



HAL
open science

Coupling Forward-Backward with Penalty Schemes and Parallel Splitting for Constrained Variational Inequalities

Hedy Attouch, Marc-Olivier Czarnecki, Juan Peypouquet

► **To cite this version:**

Hedy Attouch, Marc-Olivier Czarnecki, Juan Peypouquet. Coupling Forward-Backward with Penalty Schemes and Parallel Splitting for Constrained Variational Inequalities. *SIAM Journal on Optimization*, 2011, 21 (4), pp.1251-1274. 10.1137/110820300 . hal-00803596

HAL Id: hal-00803596

<https://hal.science/hal-00803596>

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.

COUPLING FORWARD-BACKWARD WITH PENALTY SCHEMES AND PARALLEL SPLITTING FOR CONSTRAINED VARIATIONAL INEQUALITIES

HÉDY ATTOUCH, MARC-OLIVIER CZARNECKI, AND JUAN PEYPOUQUET

ABSTRACT. We are concerned with the study of a class of forward-backward penalty schemes for solving variational inequalities

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

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 $C \subset \mathcal{H}$. Let $\Psi : \mathcal{H} \rightarrow \mathbb{R}$ be a convex differentiable function whose gradient is Lipschitz continuous, and which acts as a penalization function with respect to the constraint $x \in C$. Given a sequence (β_n) of penalization parameters which tends to infinity, and a sequence of positive time steps $(\lambda_n) \in \ell^2 \setminus \ell^1$, we consider the diagonal forward-backward algorithm

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

Assuming that (β_n) satisfies the growth condition $\limsup_{n \rightarrow \infty} \lambda_n \beta_n < 2/\theta$ (where θ is the Lipschitz constant of $\nabla \Psi$), we obtain weak ergodic convergence of the sequence (x_n) to an equilibrium for a general maximal monotone operator A . We also obtain weak convergence of the whole sequence (x_n) when A is the subdifferential of a proper lower-semicontinuous convex function. As a key ingredient of our analysis, we use the cocoerciveness of the operator $\nabla \Psi$. When specializing our results to coupled systems, we bring new light on Passty's Theorem, and obtain convergence results of new parallel splitting algorithms for variational inequalities involving coupling in the constraint. We also establish robustness and stability results that account for numerical approximation errors. An illustration to compressive sensing is given.

Key words: cocoercive operators; compressive sensing; constrained convex optimization; coupled systems; forward-backward algorithms; hierarchical optimization; maximal monotone operators; parallel splitting methods; Passty's Theorem; penalization methods; variational inequalities.

AMS subject classification. 37N40, 46N10, 49M30, 65K05, 65K10, 90B50, 90C25.

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.

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 . We are concerned with the study of a class of splitting algorithms for solving variational inequalities of the form

$$(1) \quad 0 \in Ax + N_C(x).$$

Specifically, we consider diagonal forward-backward algorithms, where at each step one has to perform a proximal (backward or implicit) step with respect to A and a gradient (forward or explicit) step with respect to a penalization function for the constraint C . As we shall see, these algorithms offer several nice features which make them convenient for numerical purposes.

As a guiding principle of our study, we use the links between algorithms and continuous dissipative dynamical systems, and their asymptotic analysis by Lyapunov methods. Indeed, our algorithms can be derived by time discretization of the continuous nonautonomous differential inclusion

$$(2) \quad 0 \in \dot{x}(t) + Ax(t) + \beta(t)\partial\Psi(x(t)),$$

which has been introduced in [6, Attouch and Czarnecki], and whose trajectories – under certain growth conditions on the function $\beta(\cdot)$ – asymptotically reach equilibria given by (1). In system (2), $\Psi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ acts as an exterior penalization function with respect to the constraint $x \in C$. The corresponding penalization parameter $\beta(t)$ tends to $+\infty$ as $t \rightarrow +\infty$.

This work is closely related and complementary to [7, Attouch, Czarnecki and Peypouquet], where the authors considered the implicit time discretization (backward-backward scheme) of (2)

$$(3) \quad x_n = (I + \lambda_n\beta_n\partial\Psi)^{-1}(I + \lambda_nA)^{-1}x_{n-1},$$

which makes sense for a general convex lower semicontinuous penalization function Ψ , and which combines proximal steps respectively relative to the operator A and the set C (see also [31] for a purely explicit scheme). In [7], convergence results to equilibria have been obtained for algorithm (3) under the key assumption that

$$(4) \quad \sum_{n=1}^{\infty} \lambda_n\beta_n \left[\Psi^* \left(\frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] < +\infty \quad \text{for each } p \in R(N_C),$$

where Ψ^* is the Fenchel conjugate of Ψ and $R(N_C)$ denotes the range of N_C . This condition is satisfied if $\sum_{n=1}^{\infty} \frac{\lambda_n}{\beta_n} < +\infty$ whenever Ψ can be bounded from below by a multiple of the square of the distance to C (see [7]). This is the case, for instance, if $C = \text{Ker}(L)$ and $\Psi(x) = \|Lx\|^2$, where L is a bounded linear operator with closed range (see for example [11, Paragraph II.7]).

By contrast, when the penalization function Ψ is differentiable (which is generally the case), it is rather natural to consider the mixed explicit-implicit discretized version of (2)

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

which provides the following forward-backward algorithm

$$(5) \quad x_n = (I + \lambda_nA)^{-1}(x_{n-1} - \lambda_n\beta_n\nabla\Psi(x_{n-1})),$$

whose study is the central subject of this paper. Some of the ideas in [7] are useful for our purposes but the mixed implicit-explicit character of the algorithm described in (5) poses a different challenge that requires more subtle techniques, in some sense.

The forward-backward schemes (in general) have the advantage of being easier to compute than the backward-backward schemes, which ensures enhanced applicability to real-life problems. Iterations have lower computational cost and can be computed exactly. They naturally lead to parallel splitting methods for solving coupled systems. However, they tend to be less stable than the implicit

ones. An abundant literature has been devoted to the study of the forward-backward algorithms, and their many applications, see [5, Attouch, Briceño and Combettes], [18, Combettes and Wajs] and the references therein. Thus the main original aspect of our approach is to show how such algorithms can be combined with penalization methods.

Diagonal algorithms combine classical algorithms (gradient, proximal, alternate minimization, Newton) with approximations methods (penalization, regularization, vanishing viscosity, among others). A rich literature has been devoted to this subject, see for example [1], [7], [8], [14], [19], [25], [30] and the references therein. A unifying view on these algorithms can be obtained by considering them as time discretization of some corresponding continuous-time nonautonomous differential inclusions

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

(see [2], [3]). In our situation, $A(t)x = Ax + \beta(t)\partial\Psi(x)$ involves multiscale aspects. Our algorithms are naturally linked with diagonal methods involving asymptotically vanishing terms (viscosity methods). Passing from one to the other relies on time rescaling, see [6]. They both involve multiscale aspects and they asymptotically lead to a hierarchical selection principle. For Tikhonov regularization see [25, Lehdili and Moudafi] and [20, Cominetti, Peypouquet and Sorin]. See [14, Cabot] for some further related results and references.

This result is even clearer when $A = \partial\Phi$ is the subdifferential of a proper lower-semicontinuous convex function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. Assuming that some qualification condition holds (for instance if Φ is continuous), the variational inequality (1) is equivalent to

$$(6) \quad x \in \text{Argmin}\{\Phi(z) : z \in \text{Argmin}\Psi\}.$$

Therefore, our results can also be considered as numerical methods for hierarchical minimization.

Our main results. The main set of hypotheses is the following:

$$(H_0) \quad \begin{cases} i) & T_{A,C} = A + N_C \text{ is maximal monotone and } \mathcal{S} = (T_{A,C})^{-1}0 \neq \emptyset; \\ ii) & \text{For each } p \in R(N_C) \text{ (the range of } N_C), \sum_{n=1}^{\infty} \lambda_n \beta_n \left[\Psi^* \left(\frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] < \infty; \\ iii) & (\lambda_n) \in \ell^2 \setminus \ell^1. \end{cases}$$

Depending on the regularity of the function Ψ we shall use a supplementary assumption on the step sizes and penalization parameters. A discussion on these hypotheses will be given later on.

We are able to prove the following results **A**, **B**, **C**: Let (x_n) be a sequence satisfying Algorithm (5) up to a numerical error ε_n (see section 6 for precise details), and let (z_n) be the sequence of weighted averages

$$(7) \quad z_n = \frac{1}{\tau_n} \sum_{k=1}^n \lambda_k x_k, \quad \text{where} \quad \tau_n = \sum_{k=1}^n \lambda_k.$$

A. The sequence (z_n) converges weakly to a solution of (1) (Theorems 5 and 12).

B. If A is strongly monotone then (x_n) converges strongly to the unique solution of (1) (Theorems 7 and 12).

C. If $A = \partial\Phi$ for some proper lower-semicontinuous convex function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, and either Φ or Ψ is inf-compact, then (x_n) converges weakly to a solution of (6) (Theorem 16).

Our results encompass the asymptotic behavior of the well-known proximal point algorithm with variable time step λ_n , see [27, Lions] and [13, Brézis and Lions]. See also [32, Peypouquet and Sorin] for a complete survey on the topic. It also brings new light to the Passty's Theorem (see section 4), which can be considered as a special instance of our algorithm. The fact that $\lambda_n \rightarrow 0$ is a key assumption for our results. Its relevance is discussed in section 4, Remark 15.

Organization of the paper. 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 [28, Opial] and [29, Passty] that are useful for proving weak convergence of a sequence in a Hilbert space without *a priori* knowledge of the limit. Section 2 contains a general abstract result which is at the core of our asymptotic analysis. Section 3 contains our main results of type **A**, **B** for Algorithm (5) when $\nabla\Psi$ is supposed to be Lipschitz continuous. Section 4 makes the links with Passty's Theorem. In section 5, we prove the weak convergence of trajectories generated by our algorithm when $A = \partial\Phi$ is the subdifferential of some proper lower-semicontinuous convex function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ (result of type **C**). In section 6 we address some application issues: we consider inexact version of our algorithm, comment on particular instances for the function Ψ and mention several domains of application including compressive sensing, which we discuss in more detail.

1. PRELIMINARIES

1.1. Some facts of convex analysis and maximal monotone operator theory. Let $\Gamma_0(\mathcal{H})$ denote the set of all proper lower-semicontinuous convex functions on a Hilbert space \mathcal{H} . The norm and inner product in \mathcal{H} are denoted by $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$, respectively. Given $F \in \Gamma_0(\mathcal{H})$ and $x \in \mathcal{H}$, the *subdifferential* of F at x is the set

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

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 $\sigma_C(x^*) = \sup_{y \in C} \langle x^*, y \rangle$. The *normal cone* to C at x is

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

if $x \in C$ and \emptyset otherwise. Observe that $\partial\delta_C = N_C$. We denote the range of N_C by $R(N_C)$.

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. It is convenient to identify a maximal monotone operator A with its graph, thus we equivalently write $x^* \in Ax$ or $[x, x^*] \in A$. The inverse $A^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ of A is defined by $x \in A^{-1}x^* \Leftrightarrow x^* \in Ax$. It is still a maximal monotone operator. 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 [12, Brézis] or [32]). The operator $(I + \lambda A)^{-1}$ is nonexpansive and everywhere defined. It is called the *resolvent* of A of index λ .

Let N_C be the normal cone to the set C and let A be a maximal monotone operator on \mathcal{H} . Suppose the monotone operator $T_{A,C} = A + N_C$ is maximal monotone and

$$\mathcal{S} = (T_{A,C})^{-1}0 \neq \emptyset.$$

By maximal monotonicity of $T_{A,C}$, a point z belongs to \mathcal{S} if, and only if

$$\langle 0 - w, z - u \rangle \geq 0 \text{ for all } (u, w) \in T_{A,C}.$$

Equivalently, a point z belongs to \mathcal{S} if, and only if the following property holds

$$(8) \quad \langle w, u - z \rangle \geq 0 \text{ for all } u \in \text{dom}(T_{A,C}) = C \cap \text{dom}(A) \text{ and all } w \in T_{A,C}u.$$

In the sequel we shall often use this relation as a characterization of the equilibria.

If $A = \partial\Phi$ for some $\Phi \in \Gamma_0(\mathcal{H})$ and if $u \in \mathcal{S}$ then there exists $p \in N_C(u)$ such that $-p \in \partial\Phi(u)$. Hence for each $x \in C$ one has

$$\Phi(x) \geq \Phi(u) + \langle -p, x - u \rangle = \Phi(u) + \sigma_C(p) - \langle p, x \rangle \geq \Phi(u)$$

because

$$(9) \quad p \in N_C(u) \Rightarrow \sigma_C(p) = \langle p, u \rangle.$$

Thus, when $A = \partial\Phi$ the maximal monotonicity of $T_{A,C}$ implies

$$S = \operatorname{Argmin}\{\Phi(x) : x \in C\}.$$

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 maximal monotone operator which is strongly monotone must contain exactly one element.

Finally recall that the subdifferential of a function in $\Gamma_0(\mathcal{H})$ is maximal monotone. For $\Psi \in \Gamma_0(\mathcal{H})$ we denote by Ψ^* the *Fenchel conjugate* of Ψ :

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

1.2. Some useful results. We now state some results that will be used throughout this paper.

The following lemma gathers results from [28, 29] (see also [32]). Though simple, it is a powerful tool for proving weak convergence in Hilbert spaces without *a priori* knowledge of the limit. Let (x_n) be any sequence in \mathcal{H} . Being given a sequence (λ_k) of positive numbers such that $\sum_k \lambda_k = +\infty$, let us define (z_n) as in (7):

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

Lemma 1 (Opial-Passty). *Let F be a nonempty subset of \mathcal{H} and assume that $\lim_{n \rightarrow \infty} \|x_n - x\|$ exists for all $x \in F$. 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 $n \rightarrow +\infty$.*

Next, let us recall the following elementary fact concerning real sequences. We include the proof for the reader's convenience.

Lemma 2. *Let (a_n) , (b_n) and (ε_n) be real sequences. Assume that (a_n) is bounded from below, (b_n) is nonnegative, $(\varepsilon_n) \in \ell^1$ and $a_{n+1} - a_n + b_n \leq \varepsilon_n$ for every $n \in \mathbb{N}$. Then (a_n) converges and $(b_n) \in \ell^1$.*

Proof. Define the sequence (w_n) by $w_n = a_n - \sum_{k=1}^{n-1} \varepsilon_k$. The sequence (w_n) is bounded from below and nonincreasing, hence convergent. It follows that $\lim_{n \rightarrow +\infty} a_n = \sum_{k=1}^{+\infty} \varepsilon_k + \lim_{n \rightarrow +\infty} w_n$. Next observe that $\sum_{k=1}^n b_k \leq a_1 - a_{n+1} + \sum_{k=1}^n \varepsilon_k$ to conclude. \blacksquare

Finally, the Baillon-Haddad Theorem (see [10]), stated below as Lemma 3, shows the relationship between the Lipschitz continuity and the cocoerciveness of the gradient of a convex differentiable function. It is a well established fact that the cocoerciveness of the operator, with respect to which the forward step is performed, is a crucial property for obtaining the convergence of the forward-backward algorithm, see [18], [5].

Lemma 3 (Baillon-Haddad). *Let $\Psi : \mathcal{H} \rightarrow \mathbb{R}$ be a convex differentiable function. The following are equivalent:*

- i) $\nabla\Psi$ is Lipschitz continuous with constant θ .
- ii) $\nabla\Psi$ is $\frac{1}{\theta}$ -cocoercive, which means that, for every x, y belonging to \mathcal{H}

$$\langle \nabla\Psi(y) - \nabla\Psi(x), y - x \rangle \geq \frac{1}{\theta} \|\nabla\Psi(y) - \nabla\Psi(x)\|^2.$$

2. A GENERAL ABSTRACT RESULT

In this section, we prove an abstract convergence result (Theorem 5) which is at the core of the asymptotic analysis of our algorithms. It is valid for a general convex lower semicontinuous penalization function Ψ . In the next section, we shall see that the assumptions of this theorem are satisfied when Ψ is a differentiable function whose gradient $\nabla\Psi$ is θ -Lipschitz continuous and the step sizes and penalization parameters satisfy a simple hypothesis with respect to θ . This approach allows us to better delineate the importance of this regularity assumption and opens the gate to possible further extensions.

Let us fix the notations. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone operator, $\Psi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ a proper lower-semicontinuous convex function with $C = \text{Argmin}(\Psi) \neq \emptyset$ and $\min(\Psi) = 0$. Finally consider (λ_n) , (β_n) two sequences of positive real numbers. We are interested in sequences (x_n) generated by

Algorithm (basic form): Fix $x_1 \in \mathcal{H}$. For each $n \in \mathbb{N}$ set

$$(10) \quad \begin{cases} x_{n+1} &= (I + \lambda_n A)^{-1}(x_n - \lambda_n \beta_n w_n), \\ w_n &\in \partial\Psi(x_n). \end{cases}$$

This is equivalent to

$$(11) \quad x_n - \lambda_n \beta_n w_n \in x_{n+1} + \lambda_n A x_{n+1} \quad \text{and} \quad \frac{x_n - x_{n+1}}{\lambda_n} - \beta_n w_n \in A x_{n+1}.$$

This algorithm is well defined if, for example, Ψ is everywhere defined, which will be the case in the next section. In this section we do not discuss any further the well posedness of the algorithm. Instead, we take for granted the existence of sequences (x_n) satisfying (10).

Set $T_{A,C} = A + N_C$, so that $\text{dom}(T_{A,C}) = C \cap \text{dom}(A)$ and $\mathcal{S} = T_{A,C}^{-1}0$. If $[u, w] \in T_{A,C}$ there exist $v \in Au$ and $p \in N_C(u)$ such that $w = v + p$. Recall that Ψ vanishes on $\text{Argmin}(\Psi) = C$.

Lemma 4. *Take $[u, w] \in T_{A,C}$ so that $w = v + p$ for some $v \in Au$ and $p \in N_C(u)$. The following inequality holds for all $n \in \mathbb{N}$:*

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

Proof. Since $\frac{x_n - x_{n+1}}{\lambda_n} - \beta_n w_n \in A x_{n+1}$ and $v \in Au$, the monotonicity of A implies

$$\langle x_n - x_{n+1} - \lambda_n(\beta_n w_n + v), x_{n+1} - u \rangle \geq 0.$$

Therefore,

$$\langle x_n - x_{n+1}, u - x_{n+1} \rangle \leq \lambda_n \langle \beta_n w_n + v, u - x_{n+1} \rangle.$$

Since

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

we have

$$(12) \quad \begin{aligned} \|x_{n+1} - u\|^2 - \|x_n - u\|^2 &\leq 2\lambda_n \langle \beta_n w_n + v, u - x_{n+1} \rangle - \|x_{n+1} - x_n\|^2 \\ &= 2\lambda_n \langle \beta_n w_n + v, u - x_n \rangle + 2\lambda_n \langle \beta_n w_n + v, x_n - x_{n+1} \rangle - \|x_{n+1} - x_n\|^2 \\ &\leq 2\lambda_n \langle \beta_n w_n + v, u - x_n \rangle + \lambda_n^2 \|\beta_n w_n + v\|^2 \\ &\leq 2\lambda_n \langle \beta_n w_n + v, u - x_n \rangle + 2\lambda_n^2 \beta_n^2 \|w_n\|^2 + 2\lambda_n^2 \|v\|^2. \end{aligned}$$

The proof will be complete if we verify that

$$(13) \quad \langle \beta_n w_n + v, u - x_n \rangle \leq \beta_n \left[\Psi^* \left(\frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] + \langle w, u - x_n \rangle.$$

To this end we first write

$$(14) \quad \langle \beta_n w_n + v, u - x_n \rangle = \beta_n \langle w_n, u - x_n \rangle + \langle v, u - x_n \rangle.$$

Since $u \in C$, and $w_n \in \partial\Psi(x_n)$ the subdifferential inequality for the convex function Ψ gives

$$(15) \quad 0 = \Psi(u) \geq \Psi(x_n) + \langle w_n, u - x_n \rangle.$$

Combining inequalities (14) and (15) and recalling that $v = w - p$ we obtain

$$(16) \quad \begin{aligned} \langle \beta_n w_n + v, u - x_n \rangle &\leq -\beta_n \Psi(x_n) + \langle v, u - x_n \rangle \\ &= -\beta_n \Psi(x_n) + \langle w - p, u - x_n \rangle \\ &= \langle p, x_n \rangle - \beta_n \Psi(x_n) - \langle p, u \rangle + \langle w, u - x_n \rangle \\ &= \beta_n \left[\left\langle \frac{p}{\beta_n}, x_n \right\rangle - \Psi(x_n) - \left\langle \frac{p}{\beta_n}, u \right\rangle \right] + \langle w, u - x_n \rangle \\ &\leq \beta_n \left[\Psi^* \left(\frac{p}{\beta_n} \right) - \left\langle \frac{p}{\beta_n}, u \right\rangle \right] + \langle w, u - x_n \rangle. \end{aligned}$$

Since $p \in N_C(u)$, one has $\langle p, c - u \rangle \leq 0$ for all $c \in C$, thus

$$\left\langle \frac{p}{\beta_n}, u \right\rangle = \sup_{c \in C} \left\langle \frac{p}{\beta_n}, c \right\rangle = \sigma_C \left(\frac{p}{\beta_n} \right).$$

Using this fact in (16) we obtain (13), which completes the proof. \blacksquare

2.1. Ergodic convergence. Consider a sequence (x_n) satisfying (10) and the sequence (z_n) of averages as defined in (7).

Theorem 5. *Let (H_0) hold. If $(\lambda_n \beta_n \|w_n\|) \in \ell^2$ then the sequence (z_n) converges weakly as $n \rightarrow \infty$ to a point in \mathcal{S} .*

Proof. By Opial-Passty Lemma 1, it suffices to prove that the two following properties hold:

O1) for each $u \in \mathcal{S}$ the sequence $(\|x_n - u\|)$ is convergent,

O2) every weak cluster point of the sequence (z_n) lies in \mathcal{S} .

For item **O1)**, let $u \in \mathcal{S}$. Then we can take $w = 0$ in Lemma 4 to obtain

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

Since the right-hand side is summable, and the sequence $(\|x_n - u\|)$ is bounded from below, we immediately deduce that $\lim_{n \rightarrow \infty} \|x_n - u\|$ exists by Lemma 2.

For item **O2)**, as we already observed in (8), since $T_{A,C}$ is maximal monotone, a point z belongs to \mathcal{S} if, and only if, $\langle w, u - z \rangle \geq 0$ for all $[u, w] \in T_{A,C}$.

Take $u \in C \cap \text{dom}(A)$ and $w \in T_{A,C}u$. Recall from Lemma 4 that

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

Summing up for $n = 1, \dots, N$, discarding the nonnegative term $\|x_{N+1} - u\|^2$ and dividing by

$2\tau_N = 2 \sum_{k=1}^N \lambda_k$ we obtain

$$(17) \quad -\frac{\|x_1 - u\|^2}{2\tau_N} \leq \frac{L}{2\tau_N} + \langle w, u - z_N \rangle$$

for some positive constant L . For example, take

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

which is finite in view of our assumptions. By passing to the limit in (17) and using that $\tau_N \rightarrow +\infty$ as $N \rightarrow +\infty$ (because $(\lambda_n) \notin \ell^1$) we obtain

$$\liminf_{n \rightarrow \infty} \langle w, u - z_n \rangle \geq 0.$$

Finally, if some subsequence (z_{n_k}) converges weakly to z , then $0 \leq \langle w, u - z \rangle$. Since this is true for each $w \in T_{A,C}u$ and $T_{A,C}$ is maximal monotone, we conclude that $z \in \mathcal{S}$. \blacksquare

Remark 6. Under the hypotheses of Theorem 5 one has $\lim_{n \rightarrow \infty} \lambda_n \beta_n \|w_n\| = 0$. Thus the sequence (y_n) defined by $y_n = x_n - \lambda_n \beta_n w_n$ satisfies $\lim_{n \rightarrow \infty} \|x_n - y_n\| = 0$. \square

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 maximal monotone operator which is strongly monotone is a singleton (thus nonempty). We now prove the strong convergence of the sequences (x_n) defined by Algorithm (10) when A is a maximal monotone operator which is strongly monotone.

Theorem 7. Let (H_0) hold. If $(\lambda_n \beta_n \|w_n\|) \in \ell^2$ and the operator A is strongly monotone then any sequence (x_n) generated by Algorithm (10) converges strongly to the unique $u \in \mathcal{S}$.

Proof. Recall that $\frac{x_n - x_{n+1}}{\lambda_n} - \beta_n w_n \in Ax_{n+1}$.

Let u be the unique element in \mathcal{S} . Hence there exists $v \in Au$ and $p \in N_C(u)$ such that $v + p = 0$. The strong monotonicity of A implies

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

We follow the arguments in the proof of Lemma 4 (with $w = 0$) to obtain successively

$$2\alpha \lambda_n \|x_{n+1} - u\|^2 + \|x_{n+1} - u\|^2 - \|x_n - u\|^2 \leq 2\lambda_n \langle \beta_n w_n + v, u - x_{n+1} \rangle - \|x_{n+1} - x_n\|^2,$$

and

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

Summation gives

$$2\alpha \sum_{n=1}^{\infty} \lambda_n \|x_{n+1} - u\|^2 \leq \|x_1 - u\|^2 + 2 \sum_{n=1}^{\infty} \lambda_n \beta_n \left[\Psi^* \left(\frac{2p}{\beta_n} \right) - \sigma_C \left(\frac{2p}{\beta_n} \right) \right] + 2 \sum_{n=1}^{\infty} \lambda_n^2 \beta_n^2 \|w_n\|^2 + \lambda_n^2 \|v\|^2 < \infty.$$

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

3. THE COCOERCIVE CASE.

When the penalty function Ψ is smooth enough, we are going to exhibit conditions that only involve the given data of the problem, and which guarantee the summability assumption on the sequence $(\lambda_n \beta_n \|w_n\|)$. Throughout this section we assume that Ψ is a differentiable function whose gradient $\nabla \Psi$ is θ -Lipschitz continuous. By virtue of Lemma 3 this is equivalent to $\nabla \Psi$ being $\frac{1}{\theta}$ -cocoercive. Then $w_n = \nabla \Psi(x_n)$ and our basic algorithm now writes

Algorithm (cocoercive case): Fix $x_1 \in \mathcal{H}$. For each $n \in \mathbb{N}$ set

$$(18) \quad x_{n+1} = (I + \lambda_n A)^{-1}(x_n - \lambda_n \beta_n \nabla \Psi(x_n)).$$

Consider a sequence (x_n) satisfying (18). Lemma 10 below is a refinement of Lemma 4 that will be used later to show our main convergence result (Theorem 12). We shall first prove two intermediate results:

Lemma 8. *Take $u \in C \cap \text{dom}(A)$ and $v \in Au$. Then for all $\eta \geq 0$ and all $n \in \mathbb{N}$ we have*

$$(19) \quad \begin{aligned} \|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \frac{\eta}{1 + \eta} \|x_{n+1} - x_n\|^2 + \frac{2\eta}{1 + \eta} \lambda_n \beta_n \Psi(x_n) \\ \leq \lambda_n \beta_n \left((1 + \eta) \lambda_n \beta_n - \frac{2}{\theta(1 + \eta)} \right) \|w_n\|^2 + 2\lambda_n \langle v, u - x_{n+1} \rangle. \end{aligned}$$

Proof. Since $v \in Au$ and $x_n - x_{n+1} - \lambda_n \beta_n w_n \in \lambda_n A x_{n+1}$, the monotonicity of A implies

$$(20) \quad \langle x_n - x_{n+1} - \lambda_n \beta_n w_n - \lambda_n v, x_{n+1} - u \rangle \geq 0$$

so that

$$\langle x_n - x_{n+1}, u - x_{n+1} \rangle \leq \lambda_n \langle \beta_n w_n + v, u - x_{n+1} \rangle,$$

which in turn gives

$$(21) \quad \|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \|x_{n+1} - x_n\|^2 \leq 2\lambda_n \langle \beta_n w_n + v, u - x_{n+1} \rangle.$$

By developping the right-hand side, we deduce the following inequality

$$(22) \quad \begin{aligned} \|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \|x_{n+1} - x_n\|^2 \\ \leq 2\lambda_n \beta_n \langle w_n, u - x_n \rangle + 2\lambda_n \langle \beta_n w_n, x_n - x_{n+1} \rangle + 2\lambda_n \langle v, u - x_{n+1} \rangle. \end{aligned}$$

We now focus on the first term of the right-hand side of (22). We give two different bounds of the term $\langle w_n, u - x_n \rangle$, each of which will be essential in the following.

First, the cocoercivness of $\nabla \Psi$ writes at points x_n and u

$$\langle \nabla \Psi(x_n) - \nabla \Psi(u), x_n - u \rangle \geq \frac{1}{\theta} \|\nabla \Psi(x_n) - \nabla \Psi(u)\|^2.$$

Since $\nabla \Psi(x_n) = w_n$ and $\nabla \Psi(u) = 0$, we have

$$(23) \quad \langle w_n, u - x_n \rangle \leq -\frac{1}{\theta} \|w_n\|^2.$$

Second, the subdifferential inequality

$$\Psi(u) \geq \Psi(x_n) + \langle \nabla \Psi(x_n), u - x_n \rangle$$

gives, since $\Psi(u) = 0$,

$$(24) \quad \langle w_n, u - x_n \rangle \leq -\Psi(x_n).$$

Take $\eta \geq 0$, and bound the first term on the right-hand side of (22) by using a convex combination of inequalities (23) and (24), namely

$$(25) \quad 2\lambda_n\beta_n\langle w_n, u - x_n \rangle \leq -\frac{2}{\theta(1+\eta)}\lambda_n\beta_n\|w_n\|^2 - \frac{2\eta}{1+\eta}\lambda_n\beta_n\Psi(x_n).$$

For the remaining term $2\lambda_n\beta_n\langle w_n, x_n - x_{n+1} \rangle$, use the identity

$$\begin{aligned} \frac{1}{1+\eta}\|x_{n+1} - x_n + (1+\eta)\lambda_n\beta_n w_n\|^2 &= \frac{1}{1+\eta}\|x_{n+1} - x_n\|^2 + (1+\eta)\lambda_n^2\beta_n^2\|w_n\|^2 \\ &\quad + 2\lambda_n\beta_n\langle w_n, x_{n+1} - x_n \rangle, \end{aligned}$$

to obtain the bound

$$(26) \quad 2\lambda_n\beta_n\langle w_n, x_n - x_{n+1} \rangle \leq \frac{1}{1+\eta}\|x_{n+1} - x_n\|^2 + (1+\eta)\lambda_n^2\beta_n^2\|w_n\|^2.$$

Inequalities (22), (25) and (26) together give

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

and the proof is complete. \blacksquare

Lemma 9. *Assume $\limsup_{n \rightarrow \infty} \lambda_n\beta_n < 2/\theta$. Then there exist $a > 0$, $b > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$, any $u \in C \cap \text{dom}(A)$ and any $v \in Au$ we have*

$$(27) \quad \|x_{n+1} - u\|^2 - \|x_n - u\|^2 + a \left[\|x_{n+1} - x_n\|^2 + \lambda_n\beta_n\Psi(x_n) + \lambda_n\beta_n\|w_n\|^2 \right] \leq 2\lambda_n\langle v, u - x_n \rangle + b\lambda_n^2\|v\|^2.$$

Proof. We begin by analyzing the last term on the right-hand side of (19). Observe that

$$\begin{aligned} 2\lambda_n\langle v, u - x_{n+1} \rangle &= 2\langle \lambda_nv, x_n - x_{n+1} \rangle + 2\lambda_n\langle v, u - x_n \rangle \\ &\leq \frac{\eta}{2(1+\eta)}\|x_{n+1} - x_n\|^2 + \frac{2(1+\eta)}{\eta}\lambda_n^2\|v\|^2 + 2\lambda_n\langle v, u - x_n \rangle. \end{aligned}$$

Replacing this in inequality (19), and adding a term $\frac{\eta}{1+\eta}\lambda_n\beta_n\|w_n\|^2$ to each side, we deduce that

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

To conclude, since $\limsup_{n \rightarrow \infty} \lambda_n\beta_n < 2/\theta$, there exists $N \in \mathbb{N}$ such that $\lambda_n\beta_n < 2/\theta$ for all $n \geq N$. Notice that

$$\lim_{\eta \rightarrow 0} \lambda_n\beta_n \left((1+\eta)\lambda_n\beta_n - \frac{2}{\theta(1+\eta)} + \frac{\eta}{1+\eta} \right) = \lambda_n\beta_n \left(\lambda_n\beta_n - \frac{2}{\theta} \right) < 0.$$

Therefore, it suffices to take $\eta_0 > 0$ small enough, then set

$$a = \frac{\eta_0}{2(1+\eta_0)} \quad \text{and} \quad b = \frac{2(1+\eta_0)}{\eta_0}$$

to obtain (27). ■

Without any loss of generality we may assume that $\lambda_n \beta_n < 2/\theta$ for all $n \in \mathbb{N}$ and that inequality (27) holds for all $n \in \mathbb{N}$.

With the notation of the preceding lemma, for $u \in C \cap \text{dom}(A)$ set

$$D_n(u) = \|x_{n+1} - u\|^2 - \|x_n - u\|^2 + a \left[\|x_{n+1} - x_n\|^2 + \frac{\lambda_n \beta_n}{2} \Psi(x_n) + \lambda_n \beta_n \|w_n\|^2 \right].$$

Notice the difference with the left-hand side of inequality (27).

Lemma 10. *Let $u \in C \cap \text{dom}(A)$. Take $w \in T_{A,C}u$, $v \in Au$ and $p \in N_C(u)$, so that $v = w - p$. The following inequality holds:*

$$(28) \quad D_n(u) \leq \frac{a\lambda_n\beta_n}{2} \left[\Psi^* \left(\frac{4p}{a\beta_n} \right) - \sigma_C \left(\frac{4p}{a\beta_n} \right) \right] + 2\lambda_n \langle w, u - x_n \rangle + b\lambda_n^2 \|v\|^2.$$

Proof. First observe that

$$\begin{aligned} 2\lambda_n \langle v, u - x_n \rangle - \frac{a\lambda_n\beta_n}{2} \Psi(x_n) &= 2\lambda_n \langle w, u - x_n \rangle + 2\lambda_n \langle p, x_n \rangle - \frac{a\lambda_n\beta_n}{2} \Psi(x_n) - 2\lambda_n \langle p, u \rangle \\ &= 2\lambda_n \langle w, u - x_n \rangle + \frac{a\lambda_n\beta_n}{2} \left[\left\langle \frac{4p}{a\beta_n}, x_n \right\rangle - \Psi(x_n) - \left\langle \frac{4p}{a\beta_n}, u \right\rangle \right] \\ &\leq 2\lambda_n \langle w, u - x_n \rangle + \frac{a\lambda_n\beta_n}{2} \left[\Psi^* \left(\frac{4p}{a\beta_n} \right) - \left\langle \frac{4p}{a\beta_n}, u \right\rangle \right]. \end{aligned}$$

Since $\frac{4p}{a\beta_n} \in N_C(u)$, the support function satisfies $\sigma_C(\frac{4p}{a\beta_n}) = \langle \frac{4p}{a\beta_n}, u \rangle$. Whence

$$2\lambda_n \langle v, u - x_n \rangle \leq \frac{a\lambda_n\beta_n}{2} \Psi(x_n) + 2\lambda_n \langle w, u - x_n \rangle + \frac{a\lambda_n\beta_n}{2} \left[\Psi^* \left(\frac{4p}{a\beta_n} \right) - \sigma_C \left(\frac{4p}{a\beta_n} \right) \right].$$

Using inequality (27) and regrouping the terms containing $\Psi(x_n)$ we obtain (28). ■

Proposition 11. *Let (H_0) hold and $\limsup_{n \rightarrow \infty} \lambda_n \beta_n < 2/\theta$. Then we have the following:*

i) *For each $u \in \mathcal{S}$, $\lim_{n \rightarrow \infty} \|x_n - u\|$ exists.*

ii) *The series $\sum_{n=1}^{\infty} \|x_{n+1} - x_n\|^2$, $\sum_{n=1}^{\infty} \lambda_n \beta_n \Psi(x_n)$ and $\sum_{n=1}^{\infty} \lambda_n \beta_n \|w_n\|^2$ are convergent.*

In particular, $\lim_{n \rightarrow \infty} \|x_{n+1} - x_n\| = 0$. If moreover $\liminf_{n \rightarrow \infty} \lambda_n \beta_n > 0$ then $\lim_{n \rightarrow \infty} \Psi(x_n) = \lim_{n \rightarrow \infty} \|w_n\| = 0$ and every weak cluster point of (x_n) lies in C .

Proof. Since $u \in \mathcal{S}$ one can take $w = 0$ in (28). By hypothesis the right-hand side is summable and all the conclusions follow using Lemma 2. ■

Recall that $z_n = \frac{1}{\tau_n} \sum_{k=1}^n \lambda_k x_k$, where $\tau_n = \sum_{k=1}^n \lambda_k$.

Theorem 12. *Let (H_0) hold and $\limsup_{n \rightarrow \infty} \lambda_n \beta_n < 2/\theta$. Then*

- *(weak ergodic convergence) A being a general maximal monotone operator, the sequence (z_n) converges weakly as $n \rightarrow \infty$ to a point in \mathcal{S} .*
- *(strong convergence) A being maximal monotone and strongly monotone, the sequence (x_n) converges strongly as $n \rightarrow \infty$ to a point in \mathcal{S} .*

Proof. By Proposition 11, item ii), we have $\sum_{n=1}^{\infty} \lambda_n \beta_n \|w_n\|^2 < \infty$. Since the sequence $(\lambda_n \beta_n)$ is bounded from above by some constant c , it follows that

$$\sum \lambda_n^2 \beta_n^2 \|w_n\|^2 \leq c \sum \lambda_n \beta_n \|w_n\|^2 < +\infty.$$

Therefore the results follow from Theorems 5 and 7, respectively. \blacksquare

4. STRONG COUPLING AND PASSTY'S THEOREM

Our method sheds new light on Passty's Theorem [29, Theorem 1], which is a classical splitting alternating algorithm for finding zeroes of the sum of two maximal monotone operators. Passty's Theorem, which is itself an extension of a result by P. L. Lions [27], is in the line of the Lie-Trotter-Kato formulae, and can be stated as follows.

Theorem 13 (Passty). *Let A_1 and A_2 be two maximal monotone operators on a Hilbert space, with maximal monotone sum $A_1 + A_2$, and such that $(A_1 + A_2)^{-1}(0) \neq \emptyset$. Let (μ_n) be a sequence of positive real numbers, which belongs to $\ell^2 \setminus \ell^1$. Every sequence (x_n) generated by the algorithm*

$$(29) \quad x_n = (I + \mu_n A_2)^{-1} (I + \mu_n A_1)^{-1} x_{n-1}$$

converges weakly in average to a zero of $A_1 + A_2$.

4.1. Splitting algorithm with strong coupling. With the notations of Passty's Theorem, let us consider $A_1 : \mathcal{H} \rightrightarrows \mathcal{H}$ and $A_2 : \mathcal{H} \rightrightarrows \mathcal{H}$ two maximal monotone operators on a Hilbert space H . Set $\mathcal{H} = H \times H$ and define the (maximal monotone) operator A on \mathcal{H} by

$$A(x^1, x^2) = (A_1 x^1, A_2 x^2).$$

Set $C = \{(x_1, x_2) : x_1 = x_2\}$ and observe that $N_C(x) = C^\perp = \{(p_1, p_2) : p_1 + p_2 = 0\}$ if $x \in C$ and is empty otherwise. We deduce that $(x_1, x_2) \in S \Leftrightarrow x_1 = x_2 = u$ for some $u \in \mathcal{H}$ and $A_1 u + A_2 u \ni 0$. Let Ψ be the (strong)¹ coupling function

$$\Psi(x^1, x^2) = \frac{1}{2} \|x^1 - x^2\|^2$$

so that

$$\nabla \Psi(x^1, x^2) = (x^1 - x^2, x^2 - x^1).$$

Set $x_n = (x_n^1, x_n^2)$, $n \in \mathbb{N}$. In our situation, the cocoercive Algorithm (18) becomes

Algorithm (strong coupling):

$$(30) \quad \begin{cases} x_{n+1}^1 &= (I + \lambda_n A_1)^{-1} (x_n^1 - \lambda_n \beta_n (x_n^1 - x_n^2)) \\ x_{n+1}^2 &= (I + \lambda_n A_2)^{-1} (x_n^2 - \lambda_n \beta_n (x_n^2 - x_n^1)). \end{cases}$$

The limiting case $\lambda_n \beta_n = 1$ is of particular interest since (30) simplifies to

$$\begin{cases} x_{n+1}^1 &= (I + \lambda_n A_1)^{-1} (x_n^2) \\ x_{n+1}^2 &= (I + \lambda_n A_2)^{-1} (x_n^1). \end{cases}$$

This implies

$$(31) \quad \begin{cases} x_{n+1}^1 &= (I + \lambda_n A_1)^{-1} (I + \lambda_{n-1} A_2)^{-1} x_{n-1}^1 \\ x_{n+1}^2 &= (I + \lambda_n A_2)^{-1} (I + \lambda_{n-1} A_1)^{-1} x_{n-1}^2 \end{cases}$$

and we recover Passty's algorithm on the sequence (x_{2n}^2) provided $\lambda_{2n+1} = \lambda_{2n} = \mu_n$. Let us consider the conditions of Theorem 12 in this situation (strong coupling case):

¹For strong coupling versus weak coupling see section 6.

- i) Take $(\lambda_n) \in \ell^2 \setminus \ell^1$.
- ii) Because of the quadratic property of the function Ψ , the condition

$$\sum_{n=1}^{\infty} \lambda_n \beta_n \left[\Psi^* \left(\frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] < \infty \text{ for all } p \in R(N_C)$$

is equivalent to

$$\sum_{n=1}^{\infty} \frac{\lambda_n}{\beta_n} < \infty.$$

- ii) An elementary computation shows that $\nabla \Psi$ is 2-Lipschitz continuous ($\theta = 2$), or equivalently that $\nabla \Psi$ is $\frac{1}{2}$ -cocoercive. The condition $\limsup_{n \rightarrow \infty} \lambda_n \beta_n < 2/\theta$ is equivalent to

$$\limsup_{n \rightarrow \infty} \lambda_n \beta_n < 1.$$

- iv) The maximal monotonicity of $A + N_C$ is equivalent to that of $A_1 + A_2$:

Proof. By Minty's Theorem it suffices to prove that the surjectivity of $I + (A + N_C)$ in $\mathcal{H} \times \mathcal{H}$ is necessary and sufficient for the surjectivity of $2I + (A_1 + A_2)$ in \mathcal{H} .² For necessity, let $(y_1, y_2) \in \mathcal{H} \times \mathcal{H}$ and take $x_1 \in \mathcal{H}$ such that $2x_1 + z_1 + z_2 = y_1 + y_2$, where $z_i \in A_i x_1$. Then set $-p_2 = p_1 = y_1 - x_1 - z_1$. Clearly $(p_1, p_2) \in N_C(x_1, x_1)$ and $(I + (A + N_C))(x_1, x_1) \ni (y_1, y_2)$. For sufficiency, let $y \in \mathcal{H}$ and choose (x_1, x_2) such that $(I + (A + N_C))(x_1, x_2) \ni (y, 0)$. This implies $x_1 = x_2$ and there exists $p \in \mathcal{H}$ such that $x_1 + A_1 x_1 + p \ni y$ and $x_1 + A_2 x_1 - p \ni 0$. Whence $2x_1 + A_1 x_1 + A_2 x_1 \ni y$. ■

Clearly, items i)–iii) are satisfied if one takes $(\lambda_n) \in \ell^2 \setminus \ell^1$ and $\lambda_n \beta_n = \gamma$ for some $\gamma < 1$. By combining the previous results we obtain the following:

Proposition 14 (Strong coupling). *Let A_1 and A_2 be two maximal operators on a Hilbert space \mathcal{H} with maximal monotone sum $A_1 + A_2$, and such that $(A_1 + A_2)^{-1}(0) \neq \emptyset$. Let (λ_n) be a sequence of positive real numbers, which belongs to $\ell^2 \setminus \ell^1$. Take $0 < \gamma < 1$.*

i) Every sequence $((x_n^1, x_n^2))$ generated by the algorithm

$$(32) \quad \begin{cases} x_{n+1}^1 &= (I + \lambda_n A_1)^{-1} ((1 - \gamma)x_n^1 + \gamma x_n^2) \\ x_{n+1}^2 &= (I + \lambda_n A_2)^{-1} ((1 - \gamma)x_n^2 + \gamma x_n^1) \end{cases}$$

converges weakly in average to some element (u, u) with u being equal to a zero of $A_1 + A_2$.

ii) Suppose moreover that $A_1 + A_2$ is strongly monotone. Then the sequences (x_n^1) and (x_n^2) converge strongly to the unique zero of $A_1 + A_2$.

Remark 15. The above algorithm, just like Passty's, is a splitting algorithm: at each step, it only requires the computation of the resolvents of the operators A_1 and A_2 , separately. We point out two noticeable differences between our algorithm and Passty's algorithm.

- (1) The algorithm (32) is a parallel splitting algorithm. By contrast, Passty's algorithm is naturally described as an alternating algorithm. Indeed they are naturally linked as shown by (31) by considering the sequences x_{2n+1}^1 and x_{2n}^1 .
- (2) In Proposition 14 it is assumed that $0 < \gamma < 1$. The case $\gamma = 1$, which corresponds to Passty's Theorem, does not fit directly into our framework, which is based on the cocoerciveness property of the coupling function. Indeed, we shall prove in the next subsection that the case $\gamma = 1$ can be obtained by adapting our approach to this specific situation.

²The maximal monotonicity of an operator is equivalent to the maximal monotonicity of any positive multiple.

- (3) To better understand the importance of the assumption $\lambda_n \rightarrow 0$ in our algorithms, take $\lambda_n \equiv \lambda > 0$, $\gamma = 1/2$, $A_1 = \nabla f_1$, $A_2 = \nabla f_2$ in (32). An elementary computation shows that the stationary points of (32) are the minimizers of the function $f_1(x^1) + f_2(x^2) + \frac{1}{2\lambda} \|x^1 - x^2\|^2$ (which is different from the original problem and approaches it only when $\lambda \rightarrow 0$). This is a specific feature of our approach that makes it different from other decoupling algorithms. \square

Barycenter. In (32) take $A_1 = \partial\delta_{C_1}$ and $A_2 = \partial\delta_{C_2}$, where C_1 and C_2 are nonempty closed convex sets in \mathcal{H} . Then the resolvent $(I + \lambda_n A_i)^{-1}$ is the projection onto C_i . The iteration described in (32) becomes

$$\begin{cases} x_{n+1}^1 &= P_{C_1} \left((1 - \gamma)x_n^1 + \gamma x_n^2 \right) \\ x_{n+1}^2 &= P_{C_2} \left((1 - \gamma)x_n^2 + \gamma x_n^1 \right). \end{cases}$$

Observe that we recover the classical barycentric projection method.

The case of M variables. This procedure can be easily generalized to M variables. Let A_1, \dots, A_M be maximal montone 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 $A(\vec{x}) = (A_1 x^1, \dots, A_M x^M)$ and set $\Psi(\vec{x}) = \frac{1}{2} \sum_{i < j} \|x^i - x^j\|^2$. The

j -th partial derivative of Ψ is given by $\frac{\partial \Psi}{\partial x^j}(\vec{x}) = (M - 1)x^j - \sum_{k \neq j} x^k$ and $\nabla \Psi$ is Lipschitz with constant $\theta = 2(M - 1)$. Let $\lambda_n \beta_n < \frac{2}{\theta} = \frac{1}{M-1}$. If we set $\alpha_n = \lambda_n \beta_n (M - 1)$ then we can write

$$x_{n+1}^j = (I + \lambda_n A_j)^{-1} \left[(1 - \alpha_n)x_n^j + \left(\frac{\alpha_n}{M - 1} \right) \sum_{k \neq j} x_n^k \right], \quad \text{for } j = 1, \dots, M.$$

We now show that, in the specific situation of the Passty's Theorem, the convergence result still holds in the limiting (equality) case. This provides an alternative proof for Theorem 13.

4.2. A parallel approach to Passty's Theorem. Under the considerations and notation of the preceding subsection, setting $\gamma = 1$, we obtain (31):

$$\begin{cases} x_{n+1}^1 &= (I + \lambda_n A_1)^{-1} (I + \lambda_{n-1} A_2)^{-1} x_{n-1}^1 \\ x_{n+1}^2 &= (I + \lambda_n A_2)^{-1} (I + \lambda_{n-1} A_1)^{-1} x_{n-1}^2, \end{cases}$$

which is precisely Passty's algorithm on the sequence (x_{2n}^2) provided $\lambda_{2n+1} = \lambda_{2n} = \mu_n$.

Proof of Theorem 13. With the notation of the preceding subsection, first remark that

$$\nabla \Psi(x) = (x^1 - x^2, x^2 - x^1) = x - sx,$$

where we have written $x = (x^1, x^2)$ and $sx = (x^2, x^1)$.

Let $u \in C \cap \text{dom}(A)$ and $v \in Au$. That is, $u = (u_1, u_1)$ with $u_1 \in \text{dom}(A_1) \cap \text{dom}(A_2)$ and $v = (v_1, v_2)$ with $v_1 \in A_1 u_1$ and $v_2 \in A_2 u_1$.

Since $w_n = \nabla \Psi(x_n) = x_n - sx_n$, inequality (20) gives

$$\langle sx_n - x_{n+1}, x_{n+1} - u \rangle \leq \lambda_n \langle v, u - x_{n+1} \rangle.$$

Since $\|x_n - u\| = \|sx_n - u\|$, we deduce that

$$(33) \quad \|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \|x_{n+1} - sx_n\|^2 \leq 2\lambda_n \langle v, x_n - x_{n+1} \rangle + 2\lambda_n \langle v, u - x_n \rangle.$$

Noticing that

$$0 \leq \|x_{n+1} - sx_n + \lambda_n v\|^2 = \|x_{n+1} - sx_n\|^2 + 2\lambda_n \langle v, x_{n+1} - sx_n \rangle + \lambda_n^2 \|v\|^2,$$

we obtain the two following inequalities

$$(34) \quad \|x_{n+1} - u\|^2 - \|x_n - u\|^2 \leq 2\lambda_n \langle v, u - sx_n \rangle + \lambda_n^2 \|v\|^2$$

$$(35) \quad \|x_{n+1} - u\|^2 - \|x_n - u\|^2 \leq 2\lambda_n \langle v, u - x_{n+1} \rangle.$$

But $su = u$ and $\langle v, u - sx_n \rangle = \langle sv, u - x_n \rangle$, thus

$$(36) \quad \|x_{n+1} - u\|^2 - \|x_n - u\|^2 \leq 2\lambda_n \langle sv, u - x_n \rangle + \lambda_n^2 \|v\|^2.$$

Write (36) at rank $2n + 1$, and (35) at rank $2n$

$$\begin{aligned} \|x_{2n+2} - u\|^2 - \|x_{2n+1} - u\|^2 &\leq 2\lambda_{2n+1} \langle sv, u - x_{2n+1} \rangle + \lambda_{2n+1}^2 \|v\|^2 \\ \|x_{2n+1} - u\|^2 - \|x_{2n} - u\|^2 &\leq 2\lambda_{2n} \langle v, u - x_{2n+1} \rangle. \end{aligned}$$

Since $\lambda_{2n+1} = \lambda_{2n}$, summing the two above inequalities we obtain

$$(37) \quad \|x_{2n+2} - u\|^2 - \|x_{2n} - u\|^2 \leq 2\lambda_{2n} \langle v + sv, u - x_{2n+1} \rangle + \lambda_{2n+1}^2 \|v\|^2.$$

Take $u_1 \in (A_1 + A_2)^{-1}(0)$ and $v = (v_1, -v_1)$. From

$$\|x_{2n+2} - u\|^2 - \|x_{2n} - u\|^2 \leq \lambda_{2n+1}^2 \|v\|^2$$

and Lemma 2 we deduce the convergence of $\|x_{2n} - u\|$. Next, from (34) and (35), we deduce the convergence of $\|x_{2n+1} - u\|$ to the same limit.

We now consider the sequences of averages. First set

$$z_n^{odd} = \frac{\sum_{k=0}^n \lambda_{2k+1} x_{2k+1}}{\sum_{k=0}^n \lambda_{2k+1}}$$

and let z_∞ be a weak cluster point of (z_n^{odd}) as $n \rightarrow \infty$. In view of (37), since $\lambda_{2n} = \lambda_{2n+1}$ we have

$$\frac{\|x_{2n+2} - u\|^2 - \|x_0 - u\|^2}{\sum_{k=0}^n \lambda_{2k+1}} \leq 2 \langle v + sv, u - z_n^{odd} \rangle + \frac{\sum_{k=0}^n \lambda_{2k+1}^2 \|v\|^2}{\sum_{k=0}^n \lambda_{2k+1}}.$$

Passing to the limit we obtain

$$0 \leq \langle v + sv, u - z_\infty \rangle,$$

thus $0 \leq \langle v_1 + v_2, u_1 - z_{\infty,1} \rangle$ and $0 \leq \langle v_1 + v_2, u_1 - z_{\infty,2} \rangle$. This being true for every $u_1 \in \text{dom}(A_1) \cap \text{dom}(A_2)$, and every $v_1 + v_2 \in (A_1 + A_2) u_1$, we deduce that $(A_1 + A_2) z_{\infty,1} \ni 0$ and $(A_1 + A_2) z_{\infty,2} \ni 0$. The Opial Lemma, implies the weak convergence of (z_n^{odd}) to a zero of $(A_1 + A_2) \times (A_1 + A_2)$. Now set

$$z_n^{even} = \frac{\sum_{k=0}^n \lambda_{2k} x_{2k}}{\sum_{k=0}^n \lambda_{2k}}.$$

From (33), we deduce that $\lim_{n \rightarrow \infty} \|x_{n+1} - sx_n\|^2 = 0$. This implies that the sequences (z_n^{even}) , and (sz_n^{odd}) have the same weak limits. We conclude that the sequence of averages

$$z_n = \frac{\sum_{k=0}^n \lambda_k x_k}{\sum_{k=0}^n \lambda_k}$$

also converges to a zero of $(A_1 + A_2) \times (A_1 + A_2)$. ■

Observe that Passty's Theorem is not contained in Theorem 12 but the proof uses the same ideas. An interesting question is whether or not there exist conditions on the function Ψ (more general than strong coupling) such that convergence can still be granted in the limiting case $\lambda_n \beta_n \equiv 2/\theta$.

5. WEAK CONVERGENCE FOR SUBDIFFERENTIALS

In this section, we consider algorithm (18) in the special case where $A = \partial\Phi$ is the subdifferential of a proper lower-semicontinuous convex function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. Assuming that $\partial\Phi + N_C$ is maximal monotone, the solution set \mathcal{S} is equal to

$$\mathcal{S} = (\partial\Phi + N_C)^{-1}(0) = \text{Argmin}_C \Phi.$$

In the following theorem, we are going to prove that the sequence (x_n) generated by algorithm (18) converges weakly to a point in \mathcal{S} . We will need to make an inf-compactness assumption on the functions Φ or Ψ . Recall that a function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is said to be *inf-compact* if,

$$\forall r > 0, \forall \lambda \in \mathbb{R}, \{x \in \mathcal{H} : \|x\| \leq r, f(x) \leq \lambda\} \text{ is relatively compact in } \mathcal{H}.$$

Let $\text{dist}(\cdot, \mathcal{S})$ denote the distance function to the closed convex set \mathcal{S} and set $d(x) = \frac{1}{2} \text{dist}(x, \mathcal{S})^2$. Then d is a convex differentiable function and $\nabla d(x) = x - P_{\mathcal{S}}x$, where $P_{\mathcal{S}}$ denotes the projection onto the set \mathcal{S} .

It is convenient to reformulate algorithm (18) as

$$(38) \quad \varphi_{n+1} := \frac{x_n - x_{n+1}}{\lambda_n} - \beta_n w_n \in \partial\Phi(x_{n+1}),$$

where $w_n = \nabla\Psi(x_n)$.

Theorem 16. *Assume that $\partial\Phi + N_C$ is maximal monotone, $\nabla\Psi$ is θ -Lipschitz continuous, and Φ or Ψ is inf-compact. Suppose that $0 < \liminf_{n \rightarrow \infty} \lambda_n \beta_n \leq \limsup_{n \rightarrow \infty} \lambda_n \beta_n < 2/\theta$, $(\lambda_n) \in \ell^2 \setminus \ell^1$ and $\sum_{n=1}^{\infty} \lambda_n \beta_n \left[\Psi^* \left(\frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] < \infty$ for all $p \in R(N_C)$. Then $\lim_{n \rightarrow \infty} d(x_n) = 0$ and the sequence (x_n) converges weakly as $n \rightarrow \infty$ to a point in \mathcal{S} . Convergence is strong if Ψ is inf-compact.*

Proof. The key part is that $\lim_{n \rightarrow \infty} d(x_n) = 0$. Once this is done, for the weak convergence, since $\lim_{n \rightarrow \infty} d(x_n) = 0$ every weak cluster point of the sequence (x_n) as $n \rightarrow \infty$ belongs to \mathcal{S} . This fact, along with part *i*) of Proposition 11, gives the weak convergence of the sequence by virtue of Opial's Lemma. Finally observe that, since $\lim_{n \rightarrow \infty} \Psi(x_n) = 0$, if Ψ is inf-compact the weak convergence of the relatively compact sequence (x_n) implies its strong convergence.

Let us prove now that $\lim_{n \rightarrow \infty} d(x_n) = 0$. Since d is convex we have

$$(39) \quad \begin{aligned} d(x_n) &\geq d(x_{n+1}) + \langle x_{n+1} - P_{\mathcal{S}}x_{n+1}, x_n - x_{n+1} \rangle \\ &= d(x_{n+1}) + \lambda_n \langle x_{n+1} - P_{\mathcal{S}}x_{n+1}, \varphi_{n+1} \rangle + \lambda_n \beta_n \langle x_{n+1} - P_{\mathcal{S}}x_{n+1}, w_n \rangle, \end{aligned}$$

where φ_{n+1} is given by (38). Let

$$\alpha = \min\{\Phi(z) : z \in C\}.$$

From the convexity of Φ and (38) we deduce that

$$\alpha = \Phi(P_{\mathcal{S}}x_{n+1}) \geq \Phi(x_{n+1}) + \langle \varphi_{n+1}, P_{\mathcal{S}}x_{n+1} - x_{n+1} \rangle$$

and so

$$(40) \quad \langle \varphi_{n+1}, x_{n+1} - P_{\mathcal{S}}x_{n+1} \rangle \geq \Phi(x_{n+1}) - \alpha.$$

Since $\nabla\Psi(P_{\mathcal{S}}x_{n+1}) = 0$, the $1/\theta$ cocoercivity of $\nabla\Psi$ implies

$$\langle w_n, x_n - P_{\mathcal{S}}x_{n+1} \rangle \geq \frac{1}{\theta} \|w_n\|^2,$$

whence

$$(41) \quad \begin{aligned} \langle w_n, x_{n+1} - P_{\mathcal{S}}x_{n+1} \rangle &\geq \frac{1}{\theta} \|w_n\|^2 + \langle w_n, x_{n+1} - x_n \rangle \\ &\geq -\frac{\theta}{4} \|x_{n+1} - x_n\|^2. \end{aligned}$$

We may suppose, without loss of generality, that $\sup_{n \in \mathbb{N}} \lambda_n \beta_n \leq 2/\theta$. Replacing (40) and (41) in (39) we obtain

$$(42) \quad d(x_{n+1}) - d(x_n) + \lambda_n [\Phi(x_{n+1}) - \alpha] \leq \frac{1}{2} \|x_{n+1} - x_n\|^2.$$

Observe that, by virtue of part *ii*) in Proposition 11 the right-hand side of the previous inequality is summable. Since x_{n+1} need not be in C one cannot guarantee that $\Phi(x_{n+1}) - \alpha \geq 0$ for all $n \in \mathbb{N}$. We shall analyze the two possible situations separately, namely

Case I: There exists $n_0 \in \mathbb{N}$ such that $\Phi(x_n) \geq \alpha$ for all $n \geq n_0$.

Case II: For each $n \in \mathbb{N}$ there exists $n' > n$ such that $\Phi(x_{n'}) < \alpha$.

In our analysis we follow the arguments in [14, Cabot], [6, Attouch-Czarnecki] which can be traced back to [9, Baillon-Cominetti].

Case I: For every $n \geq n_0$ we have

$$d(x_{n+1}) - d(x_n) \leq \frac{1}{2} \|x_{n+1} - x_n\|^2.$$

Since $d(x_n)$ is bounded from below and the right-hand side is summable, we conclude that $\lim_{n \rightarrow \infty} d(x_n)$ exists by Lemma 2. In order to verify that this limit must be 0 it suffices to find a subsequence that converges to 0. Let us sum up inequality (42) for $n = n_0, \dots, N$ to obtain

$$d(x_{N+1}) - d(x_{n_0}) + \sum_{n=n_0}^N \lambda_n [\Phi(x_{n+1}) - \alpha] \leq \frac{1}{2} \sum_{n=n_0}^N \|x_{n+1} - x_n\|^2.$$

Letting $N \rightarrow \infty$ we deduce that $\sum_{n=n_0}^{\infty} \lambda_n [\Phi(x_{n+1}) - \alpha] < +\infty$. Since $(\lambda_n) \notin \ell^1$ we must have $\liminf_{n \rightarrow \infty} \Phi(x_n) \leq \alpha$. Consider a subsequence (x_{k_n}) such that

$$\lim_{n \rightarrow \infty} \Phi(x_{k_n}) = \liminf_{n \rightarrow \infty} \Phi(x_n).$$

Clearly the sequence $(\Phi(x_{k_n}))$ is bounded. By part *i*) in Proposition 11, the sequence (x_n) is bounded and so $(\Psi(x_{k_n}))$ is bounded as well. The inf-compactness assumption ensures the existence of subsequence $(x_{k'_n})$ that converges strongly to some \bar{x} , that must belong to C by Proposition 11. From the weak lower-semicontinuity of Φ we deduce that

$$\alpha \leq \Phi(\bar{x}) \leq \lim_{n \rightarrow \infty} \Phi(x_{k'_n}) = \liminf_{n \rightarrow \infty} \Phi(x_n) \leq \alpha.$$

This shows that $\bar{x} \in \mathcal{S}$ and so $\lim_{n \rightarrow \infty} d(x_{k'_n}) = 0$ by the continuity of d .

Case II: For all sufficiently large n the number

$$\tau_n = \max\{k \leq n : \Phi(x_k) < \alpha\}$$

is well defined. Observe that we have $\lim_{n \rightarrow \infty} \tau_n = +\infty$. Take $N \in \mathbb{N}$ (large enough for τ_N to exist).

If $\tau_N < N$ then $\Phi(x_{n+1}) \geq \alpha$ for $n = \tau_N, \dots, N-1$. Inequality (42) then gives

$$d(x_N) - d(x_{\tau_N}) \leq \frac{1}{2} \sum_{n=\tau_N}^{N-1} \|x_{n+1} - x_n\|^2.$$

If $N = \tau_N$ then $d(x_{\tau_N}) = d(x_N)$. In either case we have

$$d(x_N) - d(x_{\tau_N}) \leq \frac{1}{2} \sum_{n=\tau_N}^{\infty} \|x_{n+1} - x_n\|^2$$

and so letting $N \rightarrow \infty$ we deduce that

$$0 \leq \limsup_{n \rightarrow \infty} d(x_n) \leq \limsup_{n \rightarrow \infty} d(x_{\tau_n}).$$

It suffices to prove that $\limsup_{n \rightarrow \infty} d(x_{\tau_n}) = 0$. Since $\Phi(x_{\tau_n}) < \alpha$ for all n one has $\limsup_{n \rightarrow \infty} \Phi(x_{\tau_n}) \leq \alpha$. In particular, the sequence $(\Phi(x_{\tau_n}))$ is bounded. As before, the same is true for $(\Psi(x_{\tau_n}))$ and using the inf-compactness assumption one concludes, as in case I, that a subsequence converges strongly to a point in \mathcal{S} , which guarantees that $\lim_{n \rightarrow \infty} d(x_n) = 0$. \blacksquare

6. APPLICATION ISSUES

In this section we discuss on the implementation and applicability of this method. We begin by establishing that the main conclusions remain true if the iterates are computed inexactly. Next we discuss on some specializations to the case of coupled constraints and when the objective function has a decomposable structure. These appear frequently in real problems. We finish by mentioning several domains of application and develop in further depth the problem of compressive sensing.

6.1. Inexact version. Let us assume that we can compute the iterates following the rule (10) only approximately. More precisely, assume the sequence (\hat{x}_n) satisfies

$$(43) \quad \hat{x}_{n+1} = (I + \lambda_n A)^{-1} [\hat{x}_n - \lambda_n \beta_n \nabla \Psi(\hat{x}_n) + \zeta_n] + \xi_n.$$

We shall establish that if the errors are summable then the sequences computed via (43) have the same asymptotic behavior as those computed exactly using (10). To this end we make use of the following result:

Lemma 17. [2, Proposition 6.1] *Let (F_n) be a family of nonexpansive functions on a Banach space X and assume that every sequence (z_n) satisfying $z_{n+1} = F_n(z_n)$ converges in a sense³. Then so does every sequence (\hat{z}_n) satisfying $\hat{z}_{n+1} = F_n(\hat{z}_n) + \varepsilon_n$ provided $\sum_{n=1}^{\infty} \|\varepsilon_n\| < +\infty$.*

Proposition 18. *Assume that $\lambda_n \beta_n \leq 2/\theta$ for all (sufficiently large) n . The conclusions of Theorems 12 and 16 remain true under the same hypotheses if \hat{x}_n satisfies (43) with $\sum_{n=1}^{\infty} \|\zeta_n\| < +\infty$ and $\sum_{n=1}^{\infty} \|\xi_n\| < +\infty$.*

Proof. For $n \in \mathbb{N}$ write $R_n = (I + \lambda_n A)^{-1}$ and $T_n = I - \lambda_n \beta_n \nabla \Psi$. Since A is monotone we have

$$(44) \quad \|R_n(u) - R_n(v)\| \leq \|u - v\|.$$

On the other hand,

$$(45) \quad \begin{aligned} \|T_n(u) - T_n(v)\|^2 &= \|(u - v) - \lambda_n \beta_n (\nabla \Psi(u) - \nabla \Psi(v))\|^2 \\ &= \|u - v\|^2 + \lambda_n^2 \beta_n^2 \|\nabla \Psi(u) - \nabla \Psi(v)\|^2 - 2\lambda_n \beta_n \langle u - v, \nabla \Psi(u) - \nabla \Psi(v) \rangle \\ &\leq \|u - v\|^2 + \lambda_n \beta_n \left(\lambda_n \beta_n - \frac{2}{\theta} \right) \|\nabla \Psi(u) - \nabla \Psi(v)\|^2 \\ &\leq \|u - v\|^2. \end{aligned}$$

³Here the expression *in a sense* can be replaced by: *strongly*, *weakly*, *strongly in average*, or *weakly in average*, among others (see [2]).

For $n \geq 1$ define F_n and ε_n as follows:

$$\begin{cases} F_{2n-1} = T_n \\ F_{2n} = R_n \end{cases} \quad \text{and} \quad \begin{cases} \varepsilon_{2n-1} = \zeta_n \\ \varepsilon_{2n} = \xi_n. \end{cases}$$

Consider the following nonexpansive algorithm defined by

$$z_{n+1} = F_n(z_n).$$

Then the sequence (x_n) computed via the error-free algorithm (10) satisfies

$$x_n = z_{2n}.$$

By Remark 6, the sequence (y_n) given by

$$y_n = x_n - \lambda_n \beta_n \nabla \Psi(x_n) = z_{2n-1}$$

satisfies $\lim_{n \rightarrow \infty} \|x_n - y_n\| = 0$ so that the sequences (x_n) and (z_n) have the same convergence properties to the extent of the results presented in this paper.

By Lemma 17 the sequence (\hat{z}_n) computed approximately following the rule

$$\hat{z}_{n+1} = F_n(\hat{z}_n) + \varepsilon_n$$

also has the same convergence properties as (z_n) . We conclude by noticing that $\hat{x}_n = \hat{z}_{2n}$. ■

6.2. Some specializations. A special case is when C is of the form

$$C = \{x \in \mathcal{H} : Mx \in D\},$$

where M is a bounded linear operator from \mathcal{H} to another Hilbert space \mathcal{G} and D is a nonempty closed convex subset of \mathcal{G} . This accounts for bilateral (subspace) and unilateral (conic) constraints, among others. Let us denote by $\|M\|$ the operator norm of M . Let $\psi : \mathcal{G} \rightarrow \mathbb{R}$ be a convex differentiable function such that $\nabla \psi$ is τ -Lipschitz and ψ vanishes on $D = \text{Argmin} \psi$. Observe that if D is a hyperplane then $\psi(x) = \|x - P_D(x)\|^2$ has this property and is easy to compute (P_D denotes the orthogonal projection onto D). Set $\Psi = \psi \circ M$. Then Ψ is differentiable and its gradient $\nabla \Psi$ is given by

$$\nabla \Psi(x) = M^* \nabla \psi(Mx),$$

where M^* is the adjoint of M . The function $\nabla \Psi$ is Lipschitz with constant bounded by $\tau \|M\|^2$. This situation can be easily extended to several constraints of the form given above. Let

$$C = \{x \in \mathcal{H} : M_k x \in D_k, \quad k = 1, \dots, p\},$$

where each D_k is a closed convex subset of a Hilbert space \mathcal{G}_k . For $k = 1, \dots, p$ take ψ_k such that $\nabla \psi_k$ is τ_k -Lipschitz and ψ_k vanishes on $D_k = \text{Argmin} \psi_k$ and define $\Psi(x) = \sum_{k=1}^p \psi_k(M_k x)$. Then

$$\nabla \Psi(x) = \sum_{k=1}^p M_k^* \nabla \psi_k(M_k x).$$

The Lipschitz constant of $\nabla \Psi$ can be bounded above by $\sum_{k=1}^p \tau_k \|M_k\|^2$.

Now let $\mathcal{H} = \mathcal{H}_1 \times \dots \times \mathcal{H}_N$ and assume that the objective function Φ can be decomposed as $\Phi(x) = \sum_{i=1}^N \phi_i(x^i)$, where each ϕ_i is proper, lower-semicontinuous and convex. Then

$$\partial \Phi = (\partial \phi_1, \dots, \partial \phi_N).$$

On the other hand, assume also that each M_k can be expressed as $M_k x = \sum_{i=1}^N M_k^i x^i$, where each M_k^i is a bounded linear operator on \mathcal{H}_i . Observe that in this case $\|M_k\|^2 = \sum_{i=1}^N \|M_k^i\|^2$ if all the norms are defined in the canonical fashion. As before, the Lipschitz constant of $\nabla \Psi$ can be bounded by $\sum_{k=1}^p \sum_{i=1}^N \tau_k \|M_k^i\|^2$.

Our algorithm is given by the following system of equations:

$$x_{n+1}^i = (I + \lambda_n \partial \phi_i)^{-1} \left(x_n^i - \lambda_n \beta_n \sum_{k=1}^p (M_k^i)^* \nabla \psi_k \left(\sum_{j=1}^N M_k^j x_n^j \right) + \zeta_n^i \right) + \xi_n^i, \quad \text{for } i = 1, \dots, N.$$

6.3. Domains of application. The method developed in this paper can be applied to find Wardrop equilibria for network flows and construct best approximations for the convex feasibility problem (see [5]), as well as domain decomposition methods for PDE's [4] and optimal control problems [7] or best response dynamics for potential games [4]. Here we discuss on the problem of finding the sparsest solutions of underdetermined systems of equations.

Sparse solutions for underdetermined systems of equations. Let A be a matrix of size $J \times N$ and let $b \in \mathbb{R}^J$. When the system $Ax = b$ is underdetermined, an important problem in signal compression and statistics (see [21, 33]) is to find its sparsest solutions. That is,

$$(\mathcal{P}_0) \quad \min\{\|x\|_0 : Ax = b\},$$

where $\|\cdot\|_0$ denotes the *counting* norm (number of nonzero entries). The convex relaxation of this nonconvex problem is

$$(\mathcal{P}_1) \quad \min\{\|x\|_1 : Ax = b\}.$$

Under some conditions on the matrix A (see [22]) solutions of (\mathcal{P}_0) can be found by solving (\mathcal{P}_1) .

We define $\Phi(x) = \|x\|_1$ and $\Psi(x) = \frac{1}{2}\|Ax - b\|^2$. Then $\nabla \Psi(x) = A^*(Ax - b)$, which is Lipschitz-continuous with constant $\theta = \|A^*A\|$.

As in Remark 6 we write

$$y_n = x_{n-1} - \lambda_n \beta_n \nabla \Psi(x_{n-1}) = x_{n-1} - \lambda_n \beta_n A^*(Ax_{n-1} - b).$$

Now for $s \in \mathbb{R}$ denote $\phi(s) = |s|$ so that $\Phi(x) = \sum_{i=1}^N \phi(x^i)$ and $\partial \Phi(x) = (\partial \phi(x^1), \dots, \partial \phi(x^N))$. The algorithm explicitly reads

$$(46) \quad x_n^i = (I + \lambda_n \partial \phi)^{-1} y_n^i = \begin{cases} y_n^i - \lambda_n & \text{if } y_n^i > \lambda_n \\ y_n^i + \lambda_n & \text{if } y_n^i < -\lambda_n \\ 0 & \text{if } y_n^i \in [-\lambda_n, \lambda_n], \end{cases} \quad \text{for } i = 1, \dots, N.$$

Related ℓ^1 -regularization approaches can be found in [15, 17, 23]. The LASSO method [33] is different because it considers a constraint of the form $\|x\|_1 \leq T$ and minimizes $\|Ax - b\|^2$.

A simple illustration. The following *academic* example aims at illustrating the procedure, certainly not at testing its numerical performance. Consider the system $Ax = b$, where

$$A = \begin{pmatrix} 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

The solutions form a two-dimensional affine subspace of \mathbb{R}^5 among which the sparsest point is

$$\hat{x} = (0 \ 0 \ -1 \ 0 \ 0)'$$

Here $\theta = \|A^*A\| = 4$. We implement our algorithm in SCILAB with $\lambda_n = 1/n$, $\lambda_n \beta_n \equiv 0.49 < 2/\theta$, starting from 10 randomly generated initial points in $[-2, 2]^5$. The average outcome after 50 iterations was

$$\tilde{x} = (0 \ 0 \ -0.9796 \ 0 \ 0)'$$

and the average processing time was 0.05 seconds in a laptop computer with a U9300 Intel(R) Core(TM)2 CPU and 3 GB of RAM.

Remark 19. Some variants:

(1) For the problem with inequality constraints

$$\min\{\|x\|_1 : Ax \leq b\}$$

one defines

$$\Psi(x) = \frac{1}{2} \sum_{j=1}^J \left[\langle r_j, x \rangle - b^j \right]_+^2,$$

where r_j is the j -th row of A and $[s]_+$ denotes the positive part of $s \in \mathbb{R}$. In that case

$$y_n^i = x_{n-1}^i - \lambda_n \beta_n \sum_{j=1}^J r_j^i \left[\langle r_j, x_{n-1} \rangle - b^j \right]_+$$

and one computes x_n^i from y_n^i using (46).

(2) For stable signal recovery [16], where the constraint $Ax = b$ is replaced by $\|Ax - b\| \leq \varepsilon$ one can write $\Psi(x) = \|Ax - b\|^2 - \varepsilon^2$ if $\|Ax - b\| \geq \varepsilon$ and $\Psi(x) = 0$ otherwise. \square

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 Lipschitz evolution systems in Banach spaces, *Nonlinear Anal.*, **73** (2010), no. 9, 3018-3033.
- [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., Briceño-Arias L., and Combettes P.-L., A parallel splitting method for coupled monotone inclusions, *SIAM J. Control Optim.* **48** (2010), no. 5, 3246-3270.
- [6] Attouch H. and Czarnecki M.-O., Asymptotic behavior of coupled dynamical systems with multiscale aspects, *J. Differential Equations*, **248** (2010), no. 6, 1315-1344.
- [7] Attouch H., Czarnecki M.-O. and Peypouquet J., Prox-penalization and splitting methods for constrained variational problems, *SIAM J. Optim.*, **21** (2011), no. 1, 149-173.
- [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] Baillon J.-B. and Cominetti R., A convergence result for nonautonomous subgradient evolution equations and its application to the steepest descent exponential penalty trajectory in linear programming, *J. Funct. Anal.*, **187** (2001), 263-273.
- [10] Baillon J.-B. and Haddad G., Quelques propriétés des opérateurs angle-bornés et n -cycliquement monotones, *Israel J. Math.* **26** (1977), no. 2, 137-150.
- [11] Brézis, H., *Analyse fonctionnelle: théorie et applications*, Dunod, Paris, 1999.
- [12] Brézis H., *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North Holland Publishing Company, Amsterdam, 1973.
- [13] Brézis H., Lions P.-L., Produits infinis de résolvantes, *Israel J. Math.*, **29** (1978), 329-345.
- [14] Cabot A., Proximal point algorithm controlled by a slowly vanishing term. Applications to hierarchical minimization, *SIAM J. Optim.*, **15** (2005), no. 2, 555-572.
- [15] Candès E.-J., Romberg J.-K., Tao T., Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory* **52** (2006), no. 2, 489-509.
- [16] Candès E.-J., Romberg J.-K., Tao T., Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.*, **59** (2006), no. 8, 1207-1223.
- [17] Chen S.-S., Donoho D.-L., Saunders M.-A., Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, **20** (1998), no. 1, 33-61.
- [18] Combettes P.-L., Wajs V.R., Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, **4** (2005), 1168-1200.
- [19] 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.

- [20] Cominetti R., Peyrouquet 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.
- [21] Donoho D.-L. Compressed sensing, *IEEE Trans. Inform. Theory* **52** (2006), no. 4, 1289-1306.
- [22] Donoho D.-L., Tanner J., Sparse nonnegative solutions of underdetermined linear equations by linear programming, *Proceedings of the National Academy of Sciences*, **102** (2005), no. 27, 9446-9451.
- [23] Friedlander M.-P., Tseng P., Exact regularization of convex programs, *SIAM J. Optim.*, **18** (2007), no. 4, 1326-1350.
- [24] Kaplan A. and Tichatschke R., Regularized penalty method for non-coercive parabolic optimal control problems, *Control and Cybernetics*, **27** (1998), no. 1, 5-27.
- [25] Lehdili N. and Moudafi A., Combining the proximal algorithm and Tikhonov regularization, *Optimization*, **37** (1996), 239-252.
- [26] Lemaire B., Bounded diagonally stationary sequences in convex optimization, *J. Convex Anal.*, **1** (1994), 75-86.
- [27] Lions P.-L., Une méthode itérative de résolution d'une inéquation variationnelle, *Israel J. Math.*, **31** (1978), 204-208.
- [28] Opial Z., Weak convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc.*, **73** (1967), 591-597.
- [29] 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.
- [30] Peyrouquet J., Asymptotic convergence to the optimal value of diagonal proximal iterations in convex minimization, *J. Convex Anal.*, **16** (2009), no. 1, 277-286.
- [31] Peyrouquet J., Coupling the gradient method with a general exterior penalization scheme for convex minimization, *Paper under review*.
- [32] Peyrouquet J. and Sorin S., Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time, *J. Convex Anal.*, **17** (2010), 1113-1163.
- [33] Tibshirani R., Regression shrinkage and selection via the Lasso, *Journal of the Royal Statistical Society, Series B*, **58** (1996), no. 1, 267-288.

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`