21.
- [8] Bahraoui M.-A. and Lemaire B., Convergence of diagonally stationary sequences in convex optimization, *Set-Valued Anal.*, **2** (1994), no. 1-2, 49-61.
- [9] Brézis H., Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert, North Holland Publishing Company, Amsterdam, 1973.
- [10] Brézis H., Lions P.-L., Produits infinis de résolvantes, *Israel J. Math.*, **29** (1978), 329-345.
- [11] Cabot A., Proximal point algorithm controlled by a slowly vanishing term. Applications to hierarchical minimization, *SIAM J. Optim.*, **15** (2005), no. 2, 555-572.
- [12] Combettes P.-L., Iterative construction of the resolvent of a sum of maximal monotone operators, *Journal of Convex Analysis*, **16**, no. 4, (2009).
- [13] Cominetti R., Courdurier M., Coupling general penalty schemes for convex programming with the steepest descent method and the proximal point algorithm, *SIAM J. Optim.*, **13** (2002), 745-765.
- [14] Cominetti R., Peypouquet J. and Sorin S., Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization, *J. Diff. Equations*, **245** (2008), 3753-3763.
- [15] Hettich R., Kaplan A. and Tichatschke R., Regularized penalty methods for ill-posed optimal control problems, *Control and Cybernetics*, **26**, no. 1, (1997), 5-27.
- [16] Kaplan A. and Tichatschke R., Regularized penalty method for non-coercive parabolic optimal control problems. *Control and Cybernetics*, **27** (1998), no. 1, 5-27.
- [17] Kato T., Nonlinear semi-groups and evolution equations, *J. Math. Soc. Japan*, **19** (1967), 508-520.
- [18] Kobayasi K., Kobayashi Y. and Oharu S., Nonlinear evolution operators in Banach spaces, *Osaka J. Math.*, **21** (1984), 281-310.
- [19] Lehdili N., Lemaire B., The barycentric proximal method. *Comm. Appl. Nonlinear Anal.*, **6** (1999), no. 2, 29-47.
- [20] Lehdili N. and Moudafi A., Combining the proximal algorithm and Tikhonov regularization, *Optimization*, **37** (1996), 239-252.
- [21] Lemaire B., Bounded diagonally stationary sequences in convex optimization, *J. Convex Anal.*, **1** (1994), 75-86.
- [22] Lions P.-L., Une méthode itérative de résolution d'une inéquation variationnelle, *Israel J. Math.*, **31** (1978), 204-208.
- [23] Opial Z., Weak Convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc.*, **73** (1967), 591-597.
- [24] Passty G., Ergodic convergence to a zero of the sum of monotone operators in Hilbert space, *J. Math. Anal. Appl.*, **72** (1979), no. 2, 383-390.
- [25] Peypouquet J., Asymptotic Convergence to the Optimal Value of Diagonal Proximal Iterations in Convex Minimization. *J. Convex Anal.*, **16** (2009), no. 1, 277-286.
- [26] Peypouquet J. and Sorin S., Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. *J. Convex Anal.*, **17** (2010), xx-xx.

INSTITUT DE MATHÉMATIQUES ET MODÉLISATION DE MONTPELLIER, UMR 5149 CNRS, UNIVERSITÉ MONTPELLIER 2, PLACE EUGÈNE BATAILLON, 34095 MONTPELLIER CEDEX 5, FRANCE

E-mail address: attouch@math.univ-montp2.fr, marco@math.univ-montp2.fr

DEPARTAMENTO DE MATEMÁTICA, UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA, AVENIDA ESPAÑA 1680, VALPARAÍSO, CHILE.

E-mail address: juan.peypouquet@usm.cl