LAGRANGIAN-PENALIZATION ALGORITHM FOR CONSTRAINED OPTIMIZATION AND VARIATIONAL INEQUALITIES

P. FRANKEL AND J. PEYPOUQUET

Abstract. Let X, Y be real Hilbert spaces. Consider a bounded linear operator $A : X \to Y$ and a nonempty closed convex set $\mathcal{C} \subset Y$. In this paper we propose an inexact proximal-type algorithm to solve constrained optimization problems

$$(\mathcal{P}) \qquad \inf\{f(x) \ : \ Ax \in \mathcal{C}\},$$

where f is a proper lower-semicontinuous convex function on X; and variational inequalities

$$(\mathcal{VI}) \qquad 0 \in \mathcal{M}x + A^* N_{\mathcal{C}}(Ax),$$

where $\mathcal{M} : X \rightrightarrows X$ is a maximal monotone operator and $N_{\mathcal{C}}$ denotes the normal cone to the set $\mathcal{C}$. Our method combines a penalization procedure involving a bounded sequence of parameters, with the predictor corrector proximal multiplier method of [12]. Under suitable assumptions the sequences generated by our algorithm are proved to converge weakly to solutions of $(\mathcal{P})$ and $(\mathcal{VI})$. As applications, we describe how the algorithm can be used to find sparse solutions of linear inequality systems and solve partial differential equations by domain decomposition.

Introduction

Let X, Y be real Hilbert spaces. Given a proper lower-semicontinuous function $f : X \to \mathbf{R} \cup \{+\infty\}$, a nonempty closed convex subset $\mathcal{C}$ of Y and a bounded linear operator $A : X \to Y$, consider the following problem

$$(\mathcal{P}) \qquad \min\{f(x) \ : \ Ax \in \mathcal{C}\}.$$

Here f is the *objective function* and $\mathcal{C}$ is a *set of constraints* for the *observations* of x given by Ax. Denote by S the solution set of $(\mathcal{P})$. Let us mention two simple instances of this problem:

1. Inequality constraints in mathematical programming. Let $A = (A_m^n)$ be a $M \times N$ matrix and let $b \in \mathbf{R}^M$. For the problem of minimizing $f : \mathbf{R}^N \to \mathbf{R}$ subject to $Ax \leq b$ the set $\mathcal{C}$ is given by $\mathcal{C} = \{y \in \mathbf{R}^M : y_m \leq b_m, \ m = 1, \dots, M\}$. More generally, one can require the observations Ax of the vector x to take values under given thresholds $c_1, \dots, c_J$ for valuation functions $g_1, \dots, g_J$. In that case, $\mathcal{C} = \{y \in \mathbf{R}^M : g_j(y) \leq c_j, \ j = 1, \dots, J\}$. □

2. Domain decomposition for partial differential equations. Let us consider a bounded domain $\Omega \subset \mathbf{R}^N$ which is decomposed in two non-overlapping subdomains Ω_1 and Ω_2 with a common interface Γ. Consider the problem of finding a function on Ω satisfying some elliptic differential equations on Ω_1 and Ω_2 and such that the jump when passing from Ω_1 to Ω_2 is nonnegative. For the Poisson equation with right-hand side h and Neumann boundary conditions, the variational formulation is

$$\inf\left\{\frac{1}{2}\int_{\Omega_1}|\nabla u|^2 - \int_{\Omega_1} hu + \frac{1}{2}\int_{\Omega_2}|\nabla v|^2 - \int_{\Omega_2} hv; \ (u,v) \in H^1(\Omega_1) \times H^1(\Omega_2) \text{ and } u_{|\Gamma} \geq v_{|\Gamma}\right\}.$$

1991 *Mathematics Subject Classification.* 65K05, 65K10, 46N10, 49J40, 49M27, 90C25.
Key words and phrases. Convex optimization, proximal methods, Lagrangian, domain decomposition for PDE's.
With the support of the French ANR grant ANR-08-BLAN-0294-03, Chilean FONDECYT grant 11090023 and Basal Proyect, CMM, Universidad de Chile.

Here $X = H^1(\Omega_1) \times H^1(\Omega_2)$, $Y = L^2(\Gamma)$, $A(u,v) = u_{|\Gamma} - v_{|\Gamma}$, $\mathcal{C} = \{y \in Y : y \geq 0\}$ and $f(u,v) = \frac{1}{2}\int_{\Omega_1} |\nabla u|^2 - \int_{\Omega_1} hu + \frac{1}{2}\int_{\Omega_2} |\nabla v|^2 - \int_{\Omega_2} hv$. □

This paper is concerned with a new algorithm of proximal type that provides a solution for problem $(\mathcal{P})$. It can also be applied to solve constrained variational inequalities of the form

$$(\mathcal{VI}) \qquad 0 \in \mathcal{M}x + A^* N_{\mathcal{C}}(Ax),$$

where $\mathcal{M} : X \rightrightarrows X$ is a maximal monotone operator and $N_{\mathcal{C}}$ denotes the normal cone to the set $\mathcal{C}$.

Notice that x is a solution of problem $(\mathcal{P})$ if and only if $0 \in \partial(f + \delta_{\mathcal{C}} \circ A)(x)$, where $\delta_{\mathcal{C}}$ is the indicator function of the set $\mathcal{C}$. Recalling that $\partial\delta_{\mathcal{C}} = N_{\mathcal{C}}$, we observe that if $\mathcal{M} = \partial f$ then any solution of $(\mathcal{VI})$ is a solution of $(\mathcal{P})$. Equivalence holds under qualification conditions. It occurs, for instance, if $\mathcal{C} - A(\mathrm{dom} f)$ is a neighborhood of the origin (see [9, Theorem 2.168]).

Our method has been inspired by two classical approaches:

1. Penalization. Let us introduce a penalization function $P : Y \to [0, +\infty)$ such that $P(y) = 0$ if, and only if, $y \in \mathcal{C}$. Following [7], [14] or [4], one way to approximate points in S is to apply either a diagonal or an alternating proximal point algorithm to the family (f_k) of functions given by

$$f_k(x) = f(x) + \beta_k P(Ax), \tag{1}$$

while letting $\beta_k \to +\infty$. The idea behind is that, since the proximal point algorithm tends to minimize the function f_k, once β_k is large, the cost given by $\beta_k P(Ax)$ will force Ax to be close to $\mathcal{C}$ in some sense. This approach is especially useful when the set $\mathcal{C}$ is expressed as a sublevel set of a convex function or as intersections of such sets. Several theoretical or practical choices for the function P are available. For instance, one can take $P(\cdot) = d(\cdot, \mathcal{C})$, the distance function to $\mathcal{C}$. For the case of linear inequality constraints one can use $P(y) = \sum_{m=1}^{M} \left[y_m - b_m\right]_+$, where $[r]_+$ denotes the positive part of $r \in \mathbf{R}$.

The penalization procedure described above using (1) provides a solution of $(\mathcal{P})$. However, it often involves parameters that tend either to 0 or $+\infty$, which might lead to numerical instabilities or ill-conditioning. □

2. Lagrangian duality. Let $\sigma_{\mathcal{C}}$ denote the support function of the set $\mathcal{C}$ and define the Lagrangian function $\mathcal{L}(x,\mu) = f(x) + \langle \mu, Ax \rangle - \sigma_{\mathcal{C}}(\mu)$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in Y. Observe that problem $(\mathcal{P})$ is

$$(\mathcal{P}) \qquad \inf_{x \in X} \sup_{\mu \in Y} \mathcal{L}(x,\mu) = \inf_{x \in X} \{f(x) : Ax \in \mathcal{C}\}$$

(see [8, Chapter V]). If (x^*, μ^*) is a saddle point of $\mathcal{L}$ then $Ax^* \in \mathcal{C}$ and x^* is a solution of $(\mathcal{P})$[1]. The operator $T : X \times Y \rightrightarrows X \times Y$ defined by $T(x,\mu) = (\partial f(x) + A^*\mu, \partial\sigma_{\mathcal{C}}(\mu) - Ax)$ is maximal monotone and its zeroes coincide with the saddle points of $\mathcal{L}$ (see [20]). Therefore, one can obtain solutions of $(\mathcal{P})$ by applying the proximal point algorithm to T (see [10], [21] or [19]). One drawback is the implementation complexity due to the presence of the support function $\sigma_{\mathcal{C}}$. □

[1] Also μ^* is a solution of the dual problem

$$(\mathcal{P}^*) \qquad \sup_{\mu \in Y} \inf_{x \in X} \mathcal{L}(x,\mu) = \sup_{\mu \in Y} \{f^*(-A^*\mu) - \sigma_{\mathcal{C}}(\mu)\}.$$

In order to solve problems $(\mathcal{P})$ and $(\mathcal{VI})$ we propose a Lagrangian-based approach that incorporates a sort of penalization function for the set $\mathcal{C}$. It is worth mentioning that neither divergent penalization parameters nor vanishing step sizes come into play. The method uses the prediction-correction ideas introduced in [12] for minimization problems, but keeping a multiplier for the constraint involving P. This multiplier can also be interpreted as a vector of penalization parameters with an updating rule that prevents them from growing indefinitely. The prediction-correction steps also allow to circumvent the problem of computing resolvents of sums. All the analysis is carried out in a Hilbert space setting.

This paper is organized as follows: In Section 1 we discuss on the problems $(\mathcal{P})$ and $(\mathcal{VI})$, alternative formulations and their sets of solutions. We present our Lagrangian-based algorithm with explicitly evaluated prediction/correction steps for the Lagrange multipliers and describe our main results. The convergence analysis in the context of problem $(\mathcal{VI})$ is presented in Section 2. Section 3 contains additional results for problem $(\mathcal{P})$. The remainder is devoted to applications. In Section 4 we explain how the algorithm can be used to obtain sparse solutions for a system of linear inequalities. Section 5 contains a domain decomposition method for partial differential equations with a unilateral transfer through the boundary.

1. Preliminaries

Since no confusion should arise, all inner products (in X, Y and $\mathbf{R}^M$) will be denoted by $\langle\cdot,\cdot\rangle$ and the corresponding norms by $|\cdot|$.

Let $P = (p_m)_{m=1}^M$ be a l-Lipschitz vector-valued function on Y such that each component p_m is nonnegative and convex. Assume that the set $\mathcal{C}$ is defined by

$$\mathcal{C} = \{\ y \in Y : P(y) = 0\ \}.$$

Set $\mathbf{H} = X \times Y \times Y \times \mathbf{R}^M$. In order to simplify the notation, let us write $\partial P = (\partial p_m)_{m=1}^M$. Following [20, 6], given a maximal monotone operator $\mathcal{M} : X \rightrightarrows X$ we define the monotone[2] operator $\mathcal{N}_{\mathcal{M}} : \mathbf{H} \rightrightarrows \mathbf{H}$ by

$$\mathcal{N}_{\mathcal{M}}(x, y, \mu, \nu) = (\mathcal{M}x + A^*\mu, -\mu + \langle \nu, \partial P(y)\rangle, -Ax + y, -P(y)).$$

Since each component p_m is continuous, for each fixed $\nu \in \mathbf{R}^M$ we have $\partial(\langle \nu, P(\cdot)\rangle)(y) = \langle \nu, \partial P(y)\rangle$ for all $y \in Y$. Therefore, the operator $\langle \nu, \partial P\rangle : Y \rightrightarrows Y$ is maximal monotone. Write $\mathbf{S}_{\mathcal{M}} = \mathcal{N}_{\mathcal{M}}^{-1}0$ and observe that a point $(x^*, y^*, \mu^*, \nu^*) \in \mathbf{H}$ belongs to $\mathbf{S}_{\mathcal{M}}$ if, and only if,

$$-A^*\mu^* \in \mathcal{M}x^*, \quad \mu^* \in \langle \nu^*, \partial P(y^*)\rangle, \quad Ax^* = y^*, \quad \text{and} \quad P(y^*) = 0.$$

If $(x^*, y^*, \mu^*, \nu^*) \in \mathbf{S}_{\mathcal{M}}$ then x^* satisfies $(\mathcal{VI})$ because $\langle \nu^*, \partial P(y^*)\rangle \subset N_{\mathcal{C}}(y^*)$. The converse depends on the function P. For example, if $P(\cdot) = d(\cdot, \mathcal{C})$ and x^* satisfies $(\mathcal{VI})$, then there exist y^*, μ^* and ν^* such that $(x^*, y^*, \mu^*, \nu^*) \in \mathbf{S}_{\mathcal{M}}$.

On the other hand, by introducing an auxiliary variable $y \in Y$ we can rewrite $(\mathcal{P})$ as

$$\inf\{f(x) : Ax = y \text{ and } P(y) = 0\} = \inf\{f(x) : (x, y) \in \mathbf{C}\},$$

where

$$\mathbf{C} = \{\ (x, y) \in X \times Y\ :\ Ax = y,\ y \in \mathcal{C}\ \}$$

is the set of *primal feasible points*.

Define the Lagrangian function $L : \mathbf{H} \to \mathbf{R} \cup \{+\infty\}$ by

$$L(x, y, \mu, \nu) = f(x) + \langle \mu, Ax - y\rangle + \langle \nu, P(y)\rangle. \tag{2}$$

[2] Maximality is irrelevant for our convergence analysis.

A point $w^*=(x^*,y^*,\mu^*,\nu^*)\in\mathbf{H}$ is a *saddle point* of L if

$$L(x^*,y^*,\mu,\nu)\le L(x^*,y^*,\mu^*,\nu^*)\le L(x,y,\mu^*,\nu^*) \tag{3}$$

for all $(x,y,\mu,\nu)\in\mathbf{H}$. The set of saddle points of L coincides with $\mathbf{S}_{\partial f}$ (see [20]). Observe that if (x^*,y^*,μ^*,ν^*) is a saddle point of the Lagrangian then $(x^*,y^*)\in\mathbf{C}$ and x^* is a solution of $(\mathcal{P})$.

In order to find points in $\mathbf{S}_{\mathcal{M}}$, we propose the following method. Let us take $w^0\in\mathbf{H}$ and define the sequence (w^k) inductively as follows: given $w^{k-1}=(x^{k-1},y^{k-1},\mu^{k-1},\nu^{k-1})$ we introduce a prediction $(\tilde\mu^k,\tilde\nu^k)$ for the multipliers using the proximal point algorithm. This idea is motivated by [12]. By linearity, this accounts to

$$(\mathbf{A1})\qquad \left\{\begin{array}{rcl}\tilde\mu^k&=&\mu^{k-1}+\lambda_k(Ax^{k-1}-y^{k-1})\\ \tilde\nu^k&=&\nu^{k-1}+\lambda_kP(y^{k-1}).\end{array}\right.$$

Proximal steps with respect to the state variables (x,y) read

$$-\frac{x^k-x^{k-1}}{\lambda_k}-A^*\tilde\mu^k\in\mathcal{M}x^k\qquad\text{and}\qquad -\frac{y^k-y^{k-1}}{\lambda_k}+\tilde\mu^k\in\sum_{m=1}^M\tilde\nu^k_m\partial p_m(y^k), \tag{4}$$

respectively. If $\mathcal{M}=\partial f$ these correspond to

$$\left\{\begin{array}{rcl}x^k&=&\mathrm{Argmin}_{x\in X}\left\{L(x,y^{k-1},\tilde\mu^k,\tilde\nu^k)+\frac{1}{2\lambda_k}|x-x^{k-1}|^2\right\}\\ y^k&=&\mathrm{Argmin}_{y\in Y}\left\{L(x^{k-1},y,\tilde\mu^k,\tilde\nu^k)+\frac{1}{2\lambda_k}|y-y^{k-1}|^2\right\}.\end{array}\right.$$

Due to the maximal monotonicity of $\mathcal{M}$ and $\langle\nu,\partial P\rangle$, each of the inclusions given by (4) has a unique solution by virtue of Minty's Theorem. However, since they might be difficult to solve it is important to use approximate or relaxed versions. For $\varepsilon\ge 0$ set

$$\mathcal{M}_\varepsilon x=\{x^*\in X\ :\ \langle x^*-u^*,x-u\rangle\ge-\varepsilon\quad\text{for all } u^*\in\mathcal{M}u\ \}.$$

We always have $\mathcal{M}\subset\mathcal{M}_\varepsilon$. Moreover, if $\mathcal{M}=\partial f$ then $\partial f\subset\partial_\varepsilon f\subset(\partial f)_\varepsilon$, where ∂_ε denotes the standard ε-approximate subdifferential. We consider the inclusions

$$(\mathbf{A2})\qquad -\frac{x^k-x^{k-1}}{\lambda_k}-A^*\tilde\mu^k\in\mathcal{M}_{\varepsilon_k}x^k\qquad\text{and}\qquad -\frac{y^k-y^{k-1}}{\lambda_k}+\tilde\mu^k\in\sum_{m=1}^M\tilde\nu^k_m\partial_{\varepsilon_k}p_m(y^k),$$

for $\varepsilon_k\ge 0$. Finally, the multipliers are updated using:

$$(\mathbf{A3})\qquad \left\{\begin{array}{rcl}\mu^k&=&\mu^{k-1}+\lambda_k(Ax^k-y^k)\\ \nu^k&=&\nu^{k-1}+\lambda_kP(y^k).\end{array}\right.$$

In the following sections we shall prove the weak convergence of the sequence (w^k) generated by $(\mathbf{A1})-(\mathbf{A3})$ to a point in $\mathbf{S}_{\mathcal{M}}$ under a summability assumption on the error sequence (ε_k) and a boundedness assumption on the step sizes (λ_k). For a general maximal monotone operator $\mathcal{M}$ we require Y to be finite-dimensional, an assumption that is already present in [12]. When $\mathcal{M}$ is the subdifferential of some proper lower-semicontinuous function $f:X\to\mathbf{R}\cup\{+\infty\}$, this hypothesis on the dimension of Y can be eliminated. Moreover, we also establish the existence of $\lim_{k\to+\infty}L(x^k,y^k,\mu^k,\nu^k)$ and $\lim_{k\to+\infty}f(x^k)$, which provide a key tool for upgrading convergence from weak to strong in the application described in Section 5.

2. Convergence toward $\mathbf{S}_{\mathcal{M}}$

The purpose of this section is to prove the following:

Theorem 1. *Let X be a real Hilbert space and $Y = \mathbf{R}^p$. Let $\mathbf{S}_{\mathcal{M}} \neq \emptyset$ and assume $(\varepsilon_k) \in \ell^1$ and $0 < \underline{\lambda} \leq \lambda_k \leq \overline{\lambda} < \max\{\frac{1}{\sqrt{2}\|A\|}, \frac{1}{\sqrt{2+l^2}}\}$. Any sequence (x^k, y^k, μ^k, ν^k) generated by Algorithm $(\mathbf{A1}) - (\mathbf{A3})$ converges weakly as $k \to +\infty$ to some $(x^\infty, y^\infty, \mu^\infty, \nu^\infty) \in \mathbf{S}_{\mathcal{M}}$.*

We start by deriving the fundamental estimations that will support the convergence analysis. For $w \in \mathbf{H}$, let us write

$$|\!|\!| w |\!|\!|^2 = |x|^2 + |y|^2 + |\mu|^2 + |\nu|^2.$$

Lemma 2. *Let $(x^*, y^*, \mu^*, \nu^*) \in \mathbf{S}_{\mathcal{M}}$. Then for all $k \in \mathbf{N}$ we have*

$$\begin{aligned} &|\!|\!| w^k - w^* |\!|\!|^2 - |\!|\!| w^{k-1} - w^* |\!|\!|^2 + |\tilde{\mu}^k - \mu^{k-1}|^2 + |\tilde{\nu}^k - \nu^{k-1}|^2 \\ &+ \left(1 - 2\lambda_k^2 \|A\|^2\right) |x^k - x^{k-1}|^2 + \left(1 - \lambda_k^2(2 + l^2)\right) |y^k - y^{k-1}|^2 \quad \leq \quad 2\lambda_k(M+1)\varepsilon_k. \end{aligned} \tag{5}$$

Proof. Let $(x^*, y^*, \mu^*, \nu^*) \in \mathbf{S}_{\mathcal{M}}$. From the definition of $\mathcal{M}_\varepsilon$ and $(\mathbf{A2})$ we have

$$\left\langle A^*\mu^* - \frac{x^k - x^{k-1}}{\lambda_k} - A^*\tilde{\mu}^k, x^* - x^k \right\rangle \leq \varepsilon_k,$$

and we infer that

$$|x^k - x^*|^2 - |x^{k-1} - x^*|^2 + |x^k - x^{k-1}|^2 + 2\lambda_k \langle \tilde{\mu}^k - \mu^*, A(x^k - x^*) \rangle \leq 2\lambda_k \varepsilon_k. \tag{6}$$

On the other hand, the ε_k-approximate subdifferential inequality for each $\tilde{\nu}^k p_m$ gives

$$2\lambda_k \langle \tilde{\nu}^k, P(y^*) - P(y^k) \rangle \quad \geq \quad -2\lambda_k \left\langle \frac{y^k - y^{k-1}}{\lambda_k} - \tilde{\mu}^k, y^* - y^k \right\rangle - 2\lambda_k M \varepsilon_k$$

by summation. Hence

$$|y^k - y^*|^2 - |y^{k-1} - y^*|^2 + |y^k - y^{k-1}|^2 + 2\lambda_k \langle \tilde{\nu}^k, P(y^k) - P(y^*) \rangle + 2\lambda_k \langle \tilde{\mu}^k, y^* - y^k \rangle \leq 2\lambda_k M \varepsilon_k. \tag{7}$$

Moreover we have $\mu^* \in \langle \nu^*, \partial P(y^*) \rangle$, and so

$$2\lambda_k \langle -\mu^*, y^* - y^k \rangle - 2\lambda_k \langle \nu^*, P(y^k) - P(y^*) \rangle \leq 0. \tag{8}$$

Summing up inequalities (6), (7) and (8), and using that $Ax^* = y^*$, one obtains

$$\begin{aligned} &|x^k - x|^2 - |x^{k-1} - x|^2 + |x^k - x^{k-1}|^2 \\ &+ |y^k - y|^2 - |y^{k-1} - y|^2 + |y^k - y^{k-1}|^2 \\ &+ 2\lambda_k \left[\langle \tilde{\mu}^k - \mu^*, Ax^k - y^k \rangle + \langle \tilde{\nu}^k - \nu^*, P(y^k) \rangle \right] \quad \leq 2\lambda_k (M+1)\varepsilon_k. \end{aligned} \tag{9}$$

We rewrite the term in the bracket as follows

$$\begin{aligned} &\langle \tilde{\mu}^k - \mu^*, Ax^k - y^k \rangle + \langle \tilde{\nu}^k - \nu^*, P(y^k) \rangle \\ &= \langle \tilde{\mu}^k - \mu^k, Ax^k - y^k \rangle + \langle \tilde{\nu}^k - \nu^k, P(y^k) \rangle + \langle \mu^k - \mu^*, Ax^k - y^k \rangle + \langle \nu^k - \nu^*, P(y^k) \rangle \\ &= \frac{1}{\lambda_k} \langle \tilde{\mu}^k - \mu^k, \mu^k - \mu^{k-1} \rangle + \frac{1}{\lambda_k} \langle \tilde{\nu}^k - \nu^k, \nu^k - \nu^{k-1} \rangle + \frac{1}{\lambda_k} \langle \mu^k - \mu^*, \mu^k - \mu^{k-1} \rangle + \frac{1}{\lambda_k} \langle \nu^k - \nu^*, \nu^k - \nu^{k-1} \rangle \\ &= \frac{1}{2\lambda_k} [|\tilde{\mu}^k - \mu^{k-1}|^2 - |\tilde{\mu}^k - \mu^k|^2 - |\mu^k - \mu^{k-1}|^2] + \frac{1}{2\lambda_k} [|\tilde{\nu}^k - \nu^{k-1}|^2 - |\tilde{\nu}^k - \nu^k|^2 - |\nu^k - \nu^{k-1}|^2] \\ &\quad + \frac{1}{2\lambda_k} [|\mu^k - \mu^*|^2 + |\mu^k - \mu^{k-1}|^2 - |\mu^{k-1} - \mu^*|^2] + \frac{1}{2\lambda_k} [|\nu^k - \nu^*|^2 + |\nu^k - \nu^{k-1}|^2 - |\nu^{k-1} - \nu^*|^2]. \end{aligned} \tag{10}$$

To simplify the notation, define

$$\rho_k = |x^k - x^{k-1}|^2 + |y^k - y^{k-1}|^2 + |\tilde{\mu}^k - \mu^{k-1}|^2 + |\tilde{\nu}^k - \nu^{k-1}|^2.$$

Recall that $\|w\|^2 = |x|^2 + |y|^2 + |\mu|^2 + |\nu|^2$ for $w \in \mathbf{H}$. Replacing equality (10) in (9), we deduce that

$$\|w^k - w^*\|^2 - \|w^{k-1} - w^*\|^2 + \rho_k - |\tilde{\mu}^k - \mu^k|^2 - |\tilde{\nu}^k - \nu^k|^2 \leq 2\lambda_k(M+1)\varepsilon_k.$$

To conclude, observe that

$$|\tilde{\mu}^k - \mu^k|^2 = \lambda_k^2|A(x^{k-1} - x^k) - (y^{k-1} - y^k)|^2 \leq 2\lambda_k^2\|A\|^2|x^k - x^{k-1}|^2 + 2\lambda_k^2|y^k - y^{k-1}|^2,$$

while

$$|\tilde{\nu}^k - \nu^k|^2 = \lambda_k^2|P(y^{k-1}) - P(y^k)|^2 \leq \lambda_k^2 l^2|y^k - y^{k-1}|^2.$$

Adding the last three inequalities we obtain (5). ■

In order to prove the convergence of the algorithm first recall the following elementary result for real sequences. A proof can be found, for instance, in [5, Lemma 2].

Lemma 3. *Let (a_k), (b_k) and (η_k) be real sequences. Assume that (a_k) is bounded from below, (b_k) is nonnegative and $(\eta_k) \in l^1$. Assume also that $a_{k+1} - a_k + b_k \leq \eta_k$ for every $k \in \mathbf{N}$. Then (a_k) converges and $(b_k) \in l^1$.*

An immediate consequence of Lemmas 2 and 3 is the following:

Proposition 4. *Let $\mathbf{S}_{\mathcal{M}} \neq \emptyset$ and assume $(\varepsilon_k) \in \ell^1$ and $0 < \underline{\lambda} \leq \lambda_k \leq \overline{\lambda} < \max\{\frac{1}{\sqrt{2}\|A\|}, \frac{1}{\sqrt{2+l^2}}\}$. We have the following:*

(i) *the sequences $(|x^k - x^{k-1}|^2)$, $(|y^k - y^{k-1}|^2)$, $(|Ax^k - y^k|^2)$, $(|P(y^k)|^2)$ are summable;*
(ii) *for every $(x^*, y^*, \nu^*, \mu^*) \in \mathbf{S}_{\mathcal{M}}$, $\lim_{k\to+\infty} \|(x^k, y^k, \mu^k, \nu^k) - (x^*, y^*, \mu^*, \nu^*)\|$ exists in $\mathbf{R}$.*

In order to prove the main result of this section we shall use Opial's Lemma [18], which we recall for the sake of completeness:

Lemma 5 (Opial)**.** *Let H be a Hilbert space endowed with the norm $\|\cdot\|$. Let (ξ_n) be a sequence of H such that there exists a nonempty set $\Xi \subset H$ which verifies*

(a) *for all $\xi \in \Xi$, $\lim\limits_{n\to+\infty} \|\xi_n - \xi\|$ exists,*
(b) *if $(\xi_{n_k}) \rightharpoonup \overline{\xi}$ weakly in H as $k \to +\infty$, we have $\overline{\xi} \in \Xi$.*

Then the sequence (ξ_n) converges weakly in H as $n \to +\infty$ to a point in Ξ.

We are now in position to prove the main result of this section.

Proof of Theorem 1. Let (x^k, y^k, μ^k, ν^k) be a sequence generated by Algorithm $(\mathbf{A1}) - (\mathbf{A3})$. In view of item (ii) of Proposition 4, the quantity $\|w^k - w\|$ has a limit as $n \to +\infty$ for every $w \in \mathbf{S}_{\mathcal{M}}$. This shows point (a) in Opial's Lemma. To prove point (b), suppose a subsequence of (x^k, y^k, μ^k, ν^k), still denoted (x^k, y^k, μ^k, ν^k), that converges weakly to $(x^\infty, y^\infty, \mu^\infty, \nu^\infty)$, *i.e.* (x^k) weakly converges toward x^∞ in X and (y^k, μ^k, ν^k) strongly converges toward $(y^\infty, \mu^\infty, \nu^\infty)$ in $Y \times Y \times \mathbf{R}^M$. We must show that $(x^\infty, y^\infty, \mu^\infty, \nu^\infty) \in \mathbf{S}_{\mathcal{M}}$. Using the closedness of the function $(x, y) \in X \times Y \to |Ax - y|^2 \in \mathbf{R}_+$ and the continuity of the function P and item (i) of Proposition 4, we have

$$\begin{aligned} |Ax^\infty - y^\infty|^2 &\leq \liminf_{k\to+\infty} |Ax^k - y^k|^2 = 0, \\ P(y^\infty) &= \lim_{k\to+\infty} P(y^k) = 0, \end{aligned}$$

hence $Ax^\infty - y^\infty = 0$ and $P(y^\infty) = 0$, which implies $(x^\infty, y^\infty) \in \mathbf{C}$. Let (x, x^*) be in the graph of $\mathcal{M}$. In view of $(\mathbf{A2})$, we have

$$\langle -\frac{x^k - x^{k-1}}{\lambda_k} - A^*\tilde{\mu}^k - x^*, x^k - x\rangle \geq -\varepsilon_k.$$

Notice that, in view of Proposition 4(i), $\lim_{k\to+\infty} -\frac{x^k-x^{k-1}}{\lambda_k} = 0$. Moreover $\lim_{k\to+\infty} |Ax^k - y^k| = 0$, hence the sequence $(\tilde{\mu}^k)$ strongly converges in Y toward μ^∞. Using also the continuity of the operator A^*, we can pass to the limit in the above inequality to obtain

$$\langle -A^*\mu^\infty - x^*, x^\infty - x\rangle \geq 0.$$

Using the maximality of the operator $\mathcal{M}$, this implies $-A^*\mu^\infty \in \mathcal{M}x^\infty$. Let now (y, y^*) in the graph of $\langle \nu^\infty, \partial P\rangle$, we have

$$\langle \nu^\infty, P(y^k) - P(y)\rangle \geq \langle y^*, y^k - y\rangle.$$

Moreover in view of (**A2**), we have

$$\langle \tilde{\nu}^k, P(y) - P(y^k)\rangle \quad \geq \quad \left\langle -\frac{y^k - y^{k-1}}{\lambda_k} + \tilde{\mu}^k, y - y^k \right\rangle - M\varepsilon_k.$$

Adding these two last inequalities, we obtain

$$\langle \nu^\infty - \tilde{\nu}^k, P(y^k) - P(y)\rangle \quad \geq \quad \left\langle y^* + \frac{y^k - y^{k-1}}{\lambda_k} - \tilde{\mu}^k, y^k - y \right\rangle - M\varepsilon_k.$$

In view of Proposition 4(i), $\lim_{k\to+\infty} \frac{y^k - y^{k-1}}{\lambda_k} = 0$. Moreover $\lim_{k\to+\infty} P(y^k) = 0$, hence the sequence $(\tilde{\nu}^k)$ strongly converges in Y toward ν^∞. We can pass to the limit in the above inequality to obtain

$$\langle \mu^\infty - y^*, y^\infty - y\rangle \geq 0.$$

By maximality of the operator $\langle \nu^\infty, \partial P\rangle$, this implies that $\mu^\infty \in \langle \nu^\infty, \partial P(y^\infty)\rangle$. This achieves to prove that $(x^\infty, y^\infty, \mu^\infty, \nu^\infty) \in \mathbf{S}_\mathcal{M}$. ■

Remark 6. If $\mathcal{M}$ is strongly monotone with parameter $\alpha > 0$, the algorithm can be slightly modified in order to obtain strong convergence in Theorem 1. It suffices to redefine the operator $\mathcal{M}_\varepsilon$ for $\varepsilon \geq 0$ as

$$\widetilde{\mathcal{M}}_\varepsilon x = \{x^* \in X \;:\; \langle x^* - u^*, x - u\rangle \geq \alpha\|x - u\|^2 - \varepsilon \quad \text{for all } u^* \in \mathcal{M}u\;\}.$$

The strong monotonicity of $\mathcal{M}$ implies that one still has $\mathcal{M} \subset \widetilde{\mathcal{M}}_\varepsilon$. Following the argument in Lemma 2 we deduce that

$$|\!|\!| w^k - w^* |\!|\!|^2 - |\!|\!| w^{k-1} - w^* |\!|\!|^2 + 2\alpha\lambda_k\|x^k - x^*\|^2 \leq 2\lambda_k(M+1)\varepsilon_k$$

for all $k \in \mathbf{N}$, where x^* is the unique solution of $(\mathcal{VI})$ and w^* is any element in $\mathbf{S}_\mathcal{M}$. The details are left to the reader. This immediately implies that x^k converges strongly to x^* as $k \to +\infty$.

3. Further results for $\mathcal{M} = \partial f$

If $\mathcal{M} = \partial f$ a more detailed analysis can be carried out and some results can be improved. In particular, the assumption on the dimension of Y can be omitted. Moreover, part (ii) in Proposition 9 below is used in Section 5 to upgrade convergence from weak to strong in a domain decomposition method for partial differential equations. In this section, we assume that the primal steps are computed using the approximate subdifferentials. Namely,

$$(\mathbf{A2'}) \qquad -\frac{x^k - x^{k-1}}{\lambda_k} - A^*\tilde{\mu}^k \in \partial_{\varepsilon_k} f(x^k) \qquad \text{and} \qquad -\frac{y^k - y^{k-1}}{\lambda_k} + \tilde{\mu}^k \in \sum_{m=1}^{M} \tilde{\nu}_m^k \partial_{\varepsilon_k} p_m(y^k),$$

for $\varepsilon_k \geq 0$. We shall prove the following:

Theorem 7. *Let X and Y be real Hilbert spaces. Let $\mathbf{S}_{\partial f} \neq \emptyset$ and assume $(\varepsilon_k) \in \ell^1$ and $0 < \underline{\lambda} \leq \lambda_k \leq \overline{\lambda} < \max\{\frac{1}{\sqrt{2}\|A\|}, \frac{1}{\sqrt{2+l^2}}\}$. Any sequence (x^k, y^k, μ^k, ν^k) generated by Algorithm $(\mathbf{A1})-(\mathbf{A2'})-(\mathbf{A3})$ converges weakly as $k \to +\infty$ to some $(x^\infty, y^\infty, \mu^\infty, \nu^\infty) \in \mathbf{S}_{\partial f}$.*

We begin with a reinforced version of Lemma 2:

Lemma 8. *Let (x^*, y^*, μ^*, ν^*) have the saddle-point property. Then for all $k \in \mathbf{N}$ we have*

$$\begin{aligned} &\|w^k - w^*\|^2 - \|w^{k-1} - w^*\|^2 + |\tilde{\mu}^k - \mu^{k-1}|^2 + |\tilde{\nu}^k - \nu^{k-1}|^2 \\ &+ \left(1 - 2\lambda_k^2\|A\|^2\right)|x^k - x^{k-1}|^2 + \left(1 - \lambda_k^2(2 + l^2)\right)|y^k - y^{k-1}|^2 \\ &+2\lambda_k \Big[L(x^k, y^k, \mu^*, \nu^*) - L(x^*, y^*, \mu^*, \nu^*)\Big] \;\leq\; 2\lambda_k(M+1)\varepsilon_k. \end{aligned} \tag{11}$$

Proof. The subdifferential inequality for f gives

$$\begin{aligned} 2\lambda_k(f(x) - f(x^k)) &\geq -2\lambda_k \left\langle \frac{x^k - x^{k-1}}{\lambda_k} + A^*\tilde{\mu}^k, x - x^k \right\rangle - 2\lambda_k\varepsilon_k \\ &= |x^k - x|^2 - |x^{k-1} - x|^2 + |x^k - x^{k-1}|^2 + 2\lambda_k\langle\tilde{\mu}^k, A(x^k - x)\rangle - 2\lambda_k\varepsilon_k \end{aligned}$$

for all $x \in X$. On the other hand, the subdifferential inequality for each $\tilde{\nu}^k p_m$ gives

$$\begin{aligned} 2\lambda_k\langle\tilde{\nu}^k, P(y) - P(y^k)\rangle &\geq -2\lambda_k \left\langle \frac{y^k - y^{k-1}}{\lambda_k} - \tilde{\mu}^k, y - y^k \right\rangle - 2\lambda_k M\varepsilon_k \\ &= |y^k - y|^2 - |y^{k-1} - y|^2 + |y^k - y^{k-1}|^2 + 2\lambda_k\langle\tilde{\mu}^k, y - y^k\rangle - 2\lambda_k M\varepsilon_k \end{aligned}$$

for all $y \in Y$. Summing up, one obtains

$$\begin{aligned} &|x^k - x|^2 - |x^{k-1} - x|^2 + |x^k - x^{k-1}|^2 \\ &+ |y^k - y|^2 - |y^{k-1} - y|^2 + |y^k - y^{k-1}|^2 \\ &+ 2\lambda_k \Big[L(x^k, y^k, \tilde{\mu}^k, \tilde{\nu}^k) - L(x, y, \tilde{\mu}^k, \tilde{\nu}^k)\Big] \;\leq 2\lambda_k(M+1)\varepsilon_k. \end{aligned} \tag{12}$$

Let (x^*, y^*, μ^*, ν^*) have the saddle-point property and take $x = x^*$ and $y = y^*$ in (12). Since $L(x^*, y^*, \tilde{\mu}^k, \tilde{\nu}^k) \leq L(x^*, y^*, \mu^*, \nu^*)$, we obtain

$$\begin{aligned} &|x^k - x^*|^2 - |x^{k-1} - x^*|^2 + |x^k - x^{k-1}|^2 \\ &+ |y^k - y^*|^2 - |y^{k-1} - y^*|^2 + |y^k - y^{k-1}|^2 \\ &+ 2\lambda_k \Big[L(x^k, y^k, \tilde{\mu}^k, \tilde{\nu}^k) - L(x^*, y^*, \mu^*, \nu^*)\Big] \;\leq 2\lambda_k(M+1)\varepsilon_k. \end{aligned} \tag{13}$$

We can write

$$\begin{aligned} L(x^k, y^k, \tilde{\mu}^k, \tilde{\nu}^k) - L(x^*, y^*, \mu^*, \nu^*) &= L(x^k, y^k, \tilde{\mu}^k, \tilde{\nu}^k) - L(x^k, y^k, \mu^*, \nu^*) \\ &\quad +L(x^k, y^k, \mu^*, \nu^*) - L(x^*, y^*, \mu^*, \nu^*) \\ &= \langle\tilde{\mu}^k - \mu^*, Ax^k - y^k\rangle + \langle\tilde{\nu}^k - \nu^*, P(y^k)\rangle \\ &\quad +L(x^k, y^k, \mu^*, \nu^*) - L(x^*, y^*, \mu^*, \nu^*). \end{aligned}$$

Using equality (10), complete the proof of (11) as in Lemma 2. ■

The following complements Proposition 4.

Proposition 9. *Let $\mathbf{S}_{\partial f} \neq \emptyset$ and assume $(\varepsilon_k) \in \ell^1$ and $0 < \underline{\lambda} \leq \lambda_k \leq \overline{\lambda} < \max\{\frac{1}{\sqrt{2}\|A\|}, \frac{1}{\sqrt{2+l^2}}\}$. We have the following:*

(i) *for each $(x^*, y^*, \nu^*, \mu^*) \in \mathbf{S}_{\partial f}$, the sequence $(L(x^k, y^k, \mu^*, \nu^*) - L(x^*, y^*, \mu^*, \nu^*))$ is in ℓ^1;*
(ii) $\lim_{k\to+\infty} L(x^k, y^k, \mu^k, \nu^k) = L(x^*, y^*, \mu^*, \nu^*)$ *and* $\lim_{k\to+\infty} f(x^k) = f(x^*)$.

Proof. Item (i) is an immediate consequence of Lemmas 8 and 3 because each of the terms $L(x^k, y^k, \mu^*, \nu^*) - L(x^*, y^*, \mu^*, \nu^*)$ is nonnegative in view of the saddle-point property. We deduce that $\lim_{k\to+\infty}(L(x^k, y^k, \mu^*, \nu^*) - L(x^*, y^*, \mu^*, \nu^*)) = 0$. By Proposition 4 (i), $\lim_{k\to+\infty} Ax^k - y^k = 0$, $\lim_{k\to+\infty} P(y^k) = 0$ and the sequences (μ^k) and (ν^k) are bounded. This easily implies (ii). ■

We can now prove the main result of this section.

Proof of Theorem 7. Let (x^k, y^k, μ^k, ν^k) be any sequence generated by Algorithm $(\mathbf{A1})-(\mathbf{A2'})-(\mathbf{A3})$. In view of item (ii) of Proposition 4, the quantity $|||w^k - w|||$ has a limit as $n \to +\infty$ for every $w \in \mathbf{S}_{\partial f}$. This shows point (a) in Opial's Lemma. To prove point (b), suppose a subsequence of (x^k, y^k, μ^k, ν^k), still denoted (x^k, y^k, μ^k, ν^k), converges weakly to $(x^\infty, y^\infty, \mu^\infty, \nu^\infty)$. We must show that $(x^\infty, y^\infty, \mu^\infty, \nu^\infty)$ is a saddle-point for the Lagrangian function L. Using the closedness of the functions $(x, y) \in X \times Y \to |Ax - y|^2 \in \mathbf{R}_+$ and $|P|$ and item (i) of Proposition 4, we have

$$\begin{aligned} |Ax^\infty - y^\infty|^2 &\leq \liminf_{k\to+\infty} |Ax^k - y^k|^2 = 0, \\ |P(y^\infty)| &\leq \liminf_{k\to+\infty} |P(y^k)| = 0, \end{aligned}$$

hence $Ax^\infty - y^\infty = 0$ and $P(y^\infty) = 0$, which implies $(x^\infty, y^\infty) \in \mathbf{C}$. Let us fix $(x, y) \in X \times Y$. For all $N \in \mathbf{N}$ we have

$$2\sum_{k=1}^{N} \lambda_k (L(x^k, y^k, \tilde{\mu}^k, \tilde{\nu}^k) - L(x, y, \tilde{\mu}^k, \tilde{\nu}^k)) \leq |x^0 - x|^2 + |y^0 - y|^2 + 2\overline{\lambda}(M + 1)\sum_{k=1}^{\infty} \varepsilon_k$$

in view of inequality (12). Therefore, $\liminf_{k\to+\infty}(L(x^k, y^k, \tilde{\mu}^k, \tilde{\nu}^k) - L(x, y, \tilde{\mu}^k, \tilde{\nu}^k)) \leq 0$. Notice that, since $\lim_{k\to\infty} |Ax^k - y^k| = \lim_{k\to\infty} |P(y^k)| = 0$, the sequence $(\tilde{\mu}^k, \tilde{\nu}^k)$ converges weakly to $(\mu^\infty, \nu^\infty) \in Y \times \mathbf{R}$. We deduce that

$$\begin{aligned} \lim_{k\to\infty} L(x, y, \tilde{\mu}^k, \tilde{\nu}^k) &= \lim_{k\to\infty} \Big(f(x) + \langle \tilde{\mu}^k, Ax - y\rangle + \langle \tilde{\nu}^k, P(y)\rangle \Big) \\ &= f(x) + \langle \mu^\infty, Ax - y\rangle + \langle \nu^\infty, P(y)\rangle \\ &= L(x, y, \mu^\infty, \nu^\infty). \end{aligned}$$

Moreover

$$L(x^k, y^k, \tilde{\mu}^k, \tilde{\nu}^k) = f(x^k) + \langle \tilde{\mu}^k, Ax^k - y^k\rangle + \langle \tilde{\nu}^k, P(y^k)\rangle \tag{14}$$

and the last two terms tend to 0 as $k \to +\infty$. Whence

$$\liminf_{k\to+\infty} f(x^k) = \liminf_{k\to+\infty} L(x^k, y^k, \tilde{\mu}^k, \tilde{\nu}^k) \leq \liminf_{k\to+\infty} L(x, y, \tilde{\mu}^k, \tilde{\nu}^k) = \lim_{k\to+\infty} L(x, y, \tilde{\mu}^k, \tilde{\nu}^k) = L(x, y, \mu^\infty, \nu^\infty).$$

Finally, using the fact that every limit point of (x^k, y^k) is feasible along with closedness of the function f, we infer that

$$L(x^\infty, y^\infty, \mu^\infty, \nu^\infty) = f(x^\infty) \leq \liminf_{k\to+\infty} f(x^k) \leq L(x, y, \mu^\infty, \nu^\infty).$$

We now must prove that, for every $(\mu, \nu) \in Y \times \mathbf{R}$, we have

$$L(x^\infty, y^\infty, \mu, \nu) \leq L(x^\infty, y^\infty, \mu^\infty, \nu^\infty).$$

This is clear since $Ax^\infty - y^\infty = 0$ and $P(y^\infty) = 0$. We have proved that every weak cluster point of the sequence (x^k, y^k, μ^k, ν^k) is a saddle-point for the Lagrangian function L and the result follows from Opial's Lemma. ■

Remark 10. Our penalization scheme is exact in the following sense: Let $(x^*, y^*, \mu^*, \nu^*) \in \mathbf{S}_{\partial f}$ and let

$$\widehat{x} \in \text{Argmin}\{f(x) + \langle \nu, P(Ax) \rangle\},$$

with $\nu_m > \nu_m^*$ for $m = 1, \ldots, M$. Then $\widehat{x}$ is a solution of $(\mathcal{P})$. Indeed, from the definition of $\widehat{x}$ and the saddle-point property (3), we have

$$f(\widehat{x}) + \langle \nu, P(A\widehat{x}) \rangle \leq f(x^*) \leq f(\widehat{x}) + \langle \nu^*, P(A\widehat{x}) \rangle.$$

Since $\nu_m > \nu_m^*$ for each $m = 1, \ldots, M$ one must have $P(A\widehat{x}) = 0$ and $f(\widehat{x}) \leq f(x^*)$, which implies $\widehat{x}$ is a solution of $(\mathcal{P})$.

4. Sparse solutions for linear inequality systems

Let $A = (A_m^n)$ be a $M \times N$ matrix and let $b \in \mathbf{R}^M$ and consider the problem

$$\min\{ \ \|x\|_1 \ : \ Ax \leq b \ \}. \tag{15}$$

This is the convex relaxation of the nonconvex problem (see [17]) of finding the sparsest solutions to the system of inequalities $Ax \leq b$, which is stated as

$$\min\{ \ \|x\|_0 \ : \ Ax \leq b \ \},$$

where $\| \cdot \|_0$ denotes the *counting norm* (number of nonzero entries). The interested reader may consult [11], [13], [16].

The problem defined in (15) can be restated as

$$\min\{ \ \|x\|_1 \ : \ Ax = y, \ y \leq b \ \}.$$

For $m = 1, \ldots, M$ take

$$p_m(y) = \Big[y_m - b_m \Big]_+ .$$

Begin with $(x^{k-1}, y^{k-1}, \mu^{k-1}, \nu^{k-1})$ and apply the multiplier prediction steps following (**A1**):

$$\tilde{\mu}^k = \mu^{k-1} + \lambda_k (Ax^{k-1} - y^{k-1})$$

and for $m = 1, \ldots, M$

$$\tilde{\nu}_m^k = \begin{cases} \nu_m^{k-1} & \text{if} \quad y_m^{k-1} \leq b_m \\ \nu_m^{k-1} + \lambda_k (y_m^{k-1} - b_m) & \text{otherwise.} \end{cases}$$

Next, the exact primal step with respect to the x-variable

$$-\frac{x^k - x^{k-1}}{\lambda_k} - A^* \tilde{\mu}^k \in \partial f(x^k)$$

reduces to

$$x_n^k = \begin{cases} x_n^{k-1} - \lambda_k (A^* \tilde{\mu}^k)_n - \lambda_k & \text{if} \quad x_n^{k-1} - \lambda_k (A^* \tilde{\mu}^k)_n \ > \ \lambda_k \\ x_n^{k-1} - \lambda_k (A^* \tilde{\mu}^k)_n + \lambda_k & \text{if} \quad x_n^{k-1} - \lambda_k (A^* \tilde{\mu}^k)_n \ < \ -\lambda_k \\ 0 & \text{if} \quad x_n^{k-1} - \lambda_k (A^* \tilde{\mu}^k)_n \ \in \ [-\lambda_k, \lambda_k] \end{cases}$$

for $n = 1, \ldots, N$. On the other hand, for the y-variable we have

$$-\frac{y^k - y^{k-1}}{\lambda_k} + \tilde{\mu}^k \in \sum_{m=1}^{M} \tilde{\nu}_m^k \partial p_m(y^k),$$

which we rewrite as

$$y_m^k = \begin{cases} y_m^{k-1} + \lambda_k \tilde{\mu}_m^k - \lambda_k \tilde{\nu}_m^k & \text{if} \quad y_m^{k-1} + \lambda_k \tilde{\mu}_m^k - b_m \ > \ \lambda_k \tilde{\nu}_m^k \\ y_m^{k-1} + \lambda_k \tilde{\mu}_m^k & \text{if} \quad y_m^{k-1} + \lambda_k \tilde{\mu}_m^k - b_m \ < \ 0 \\ b_m & \text{if} \quad y_m^{k-1} + \lambda_k \tilde{\mu}_m^k - b_m \ \in \ [0, \lambda_k \tilde{\nu}_m^k] \end{cases}$$

for $m = 1, \dots, M$.

Finally we update the multipliers

$$\mu^k = \mu^{k-1} + \lambda_k (Ax^k - y^k)$$

and for $m = 1, \dots, M$

$$\nu_m^k = \begin{cases} \nu_m^{k-1} & \text{if } \; y_m^k \le b_m \\ \nu_m^{k-1} + \lambda_k (y_m^k - b_m) & \text{otherwise.} \end{cases}$$

A simple illustration. With no intention to test the numerical performance of the method we present the following academic example to illustrate the implementation. Let

$$A = \begin{pmatrix} -1 & 0 & -1 & 1 & 0 & 1 \\ 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ -1 & 1 & 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & -1 & -1 & 1 \\ -1 & -1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & -1 & -1 & 1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} -2 \\ -1 \\ -1 \\ 0 \\ -2 \\ -1 \\ 0 \end{pmatrix}.$$

The sparsest solution of the system of inequalitites given by $Ax \le b$ is

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

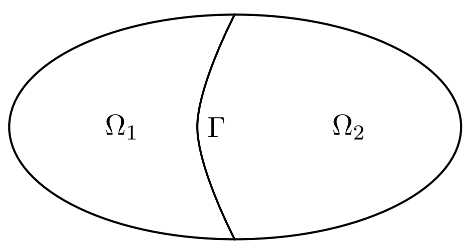
We implement our algorithm in SCILAB with $\lambda_k \equiv 0.4$, starting from 10 randomly generated initial points in $[-2, 2]^6$. The average outcome after 20 iterations was

$$\tilde{x} = (0\ 0\ 1.0052\ 0\ 0\ \ -0.9913)'$$

and the average processing time was 0.1 seconds in a laptop computer with a U9300 Intel(R) Core(TM)2 CPU and 3 GB of RAM.

5. Domain decomposition for partial differential equations

Let us consider a bounded domain $\Omega = \Omega_1 \cup \Omega_2 \cup \Gamma$ of $\mathbf{R}^N$ which can be decomposed in two non overlapping Lipschitz subdomains Ω_1 and Ω_2 with a common interface Γ. We assume that $\mathcal{H}^{N-1}(\Gamma) > 0$, where $\mathcal{H}^{N-1}$ is the Hausdorff measure of dimension $N - 1$. This situation is illustrated in the next figure.



Let $h \in L^2(\Omega)$. We consider the following problem

$$\begin{array}{ll} \text{Minimize} & \left\{ \frac{1}{2} \int_{\Omega_1} |\nabla u|^2 - \int_{\Omega_1} hu + \frac{1}{2} \int_{\Omega_2} |\nabla v|^2 - \int_{\Omega_2} hv \right\}; \\ \text{subject to} & (u, v) \in H^1(\Omega_1) \times H^1(\Omega_2) \text{ and } u_{|\Gamma} \ge v_{|\Gamma}. \end{array} \tag{16}$$

This kind of minimization problems often arises in the description of phenomena where the boundary is free, *i.e.* no external action is exerted on $\partial\Omega$, and involving discontinuities through the interface Γ. Here we consider the problem where the jump when passing from Ω_1 to Ω_2 is nonnegative. The case with no condition on the jump through the interface is treated in [1] with

Dirichlet conditions on the boundary of Ω and in [15] with Neumann conditions. In [6] (respectively [3]) the authors consider the problem with a no-jump condition through the interface and Dirichlet conditions on the boundary of Ω (respectively Neumann conditions), which amounts to solving a Dirichlet problem (respectively Neumann problem) on the whole set Ω by decomposition.
Notice that Problem (16) is not coercive. Under the assumptions $\int_\Omega h = 0$ and $\int_{\Omega_1} h < 0$, we can use [2, Theorem 15.1.2] to prove the existence of solutions.
Let us now show how the algorithm described by (**A1**)−(**A2′**)−(**A3**) can be applied to solve problem (16). The space $X = H^1(\Omega_1) \times H^1(\Omega_2)$ is equipped with the scalar product $\langle (u_1, v_1), (u_2, v_2) \rangle_X = \int_{\Omega_1} (\nabla u_1 . \nabla u_2 + u_1 u_2) + \int_{\Omega_2} (\nabla v_1 . \nabla v_2 + v_1 v_2)$ and the corresponding norm. The space $Y = L^2(\Gamma)$ is equipped with the scalar product $\langle y_1, y_2 \rangle_Y = \int_\Gamma y_1 y_2$ and the associated norm. We denote respectively $A_1 : H^1(\Omega_1) \to Y$ and $A_2 : H^1(\Omega_2) \to Y$ the trace operators on Γ. Problem (16) can be reformulated as problem $(\mathcal{P})$ with the following notations

$$(\mathcal{P}) \qquad \min \{ f(u) + g(v); \quad (u, v) \in X, \quad A(u, v) - y = 0, \quad y \in \mathcal{C} \},$$

where

$$f(u) = \frac{1}{2} \int_{\Omega_1} |\nabla u|^2 - \int_{\Omega_1} hu \quad \text{and} \quad g(v) = \frac{1}{2} \int_{\Omega_2} |\nabla v|^2 - \int_{\Omega_2} hv, \tag{17}$$

the operator $A : X \longrightarrow Y$ is defined by $A(u, v) = A_1 u - A_2 v$ and the set $\mathcal{C}$ is the closed convex cone of the space Y defined by $\mathcal{C} = \{ y \in Y; \quad y \geq 0 \}$. We now describe the computation of the primal steps. The auxiliary varibles y and ν are used in the computation of the Lagrange multiplier approximations $\tilde{\mu}^k$ and μ^k. Their definition depends on the particular choice of the function P. One can take $P(y) = d(y, \mathcal{C})$, which in this case is the L^2-norm of the negative part of y.

Description of the primal steps. A derivative computation allows to express the exact primal steps

$$\left\{ \begin{array}{l} u^k = \operatorname{Argmin} \left\{ f(u) + \langle \tilde{\mu}^k, A_1 u \rangle + \frac{1}{2\lambda_k} |u - u^{k-1}|^2; \quad u \in H^1(\Omega_1) \right\} \\ v^k = \operatorname{Argmin} \left\{ g(v) - \langle \tilde{\mu}^k, A_2 v \rangle + \frac{1}{2\lambda_k} |v - v^{k-1}|^2; \quad v \in H^1(\Omega_2) \right\}, \end{array} \right. \tag{18}$$

as

$$\left\{ \begin{array}{rcl} \int_{\Omega_1} \nabla u^k \cdot \nabla u + \frac{1}{\lambda_k} \int_{\Omega_1} \nabla (u^k - u^{k-1}) \cdot \nabla u + \frac{1}{\lambda_k} \int_{\Omega_1} (u^k - u^{k-1}) u & = & \int_{\Omega_1} hu - \int_\Gamma \tilde{\mu}^k A_1 u \\ \int_{\Omega_2} \nabla v^k \cdot \nabla v + \frac{1}{\lambda_k} \int_{\Omega_2} \nabla (v^k - v^{k-1}) \cdot \nabla v + \frac{1}{\lambda_k} \int_{\Omega_2} (v^k - v^{k-1}) v & = & \int_{\Omega_2} hv + \int_\Gamma \tilde{\mu}^k A_2 v, \end{array} \right.$$

for all $u \in H^1(\Omega_1)$ and $v \in H^1(\Omega_2)$. These are the variational weak formulations of the following mixed Dirichlet-Neumann boundary value problems respectively on Ω_1

$$\left\{ \begin{array}{rcll} -(1 + \frac{1}{\lambda_k}) \Delta u^k + \frac{1}{\lambda_k} u^k & = & h - \frac{1}{\lambda_k} \Delta u^{k-1} + \frac{1}{\lambda_k} u^{k-1} & \text{on} \quad \Omega_1 \\ (1 + \frac{1}{\lambda_k}) \frac{\partial u_k}{\partial \nu} & = & \frac{1}{\lambda_k} \frac{\partial u^{k-1}}{\partial \nu} & \text{on} \quad \partial\Omega_1 \cap \partial\Omega \\ (1 + \frac{1}{\lambda_k}) \frac{\partial u_k}{\partial \nu} & = & \frac{1}{\lambda_k} \frac{\partial u^{k-1}}{\partial \nu} - \tilde{\mu}^k & \text{on} \quad \Gamma, \end{array} \right.$$

and Ω_2

$$\left\{ \begin{array}{rcll} -(1 + \frac{1}{\lambda_k}) \Delta v^k + \frac{1}{\lambda_k} v^k & = & h - \frac{1}{\lambda_k} \Delta v^{k-1} + \frac{1}{\lambda_k} v^{k-1} & \text{on} \quad \Omega_2 \\ (1 + \frac{1}{\lambda_k}) \frac{\partial v_k}{\partial \nu} & = & \frac{1}{\lambda_k} \frac{\partial v^{k-1}}{\partial \nu} & \text{on} \quad \partial\Omega_2 \cap \partial\Omega \\ (1 + \frac{1}{\lambda_k}) \frac{\partial v_k}{\partial \nu} & = & \frac{1}{\lambda_k} \frac{\partial v^{k-1}}{\partial \nu} + \tilde{\mu}^k & \text{on} \quad \Gamma. \end{array} \right.$$

Convergence. Since this matter is out of the scope of this paper, we shall not enter into the details concerning the existence of saddle points here. Instead we shall assume that there are such points. Under these conditions, any sequence (u^k, v^k) generated by (18) converges *strongly* in

$H^1(\Omega_1)\times H^1(\Omega_2)$ to a solution $(\overline{u},\overline{v})$ of problem (16). Indeed, let $((u^k,v^k),y^k,\mu^k,\nu^k)$ be a sequence generated by $(\mathbf{A1})-(\mathbf{A2'})-(\mathbf{A3})$ so that (u^k,v^k) satisfies (18). In view of Theorem 7, (u^k,v^k) converges weakly in $H^1(\Omega_1)\times H^1(\Omega_2)$ to a minimum point $(\overline{u},\overline{v})$ of problem $(\mathcal{P})$. For the strong convergence, observe that, by the Rellich-Kondrachov Theorem, the sequence (u^k,v^k) converges to $(\overline{u},\overline{v})$ strongly in $L^2(\Omega_1)\times L^2(\Omega_2)$. Moreover, from Proposition 9 (ii), we have $\lim_{k\to+\infty} f(u^k)+g(v^k)=f(\overline{u})+g(\overline{v})$, which in turn implies that

$$\lim_{k\to+\infty}\int_{\Omega_1}|\nabla u^k|^2+\int_{\Omega_2}|\nabla v^k|^2=\int_{\Omega_1}|\nabla\overline{u}|^2+\int_{\Omega_2}|\nabla\overline{v}|^2.$$

As a consequence, we have $\lim_{k\to+\infty}|(u^k,v^k)|_{H^1(\Omega_1)\times H^1(\Omega_2)}=|(\overline{u},\overline{v})|_{H^1(\Omega_1)\times H^1(\Omega_2)}$ and we conclude that the convergence is strong.

Observe that the algorithm allows to solve the initial problem on Ω by solving separately problems on Ω_1 and Ω_2.

Acknowledgements

The authors are deeply indebted to Professor H. Attouch for his valuable comments and advice.

References

[1] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, Alternating proximal algorithms for weakly coupled convex minimization problems. Applications to dynamical games and PDE's, *J. Convex Anal.*, **15** (2008), No 3, pp. 485-506.

[2] H. Attouch, G. Buttazzo, G. Michaille, Variational analysis in Sobolev and BV spaces. Applications to PDE's and optimization, *MPS/SIAM Series on Optimization*, 6. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006.

[3] H. Attouch, A. Cabot, P. Frankel, J. Peypouquet, Alternating proximal algorithms for constrained variational inequalities. Applications to domain decomposition for PDE's, *Nonlinear Anal.*, **74** (2011), no. 18, 7455-7473.

[4] H. Attouch, M.O. Czarnecki, J. Peypouquet, Prox-penalization and splitting methods for constrained variational problems, *SIAM J. Opt.*, **21** (2011), pp. 149–173.

[5] H. Attouch, M.O. Czarnecki, J. Peypouquet, Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities. To appear in *SIAM J. Opt.*

[6] H. Attouch, M. Soueycatt, Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games , PDE's and control, *Pacific J. Opt.*, **5** (2009), No 1, pp. 17-37.

[7] A. Auslender, J.-P. Crouzeix, P. Fedit, Penalty-proximal methods in convex programming, *J. Optim. Theory Appl.* **55** (1987), no. 1, 1-21.

[8] D. Azé, Eléments d'analyse convexe et variationnelle, Ellipses, Paris, 1997.

[9] F. Bonnans, A. Shapiro, Perturbation Analysis of Optimization Problems, *Springer Series in Operations Research*, Springer-Verlag New York Inc., 2000.

[10] H. Brézis, P.-L. Lions, Produits infinis de résolvantes, *Israel J. Math.*, **29** (1978), 329-345.

[11] Candès E.-J., Romberg J.-K., Tao T., Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.*, **59** (2006), no. 8, 1207-1223.

[12] G. Chen, M. Teboulle, A proximal-based decomposition method for convex minimization problems, *Math. Programming* **64** (1994), no. 1, Ser. A, 81-101.

[13] S.-S. Chen, D.-L. Donoho, M.-A. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, **20** (1998), no. 1, 33-61.

[14] R. Cominetti, M. Courdurier, Coupling general penalty schemes for convex programming with the steepest descent method and the proximal point algorithm, *SIAM J. Optim*, **13** (2002), 745-765.

[15] P. Frankel, Alternating proximal algorithm with costs-to-move, dual description and application to PDE's, *to appear in Discrete Contin. Dynam. Systems - Series S.*

[16] M.-P. Friedlander, P. Tseng, Exact regularization of convex programs, *SIAM J. Optim.*, **18** (2007), no. 4, 1326-1350.

[17] J.-B. Hiriart-Urruty, H.-Y. Le, Convexifying the set of matrices of bounded rank. Applications to the quasiconvexification and convexification of the rank function, *preprint.*

[18] Z. Opial, Weak convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc.*, **73** (1967), pp. 591-597.

[19] J. Peypouquet, S. Sorin, Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. *J. Convex Anal.* **17** (2010), no. 3-4, 1113–1163.
[20] R.T. Rockafellar, Monotone operators associated with saddle-functions and minimax problems. 1970 Nonlinear Functional Analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 1, Chicago, Ill., 1968) pp. 241–250
[21] R.T. Rockafellar, Monotone Operators and the Proximal Point Algorithm, *SIAM J. Control Optim.* **14** (1976), 877-898.

Département de Mathématiques, Université Montpellier II, CC 051, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France
E-mail address: p.frankel30@orange.fr

Departamento de Matemática, Universidad Técnica Federico Santa María, Avenida España 1680, Valparaíso, Chile
E-mail address: juan.peypouquet@usm.cl