Journal of Convex Analysis
Volume 27 (2020), No. 3, 893–922

A Unified Splitting Algorithm for Composite Monotone Inclusions

Ernesto Oré-Albornoz
IMCA, Universidad Nacional de Ingeniería, Lima, Peru;
and: LIMOS, Université Clermont Auvergne, Clermont-Ferrand, France

Philippe Mahey
LIMOS, Université Clermont Auvergne, Clermont-Ferrand, France
mahey@isima.fr

Eladio Ocaña-Anaya
IMCA, Instituto de Matemática y Ciencias Afines,
Universidad Nacional de Ingeniería, Lima, Peru

Received: June 25, 2018
Revised manuscript received: May 3, 2019
Accepted: June 19, 2019

Operator splitting methods have been recently concerned with inclusions problems based on composite operators made of the sum of two monotone operators, one of them associated with a linear transformation. We analyze here a general and new splitting method which indeed splits both operator proximal steps, and avoiding costly numerical algebra on the linear operator. The family of algorithms induced by our generalized setting includes known methods like Chambolle-Pock primal-dual algorithm and Shefi-Teboulle Proximal Alternate Direction Method of Multipliers. The study of the ergodic and non ergodic convergence rates show similar rates with the classical Douglas-Rachford splitting scheme. We end with an application to a multi-block convex optimization model which leads to a generalized Separable Augmented Lagrangian Algorithm.

Keywords: Splitting methods, monotone inclusions, convergence analysis.

2010 Mathematics Subject Classification: 90C25, 90C30, 65K13.

1. Introduction

Composite models involving sums and compositions of linear and monotone operators are very common and still challenging problems like in constrained separable convex optimization or composite variational inequalities. We will consider here composite monotone inclusions of the form (X and Y are Hilbert spaces):

$$0 \in S(x) + A^*T(Ax) \tag{1}$$

where $S : X \mapsto X$ and $T : Y \mapsto Y$ are maximal monotone operators and $A : X \mapsto Y$ is a linear transformation (associated with its adjoint operator $A^* : Y \mapsto X$).

Most existing monotone operator splitting methods can deal with composite models, for example the Douglas-Rachford family (see [14]) and its special decomposition versions, the Alternate Direction Method of Multipliers (ADMM) (see

ISSN 0944-6532 / \$ 2.50

[11, 12]) and the Partial Inverse or Proximal Decomposition Algorithm (PDA) (see [16, 19, 24]).

Lions and Mercier [14] analyzed the Douglas-Rachford's method (including the limiting case of Peaceman-Rachford splitting, PRS) for the case of the sum of two maximal monotone operators $(S+T)$, alternating between proximal steps applied to each operator separately. Gabay [12] analyzed the case $S + A^*TA$ where A is an injective linear transformation (and A^* its adjoint), yielding the celebrated Alternative Direction Method of Multipliers (ADMM). Spingarn [23] studied the case when the operator is the sum of the normal cone of a closed subspace M and a maximal monotone operator T. Later, Pennanen [19] showed how to reformulate that model as a monotone inclusion

The first study which explicitly considered an algorithm to solve the composite inclusion which avoids the use of projection (or proximal) steps on the range of A was proposed in [5] (an extension of Spingarn's Partial Inverse to composite models was proposed too in [1]). The corresponding algorithms solve the dual problem at the same time, which is defined by:

$$0 \in -AS^{-1}(-A^*y) + T^{-1}(y)$$

Many applications surge in the minimization of separable convex functions like:

$$\text{Minimize} \quad f(x) + g(Ax) \tag{2}$$

where $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g\colon \mathbb{R}^m \to \overline{\mathbb{R}}$ are proper lower semi-continuous convex functions and A is a given $(m\times n)$ matrix (the adjoint operator A^* is thus identified with the transpose matrix A^t).

Because many important applications in different fields can be better described in a finite dimensional vector space and to avoid too lengthy convergence proofs, we will work on finite dimensions throughout the paper.

The dual problem in the sense of Rockafellar-Fenchel theory is:

$$\text{Minimize} \quad f^*(-A^ty) + g^*(y)$$

where $f^*(v^*) = \sup_x \langle x, v^*\rangle - f(x)$ is the conjugate function of f.

Recently Chambolle-Pock [6] studied model (2) and introduced new splitting schemes applied to a Lagrangian formulation of the primal minimization problem. They applied a primal-dual version of (ADMM) to the following saddle-point formulation:

$$\min_x \; \max_y \quad f(x) - g^*(y) + \langle Ax, y\rangle$$

Observe that, using the subdifferential operators $S = \partial f$ and $T = \partial g$, we could as well define a Lagrangian operator associated with the composite inclusion (1):

$$\overline{L}(x,y) = [S(x) + A^ty] \times [T^{-1}(y) - Ax] \tag{3}$$

Chambolle and Pock's algorithm relies on two Proximal steps on f and g with an additional extrapolation step (in a similar fashion of Varga's iterative principle [25]) as summarized in the following:

$$\begin{cases} x^{k+1} &= (I+\tau\partial f)^{-1}(x^k - \tau A^t \bar{y}^k) \\ y^{k+1} &= (I+\sigma\partial g^*)^{-1}(y^k + \sigma A x^{k+1}) \\ \bar{y}^{k+1} &= y^{k+1} + \theta(y^{k+1} - y^k) \end{cases}$$

where $(I+\tau\partial f)^{-1}$ is the resolvent operator of the subdifferential operator $S = \partial f$ which is known to be defined on the whole space and supposed to be easily computable in a so-called 'backward' proximal step as detailed below.

The difference and presumed advantage of that formulation is the symmetry (considering that x and y can be updated in reverse order) and a potentially decomposable algorithm which depends on three parameters. Their convergence result states that we should choose their values such that $\sigma\tau\|A\|^2 < 1$.

Observe now that (CPA) can be rewritten using Augmented Lagrangian-like functions by using the Moreau identity (see [17]) :

$$(I+\sigma\partial g^*)^{-1}(y) + \sigma(I+\sigma^{-1}\partial g)^{-1}(\sigma^{-1}y) = y$$

Resuming the transformed steps into the following iteration :

Algorithm (**CPA**)

$$\begin{cases} x^{k+1} &= \operatorname{argmin}_x\ f(x) + \frac{1}{2\tau}\|x - x^k + \tau A^t \bar{y}^k\|^2 \\ z^{k+1} &= \operatorname{argmin}_z\ g(z) + \frac{\sigma}{2}\|z - Ax^{k+1} - \sigma^{-1}y^k\|^2 \\ y^{k+1} &= y^k + \sigma(A\bar{x}^{k+1} - z^{k+1}) \\ \bar{y}^{k+1} &= y^{k+1} + \theta(y^{k+1} - y^k) \end{cases}$$

Chambolle and Pock confirmed the expected rate of convergence in $O(1/k)$ and even obtain the accelerated rate of $O(1/k^2)$ following the FISTA scheme of Beck and Teboulle [3] (thus reaching Nesterov's optimal rates in convex programming [18]).

In a recent survey, Shefi and Teboulle [22] have presented a unified scheme algorithm for solving model (2) based on the introduction of additional proximal terms like in Rockafellar's Proximal Method of Multipliers [20]. The resulting schemes include a version of a Proximal (ADMM) and other known algorithms like Chambolle-Pock's method (CPA). Indeed, a generic sequential algorithm proposed by Shefi and Teboulle is the following three steps scheme :

Algorithm (**STA**)

$$\begin{cases} x^{k+1} &= \operatorname{argmin}_x\ f(x) + \frac{\sigma}{2}\|Ax - z^k + \sigma^{-1}y^k\|^2 + \frac{1}{2}\|x - x^k\|^2_{M_1} \\ z^{k+1} &= \operatorname{argmin}_z\ g(z) + \frac{\sigma}{2}\|Ax^{k+1} - z + \sigma^{-1}y^k\|^2 + \frac{1}{2}\|z - z^k\|^2_{M_2} \\ y^{k+1} &= y^k + \sigma(Ax^{k+1} - z^{k+1}) \end{cases}$$

where $\|.\|_M$ is the norm induced by a symmetric positive definite matrix M, i.e. $\|x\|^2_M = \langle x, Mx\rangle$. The Algorithm (STA) makes use of alternate minimization steps on the Augmented Lagrangian function associated with the coupling subspace $Ax - z = 0$. It is noted in [22] that (CPA) with the choice $\theta = 1$ corresponds exactly to (STA) with $M_1 = \tau^{-1}I - \sigma A^t A$ and $M_2 = 0$ (which implies again that $\sigma\tau\|A\|^2 < 1$).

Later, Condat [7] extended the model (2) and algorithm (CPA) to the case $f = F + h$ where $F\colon \mathbb{R}^n \to \mathbb{R}$ is convex and smooth. He relaxed the restriction on the parameters allowing $\sigma\tau\|A\|^2 = 1$ and also includes the Douglas-Rachford family in the case of $A = I$ (therefore we can say that Chambolle-Pock's method generalized Douglas-Rachford's splitting scheme). Condat showed too that Chambolle-Pock's method is the proximal point method applied to the Lagrangian operator associated with the primal and dual pair of inclusions.

In this paper we will further extend the algorithms surveyed by Shefi and Teboulle, in order to solve the following convex optimization problem

$$\text{(P)} \qquad \min_{(x,z)} \quad [f(x) + g(z) : Ax + Bz = 0],$$

where f and g are again convex lsc functions and, A and B are two matrices of order $m \times n$ and $m \times p$, respectively. It is clear that this problem includes problem (2) by considering $B = -I_{p\times p}$.

The primal variational formulation of (P) is the following

$$\text{Find } (x,z) \in \mathbb{R}^n \times \mathbb{R}^p \text{ such that } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial f(x) \\ \partial g(z) \end{pmatrix} + \begin{pmatrix} A^t \\ B^t \end{pmatrix} \partial\delta_{\{0\}}(Ax+Bz), \tag{4}$$

where δ_C is the indicator function of C which is 0 on C and $+\infty$ otherwise.

The dual variational formulation of (P) is the following

$$\text{Find } \; y \in \mathbb{R}^m \; \text{ such that } \; 0 \in \; - A(\partial f)^{-1}(-A^t y) - B(\partial g)^{-1}(-B^t y). \tag{5}$$

In Section 2, we propose a generalized proximal point method (GPPM) which was developed implicitly by Condat [7], where we consider specific assumptions to relax the condition of symmetric positive definiteness of the matrix associated with the resolvent, to authorize matrices which are only symmetric positive semidefinite, maintaining the properties of convergence of the proximal method.

In Section 3, we apply GPPM in order to find a zero of the Lagrangian map associated with problem (P), selecting an appropriate symmetric positive semi definite matrix in order to obtain a Generalized Splitting Scheme (GSS), which includes various known algorithms, for instance both types of algorithms studied by Shefi and Teboulle [22] correspond indeed to particular choices of the parameters in GSS.

In Section 4, we define a 1-co-coercive operator G_P^T related to GPPM, which set of fixed points is related to the zeroes of T. When T is the Lagrangian operator and the matrix P has a special structure as considered in Section 3, we show examples where we can get that operator explicitly, in particular we can recover the Douglas-Rachford operator.

In Section 5, we investigate the rate of converge of the GSS scheme, in the ergodic and non ergodic sense, analyzing the convergence of the sequences of the optimal values and the constraints violations associated with problem (P).

Finally, section 6 applies the GSS scheme to some general multi-block convex optimization problem with a composite structure. We show the relationship with a separable Augmented Lagrangian algorithm (SALA) introduced in [15].

2. A generalized proximal point method

We begin by quoting some basic properties around monotonicity of point-to-set operators on $\mathbb{R}^n$. For a set-valued operator $T\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, denoting its graph set by $\mathrm{gr}(T) = \{(x, y) \mid y \in T(x)\}$, we consider the main properties used in this paper:

Definition 2.1. An operator T is *monotone* if for any (x, x^*), $(\bar{x}, \bar{x}^*) \in \mathrm{gr}\,(T)$, one has

$$\langle x^* - \bar{x}^*, x - \bar{x}\rangle \geq 0$$

It is *maximal monotone* if $\mathrm{gr}(T)$ is not strictly contained in the graph of another monotone operator.

Definition 2.2. An operator T is *strongly monotone with radius* $\rho > 0$ (or shortly ρ-strongly monotone) if $T - \rho I$ is monotone, i.e. for any (x, x^*), $(\bar{x}, \bar{x}^*) \in \mathrm{gr}\,(T)$, one has

$$\langle x^* - \bar{x}^*, x - \bar{x}\rangle \geq \rho\|x - \bar{x}\|^2.$$

For single-valued operators, we get the following properties:

Definition 2.3. An operator T is *Lipschitz continuous with constant* L (or shortly L-Lipschitz) if

$$\forall x, x' \;\; \|T(x) - T(x')\| \leq L\|x - x'\|$$

It is nonexpansive if $L \leq 1$.

Definition 2.4. An operator T is α*-averaged* if $T = (1 - \alpha)I + \alpha R$, where R is a nonexpansive operator.

A 1/2-averaged operator is also called *firmly nonexpansive.* For example, the resolvent of a maximal monotone operator $J^T = (I + T)^{-1}$ is firmly nonexpansive (and defined on the whole space).

Definition 2.5. An operator T is *co-coercive with constant* β (or shortly β-co-coercive) if T^{-1} is β-strongly monotone

In this case βT is also firmly nonexpansive.

The classical *Proximal Point method* is used to solve a monotone inclusion

$$\text{(V)} \qquad \text{Find } \; x \in \mathbb{R}^r \; \text{ such that } \; 0 \in T(x)$$

where $T\colon \mathbb{R}^r \rightrightarrows \mathbb{R}^r$ is a maximal monotone operator. We denote by $\mathrm{sol}\,(V)$ the solution set of problem (V). It is closed, convex and may be empty. The iteration exploits the contractive properties of the resolvent operator $J^T_\tau = (I + \tau T)^{-1}$ to define a sequence given by $x^{k+1} = J_\tau^T(x^k)$ which converges weakly to a solution of (V) if it is nonempty.

Following former ideas developed by Condat [7] in the proof of convergence of a specialized splitting method closely related to (CPA), we define the generalized Proximal Point iteration by substituting the classical resolvent by

$$J_P^T := (T+P)^{-1}P \tag{6}$$

where P is an $r \times r$ symmetric positive semidefinite matrix.

Since T is monotone, then for any (x, x^*), $(\bar{x}, \bar{x}^*) \in \operatorname{graph}(J_P^T)$, one has

$$\langle x^* - \bar{x}^*, Px - P\bar{x}\rangle \geq \langle Px^* - P\bar{x}^*, x^* - \bar{x}^*\rangle \geq 0. \tag{7}$$

We deduce immediately the following properties :

- $T+P$ and thereby its inverse $(T+P)^{-1}$ are monotone.
- $R := P + I_{r\times r} - Q$ is a symmetric positive definite matrix, whenever Q is the orthogonal projection onto the image of P, which implies in particular that Q satisfies $QP = PQ = P$ and $Q^2 = Q$.
- $J_P^T = J_P^T Q$, where Q is as above.

As R is symmetric positive definite, it induces an inner product on $\mathbb{R}^r$ by setting $\langle u, v\rangle_R := \langle Ru, v\rangle$ for all $u, v \in \mathbb{R}^r$ with its corresponding norm $\|u\|_R := \sqrt{\langle u, u\rangle_R}$ for all $u \in \mathbb{R}^r$. Hence, from (7), for all $x, \bar{x} \in \operatorname{dom}(QJ_P^T) = \operatorname{dom}(J_P^T)$,

$$\langle QJ_P^T(x) - QJ_P^T(\bar{x}), x - \bar{x}\rangle_R \geq \|QJ_P^T(x) - QJ_P^T(\bar{x})\|_R^2,$$

which implies that QJ_P^T is *R-co-coercive* on domain of J_P^T (we will use the shortcut R-co-coercive for 1-co-coercive with respect to metric R throughout the text).

We deduce immediately the following relationship between the solution set of problem (V) and the fixed points of J_P^T and QJ_P^T.

Proposition 2.6. *With the same notation as before, we have*

- $x \in \operatorname{sol}(V)$ *if and only if* x *is a fixed point of* J_P^T.
- v *is a fixed point of* QJ_P^T *if and only if* $v = Qx$ *for some* $x \in \operatorname{sol}(V) \cap J_P^T(v)$.

Proof. The first property is directly by definition. The second one follows from the fact that $v \in QJ_P^T v$ if and only if there exists x such that $x \in J_P^T(v)$ satisfying $v = Qx$. It follows that $x \in J_P^T(v) = J_P^T(Qx) = J_P^T(x)$. Using the first equivalence we deduce that x belongs to $\operatorname{sol}(V)$. □

Concerning the *regularity* of J_P^T, we have

- If P is positive definite, then $Q = I_{r\times r}$ and $R = P$. We deduce that $J_P^T = QJ_P^T$ and then J_P^T is P-co-coercive on the whole of its domain.
- If P is not positive definite, then J_P^T may not be single valued. But if it is single valued, then it is continuous on the whole of its domain.

We consider now a relaxed version of the generalized proximal iteration. In connection with the resolvent operator J_P^T and a real positive parameter ρ, we consider for an arbitrary point $x_0 \in \mathrm{dom} J_P^T$, the sequence $\{x^k\}$ defined by

$$x^{k+1} \in \rho J_P^T(x^k) + (1-\rho)x^k. \tag{8}$$

Notice that this sequence is well defined whenever

$$\mathrm{range}\,(\rho J_P^T + (1-\rho)I) \subseteq \mathrm{dom}(J_P^T).$$

Concerning the convergence of $\{x^k\}$, we distinguish the following situations:

- If P is positive definite, then J_P^T is P-co-coercive (hence single valued) with full domain which implies that $\{x^k\}$ converges, for $\rho \in (0,2)$, assuming $\mathrm{sol}\,(V)$ nonempty. In fact, given $x^* \in \mathrm{sol}\,(V)$, the convergence follows from the inequality
$$\|x^k - x^*\|_P^2 \geq \frac{2-\rho}{\rho}\|x^{k+1} - x^k\|_P^2 + \|x^{k+1} - x^*\|_P^2.$$
- In general, since QJ_P^T is R-co-coercive, then for $\rho \in (0,2)$ and assuming that QJ_P^T has closed domain and nonempty fixed point set (which is equivalently to say that $\mathrm{sol}\,(V)$ is nonempty), the sequence $\{Qx^k\}$ is convergent. The convergence of $\{x^k\}$ needs additional assumptions as we show in the following proposition.

Proposition 2.7. *Let $T\colon \mathbb{R}^r \rightrightarrows \mathbb{R}^r$ be maximal monotone and P be an $r \times r$ positive semidefinite matrix. Assuming J_P^T single valued (which implies that it is continuous) with closed domain and* $\mathrm{sol}\,(V)$ *not empty. Then, for $\rho \in (0,2)$, the sequence $\{x^k\}$ converges to some point belonging to* $\mathrm{sol}\,(V)$.

Proof. Since QJ_P^T is R-co-coercive, it is single valued on its domain; and since $J_P^T = J_P^T Q$, then from (8) we obtain that

$$Qx^{k+1} = \rho Q J_P^T(Qx^k) + (1-\rho)Qx^k. \tag{9}$$

Using again the fact that QJ_P^T is R-co-coercive and, by assumptions with closed domain, $\rho \in (0,2)$ and $\mathrm{sol}\,(V)$ nonempty, then $\{Qx^k\}$ converges to some point a, which is a fixed point of QJ_P^T. From Proposition 2.6 and the single valuedness assumption, $J_P^T(a) \in \mathrm{sol}\,(V)$.

On the other hand, using the triangular inequality in (8) we have

$$\|x^{k+1} - J_P^T(a)\| \leq \rho\|J_P^T(Qx^k) - J_P^T(a)\| + |1-\rho|\|x^k - J_P^T(a)\|.$$

Since J_P^T is continuous, the sequence $\|J_P^T(Qx^k) - J_P^T(a)\|$ converges to 0. We deduce that $\{x^k\}$ converges to $J_P^T(a)$. □

Some examples of specially tailored co-coercive operators will be discussed in Section 4.

Remark 2.8. The hypothesis in the last proposition over J_P^T seem to be restrictive. If T is strongly monotone, the single valuedness with full domain of J_P^T are deduced. In fact, for an arbitrary positive semidefinite matrix P there is a non strong monotone map T such that J_P^T is single valued with full domain. Indeed, decomposing P as

$$P = EDE^t$$

where D is the diagonal matrix consisting of eigenvalues of P and E an orthogonal matrix, i.e, satisfying $EE^t = E^tE = I$, and considering a singular diagonal matrix $\hat{D}$ such that $D + \hat{D}$ is not singular, we have that the map

$$T = E\hat{D}E^t$$

is maximal monotone but not strongly monotone. Furthermore, the resolvent map $J_P^T = E(D + \hat{D})^{-1}DE^t$ is single valued with full domain. □

In the next section, we work with an special map which under some conditions its resolvent map is single valued with full domain.

3. Generalized splitting algorithms

With the convex minimization problem (P) defined in Section 1, we associate its Lagrangian function defined as

$$l(x, z, y) = f(x) + g(z) + \langle y, Ax + Bz\rangle \tag{10}$$

and then its saddle-point problem in the variational setting

$(\mathrm{V_L})$ Find $(\bar{x}, \bar{z}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ such that $0 \in L(\bar{x}, \bar{z}, \bar{y})$,

where L is the maximal monotone map defined on $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ as

$$L(x, z, y) := (\partial_{x,z} l) \times (\partial_y[-l]) = \begin{pmatrix} \partial f(x) \\ \partial g(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}. \tag{11}$$

The map L, as the sum of maximal monotone operators and a skew-symmetric linear operator, satisfies similar inequalities as the subdifferential of a convex-concave bifunction. These inequalities will be used in order to obtain the rate of convergence studied in Section 5.

Proposition 3.1. *For any* (d, d^*), $(\bar{d}, \bar{d}^*) \in \operatorname{graph}(L)$, *considering* $d = (x, z, y)$ *and* $\bar{d} = (\bar{x}, \bar{z}, \bar{y})$, *it holds*

$$\langle d - \bar{d}, d^*\rangle \geq l(x, z, \bar{y}) - l(\bar{x}, \bar{z}, y) \geq \langle d - \bar{d}, \bar{d}^*\rangle.$$

These inequalities can still be verified if we consider $(d, d^*) \in \operatorname{graph}(L)$ *and* $\bar{d} \in \operatorname{dom}(f) \times \operatorname{dom}(g) \times \mathbb{R}^m$, *for the first inequality; and* $(\bar{d}, \bar{d}^*) \in \operatorname{graph}(L)$ *and* $d \in \operatorname{dom}(f) \times \operatorname{dom}(g) \times \mathbb{R}^m$, *for the second inequality.*

It is well known that under some regularity conditions, problem (V_L) admits a saddle-point if and only if problem (P) admits an optimal solution. One instance of such regularity condition is :

(H) There exist $x \in \mathrm{ri}(\mathrm{dom} f)$ and $z \in \mathrm{ri}(\mathrm{dom} g)$ such that $Ax + Bz = 0$,

where $\mathrm{ri}(C)$ stands for the relative interior of set C which is the interior set within the affine hull of C.

We now apply to problem (V_L) the relaxed proximal method described in the previous section for a specially tailored matrix P in order to provide a separable structure to the algorithm.

3.1. The separable structure on the main step

In this part we describe the main iteration step of the relaxed proximal method given in (8) providing a decomposable structure.

We will choose an appropriate symmetric matrix P in order to split $(L+P)^{-1}$ or equivalently $J_P^L = (L+P)^{-1}P$, into a separable structure leaving f and g separated.

To that end, given $(\tilde{x}, \tilde{z}, \tilde{y}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, we analyze the solution of the following inclusion system: Find (x, z, y) such that

$$\begin{pmatrix} \partial f(x) \\ \partial g(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix} + \begin{pmatrix} P_{11} & P_{21}^t & P_{31}^t \\ P_{21} & P_{22} & P_{32}^t \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix} \ni \begin{pmatrix} \tilde{x} \\ \tilde{z} \\ \tilde{y} \end{pmatrix}.$$

We introduce now two parameters $\alpha, \beta \in \mathbb{R}$, and a positive definite matrix M to simplify the third row-block of P into $P_3 = [(1+\alpha)A \quad (1+\beta)B \quad M^{-1}]$. So, the last inclusion can be expressed as

$$y = M\tilde{y} - \alpha MAx - \beta MBz \tag{12}$$

and hence, replacing it in the second block-system, this results in

$$\partial g(z) + (2+\beta)B^t(M\tilde{y} - \alpha MAx - \beta MBz) + P_{21}x + P_{22}z \ni \tilde{z}.$$

So, in order to express this last system eliminating primal variable x, we need to consider $P_{21} = \alpha(2+\beta)B^t MA$, obtaining

$$z \in (\partial g + P_{22} - \beta(2+\beta)B^t MB)^{-1}(\tilde{z} - (2+\beta)B^t M\tilde{y}). \tag{13}$$

Using again (12), now in the first block system, we get

$$\partial f(x) + (2+\alpha)A^t(M\tilde{y} - \alpha MAx - \beta MBz) + P_{11}x + \alpha(2+\beta)B^t MAz \ni \tilde{x}$$

which is equivalent to

$$x \in (\partial f + P_{11} - \alpha(2+\alpha)A^t MA)^{-1}(\tilde{x} - (2+\alpha)A^t M\tilde{y} - 2(\alpha-\beta)A^t MBz). \tag{14}$$

Summarizing the previous sequence in order to get a separable structure, we must first solve system (13), then system (14) and finally system (12). The corresponding matrix P, of order $(r \times r)$ with $r = n + p + m$, is then of the form

$$P := \begin{pmatrix} C_1 & \alpha(2+\beta)A^tMB & (1+\alpha)A^t \\ \alpha(2+\beta)B^tMA & C_2 & (1+\beta)B^t \\ (1+\alpha)A & (1+\beta)B & M^{-1} \end{pmatrix} \tag{15}$$

where $C_1(n \times n), C_2(p \times p)$ are arbitrary symmetric matrices,

From the maximality of ∂f and ∂g, the inclusions in (13) and (14) are indeed equalities if the matrices defined as

$$W_1 := C_1 - \alpha(2+\alpha)A^tMA \quad \text{and} \quad W_2 := C_2 - \beta(2+\beta)B^tMB,$$

are positive definite. In that case $(L+P)^{-1}$ is single-valued with full domain and therefore J_P^L is continuous with full domain.

It is clear that P is symmetric. It is positive semidefinite (resp. positive definite) if and only if the matrix

$$U := \begin{pmatrix} C_1 - (1+\alpha)^2A^tMA & (\alpha-\beta-1)A^tMB \\ (\alpha-\beta-1)B^tMA & C_2-(1+\beta)^2B^tMB \end{pmatrix} \tag{16}$$

is positive semidefinite (resp. positive definite).

We now list some conditions in order to get a positive semidefinite matrix U:

(A1) If $C_1 - [(1+\alpha)^2 + (\alpha-\beta-1)^2]A^tMA$ and $C_2 - [(1+\beta)^2+1]B^tMB$ are positive semidefinite then U is positive semidefinite.

(A2) If $C_1 - [(1+\alpha)^2+1]A^tMA$ and $C_2 - [(1+\beta)^2 + (\alpha-\beta-1)^2]B^tMB$ are positive semidefinite then U is positive semidefinite.

(A3) If $\beta \leq \alpha - 1$, and $C_1 - [(1+\alpha)^2 + (\alpha-\beta-1)]A^tMA$ and $C_2 - [(1+\beta)^2 + (\alpha-\beta-1)]B^tMB$ are positive semidefinite then U is positive semidefinite.

(A4) If $\beta = \alpha - 1$. Then $C_1 - (1+\alpha)^2A^tMA$ and $C_2 - \alpha^2B^tMB$ are positive semidefinite if only if U is positive semidefinite.

In order to calculate the sequence in (8), we first calculate $(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) = J_P^L(x^k, z^k, y^k)$, which is equal to

$$(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) = (L+P)^{-1} \begin{pmatrix} C_1x^k + \alpha(2+\beta)A^tMBz^k + (1+\alpha)A^ty^k \\ \alpha(2+\beta)B^tMAx^k + C_2z^k + (1+\beta)B^ty^k \\ (1+\alpha)Ax^k + (1+\beta)Bz^k + M^{-1}y^k \end{pmatrix}$$

Then from (13), we have

$$\tilde{z}^{k+1} = \bar{J}^g_{W_2}\left(\tilde{z} - (\beta+2)B^tMAx^k\right) \tag{17}$$

where $\tilde{z} = C_2z^k - (2+\beta)(1+\beta)B^tMBz^k - B^ty^k$ and $\bar{J}^g_{W_2} = (\partial g + W_2)^{-1}$ is the generalized resolvent operator associated with the convex function g.

From (14), we have

$$\tilde{x}^{k+1} = \bar{J}^f_{W_1}\left(\tilde{x} - 2(\alpha-\beta)A^tMB\tilde{z}^{k+1}\right) \tag{18}$$

where $\tilde{x} = C_1x^k - (2+\alpha)(1+\alpha)A^tMAx^k + (\alpha - 2\beta - 2)A^tMBz^k - A^ty^k$ and $\bar{J}^f_{W_1} = (\partial f + W_1)^{-1}$ is the generalized resolvent operator associated with the convex function f; and from (12), we get

$$\tilde{y}^{k+1} = y^k + (1+\alpha)MAx^k + (1+\beta)MBz^k - \alpha MA\tilde{x}^{k+1} - \beta MB\tilde{z}^{k+1}. \tag{19}$$

The sequence in (8) is completed with an extrapolation step for a given $\rho \in (0,2)$:

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, z^k, y^k). \tag{20}$$

We obtain the following proposition directly from Proposition 2.7.

Proposition 3.2. *Let $\rho \in (0,2)$. Assume that $C_1 \in \mathbb{R}^{n\times n}$, $C_2 \in \mathbb{R}^{p\times p}$ and $M \in \mathbb{R}^{m\times m}$ are symmetric, with M positive definite; and α, $\beta \in \mathbb{R}$, such that W_1 and W_2 are positive definite and satisfying one of conditions* (A1)-(A4)*. If* $\operatorname{sol}(V_L)$ *is nonempty, then for an arbitrary $(x^0, z^0, y^0) \in \mathbb{R}^{n+p+m}$, the sequence (x^k, z^k, y^k) defined by the sequential update formulas* ($17 \to 18 \to 19 \to 20$) *converges to some element of* $\operatorname{sol}(V_L)$.

3.2. The generalized splitting scheme

We will now further reformulate the iteration to show the alternating steps on separable Augmented Lagrangian functions. We introduce the parameter $\gamma = \alpha - \beta$ and the matrices defined as

$$V_1 := W_1 - A^tMA \quad \text{and} \quad V_2 := W_2 - B^tMB. \tag{21}$$

The conditions (A1)–(A4) become:

- **(A1')** If $V_1 - (\gamma-1)^2A^tMA$ and $V_2 - B^tMB$ are positive semidefinite then U is positive semidefinite.
- **(A2')** If $V_1 - A^tMA$ and $V_2 - (\gamma-1)^2B^tMB$ are positive semidefinite then U is positive semidefinite.
- **(A3')** If $\gamma \geq 1$. Then $V_1 - (\gamma-1)A^tMA$ and $V_2 - (\gamma-1)B^tMB$ are positive semidefinite then U is positive semidefinite.
- **(A4')** If $\gamma = 1$. Then V_1 and V_2 are positive semidefinite if only if U is positive semidefinite.

We introduce a new primal-dual auxiliary variable

$$u^k := y^k + (\alpha - \gamma + 1)MAx^k + (1+\beta)MBz^k,$$

to obtain the following updates:

$$z^{k+\frac{1}{2}} = V_2z^k - B^tu^k \tag{22}$$

$$\tilde{z}^{k+1} = J^g_{W_2}[z^{k+\frac{1}{2}} - B^tMAx^k] \tag{23}$$

$$x^{k+\frac{1}{2}} = V_1x^k - \gamma A^tMAx^k + (\gamma-1)A^tMBz^k - A^tu^k \tag{24}$$

$$\tilde{x}^{k+1} = J^f_{W_1}[x^{k+\frac{1}{2}} - 2\gamma A^tMB\tilde{z}^{k+1}] \tag{25}$$

$$\tilde{u}^{k+1} = u^k + \gamma MAx^k + (1-\gamma)MA\tilde{x}^{k+1} + MB\tilde{z}^{k+1} \tag{26}$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1-\rho)(x^k, z^k, u^k) \tag{27}$$

which is equivalent to the following sequential minimization subproblems:

Generalized Splitting Scheme (GSS)

$$\tilde{z}^{k+1} \in \operatorname{argmin} \left\{ g(z) + \tfrac{1}{2}\|Bz + M^{-1}u^k + Ax^k\|^2_M + \tfrac{1}{2}\|z - z^k\|^2_{V_2} \right\} \tag{28}$$

$$v^{k+\frac{1}{2}} = \gamma Ax^k - (\gamma-1)Bz^k + M^{-1}u^k \tag{29}$$

$$\tilde{x}^{k+1} \in \operatorname{argmin} \left\{ f(x) + \tfrac{1}{2}\|Ax + v^{k+\frac{1}{2}} + 2\gamma B\tilde{z}^{k+1}\|^2_M + \tfrac{1}{2}\|x - x^k\|^2_{V_1} \right\} \tag{30}$$

$$\tilde{u}^{k+1} = u^k + M(\gamma Ax^k + (1-\gamma)A\tilde{x}^{k+1} + B\tilde{z}^{k+1}) \tag{31}$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1-\rho)(x^k, z^k, u^k). \tag{32}$$

From Proposition 3.2, we obtain the proposition of convergence of (GSS)

Proposition 3.3. *Let $\rho \in (0, 2)$. Assume that $V_1 \in \mathbb{R}^{n\times n}$, $V_2 \in \mathbb{R}^{p\times p}$ and $M \in \mathbb{R}^{m\times m}$ are symmetric, with M positive definite such that $V_1 + A^tMA$ and $V_2 + B^tMB$ are positive definite. Let $\gamma \in \mathbb{R}$ be such that one of the conditions $(A1')-(A4')$ is satisfied. If $\operatorname{sol}(V_L)$ is nonempty, then for an arbitrary $(x^0, z^0, u^0) \in \mathbb{R}^{n+p+m}$, the sequence (x^k, z^k, u^k) in (28)–(32) converges to some element of $\operatorname{sol}(V_L)$.*

We analyze now the special cases when $\gamma = 0$ and $\gamma = 1$, which correspond to the two types of algorithms proposed by Shefi and Teboulle [22].

3.2.1. Case $\gamma = 0$

From $(A1')$, if both matrices $V_1 - A^tMA$ and $V_2 - B^tMB$ are positive semi-definite then P is a positive semi-definite matrix.

Switching the order (28) for (30), we get the following algorithm where the primal updates are performed in parallel:

$$\tilde{x}^{k+1} \in \operatorname{argmin} \left\{ f(x) + \tfrac{1}{2}\|Ax + Bz^k + M^{-1}u^k\|^2_M + \tfrac{1}{2}\|x - x^k\|^2_{V_1} \right\} \tag{33}$$

$$\tilde{z}^{k+1} \in \operatorname{argmin} \left\{ g(z) + \tfrac{1}{2}\|Ax^k + Bz + M^{-1}u^k\|^2_M + \tfrac{1}{2}\|z - z^k\|^2_{V_2} \right\} \tag{34}$$

$$\tilde{u}^{k+1} = u^k + M(A\tilde{x}^{k+1} + B\tilde{z}^{k+1}) \tag{35}$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1-\rho)(x^k, z^k, u^k) \tag{36}$$

If $B = -I_{p\times p}$, $M = cI_{p\times p}$ and $\rho = 1$, we obtain the algorithm STA type I proposed by Shefi and Teboulle [22].

Summarizing, from Proposition 3.3, we obtain the following proposition of convergence of the sequence defined by (33)–(36).

Proposition 3.4. *Let $\rho \in (0,2)$. Assume that $V_1 \in \mathbb{R}^{n\times n}$, $V_2 \in \mathbb{R}^{p\times p}$ and $M \in \mathbb{R}^{m\times m}$ are symmetric, with M positive definite, such that $V_1 + A^tMA$ and $V_2 + B^tMB$ are positive definite and $V_1 - A^tMA$ and $V_2 - B^tMB$ are positive semi-definite. If* $\operatorname{sol}(V_L)$ *is nonempty, then for an arbitrary $(x^0, z^0, u^0) \in \mathbb{R}^{n+p+m}$, the sequence (x^k, z^k, u^k) in (33)–(36) converges to some element of* $\operatorname{sol}(V_L)$.

3.2.2. Case $\gamma = 1$

From (A4'), it follows that V_1 and V_2 are positive semi-definite if only if P is a positive semi-definite matrix. In this case GSS becomes:

$$\tilde{z}^{k+1} \in \operatorname{argmin}\left\{g(z) + \tfrac{1}{2}\|Ax^k + Bz + M^{-1}u^k\|_M^2 + \tfrac{1}{2}\|z - z^k\|_{V_2}^2\right\} \tag{37}$$

$$\tilde{u}^{k+1} = u^k + M(Ax^k + B\tilde{z}^{k+1}) \tag{38}$$

$$\tilde{x}^{k+1} \in \operatorname{argmin}\left\{f(x) + t\tfrac{1}{2}\|Ax + B\tilde{z}^{k+1} + M^{-1}\tilde{u}^{k+1}\|_M^2 + t\tfrac{1}{2}\|x - x^k\|_{V_1}^2\right\} \tag{39}$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1-\rho)(x^k, z^k, u^k) \tag{40}$$

If $B = -I_{p\times p}$, $M = \tau I_{p\times p}$, $V_2 = 0$ and $V_1 = \sigma^{-1}I_{n\times n} - \tau A^tTA$ with $1 \geq \sigma\tau\|A\|^2$, then we obtain the over relaxed algorithm proposed by Chambolle-Pock [6].

Considering $\rho = 1$ and defining, $\bar{x}^k := x^k$, $\bar{z}^k := z^{k+1}$ and $\bar{u}^k := u^{k+1}$, then substituting in (37)-(39) and switching the order, we get the following algorithm

$$\bar{x}^{k+1} \in \operatorname{argmin}\left\{f(x) + \tfrac{1}{2}\|Ax + B\bar{z}^k + M^{-1}\bar{u}^k\|_M^2 + \tfrac{1}{2}\|x - \bar{x}^k\|_{V_1}^2\right\} \tag{41}$$

$$\bar{z}^{k+1} \in \operatorname{argmin}\left\{g(z) + \tfrac{1}{2}\|A\bar{x}^{k+1} + Bz + M^{-1}\bar{u}^k\|_M^2 + \tfrac{1}{2}\|z - \bar{z}^k\|_{V_2}^2\right\} \tag{42}$$

$$\bar{u}^{k+1} = \bar{u}^k + M(A\bar{x}^{k+1} + B\bar{z}^{k+1}) \tag{43}$$

If $B = -I_{p\times p}$ and $M = cI_{p\times p}$, we obtain the algorithm STA type II proposed by Shefi and Teboulle [22], which is called the Proximal Alternating Direction Method (PADM).

Further transformations applied to (37)–(40) lead us to consider two interesting algorithms. The first of them is obtained by considering $V_2 = 0$, and considering the auxiliary variables $\hat{x}^{k+1}, \hat{z}^k$, $\hat{u}^k$, $\hat{s}^k$ to update the relaxed sequences $\hat{x}^{k+1} := \frac{1}{\rho}x^{k+1} + (1-\frac{1}{\rho})x^k = \tilde{x}^{k+1}$, $\hat{z}^k := \frac{1}{\rho}z^{k+1} + (1-\frac{1}{\rho})z^k = \tilde{z}^{k+1}$, $\hat{u}^k := \tilde{u}^{k+1}$ and $\hat{s}^k := x^k$, getting

$$\begin{aligned}
\hat{x}^{k+1} &\in \operatorname{argmin}\left\{f(x) + \tfrac{1}{2}\|Ax + B\hat{z}^k + M^{-1}\hat{u}^k\|_M^2 + \tfrac{1}{2}\|x - \hat{s}^k\|_{V_1}^2\right\} && (44)\\
\hat{z}^{k+1} &\in \operatorname{argmin}\left\{g(z) + \tfrac{1}{2}\|\rho A\hat{x}^{k+1} + Bz + M^{-1}\hat{u}^k + (\rho-1)B\hat{z}^k\|_M^2\right\} && (45)\\
\hat{u}^{k+1} &= \hat{u}^k + \rho MA\hat{x}^{k+1} + (\rho-1)MB\hat{z}^k + MB\hat{z}^{k+1} && (46)\\
\hat{s}^{k+1} &= \rho\hat{x}^{k+1} + (1-\rho)\hat{s}^k && (47)
\end{aligned}$$

The second interesting algorithm is obtained by considering the auxiliary variables $\check{x}^k, \check{z}^k$, $\check{u}^k$, $\check{s}^k$ to update the relaxed sequences $\check{x}^k := \frac{1}{\rho}x^{k+1} + (1-\frac{1}{\rho})x^k = \tilde{x}^{k+1}$,

$\check{z}^k := \frac{1}{\rho} z^{k+1} + (1 - \frac{1}{\rho}) z^k = \tilde{z}^{k+1}$, $\check{u}^k := \tilde{u}^{k+1}$ and $\check{s}^k := x^k$, getting

$$\check{z}^{k+1} \in \operatorname{argmin} \left\{ g(z) + \tfrac{1}{2} \| \rho A \check{x}^k + Bz + M^{-1} \check{u}^k + (\rho - 1) B \check{z}^k \|_M^2 \right\} \tag{48}$$

$$\check{u}^{k+1} = \check{u}^k + \rho M A \check{x}^k + (\rho - 1) M B \check{z}^k + M B \check{z}^{k+1} \tag{49}$$

$$\check{s}^{k+1} = \rho \check{x}^k + (1 - \rho) \check{s}^k \tag{50}$$

$$\check{x}^{k+1} \in \operatorname{argmin} \left\{ f(x) + \tfrac{1}{2} \| Ax + B \check{z}^{k+1} + M^{-1} \check{u}^{k+1} \|_M^2 + \tfrac{1}{2} \| x - \check{s}^{k+1} \|_{V_1}^2 \right\} \tag{51}$$

So, by considering in these two last algorithms $B = -I_{p \times p}$, $M = cI_{p \times p}$ and $V_1 = 0$, the sequences $\hat{s}^k$ and $\check{s}^k$ becomes unnecessary. Moreover, (44)-(47) become the generalized ADMM proposed by Eckstein [11], and (48)-(51) become the algorithm 2 consider in [9].

From Proposition 3.3, we obtain the convergence of the sequence (37)-(40)

Proposition 3.5. *Let $\rho \in (0, 2)$. Assume that $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ and $M \in \mathbb{R}^{m \times m}$ are symmetric, with V_1 and V_2 positive semi-definite and M positive definite such that $V_1 + A^t M A$ and $V_2 + B^t M B$ are positive definite. If $\operatorname{sol}(V_L)$ is nonempty, then for an arbitrary $(x^0, z^0, u^0) \in \mathbb{R}^{n+p+m}$, the sequence (x^k, z^k, u^k) defined in (37)–(40) converges to some element of $\operatorname{sol}(V_L)$.*

4. The co-coercive map associated with GPPM

Lions and Mercier [14] have transformed an inclusion problem for the sum of two maximal monotone operators $(S + T)$ into a fixed-point equation with respect to an appropriate operator, the *Douglas-Rachford* operator, which is 1-co-coercive map and, in order to compute its value at each point of its domain, only local calculations of proximal terms on S and T separately are needed. Eckstein [11] later showed the relationship between the splitting algorithm (ADMM) and the fixed-point method applied to a Douglas-Rachford operator, after a suitable linear transformation.

In our general setting, we show in this section that the sequence generated by the generalized proximal point method (GPPM) corresponding to map J_P^T for arbitrary maximal monotone operator T and arbitrary symmetric positive semidefinite matrix P is nothing else but the sequence generated by the fixed point method corresponding to map G_S^T defined in (53), after a linear transformation S (satisfying $P = S^t S$). It leads thus in some sense to a generalization of the Douglas-Rachford operator, keeping the property of 1-co-coercivity.

As pointed out in section 3.1, the sequence generated by GPPM for $T = L$ defined in (11) and P defined in (15) corresponds to the sequence generated by the generalized splitting scheme (GSS) defined in ($17 \to 18 \to 19 \to 20$).

In Section 2, we have shown that the sequence generated by GPPM is nothing else but, under the linear transformation Q, the sequence generated by the fixed point method corresponding to the R-co-coercive map QJ_P^T (see (9)). Nevertheless, for arbitrary symmetric positive semidefinite matrix P, matrices Q and R are difficult to calculate; of course, when P is symmetric positive definite, then $Q = I$ and $R = P$. Alternately by considering S such that $P = S^tS$, we define G_S^T an operator easier to implement than QJ_P^T and having similar properties, for example, a 1-co-coercive operator. In particular, using G_S^T instead of QJ_P^T, we will give an alternative proof of Proposition 2.7.

Finally, by considering $S = S_3$ defined in Remark 4.6, one obtains $G_{S_3}^L = S_3^t(L + S_3^tS_3)^{-1}S_3$ which corresponds, under a reparametrization, to the classical Douglas-Rachford operator defined by $M^{-\frac{1}{2}}S_3^t(L+S_3^tS_3)^{-1}S_3M^{\frac{1}{2}}$. In other words, the Douglas-Rachford operator and its fundamental properties of co-coercivity and splittability will be show to be a special case of our generalized setting based on the Lagrangian monotone inclusion.

Associated with the $r \times r$ symmetric positive semidefinite matrix P introduced in the former section, let consider a $q \times r$ matrix S satisfying

$$P = S^tS \tag{52}$$

and then the map $G_S^T \colon \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ defined as

$$G_S^T := S(T + S^tS)^{-1}S^t. \tag{53}$$

It follows that
$$SJ_P^T = G_S^TS \tag{54}$$

and hence, from (7), we get for all $w, w' \in \mathbb{R}^r$:

$$\langle G_S^T(Sw) - G_S^T(Sw'), Sw - Sw'\rangle \geq \|G_S^T(Sw) - G_S^T(Sw')\|^2.$$

Since for any $s, s' \in \mathbb{R}^q$ there exist $w, w' \in \mathbb{R}^r$ such that $S^tSw = S^ts$ and $S^tSw' = S^ts'$, we get

$$\langle G_S^T(s) - G_S^T(s'), s - s'\rangle \geq \|G_S^T(s) - G_S^T(s')\|^2$$

which means that G_S^T is 1-co-coercive.

The following proposition shows in particular that G_S^T is the Moreau-Yosida regularization of $ST^{-1}S^t$. This will be used in the examples considered in this Section and in Section 6 (Proposition 6.1).

Proposition 4.1. *Let $T \colon \mathbb{R}^r \rightrightarrows \mathbb{R}^r$ be an arbitrary map, S and M two matrices of order $q \times r$ and $q \times q$, respectively, with M invertible. For $z \in \mathbb{R}^q$ the value $(ST^{-1}S^t + M)^{-1}Mz$ is nonempty if and only if $(T + S^tM^{-1}S)^{-1}S^tz$ is nonempty. Furthermore, it holds that*

$$(ST^{-1}S^t + M)^{-1}Mz = z - M^{-1}S(T + S^tM^{-1}S)^{-1}S^tz.$$

Proof. The proof results from the following two properties:

- $x \in (ST^{-1}S^t + M)^{-1}Mz$ if and only if there exists $y \in \mathbb{R}^m$ such that $S^tx \in T(y)$ and $z - M^{-1}Sy = x$.
- $y^* \in (T + S^tM^{-1}S)^{-1}S^tz$ if and only if exists $x^* \in \mathbb{R}^r$ such that $S^tx^* \in T(y^*)$ and $z - M^{-1}Sy^* = x^*$. $\square$

Similar to Proposition 2.6, we get the relationship between the solution set of problem (V) and the fixed points of G_S^T.

Proposition 4.2. *With the same notations as before, we have*

- *If $z \in \mathrm{sol}\,(V)$, then Sz is a fixed point of G_S^T.*
- *If w is a fixed point of G_S^T, then $w = Sq$ for some $q \in \mathrm{sol}\,(V) \cap (T+P)^{-1}S^t w$.*

We deduce that the set of fixed point of G_S^T is exactly

$$S(\mathrm{sol}\,(V)) = \{Sw : w \in \mathrm{sol}\,(V)\}.$$

Applying S to the sequence $\{w^k\}$ defined in (8) and considering the permutation property (54), we get:

$$Sw^{k+1} = \rho G_S^T(Sw^k) + (1-\rho)Sw^k. \tag{55}$$

This equation gives us another alternative proof of convergence of the sequence $\{w^k\}$ under the same conditions of Proposition 2.7. In fact, since G_S^T is 1-co-coercive and from (55), we have that given $w^* \in \mathrm{sol}\,(V)$

$$\|Sw^k - Sw^*\|^2 - \frac{2-\rho}{\rho}\|Sw^{k+1} - Sw^k\|^2 - \|Sw^{k+1} - Sw^*\|^2 \geq 0 \tag{56}$$

Since $\mathrm{rank}\, S^tS = \mathrm{rank}\, S^t$, the domain of G_S^T is equal to the domain of J_P^T which is closed, using this fact and from (56) we deduce that Sw^k converges to some point b, which is a fixed point of G_S^T. On the other hand, using the triangular inequality and considering $\tilde{w} := (T+P)^{-1}S^t b$, we get

$$\|w^{k+1} - \tilde{w}\| \leq \rho\|(T+P)^{-1}S^t(Sw^k) - \tilde{w}\| + |1-\rho|\|w^k - \tilde{w}\|.$$

From the continuity of J_P^T, we deduce the continuity of $(T+P)^{-1}S^t = J_P^T S^+$, where S^+ denotes the Moore-Penrose pseudo-inverse matrix of S. Therefore we deduce that $\{w^k\}$ converges to $\tilde{w}$.

We now give some explicit expressions of G_S^L for the Lagrangian operator L and matrix S such that $P = S^tS$, considered in Section 3.

4.1. Examples of co-coercive operators G_S^L

Example 4.3. Let $\gamma = 1$ ($\beta = \alpha - 1$), We consider in (15),

$$C_1 = V_1 + (1+\alpha)^2 A^t M A \quad \text{and} \quad C_2 = V_2 + \alpha^2 B^t M B,$$

where V_1 and V_2 are as (21) assumed positive semidefinite matrices. In (37)–(40) matrices V_1 and V_2 are associated with the additional proximal term that will be used in ADMM, which, as we have shown in Subsection 3.2.2, it is related to Shefi-Teboulle algorithm type II [22]. We get:

$$P = \begin{pmatrix} V_1 + (1+\alpha)^2 A^t M A & (1+\alpha)\alpha A^t M B & (1+\alpha)A^t \\ (1+\alpha)\alpha B^t M A & V_2 + \alpha^2 B^t M B & \alpha B^t \\ (1+\alpha)A & \alpha B & M^{-1} \end{pmatrix}.$$

The matrix
$$S_1 = \begin{pmatrix} V_1^{\frac{1}{2}} & 0 & 0 \\ 0 & V_2^{\frac{1}{2}} & 0 \\ (1+\alpha)M^{\frac{1}{2}}A & \alpha M^{\frac{1}{2}}B & M^{-\frac{1}{2}} \end{pmatrix}$$

verifies (52) and the corresponding map $G_{S_1}^L$, that applies $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ into itself, is defined as

$$G_{S_1}^L(\hat{x}, \hat{z}, \hat{y}) = \begin{pmatrix} V_1^{\frac{1}{2}} x \\ V_2^{\frac{1}{2}} z \\ M^{\frac{1}{2}}Ax + M^{\frac{1}{2}}Bz + \hat{y} \end{pmatrix}$$

where
$$x = (\partial f + V_1 + A^t MA)^{-1}(V_1^{\frac{1}{2}}\hat{x} - A^t M^{\frac{1}{2}}(\hat{y} + 2M^{\frac{1}{2}}Bz))$$
$$z = (\partial g + V_2 + B^t MB)^{-1}(V_2^{\frac{1}{2}}\hat{z} - B^t M^{\frac{1}{2}}\hat{y}).$$

Note that $G_{S_1}^L$ has full domain if $V_1 + A^t MA$ and $V_2 + B^t MB$ are assumed positive definite matrices. □

In the two following remarks we will use the following notation for arbitrary maps T_1 and T_2 and vectors x and y of appropriated dimensions:

$$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} T_1^{-1}x \\ T_2^{-1}y \end{pmatrix}.$$

Remark 4.4. The map $G_{S_1}^L$ is the Douglas–Rachford operator [14], applied to the two maps

$$-\begin{pmatrix} V_1^{\frac{1}{2}} & 0 \\ 0 & I_{p\times p} \\ M^{\frac{1}{2}}A & 0 \end{pmatrix} \begin{bmatrix} \partial f \\ 0 \end{bmatrix}^{-1} \left(-\begin{pmatrix} V_1^{\frac{1}{2}} & 0 & A^t M^{\frac{1}{2}} \\ 0 & I_{p\times p} & 0 \end{pmatrix} \right)$$

and

$$-\begin{pmatrix} -I_{n\times n} & 0 \\ 0 & -V_2^{\frac{1}{2}} \\ 0 & M^{\frac{1}{2}}B \end{pmatrix} \begin{bmatrix} 0 \\ \partial g \end{bmatrix}^{-1} \left(-\begin{pmatrix} -I_{n\times n} & 0 & 0 \\ 0 & -V_2^{\frac{1}{2}} & B^t M^{\frac{1}{2}} \end{pmatrix} \right)$$

The corresponding sum of these two maps is exactly the *dual variational map* (5) associated with the following optimization problem

$$\min_{(x_1,x_2,z_1,z_2)\in\mathcal{F}} (f,0)(x_1,x_2) + (0,g)(z_1,z_2)$$

where $\mathcal{F}$ is the set of all triples (x_1, x_2, z_1, z_2) satisfying

$$\begin{pmatrix} V_1^{\frac{1}{2}} & 0 \\ 0 & I_{p\times p} \\ M^{\frac{1}{2}}A & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} -I_{n\times n} & 0 \\ 0 & -V_2^{\frac{1}{2}} \\ 0 & M^{\frac{1}{2}}B \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 0. \qquad □$$

Remark 4.5. In the case that $V_2 = 0$, which corresponds to Chambolle-Pock algorithm as we showed in section 3.2.2, we can restrict the map $G^L_{S_1}$, and obtain the map D_1 that applies $\mathbb{R}^n \times \mathbb{R}^m$ into itself, where $D_1(x,u)$ is

$$\begin{pmatrix} V_1^{\frac{1}{2}}(\partial f + V_1 + A^tMA)^{-1}[V_1^{\frac{1}{2}}x - A^tM^{\frac{1}{2}}(u+2z)] \\ M^{\frac{1}{2}}A(\partial f + V_1 + A^tMA)^{-1}[V_1^{\frac{1}{2}}x - A^tM^{\frac{1}{2}}(u+2z)] + z + u \end{pmatrix}$$

where $$z = M^{\frac{1}{2}}B(\partial g + B^tMB)^{-1}B^tM^{\frac{1}{2}}(-u).$$

Note if B is injective, then D_1 has full domain.

The map D_1 can be obtained in the form (53), considering that when $V_2 = 0$, the matrix

$$S_2 = \begin{pmatrix} V_1^{\frac{1}{2}} & 0 & 0 \\ (1+\alpha)M^{\frac{1}{2}}A & \alpha M^{\frac{1}{2}}B & M^{-\frac{1}{2}} \end{pmatrix}$$

verifies (52), and we obtain that $D_1 = G^L_{S_2}$.

The map D_1 can also be obtained as the Douglas–Rachford operator, applied to the two maps

$$-\begin{pmatrix} V_1^{\frac{1}{2}} \\ M^{\frac{1}{2}}A \end{pmatrix} (\partial f)^{-1} \left(-\begin{pmatrix} V_1^{\frac{1}{2}} & A^tM^{\frac{1}{2}} \end{pmatrix} \right)$$

and $$-\begin{pmatrix} -I_{n\times n} & 0 \\ 0 & M^{\frac{1}{2}}B \end{pmatrix} \begin{bmatrix} 0 \\ \partial g \end{bmatrix}^{-1} \left(-\begin{pmatrix} -I_{n\times n} & 0 \\ 0 & B^tM^{\frac{1}{2}} \end{pmatrix} \right)$$

The corresponding sum of these two maps is exactly the dual variational map associated with the following optimization problem

$$\min_{(x,z_1,z_2)\in\mathcal{F}} f(x) + (0,g)(z_1,z_2)$$

where $\mathcal{F}$ is the set of all triple (x,z_1,z_2) satisfying

$$\begin{pmatrix} V_1^{\frac{1}{2}} \\ M^{\frac{1}{2}}A \end{pmatrix} x + \begin{pmatrix} -I_{n\times n} & 0 \\ 0 & M^{\frac{1}{2}}B \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 0. \qquad \square$$

Remark 4.6. In the case $V_1 = 0$ and $V_2 = 0$, we can restrict the map $G^L_{S_1}$, and obtain the map D_2 that applies $\mathbb{R}^m$ into itself, where $D_2(x,u)$ is

$$M^{\frac{1}{2}}A(\partial f + A^tMA)^{-1}A^tM^{\frac{1}{2}}[-u-2z] + z + u,$$

where $$z = M^{\frac{1}{2}}B(\partial g + B^tMB)^{-1}B^tM^{\frac{1}{2}}(-u).$$

Note that if A and B are injective, then D_2 has full domain.

The map D_2 can be obtained in the form (53), considering that when $V_1 = V_2 = 0$, the matrix

$$S_3 = \begin{pmatrix} (1+\alpha)M^{\frac{1}{2}}A & \alpha M^{\frac{1}{2}}B & M^{-\frac{1}{2}} \end{pmatrix}$$

verifies (52), and we obtain that $D_2 = G^L_{S_3}$.

The map D_2 can also be obtained as the Douglas–Rachford operator [14], applied to the two maps

$$-M^{\frac{1}{2}}A(\partial f)^{-1}(-A^tM^{\frac{1}{2}}) \quad \text{and} \quad -M^{\frac{1}{2}}B(\partial g)^{-1}(-B^tM^{\frac{1}{2}}).$$

The corresponding sum of these two maps is exactly the *dual variational map* (5) associated with the following optimization problem

$$\min_{(x,y)} \quad [f(x)+g(z) : M^{\frac{1}{2}}Ax + M^{\frac{1}{2}}Bz = 0].$$

Alternatively we can consider, instead D_2, the map $\tilde{D}_2 := M^{-\frac{1}{2}}D_2M^{\frac{1}{2}}$, i.e

$$\tilde{D}_2(\bar{u}) = A(\partial f + A^tMA)^{-1}A^tM[-\bar{u}-2z] + z + \bar{u}$$

where

$$z = B(\partial g + B^tMB)^{-1}B^tM(-\bar{u}),$$

which is M-co-coercive. □

Example 4.7. Let $\gamma = 0$ ($\alpha = \beta$). We consider in (15),

$$C_1 = (1+(\alpha+1)^2)A^tMA + R \quad \text{and} \quad C_2 = (1+(\alpha+1)^2)B^tMB$$

where R is a positive semidefinite matrix. Then V_1 and V_2 in (21) are equal to

$$V_1 = A^tMA + R \text{ and } V_2 = B^tMB.$$

These matrices are associated with the additional proximal term considered in (33)-(36), which, as we have shown in Subsection 3.2.1, it is related to Shefi-Teboulle algorithm type I [22]. We get :

$$P = \begin{pmatrix} (1+(\alpha+1)^2)A^tMA + R & \alpha(2+\alpha)A^tMB & (1+\alpha)A^t \\ \alpha(2+\alpha)B^tMA & (1+(\alpha+1)^2)B^tMB & (1+\alpha)B^t \\ (1+\alpha)A & (1+\alpha)B & M^{-1} \end{pmatrix}.$$

The matrix $$S_4 = \begin{pmatrix} R^{\frac{1}{2}} & 0 & 0 \\ M^{\frac{1}{2}}A & -M^{\frac{1}{2}}B & 0 \\ (1+\alpha)M^{\frac{1}{2}}A & (1+\alpha)M^{\frac{1}{2}}B & M^{-\frac{1}{2}} \end{pmatrix}$$

verifies (52) and hence the value $G^L_{S_4}(\hat{x},\hat{z},\hat{y})$ of the corresponding map $G^L_{S_4}$, that applies $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ into itself, is

$$\begin{pmatrix} R^{\frac{1}{2}}x \\ M^{\frac{1}{2}}Ax - M^{\frac{1}{2}}Bz \\ M^{\frac{1}{2}}Ax + M^{\frac{1}{2}}Bz + \hat{y} \end{pmatrix}$$

where $$x = (\partial f + 2A^tMA + R)^{-1}(R^{\frac{1}{2}}\hat{x} + A^tM^{\frac{1}{2}}(\hat{z}-\hat{y}))$$

$$z = (\partial g + 2B^tMB)^{-1}B^tM^{\frac{1}{2}}(-\hat{z}-\hat{y}).$$

Note that $G^L_{S_4}$ has full domain if $2A^tMA + R$ and $2B^tMB$ are assumed positive definite matrices.

Remark 4.8. In the case that $R = 0$, we can restrict the map $G^L_{S_4}$, and obtain the map D_3 that applies $\mathbb{R}^m \times \mathbb{R}^m$ into itself, where $D_3(\hat{z}, \hat{y})$ is

$$\begin{pmatrix} M^{\frac{1}{2}}A(\partial f + 2A^tMA)^{-1}A^tM^{\frac{1}{2}}(\hat{z}-\hat{y}) + M^{\frac{1}{2}}(\partial g + 2B^tMB)^{-1}B^tM^{\frac{1}{2}}(-\hat{z}-\hat{y}) \\ M^{\frac{1}{2}}A(\partial f + 2A^tMA)^{-1}A^tM^{\frac{1}{2}}(\hat{z}-\hat{y}) - M^{\frac{1}{2}}(\partial g + 2B^tMB)^{-1}B^tM^{\frac{1}{2}}(-\hat{z}-\hat{y}) + \hat{y} \end{pmatrix}$$

The map D_3 can be obtained as the form (53), considering that when $V_1 = A^tMA$ and $V_2 = B^tMB$, the matrix

$$S_5 = \begin{pmatrix} M^{\frac{1}{2}}A & -M^{\frac{1}{2}}B & 0 \\ (1+\alpha)M^{\frac{1}{2}}A & (1+\alpha)M^{\frac{1}{2}}B & M^{-\frac{1}{2}} \end{pmatrix}.$$

verifies (52), then we obtain that $D_3 = G^L_{S_5}$. □

5. Rate of convergence

The global rate of convergence of ADMM and other monotone operator splitting algorithms has motivated many research contributions that we cannot survey here (see [9] for example). We will recover these results for the generalized splitting scheme GSS with no further refinements (like uniform or strong convexity) and will remain in the framework of finite-dimensional spaces (see [2] for similar results in Hilbert spaces).

Davis and Yin [9] have show the ergodic and nonergodic convergence rate of the feasibility and objective function error related to the relaxed PRS and relaxed ADMM, which is a particular case of our general scheme as remarked in Subsection 3.2.2. Similarly, in this Section, without regularity assumption, we show the ergodic and nonergodic convergence rate of the constraint violations (feasibility) and objective function error related to the chain of steps $17 \rightarrow\ 18 \rightarrow 19 \rightarrow\ 20$, defined in Subsection 3.1, which is our main sequence asociated with primal problem (P) defined in the first section.

With the same expressions of matrices P and U defined in (15) and (16), respectively, we get the following identity by using S satisfying $P = S^tS$ and explicit expressions of P and U,

$$\begin{aligned} \|(x,z,y)\|^2_P &= \|S(x,z,y)\|^2 \\ &= \|(x,z)\|^2_U + \|M^{\frac{1}{2}}((1+\alpha)Ax + (1+\beta)Bz) + M^{-\frac{1}{2}}y\|^2. \end{aligned} \tag{57}$$

Notice that for $\gamma = 0$ ($\beta = \alpha$),

$$\|(x,z)\|^2_U = \|x\|^2_{V_1 - A^tMA} + \|z\|^2_{V_2 - B^tMB} + \|Ax - Bz\|^2_M$$

and for $\gamma = 1$ ($\beta = \alpha - 1$),

$$\|(x,z)\|^2_U = \|x\|^2_{V_1} + \|z\|^2_{V_2}. \tag{58}$$

Back to the sequence ($17 \to 18 \to 19 \to 20$) and considering $w^k = (x^k, z^k, y^k)$, it holds from definition that

$$J_P^L w^k = (\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) \text{ and } w^{k+1} = \rho J_P^L w^k + (1-\rho) w^k. \tag{59}$$

The following proposition will be used later in Subsection 5.2 in order to estimate an upper bound of the optimal value of problem (P).

Proposition 5.1. *We consider $w = (x, z, y) \in \mathrm{dom}(f) \times \mathrm{dom}(g) \times \mathbb{R}^m$. Using the same notation as before, the following inequality holds:*

$$\|w^k - w\|_P^2 - \frac{2-\rho}{\rho}\|w^{k+1} - w^k\|_P^2 - \|w^{k+1} - w\|_P^2 \geq 2\rho\left[l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y) - l(x, z, \tilde{y}^{k+1})\right]$$

Proof. Let $w = (x, z, y) \in \mathrm{dom}(f) \times \mathrm{dom}(g) \times \mathbb{R}^m$.
Since $P(w^k - J_P^L w^k) \in L(J_P^L w^k)$, then using Proposition 3.1, it holds that

$$\left\langle J_P^L w^k - w, P(w^k - J_P^L w^k)\right\rangle \geq l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y) - l(x, z, \tilde{y}^{k+1}). \tag{60}$$

On the other hand, from the symmetry of P, we have

$$2\rho\left\langle J_P^L w^k - w, P(w^k - J_P^L w^k)\right\rangle = \|w^k - w\|_P^2 - \frac{2-\rho}{\rho}\|w^{k+1} - w^k\|_P^2 - \|w^{k+1} - w\|_P^2$$

So, replacing this last expression in (60), we get the desired inequality. $\square$

In particular, from the inequality in the last proposition, we get

$$\|w^k - w\|_P^2 - \|w^{k+1} - w\|_P^2 \geq 2\rho\left[l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y) - l(x, z, \tilde{y}^{k+1})\right]. \tag{61}$$

This inequality will be used in Theorem 5.2 for approximating the optimal value of problem (P).

We note that Proposition 5.1 is a general version of the inequality given in Proposition 2 of [9] by considering $A = I = -B$, $M = \gamma^{-1} I$ and P as in Remark 4.6, $w = (x, x, 0)$ (which implies $M^{-\frac{1}{2}} S_3 w = x$), $z = M^{-\frac{1}{2}} S_3 z^k$, and

$$z^+ = (T_{PRS})_\lambda(z) = (M^{-\frac{1}{2}} G_{S_3}^L M^{\frac{1}{2}})_{2\lambda}(M^{-\frac{1}{2}} S_3 z^k) = M^{-\frac{1}{2}} S_3 w^{k+1}.$$

Similarly, Proposition 5.1 is also a general version of the one given in Proposition 11 of [9] by considering $M = \gamma I$ and P as in Remark 4.6; $(\bar{x}^*, \bar{z}^*, \bar{y}^*)$ and z^* fixed points of $G_{S_3}^L$ and $(T_{PRS})_\lambda = (M^{\frac{1}{2}} G_{S_3}^L M^{-\frac{1}{2}})_{2\lambda}$, respectively; w^k satisfying $M^{\frac{1}{2}} S_3 w^k = z^k$ and $w = (\bar{x}^*, \bar{z}^*, 0)$ such that

$$M^{\frac{1}{2}} S_3 w = M^{\frac{1}{2}} S_3(\bar{x}^*, \bar{z}^*, \bar{y}^*) - \bar{y}^* = z^* - w^*$$

where $w^* = J_{\gamma(-B)(\partial g)^{-1}(-B^t)}(z^*)$.

5.1. Bounding the fixed-point residual

The *fixed-point residual* of operator $\rho G_S^T+(1-\rho)I_{q\times q}$ is the sequence whith general term

$$\|(\rho G_S^T+(1-\rho)I_{q\times q})Sw^k - Sw^k\|^2$$

which, from (55), is equal to $\|Sw^{k+1} - Sw^k\|^2$.

Since $\rho \in (0,2)$, then $\rho G_S^T+(1-\rho)I_{q\times q}$ is non expansive and hence the sequence $\{\|Sw^{k+1} - Sw^k\|\}$ is non increasing. Summing over $k = 0,\cdots,N-1$ in (56), we get

$$\|Sw^k - Sw^{k-1}\|^2 \leq \frac{\rho}{(2-\rho)k}\|Sw^0 - Sw^*\|^2. \tag{62}$$

On the other hand, using the Jensen's inequality, we get

$$\|Sw^k - Sw^0\|^2 \leq 2\|Sw^k - Sw^*\|^2 + 2\|Sw^0 - Sw^*\|^2 \leq 4\|Sw^0 - Sw^*\|^2$$

and hence

$$\left\|\frac{1}{N}\sum_{k=1}^{N}(Sw^k - Sw^{k-1})\right\|^2 = \frac{1}{N^2}\|Sw^N - Sw^0\|^2 \leq \frac{4}{N^2}\|Sw^0 - Sw^*\|^2. \tag{63}$$

Notice that upper bounds (62) and (63) can also be deduced respectively from Theorem 1 "Notes on Theorem 1" and Theorem 2 developed in D. Davis and W. Yin [9].

5.2. Bounding the saddle-point gap

We consider the following *ergodic sequences* defined as: for $N \geq 1$,

$$\overline{x}_N := \frac{1}{N}\sum_{k=1}^{N}\tilde{x}^k, \quad \overline{z}_N := \frac{1}{N}\sum_{k=1}^{N}\tilde{z}^k \quad \text{and} \quad \overline{y}_N := \frac{1}{N}\sum_{k=1}^{N}\tilde{y}^k.$$

Theorem 5.2. *With the same notations as before, we get the following rate of convergence:*

- *Ergodic Convergence: for any* $w = (x,z,y) \in \mathrm{dom}(f)\times\mathrm{dom}(g)\times\mathbb{R}^m$

$$l(\overline{x}_k,\overline{z}_k,y) - l(x,z,\overline{y}_k) \leq \frac{1}{2\rho k}\|Sw^0 - Sw\|^2. \tag{64}$$

- *Nonergodic Convergence: for any* $w^* = (x^*,z^*,y^*) \in \mathrm{sol}\,(V_L)$

$$l(\tilde{x}^{k+1},\tilde{z}^{k+1},y^*) - l(x^*,z^*,\tilde{y}^{k+1}) \leq \frac{1+|1-\rho|}{\rho\sqrt{\rho(2-\rho)(k+1)}}\|Sw^0 - Sw^*\|^2. \tag{65}$$

Proof. Summing (61) over $k = 0,\cdots,N-1$, and applying the Jensen's inequality to the convex functions $l(\cdot,\cdot,y) - l(x,z,\cdot)$ for arbitrary fixed elements $x \in \mathrm{dom}(f)$,

$z \in \operatorname{dom}(g)$ and $y \in \mathbb{R}^m$, where l is the lagrangian function defined in (10) of Section 3, we deduce the desired ergodic convergence.

Given $w^* \in \operatorname{sol}(V_L)$ and considering $w = w^*$ in (60), we get

$$\langle G_S^L Sw^k - Sw^*, Sw^k - G_S^L Sw^k\rangle \geq l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y^*) - l(x^*, z^*, \tilde{y}^{k+1}) \geq 0$$

and hence, from the Cauchy-Schwarz inequality and (55), we obtain

$$\frac{1}{\rho}\|G_S^L Sw^k - Sw^*\|\|Sw^{k+1} - Sw^k\| \geq l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y^*) - l(x^*, z^*, \tilde{y}^{k+1}). \tag{66}$$

On other hand, from (55) and since $\{\|Sw^{k+1} - Sw^*\|\}$ is non increasing, we get

$$\begin{aligned}\|G_S^L Sw^k - Sw^*\| &= \|\frac{1}{\rho}(Sw^{k+1} - Sw^*) + (1 - \frac{1}{\rho})(Sw^k - Sw^*)\| \\ &\leq \frac{1 + |1-\rho|}{\rho}\|Sw^0 - Sw^*\|.\end{aligned}$$

So, replacing this last expression and inequality (62) in expression (66), we deduce the desired nonergodic convergence. □

5.3. Bounding constraint violations

We consider, for $N \geq 1$, $\hat{x}_N := \frac{1}{N}\sum_{k=1}^{N} x^{k-1}$ and $\hat{z}_N := \frac{1}{N}\sum_{k=1}^{N} z^{k-1}$.

We get the following theorem

Theorem 5.3. *With the same notations as before, for any $w^* \in \operatorname{sol}(V_L)$, we get the following rate of convergence:*

- *Ergodic Convergence:*

$$\|(\overline{x}_k - \hat{x}_k, \overline{z}_k - \hat{z}_k)\|_U^2 + \|A\overline{x}_k + B\overline{z}_k\|_M^2 \leq \frac{4}{\rho^2 k^2}\|Sw^0 - Sw^*\|^2.$$

- *Nonergodic Convergence:*

$$\|(\tilde{x}^k - x^{k-1}, \tilde{z}^k - z^{k-1})\|_U^2 + \|A\tilde{x}^k + B\tilde{z}^k\|_M^2 \leq \frac{1}{(2-\rho)\rho k}\|Sw^0 - Sw^*\|^2.$$

Proof. From (59) we have $w^k - w^{k-1} = \rho(\tilde{x}^k - x^{k-1},\, \tilde{z}^k - z^{k-1}, \tilde{y}^k - y^{k-1})$ and hence, from (19), we get

$$\begin{aligned}&w^k - w^{k-1} = \\ &= \rho(\tilde{x}^k - x^{k-1},\, \tilde{z}^k - z^{k-1},\, M[(1+\alpha)Ax^{k-1} + (1+\beta)Bz^{k-1} - \alpha A\tilde{x}^k - \beta B\tilde{z}^k]).\end{aligned} \tag{67}$$

Summing over $k = 1, \cdots, N$, we obtain

$$\begin{aligned}&\frac{1}{N}\sum_{k=1}^{N}(w^k - w^{k-1}) = \\ &= \rho(\overline{x}_N - \hat{x}_N,\, \overline{z}_N - \hat{z}_N,\, M[(1+\alpha)A\hat{x}_N + (1+\beta)B\hat{z}_N - \alpha A\overline{x}_N - \beta B\overline{z}_N]).\end{aligned}$$

Then from (57), we get

$$\frac{1}{\rho^2}\left\|\frac{1}{N}\sum_{k=1}^{N}(w^k - w^{k-1})\right\|_P^2 = \|(\overline{x}_N - \hat{x}_N, \overline{z}_N - \hat{z}_N)\|_U^2 + \|A\overline{x}_N + B\overline{z}_N\|_M^2$$

and hence, given $w^* \in \operatorname{sol}(V_L)$, we deduce from (63) the ergodic rate of convergence for constraint violations.

Using (67), from (57), we get

$$\frac{1}{\rho^2}\|w^k - w^{k-1}\|_P^2 = \|(\tilde{x}^k - x^{k-1}, \tilde{z}^k - z^{k-1})\|_U^2 + \|A\tilde{x}^k + B\tilde{z}^k\|_M^2$$

and hence, from (62), we deduce the nonergodic rate of convergence for constraint violations. □

We note that the particular case $\gamma = 1$, $V_1 = 0$ and $V_2 = 0$, which implies that $U = 0$, the two terms $\|(\overline{x}_k - \hat{x}_k, \overline{z}_k - \hat{z}_k)\|_U^2$ and $\|(\tilde{x}^k - x^{k-1}, \tilde{z}^k - z^{k-1})\|_U^2$ of inequalities in Theorem 5.3 are null and hence we recover the Theorem 15 of [9].

Remark 5.4. From Proposition 4.1, we deduce that G_S^T is the classical resolvent of $(ST^{-1}S^t)^{-1}$ and hence from Rockafellar [21] we can obtain the linear convergence of $w^{k+1} = G_S^T(w^k)$ if $ST^{-1}S^t$ is Lipschitz continuos at 0, i.e., satisfying

$$\|z - z^*\| \leq a\|w\| \text{ whenever } z \in ST^{-1}S^t(w) \text{ and } \|w\| \leq \tau,$$

for some two positive parameters a and τ.

In particular if T is α-strong monotone then $ST^{-1}S^t$ is $\frac{\|S\|^2}{\alpha}$-Lipschitz and hence the linear convergence of $\|Sx^k\|$ is deduced for x^k defined in (8) with $\rho = 1$.

Considering $\rho = 1$ $\gamma = 1$, $V_2 = 0$ and $B = I$ in GSS algorithm, we have that using the map defined in (3), this algorithm can be deduced from

$$(x^{k+1}, u^{k+1}) = J_{\overline{P}}^{\overline{L}}(x^k, u^k), \quad \text{with} \quad \overline{P} = \begin{pmatrix} V_1 + A^tMA & A^t \\ A & M^{-1} \end{pmatrix}.$$

Therefore if f is strongly convex and g Lipschitz cntinuous then $\overline{L}$ is strongly monotone and hence from the previous paragraph discussion, we deduce the linear convergence of (x^k, u^k). Alternately this convergence result can be deduced from Remark 4.5, where we show that $G_{S_2}^L$ coincides with the Douglas-Rachford map and so apply the convergencia result given by D. Davis [8]. □

6. Application to multi-block optimization problems

To conclude our study, we consider the application of the generalized scheme GSS to the decomposition of some block structured convex optimization problems.

For $i \in \{1, ..., q\}$, let $f_i \colon \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ and $g \colon \mathbb{R}^m \to \overline{\mathbb{R}}$ are proper lsc convex functions, A_i and B matrices of order $p \times n_i$ and $p \times m$, respectively. We consider the following S-Model problem :

$$\inf_{(x_1, \cdots, x_q, z)} \sum_{i=1}^{q} f_i(x_i) + g(z) \quad \text{s.t} \quad \sum_{i=1}^{q} A_i x_i + Bz = 0.$$

This problem has been analyzed by many authors (see [13] for instance). We rewrite it into two different forms, (B_1) and (B_2), but with the same structure as (BP) defined below, then we rewrite (BP) as problem $(\overline{P})$ also defined below. Finally, we apply the algorithm (37)–(40) to this last problem.

The S- Model problem is equivalent to

$$\inf_{(x_1, \cdots, x_q, z)} \sum_{i=1}^{q} f_i(x_i) + g(z) + \delta_{\{0\}}(\sum_{i=1}^{q} A_i x_i + Bz). \tag{B_1}$$

In this formulation the function g can be viewed as a function f_i. The associated dual problem of (B_1) is

$$\inf_{y^*} \sum_{i=1}^{q} (f_i^* \circ A_i^t) y^* + (g^* \circ B^t) y^*. \tag{Ds}$$

Now, by considering $n = \sum_{i=1}^{q} n_i$ and $f \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ defined as $f(x) := \sum_{i=1}^{q} f_i(x_i)$, the problem (Ds) can be written as

$$\inf_{y^*} \; f^* \circ \begin{bmatrix} A_1^t \\ \vdots \\ A_q^t \end{bmatrix} y^* + (g^* \circ B^t) y^*.$$

This is a composite problem whose associated dual problem is

$$\inf_{(x_1, \cdots, x_q)} \sum_{i=1}^{q} f_i(x_i) + (g^* \circ B^t)^* \circ (-\sum_{i=1}^{q} A_i x_i). \tag{B_2}$$

We observe that in this last problem we reduce the number of variables considered in the S-Model problem and function g acts now as regularization function.

Using the same notation as before, we define a problem having the same structures as problems (B_1) and (B_2) :

$$V_P = \inf_{(x_1, \cdots, x_q)} \sum_{i=1}^{q} f_i(x_i) + (g^* \circ B^t)^* \circ (\sum_{i=1}^{q} A_i x_i). \tag{BP}$$

In order to apply the splitting algorithm to problem (BP), we reformulate it to an appropriate optimization problem. To do it, consider

$$K := (I_{p \times p} \cdots I_{p \times p}) \in \mathbb{R}^{p \times pq} \quad \text{and} \quad A := \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_q \end{pmatrix} \in \mathbb{R}^{pq \times n}.$$

So, problem (BP) can be formulated as

$$\inf_{x\in\mathbb{R}^n,\, z\in\mathbb{R}^{pq}} \quad \left[f(x) + (g^* \circ B^t)^* \circ Kz : Ax - z = 0\right]. \qquad (\overline{P})$$

Notice that this last formulation problem have a good separable structure.

We apply to problem $(\overline{P})$ the algorithm (37)-(40) developed in Subsection 3.2.2. We assume that g verifies the following identity

$$\partial[(g^* \circ B^t)^* \circ K] = K^t(B(\partial g)^{-1}B^t)^{-1}K.$$

The saddle-point problem of $(\overline{P})$ is

$$\text{Find } (\bar{x}, \bar{z}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^{pq} \times \mathbb{R}^{pq} \text{ such that } \; 0 \in \bar{L}(\bar{x}, \bar{z}, \bar{y}) \qquad (V_{\bar{L}})$$

where $\bar{L}$ is the maximal monotone map defined on $\mathbb{R}^n \times \mathbb{R}^{pq} \times \mathbb{R}^{pq}$ as

$$\bar{L}(x, z, y) := \begin{pmatrix} \partial f(x) \\ K^t(B(\partial g)^{-1}B^t)^{-1}Kz \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & I \\ -A & I & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}.$$

For $i \in \{1, ..., q\}$, let M_i be an $p \times p$ symmetric positive definite matrix and Q_i be an $n_i \times n_i$ symmetric positive semi-definite matrix.

In order to take advantage of the separability of f, we take $V_1 = diag([Q_1, ..., Q_q])$ and $M = diag([M_1, ..., M_q])$, and we consider $V_2 = 0$ in order to calculate z^{k+1} using alone the resolvent of ∂g. So, the related algorithm (37)-(40) take the following structure:

$$\tilde{z}^{k+1} = (K^t(B(\partial g)^{-1}B^t)^{-1}K + M)^{-1}(MAx^k + y^k) \qquad (68)$$

$$\tilde{y}^{k+1} = y^k + M(Ax^k - \tilde{z}^{k+1}) \qquad (69)$$

$$\tilde{x}^{k+1} = (\partial f + A^tMA + V_1)^{-1}(V_1x^k + A^tM\tilde{z}^{k+1} - A^t\tilde{y}^{k+1}) \qquad (70)$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, z^k, y^k) \qquad (71)$$

Because the diagonal structure of expression (70) the calculation of $\tilde{x}^{k+1}$ is realized in parallel: for $i \in \{1, \cdots, q\}$,

$$\tilde{x}_i^{k+1} = (\partial f_i + A_i^tM_iA_i + Q_i)^{-1}(Q_ix_i^k + A_i^tM_i\tilde{z}_i^{k+1} - A_i^t\tilde{y}_i^{k+1}).$$

Now, in order to calculate $\tilde{z}^{k+1}$, the following identity is relevant:

Proposition 6.1. *With the same notations as before, the following identity holds:*

$$(K^t(B(\partial g)^{-1}B^t)^{-1}K + M)^{-1}M = I - M^{-1}K^t\Sigma(I - B(\partial g + B^t\Sigma B)^{-1}B^t\Sigma)K$$

where Σ is a $p \times p$ matrix defined by

$$\Sigma := (KM^{-1}K^t)^{-1} = \left(\sum_{i=1}^{q} M_i^{-1}\right)^{-1}.$$

Proof. From Proposition 4.1, we have

$$(K^t(B(\partial g)^{-1}B^t)^{-1}K + M)^{-1}M = I - M^{-1}K^t(B(\partial g)^{-1}B^t + KM^{-1}K^t)^{-1}K$$

and hence by combining it with the following identity

$$(B(\partial g)^{-1}B^t + \Sigma^{-1})^{-1}\Sigma^{-1} = I - \Sigma B(\partial g + B^t\Sigma B)^{-1}B^t$$

obtained also from Proposition 4.1, we get the desired identity. □

So, using the identity of this last proposition, we can obtain an equivalent expression of $\tilde{y}^{k+1}$ in (69) but with a more tractable expression for computational purpose:

$$\tilde{y}^{k+1} = K^t\Sigma(I - B(\partial g + B^t\Sigma B)^{-1}B^t\Sigma)K(Ax^k + M^{-1}y^k). \tag{72}$$

It follows in particular that $\tilde{y}^{k+1} \in \operatorname{range} K^t$ and, by considering $y^k \in \operatorname{range} K^t$ in (71), we have that $y^{k+1} \in \operatorname{range} K^t$ and hence all the block components of $\tilde{y}^{k+1}$ (similarly of y^{k+1}) are equal. We denote by $\tilde{y}_c^{k+1}$ (resp y_c^{k+1}) such a block component of $\tilde{y}^{k+1}$ (resp y^{k+1}). Then,

$$\tilde{y}_c^{k+1} = \Sigma(I - B(\partial g + B^t\Sigma B)^{-1}B^t\Sigma)K(Ax^k + M^{-1}K^t y_c^k) \tag{73}$$

By denoting $\zeta^{k+1} := (\partial g + B^t\Sigma B)^{-1}B^t(\Sigma\sum_{j=1}^{q}(A_jx_j^k) + y_c^k)$ we obtain, from (73),

$$\tilde{y}_c^{k+1} = y_c^k + \Sigma(\sum_{j=1}^{q}(A_jx_j^k) - B\zeta^{k+1}). \tag{74}$$

On the other hand, from (69), we get

$$\tilde{z}^{k+1} = Ax^k + M^{-1}K^t(y_c^k - \tilde{y}_c^{k+1})$$

which combining with (74), we deduce that for $i \in \{1, \cdots, q\}$,

$$\tilde{z}_i^{k+1} = A_ix_i^k - M_i^{-1}\Sigma\left(\sum_{j=1}^{q}(A_jx_j^k) - B\zeta^{k+1}\right).$$

Therefore we obtain the following algorithm, called *"Proximal Multi-Block Algorithm"*.

Proximal Multi-block Algorithm (PMA)

For $i \in \{1, \cdots, q\}$ set $Q_i \in \mathbb{R}^{n_i\times n_i}$ symmetric positive semi-definite, $M_i \in \mathbb{R}^{p\times p}$ symmetric positive definite. Set $\Sigma = \left(\sum_{i=1}^{q} M_i^{-1}\right)^{-1}$. Then for an arbitrary $(x^0, z^0, y_c^0) \in \mathbb{R}^n \times \mathbb{R}^{pq} \times \mathbb{R}^p$

Step 1. Find ζ^{k+1} such that

$$\zeta^{k+1} = \operatorname{argmin}\left\{g(w) + \frac{1}{2}\|Bw - \sum_{j=1}^{q}(A_jx_j^k) - \Sigma^{-1}y_c^k\|_\Sigma^2\right\}$$

Step 2. Find $\tilde{z}^{k+1}$

For all $i \in \{1, ..., q\}$ **do**

Find $\tilde{z}_i^{k+1}$ such that

$$\tilde{z}_i^{k+1} = A_i x_i^k - M_i^{-1}\Sigma\left(\sum_{j=1}^{q}(A_j x_j^{k+1}) - B\zeta^{k+1}\right).$$

Step 3. Find $\tilde{y}_c^{k+1}$ such that

$$\tilde{y}_c^{k+1} = y_c^k + \Sigma\left(\sum_{j=1}^{q}(A_j x_j^k) - B\zeta^{k+1}\right).$$

end for

Step 4. Find $\tilde{x}^{k+1}$

For all $i \in \{1, ..., q\}$ **do**

Find $\tilde{x}_i^{k+1}$ such that

$$\tilde{x}_i^{k+1} = \text{argmin}\left\{f_i(x_i) + \tfrac{1}{2}\|A_i x_i - \tilde{z}_i^{k+1} + M_i^{-1}\tilde{y}_c^{k+1}\|_{M_i}^2 + \tfrac{1}{2}\|x_i - x_i^k\|_{Q_i}^2\right\}$$

end for

Step 5. Find $(x^{k+1}, z^{k+1}, y_c^{k+1})$

$$(x^{k+1}, z^{k+1}, y_c^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}_c^{k+1}) + (1-\rho)(x^k, z^k, y_c^k).$$

The next proposition gives conditions in order to guarantee the convergence of PMA. The proof is a direct consequence of Proposition 3.5.

Proposition 6.2. *Let $\rho \in (0, 2)$. For $i \in \{1, ..., q\}$, assume that $Q_i \in \mathbb{R}^{n_i \times n_i}$ and $M_i \in \mathbb{R}^{p \times p}$ are symmetric, with Q_i positive semi-definite and M_i positive definite such that $Q_i + A_i^t M_i A_i$ is positive definite. If* $\text{sol}\,(V_{\bar{L}})$ *is nonempty, then for an arbitrary $(x^0, z^0, y_c^0) \in \mathbb{R}^n \times \mathbb{R}^{pq} \times \mathbb{R}^p$, the sequence $(x^k, z^k, K^t y_c^k)$ generated by* (PMA) *converges to some element of* $\text{sol}\,(V_{\bar{L}})$.

The *Separable Augmented Lagrangian Algorithm (SALA)* with multidimensional scaling has been proposed in [10] to solve a special case of the S-Model where $g = 0$ and $B = 0$. This algorithm can be recovered if instead of applying the algorithm (37)–(40) to problem $(\overline{P})$, we consider the algorithm (41)–(43) with $V_1 = V_2 = 0$. Therefore SALA is a particular version of (PMA).

The advantages of (PMA) are twofold: 1) the inclusion of the relaxing term $\rho \in (0, 2)$, which enables the accelaration of the algorithm, and 2) the additional proximal term $\|x_i - x_i^k\|_{Q_i}^2$ considered in the sub problems of Step 4, which improves the strong convexity of the proximal sub problem when we choose an adequate matrix Q_i. More specifically, considering σ_i and τ_i positive numbers holding $\sigma_i \tau_i \|A_i\|^2 \leq 1$ and choosing M_i and Q_i matrices defined as

$$M_i = \sigma_i I_{p \times p} \text{ and } Q_i = \tau_i^{-1} I_{n_i \times n_i} - \sigma_i A_i^t A_i,$$

the conditions about matrices Q_i, M_i and $Q_i + A_i^t M_i A_i$ in Proposition 6.2 are verified and hence the subproblem in Step 4 of the Algorithm (PMA) becomes

$$\tilde{x}_i^{k+1} = \text{argmin}\left\{f_i(x_i) + \tfrac{1}{2\tau_i}\|x_i - x_i^k - \tau_i[\sigma_i A_i^t \tilde{z}_i^{k+1} - \sigma_i A_i^t A_i x_i^k - A_i^t \bar{y}^{k+1}]\|^2\right\}$$

which has an explicit solution in some particular cases, for instance $f_i(x_i) = \|x_i\|_1$.

7. Conclusions

We have introduced a generalized splitting scheme for monotone composite inclusions involving the sum of two monotone operators. It is based on a generalization of the Proximal iteration allowing positive semidefinite matrices in the resolvent operator. The new scheme includes different scaling parameters and induces separable augmented lagrangian subproblems which are themselves regularized in the primal and dual variables. It can be seen as an extension of Chambolle and Pock's saddle-point splitting scheme and it allows to recover different splitting algorithms studied in the literature. The rate of convergence is in $O(1/k)$ in the ergodic sense which is indeed the expected rate for that family of methods. Further investigation is ongoing to consider the possibility to obtain linear convergence rates in the strongly convex case and the challenge remains, as for all these theoretical algorithms, to be able to tune the parameters and allow approximate solutions in the subproblems to design practical and efficient algorithms.

We have presented the application of the generalized splitting scheme to block-structured convex optimization problems. It can be applied to a wide variety of models with any number of blocks and linear but not necessarily full-rank coupling constraints. It remains to show how to adapt the generalized splitting to composite models with very large scale linear operators like the ones studied in Statistical Learning (see for example [4]).

Acknowledgement. The work of E. Ore Albornoz was partially supported by a FONDECYT grant number 217-2014 and a PGMO (Programme Gaspard Monge en Optimisation) grant number 2014-0020H.

References

[1] M. A. Alghamdi, A. Alotaibi, P. L. Combettes, N. Shahzad: *A primal-dual method of partial inverses for composite inclusions*, Optimization Letters 8 (2014) 2271–2284.

[2] H. Attouch, M. Soueycatt: *Augmented Lagrangian and proximal alternating direction method of multipliers in Hilbert spaces. Applications to games, PDE's and control*, Pacific J. Optimization 5 (2008) 17–37.

[3] A. Beck, M. Teboulle: *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. 2 (2009) 183–202.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein: *Distributed optimization and statistical learning with the alternating direction method of multipliers*, Foundations and Trends in Machine Learning 3 (2011) 1–122.

[5] L. M. Briceno-Arias, P. L. Combettes: *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM J. Optimization 21 (2011) 1230–1250.

[6] A. Chambolle, T. Pock: *A first-order primal-dual algorithm for convex programs with applications to imaging*, J. Math. Imaging Vision 40 (2011) 1–26.

[7] L. Condat: *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optimization Theory Appl. 158 (2013) 460–479.

[8] D. Davis: *Convergence rate analysis of forward-Douglas-Rachford splitting scheme*, arXiv:1410.2654 (2014).

[9] D. Davis, W. Yin: *Convergence rate analysis of several splitting schemes*, in: *Splitting Methods in Communication, Imaging, Science and Engineering*, R. Glowinski, S. J. Osher and W. Yin (eds.), Springer, Cham (2016) 115–163.

[10] J. P. Dussault, O. M. Gueye, P. Mahey: *Separable augmented Lagrangian algorithm with multidimensional scaling for monotropic programming*, J. Optimization Theory Appl. 127 (2005) 329–345.

[11] J. Eckstein, D. P. Bertsekas: *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Programming 55 (1992) 293–318.

[12] D. Gabay: *Applications of the method of multipliers to variational inequalities*, in: *Augmented Lagrangian Methods: Application to Numerical Solutions of Boundary-Value Problems*, M. Fortin, R. Glowinski (eds.), Studies in Mathematics and its Applications 15, North-Holland, Amsterdam (1983) 299–331.

[13] A. Lenoir, P. Mahey: *A survey of monotone operator splitting methods for the decomposition of convex programs*, RAIRO Oper. Res. 51 (2017) 17–41.

[14] P. L. Lions, B. Mercier: *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Analysis 16 (1979) 964–979.

[15] P. Mahey, J. P. Dussault, A. Benchakroun, A. Hamdi: *Adaptive scaling and convergence rates of a separable augmented Lagrangian algorithm*, in: *Optimization*, Lecture Notes in Economics and Mathematical Systems 481, V. H. Nguyen, J. J. Strodiot, P. Tossings (eds.), Springer, Berlin (2000) 278–287.

[16] P. Mahey, S. Oualibouch, D. T. Pham: *Proximal decomposition on the graph of a maximal monotone operator*, SIAM J. Optimization 5 (1995) 454–466.

[17] J. J. Moreau: *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. Française 93 (1965) 273–299.

[18] Y. Nesterov: *Smooth minimization of nonsmooth functions*, Math. Programming, Series A 103 (2005) 127–152.

[19] T. Pennanen: *A splitting method for composite mappings*, Num. Functional Analysis Optimization 23 (2002) 875–890.

[20] R. T. Rockafellar: *Augmented lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res. 1 (1976) 97–116.

[21] R. T. Rockafellar: *Monotone operators and the proximal point algorithm*, SIAM J. Control Optimization 14 (1976) 877–898.

[22] R. Shefi, M. Teboulle: *Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization*, SIAM J. Optimization 24 (2014) 269–297.

[23] J. E. Spingarn: *Partial inverse of a monotone operator*, Appl. Math. Optimization 10 (1983) 247-265.

[24] J. E. Spingarn: *Applications of the method of partial inverses to convex programming:decomposition*, Math. Programming 32 (1985) 199–223.

[25] R. S. Varga: *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, (1966).