M2AN - MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE

P. MAHEY
PHAM DINH TAO

Partial regularization of the sum of two maximal monotone operators

M2AN - Modélisation mathématique et analyse numérique, tome 27, nº 3 (1993), p. 375-392
<http://www.numdam.org/item?id=M2AN_1993__27_3_375_0>

MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS
MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE

(Vol. 27, n° 3, 1993, p. 375 à 392)

PARTIAL REGULARIZATION OF THE SUM OF TWO MAXIMAL MONOTONE OPERATORS (*)

by P. MAHEY (1) and PHAM DINH TAO (2)

Communicated by P. L. LIONS

Abstract. — *To find a zero of the sum of two maximal monotone operators, we analyze a two-steps algorithm where the problem is first approximated by a regularized one and the regularization parameter is then reduced to converge to a solution of the original problem. We give a formal proof of the convergence which, in that case, is not ergodic. The main result is a generalization of one given by Brezis [4] who has considered operators of the form* $I + A + B$. *Additional insight on the underlying existence problems and on the kind of convergence we aim at are given with the hypothesis that one of the two operators is strongly monotone. A general scheme for the decomposition of large scale convex programs is then induced.*

Résumé. — *On analyse ici un algorithme qui recherche un zéro de la somme de deux opérateurs maximaux monotones. On résoud une séquence de problèmes régularisés dont la solution converge vers la solution cherchée quand le paramètre de régularisation tend vers zéro. On évite alors de se restreindre à la convergence ergodique. Le résultat principal est une généralisation d'un théorème de Brézis qui a considéré des opérateurs de la forme* $I + A + B$. *On peut alors raffiner ces résultats dans le cas fortement monotone. Finalement, on propose un schéma général d'algorithme de décomposition pour la programmation convexe.*

1. INTRODUCTION

We consider the following inclusion problem in a finite dimensional space X $(X = \mathbb{R}^n)$:

$$\text{Find} \quad x \in X \quad \text{such that} \quad 0 \in (A + B)x \qquad (P)$$

where A and B are two maximal monotone operators.

We analyze here the convergence of some specific splitting algorithms, i.e. such that separate steps on A and B are made to avoid the difficulties

(*) Manuscrit reçu le 12 mai 1992.
(1) Artemis-IMAG, BP 53X, 38041 Grenoble.
(2) LMAI-INSA Rouen, BP 8, 76131 Mont Saint-Aignan.

derived from the coupling between the two operators like in decomposition methods (other splitting algorithms are described in [9]).

Maximal monotone operators in Hilbert spaces have been extensively studied, mainly in the context of evolution equations, by Brezis [4] who, in particular, has given some conditions for the sum of two maximal monotone operators to be maximal monotone. Indeed, this fact happens when one of the operators is Lipschitzian. This result induces a strategy to solve (P) which consists in substituting one of the operators, say A, by a regularized approximation, for instance its Moreau-Yosida approximation A_λ, i.e., for some $\lambda > 0$:

$$A_\lambda = \frac{1}{\lambda}(I - (I + \lambda A)^{-1}).$$

Then, the regularized problem is :

$$\text{Find } x \in X \text{ such that } 0 \in (A_\lambda + B)x. \qquad (P_\lambda)$$

Properties of A_λ are well-known (see Moreau [13] and Brezis [4]). In particular, it tends in some way to the original operator A when $\lambda \downarrow 0$. We analyze here the convergence of the following iteration :

$$x_k^{t+1} = (I + \lambda_k B)^{-1} (I + \lambda_k A)^{-1} x_k^t \qquad (1)$$

where $\{\lambda_k\}$ is an a priori defined sequence of positive numbers.

It is shown in Lions [8] and Passty [14] that, if both subscripts k and t are incremented together, i.e., if the iteration takes the form :

$$z_{k+1} = (I + \lambda_k B)^{-1} (I + \lambda_k A)^{-1} x_k \qquad (2)$$

with $\lambda_k \downarrow 0$ and $\sum \lambda_k = +\infty$, then the sequence of weighted averages

$$z_k = \sum_{i=1}^{k} \lambda_i x_i / \sum_{i=1}^{k} \lambda_i$$

converges to a solution of (P) if one exists. Our purpose is to show the convergence of a two-steps version of the iterative process (1) where we iterate first on subscript t for a fixed k, then increment k and perform another cycle of iterations. We show in particular that the sequence generated by 1 for a fixed k converges to a solution of (P_λ) when it exists. The important thing is that the solution x_λ converges to a solution of (P) when $\lambda \downarrow 0$, avoiding ergodic convergence. In fact, the parameter λ acts like a penalty parameter in a penalty method for constrained programming, which means that we cannot set it to a too small value at first to avoid ill-conditioning but we may reduce it if some convergence criterion is not met. In fact, the iterative process (1) can be seen as a decomposed version of the

Proximal Point Algorithm but the need of reducing the parameter λ to zero forbids to apply the classical results of convergence given by Rockafellar [17] in the case of a single operator.

As in all iterative methods where different subproblems are to be solved at each step, two important questions arise beside the necessity of convergence : the problem of existence of solutions of the successive subproblems which is not necessarily guaranteed by the existence of a solution of (P) and the computational simplicity of these subproblems without which there is no interest to manipulate the original problem in that way. We shall see that these questions are linked to the choice of the operator to be regularized. Sufficient conditions for existence of solutions in the subproblems and for the convergence of the whole sequence are obtained when one of the operators is strongly monotone. In this case, it is natural to regularize the other one.

The interest for Proximal methods has increased recently motivated by the theoretical works of Rockafellar [17] and Spingarn [19]. The numerical behaviour in the applications, initially limited to variational inequalities as in the pioneering work of Martinet [11], have been analyzed in the context of Mathematical Programming (see [3], [7], [12], [21] for example). The positive aspects that come out of these experience are :

The numerical stability.

The ability of dealing with nonsmoothness.

The nice effect of the regularization on the decentralization of subproblems in the decomposition of large-scale programs (see [2], [10] and [15]).

The negative aspects are mostly :

The complexity of the proximal computation which limits the domain of applications.

The slow rate of convergence.

This latter drawback is minimized by the constatation that, in some specific applications as the Fermat-Weber location problem ([12]) or the numerical solution of evolution equations for low level vision ([7]), the proximal algorithms seem to exhibit the best stability and efficiency.

A particular motivation we have in mind beside the general problem (P) is the case where $A = \partial f$, the subdifferential mapping of a convex lsc function f and B is the subdifferential mapping of the indicator function χ_C of a closed convex subset C of X. Then, under mild conditions, (P) is the optimality condition for the following convex program :

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & x \in C\,. \end{array} \tag{3}$$

When C is an appropriate subspace of a product space, this formulation is attractive to decompose large-scale problems. The subspace represents the coupling between the subsystems. The iteration 1 decomposes in two steps, one proximal iteration on f which maintains separability and a projection on C which satisfies the coupling. This idea has been used by Pierra [15] in this context, but the convergence proof we propose here, beside the fact that it applies to the splitting of general operators, is more straightforward and less restrictive than his proof.

2. PRELIMINARY RESULTS

In this section, A and B are maximal monotone operators on X, finite dimensional, with respective domains $D(A)$ and $D(B)$, and we denote by A_λ (resp. B_λ) their Moreau-Yosida approximations, i.e. $A_\lambda = \frac{1}{\lambda}(I - (I + \lambda A)^{-1})$. The operator $(I + \lambda A)^{-1}$ is generally called the resolvent. We recall below some important properties of these operators which proofs can be found in Brezis [4] :

PROPOSITION 1 :

1. *A is closed in the sense that its graph $Gr(A) = \{(x, y) \in X \times X \mid y \in A(x)\}$ is closed in $X \times X$. Furthermore, for any $x \in D(A)$, the set $A(x)$ is closed, convex and nonempty.*
2. *$(I + \lambda A)^{-1}$ is a contraction defined on the whole space X for any $\lambda > 0$.*
3. *A_λ is maximal monotone and lipschitzian with ratio $1/\lambda$.*
4. *$\forall x \in X,\ A_\lambda x \in A((I + \lambda A)^{-1} x)$.*
5. *$\overline{D(A)}$ is convex, the range of $(I + \lambda A)^{-1}$ is $D(A)$ and*

$$\lim_{\lambda \downarrow 0} (I + \lambda A)^{-1} x = \operatorname{Proj}_{\overline{D(A)}} x .$$

6. *For any $\lambda > \mu > 0$, for any $x \in D(A)$, $\|A_\lambda x\| \leq \|A_\mu x\| \leq \|A_0 x\|$, where $A_0 x$ is the minimum norm element of Ax.*

Moreover, when $\lambda \downarrow 0$, we have the following limits :

$$\forall x \in D(A),\ \lim \|A_\lambda x\| = \|A_0 x\|$$

$$If\ x \notin D(A),\ \lim \|A_\lambda x\| = +\infty$$

The Proximal Point algorithm is based on proposition 1.2. Rockafellar [17] has analyzed the convergence of the proximal iteration :

$$x_{k+1} = (I + \lambda_k A)^{-1} x_k \tag{4}$$

The main result tells us that, when λ_k is bounded away from zero, problem (P) has a solution if and only if the sequence $\{x_k\}$ is bounded. Then it converges to a point x^∞ such that $0 \in Ax^\infty$ and $\lim_{k \to \infty} \|A_{\lambda_k} x_k\| = 0$.

Rockafellar concludes that the convergence becomes faster when λ_k increases. Observe that the Proximal Point method, as it searches a fixed point of $(I + \lambda A)^{-1}$, converges to a zero of the operator A_λ which is also a zero of A. This is no more true if we look for a zero of $A + B$ and substitute $A + B$ by $A_\lambda + B$. In fact, we need a sequence $\{\lambda_k\}$ which decreases to zero. Brezis and Lions [5] have given the following conditions for the convergence of the sequence $\{x_k\}$ generated by (4) to a zero of A :

- a sufficient condition for the convergence is : $\sum \lambda_k^2 \to \infty$;
- if $\sum \lambda_k^2$ is bounded and $\sum \lambda_k \to \infty$, then the sequence $\{x_k\}$ converges to a point which is generally not a solution, but on the other hand, the sequence of average solutions $\{z_k\}$, with $z_k = \dfrac{\sum_{i=1}^{k} \lambda_i x_i}{\sum_{i=1}^{k} \lambda_i}$, does converge ;
- if $A = \partial f$, the subdifferential of a convex lsc mapping, then it suffices to suppose that $\sum \lambda_k \to \infty$ to get the convergence of the sequence $\{x_k\}$.

3. REGULARIZATION TECHNIQUES

The regularization we are interested in is the substitution of a maximal monotone operator by its Moreau-Yosida approximation. It is well-known (see [4] or [18]) that the sum of two maximal monotone operators A and B is maximal monotone when $\operatorname{int}(D(A)) \cap D(B) \neq \emptyset$.

This condition may be refined when $A = \partial f$ and $B = \partial g$ are the subdifferential mappings of two proper convex lsc functions (see [16] and [20]). In that case, we have the following inclusions :

$$ri\,(\operatorname{dom} f) \subset D(A) \subset \operatorname{dom} f$$
$$ri\,(\operatorname{dom} g) \subset D(B) \subset \operatorname{dom} g$$

where $\operatorname{dom} f = \{x \in X \mid f(x) < +\infty\}$ is the effective domain of f and ri denotes the relative interior. Then, $A + B$ is maximal (i.e. coincides with $\partial(f + g)$) if $ri\,(\operatorname{dom} f) \cap ri\,(\operatorname{dom} g) \neq \emptyset$. Moreover, if one function, for instance f, is polyhedral, $D(A) = \operatorname{dom} f$ and we can omit ri in the precedent condition.

These sufficient conditions for the maximality of $A + B$ could be compared to the necessary condition of existence of solutions of (P):

$$D(A + B) = D(A) \cap D(B) \neq \emptyset\,.$$

Anyway, the critical cases where $A + B$ is not maximal remain of important interest in the theory or regularization and decomposition. We shall develop these ideas further in section 5.

A natural way to regularize the sum of two operators is to regularize one of them. We are faced with three distinct strategies : to regularize A only, to regularize B only or to regularize both.

In the first case, we approximate (P) by (P_λ) which can be transformed in the following way :

$$0 \in x - (I + \lambda A)^{-1} x + \lambda B x$$

which, in turn, as B is maximal monotone, is equivalent to :

$$x = (I + \lambda B)^{-1} (I + \lambda A)^{-1} x . \tag{5}$$

Thus, x is a fixed point of the operator $T_\lambda = (I + \lambda B)^{-1} (I + \lambda A)^{-1}$, which is a contraction, inducing the fixed point iteration :

$$\begin{aligned} &x^0 \in X \\ &x^{t+1} = (I + \lambda B)^{-1} (I + \lambda A)^{-1} x^t . \end{aligned} \tag{6}$$

Before discussing the convergence of iteration (6), we recall some useful results on the existence of solutions for problems (P) and (P_λ).

3.1. Existence results

Existence of solutions is linked with the property of surjectivity of the operator ; we know for example that, if T is a maximal monotone operator such that $D(T)$ is bounded, then T is surjective (see Brezis [4]).

If T is the sum of two maximal monotone operators, it is clear that its domain is bounded if one of the domains is bounded. On the other hand, we have seen before that the sum of two maximal monotone operators is maximal when one of them is single-valued and Lipschitzian. This implies that (P_λ) has a solution if $D(B)$ is bounded (this last result is similar to the conditions of existence for variational inequalities involving single-valued and hemicontinuous monotone operators on bounded convex sets given in Stampacchia [20]).

Another interesting situation which we develop hereafter is the case where one operator is strongly monotone :

If B is strongly monotone, i.e., if $\exists \alpha > 0$ such that :

$$\langle x - x', y - y' \rangle \geq \alpha \|x - x'\|^2 , \ \forall x, x' \text{ and } \forall y \in Bx , \ \forall y' \in Bx'$$

then (P_λ) has a unique solution for any $\lambda > 0$. Indeed, the operator T_λ is now a strict contraction :

$$\|T_\lambda(x) - T_\lambda(x')\| \leq$$
$$\leq \frac{1}{1 + \alpha\lambda} \|(I + \lambda A)^{-1} x - (I + \lambda A)^{-1} x'\| \leq \frac{1}{1 + \alpha\lambda} \|x - x'\| \tag{7}$$

Note that, in this case, (P) has at most one solution.

Furthermore, as B is strongly monotone, $C = B - \alpha I$ is maximal monotone and (P_λ) is equivalent to :

$$0 \in x_\lambda + \frac{1}{\alpha} C x_\lambda + \frac{1}{\alpha} A_\lambda x_\lambda .$$

We observe here that, when $\alpha = 1$, the present situation fits in the model analyzed by Brezis [4], page 34 :

Note that a similar results to (7) exists if we regularize B in the place of A. We will show in Theorem 3 that the strongly monotone case allows some refinements in the final convergence theorem.

3.2. Convergence of the proximal iteration

THEOREM 1 : *For some $\lambda > 0$, assume that the sequence $\{x^t\}$ generated by* (6) *is bounded, then it converges to a point x_λ which is a solution of the regularized problem* (P_λ).

Proof : Each resolvent is nonexpansive as was seen before (Proposition 1.2), so that the composed operator T_λ is nonexpansive too. As the iterates are all in a bounded set, by a theorem of Browder *et al.* (see [6]), the sequence (6) converges to a fixed point of T_λ, say x_λ, which means that :

$$x_\lambda = (I + \lambda B)^{-1} (I + \lambda A)^{-1} x_\lambda .$$

∎

Remark : If we know that (P_λ) has a solution x_λ (this is the case when one operator is strongly monotone), then it is a fixed point of T_λ and for any t, we have :

$$\|x^t - x_\lambda\| = \|T^t x^0 - T^t x_\lambda\| \leq \|x^0 - x_\lambda\| .$$

We have seen in section 3.1 that the choice of the operator to be regularized is not indifferent. More insight on that question can be given by the application of the regularization to the convex problem (3) (see too Theorem 3) :

a) Regularizing A :

Then (P_λ) is the optimality condition for the regularized convex program :

$$\begin{array}{c} \text{Minimize} \quad f_\lambda(x) \\ x \in C \end{array} \tag{8}$$

where $f_\lambda(x) = \inf \left\{ f(z) + \frac{1}{2\lambda} \|z - x\|^2 \right\}$. It is known [13] that for every x, there is a unique z_x such that $f_\lambda(x) = f(z_x) + \frac{1}{2\lambda}\|z_x - x\|^2$, and that $z_x = (I + \lambda\, \partial f)^{-1} x$ The regularized function f_λ is smooth and its gradient is $A_\lambda x = \frac{1}{\lambda}(x - z_x)$

A sufficient condition for the existence of solutions of (P) and (P_λ) is that C is a bounded set Observe that (3) has a unique solution if f is essentially strictly convex (i e f is strictly convex on any convex subset of $D(\partial f)$) But it can be seen from results on the Legendre transform that the essentially strict convexity of f implies the essentially strict convexity of f_λ, which means that (8) has a unique solution, too

Now, iteration (6) takes the form

$$
\begin{aligned}
& x^0 \in C \\
& z^{t+1} = \operatorname{Argmin}\left\{f(z) + \frac{1}{2\lambda}\|z - x^t\|^2\right\} \\
& x^{t+1} = \operatorname{Proj}_C z^{t+1}
\end{aligned}
\tag{9}
$$

b) Regularizing B

If we regularize B, we get the following equation

$$0 \in Ax + B_\lambda x \tag{10}$$

and, as $B_\lambda x = \frac{1}{\lambda}(x - \operatorname{Proj}_C x)$, (10) is the optimality condition for the unconstrained problem

$$
\begin{aligned}
&\text{Minimize} \quad f(x) + \frac{1}{2\lambda} d(x, C)^2 \\
&\qquad\qquad x \in X
\end{aligned}
\tag{11}
$$

where $d(x, C)$ is the Euclidian distance between x and C In fact, the constraint is treated here like in a penalty method Again, (11) has a solution if C is bounded Equation (10) leads to the following fixed point equation

$$z_\lambda = (I + \lambda A)^{-1} (I + \lambda B)^{-1} z_\lambda \tag{12}$$

and we obtain the same sequences $\{z^t\}$ as in (9) if we initialize by $z^1 = (I + \lambda A)^{-1} x^0$

In conclusion, we are faced with two dependent sequences, $\{x^t\}$ and $\{z^t\}$, the first one in C and the second one outside C such that $x^t = \operatorname{Proj}_C z^t$ Both converge from Theorem 1 and if x_λ and z_λ are their respective limit points, we have $x_\lambda = \operatorname{Proj}_C z_\lambda$

Remarks :

1. If both operators were regularized, we should add the penalty term to (8) but no direct fixed point iteration could be induced.
2. Some additional insight on these models is given at the end of section 4.

4. MAIN CONVERGENCE RESULT

We analyze here the convergence of the sequence $\{x_k\}$ defined by : $x_k \in X$ and x_k solves problem (P_λ) for $\lambda = \lambda_k$, where $\{\lambda_k\}$ is a sequence of positive real numbers such that $\lambda_k \downarrow 0$, when $k \to \infty$. We assume in this section that (P_λ) has a solution for any $\lambda_k > 0$. Then, for each k, x_k satisfies :

$$x_k = (I + \lambda_k B)^{-1} (I + \lambda_k A)^{-1} x_k . \tag{13}$$

To simplify our notations, we denote by A_k the operator A_{λ_k}. Then we denote by $\{y_k\}$ the sequence associated to the sequence $\{x_k\}$ such that $y_k = A_k x_k$, $\forall k$. That sequence plays a central role in the convergence results we present hereafter.

THEOREM 2 :

1. *Let x be a solution of (P) and $y \in Ax \cap (-Bx)$. Then, $\|y_k\| \leq \|y\|$ for any $\lambda_k > 0$.*
2. *Let x be a limit point of $\{x_k\}$ and assume that the sequence $\{y_k\}$ is bounded. Then, x solves (P).*
3. *Assume that the sequence $\{x_k\}$ has a limit point. Then, (P) admits a solution if and only if the sequence $\{y_k\}$ is bounded.*
4. *If (P) has a unique solution x^* and if the sequence $\{x_k\}$ is bounded, then $x_k \to x^*$ and $y_k \to y^*$ which is the element of minimum norm in $Ax^* \cap (-Bx^*)$.*

Proof :

1. As $y \in -Bx$ and $y_k \in -Bx_k$, the monotonicity of B implies that :

$$\langle y_k - y, x_k - x \rangle \leq 0 .$$

But, as $y_k = A_k x_k = \dfrac{1}{\lambda_k} (x_k - (I + \lambda_k A)^{-1} x_k)$, we can write :

$$x_k = \lambda_k y_k + (I + \lambda_k A)^{-1} x_k . \tag{14}$$

Then, $\langle y_k - y, \lambda_k y_k \rangle \leq - \langle y_k - y, (I + \lambda_k A)^{-1} x_k - x \rangle$. The right-hand side is non positive because $y_k = A_k x_k \in A(I + \lambda_k A)^{-1} x_k$ (Proposition 1.4), $y \in Ax$ and A is monotone. Then :

$$\|y_k\|^2 \leq \|y\| \, \|y_k\| .$$

2. We show first that any limit point x of the sequence $\{x_k\}$ is in $\overline{D(A)} \cap \overline{D(B)}$: from (13) and Proposition 1.5, we see that $x_k \in D(B)$. Again, from (14) and Proposition 1.5, we see that $x_k \in \lambda_k y_k + D(A)$. Then, as y_k is bounded, the first term of the sum tends to zero when $\lambda_k \downarrow 0$ and the limit point x must be in $\overline{D(A)}$. Then, $x \in \overline{D(A)} \cap \overline{D(B)}$.

As x_k solves (P_λ) for $\lambda = \lambda_k$, it satisfies :

$$- A_k x_k \in Bx_k \quad \text{or equivalently :} \quad - y_k \in Bx_k .$$

By assumption, the sequence $\{y_k\}$ is bounded and we can extract a convergent subsequence which we denote too $\{y_k\}$ to avoid overloaded notations. Let y be its limit.

On the other hand, if we put $z_k = (I + \lambda_k A)^{-1} x_k$, then, from Proposition 1.4, we know that $y_k \in Az_k$ for any k. To prove that $\lim z_k = x$, we compute $\|z_k - x\|$:

$$\begin{aligned}\|z_k - (I + \lambda_k A)^{-1} x + (I + \lambda_k A)^{-1} x - x\| &\leq \\ \leq \|(I + \lambda_k A)^{-1} x_k - (I + \lambda_k A)^{-1} x\| &+ \|(I + \lambda_k A)^{-1} x - x\| .\end{aligned}$$

The first norm is bounded by $\|x_k - x\|$ because $(I + \lambda_k A)^{-1}$ is a contraction. The second norm tends to zero when λ_k tends to zero (Proposition 1.5 with the fact that $x \in \overline{D(A)}$).

Finally, as $y_k \to y$ and $z_k \to x$, A being a closed map, we must have $y \in Ax$.

But, $- y_k \in Bx_k$, and from the closedness of B, we conclude that $- y \in Bx$, which means that x solves (P).

3. Immediate consequence of the two first parts.

4. Suppose now that problem (P) possesses a unique solution x^*. As the sequence $\{x_k\}$ is bounded, we know from part 2 that its limit point solves (P). Then the sequence $\{x_k\}$ has a unique limit point which is exactly x^* and $x_k \to x^*$. Let y^* be the element of minimum norm in $Ax^* \cap (- Bx^*)$ (note that y^* exists and is unique according to Proposition 1.1). We already know from the first result of the theorem that, for any k :

$$\|y_k\| \leq \|y^*\| .$$

Let y be a limit point of the sequence $\{y_k\}$. Then, $\|y\| \leq \|y^*\|$. But, as in the second section of the present proof, we have $y \in Ax^* \cap (- Bx^*)$. It follows that $y = y^*$ and the whole sequence $\{y_k\}$ converges to y^*. ■

Remarks : We have used the fact in section 2 of the above proof that $x_k \in D(B)$ for any k. Then, the sequence $\{x_k\}$ is bounded if $D(B)$ is bounded. This is the case for example if (P) is the optimality condition for problem (3) with C bounded.

On the other hand, if $D(A)$ is bounded, the existence of a solution of (P) implies that $\{x_k\}$ is bounded. Indeed, it follows from the first part of the theorem that the sequence $\{y_k\}$ is bounded and relation (14) with the same arguments as in the proof of the second part shows that x_k remains bounded.

An important case where unicity of the solutions is observed is the strongly monotone case. Indeed, if B is strongly monotone and A is regularized, problem (P) has at most one solution and problem (P_λ) has a unique solution for any $\lambda > 0$ (see section 3.1). Then, part 4 of the precedent theorem can be refined in the following theorem which is, in fact, inspired by a theorem given in Brezis ([4], p. 35) and which proof will therefore be omitted here :

THEOREM 3 : *Suppose that B is strongly monotone with constant α and that A is regularized. Then, (P) admits a unique solution x^* if and only if the sequence $\{y_k\}$ is bounded. In this case, $x_k \to x^*$ and $y_k \to y^*$ which is the element of minimum norm in $Ax^* \cap (-Bx^*)$.*

Moreover, we have the following estimation :

$$\|x_k - x^*\| \leqslant (\alpha^{-1} \lambda_k \|y^*\| \, \|y_k - y^*\|)^{1/2} = 0(\sqrt{\lambda}) .$$

Application : Strongly convex programming

Let us come back to the convex program (3) with the additional requirement that f is strongly convex on the closed convex set C. This means now that $B = \partial f$ is strongly monotone. In this particular case, we know that (P) has a unique solution. In addition, Theorem 3 implies the following convergence results :

1. B is regularized. Then, (P_λ) is :

Minimize f_λ on $x \in C$, which admits a unique solution x_λ for any $\lambda > 0$.

2. A is regularized. Then (P_λ) is :

Minimize $f(x) + \dfrac{1}{2\lambda} d(x, C)^2$ which admits a unique solution x'_λ for any $\lambda > 0$.

When $\lambda \downarrow 0$, $x_\lambda \to x^*$ and $x'_\lambda \to x^*$. Furthermore, wet get the estimations in a neighbourhood of x^* :

$$\|x_\lambda - x^*\| = 0(\sqrt{\lambda})$$
$$\|x'_\lambda - x^*\| = 0(\sqrt{\lambda}) .$$

Observe that these last results mean that the whole sequence generated by that specific penalty method converges at the speed of $\sqrt{\lambda}$. We have then given an elegant and clearcut proof of the convergence for both regularizations.

5. MORE ON THE MAXIMALITY OF $A + B$

We state now problems (P) and (P_λ) in the following equivalent form :

$$(P): \ x = (I + A + B)^{-1} x$$
$$(P_\lambda): x_\lambda = (I + A_\lambda + B)^{-1} x_\lambda .$$

Let consider a convergent subsequence $\{x_{\phi(\lambda)}\}$ of the main sequence $\{x_k\}$ as $\lambda \downarrow 0$, and let x be its limit point. Let define also $u_\lambda = (I + A_\lambda + B)^{-1} x$.

Brezis-Crandall-Pazy's theorem (see [4]) asserts that $(I + A + B)^{-1}$ is nonempty if and only if $\{A_\lambda u_\lambda\}$ is bounded. In this case, we get :

$$\lim_{\lambda \downarrow 0} u_\lambda = (I + A + B)^{-1} x .$$

On the other hand, we can write :

$$\begin{aligned} &\left\| (I + A_\lambda + B)^{-1} x_\lambda - (I + A + B)^{-1} x \right\| \leq \\ &\qquad \left\| (I + A_\lambda + B)^{-1} x_\lambda - (I + A_\lambda + B)^{-1} x \right\| + \left\| u_\lambda - (I + A + B)^{-1} x \right\| \\ &\qquad \leq \|x_\lambda - x\| + \left\| u_\lambda - (I + A + B)^{-1} x \right\| \end{aligned}$$

which implies that $x = (I + A + B)^{-1} x$.

The condition $\{A_\lambda u_\lambda\}$ bounded is obviously closely related to the necessary condition for existence of solutions of problem (P) : $\{A_\lambda x_\lambda\}$ bounded (*cf.* Theorem 2). The main difference is that $\{u_\lambda\}$ is defined from the limits of convergent subsequences of $\{x_k\}$ which turns the test for boundedness of $\{A_\lambda u_\lambda\}$ quite impractical. On the other hand, the assumption « $\{A_\lambda u_\lambda\}$ bounded » is equivalent to « $\{A_\lambda x_\lambda\}$ bounded » in the following sense :

$$x_{\phi(\lambda)} \to x \quad \text{and} \quad \{A_\lambda u_\lambda\} \text{ bounded} \Rightarrow \{A_\lambda x_\lambda\} \text{ bounded}$$
$$x_{\phi(\lambda)} \to x \quad \text{and} \quad \{A_\lambda x_\lambda\} \text{ bounded} \Rightarrow \{A_\lambda u_\lambda\} \text{ bounded}$$

Brezis-Crandall-Pazy's results (BCP) as ours (Theorem 2) have local and optimal features. Moreover, if $A + B$ is assumed maximal (a global assumption), any cluster point x of $\{x_\lambda\}$ is a solution of (P). Observe that the global form of BCP's theorem may be written as :

$$\lim_{\lambda \downarrow 0+} (I + A_\lambda + B)^{-1} x = (I + A + B)^{-1} x , \quad \forall x \Leftrightarrow A + B \text{ maximal} .$$

On the other hand, it is shown in [1] that the maximality of $A + B$ is equivalent to the pointwise convergence of the graph of $A_\lambda + B$ towards the graph of $A + B$. We can now understand easier why the global feature of the maximality assumption on $A + B$ is too restrictive ; indeed, what we are looking for is the convergence of the zeros of $A_\lambda + B$ towards the zeros of $A + B$ and not of the entire graph. Moreover, the following open question which concerns the maximal continuation of the monotone operator $A + B$ is of interest when dealing with decomposition methods :

What happens if $\{A_\lambda x_\lambda\}$ is not bounded ?

We already know that, in that case, (P) has no solution. Moreover, $A + B$ is not maximal if $\{x_\lambda\}$ has a cluster point x (this happens when $\{x_\lambda\}$ is bounded). The problem is to know if x is a zero of a maximal continuation of $A + B$.

Observe that, if $\lim_{\lambda \downarrow 0} x_{\phi(\lambda)} = x \in D(A)$, then $A_{\phi(\lambda)} x \to A_0 x$ when $\lambda \downarrow 0$ (Proposition 1.4), and as :

$$\begin{aligned} \|A_{\phi(\lambda)} x_{\phi(\lambda)} - A_0 x\| &\leq \|A_{\phi(\lambda)} x_{\phi(\lambda)} - A_{\phi(\lambda)} x\| + \|A_{\phi(\lambda)} x - A_0 x\| \\ &\leq \frac{1}{\phi(\lambda)} \|x_{\phi(\lambda)} - x\| + \|A_{\phi(\lambda)} x - A_0 x\| \end{aligned}$$

we obtain the following result which concerns the speed of convergence of $x_{\phi(\lambda)}$ towards x :

$$\phi(\lambda) = 0(\|x_{\phi(\lambda)} - x\|) .$$

The above discussion is illustrated on the following example :

Let consider the problem (Q) in $\mathbb{R}^3$ of finding the closest point to the origin in the intersection of a two-dimensional subspace L and a cylinder C lying on that subspace (see *fig.* 1). We consider below both cases whether $0 \in L \cap C$ or not.

The problem (Q) is then :

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}\|x\|^2 \\ \text{subject to} & x \in L \cap C \end{array}$$

and we may write it as the problem (P') of finding a zero of two maximal monotone operators A' and B' such that : $A' = I$ and $B' = \partial\chi_{L \cap C}$. It is easy to show that $A' + B'$ is maximal (indeed, for any $x \in L \cap C$, $(A' + B')x = \{x\} + (L \cap C)^\perp$, a two-dimensional plane orthogonal to the boundary $L \cap C$).

On the other hand, if we want to achieve a decomposition of the constraints, we can split them in the following way : (P) Find a zero of $A + B$ where $A = I + \partial\chi_L$, and $B = \partial\chi_C$. Then, we get that Dom $(A + B) = L \cap C$ and, for any $x \in L \cap C$, $(A + B)x = \{x\} + L^\perp$, which is the one-dimensional line orthogonal to L at x. Two situations may arise : either

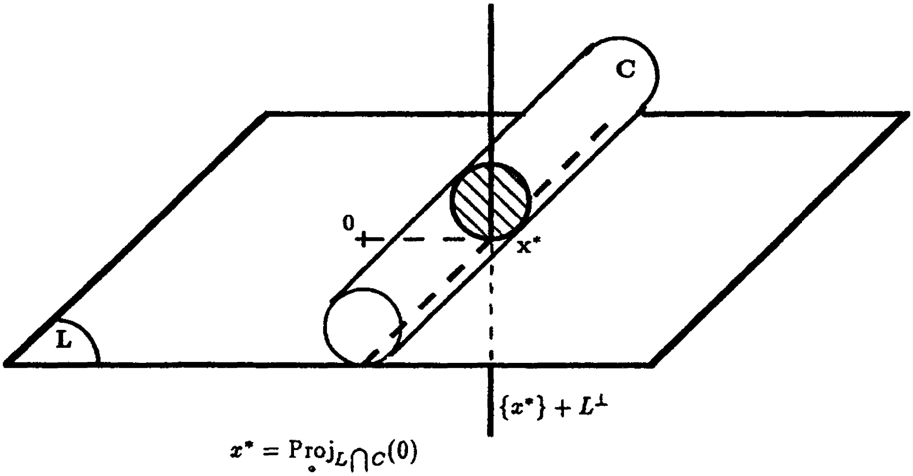

Figure 1. — Example in $\mathbb{R}^3$.

$0 \in L \cap C$ or not. In both cases, $A + B$ is not maximal because $(A + B)x$ is strictly contained in $(A' + B')x$ and, moreover, in the second case, $A + B$ has no zero at all. We show that, even in that case, the splitting method works and converges to the right point.

As in the first case x_λ is the optimal solution of (Q) for any λ, we analyze directly the second case when $A + B$ has no zero : let $x_k \in C$ and z_k be the unique point in L such that :

$$z_k = \mathrm{Argmin}_{x \in L}\left(\frac{1}{2} \|x\|^2 + \frac{1}{2\lambda} \|x - x_k\|^2 \right) .$$

Hence, we can see that z_k is the projection on L of $\dfrac{1}{1+\lambda} x_k$ so that :

$$x_{k+1} = \mathrm{Proj}_C \left(\mathrm{Proj}_L \left(\frac{x_k}{1+\lambda} \right) \right) .$$

For $\lambda > 0$, the sequence $\{x_k\}$ converges to the point x_λ whose construction is shown on figure 2. Observe that the sequence $\{A_\lambda x_\lambda\}$ diverges. It can be proved easily that x_λ is the intersection of the section of C closest to the origin and a hyperbola whose axes are the horizontal axis and a vertical line passing through the solution x^*. When λ tends to zero, the hyperbola degenerates into the two axes and x_λ tends to x^*, which can now be interpreted as a zero of the maximal continuation $A' + B'$ of $A + B$.

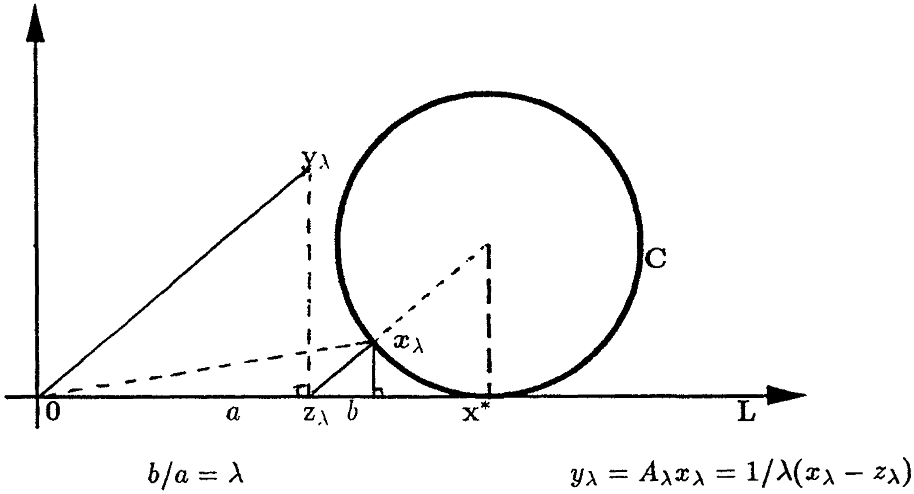

Figure 2. — Construction of x_λ.

6. APPLICATION TO THE DECOMPOSITION OF CONVEX PROGRAMS

We consider first the following problem in $\mathbb{R}^n$:

$$\begin{aligned} &\text{Minimize} \quad \sum_{i=1}^{p} f_i(x) \\ &x \in S = \bigcap_{i=1}^{p} S_i \end{aligned} \tag{15}$$

where each f_i is a proper, convex and lsc function on S_i, which are closed and bounded convex sets of $\mathbb{R}^n$. We assume too that S is not empty.

To decompose it, we create p copies of the variables denoted by $x_1, \ldots, x_p$, with $x_i \in \mathbb{R}^n$ for all i. Then problem (15) is equivalent to solve in the product space $X = (\mathbb{R}^n)^p$ the following problem :

$$\begin{aligned} &\text{Minimize} \quad \sum_{i=1}^{p} f_i(x_i) \\ &x_i \in S_i\,, \quad i = 1, \ldots, p \\ &x = (x_1, \ldots, x_p) \in L = \{x \in X \mid x_1 = \cdots = x_p\}\,. \end{aligned} \tag{16}$$

To apply the precedent algorithm to (16), let, for any i, F_i be the proper convex lsc function defined on $\mathbb{R}^n$ such that :

$$F_i(x_i) = f_i(x_i) + g_i(x_i)$$

where g_i is the indicator function of the set S_i, and F be the proper convex lsc function defined on X such that :

$$F(x) = \sum_{i=1}^{p} F_i(x_i).$$

Then $A = \partial F$ and $B = L^{\perp}$, the orthogonal subspace to L.

Consequently, it can be easily verified that :

$$(I + \lambda A)^{-1} x = (z_1, \dots, z_p),$$

with z_i

$$= \operatorname{Argmin} \left\{ f_i(\xi_i) + \frac{1}{2\lambda} \| \xi_i - x_i \|^2 \mid \xi_i \in S_i \right\}$$

$$(I + \lambda B)^{-1} x = \operatorname{Proj}_L x = 1/p \sum x_i .$$

Indeed, each step of the algorithm splits in p convex subproblems, each one solved on an isolate set S_i :

Initialize $\lambda > 0$, $x^0 \in \mathbb{R}^n$;

a) Solve the independent subproblems for each i :

$$\begin{aligned} &\text{Minimize} \quad f_i(x_i) + \frac{1}{2\lambda} \|x_i - x^t\|^2 \\ &x_i \in S_i . \end{aligned} \tag{17}$$

Let z_i^t be the optimal solution of (17).

b)

$$x^{t+1} = 1/p \sum_{i=1}^{p} z_i^t . \tag{18}$$

This algorithm has been proposed in Pierra [15]. As (17) has a unique optimal solution, from Theorem 1, we conclude that the sequence $\{x^t\}$ converges to a limit point x_λ and, when $\lambda \downarrow 0$, from Theorem 2, x_λ tends to a solution of (15). In practice, we need a test to decide whether we must reduce λ and solve another cycle of iterations (17)-(18) or not.

Other situations where decomposition is induced on a certain product space are reviewed in Mahey *et al.* [10]. In particular, when f is the dual function of a convex constrained separable problem, the proximal algorithm looks like a separable version of the Augmented Lagrangian method. In this sense, it can be compared to similar approaches given in Bertsekas *et al.* [2] and in Spingarn [19]. Numerical experiments on these methods for the decomposition of large-scale convex programs will be published elsewhere.

7. CONCLUSION

Regularization techniques for inclusion problems depending on the sum of two maximal operators lead to an iterative process composed of two separate proximal steps. To avoid ergodic convergence, we need to iterate with a fixed sufficiently small parameter.

If we come back to the doubly subscripted sequence defined by (1), we know from elementary analysis that there exists a subsequence $\{x_k^{l(k)}\}$ which converges to a solution x^* of (P). This means that iteration (2) can converge without the ergodic artifice. We conjecture that the sequence $\{\lambda_k\}$ must decrease to zero very slowly (no counterexample with $\sum \lambda_k^2 \to \infty$ has been found to our knowledge). Some of these numerical aspects will be presented in a forthcoming paper.

The splitting of the operators induces some classical iterative schemes for the decomposition of convex programs which, though theoretically slow, are numerically stable and quite simple to implement in comparison with some direct approach like Spingarn's Partial Inverse method [19].

ACKNOWLEDGMENTS

The authors are indebted to P. L. Lions and an anonymous referee for arising some questions which have motivated the analysis in section 5.

REFERENCES

[1] H. ATTOUCH, *Variational convergence for functions and operators,* Pitman, 1984.

[2] D. P. BERTSEKAS and J. N. TSITSIKLIS, *Parallel and Distributed Computation,* Prentice-Hall Int., 1989.

[3] M. A. BOUGHAZI, *Contribution à l'étude des algorithmes d'optimisation en Analyse des Données,* Thèse de Doctorat, Université de Grenoble, 1987.

[4] H. BREZIS, Opérateurs maximaux monotones, *Mathematics Studies 5,* North Holland, 1973.

[5] H. BREZIS and P. L. LIONS, Produits infinis de résolvantes, *Israel J. of Math. 29,* 4, 1978, pp. 329-345.

[6] F. BROWDER and W. PETRYSHIN, Construction of fixed points of nonlinear mappings in Hilbert spaces, *J. of Math. Anal. and Appl. 20,* 1967, pp. 197-228.

[7] J. LEMORDANT and T. PHAM DINH, *Algorithme proximal pour la résolution numérique d'équations d'évolution en vison de bas niveau,* preprint.

[8] P. L. LIONS, Une méthode itérative de résolution d'une inéquation variationnelle, *Israel J. of Math. 31,* 2, 1978, pp. 204-208.

[9] P. L. LIONS and B. MERCIER, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal. 16,* 6, 1979, pp. 964-979.

[10] P. MAHEY, H. VAN NGUYEN and D. T. PHAM, *Proximal methods and decomposition of large convex programs,* ARTEMIS RR877, 1992.

[11] B. MARTINET, *Algorithmes pour la résolution de problèmes d'optimisation et de minimax,* Thèse d'État, Univ. de Grenoble, 1972.

[12] C. MICHELOT, *Problème de localisation : propriétés géométriques et résolution par des méthodes d'optimisation,* Thèse de Doctorat, Université de Bourgogne, 1988.

[13] J. J. MOREAU, Proximité et dualité dans un espace Hilbertien, *Bull. Soc. Math. France 93,* 1965, pp. 273-299.

[14] G. B. PASSTY, Ergodic convergence to a zero of the sum of monotone operators in Hilbert spaces, *J. of Math. Anal. and Appl. 72,* 1979, pp. 383-390.

[15] G. PIERRA G., Decomposition through formalization in a product space, *Math. Prog. 28,* 1984, pp. 96-115.

[16] R. T. ROCKAFELLAR, *Convex analysis,* Princeton University Press, 1970.

[17] R. T. ROCKAFELLAR, Monotone operators and the proximal point algorithm in convex programming, *SIAM J. on Control and Optim. 14,* 1976, pp. 877-898.

[18] R. T. ROCKAFELLAR, On the maximality of sums of nonlinear monotone operators, *Trans. Amer. Math. Soc. 149,* 1970, pp. 75-88.

[19] J. E. SPINGARN, Partial inverse of a monotone operator, *Appl. Math. Optim. 10,* 1983, pp. 247-265.

[20] G. STAMPACCHIA, Variational inequalities, in *Theory and Applications of Monotone Operators,* A. Ghizetti, 1969.

[21] A. YASSINE, *Études adaptatives et comparatives de certains algorithmes en optimisation. Implémentations effectives et applications,* Thèse de Doctorat, Université de Grenoble, 1989.