ESAIM: Control, Optimisation and Calculus of Variations September 1997, Vol. 2, pp. 281–306
URL: http://www.emath.fr/cocv/

AN EXTENSION OF THE AUXILIARY PROBLEM PRINCIPLE TO NONSYMMETRIC AUXILIARY OPERATORS

A. RENAUD AND G. COHEN

Abstract. To find a zero of a maximal monotone operator, an extension of the Auxiliary Problem Principle to nonsymmetric auxiliary operators is proposed. The main convergence result supposes a relationship between the main operator and the nonsymmetric component of the auxiliary operator. When applied to the particular case of convex-concave functions, this result implies the convergence of the parallel version of the Arrow-Hurwicz algorithm under the assumptions of Lipschitz and partial Dunn properties of the main operator. The latter is systematically enforced by partial regularization.

In the strongly monotone case, it is shown that the convergence is linear in the average. Moreover, if the symmetric part of the auxiliary operator is linear, the Lipschitz property of the inverse suffices to ensure a linear convergence rate in the average.

1. Introduction

Given a maximal monotone operator Ψ over a Hilbert space X, which may be multivalued (that is, $\Psi(x)$ may be a subset of X rather than a single point), consider the problem of finding a zero of Ψ (that is, a point $x^\sharp \in X$ such that $0 \in \Psi(x^\sharp)$). This paper is devoted to studying the following general algorithmic scheme to solve this problem: at stage $k+1$, knowing x^k, obtain x^{k+1} by solving another zero-finding problem, namely,

$$0 \in \left(\frac{1}{\varepsilon} \left(\Xi(x^{k+1}) - \Xi(x^k) \right) + \Psi(x^{k+1}) \right) \tag{1.1}$$

where

- Ξ is a *single-valued* auxiliary (that is, user chosen) operator which may be nonsymmetric (i.e. is not necessarily the derivative of a mapping);
- ε is a positive scalar (again user chosen) playing the role of a stepsize as in a gradient algorithm (which, by the way, can be cast in the framework of (1.1)).

This scheme appears to be a generalization of the proximal algorithm (see [16]) introduced by Martinet [19] and extended by Rockafellar [25] to maximal monotone operators (in these works, Ξ is simply the identity). When choosing a symmetric auxiliary operator (i.e. when Ξ is the derivative K'

Research Engineer, Électricité de France, Direction des Études et Recherches, Service Études de Réseaux. E-mail: renaud@clr34el.der.edf.fr.

Senior Scientist, Centre Automatique et Systèmes, École des Mines de Paris, and Scientific Advisor, INRIA, Rocquencourt, France. E-mail: cohen@cas.ensmp.fr.

Received by the journal September 9, 1996. Accepted for publication June 5, 1997.

of a real-valued mapping K), Algorithm (1.1) yields the general form of the proximal algorithm obtained by using Bregman functions [3]. This algorithm has been studied by Censor and Zenios [5] and by Chen and Teboulle [6] in the symmetric case — when Ψ is the subdifferential of a convex function. Eckstein [12] also considered this type of algorithm in the general case and applied it to decomposition.

Algorithm (1.1) may also be considered as a generalization to nonsymmetric auxiliary operators of the Auxiliary Problem Principle (APP) introduced by Cohen ([7], [8]). In the symmetric case, for example, to solve

$$\min_{x\in X^{\mathrm{ad}}} J(x) \tag{1.2}$$

where $J = J_1+J_2$, with J_1 differentiable, over the closed convex subset X^{ad}, this principle proposes the general form of algorithm: knowing x^k, derive x^{k+1} from the resolution of

$$\min_{x\in X^{\mathrm{ad}}} \left(\frac{1}{\varepsilon}\left(K(x) - \left\langle K'(x^k), x\right\rangle\right) + \left\langle J_1'(x^k), x\right\rangle + J_2(x)\right) \tag{1.3}$$

where K is a real-valued strongly convex function. For the auxiliary operator $\Xi = K' - \varepsilon J_1'$, Algorithm (1.1) applied to the subdifferential of $J + I_{X^{\mathrm{ad}}}$, where $I_{X^{\mathrm{ad}}}$ is the indicator function of X^{ad}, yields Algorithm (1.3). For $K : x \mapsto (1/2)\|x\|^2$ and $J = J_1$, this algorithm turns out to be an explicit gradient algorithm with projection on X^{ad}. For J reduced to J_2, it is nothing else but a proximal algorithm [19], or otherwise stated, an implicit gradient algorithm. This principle thus proposes a form of continuum between proximal and gradient algorithms. Provided that K is strongly convex with constant b, if J_1' is L-Lipschitz continuous and if J_1 and J_2 are convex l.s.c., the convergence of this type algorithm has been proved by Cohen [8] for $0 < \varepsilon < 2b/L$. The assumptions upon J_1 are strongly related to those which ensure the convergence of the explicit gradient algorithm; those upon J_2 are akin to the assumptions which ensures the convergence of the proximal algorithm.

This partial linearization framework has a particular interest to build up decomposition algorithms (see [7]). Assume that $X = \times_{h\in\mathcal{H}} X_h$ where $\mathcal{H}$ is a finite set. If $X^{\mathrm{ad}} = \times_{h\in\mathcal{H}} X_h^{\mathrm{ad}}$ and if J_2 is additive with respect to this decomposition, then Problem (1.3) splits up into independent subproblems provided that the auxiliary function K be chosen additive. "Large step" decomposition algorithms[1] may thus be obtained even if the whole function is nondifferentiable provided that the linearized part J_1 has a Lipschitz derivative.

Extensions of the APP to variational inequalities have been proposed by Cohen [9] and Mataoui [20]. Considering the problem of finding an $x^\sharp$ such

[1]The expression "large step" refers to the fact that the stepsize ε in (1.1) is kept away from 0 asymptotically, that is when the iteration index k goes to infinity; this is more important than the fact that ε is kept constant or that it is made dependent on k; "large steps" should be opposed to "small steps", an expression which refers to algorithms in which ε^k must tend to 0 asymptotically, as encountered for example in some subgradient algorithms.

that

$$x^\sharp \in X, \quad \forall x \in X^{\mathrm{ad}}, \quad \left\langle \Psi_0(x^\sharp), x - x^\sharp \right\rangle \geq 0, \tag{1.4}$$

where Ψ_0 is a monotone and Lipschitz continuous operator, Cohen [9] has shown that the algorithm: knowing x^k, derive x^{k+1} from the resolution of

$$\min_{x \in X^{\mathrm{ad}}} \left(\frac{1}{\varepsilon} \left(K(x) - \left\langle K'(x^k), x - x^k \right\rangle \right) + \left\langle \Psi_0(x^k), x - x^k \right\rangle \right) \tag{1.5}$$

constructs a sequence $\{x^k\}$ which strongly converges towards a solution of (1.4) under the assumption that Ψ_0 is strongly monotone. Mataoui [20] has extended this result, showing the weak convergence by only assuming what he has called the Dunn property of Ψ_0. Several other names for this property (or very similar properties) can be found in the literature: co-coercivity [28], firm nonexpansiveness ([16]–[13]). The Dunn property is nothing else but the strong monotonicity of the inverse. Mataoui [20] shows that it can be enforced by Yosida regularization.

Algorithm (1.5) can also be derived from (1.1). For $\Psi = \partial[J + I_{X^{\mathrm{ad}}}] + \Psi_0$ and $\Xi = K' - \varepsilon\Psi_0$, Algorithm (1.1) yields (1.5). More generally, to find a zero of an operator $\Psi = \Psi_1 + \Psi_2$, by choosing $\Xi = K' - \varepsilon\Psi_1$ in Algorithm (1.1), one obtains

$$0 \in \left(\frac{1}{\varepsilon} \left(K'(x^{k+1}) - K'(x^k) \right) + \Psi_1(x^k) + \Psi_2(x^{k+1}) \right). \tag{1.6}$$

This *partial relaxation* algorithm can also be cast into the APP framework and can be used to derive decomposition algorithms.

Notice that when $K' = I$ (I denotes identity), Algorithm (1.6) takes the form

$$x^{k+1} = (I + \varepsilon\Psi_2)^{-1} (I - \varepsilon\Psi_1)(x^k). \tag{1.7}$$

The weak convergence of this "splitting algorithm" (proposed by Lions and Mercier [17]), has been proved by Gabay [14] for Ψ_2 monotone when Ψ_1 enjoys the Dunn property (see also [26]). Besides, in parallel with our work, Zhu and Marcotte have shown the weak convergence of Algorithm (1.6) under the same assumptions upon Ψ_1 and Ψ_2.

The work of Patriksson [22] in this field must also be mentioned. He proposes a framework which introduces nonsymmetric auxiliary operators and which, in a way, may be considered as being more general than (1.1). The possibility of performing a linear search is for example considered. For this reason, the formalism (1.1) cannot be considered as being particularly original.

The main contribution of this paper is to be found in the type of assumptions to ensure convergence and in the results regarding the convergence rate. It is shown that it is the relationship between the auxiliary operator Ξ and the main operator Ψ, rather than the individual properties of each operator, which plays the central role (see Theorem 3.4). In particular, for the case of the "splitting" algorithm (1.7), convergence does not directly follows from the properties of Ψ_1 but from a relationship between Ψ_1 and the whole operator Ψ: the Dunn property, or even the monotonicity of Ψ_1,

is not required to prove convergence. Moreover, even if $\Psi_1 = \Psi$ and $\Psi_2 = 0$, the convergence of (1.7), which boils down to

$$x^{k+1} = x^k - \varepsilon\Psi(x^k) \tag{1.8}$$

may be ensured without the Dunn property provided that Ψ has a specific structure. This type of situation occurs in particular when Ψ is the operator associated with a convex-concave function (§4.2). In this case, Algorithm (1.8) is the parallel version of the Arrow-Hurwicz algorithm. In the Lipschitz continuous case, from the general convergence result, the convergence of this Arrow-Hurwicz algorithm is derived under the assumption that a "partial Dunn property" is satisfied (see Definition 4.4). This property can be enforced by a partial regularization (see Definition 4.6). In the particular case when the convex-concave mapping is a Lagrangian, the partial regularization with respect to the dual variables yields an Augmented Lagrangian. Therefore, the parallel version of the Arrow-Hurwicz algorithm applied to an Augmented Lagrangian converges in the Lipschitz case. As far as we know, these results are original.

Our study of the convergence rate of Algorithm (1.1) shows that the strong monotonicity of Ψ ensures a "linear in the average" rate of convergence. Moreover, this strong monotonicity assumption can be weakened if the symmetric part of Ξ is linear. Then, the Lipschitz property of the inverse yields linear convergence. The result obtained by Luque [18] for the proximal algorithm is thus generalized.

This paper is structured as follows. After some basic results have been recalled (§2), the convergence results concerning Algorithm (1.1) are proposed (§3). In the second part of the paper, some results about partial regularization (§4) are given and the particular case of saddle points problems is then studied. The third part (§5) is devoted to the study of the convergence rate of Algorithm (1.1).

2. Basic results

2.1. Basic definitions

Let X be a Hilbert space, let $\langle\cdot,\cdot\rangle$ denote its scalar product and let $\|\cdot\|$ be the related norm. Let I be the identity mapping over X and $\Psi : X \to 2^X$ be a point-to-set operator. The operator is single-valued if, for all x, $\Psi(x)$ is reduced to a point. No distinction will be made in this case between this point and the singleton $\Psi(x)$.

The operator Ψ is *monotone* if

$$\forall x_1, x_2 \in X,\ \forall(\psi_1,\psi_2) \in \Psi(x_1)\times\Psi(x_2) :\ \langle\psi_1-\psi_2, x_1-x_2\rangle \geq 0.$$

If moreover

$$x_1 \neq x_2 \Rightarrow \langle\psi_1-\psi_2, x_1-x_2\rangle > 0,$$

then Ψ is *strictly* monotone. Finally, if there exists a constant $a \in \mathbb{R}_+^*$ such that $\Psi - aI$ is monotone, then Ψ is *strongly* monotone with constant a.

The domain of Ψ is defined as follows

$$\mathrm{dom}\Psi = \{x \in X \mid \Psi(x) \neq\varnothing\}\,.$$

DEFINITION 2.1. A monotone operator is maximal if its graph cannot be included in a strictly larger graph of a monotone operator.

DEFINITION 2.2. A single-valued operator Ψ is hemicontinuous if for all $x_1, x_2 \in X$, the mapping

$$\Psi_{x_1,x_2} : [0,1] \to X,\ t \mapsto \Psi(tx_1 + (1-t)x_2) \tag{2.1}$$

is continuous in the weak topology.

A monotone operator which is hemicontinuous is necessarily maximal (see [24]).

2.2. Existence of zeros of maximal monotone operators

The point $x^\sharp \in X$ is a zero of Ψ if $0 \in \Psi(x^\sharp)$.

ASSUMPTION 2.3. There exist $\overline{x} \in X$ and $\rho > 0$ such that

$$\forall x \in X,\ \forall \psi \in \Psi(x) :\ \{\|x - \overline{x}\| > \rho\} \Rightarrow \{\langle \psi, x - \overline{x}\rangle \geq 0\}.$$

REMARK 2.4. A strongly monotone operator satisfies this assumption and so does an operator the domain of which is bounded.

PROPOSITION 2.5 (REF. [24]). *Let Ψ be a maximal monotone operator. The operator has a zero if and only if it satisfies Assumption 2.3.*

2.3. Sum of maximal operators

PROPOSITION 2.6. *Let $\Psi_1 : X \to 2^X$ be a maximal monotone operator. Let $\Psi_2 : X \to X$ be a hemicontinuous operator. If $\Psi = \Psi_1 + \Psi_2$ is monotone, it is maximal monotone.*

Proof. One has to prove that, for every $x \in X$ and for every ψ such that

$$\forall x_0 \in X,\ \forall \psi_0 \in \Psi(x_0),\ \langle \psi_0 - \psi, x_0 - x\rangle \geq 0, \tag{2.2}$$

then $\psi \in \Psi(x)$. Let $x_0 \in X$, $t \in [0,1]$ and $\psi_{1,t} \in \Psi_1(x + t(x_0 - x))$. The monotonicity of Ψ_1 and (2.2) yield

$$\begin{aligned}\left\langle \psi_{1,1} + \Psi_2\left(x + t(x_0 - x)\right) - \psi,\, x_0 - x\right\rangle & \\ \geq \left\langle \psi_{1,t} + \Psi_2\left(x + t(x_0 - x)\right) - \psi, x_0 - x\right\rangle & \\ \geq 0. &\end{aligned}$$

Since Ψ_2 is hemicontinuous, when $t \to 0$, one obtains

$$\left\langle \psi_{1,1} - \left(\psi - \Psi_2(x)\right), x_0 - x\right\rangle \geq 0.$$

One can complete the proof using the maximality of Ψ_1. □

REMARK 2.7. Proposition 2.6 does not assume the monotonicity of Ψ_2. For this reason, it does not follow from the classical result on the maximality of the sum of monotone operators (see [24]).

2.4. Dunn property

2.4.1. Definition.

Definition 2.8 (Dunn property). Let $A \in \mathbb{R}_+^*$ and $\Psi : X \to X$ be a single-valued mapping; Ψ has the Dunn property with constant A, if

$$\forall x_1, x_2 \in X : \langle \Psi(x_1) - \Psi(x_2), x_1 - x_2 \rangle \geq \frac{1}{A} \|\Psi(x_1) - \Psi(x_2)\|^2.$$

This property can also be stated as the strong monotonicity of the (point-to-set) inverse Ψ^{-1} with constant $1/A$. Observe that Ψ is then Lipschitz with constant A (thus the fact that it is single-valued is implied by the Dunn property itself).

If Ψ is the derivative of a convex function, Ψ has the Dunn property with constant A if and only if it is A-Lipschitz (see [1, 15]). However, in the nonsymmetric case, this equivalence is not true. The rotation $(x, y) \mapsto (-y, x)$ is Lipschitz but does not enjoy the Dunn property. Nevertheless, if Ψ is strongly monotone with constant a and Lipschitz with constant L, Mataoui [20] shows that it enjoys the Dunn property with constant L^2/a. The Yosida regularization [27] also enforces the Dunn property [20].

Definition 2.9 (Regularization). Let $\Psi : X \to 2^X$ be a maximal monotone operator and $c > 0$. The operator

$$\Psi_c : X \to X,\ x \mapsto \frac{1}{c}(x - s(x))$$

where $s(x)$ is such that

$$0 \in \left(\frac{1}{c}\big(s(x) - x\big) + \Psi\big(s(x)\big) \right)$$

is the regularization of Ψ.

Proposition 2.10 (Ref. [20]). *The regularization Ψ_c of Ψ is monotone and enjoys the Dunn property with constant $1/c$.*

2.4.2. Convergence based on the Dunn property.

If Ψ_1 has the Dunn property with constant A and if $0 < \varepsilon < 2/A$, Algorithm (1.7) weakly converges (see [14]). Moreover, when K is a strongly convex function with constant b, the convergence of (1.5) and (1.6) has been shown by Mataoui [20] and Zhu and Marcotte [28] if $0 < \varepsilon < 2b/A$.

3. New convergence results for Algorithm (1.1)

The convergence results which are given in this section are very different when the operator Ψ is symmetric and when it is not. In the symmetric case, the Dunn property of the auxiliary operator Ξ over every bounded set suffices to ensure convergence. If Ψ is not symmetric, assumptions that relate the geometries of Ψ and Ξ have to be made. The proof of these convergence results can be found in Appendix A.

3.1. Symmetric case

Consider the minimization problem (1.2) where $J : X \to \mathbb{R}$ is a convex lower semi-continuous (l.s.c.) mapping and $X^{\rm ad}$ is a nonempty closed convex subset. The subdifferential of $J + I_{X^{\rm ad}}$ where $I_{X^{\rm ad}}$ is the indicator function of $X^{\rm ad}$ is a maximal monotone operator (see [21]. A zero of this maximal monotone operator is a solution of (1.2).

Algorithm (1.1) when applied to this problem yields

$$0 \in \left(\frac{1}{\varepsilon} \left(\Xi(x^{k+1}) - \Xi(x^k) \right) + \partial [J + I_{X^{\rm ad}}](x^{k+1}) \right). \tag{3.1}$$

The idea of using such a nonsymmetric auxiliary operator to minimize a function may seem rather odd. Nevertheless, the relaxation a nonsymmetric part of $\partial[J + I_{X^{\rm ad}}]$ at each iteration is not as eccentric as it may seem. In the linear case, — when $J = \frac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle$, with A linear and $b \in X$, and $X^{\rm ad} = X$ —, by choosing $\Xi = I - \varepsilon U$, where U is an upper triangular part of A, Algorithm (3.1) turns out to be a Gauss-Siedel algorithm: a triangular system has to be solved at each iteration. Using this type of strategy, in the nonlinear case, sequential version of gradient or relaxation algorithms may be obtained. To illustrate this point, let us consider the following example

$$\min_{(y,z) \in Y^{\rm ad} \times Z^{\rm ad}} F(y,z) \stackrel{\rm def}{=} f(y) + \langle y, h(z)\rangle + g(z), \tag{3.2}$$

where Y and Z are two Hilbert spaces, $h : Z \to Y$ and $f : Y \to \mathbb{R}$ and $g : Z \to \mathbb{R}$. One assumes that F is a convex mapping and that $Y^{\rm ad}$ and $Z^{\rm ad}$ are two nonempty closed convex subsets.

By choosing a (generally) nonsymmetric operator $\Xi : (y,z) \mapsto (y - \varepsilon h(z), \frac{\varepsilon}{\rho} z)$, Algorithm (3.1) applied to this problem turns out to be a sequential relaxation algorithm: knowing (y^k, z^k),

$$\min_{y \in Y^{\rm ad}} \left(\frac{1}{2\varepsilon} \|y - y^k\|^2 + F(y, z^k) \right) \text{ provides } y^{k+1}, \tag{3.3a}$$

$$\min_{z \in Z^{\rm ad}} \left(\frac{1}{2\rho} \|z - z^k\|^2 + F(y^{k+1}, z) \right) \text{ provides } z^{k+1}. \tag{3.3b}$$

If f and g are differentiable, with

$$\Xi : (y,z) \mapsto \big(y - \varepsilon h(z) - \varepsilon f'(y), (\varepsilon/\rho) z - \varepsilon g'(z)\big),$$

one obtains a sequential gradient algorithm

$$y^{k+1} = {\rm proj}_{Y^{\rm ad}} \left(y^k - \varepsilon \frac{\partial F}{\partial y}(y^k, z^k) \right), \tag{3.4a}$$

$$z^{k+1} = {\rm proj}_{Z^{\rm ad}} \left(z^k - \rho \frac{\partial F}{\partial z}(y^{k+1}, z^k) \right) \tag{3.4b}$$

where, e.g., ${\rm proj}_{Y^{\rm ad}}$ denotes the projection on $Y^{\rm ad}$; clearly, partial linearizations of f and g would yield algorithms halfway between (3.3) and (3.4).

Proposition 3.1. *Assume that*

- $J : X \to \mathbb{R}$ *is a convex, l.s.c. function which is coercive over* $X^{\rm ad}$,
- *the auxiliary operator* $\Xi : X \to X$ *has the Dunn property over every bounded subset of* X.

Then, Problem (3.1) *has a solution* x^{k+1}. *The sequence* $\{J(x^k)\}$ *decreases and converges towards* $J(x^\sharp)$ *where* $x^\sharp$ *is a solution of* (1.2). *Furthermore, the sequence* $\{x^k\}$ *has cluster points in the weak topology, and every such cluster point is a solution of* (1.2).

REMARK 3.2. The Dunn property of Ξ over every bounded set is satisfied if this operator is strongly monotone and is Lipschitz continuous over every bounded set (see [20]).

REMARK 3.3. Assume that the mapping h in Problem (3.2) has the Lipschitz property with constant L. The auxiliary operator Ξ which has been chosen to obtain (3.3) is such that for all y_1 , $y_2 \in Y$ and for all z_1 , $z_2 \in Z$, one has that

$$\begin{aligned}\big\langle \Xi(y_1,z_1) \; &- \Xi(y_2,z_2), (y_1,z_1)-(y_2,z_2)\big\rangle \\ &= \|y_1-y_2\|^2 - \varepsilon\,\langle h(z_1)-h(z_2), y_1-y_2\rangle + \frac{\varepsilon}{\rho}\|z_1-z_2\|^2 \\ &\geq \left(1-\frac{1}{2\alpha}\right)\|y_1-y_2\|^2 + \varepsilon\left(\frac{1}{\rho}-\frac{\alpha\varepsilon L^2}{2}\right)\|z_1-z_2\|^2,\end{aligned}$$

for all $\alpha > 0$. Therefore, if $\varepsilon\rho < 4/L^2$, the operator Ξ is strongly monotone and has the Lipschitz property: the assumptions of Proposition 3.1 are satisfied; Algorithm 3.3 converges.

Proposition 3.1 also ensures the convergence of Algorithm 3.4 provided that h, f' and g' have the Lipschitz property.

3.2. General case

If $M : X \to X$ is a linear continuous, symmetric, strongly monotone operator, then $\|x\|_M$ denotes the related norm $\langle x, Mx\rangle^{1/2}$.

THEOREM 3.4. *Let* $\Psi : X \to 2^X$ *be a maximal monotone operator. Assume that* $\Xi = K' + \varepsilon\Sigma$ *where:*

- K' *is the derivative of a convex mapping* K; K' *is supposed to be Lipschitz over any bounded subset and to satisfy*

$$\forall x_1, x_2 \in \mathrm{dom}\Psi,\ \langle K'(x_1)-K'(x_2), x_1-x_2\rangle \geq \|x_1-x_2\|_M^2, \tag{3.5}$$

- $\Sigma : X \to X$ *is hemicontinuous.*

Suppose that there exists $A > 0$ *such that*

$$\forall x_1, x_2 \in \mathrm{dom}\Psi,\ \forall(\psi_1,\psi_2) \in \Psi(x_1)\times\Psi(x_2) :$$
$$\langle \psi_1-\psi_2, x_1-x_2\rangle \geq \frac{1}{A}\,\|\Sigma(x_1)-\Sigma(x_2)\|_{M^{-1}}^2, \tag{3.6}$$

and that $0 < \varepsilon < 2/A$; *then Problem* (1.1) *has a unique solution. Assuming* Ψ *has a zero* $x^\sharp$, *the sequence* $\{x^k\}$ *has cluster points in the weak topology and every such cluster point is a zero of* Ψ. *Moreover:*

- $\{\Sigma(x^k)\}$ *strongly converges towards* $\Sigma(x^\sharp)$ *(which, after* (3.6)*, is unique even if* $x^\sharp$ *is not) and* $\{\Xi(x^{k+1})-\Xi(x^k)\}$ *strongly converges towards* 0*;*
- *if* K' *is continuous when the domain and the range spaces are both equipped with the weak topology, the whole sequence* $\{x^k\}$ *converges weakly towards a zero of* Ψ*;*

- *if Ψ is strongly monotone, $\{x^k\}$ strongly converges towards the unique zero of Ψ.*

REMARK 3.5. Assumptions of Theorem 3.4 ensure the continuity of K' in the strong topology. In a finite dimensional space, the convergence of the whole sequence towards a solution is thus established without any further assumptions.

REMARK 3.6. It has already been pointed out that (1.6) is nothing but (1.1) for $\Xi = K' - \varepsilon\Psi_1$ (that is, for $\Sigma = -\Psi_1$). In this case, for $M = I$, assumption (3.6) may be written as follows

$$\forall x_1, x_2 \in X^{\mathrm{ad}} : \langle \Psi(x_1) - \Psi(x_2), x_1 - x_2 \rangle \geq \frac{1}{A} \|\Psi_1(x_1) - \Psi_1(x_2)\|^2. \quad (3.7)$$

Under this assumption, if K is strongly convex with constant b, if $0 < \varepsilon < 2b/A$, and if Ψ_1 is hemicontinuous, convergence will be ensured by Theorem 3.4.

Therefore, convergence can be claimed without assuming the monotonicity of Ψ_1 *and* Ψ_2 but only the monotonicity of Ψ. Moreover, this convergence does not only result from the properties of Ψ (or of Ψ_1) but also from a relationship between the geometry of Ψ and that of Ψ_1 as expressed by (3.7).

The property (3.7) is the Dunn property whenever $\Psi = \Psi_1$. Therefore, Theorem 3.4 turns out to be a generalization of the convergence result obtained by Mataoui [20]. Zhu and Marcotte proved the convergence of a scheme like (1.6) under the assumption that Ψ_1 enjoys the Dunn property and that Ψ_2 is monotone. In such a case, (3.7) is also satisfied.

REMARK 3.7. In order to derive Algorithm (1.8) from (1.6), one can choose $K' = I$ and $\Psi_1 = \Psi$, that is, $\Psi_2 = 0$, and then, the assumption which ensures convergence of (1.8) is the Dunn property of Ψ. But, one can also obtain (1.8) from (1.6) with any nonzero Ψ_2, provided this is symmetric (choose $K' = I - \varepsilon\Psi_2$, that is, $\Sigma = -\Psi_1$). Then, condition (3.6) may turn out to be weaker than the Dunn property of Ψ. For example, in $\mathbb{R}^2$, consider

$$\Psi(x,y) = \begin{pmatrix} x - y \\ x \end{pmatrix}, \quad \Psi_1(x,y) = \begin{pmatrix} 0 \\ 2x \end{pmatrix}, \quad \Psi_2(x,y) = \begin{pmatrix} x - y \\ -x \end{pmatrix}.$$

4. APPLICATION TO SADDLE POINT PROBLEMS

In the general case, the Dunn property ensures convergence of Algorithm (1.8). We aim at showing that, by taking advantage of the specific structure of the operators related to convex-concave functions, this Dunn assumption may be weakened. The application of Theorem 3.4 proves convergence of this algorithm provided that the partial Dunn property holds true (see Definition 4.4, below). This property can be enforced by partial regularization.

4.1. PARTIAL REGULARIZATION

4.1.1. DEFINITION. Let Y and Z be two Hilbert spaces such that $X = Y \times Z$. For all $x \in X$, $\Psi_Y(x)$ and $\Psi_Z(x)$ denote the projection of $\Psi(x)$ onto Y and Z, respectively.

DEFINITION 4.1. The operator Ψ is Z-regularizable if it is monotone and if the following two assumptions hold:

$$\forall y \in Y,\ \{\mathrm{dom}\Psi_Z(y,\cdot) \neq \varnothing\} \Rightarrow \{\Psi_Z(y,\cdot) \text{ is maximal monotone}\}, \quad (4.1)$$

$$\forall (y,z) \in Y \times Z,\ \Psi(y,z) = \Psi_Y(y,z) \times \Psi_Z(y,z). \quad (4.2)$$

DEFINITION 4.2. Assume that Ψ is Z-regularizable. Let c be a positive number. The partial regularization of Ψ over Z is defined as follows: $\Psi_c^Z : X \to 2^X$,

$$x = (y,z) \mapsto \Psi_c^Z(x) \stackrel{\text{def}}{=} \begin{cases} \begin{pmatrix} c\Psi_Y(y,s) \\ -\frac{1}{c}(s-z) \end{pmatrix} & \text{if } \mathrm{dom}\Psi_Z(y,\cdot) \neq \varnothing \\ \varnothing & \text{otherwise} \end{cases} \quad (4.3)$$

where s is the unique solution of

$$0 \in \frac{1}{c}(s-z) + \Psi_Z(y,s). \quad (4.4)$$

Problem (4.4) consists in finding a zero of the sum of two maximal monotone operators, one of them, proportional to identity, having the whole space as its domain. Hence, this sum is a maximal monotone operator (see [24]). Moreover, because identity is strongly monotone, Problem (4.4) has a unique solution (see Proposition 2.5).

PROPOSITION 4.3. *Suppose Ψ is Z-regularizable and has zeros. Then, the zeros of Ψ_c^Z are those of Ψ and conversely.*

Proof. Let $x^\sharp = (y^\sharp, z^\sharp) \in X$ be a zero of Ψ. Then, from (4.4) it follows that $s_Z(x^\sharp) = z^\sharp$ and thus $x^\sharp$ is a zero of Ψ_c^Z.

Conversely, if $x^\sharp = (y^\sharp, z^\sharp) \in X$ is a zero of Ψ_c^Z, the projection of $\Psi_c^Z(x^\sharp)$ onto Z shows that $s(x^\sharp) = z^\sharp$ and therefore, from (4.4), $0 \in \Psi_Z(x^\sharp)$. The projection of $\Psi_c^Z(x^\sharp)$ onto Y shows that $0 \in \Psi_Y(x^\sharp)$. Finally, from assumption (4.2), we conclude that $0 \in \Psi(x^\sharp)$. □

4.1.2. PARTIAL REGULARIZATION AND PARTIAL DUNN PROPERTY.

DEFINITION 4.4. The operator $\Psi : X \to 2^X$ enjoys the partial Dunn property over Z with constant $A_Z \in \mathbb{R}_+^*$ if

$$\forall (x_1,x_2) \in X \times X,\ \forall (\psi_1,\psi_2) \in \Psi(x_1) \times \Psi(x_2) :$$
$$\langle \psi_1 - \psi_2, x_1 - x_2 \rangle \geq \frac{1}{A_Z} \|\psi_{1,Z} - \psi_{2,Z}\|^2. \quad (4.5)$$

REMARK 4.5. If (4.5) holds, $\Psi_Z(y,\cdot) : z \mapsto \Psi_Z(y,z)$ has the Dunn property and is thus a single-valued operator. Therefore, (4.5) can be rephrased:

$$\forall (x_1,x_2) \in X \times X,\ \forall (\psi_1,\psi_2) \in \Psi(x_1) \times \Psi(x_2) :$$
$$\langle \psi_1 - \psi_2, x_1 - x_2 \rangle \geq \frac{1}{A_Z} \|\Psi_Z(x_1) - \Psi_Z(x_2)\|^2.$$

PROPOSITION 4.6. *Suppose Ψ is Z-regularizable and has a zero. The operator Ψ_c^Z is monotone and has the partial Dunn property over Z with constant $1/c$.*

Proof. Let $x_1, x_2 \in X$. Let $(\xi_1, \xi_2) \in \Psi_c^Z(x_1) \times \Psi_c^Z(x_2)$. The definition of Ψ_c^Z shows that there exists $(\psi_{1,Y}, \psi_{2,Y}) \in \Psi_Y(y_1, s_1) \times \Psi_Y(y_2, s_2)$ such that

$$\begin{aligned}\langle \xi_1 - \xi_2, x_1 - x_2 \rangle \\ = \langle \psi_{1,Y} - \psi_{2,Y}, y_1 - y_2 \rangle + \underbrace{\frac{1}{c} \langle (s_2 - z_2) - (s_1 - z_1), z_1 - z_2 \rangle}_{\Delta}. \quad (4.6)\end{aligned}$$

One has that

$$\Delta = \frac{1}{c} \|(s_1 - z_1) - (s_2 - z_2)\|^2 - \frac{1}{c} \langle (s_1 - z_1) - (s_2 - z_2), s_1 - s_2 \rangle .$$

Moreover, the definition (4.4) of s_1 and s_2 subsumes the existence of $(\psi_{1,Z}, \psi_{2,Z}) \in \Psi_Z(y_1, s_1) \times \Psi_Z(y_2, s_2)$ such that

$$\Delta = \frac{1}{c} \|(s_1 - z_1) - (s_2 - z_2)\|^2 + \langle \psi_{1,Z} - \psi_{2,Z}, s_1 - s_2 \rangle .$$

Substituting this equality in (4.6) and using Assumption (4.2), one gets

$$\langle \xi_1 - \xi_2, x_1 - x_2 \rangle = \frac{1}{c} \|(s_1 - z_1) - (s_2 - z_2)\|^2 + \langle \psi_1 - \psi_2, (y_1, s_1) - (y_2, s_2) \rangle ,$$

where $\psi_1 \in \Psi(y_1, s_1)$ and $\psi_2 \in \Psi(y_2, s_2)$. The proof is completed by appealing to the monotonicity of Ψ and Definition 4.2. □

4.2. Saddle point problems

Consider the following problem

$$\min_{v \in \mathcal{V}^{\mathrm{ad}}} \max_{\pi \in \Pi^{\mathrm{ad}}} \Lambda(v, \pi), \quad (4.7)$$

where

- Λ is a convex-concave l.s.c.-u.s.c. function,
- $\mathcal{V}^{\mathrm{ad}} \subset \mathcal{V}$ and $\Pi^{\mathrm{ad}} \subset \Pi$ are two closed convex subsets of the Hilbert spaces $\mathcal{V}$ and Π.

Consider

$$\Gamma : (v, \pi) \mapsto \Gamma_v(v, \pi) \times \Gamma_\pi(v, \pi),$$

where

$$\Gamma_v(v, \pi) \stackrel{\mathrm{def}}{=} \partial_v [\Lambda + I_{\mathcal{V}^{\mathrm{ad}} \times \Pi}](v, \pi), \quad (4.8a)$$

$$\Gamma_\pi(v, \pi) \stackrel{\mathrm{def}}{=} \partial_\pi [-\Lambda + I_{\mathcal{V} \times \Pi^{\mathrm{ad}}}](v, \pi). \quad (4.8b)$$

This mapping is maximal monotone (see [2]). The zeros of Γ are the solutions of (4.7).

4.2.1. Algorithm. Assume that $\Lambda = \mathcal{L} + \mathcal{M}$ where $\mathcal{L}$ and $\mathcal{M}$ are two convex-concave l.s.c.-u.s.c. mappings; $\mathcal{L}$ is supposed to be Gâteaux-differentiable with a hemicontinuous derivative. To compute solutions of (4.7), consider the algorithm: knowing (v^k, π^k), derive (v^{k+1}, π^{k+1}) from the resolution of

$$\begin{aligned}\max_{\pi \in \Pi^{\mathrm{ad}}} \min_{v \in \mathcal{V}^{\mathrm{ad}}} \Big(\frac{1}{\varepsilon} \big(G(v) - \langle G'_v(v^k), v \rangle \big) + \frac{1}{\rho} \big(H(\pi) - \langle H'_\pi(\pi^k), \pi \rangle \big) \\ + \langle \mathcal{L}'_v(v^k, \pi^k), v \rangle + \langle \mathcal{L}'_\pi(v^k, \pi^k), \pi \rangle + \mathcal{M}(v, \pi) \Big), \quad (4.9)\end{aligned}$$

where ε and ρ are positive numbers and where G (respectively, H) is a strongly convex (respectively, strongly concave) auxiliary function. This algorithm falls into the framework of the Auxiliary Problem Principle. For $G' = -H' = I$ and $\mathcal{L} = \Lambda$, one obtains a gradient algorithm which is nothing but a parallel form of the Arrow-Hurwicz algorithm:

$$v^{k+1} = \mathrm{proj}_{\mathcal{V}^{\mathrm{ad}}}\left(v^k - \varepsilon \Lambda'_v(v^k, \pi^k)\right), \tag{4.10a}$$

$$\pi^{k+1} = \mathrm{proj}_{\Pi^{\mathrm{ad}}}\left(\pi^k + \rho \Lambda'_\pi(v^k, \pi^k)\right). \tag{4.10b}$$

In the case when $\mathcal{M} = \Lambda$, one obtains a classical proximal algorithm. Moreover, Algorithm (4.9) may be viewed as an application of the general scheme (1.1) to Γ. By picking up the following auxiliary operator

$$\Xi : (v, \pi) \mapsto \Xi(v, \pi) = \begin{pmatrix} G'(v) - \varepsilon \mathcal{L}'_v(v, \pi) \\ -\frac{\varepsilon}{\rho} H'(\pi) + \varepsilon \mathcal{L}'_\pi(v, \pi) \end{pmatrix}, \tag{4.11}$$

Algorithm (1.1) takes the following form: 0 belongs to

$$\begin{pmatrix} \frac{1}{\varepsilon}\left(G'(v^{k+1}) - G'(v^k)\right) + \mathcal{L}'_v(v^k, \pi^k) + \partial_v[\mathcal{M} + I_{\mathcal{V}^{\mathrm{ad}} \times \Pi}](v^{k+1}, \pi^{k+1}) \\ \frac{1}{\rho}\left(-H'(\pi^{k+1}) + H'(\pi^k)\right) - \mathcal{L}'_\pi(v^k, \pi^k) + \partial_\pi[-\mathcal{M} + I_{\mathcal{V} \times \Pi^{\mathrm{ad}}}](v^{k+1}, \pi^{k+1}) \end{pmatrix}.$$

These are nothing but the optimality conditions of Problem (4.9).

4.2.2. Convergence.

Theorem 4.7. *Consider Problem* (4.7) *where* $\Lambda = \mathcal{L} + \mathcal{M}$ *is a convex-concave mapping such that:*

- Λ *has a saddle point* $x^\sharp = (v^\sharp, \pi^\sharp)$ *over the closed convex subset* $\mathcal{V}^{\mathrm{ad}} \times \Pi^{\mathrm{ad}}$,
- $\mathcal{M} : \mathcal{V} \times \Pi \to \mathbb{R}$ *is convex-concave l.s.c.-u.s.c. over* $\mathcal{V}^{\mathrm{ad}} \times \Pi^{\mathrm{ad}}$,
- $\mathcal{L} : \mathcal{V} \times \Pi \to \mathbb{R}$ *is convex-concave l.s.c.-u.s.c. and differentiable with a hemicontinuous derivative over* $\mathcal{V}^{\mathrm{ad}} \times \Pi^{\mathrm{ad}}$.

Assume that:

- $\mathcal{L}'_v(\cdot, \pi)$ *is* L*-Lipschitz (uniformly in* π*),*
- $\mathcal{L}'_\pi(\cdot, \pi)$ *is* C*-Lipschitz (uniformly in* π*),*

and that the following condition is satisfied

$$\forall (v_1, \pi_1), (v_2, \pi_2) \in \mathcal{V}^{\mathrm{ad}} \times \Pi^{\mathrm{ad}},\ \forall (\gamma_1, \gamma_2) \in \Gamma(v_1, \pi_1) \times \Gamma(v_2, \pi_2),$$
$$\langle \gamma_1 - \gamma_2, (v_1, \pi_1) - (v_2, \pi_2) \rangle \geq \frac{1}{D} \left\| \mathcal{L}'_\pi(v_1, \pi_1) - \mathcal{L}'_\pi(v_2, \pi_2) \right\|^2, \tag{4.12}$$

where Γ *is the operator defined by* (4.8).

Consider Algorithm (4.9) *where* G *and* $-H$ *are strongly convex, with constants* b_G *and* b_H*, and differentiable with Lipschitz continuous derivatives over every bounded subset.*

Then, if ε *and* ρ *satisfy*

$$0 < \rho < \frac{b_H}{2D + C^2/\alpha}, \quad 0 < \varepsilon < \frac{b_G}{\alpha + L}, \tag{4.13}$$

for some $\alpha > 0$, v^{k+1} *and* π^{k+1} *do exist in* (4.9) *and are unique. The sequence* $\{(v^k, \pi^k)\}$ *has cluster points in the weak topology. Every such cluster point is a saddle point of* Λ *over* $\mathcal{V}^{\mathrm{ad}} \times \Pi^{\mathrm{ad}}$. *If* $\mathcal{L}'$, G' *and* H' *are*

continuous when the domain and range spaces are equipped with the weak topology, the whole sequence $\{(v^k, \pi^k)\}$ *converges weakly towards a solution.*

Proof. Under the foregoing assumptions, the operator Γ defined by (4.8) is a maximal monotone operator. Algorithm (4.9) is a particular instance of (1.1) for the auxiliary operator Ξ defined by (4.11). To complete this proof, one has thus to show that, for this auxiliary operator, the assumptions of Theorem 3.4 are satisfied.

The operator Ξ is equal to $K' + \varepsilon\Sigma$ with

$$K : (v, \pi) \mapsto K(v, \pi) \stackrel{\text{def}}{=} G(v) - \frac{\varepsilon}{\rho} H(\pi) - \varepsilon \mathcal{L}(v, \pi), \tag{4.14}$$

$$\Sigma : (v, \pi) \mapsto \Sigma(v, \pi) \stackrel{\text{def}}{=} \begin{pmatrix} 0 \\ 2\mathcal{L}'_\pi(v, \pi) \end{pmatrix}. \tag{4.15}$$

Let $x_1 \stackrel{\text{def}}{=} (v_1, \pi_1) \in \mathcal{V}^{\text{ad}} \times \Pi^{\text{ad}}$ and $x_2 \stackrel{\text{def}}{=} (v_2, \pi_2) \in \mathcal{V}^{\text{ad}} \times \Pi^{\text{ad}}$. The strong monotonicity of G' and of $-H'$ yields

$$\begin{aligned} \overbrace{\left\langle K'(x_1) - K'(x_2), x_1 - x_2 \right\rangle}^{\Delta_K} \geq b_G \|v_1 - v_2\|^2 + \frac{\varepsilon}{\rho} b_H \|\pi_1 - \pi_2\|^2 \\ - \varepsilon \underbrace{\left\langle \mathcal{L}'_v(v_1, \pi_1) - \mathcal{L}'_v(v_2, \pi_2), v_1 - v_2 \right\rangle}_{\delta_1} \\ - \varepsilon \underbrace{\left\langle \mathcal{L}'_\pi(v_1, \pi_1) - \mathcal{L}'_\pi(v_2, \pi_2), \pi_1 - \pi_2 \right\rangle}_{\delta_2}. \end{aligned}$$

The concavity of $\mathcal{L}(v, \cdot)$ and the Lipschitz property of $\mathcal{L}'_\pi(\cdot, \pi)$ yield

$$\begin{aligned} \left\langle -\mathcal{L}'_\pi(v_1, \pi_1), \pi_1 - \pi_2 \right\rangle &= \left\langle \mathcal{L}'_\pi(v_2, \pi_1) - \mathcal{L}'_\pi(v_1, \pi_1), \pi_1 - \pi_2 \right\rangle \\ &\quad - \left\langle \mathcal{L}'_\pi(v_2, \pi_1), \pi_1 - \pi_2 \right\rangle \\ &\geq -C\|v_1 - v_2\|\|\pi_1 - \pi_2\| - \mathcal{L}(v_2, \pi_1) + \mathcal{L}(v_2, \pi_2); \\ \left\langle -\mathcal{L}'_\pi(v_2, \pi_2), \pi_2 - \pi_1 \right\rangle &\geq -C\|v_1 - v_2\|\|\pi_1 - \pi_2\| - \mathcal{L}(v_1, \pi_2) + \mathcal{L}(v_1, \pi_1). \end{aligned}$$

Since $\mathcal{L}(\cdot, \pi)$ is convex and $\mathcal{L}'_v(\cdot, \pi)$ is Lipschitz, one has that

$$\begin{aligned} -\left\langle \mathcal{L}'_v(v_1, \pi_1), v_1 - v_2 \right\rangle &\geq \mathcal{L}(v_2, \pi_1) - \mathcal{L}(v_1, \pi_1) - \frac{L}{2}\|v_1 - v_2\|^2; \\ \left\langle \mathcal{L}'_v(v_2, \pi_2), v_1 - v_2 \right\rangle &\geq \mathcal{L}(v_1, \pi_2) - \mathcal{L}(v_2, \pi_2) - \frac{L}{2}\|v_1 - v_2\|^2. \end{aligned}$$

The sum of these four inequalities yields

$$-\delta_1 - \delta_2 \geq -L\|v_1 - v_2\|^2 - 2C\|v_1 - v_2\|\|\pi_1 - \pi_2\|,$$

and thus

$$\Delta_K \geq (b_G - \varepsilon(L + \alpha))\|v_1 - v_2\|^2 + \frac{\varepsilon}{\rho}\left(b_H - \rho\frac{C^2}{\alpha}\right)\|\pi_1 - \pi_2\|^2 \tag{4.16}$$

for all $\alpha \in \mathbb{R}_+^*$. If (4.13) holds true, then

$$M : (v, \pi) \mapsto (b_G - \varepsilon(L + \alpha))\, v + (\varepsilon/\rho)\left(b_H - \rho C^2/\alpha\right)\pi$$

is strongly monotone. With this definition of M, (4.16) is nothing but (3.5). Let $(\gamma_1, \gamma_2) \in \Gamma(x_1) \times \Gamma(x_2)$. Inequality (4.12), the definition (4.15) of Σ and that of M imply that

$$\begin{aligned}\langle \gamma_1 - \gamma_2, x_1 - x_2 \rangle &\geq \frac{1}{D} \left\| \mathcal{L}'_\pi(v_1, \pi_1) - \mathcal{L}'_\pi(v_2, \pi_2) \right\|^2 \\ &\geq \frac{1}{4D} \frac{\varepsilon}{\rho} \left(b_H - \rho \frac{C^2}{\alpha} \right) \|\Sigma(x_1) - \Sigma(x_2)\|^2_{M^{-1}} .\end{aligned}$$

Because $b_H > \rho C^2/\alpha$, inequality (3.6) is satisfied for an appropriate definition of A. With this value of A, assumptions on ε of Theorem 3.4 hold true if

$$0 < \varepsilon < \frac{2}{4D} \frac{\varepsilon}{\rho} \left(b_H - \rho \frac{C^2}{\alpha} \right).$$

With the condition upon ρ, this inequality is satisfied. Therefore, Theorem 3.4 implies Theorem 4.7. □

COROLLARY 4.8. *Consider Problem* (4.7) *where* Λ *is a convex-concave l.s.c.-u.s.c. mapping which has a saddle point* $(v^\sharp, \pi^\sharp)$ *over the closed convex set* $\mathcal{V}^{\mathrm{ad}} \times \Pi^{\mathrm{ad}}$. *Assume that:*

- Λ *is convex-concave, differentiable with a hemicontinuous derivative which enjoys the partial Dunn property over* Π *with constant* D,
- $\Lambda'_v(\cdot, \pi)$ *is* L*-Lipschitz (uniformly in* π*).*

Then, $\Lambda'_\pi(\cdot, \pi)$ *is Lipschitz with constant* $C = \sqrt{DL}$. *Moreover, if* ε *and* ρ *satisfy*

$$0 < \rho < \frac{1}{D(2 + L/\alpha)}, \quad 0 < \varepsilon < \frac{1}{\alpha + L}, \tag{4.17}$$

for some $\alpha > 0$, *then the sequence* $\{(v^k, \pi^k)\}$ *obtained with Algorithm* (4.10) *has cluster points in the weak topology. Every such cluster point is a saddle point of* Λ *over* $\mathcal{V}^{\mathrm{ad}} \times \Pi^{\mathrm{ad}}$. *If* Λ' *is continuous when the domain and range spaces are equipped with the weak topology, the whole sequence converges towards a saddle point.*

Proof. Let $v_1, v_2 \in \mathcal{V}^{\mathrm{ad}}$ and $\pi \in \Pi^{\mathrm{ad}}$. The Lipschitz property of $\Lambda'_v(\cdot, \pi)$ and the partial Dunn property of Λ yield

$$\begin{aligned}L\|v_1 - v_2\|^2 &\geq \|\Lambda'_v(v_1, \pi) - \Lambda'_v(v_2, \pi)\| \|v_1 - v_2\| \\ &\geq \langle \Lambda'_v(v_1, \pi) - \Lambda'_v(v_2, \pi), v_1 - v_2 \rangle \\ &\geq \frac{1}{D} \|\Lambda'_\pi(v_1, \pi) - \Lambda'_\pi(v_2, \pi)\|^2.\end{aligned}$$

This inequality shows that $\Lambda'_\pi(\cdot, \pi)$ is Lipschitz with constant $\sqrt{DL}$. This point being fixed, Corollary 4.8 is a straightforward application of Theorem 4.7. □

5. CONVERGENCE RATE

The convergence of Algorithm (1.1) is proved by showing that a certain Lyapounov function decreases. This function is generally not the square of the Hilbert space distance of the current point from the optimum, and this does not generally decrease at each iteration. Obtaining a classical linear

convergence result is thus out of question. Nevertheless, if Ψ is strongly monotone, linear convergence "*in the average*" (see Definition 5.1 below) can be proved. Moreover, this strong monotonicity assumption can be weakened if the symmetric part of Ξ is linear. Then, the Lipschitz property of the inverse yields a convergence rate again linear in the average.

The evaluation of this convergence rate in the particular case of symmetric auxiliary operators ($\Xi = K'$) shows that the ratio of the strong monotonicity constant (or of the inverse of the Lipschitz constant of the inverse) of Ψ by the Lipschitz constant of Ξ is decisive for the convergence rate.

Results about convergence rates are stated in §5.2 and 5.3 and proofs are found in Appendices B and C, respectively.

5.1. Mean convergence rate and Lyapounov functions

Definition 5.1. The sequence $\{x^k\}$ converges in the average linearly towards $x^\sharp$, with the rate $\kappa \in [0,1)$, if

$$\forall k \in \mathbb{N}: \limsup_{n\to+\infty} \sqrt[n]{\frac{\|x^{k+n}-x^\sharp\|}{\|x^k-x^\sharp\|}} \leq \kappa. \tag{5.1}$$

Proposition 5.2. *Let $\varphi : X \to \mathbb{R}$ be a convex mapping over X^{ad} which has a minimum $x^\sharp$ such that:*

$$\exists \widehat{a} \in \mathbb{R}_+^*,\ \forall x \in X^{\mathrm{ad}}:\ \varphi(x) \geq \frac{\widehat{a}}{2}\|x-x^\sharp\|^2; \tag{5.2}$$

$$\exists \widehat{L} \in \mathbb{R}_+^*,\ \forall x \in X^{\mathrm{ad}}:\ \varphi(x) \leq \frac{\widehat{L}}{2}\|x-x^\sharp\|^2. \tag{5.3}$$

Let $\{x^k\}$ be a sequence of X^{ad} such that

$$\exists \eta \in (0,1];\ \forall k \in \mathbb{N}:\ \varphi(x^{k+1})-\varphi(x^k) \leq -\eta\left(\varphi(x^k)-\varphi(x^\sharp)\right). \tag{5.4}$$

Then $\{x^k\}$ converges in the average linearly towards $x^\sharp$ with the rate $\sqrt{1-\eta}$.

Proof. Setting $x = x^\sharp$, inequalities (5.2) and (5.3) show that $\varphi(x^\sharp) = 0$. From (5.4), it follows that

$$\varphi(x^{k+1})-\varphi(x^\sharp) \leq (1-\eta)\left(\varphi(x^k)-\varphi(x^\sharp)\right)$$

and thus, for all $n \in \mathbb{N}^*$,

$$\varphi(x^{k+n})-\varphi(x^\sharp) \leq (1-\eta)^n\left(\varphi(x^k)-\varphi(x^\sharp)\right).$$

Since $\varphi(x^\sharp) = 0$, (5.2) and (5.3) yield

$$\frac{\widehat{a}}{2}\|x^{k+n}-x^\sharp\|^2 \leq \varphi(x^{k+n})-\varphi(x^\sharp) \quad \text{and} \quad \varphi(x^k)-\varphi(x^\sharp) \leq \frac{\widehat{L}}{2}\|x^k-x^\sharp\|^2.$$

The last three inequalities imply that

$$\forall n \in \mathbb{N}^*,\ \|x^{k+n}-x^\sharp\| \leq (1-\eta)^{n/2}\sqrt{\frac{\widehat{L}}{\widehat{a}}}\|x^k-x^\sharp\|,$$

from which, for all $\beta < \eta$, it follows that

$$\exists l \in \mathbb{N}^*:\ \forall n > l,\ \|x^{k+n}-x^\sharp\| \leq (1-\beta)^{n/2}\|x^k-x^\sharp\|,$$

and thus

$$\limsup_{n\to+\infty} \sqrt[n]{\frac{\|x^{k+n}-x^\sharp\|}{\|x^k-x^\sharp\|}} \leq \sqrt{1-\beta}.$$

Since this inequality holds for every $\beta < \eta$, it holds for $\beta = \eta$. □

5.2. Strong monotonicity

5.2.1. General case.

Assumption 5.3. The assumptions of Theorem 3.4 are satisfied and

- the operator Ψ has a unique zero $x^\sharp$ such that

$$\exists a \in \mathbb{R}_+^* : \ \forall x \in \mathrm{dom}\Psi,\ \forall \psi_x \in \Psi(x),\ \langle \psi_x, x-x^\sharp\rangle \geq a\|x-x^\sharp\|^2, \quad (5.5)$$

- the operator Σ is Lipschitz continuous over every bounded subset of the domain of Ψ.

Remark 5.4. Inequality (5.5) obviously holds if Ψ is strongly monotone. Yet, it may be satisfied by an operator which does not share this property. For example, the subdifferential of the function $x \to |x| + I_B(x)$, where B is a closed bounded interval of $\mathbb{R}$, satisfies (5.5).

Theorem 5.5. *Under Assumption 5.3, Algorithm* (1.1) *converges strongly and in the average linearly towards* $x^\sharp$.

5.2.2. Symmetric auxiliary operator.

Proposition 5.6. *In addition to Assumption 5.3, suppose that* $\Sigma = 0$ *and that* K' *is Lipschitz continuous with constant* B *over the domain of* Ψ. *Algorithm* (1.1) *becomes*

$$0 \in \left(\frac{1}{\varepsilon}\left(K'(x^{k+1}) - K'(x^k)\right) + \Psi(x^{k+1})\right). \quad (5.6)$$

It converges strongly and in the average linearly with the rate $\sqrt{1/(1+2\varepsilon a/B)}$.

Remark 5.7. Since $\Sigma = 0$ here, one may consider that A in Theorem 3.4 is arbitrarily small and that ε may be taken arbitrarily large. Hence the rate of the average linear convergence can be made arbitrarily good by increasing ε. This is a well known fact in the context of the proximal point algorithm to which (1.1) is reduced in the case considered here. However, from the practical point of view, (1.1) cannot always be solved easily without further specialization. This is the topic of the next remark which explains how a practical bound then arises for ε.

Remark 5.8. Consider the auxiliary function

$$K : x \mapsto \frac{\sigma}{2}\|x\|^2 - \varepsilon J_1, \quad (5.7)$$

where J_1 is a differentiable function with an A_1-Lipschitz derivative. By introducing the term $-\varepsilon J_1$, one aims at canceling the coupling terms that would prevent decomposition in the original problem. In order to ensure the strong convexity of K (and thus the convergence of (5.6)), one is led to choose $\sigma > \varepsilon A_1$. For $\sigma = \varepsilon A_1 + \varepsilon\mu$ with μ positive and "small", one obtains $(B_K/2\varepsilon) = A_1 + \mu/2$. The convergence rate of (5.6) becomes then

$$\sqrt{1/\big(1+a/(A_1+\mu/2)\big)}.$$

Since μ can be chosen as small as necessary, it is the ratio a/A_1 between the strong convexity constant — of the whole operator Ψ — and the Lipschitz constant of J_1' which determines the best possible convergence rate. Therefore, it is once more the relationship between the whole operator and the part which enters the relaxation scheme at each iteration that plays the major role.

5.2.3. Partial relaxation and nonsymmetric auxiliary operator. We now consider Algorithm (1.6). As it has already been noticed, this algorithm is a particular case of (1.1) in which $\Xi = K' - \varepsilon\Psi_1$. Because the operator Ψ_1 is not necessarily symmetric, generally Ξ is not symmetric. Therefore, Algorithm (1.6) is not a particular case of (5.6).

Proposition 5.9. *Assume that the operator Ψ is such that $\Psi = \Psi_1 + \Psi_2$ where*

- *Ψ_1 is a single-valued monotone operator which has the Dunn property with constant A_1 over the domain of Ψ;*
- *Ψ_2 is maximal monotone.*

Assume that K has a B-Lipschitz derivative and is strongly convex with constant b over the domain of Ψ, then if $0 < \varepsilon < 2b/A_1$, Algorithm (1.6) converges in the average linearly with the rate

$$\sqrt{1/\big(1 + (2\varepsilon/B)a(1 - \varepsilon A_1/2b)\big)}.$$

For $\varepsilon = b/A_1$, this convergence rate is $\sqrt{1/(1 + ba/BA_1)}$.

5.3. Lipschitz inverse

5.3.1. General case.

Assumption 5.10. The assumptions of Theorem 3.4 are satisfied and

- the operator Ψ has a unique zero $x^\sharp$, and it satisfies

$$\exists \nu \in \mathbb{R}_+^* : \forall x \in \mathrm{dom}\Psi,\ \forall \psi_x \in \Psi(x),\ \|\psi_x\| \geq \nu \|x - x^\sharp\|; \qquad (5.8)$$

- the auxiliary operator Σ is Lipschitz over every bounded subset of the domain of Ψ;
- the auxiliary function K is quadratic and has a continuous derivative.

Remark 5.11. Note that if (5.5) holds true, then (5.8) is satisfied with the same constant. If Ψ has the Dunn property and satisfies (5.8), (5.5) holds true. Nevertheless, (5.8) is weaker than (5.5). The rotation over $\mathbb{R}^2$

$$\Psi_r : (y, z) \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}$$

satisfies (5.8) but not (5.5).

Theorem 5.12. *Under Assumption 5.10, Algorithm (1.1) strongly converges towards the unique zero of Ψ and its convergence is in the average linear.*

5.3.2. Symmetric auxiliary operator.

Proposition 5.13. *Under Assumption 5.10, consider Algorithm* (5.6) *where K' is linear continuous with norm B. This algorithm strongly converges towards the unique zero of Ψ. This convergence is in the average linear with the convergence rate $\sqrt{1/(1+(\varepsilon\nu/B)^2)}$.*

Appendix: proof of technical results

Appendix A. Convergence

A.1. Symmetric case

Proof of Proposition 3.1. Because J is convex and l.s.c., the mapping $x \mapsto \partial[J+I_{X^{\mathrm{ad}}}](x)$ is maximal monotone. Because Ξ has the Dunn property (and thus the Lipschitz property) over every bounded subset of X, it is hemicontinuous over X. Proposition 2.6 thus shows that the operator

$$\Phi^k : x \mapsto \varepsilon\partial[J+I_{X^{\mathrm{ad}}}](x) + \Xi(x) - \Xi(x^k),$$

is maximal monotone. Let $x \in X^{\mathrm{ad}}$ and $\phi_x^k \in \Phi^k(x)$. Because Ξ is monotone and J is convex, one has that

$$\langle \phi_x^k, x - x^k \rangle \geq \varepsilon\big(J(x) - J(x^k)\big).$$

Since J is coercive, Φ^k satisfies Assumption 2.3. Problem (3.1) has thus a solution (see Proposition 2.5).

Let j^{k+1} be an element of $\partial[J+I_{X^{\mathrm{ad}}}](x^{k+1})$ such that

$$\frac{1}{\varepsilon}\big(\Xi(x^{k+1}) - \Xi(x^k)\big) + j^{k+1} = 0. \tag{A.1}$$

From the definition of the subdifferential, it follows that

$$\begin{aligned}\frac{1}{\varepsilon}\langle \Xi(x^{k+1}) - \Xi(x^k), x^k - x^{k+1}\rangle &= \langle j^{k+1}, x^{k+1} - x^k\rangle \\ &\geq J(x^{k+1}) - J(x^k).\end{aligned} \tag{A.2}$$

Because Ξ is monotone, the sequence $\{J(x^k)\}$ is decreasing. Because $\{J(x^k)\}$ is decreasing and bounded from below by $J(x^\sharp)$, it converges. This, and the coercivity of J, prove that $\{x^k\}$ is bounded. This sequence has thus cluster points in the weak topology.

Let A be the Dunn constant of Ξ over a bounded set which contains all the x^k. From (A.2), it follows that

$$-\frac{1}{\varepsilon A}\|\Xi(x^{k+1}) - \Xi(x^k)\|^2 \geq J(x^{k+1}) - J(x^k). \tag{A.3}$$

Because $\{J(x^k)\}$ converges, the sequence $\{\|\Xi(x^{k+1}) - \Xi(x^k)\|\}$ converges towards 0. From (A.1), it follows then that $\{j^k\}$ strongly converges to zero. Because the graph of a maximal monotone operator is closed in $X_{\mathrm{weak}} \times X_{\mathrm{strong}}$ (see [4]), every cluster point of $\{x^k\}$ in the weak topology is a zero of $\partial[J+I_{X^{\mathrm{ad}}}]$ and thus a solution of (1.2).

A.2. General case

LEMMA A.1. *Under the assumptions of Theorem 3.4, Problem* (1.1) *has a unique solution: there is a unique* x^{k+1} *such that there exists* $\psi^{k+1} \in \Psi(x^{k+1})$ *such that*

$$0 = \frac{1}{\varepsilon}\big(\Xi(x^{k+1}) - \Xi(x^k)\big) + \psi^{k+1}. \tag{A.4}$$

Proof. Since K' is Lipschitz over every bounded subset and Σ is hemicontinuous, $\Xi = K' + \varepsilon\Sigma$ is hemicontinuous over X. The operator Ψ is maximal monotone. Proposition 2.6 ensures that if $(1/\varepsilon)\Xi + \Psi$ is monotone, it is maximal monotone. In addition, from Proposition 2.5, it follows that if $(1/\varepsilon)\Xi + \Psi$ is strongly monotone, Problem (1.1) has a unique solution.

Let us prove the strong monotonicity of this operator. Let $x_1, x_2 \in \mathrm{dom}\Psi$ and let $(\psi_1, \psi_2) \in \Psi(x_1) \times \Psi(x_2)$. From (3.5) and (3.6), it follows that

$$\Big\langle \frac{1}{\varepsilon}\Xi(x_1) + \psi_1 - \frac{1}{\varepsilon}\Xi(x_2) - \psi_2, x_1 - x_2 \Big\rangle \geq$$
$$\frac{1}{\varepsilon}\|x_1 - x_2\|_M^2 + \langle \Sigma(x_1) - \Sigma(x_2), x_1 - x_2\rangle + \frac{1}{A}\|\Sigma(x_1) - \Sigma(x_2)\|_{M^{-1}}^2 .$$

Furthermore,

$$\langle \Sigma(x_1) - \Sigma(x_2), x_1 - x_2\rangle \geq -\|\Sigma(x_1) - \Sigma(x_2)\|_{M^{-1}}\|x_1 - x_2\|_M \tag{A.5}$$
$$\geq -\frac{1}{A}\|\Sigma(x_1) - \Sigma(x_2)\|_{M^{-1}}^2 - \frac{A}{4}\|x_1 - x_2\|_M^2 .$$

Since $\varepsilon < 2/A$, the previous inequalities prove that $\Xi/\varepsilon + \Psi$ is strongly monotone. Consequently, Problem (1.1) has a unique solution. □

LEMMA A.2. *Suppose that the assumptions of Theorem 3.4 hold true and consider the function* φ_0 :

$$x \mapsto \frac{1}{\varepsilon}\Big(K(x^\sharp) - K(x) - \big\langle K'(x), x^\sharp - x\big\rangle\Big) + \big\langle \Sigma(x^\sharp) - \Sigma(x), x^\sharp - x\big\rangle \tag{A.6}$$

where $x^\sharp$ *is a zero of* Ψ. *The sequence* $\{x^k\}$ *generated by Algorithm* (1.1) *satisfies, for all* $\alpha > 0$,

$$\varphi_0(x^{k+1}) - \varphi_0(x^k) \leq -\frac{1-\alpha}{2\varepsilon}\|x^{k+1} - x^k\|_M^2 + \frac{\varepsilon}{2\alpha}\|\Sigma(x^k) - \Sigma(x^\sharp)\|_{M^{-1}}^2$$
$$- \big\langle \psi^{k+1}, x^{k+1} - x^\sharp\big\rangle. \tag{A.7}$$

Proof. From the definition of φ_0, one has that

$$\varphi_0(x^{k+1}) - \varphi_0(x^k) = \frac{1}{\varepsilon}\Big(K(x^k) - K(x^{k+1}) - \big\langle K'(x^k), x^k - x^{k+1}\big\rangle\Big)$$
$$+ \frac{1}{\varepsilon}\big\langle K'(x^k) - K'(x^{k+1}), x^\sharp - x^{k+1}\big\rangle + \big\langle \Sigma(x^k) - \Sigma(x^{k+1}), x^\sharp - x^{k+1}\big\rangle$$
$$- \big\langle \Sigma(x^k) - \Sigma(x^\sharp), x^k - x^{k+1}\big\rangle.$$

Since $\Xi = K' + \varepsilon\Sigma$, equality (A.4) and (3.5) yield

$$\varphi_0(x^{k+1}) - \varphi_0(x^k) \leq -\frac{1}{2\varepsilon}\|x^k - x^{k+1}\|_M^2 - \big\langle \psi^{k+1}, x^{k+1} - x^\sharp\big\rangle$$
$$+ \|\Sigma(x^k) - \Sigma(x^\sharp)\|_{M^{-1}}\|x^k - x^{k+1}\|_M. \tag{A.8}$$

Since, for all $\alpha > 0$, one has that

$$\|\Sigma(x^k) - \Sigma(x^\sharp)\|_{M^{-1}} \|x^k - x^{k+1}\|_M \leq \frac{\varepsilon}{2\alpha} \|\Sigma(x^k) - \Sigma(x^\sharp)\|^2_{M^{-1}} + \frac{\alpha}{2\varepsilon} \|x^k - x^{k+1}\|^2_M,$$

inequality (A.8) implies (A.7). □

Proof of Theorem 3.4. The existence and the uniqueness of the solution of (1.1) is established by Lemma A.1. To complete the proof, we proceed in four stages:

- we introduce a Lyapounov function φ;
- we show that the sequence $\{\varphi(x^k)\}$ decreases;
- we derive that every cluster point of $\{x^k\}$ is a zero of Ψ;
- we finally prove additional convergence results under further assumptions.

Choice of a Lyapounov function. Let $\varphi : X \to \mathbb{R}$ be defined as

$$\varphi : x \mapsto \varphi_0(x) + \frac{1}{A} \|\Sigma(x^\sharp) - \Sigma(x)\|^2_{M^{-1}}, \tag{A.9}$$

where φ_0 is defined by (A.6) and where $x^\sharp$ is a zero of Ψ. Inequality (3.5) and the Schwarz inequality yield

$$\begin{aligned} \varphi(x) &\geq \frac{1}{2\varepsilon} \|x - x^\sharp\|^2_M - \|\Sigma(x) - \Sigma(x^\sharp)\|_{M^{-1}} \|x - x^\sharp\|_M \\ &\quad + \frac{1}{A} \|\Sigma(x) - \Sigma(x^\sharp)\|^2_{M^{-1}} \\ &\geq \frac{1}{2} \left(\frac{1}{\varepsilon} - \frac{A}{2} \right) \|x - x^\sharp\|^2_M \end{aligned} \tag{A.10}$$

by using (A.5) again. Since $0 < \varepsilon < 2/A$, this inequality proves that φ is coercive and bounded from below by zero.

The sequence $\{\varphi(x^k)\}$ is decreasing. Because $0 \in \Psi(x^\sharp)$, inequalities (A.7) and (3.6) imply that

$$\varphi(x^{k+1}) - \varphi(x^k) \leq -\frac{1}{2\varepsilon} (1 - \alpha) \|x^k - x^{k+1}\|^2_M - \frac{1}{A} \left(1 - \frac{\varepsilon A}{2\alpha} \right) \|\Sigma(x^k) - \Sigma(x^\sharp)\|^2_{M^{-1}}$$

for any $\alpha \in \mathbb{R}^*_+$. Let $\alpha = (1 + \varepsilon A/2)/2$; then,

$$\varphi(x^{k+1}) - \varphi(x^k) \leq -\frac{1}{(1 + \varepsilon A/2) A} \left(1 - \frac{\varepsilon A}{2} \right) \|\Sigma(x^k) - \Sigma(x^\sharp)\|^2_{M^{-1}} - \frac{1}{4\varepsilon} \left(1 - \frac{\varepsilon A}{2} \right) \|x^k - x^{k+1}\|^2_M. \tag{A.11}$$

Since $0 < \varepsilon < 2/A$, the sequence $\{\varphi(x^k)\}$ is decreasing. Since φ is bounded from below, this sequence converges. Because φ is coercive, we can conclude that $\{x^k\}$ is bounded and thus that it has cluster points in the weak topology.

Furthermore, since $\{\varphi(x^k)\}$ converges, the difference of two successive terms tends to zero, and then, (A.11) implies that

$$\lim_{k\to+\infty} \|x^k - x^{k+1}\| = 0; \tag{A.12}$$

$$\lim_{k\to+\infty} \|\Sigma(x^k) - \Sigma(x^\sharp)\| = 0. \tag{A.13}$$

Every cluster point of $\{x^k\}$ *is a zero of* Ψ. Because K' is Lipschitz continuous over every bounded subset, (A.12) implies that

$$\lim_{k\to+\infty} \|K'(x^k) - K'(x^{k+1})\| = 0.$$

With (A.13), it follows that $\lim_{k\to+\infty} \|\Xi(x^k) - \Xi(x^{k+1})\| = 0$ and, consequently, $\{\psi^k\}$ strongly converges towards 0. Because the graph of a maximal monotone operator is closed in $X_{\text{weak}} \times X_{\text{strong}}$ (see Ref. [4]), we can conclude that every cluster point of $\{x^k\}$ in the weak topology is a zero of Ψ.

Uniqueness of the cluster point. Assume that K' is continuous when both the domain and the range spaces are equipped with the weak topology. Let x_1 and x_2 be two cluster points of the sequence $\{x^k\}$. These points are zeros of Ψ. Thus, we can consider the Lyapounov function φ_{x_1} (respectively, φ_{x_2}) defined by (A.9) when choosing $x^\sharp = x_1$ (respectively, $x^\sharp = x_2$).

Let l_1 and l_2 denote the limits of the sequences $\{\varphi_{x_1}(x^k)\}$ and $\{\varphi_{x_2}(x^k)\}$, respectively. Consider a subsequence $\{x^{k_i}\}$ of $\{x^k\}$ which converges weakly towards x_2. The strong convergence of $\{\Sigma(x^{k_i})\}$ towards $\Sigma(x^\sharp)$ ([2]), and the foregoing assumption of continuity upon K' yield

$$\begin{aligned} l_1 - l_2 &= \lim_{i\to+\infty} \left(\varphi_{x_1}(x^{k_i}) - \varphi_{x_2}(x^{k_i})\right) \\ &= \lim_{i\to+\infty} \left(K(x_1) - K(x_2) - \left\langle K'(x^{k_i}), x_1 - x_2\right\rangle\right) \\ &= K(x_1) - K(x_2) - \langle K'(x_2), x_1 - x_2\rangle \\ &= \varphi_{x_1}(x_2). \end{aligned}$$

Points x_1 and x_2 playing symmetric parts, it follows that

$$-\varphi_{x_1}(x_2) = l_2 - l_1 = \varphi_{x_2}(x_1),$$

hence $\varphi_{x_1}(x_2) = 0$. With help of inequality (A.10), one then concludes that $x_1 = x_2$.

Strong convergence. Since $\{\psi^k\}$ converges strongly towards 0, if Ψ is strongly monotone, the sequence $\{x^k\}$ converges strongly.

REMARK A.3. Theorem 3.4 does not assume the monotonicity of Ξ. However, we proved the strong monotonicity of $\Xi/\varepsilon + \Psi$.

REMARK A.4. The term $(1/\varepsilon)\left(K(x^\sharp) - K(x) - \left\langle K'(x), x^\sharp - x\right\rangle\right)$, which appears in the definition of the Lyapounov function (see (A.6), (A.9)), is the Lyapounov function which is used in most of the convergence proofs of the "Auxiliary Problem Principle" introduced by Cohen (see [7]). This term also occurs in the study of proximal algorithm using Bregman functions by Chen and Teboulle [6]. At least when K' is linear, it may be understood as a measure of the distance to $x^\sharp$ in the metric defined by K'.

[2]Remember $\Sigma(x^\sharp)$ is unique even if $x^\sharp$ is not.

Appendix B. Convergence rate in the strongly monotone case

Lemma B.1. *Suppose Assumption 5.3 holds true and consider the function*

$$\varphi_1 : x \mapsto \varphi_0(x) + \frac{\varepsilon}{2}\|\Sigma(x) - \Sigma(x^\sharp)\|^2_{M^{-1}} + \left(1 - \frac{\varepsilon A}{2}\right) a\|x - x^\sharp\|^2, \quad \text{(B.1)}$$

where φ_0 is defined by (A.6). *The sequence $\{x^k\}$ generated by Algorithm* (1.1) *is such that*

$$\varphi_1(x^{k+1}) - \varphi_1(x^k) \leq -\left(1 - \frac{\varepsilon A}{2}\right) a\|x^k - x^\sharp\|^2. \quad \text{(B.2)}$$

Furthermore, φ_1 satisfies inequalities (5.2) *and* (5.3) *over a closed convex subset which contains all the x^k.*

Proof. Since $0 \in \Psi(x^\sharp)$ and $\varepsilon < 2/A$, the inequalities (3.6) and (5.5) yield

$$\begin{aligned}\left\langle \psi^{k+1}, x^{k+1} - x^\sharp \right\rangle &= \frac{\varepsilon A}{2}\left\langle \psi^{k+1}, x^{k+1} - x^\sharp \right\rangle + \left(1 - \frac{\varepsilon A}{2}\right)\left\langle \psi^{k+1}, x^{k+1} - x^\sharp \right\rangle \\ &\geq \frac{\varepsilon}{2}\|\Sigma(x^{k+1}) - \Sigma(x^\sharp)\|^2_{M^{-1}} + \left(1 - \frac{\varepsilon A}{2}\right) a\|x^{k+1} - x^\sharp\|^2.\end{aligned}$$

For $\alpha = 1$, the inequality (A.7) takes the following form

$$\varphi_0(x^{k+1}) - \varphi_0(x^k) \leq \frac{\varepsilon}{2}\|\Sigma(x^k) - \Sigma(x^\sharp)\|^2_{M^{-1}} - \left\langle \psi^{k+1}, x^{k+1} - x^\sharp \right\rangle.$$

These two inequalities and the definition of φ_1 yield (B.2).

It remains to prove (5.2) and (5.3) for φ_1. Theorem 3.4 proves that the sequence $\{x^k\}$ is bounded. Hence, there exists a closed and bounded convex subset to which the x^k (and $x^\sharp$) belong and over which K' and Σ are Lipschitz continuous. Consequently, (5.3) is satisfied by φ_1 for some $\widehat{L}$.

Let $x \in \mathrm{dom}\Psi$. The definition (B.1) of φ_1 and a standard consequence of (3.5) yield

$$\begin{aligned}\varphi_1(x) &\geq \frac{1}{2\varepsilon}\|x - x^\sharp\|^2_M - \|\Sigma(x) - \Sigma(x^\sharp)\|_{M^{-1}}\|x - x^\sharp\|_M \\ &\quad + \frac{\varepsilon}{2}\|\Sigma(x) - \Sigma(x^\sharp)\|^2_{M^{-1}} + \left(1 - \frac{\varepsilon A}{2}\right) a\|x - x^\sharp\|^2 \\ &\geq \left(1 - \frac{\varepsilon A}{2}\right) a\|x - x^\sharp\|^2.\end{aligned}$$

Since $\varepsilon < 2/A$, (5.2) is satisfied by φ_1 for some $\widehat{a}$. □

Proof of Theorem 5.5. We use Proposition 5.2 and the technical Lemma B.1. This lemma shows that φ_1 defined by (B.1) satisfies (5.2) and (5.3). Since $0 < \varepsilon < 2/A$, inequalities (B.2) and (5.3) show that (5.4) holds true.

Proof of Proposition 5.6. Since $\Sigma = 0$, the mapping φ_1 defined by (B.1) takes the following form:

$$\varphi_1 : x \mapsto \frac{1}{\varepsilon}\left(K(x^\sharp) - K(x) - \left\langle K'(x), x^\sharp - x \right\rangle\right) + a\left(1 - \frac{\varepsilon A}{2}\right)\|x - x^\sharp\|^2.$$

Lemma B.1 shows that this function satisfies (5.2) and (5.3). Since K' is B-Lipschitz, for all $x \in \mathrm{dom}\Psi$, one has that

$$\varphi_1(x) \leq \left(\frac{B}{2\varepsilon} + a\left(1 - \frac{\varepsilon A}{2}\right)\right)\|x - x^\sharp\|^2.$$

This inequality and (B.2) show that

$$\varphi_1(x^{k+1}) - \varphi_1(x^k) \leq -\frac{a(1-\varepsilon A/2)}{B/2\varepsilon + a(1-\varepsilon A/2)}\varphi_1(x^k).$$

Moreover, because (3.6) is satisfied for every $A > 0$, one has that

$$\varphi_1(x^{k+1}) - \varphi_1(x^k) \leq -\frac{a}{B/2\varepsilon + a}\varphi_1(x^k).$$

This proof is completed by using Proposition 5.2.

Proof of Proposition 5.9. Once more, Proposition 5.2 will be used. For $M = bI$, (3.5) holds true. As for (3.6), it is implied by the Dunn property of Ψ_1 with constant A_1 and the fact that $\Sigma = -\Psi_1$: this leads to take $A = A_1/b$ in (3.6). With this evaluation of A, consider the Lyapounov function φ_1 defined by (B.1). Inequality (B.2) of Lemma B.1 now reads:

$$\varphi_1(x^{k+1}) - \varphi_1(x^k) \leq -\left(1 - \frac{\varepsilon A_1}{2b}\right) a\|x^k - x^\sharp\|^2. \tag{B.3}$$

The operator K' is B-Lipschitz and Ψ_1 has the Dunn property. From the definition (B.1) of φ_1 and from the present form of (3.6), it follows that, for all $x \in \mathrm{dom}\Psi$,

$$\varphi_1(x) \leq \frac{B}{2\varepsilon}\|x - x^\sharp\|^2 + \left(1 - \frac{\varepsilon A_1}{2b}\right)\left(a\|x - x^\sharp\|^2 - \left\langle \Psi_1(x) - \Psi_1(x^\sharp), x - x^\sharp\right\rangle\right).$$

Since $\varepsilon < 2b/A_1$, the monotonicity of Ψ_1 yields

$$\varphi_1(x^k) \leq \left(\frac{B}{2\varepsilon} + a\left(1 - \frac{\varepsilon A_1}{2b}\right)\right)\|x^k - x^\sharp\|^2.$$

This and (B.3) imply that

$$\begin{aligned}\varphi_1(x^{k+1}) - \varphi_1(x^k) &\leq -\frac{(1-\varepsilon A_1/2b)a}{B/2\varepsilon + a(1-\varepsilon A_1/2b)}\varphi_1(x^k) \\ &= -\left(1 - \frac{1}{1 + (2\varepsilon/B)a(1-\varepsilon A_1/2b)}\right)\varphi_1(x^k).\end{aligned}$$

Because $0 < \varepsilon < 2b/A_1$, with this inequality, Proposition 5.2 can be used, given that (5.2) and (5.3) are also satisfied thanks to Lemma B.1. The proof is thus complete.

Appendix C. Convergence rate and Lipschitz continuity of the inverse

Lemma C.1. *Under Assumption 5.10, consider the mapping*

$$\varphi_2 : x \mapsto \frac{1}{\varepsilon}\left\langle (K')^{-1}\left(\Xi(x) - \Xi(x^\sharp)\right), \Xi(x) - \Xi(x^\sharp)\right\rangle + \frac{\varepsilon}{B}\left(1 - \frac{\varepsilon A}{2}\right)\nu^2\|x - x^\sharp\|^2. \tag{C.1}$$

The sequence $\{x^k\}$ *that Algorithm* (1.1) *generates is such that*

$$\varphi_2(x^{k+1}) - \varphi_2(x^k) \leq -\frac{\varepsilon}{B}\left(1 - \frac{\varepsilon A}{2}\right)\nu^2\|x^k - x^\sharp\|^2. \tag{C.2}$$

Moreover, φ_2 *satisfies* (5.2) *and* (5.3) *over a closed convex set which contains the whole sequence.*

Proof. The assumptions of Theorem 3.4 are satisfied, the sequence $\{x^k\}$ is thus bounded. Consequently, there exists a closed bounded convex subset which contains all the x^k and $x^\sharp$. Over this subset, Ξ is Lipschitz continuous. Moreover, K is strongly convex. Therefore, $(K')^{-1}$ is Lipschitz continuous. Since Ξ and $(K')^{-1}$ are Lipschitz continuous, φ_2 satisfies (5.3). The function K being convex, the term $\langle (K')^{-1}\left(\Xi(x) - \Xi(x^\sharp)\right), \Xi(x) - \Xi(x^\sharp)\rangle$ is nonnegative. Since $0 < \varepsilon < 2/A$, φ_2 satisfies (5.2).

Let $\zeta^k = (1/\varepsilon)\langle (K')^{-1}\left(\Xi(x^k) - \Xi(x^\sharp)\right), \Xi(x^k) - \Xi(x^\sharp)\rangle$. One has that

$$\begin{aligned}\zeta^{k+1} - \zeta^k = {} & \frac{2}{\varepsilon}\langle (K')^{-1}\left(\Xi(x^{k+1}) - \Xi(x^k)\right), \Xi(x^{k+1}) - \Xi(x^\sharp)\rangle \\ & - \frac{1}{\varepsilon}\langle (K')^{-1}\left(\Xi(x^{k+1}) - \Xi(x^k)\right), \Xi(x^{k+1}) - \Xi(x^k)\rangle.\end{aligned}$$

Since there exists $\psi^{k+1} \in \Psi(x^{k+1})$ such that $\Xi(x^{k+1}) - \Xi(x^k) = -\varepsilon\psi^{k+1}$, one has that

$$\begin{aligned}\zeta^{k+1} - \zeta^k = {} & \underbrace{-2\langle \psi^{k+1}, (K')^{-1}\left(\Xi(x^{k+1}) - \Xi(x^\sharp)\right)\rangle}_{\delta} \\ & - \varepsilon\langle (K')^{-1}\left(\psi^{k+1}\right), \psi^{k+1}\rangle.\end{aligned} \tag{C.3}$$

Inequality (3.6) yields

$$\begin{aligned}\delta &= -2\langle \psi^{k+1}, x^{k+1} - x^\sharp\rangle \\ &\quad - 2\varepsilon\langle (K')^{-1}\psi^{k+1}, \Sigma(x^{k+1}) - \Sigma(x^\sharp)\rangle \\ &\leq -2\langle \psi^{k+1}, x^{k+1} - x^\sharp\rangle \\ &\quad + 2\varepsilon\left\|(K')^{-1}\left(\psi^{k+1}\right)\right\|_M \|\Sigma(x^{k+1}) - \Sigma(x^\sharp)\|_{M^{-1}} \\ &\leq -\frac{2}{A}\|\Sigma(x^{k+1}) - \Sigma(x^\sharp)\|^2_{M^{-1}} \\ &\quad + 2\varepsilon\left\|(K')^{-1}\left(\psi^{k+1}\right)\right\|_M \|\Sigma(x^{k+1}) - \Sigma(x^\sharp)\|_{M^{-1}} \\ &\leq \frac{\varepsilon^2}{2}A\left\|(K')^{-1}\left(\psi^{k+1}\right)\right\|^2_M.\end{aligned}$$

From (3.5), it follows that

$$\delta \leq \frac{\varepsilon^2 A}{2}\langle (K')^{-1}\left(\psi^{k+1}\right), \psi^{k+1}\rangle. \tag{C.4}$$

The operator K' is B-Lipschitz and symmetric. Thus, it has the Dunn property with constant B. One has that

$$\begin{aligned}\left\langle (K')^{-1}(\psi^{k+1}), \psi^{k+1}\right\rangle &= \left\langle K'\left((K')^{-1}(\psi^{k+1})\right), (K')^{-1}(\psi^{k+1})\right\rangle \\ &\geq \frac{1}{B}\left\|K'\left((K')^{-1}(\psi^{k+1})\right)\right\|^2 \\ &= \frac{1}{B}\|\psi^{k+1}\|^2.\end{aligned}$$

This inequality, together with (C.3) and (C.4), implies that

$$\begin{aligned}\zeta^{k+1} - \zeta^k &\leq -\varepsilon\left(1 - \frac{\varepsilon A}{2}\right)\left\langle (K')^{-1}(\psi^{k+1}), \psi^{k+1}\right\rangle \\ &\leq -\frac{\varepsilon}{B}\left(1 - \frac{\varepsilon A}{2}\right)\|\psi^{k+1}\|^2.\end{aligned}$$

This inequality, the definition (C.1) of φ and (5.8) finally yield (C.2). □

Proof of Theorem 5.12. The mapping φ defined by (C.1) satisfies the conditions (5.2) and (5.3) of Proposition 5.2 (see Lemma C.1). The condition (5.4) is the result of (5.3) and of (C.2) also proved in Lemma C.1.

Proof of Proposition 5.13. Because $\Sigma = 0$, the Lyapounov function φ which is defined by (C.1) takes the following form:

$$\varphi : x \mapsto \frac{1}{\varepsilon}\langle K'(x) - K'(x^\sharp), x - x^\sharp\rangle + \frac{\varepsilon}{B}\left(1 - \frac{\varepsilon A}{2}\right)\nu^2\|x - x^\sharp\|^2.$$

Since K' is strongly monotone and Lipschitz continuous, φ satisfies (5.2) and (5.3). Moreover, for $x \in \mathrm{dom}\Psi$, one has that

$$\varphi(x) \leq \left(\frac{B}{\varepsilon} + \frac{\varepsilon}{B}\left(1 - \frac{\varepsilon A}{2}\right)\nu^2\right)\|x - x^\sharp\|^2. \tag{C.5}$$

The sequence $\{x^k\}$ generated by Algorithm (1.1) satisfies (C.2). Inequality (C.5) thus implies

$$\varphi(x^{k+1}) - \varphi(x^k) \leq -\frac{\varepsilon^2\nu^2(1 - \varepsilon A/2)}{B^2 + \varepsilon^2\nu^2(1 - \varepsilon A/2)}\varphi(x^k).$$

Since Assumption (3.6) is satisfied for every $A > 0$, this may be reduced to

$$\varphi(x^{k+1}) - \varphi(x^k) \leq -\frac{\varepsilon^2\nu^2/B^2}{1 + \varepsilon^2\nu^2/B^2}\varphi(x^k).$$

The proof is completed by using this inequality and Proposition 5.2.

References

[1] B. Baillon, G. Haddad: Quelques Propriétés des Opérateurs Angles-Bornés et n-cycliquement Monotones, *Israel Journal of Mathematics*, **26**, 1977, 137–150.

[2] V. Barbu, T. Precupanu: *Convexity and Optimization in Banach Spaces*, Mathematics and its applications, East European Series, D. Reidel Publishing Company, 1986.

[3] L. Bregman: The Relaxation Method of Finding the Common Point of Convex Sets and its Application to the Solution of Problems in Convex Programming, *U.S.S.R. Comput. Math. and Math. Phys.*, **7**, 1967, 200–217.

[4] H. Brezis: *Opérateurs Maximaux Monotones*, Lectures Notes 5, North-Holland, 1973.

[5] Y. Censor, S.A. Zenios: Proximal Minimization Algorithm with D-Functions, *Journal of Optimization Theory and Application*, **73**, 1992, 451–464.
[6] G. Chen, M. Teboulle: Convergence Analysis of a Proximal-like Minimization Algorithm using Bregman Functions, *SIAM Journal of Optimization*, **3**, 1993, 538–543.
[7] G. Cohen: Optimization by Decomposition and Coordination: a Unified Approach, *IEEE Transactions on Automatic Control*, **23**, 1978, 222–232.
[8] G. Cohen: Auxiliary Problem Principle and Decomposition of Optimization Problems, *Journal of Optimization Theory and Applications*, **32**, 1980, 277–305.
[9] G. Cohen: Auxiliary Problem Principle Extended to Variational Inequalities, *Journal of Optimization Theory and Applications*, **59**, 1988, 369–390.
[10] G. Cohen: *Décomposition et Coordination en Optimisation Déterministe Différentiable et Non Différentiable*, Thesis Dissertation, University of Paris Dauphine, 1984.
[11] J.C. Dunn: Convexity, Monotonicity and Gradient Process in Hilbert Spaces, *Journal of Mathematical Analysis and Applications*, **53**, 1976, 145–158.
[12] J. Eckstein: Nonlinear Proximal Point Algorithms using Bregman Functions, *Math. Oper. Research*, **18**, 1993, 202–226.
[13] J. Eckstein, D.P. Bertsekas: On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators, *Mathematical Programming*, **55**, 1992, 293–318.
[14] D. Gabay: Applications of the Method of Multipliers to Variational Inequalities, in *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Valued Problems*, M. Fortin and R. Glowinski, eds., North-Holland, Amsterdam, 1983, 299–331.
[15] J.-B. Hiriart-Urruty, C. Lemaréchal: *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, Germany, 1993.
[16] B. Lemaire: The Proximal Algorithm, in *International Series of Numerical Mathematics*, J.-P. Penot ed., Birkhäuser-Verlag, Basel, **87**, 1989, 73–87.
[17] P.-L. Lions, B. Mercier: Splitting Algorithm for the Sum of Two Nonlinear Operators, *SIAM J. Numer. Anal.*, **16**, 1979, 964–979.
[18] F.J.R. Luque: Asymptotic Convergence Analysis of the Proximal Point Algorithm, *SIAM Journal of Control and Optimization*, **22**, 1984, 277–293.
[19] B. Martinet: Régularisation d'Inéquations Variationnelles par Approximations Successives, *Revue d'Automatique, d'Informatique et de Recherche Opérationnelle*, Série Rouge, **3**, 1970, 154–159.
[20] M.A. Mataoui: *Contributions à la Décomposition et à l'Agrégation des Problèmes Variationnels*, Thesis Dissertation, École des Mines de Paris, Paris, France, 1990.
[21] J.-J. Moreau: Proximité et Dualité dans un Espace Hilbertien, *Bull. Soc. Math. France*, **93**, 1965, 273–299.
[22] M. Patriksson: *A Unified Framework of Descent Algorithm for Nonlinear Programs and Variational Inequalities*, Thesis Dissertation, Linköping University, 1993.
[23] R.T. Rockafellar: Local Boundedness of Nonlinear Monotone Operators, *Michigan Mathematics Journal*, **16**, 1969, 397–407.
[24] R.T. Rockafellar: On the Maximality of Sums of Nonlinear Monotone Operators, *Transactions of the American Mathematical Society*, **149**, 1970, 75–88.
[25] R.T. Rockafellar: Monotone Operators and the Proximal Point Algorithm, *SIAM Journal of Control and Optimization*, **14**, 1976, 877–898.
[26] P. Tseng: Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities, *SIAM Journal of Control and Optimization*, **29**, 1991, 119–138.
[27] K. Yosida: *Functional Analysis*, Springer-Verlag, 1964.
[28] D. Zhu, P. Marcotte: New Classes of Generalized Monotonicity, *Journal of Optimization Theory and Applications*, **87**, 1995, 457–471.