

# Inertial-relaxed splitting for composite monotone inclusions

E. Oré-Albornoz<sup>\*</sup>      P. Mahey<sup>†</sup>      E. Ocaña-Anaya<sup>\*</sup>

October 16, 2021

## Abstract

In a similar spirit of the extension of the proximal point method developed by Alves et al. [2], we propose in this work an Inertial-Relaxed primal-dual splitting method to address the problem of decomposing the minimization of the sum of three convex functions, one of them being smooth, and considering a general coupling subspace. A unified setting is formalized and applied to different average maps whose corresponding fixed points are related to the solutions of the inclusion problem associated with our extended model. An interesting feature of the resulting algorithms we have designed is that they present two distinct versions with a Gauss-Seidel or a Jacobi flavour, extending in that sense former proximal ADMM methods, both including inertial and relaxation parameters.

Finally we show computational experiments on a class of the fused LASSO instances of medium size.

## 1 Introduction

We will propose in this paper new versions of existing splitting methods for the following general convex minimization model involving the sum of three convex functions, one of them being smooth and the another ones composed with a linear operator :

$$\begin{aligned} \text{Minimize} \quad & f(x) + g(z) + h(x) \\ & Ax + Bz = 0 \end{aligned} \tag{1}$$

where  $f : \mathbb{R}^n \mapsto \mathbb{R}$  and  $g : \mathbb{R}^p \mapsto \mathbb{R}$  are proper convex lsc functions,  $A$  and  $B$  are  $(m \times n)$  and  $(m \times p)$  matrices, respectively, and  $h : \mathbb{R}^n \mapsto \mathbb{R}$  is convex and  $(\frac{1}{\beta})$ -Lipschitz-differentiable, with  $\beta$  a real positive.

---

<sup>\*</sup>IMCA, Instituto de Matemática y Ciencias Afines, Universidad Nacional de Ingeniería, Lima, Perú

<sup>†</sup>LIMOS, CNRS & Université Clermont Auvergne, France

That model has received a lot of attention recently, most of the work aiming at extending known splitting schemes adapted to the two functions case where  $h = 0$ . Here again, we will explore the corresponding splitting issues, thus designing algorithms which involve forward or backward steps associated with each function separately.

The celebrated Alternate Direction Method of Multipliers (*ADMM*) is one of the most important first order splitting method to solve (1) when  $h = 0$  (see [14] for the original introduction or [5] for a survey with potential applications to signal processing). It can be seen as a dual approach based on the composite Augmented Lagrangian function where the dual multipliers are denoted by  $y \in \mathbb{R}^m$  :

$$l(x, z, y) = f(x) + g(z) + \frac{1}{2} \|Ax + Bz + M^{-1}y\|_M^2$$

where  $M$  is a positive definite  $(m \times m)$  symmetric matrix and  $\|a\|_M^2 = a^T M a$  the corresponding elliptic norm. ADMM basically consists in alternating the minimization of that Lagrangian w.r.t.  $x$  and  $z$  separately, followed by the dual update of the  $y$  variables.

Many variants of ADMM have been developed, unified variants of ADMM are condensed in the Shefi-Teboulle's algorithms [22] in the case ( $B = -I_{p \times p}$ ) where additional primal proximal terms are added in the respective Lagrangian functions like in the Proximal Method of Multipliers [21]. These can be easily extended to the general case with any matrix  $B$  and called hereafter *Proximal Primal-Dual Splitting (PPDS)* with two distinct interpretations as a Gauss-Seidel and a Jacobi-like versions.

Proximal primal-dual Algorithm (PPDS)

$$\begin{cases} x^{k+1} & \in \operatorname{argmin} \left\{ f(x) + \frac{1}{2} \|Ax + Bz^k + M^{-1}y^k\|_M^2 + \frac{1}{2} \|x - x^k\|_{V_1}^2 \right\} \\ z^{k+1} & \in \operatorname{argmin} \left\{ g(z) + \frac{1}{2} \|A\eta^k + Bz + M^{-1}y^k\|_M^2 + \frac{1}{2} \|z - z^k\|_{V_2}^2 \right\} \\ y^{k+1} & = y^k + M(Ax^{k+1} + Bz^{k+1}) \end{cases}$$

where choosing  $\eta^k$  as below gives us two algorithmic versions:

$$\eta^k := \begin{cases} x^k & \text{for Jacobi version algorithm} \\ x^{k+1} & \text{for Gauss-Seidel version algorithm} \end{cases}$$

where  $V_1$  and  $V_2$  are chosen such that  $V_1$  and  $V_2$  are positive semi-definite matrices for the Gauss-Seidel version and,  $V_1 - A^t M A$  and  $V_2 - B^t M B$  are positive semi-definite matrices for the Jacobi version.

We note that both algorithms result from applying the preconditioned proximal points (corresponding to two appropriate positive semidefinite matrices, see [19]) to the following Lagrangian inclusion problem associated with (1) (case  $h = 0$ ) :

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

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

$$L(x, z, y) := \begin{pmatrix} \partial f(x) \\ \partial g(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}. \quad (2)$$

The convergence of the previous algorithm is obtained from its relationship with the fixed point formulation applied to particular 1/2-averaged maps<sup>1</sup> (each version corresponding to different averaged map), in the same way as ADMM is related to the Douglas-Rachford map (see [13] for instance).

We will use the same strategy below to propose new and generalized splitting algorithms by inspecting averaged maps associated with model (1).

When an smooth part is added to the model, represented by a function  $h$ , the aim is to further improve these algorithms by inserting forward gradient steps without destroying the splitting strategy. Condat [9] (and independently Vũ [23]), has developed two forms of algorithms considering two different *Primal-Dual Forward-Backward Splitting*, *PDFB*, whose corresponding Lagrangian maps have less variables than the map  $L$  defined by (2).

One of these algorithms is (considering here the simplified formulation with  $B = -I_{p \times p}$ ) :

Condat-Vũ Algorithm, Form I

$$\begin{cases} \tilde{x}^{k+1} &= (\tau \partial f + I_{n \times n})^{-1}(x^k - \tau \nabla h(x^k) - \tau A^t y^k) \\ \tilde{y}^{k+1} &= (\sigma \partial g^* + I_{m \times m})^{-1}(y^k + \sigma A(2\tilde{x}^{k+1} - x^k)) \\ (x^{k+1}, y^{k+1}) &= \rho^k(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1 - \rho^k)(x^k, y^k) \end{cases}$$

The other one switches the roles of primal and dual variables :

Condat-Vũ Algorithm, Form II

$$\begin{cases} \tilde{y}^{k+1} &= (\sigma \partial g^* + I_{m \times m})^{-1}(y^k + \sigma A x^k) \\ \tilde{x}^{k+1} &= (\tau \partial f + I_{n \times n})^{-1}(x^k - \tau \nabla h(x^k) - \tau A^t(2\tilde{y}^{k+1} - y^k)) \\ (x^{k+1}, y^{k+1}) &= \rho^k(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1 - \rho^k)(x^k, y^k) \end{cases}$$

Both algorithms include the relaxation parameter  $\rho^k$ , which is known to accelerate convergence for values in  $(1, 2)$  (see [10]).

On the other hand, and without the relaxation effect ( $\rho^k = 1$ ), Chambolle and Pock [8] have proposed a *Primal-Dual Splitting with Inertial step (IPDS)* method, showing to be closely related to Condat-Vũ's algorithm, Form I, but with an *inertial*

---

<sup>1</sup>A mapping  $G : X \mapsto X$  is said to be  $\alpha$  - averaged (for some  $\alpha \in (0, 1)$ ) if there exists a nonexpansive map  $N$  such that  $G = \alpha I + (1 - \alpha)N$  see [4] for more details.

accelerating term as following :

Inertial Chambolle-Pock Primal-Dual Algorithm, IPDS

$$\begin{cases} (x_w^k, y_w^k) &= (x^k, y^k) + \lambda^k(x^k - x^{k-1}, y^k - y^{k-1}) \\ x^{k+1} &= (\tau \partial f + I_{n \times n})^{-1}(x_w^k - \tau \nabla h(x_w^k) - \tau A^t y_w^k) \\ y^{k+1} &= (\sigma \partial g^* + I_{m \times m})^{-1}(y_w^k + \sigma A(2x^{k+1} - x_w^k)) \end{cases}$$

The inertial parameter  $\lambda^k$  has indeed a different effect than the relaxation strategy; it has been introduced in [1] and contains many similar features of Güler's accelerated Proximal Point algorithm [15], the latter being related with early Nesterov's optimal gradient methods for convex minimization [18].

One of the feature of the generalized primal-dual splitting proposed in this paper is the inclusion of both the relaxation ( $\rho^k$ ) and inertial ( $\lambda^k$ ) parameters in the primal updates, thus inspired by Alves et al [2], where the authors consider a Relative-error Inertial-Relaxed variant of Proximal Points to produce variants of ADMM. Analogously, we will consider Inertial-Relaxed variants of fixed point formulations applied to different averaged maps.

When  $h = 0$ , Condat-Vũ and Chambolle-Pock algorithms without relaxed or inertial terms can be deduced from *PPDS* in Gauss-Seidel version (see [22]). More exactly: from the Gauss-Seidel version of *PPDS* (in the case  $B = -I_{p \times p}$ ) and considering  $M = \sigma I_{m \times m}$ ,  $V_1 = \tau^{-1} I_{n \times n} - \sigma A^t A$ ,  $V_2 = 0$ , then, after a change of variables  $(\tilde{x}^k, \tilde{z}^k, \tilde{y}^k) = (x^{k+1}, z^k, y^k)$ , we can reobtain Condat-Vũ algorithm, form II. Analogously, interchanging the order in the model problem, we obtain the switched version: Condat-Vũ algorithm, form I.

Alternatively, another class of splitting algorithms for (1) (in the case  $B = -I_{p \times p}$ , but with the three functions) called Primal-Dual Three Operator (*PD3O*) has been analyzed by Yan [24], extending a former work by Davis and Yin [12] who supposed  $A = I_{n \times n}$ .

PD3O Algorithm

$$\begin{cases} x^k &= (\tau \partial f + I_{n \times n})^{-1}(z^k) \\ y^{k+1} &= (\sigma \partial g^* + I_{m \times m})^{-1}((I_{m \times m} - \tau \sigma A A^t)y^k - \sigma A(2x^k - z^k - \tau \nabla h(x^k))) \\ z^{k+1} &= x^k + \tau \nabla h(x^k) - \tau A^t y^{k+1} \end{cases}$$

Observe that in the case  $h = 0$  and after the change of variables  $(\tilde{x}^k, \tilde{y}^k) = (x^k, y^{k+1})$ , *PD3O* gets back to Condat-Vũ algorithm, form I, and is thus again a consequence of the *PPDS* scheme. In the general case, Yan [24] showed that *PD3O* has a broader domain of convergence and a better numerical behavior compared to Condat-Vũ's algorithm. So we consider the extension of *PD3O* (instead of Condat-Vũ's algorithm) for the general model (1). The extended algorithm that will be developed includes in particular the switched version of *PD3O* (similar to Condat-Vũ's Algorithm, Form II) and its parallel version (similar to *PPDS* Jacobi version).

Davis-Yin's 3 operator splitting [12] has been recently improved in [20] who proposed an adaptive stepsize tuning to compensate the difficulty to estimate the Lipschitz constant. In [7, 6], the authors consider an extension of Spingarn's Partial Inverse method to the 3 functions model coupled by a subspace constraint. Quite recently, a more intricate model with 4 operators is analyzed in [3] and inexact computations are allowed.

We should cite too [16] where the authors developed another class of splitting methods for a more general model including (1) that extends *PPDS* in the Gauss-Seidel version. Finally, more composite models and different extensions of Chambolle-Pock and Condat-Vũ's schemes can be found in the recent survey by Condat et al [10].

In summary, associated with the extended model (1), we will construct first order splitting algorithms that unified *PD3O* (getting a switching and parallel version) and the *PPDS* algorithms (inserting forward gradient steps), including in all of them relaxed and inertial parameters. To achieve that goal first, in section 2, we will construct two types of averaged maps associated with our extended sequential and parallel splitting algorithms, respectively. Then in order to include Inertial and relaxation parameters, in section 3 we rewrite an Inertial-Relaxed variant of the corresponding fixed point applied to averaged maps. In the next section 4, applying these variants of fixed point to the averaged maps constructed in section 2, we obtain the desirable general splitting algorithm that includes Inertial-Relaxed terms. In section 5, we choose special multidimensional scaling matrices parameters to better tune the general algorithm obtained before, in order to show the equivalence with the existing algorithms. Finally in the last section, a limited set of computational comparisons between the algorithms will be presented.

## 2 Deriving three candidate averaged maps

Assimilating the sum of  $f$  and  $h$  as a unique function in model (1), under regularity conditions, the problem is equivalent to the following saddle-point inclusion problem

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

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

$$\mathcal{L}(x, z, y) := \begin{pmatrix} \partial f(x) + \nabla h(x) \\ \partial g(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}.$$

The difficulty here is the necessity to split the regularization steps between  $f, h, g$  and the composite effect of matrices  $A$  and  $B$ . A direct application of the approach studied in [19] does not solve the difficulty, since for any matrix  $D$ , the map  $G_D^{\mathcal{L}} = D(\mathcal{L} + D^t D)^{-1} D^t$  which is  $\frac{1}{2}$ -averaged, does not separate the map  $\nabla h(x)$ .

Alternatively we obtain a equivalent problem of  $(V_{\mathcal{L}})$  to be able to apply it the Davis-Yin  $\alpha$ -averaged map [11], where  $\alpha \in (\frac{1}{2}, 1)$ , getting two distinct  $\alpha$ -averaged maps which provide the complete splitting even when  $\nabla h(x) \neq 0$ . These new maps are variants of  $G_D^{\mathcal{L}}$ , choosing  $D$  as the matrices  $\bar{Q}$  and  $\hat{Q}$  defined below in (4), respectively, which, in the case of  $\nabla h(x) = 0$ , are respectively equal (see [19]). On the other hand, regardless the Davis-Yin map, we obtained in (9),  $\hat{S}\mathcal{G}\hat{S}^t$  another class of  $\alpha$ -averaged map with a parallel splitting structure, which is a variant of  $G_D^{\mathcal{L}}$ , with  $D$  equal to  $\hat{S}$ , considering the map  $\mathcal{G}$  (that solve (10)) instead to  $(\mathcal{L} + \hat{S}^t\hat{S})^{-1}$ .

## 2.1 Two averaged maps related to the Davis-Yin map

Following the strategy given by Davis and Yin [11] to identify a single averaged map associated with an explicit 3-operators inclusion, we will now reformulate  $(V_{\mathcal{L}})$  above as a single inclusion with three operators, as seen in the next proposition.

**Proposition 1** *Let  $M$  an  $m \times m$  symmetric positive definite matrix and  $V_1$  and  $V_2$  symmetric positive semidefinite matrices of order  $n \times n$  and  $p \times p$ , respectively, such that  $V_1 + A^t M A$  is positive definite. Using the notation  $w = \begin{bmatrix} x \\ z \\ y \end{bmatrix}$  for the primal-dual space variables belonging to  $\mathbb{R}^{n+p+m}$ , problem  $(V_{\mathcal{L}})$  can be written as*

$$0 \in (\bar{A} S^{-1} \bar{A}^t)^{-1}(w) + [(-\bar{B}) T^{-1} (-\bar{B}^t)]^{-1}(w) + \bar{A} (\bar{A}^t \bar{A})^{-1} C (\bar{A}^t \bar{A})^{-1} \bar{A}^t(w),$$

where

$$\bar{A} := \begin{pmatrix} V_1^{\frac{1}{2}} & 0 \\ 0 & I_{p \times p} \\ M^{\frac{1}{2}} A & 0 \end{pmatrix}, \quad \bar{B} := \begin{pmatrix} -I_{n \times n} & 0 \\ 0 & -V_2^{\frac{1}{2}} \\ 0 & M^{\frac{1}{2}} B \end{pmatrix},$$

and for any  $(x, \xi) \in \mathbb{R}^n \times \mathbb{R}^p$ ,  $(\chi, z) \in \mathbb{R}^n \times \mathbb{R}^p$

$$S(x, \xi) := \begin{pmatrix} \partial f(x) \\ 0 \end{pmatrix}, \quad C(x, \xi) := \begin{pmatrix} \nabla h(x) \\ 0 \end{pmatrix} \quad \text{and} \quad T(\chi, z) := \begin{pmatrix} 0 \\ \partial g(z) \end{pmatrix}.$$

**Proof.** Following the lifting strategy introduced in [12], problem (1) is now lifted, adding the dummy variables  $\xi \in \mathbb{R}^p$  and  $\chi \in \mathbb{R}^n$  and using the notation  $(f_1, f_2)(x_1, x_2) = f_1(x_1) + f_2(x_2)$  and gets the following condensed form :

$$\min_{(x, \xi, \chi, z) \in \mathcal{F}} (f + h, 0)(x, \xi) + (0, g)(\chi, z)$$

where  $\mathcal{F}$  is the set of all  $(x, \xi, \chi, z)$  satisfying

$$\begin{pmatrix} V_1^{\frac{1}{2}} & 0 \\ 0 & I_{p \times p} \\ M^{\frac{1}{2}} A & 0 \end{pmatrix} \begin{pmatrix} x \\ \xi \end{pmatrix} + \begin{pmatrix} -I_{n \times n} & 0 \\ 0 & -V_2^{\frac{1}{2}} \\ 0 & M^{\frac{1}{2}} B \end{pmatrix} \begin{pmatrix} \chi \\ z \end{pmatrix} = 0.$$

Considering the notations given in the hypothesis, it holds that under regularity conditions, the last problem is equivalent to the following inclusion problem (in its dual form), using  $w^* \in \mathbb{R}^{n+p+m}$ ,

$$0 \in (-\overline{A})(S + C)^{-1}(-\overline{A}^t)(w^*) + (-\overline{B})T^{-1}(-\overline{B}^t)(w^*)$$

The key trick now is to observe that the composite inclusion above is a valid dual formulation of a primal inclusion which splits  $S$  and  $C$ , associated with the composite transformation  $\overline{A}Q\overline{A}^t$ , and it gives (now in  $\mathbb{R}^{n+p}$ ) :

$$0 \in S(\hat{w}) + C(\hat{w}) + \overline{A}^t((- \overline{B})T^{-1}(-\overline{B}^t))^{-1}\overline{A}(\hat{w})$$

Now, since  $V_1 + A^tMA$  is invertible, then  $\overline{A}$  is an injective matrix, so we have

$$0 \in S(\hat{w}) + \overline{A}^t [((- \overline{B})T^{-1}(-\overline{B}^t))^{-1} + \overline{A}(\overline{A}^t\overline{A})^{-1}C(\overline{A}^t\overline{A})^{-1}\overline{A}^t] \overline{A}(\hat{w})$$

taking again the dual, but using different dual variables denoted now by  $\hat{w}^*$

$$\begin{aligned} 0 \in & (-I)(\overline{A}S^{-1}\overline{A}^t)(-I)(\hat{w}^*) \\ & + [((- \overline{B})T^{-1}(-\overline{B}^t))^{-1} + \overline{A}(\overline{A}^t\overline{A})^{-1}C(\overline{A}^t\overline{A})^{-1}\overline{A}^t]^{-1}(\hat{w}^*) \end{aligned}$$

It is now straightforward to perform a last dual inclusion associated with the former one to obtain the desired equivalent inclusion problem.  $\blacksquare$

In the previous proposition, the assumption that  $V_1 + A^tMA$  is positive definite (which is equivalent to  $\overline{A}$  being injective) implies that  $(\overline{A}S^{-1}\overline{A}^t)^{-1}$  is maximal monotone and  $\overline{A}(\overline{A}^t\overline{A})^{-1}C(\overline{A}^t\overline{A})^{-1}\overline{A}^t$  is a  $\frac{\beta}{\|(V_1 + A^tMA)^{-1}\|}$ -cocoercive map (since  $h$  is convex and  $(\frac{1}{\beta})$ -Lipschitz-differentiable). Similarly, the remaining map  $[(-\overline{B})T^{-1}(-\overline{B}^t)]^{-1}$  involved in the reformulation of  $(V_{\mathcal{L}})$  as a sum of three maps, is too maximal monotone under the additional assumption that  $V_2 + B^tMB$  is positive definite (equivalent to  $\overline{B}$  injective). Therefore, under both assumptions that  $V_1 + A^tMA$  and  $V_2 + B^tMB$  are positive definite, we can apply the Davis-Yin map (with scalar parameter  $\gamma = 1$ ) to these inclusions and obtain two distinct maps (the other one is obtained switching the position of  $(\overline{A}S^{-1}\overline{A}^t)^{-1}$  and  $[(-\overline{B})T^{-1}(-\overline{B}^t)]^{-1}$ ) related to the principal problem  $(V_{\mathcal{L}})$ . After some calculations, these maps are written as  $G_1$  and  $G_2$ , they apply  $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$  into itself and are defined respectively as

$$G_1(x, z, y) = \begin{pmatrix} V_1^{\frac{1}{2}}(\tilde{x} - \bar{r}) \\ V_2^{\frac{1}{2}}\tilde{z} \\ -M^{\frac{1}{2}}A\tilde{x} - M^{\frac{1}{2}}B\tilde{z} + y \end{pmatrix}, \quad G_2(x, z, y) = \begin{pmatrix} V_1^{\frac{1}{2}}\hat{x} \\ V_2^{\frac{1}{2}}\hat{z} \\ M^{\frac{1}{2}}A\hat{x} + M^{\frac{1}{2}}B\hat{z} + y \end{pmatrix}$$

where for  $G_1$

$$\tilde{x} = (\partial f + V_1 + A^tMA)^{-1} \left( V_1^{\frac{1}{2}}x + A^tM^{\frac{1}{2}}y \right)$$

$$\begin{aligned}\tilde{r} &= (V_1 + A^t M A)^{-1} \nabla h(\tilde{x}) \\ \tilde{z} &= (\partial g + V_2 + B^t M B)^{-1} \left( V_2^{\frac{1}{2}} z + B^t M A(\tilde{r} - 2\tilde{x}) + B^t M^{\frac{1}{2}} y \right).\end{aligned}$$

and for  $G_2$

$$\begin{aligned}\hat{x} &= (\partial f + V_1 + A^t M A)^{-1} \left( V_1^{\frac{1}{2}} x - A^t M^{\frac{1}{2}} (y + 2M^{\frac{1}{2}} B \hat{z}) - \hat{r} \right) \\ \hat{r} &= \nabla h \left( (V_1 + A^t M A)^{-1} (V_1^{\frac{1}{2}} x - A^t M B \hat{z}) \right) \\ \hat{z} &= (\partial g + V_2 + B^t M B)^{-1} \left( V_2^{\frac{1}{2}} z - B^t M^{\frac{1}{2}} y \right).\end{aligned}$$

Note that these maps are switching the order of calling the proximal terms of  $\partial f$  and  $\partial g$ , thus inducing two algorithms where one is the switched form the other, similar to Condat-Vũ's algorithms where they consider two forms (due to the lack of symmetry with respect to the primal and dual variables). The next proposition resumes the averageness properties of  $G_1$  and  $G_2$

**Proposition 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. Considering that  $\|(V_1 + A^t M A)^{-1}\| \in ]0, 2\beta[$ . Then,  $G_1$  and  $G_2$  are  $\alpha$ -averaged with full domain, where*

$$\alpha := \frac{2\beta}{4\beta - \|(V_1 + A^t M A)^{-1}\|} \in ]\frac{1}{2}, 1[.$$

**Proof.** The fullness of the domains of  $G_1$  and  $G_2$  is deduced from the maximality of  $\partial f$  and  $\partial g$  and the fullness of the domain of  $\nabla h$ .

As said before,  $G_1$  is the Davis-Yin map (with parameter  $\gamma = 1$ ) associated with the sum of three operators inclusion problem of Proposition 1, and  $G_2$  its switched version. Therefore it holds that

$$G_1 = I - W_2 + W_1(2W_2 - I - \overline{C} \circ W_2) \quad \text{and} \quad G_2 = I - W_1 + W_2(2W_1 - I - \overline{C} \circ W_1),$$

where  $W_1 := J^{((- \overline{B})T^{-1}(- \overline{B}^t))^{-1}}$  and  $W_2 := J^{(\overline{A} \overline{S}^{-1} \overline{A}^t)^{-1}}$ , are the classic resolvents of operator  $[(- \overline{B})T^{-1}(- \overline{B}^t)]^{-1}$  and  $(\overline{A} \overline{S}^{-1} \overline{A}^t)^{-1}$  respectively, and defining  $\widehat{\beta} := \frac{\beta}{\|(V_1 + A^t M A)^{-1}\|}$ ,  $\overline{C} := \overline{A} (\overline{A}^t \overline{A})^{-1} C (\overline{A}^t \overline{A})^{-1} \overline{A}^t$  is  $\widehat{\beta}$ -cocoercive.

Since by hypothesis  $1 < 2\widehat{\beta}$ , the average properties of  $G_1$  and  $G_2$  follow.  $\blacksquare$

Observe that these new maps  $G_1$  and  $G_2$  can also be seen as a generalization of Davis-Yin maps, since they maintain the averageness and splitting properties (Davis-Yin map can be recovered in the case  $A = I_{n \times n}$  and  $B = -I_{p \times p}$ , considering



$V_1 = 0$ ,  $V_2 = 0$  and  $M = \lambda I_{m \times m}$  which allows to restrict their domain to  $\mathbb{R}^m$ ). The fixed point set of  $G_1$  and  $G_2$ , which are related to  $\text{sol}(V_{\mathcal{L}})$ , are

$$\left\{ \bar{Q} \begin{pmatrix} x^* \\ z^* \\ y^* \end{pmatrix} - \begin{pmatrix} V_1^{\frac{1}{2}}(V_1 + A^t M A)^{-1} \nabla h(x^*) \\ 0 \\ M^{\frac{1}{2}} A (V_1 + A^t M A)^{-1} \nabla h(x^*) \end{pmatrix} : (x^*, z^*, y^*) \in \text{sol}(V_{\mathcal{L}}) \right\}, \quad (3)$$

and

$$\widehat{Q}(\text{sol}(V_{\mathcal{L}})) := \{(V_1^{\frac{1}{2}} x^*, V_2^{\frac{1}{2}} z^*, M^{\frac{1}{2}} A x^* + M^{-\frac{1}{2}} y^*) : (x^*, z^*, y^*) \in \text{sol}(V_{\mathcal{L}})\}, \quad (4)$$

where  $\bar{Q}$  and  $\widehat{Q}$  are the matrices defined as

$$\bar{Q} = \begin{pmatrix} V_1^{\frac{1}{2}} & 0 & 0 \\ 0 & V_2^{\frac{1}{2}} & 0 \\ 0 & -M^{\frac{1}{2}} B & -M^{-\frac{1}{2}} \end{pmatrix} \quad \text{and} \quad \widehat{Q} = \begin{pmatrix} V_1^{\frac{1}{2}} & 0 & 0 \\ 0 & V_2^{\frac{1}{2}} & 0 \\ M^{\frac{1}{2}} A & 0 & M^{-\frac{1}{2}} \end{pmatrix}. \quad (5)$$

These matrices are also involved in the fixed point algorithms derived from  $G_1$  and  $G_2$  and in the corresponding splitting algorithms, as we will show in section 4.

## 2.2 A new averaged map with a parallel structure

We can easily obtain a splitting algorithm in a parallel form directly from the application of known algorithms to a special reformulation of the composite model. For instance, rewriting the principal model (1) as a sum of three blocks, namely :

$$\text{Minimize}_{(x,z)} (f(x) + g(z)) + \delta_{\{(x,z): Ax+Bz=0\}}(x, z) + h(x) \quad (6)$$

and then applying Davis-Yin algorithm, from the separable structure of  $(f(x) + g(z))$ , we obtain a splitting algorithm that considers the calculation of proximal terms on  $f$  and  $g$  in a parallel way, but the iterations also require the computation of the projection over the subspace  $\{(x, z) : Ax + Bz = 0\}$ .

Alternately, since *PD3O* for Jacobi version does not require any implementation of a projection, we consider its extension in order to obtain a parallel algorithm. To that purpose, we will need to construct another averaged map related to our principal model whose proximal step subproblems on  $f$  and  $g$  can also be calculated in parallel ways. We will see too below that the computation of the projection step which breaks the parallel features can be avoided.

Let  $M$  be a  $m \times m$  symmetric positive definite matrix and  $R_1$  and  $R_2$  symmetric positive semidefinite matrices of order  $n \times n$  and  $p \times p$  respectively, such that  $R_1 + 2A^t M A$  and  $R_2 + 2B^t M B$  are positive definite matrices (these hypotheses are denoted below by *Hypo* -  $M$ ).

We consider the map  $G_3$ , that applies  $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m$  into itself, defined as

$$G_3(\hat{x}, \hat{z}, \hat{w}, \hat{y}) = \begin{pmatrix} R_1^{\frac{1}{2}}x \\ R_2^{\frac{1}{2}}z \\ M^{\frac{1}{2}}Ax - M^{\frac{1}{2}}Bz \\ M^{\frac{1}{2}}Ax + M^{\frac{1}{2}}Bz + \hat{y} \end{pmatrix}$$

where

$$\begin{aligned} x &= (\partial f + R_1 + 2A^tMA)^{-1} \left( R_1^{\frac{1}{2}}\hat{x} + A^tM^{\frac{1}{2}}(\hat{w} - \hat{y}) - \bar{r} \right) \\ \bar{r} &= \nabla h \left( (R_1 + 2A^tMA)^{-1}(R_1^{\frac{1}{2}}\hat{x} + A^tM^{\frac{1}{2}}\hat{w}) \right) \\ z &= (\partial g + R_2 + 2B^tMB)^{-1} \left( R_2^{\frac{1}{2}}\hat{z} + B^tM^{\frac{1}{2}}(-\hat{w} - \hat{y}) \right). \end{aligned}$$

Notice that the evaluation of this map at any point just considers the parallel calculations of the subproblems related to  $f$  and  $g$ , and it is not necessary to compute explicitly the projection on the coupling subspace  $\{(x, z) : Ax + Bz = 0\}$ . The fixed points of  $G_3$  are also related to  $\text{sol}(V_{\mathcal{L}})$  and contained in

$$\widehat{S}(\text{sol}(V_{\mathcal{L}})) = \{(R_1^{\frac{1}{2}}x^*, R_2^{\frac{1}{2}}z^*, M^{\frac{1}{2}}Ax^* - M^{\frac{1}{2}}Bz^*, M^{-\frac{1}{2}}y^*) : (x^*, z^*, y^*) \in \text{sol}(V_{\mathcal{L}})\} \quad (7)$$

where  $\widehat{S}$  is a matrix defined as

$$\widehat{S} = \begin{pmatrix} R_1^{\frac{1}{2}} & 0 & 0 \\ 0 & R_2^{\frac{1}{2}} & 0 \\ M^{\frac{1}{2}}A & -M^{\frac{1}{2}}B & 0 \\ 0 & 0 & M^{-\frac{1}{2}} \end{pmatrix}. \quad (8)$$

In what follows, we show that  $G_3$  is an averaged map, beginning with a partial result. Denote by  $\mathcal{G}$  the map that applies  $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$  into itself and is defined by

$$\mathcal{G}(\tilde{x}, \tilde{z}, \tilde{y}) := \begin{pmatrix} x \\ z \\ MAx + MBz + M\tilde{y} \end{pmatrix}$$

where

$$\begin{aligned} x &= (\partial f + R_1 + 2A^tMA)^{-1} (\tilde{x} - A^tM\tilde{y} - \tilde{r}) \\ \tilde{r} &= \nabla h \left( (R_1 + 2A^tMA)^{-1}\tilde{x} \right) \\ z &= (\partial g + R_2 + 2B^tMB)^{-1} (\tilde{z} - B^tM\tilde{y}). \end{aligned}$$

After calculations, this map has the following properties which help us to show later the averaged properties of  $G_3$ .

**Proposition 3** *Under the Hypo – M hypotheses, it holds that*

$$G_3 = \widehat{S}\widehat{\mathcal{G}}\widehat{S}^t. \quad (9)$$

Moreover, for any  $u := (\widehat{x}, \widehat{z}, \widehat{y})$ , the following inclusion is valid :

$$L(\mathcal{G}u) + \widehat{S}^t\widehat{S}(\mathcal{G}u) \ni u + \begin{pmatrix} -\nabla h(\eta) \\ 0 \\ 0 \end{pmatrix} \quad (10)$$

where  $\eta = (R_1 + 2A^tMA)^{-1}(\widehat{x})$ .

**Proof.** The equivalence  $G_3 = \widehat{S}\widehat{\mathcal{G}}\widehat{S}^t$  is following from their definitions. To show the second part, given  $u := (\widehat{x}, \widehat{z}, \widehat{y})$ , considering  $(x, z, \nu) = \mathcal{G}u$ , it holds that

$$\partial f(x) + A^t\nu + (R_1 + A^tMA)x - A^tMBz \ni \widehat{x} - \nabla h((R_1 + 2A^tMA)^{-1}(\widehat{x}))$$

$$\begin{aligned} \partial g(z) + B^t\nu - B^tMAx + (R_2 + B^tMB)z &\ni \widehat{z} \\ -Ax - Bz + M^{-1}\nu &= \widehat{y} \end{aligned}$$

Then the last relations show the desirable result ■

Using the last proposition, we show the averaged properties of  $G_3$ .

**Proposition 4** *Let again assume that the Hypo – M hypotheses are satisfied. Considering that  $\|(R_1 + 2A^tMA)^{-1}\| \in ]0, 2\beta[$ , the map  $G_3$  is  $\alpha$ -averaged with full domain, where  $\alpha := \frac{2\beta}{4\beta - \|(R_1 + 2A^tMA)^{-1}\|} \in ]\frac{1}{2}, 1[$ .*

**Proof.** To ease the formulation, we use the following notations :  $\mu_1 = (\widehat{x}_1, \widehat{z}_1, \widehat{y}_1)$  and  $\mu_2 = (\widehat{x}_2, \widehat{z}_2, \widehat{y}_2)$ . The evaluations of  $\widehat{\mathcal{G}}\widehat{S}^t\widehat{S}$  using  $\mu_1$  and  $\mu_2$  are, for  $i = 1, 2$  :

$$\widehat{\mathcal{G}}\widehat{S}^t\widehat{S}\mu_i = \begin{pmatrix} x_i \\ z_i \\ MAx_i + MBz_i + \widehat{y}_i \end{pmatrix},$$

where

$$x_i := (\partial f + R_1 + 2A^tMA)^{-1}((R_1 + A^tMA)\widehat{x}_i - A^tMB\widehat{z}_i - A^t\widehat{y}_i - \nabla h(\eta_i))$$

$$\eta_i := (R_1 + 2A^tMA)^{-1}((R_1 + A^tMA)\widehat{x}_i - A^tMB\widehat{z}_i)$$

$$z_i := (\partial g + R_2 + 2B^tMB)^{-1}((R_2 + B^tMB)\widehat{z}_i - B^tMA\widehat{x}_i - B^t\widehat{y}_i).$$

From (10) considering  $u$  equal to  $\widehat{S}^t\widehat{S}\mu_1$  and  $\widehat{S}^t\widehat{S}\mu_2$ , using the monotonicity of operator  $L$ , and expression (9), we obtain that

$$\begin{aligned} \left\langle G_3\widehat{S}\mu_1 - G_3\widehat{S}\mu_2, (\widehat{S}\mu_1 - G_3\widehat{S}\mu_1) - (\widehat{S}\mu_2 - G_3\widehat{S}\mu_2) \right\rangle + \\ \langle x_1 - x_2, \nabla h(\eta_2) - \nabla h(\eta_1) \rangle \geq 0 \quad (11) \end{aligned}$$

Now we find an appropriate upper bound for  $\langle x_1 - x_2, \nabla h(\eta_2) - \nabla h(\eta_1) \rangle$  considering  $\|\hat{S}\mu_1 - G_3\hat{S}\mu_1 - \hat{S}\mu_2 + G_3\hat{S}\mu_2\|$ . Rewriting  $\langle x_1 - x_2, \nabla h(\eta_2) - \nabla h(\eta_1) \rangle$  as

$$\langle \eta_1 - \eta_2, \nabla h(\eta_2) - \nabla h(\eta_1) \rangle + \langle x_1 - x_2 + \eta_2 - \eta_1, \nabla h(\eta_2) - \nabla h(\eta_1) \rangle$$

then since  $\nabla h$  is co-coercive and using cauchy-inequality, it holds that

$$\begin{aligned} \langle x_1 - x_2, \nabla h(\eta_2) - \nabla h(\eta_1) \rangle &\leq -\beta \|\nabla h(\eta_2) - \nabla h(\eta_1)\|^2 + \frac{1}{4\beta} \|x_1 - x_2 + \eta_2 - \eta_1\|^2 \\ &\quad + \beta \|\nabla h(\eta_2) - \nabla h(\eta_1)\|^2 \\ &= \frac{1}{4\beta} \|x_1 - x_2 + \eta_2 - \eta_1\|^2 \end{aligned} \quad (12)$$

On other hands, it holds that  $\|\hat{S}\mu_1 - G_3\hat{S}\mu_1 - \hat{S}\mu_2 + G_3\hat{S}\mu_2\|$  is equal to

$$\|\hat{S} [\mu_1 - \mathcal{G}\hat{S}^t\hat{S}\mu_1 - \mu_2 + \mathcal{G}\hat{S}^t\hat{S}\mu_2]\|^2 = \|(p_1, p_2, p_3, p_4)\|^2 \quad (13)$$

where

$$\begin{aligned} p_1 &= R_1^{\frac{1}{2}}(\hat{x}_1 - x_1 - \hat{x}_2 + x_2) \\ p_2 &= R_2^{\frac{1}{2}}(\hat{z}_1 - z_1 - \hat{z}_2 + z_2) \\ p_3 &= M^{\frac{1}{2}}A(\hat{x}_1 - x_1 - \hat{x}_2 + x_2) - M^{\frac{1}{2}}B(\hat{z}_1 - z_1 - \hat{z}_2 + z_2) \\ p_4 &= M^{-\frac{1}{2}}(-MAx_1 - MBz_1 + MAx_2 + MBz_2). \end{aligned}$$

Denoting  $K = \begin{pmatrix} R_1^{\frac{1}{2}} & A^t M^{\frac{1}{2}} & A^t M^{\frac{1}{2}} \end{pmatrix}$ , we have that

$$K(p_1, p_3, p_4)^t = KK^t(\eta_1 - \eta_2 - x_1 + x_2),$$

and then using the last relation and (13), we obtain that

$$\|\hat{S}\mu_1 - G_3\hat{S}\mu_1 - \hat{S}\mu_2 + G_3\hat{S}\mu_2\|^2 \geq \frac{1}{\|(KK^t)^{-1}\|} \|x_1 - x_2 + \eta_2 - \eta_1\|^2.$$

Therefore from (12), we obtain the desired upper bound

$$\langle x_1 - x_2, \nabla h(\eta_2) - \nabla h(\eta_1) \rangle \leq \frac{\|(KK^t)^{-1}\|}{4\beta} \|\hat{S}\mu_1 - G_3\hat{S}\mu_1 - \hat{S}\mu_2 + G_3\hat{S}\mu_2\|^2.$$

Finally using this upper bound in (11) we obtain that

$$\|G_3\hat{S}\mu_1 - G_3\hat{S}\mu_2\|^2 \leq \|\hat{S}\mu_1 - \hat{S}\mu_2\|^2 - \frac{1-\alpha}{\alpha} \|\hat{S}\mu_1 - G_3\hat{S}\mu_1 - \hat{S}\mu_2 + G_3\hat{S}\mu_2\|^2$$

where

$$\alpha := \frac{2\beta}{4\beta - \|(R_1 + 2A^tMA)^{-1}\|}.$$

Then since for any  $s, s' \in \mathbb{R}^q$  there exist  $u_1, u_2 \in \mathbb{R}^r$  such that  $\hat{S}^t\hat{S}u_1 = \hat{S}^ts$  and  $\hat{S}^t\hat{S}u_2 = \hat{S}^ts'$ , from the last inequality and (9) we get that  $G$  is  $\alpha$ -average.  $\blacksquare$

### 3 Inertial-relaxed fixed point algorithms

In practice, any variant of a basic fixed point algorithm applied to some averaged map will yield a valid variant of the splitting algorithm related to it.

Considering a maximal monotone operator  $T$ , we describe here an Inertial-Relaxed variant to fixed point algorithm with relative error, inspired by a recent work of Alves et al. [2]. We recall first their variant of the Proximal Point algorithm to solve the inclusion  $0 \in T(x)$  for a given maximal monotone operator  $T$ , called Relative Error Inertial Relaxed Hybrid Proximal Point (RIRHPP) :

#### Relative-error Inertial-Relaxed HPP (RIRHPP)

**Initialization:** Choose  $z^0 = z^{-1} \in \mathbb{R}^r$ ,  $\bar{\lambda}, \theta \in [0, 1[$  and  $\bar{\rho} \in ]0, 2[$

**For**  $k = 0, 1, \dots$  **do**

– Choose  $\lambda^k \in [0, \bar{\lambda}[$  (Inertial parameter) and define

$$w^k = z^k + \lambda^k(z^k - z^{k-1})$$

– Inexact Subproblem:

Find  $(\tilde{z}^k, v^k) \in \mathbb{R}^r \times \mathbb{R}^r$  and  $c^k \geq 0$  such that

$$v^k \in T(\tilde{z}^k), \quad \|c^k v^k + \tilde{z}^k - w^k\|^2 \leq \theta^2 (\|\tilde{z}^k - w^k\|^2 + \|c^k v^k\|^2) \quad (14)$$

– If  $v^k = 0$ , then **STOP**. Otherwise, choose  $\rho^k \in [0, \bar{\rho}]$  (Relaxing parameter) and set

$$z^{k+1} = w^k - \rho^k \frac{\langle w^k - \tilde{z}^k, v^k \rangle}{\|v^k\|^2} v^k.$$

**end for**

We notice that *RIRHPP* can be rewritten in terms of the resolvent of  $T$ , so that we can extend this algorithm in order to find a fixed point of a 1-co-coercive map with full domain since, by Minty's Theorem, any 1-co-coercive map with full domain is the resolvent of a maximal monotone operator.

Thus, given the problem of finding a fixed point of a  $\alpha$ -averaged map  $F$  with full domain, we consider alternatively  $(1 - \frac{1}{2\alpha})I + \frac{1}{2\alpha}F$  a 1-co-coercive map with full domain, which has the same fixed points. Then we extend algorithm *RIRHPP* to a 1-co-coercive map with full domain. In summary we obtain the following algorithm where, for simplicity, we do not include the relative error feature of (RIRHPP) and consider fixed inertial-relaxed parameters :

#### Inertial-Relaxed Fixed Point (IRFP)

**Initialization:** Choose  $z^0 = z^{-1} \in \mathbb{R}^n$ , also  $(\bar{\lambda}, \bar{\rho}) \in ]0, 1[ \times ]0, 2[$  satisfying (H1) and let  $\hat{\rho} := \frac{\bar{\rho}}{2\alpha}$ .

**For**  $k = 0, 1, 2, \dots$  **do**

– Choose  $\lambda \in [0, \bar{\lambda}[$  and define (inertial term)

$$w^k = z^k + \lambda(z^k - z^{k-1}) \quad (15)$$

– Choose  $\rho \in ]0, \bar{\rho}]$  and calculate (relaxed term of fixed point algorithm)

$$z^{k+1} = (1 - \rho)w^k + \rho F(w^k). \quad (16)$$

**end for**

We consider similar conditions on the bounds  $\bar{\lambda}$  and  $\bar{\rho}$  as the ones given in [2] for *RIRHPP* i.e.

(H1)  $(\bar{\lambda}, \bar{\rho}) \in ]0, 1[ \times ]0, 2[$  and the upper bound  $\bar{\rho}$  is a function of  $\bar{\lambda}$  given by

$$\bar{\rho} = \frac{2(\bar{\lambda} - 1)^2}{2(\bar{\lambda} - 1)^2 + 3\bar{\lambda} - 1}. \quad (17)$$

In the case that no inertial term is used ( $\lambda = 0$ ), we bound the relaxation parameter by  $\bar{\rho} < \frac{1}{\alpha}$  (which is known to guarantee convergence in that case).

The convergence of *IRFP* can be directly derived from [2], but we give below an equivalent direct proof.

**Proposition 5** *Set  $\bar{\lambda}$  and  $\bar{\rho}$  satisfying (H1). Given  $F$  an  $\alpha$ -average map with full domain with at least one fixed point, then the sequences  $\{w^k\}$  and  $\{z^k\}$  computed by algorithm (IRFP) both converge to the same fixed point of  $F$ .*

**Proof.** Since  $F$  is  $\alpha$ -averaged, then  $(1 - \rho)I + \rho F$  is  $\alpha\rho$ -averaged with the same fixed point. Using (15), and given  $x^*$  any fixed point of  $F$ , we have

$$\|w^k - x^*\|^2 \geq \|z^{k+1} - x^*\|^2 + \frac{1 - \alpha\rho}{\alpha\rho} \|w^k - z^{k+1}\|^2. \quad (18)$$

From (16), we have  $w^k - x^* = (1 + \lambda)(z^k - x^*) - \lambda(z^{k-1} - x^*)$ , and using the property of  $\|\cdot\|^2$  we get

$$\|w^k - x^*\|^2 = (1 + \lambda)\|z^k - x^*\|^2 - \lambda\|z^{k-1} - x^*\|^2 + \lambda(1 + \lambda)\|z^k - z^{k-1}\|^2.$$

Defining  $\varphi^k := \|z^k - x^*\|^2$  and using the last equality in (18) and  $\rho \leq \hat{\rho}$ , it holds

$$\varphi^k - \varphi^{k+1} \geq \lambda(\varphi^{k-1} - \varphi^k) + \frac{1 - \alpha\hat{\rho}}{\alpha\hat{\rho}} \|w^k - z^{k+1}\|^2 - \lambda(1 + \lambda)\|z^k - z^{k-1}\|^2. \quad (19)$$

Now using the fact that, for  $\bar{\lambda}$  and  $\bar{\rho}$  satisfying (H1),  $\sum \|z^k - z^{k-1}\|^2$  will converge, then applying Lemma A.5 of [2], it holds that  $\sum \|w^k - z^{k+1}\|^2$  and  $\{\varphi^k\}$  both converge. Therefore, using that  $\{z^k\}$  is bounded,  $\{\|z^k - z^{k-1}\|^2\}$  and  $\{\|w^k - z^{k+1}\|^2\}$  both converge to zero. It holds that given  $z'$  a cluster point of  $\{z^k\}$ , from (15) we have  $0 = \rho(-z' + F(z'))$  then  $z'$  is a fixed point of  $F$ , then considering  $z^* = z'$ , we have that  $\{\|z^k - z'\|\}$  converges to zero which implies that  $\{w^k\}$  and  $\{z^k\}$  converge to  $z'$ .

So it just remains to prove that  $\sum \|z^k - z^{k-1}\|^2$  converges. From (16), we have  $w^k - z^{k+1} = (1 - \lambda)(z^k - z^{k+1}) + \lambda(z^k - z^{k+1} + z^k - z^{k-1})$ , and using the property of  $\|\cdot\|^2$  we have

$$\begin{aligned} \|w^k - z^{k+1}\| &= (1 - \lambda)\|z^k - z^{k+1}\|^2 + \lambda\|2z^k - z^{k+1} - z^{k-1}\|^2 - \lambda(1 - \lambda)\|z^k - z^{k-1}\|^2 \\ &\geq (1 - \lambda)\|z^k - z^{k+1}\|^2 - \lambda(1 - \lambda)\|z^k - z^{k-1}\|^2, \end{aligned}$$

then replacing in (19), and denoting  $\eta = \lambda(\alpha\hat{\rho})^{-1} + \lambda^2(2\alpha\hat{\rho} - 1)(\alpha\hat{\rho})^{-1}$ , it holds

$$\varphi^k - \varphi^{k+1} \geq \lambda(\varphi^{k-1} - \varphi^k) + \frac{(1 - \alpha\hat{\rho})(1 - \lambda)}{\alpha\hat{\rho}}\|z^k - z^{k+1}\|^2 - \eta\|z^k - z^{k-1}\|^2.$$

Defining  $\mu^k := \varphi^k - \lambda\varphi^{k-1} + \eta\|z^k - z^{k-1}\|^2$ , the last inequality is rewritten as

$$\mu^k \geq \mu^{k+1} + q(\lambda)\|z^k - z^{k+1}\|^2, \quad (20)$$

where  $q$  is a quadratic function defined by

$$q(\lambda) := (1 - \alpha\hat{\rho})(\alpha\hat{\rho})^{-1} - (2 - \alpha\hat{\rho})(\alpha\hat{\rho})^{-1}\lambda - (2\alpha\hat{\rho} - 1)(\alpha\hat{\rho})^{-1}\lambda^2.$$

Since  $\bar{\rho} < 2$ , then  $q(0) > 0$ , independently of the sign of the principal coefficient of  $q$ , it holds that  $q(\lambda)$  is strictly positive when  $\lambda \in [0, \bar{\lambda}(\bar{\rho})]$ , where  $\bar{\lambda}(\bar{\rho})$  is the smallest positive root of  $q$  (see Lemma A.3 in [2]), which is equal to  $\bar{\lambda}(\bar{\rho}) = \bar{\lambda}(\bar{\rho})$  (recall that  $2\alpha\hat{\rho} = \bar{\rho}$ ), with  $\bar{\lambda} : (0, 2) \rightarrow (0, 1)$  given by

$$\bar{\lambda}(\bar{\rho}) := \frac{2(2 - \bar{\rho})}{4 - \bar{\rho} + \sqrt{16\bar{\rho} - 7\bar{\rho}^2}},$$

whose inverse function is  $\bar{\rho} : (0, 1) \rightarrow (0, 2)$  equal to

$$\bar{\rho}(\bar{\lambda}) := \frac{2(\bar{\lambda} - 1)^2}{2(\bar{\lambda} - 1)^2 + 3\bar{\lambda} - 1}$$

Therefore if we choose  $(\bar{\lambda}, \bar{\rho}) \in (0, 1) \times (0, 2)$  satisfying (H1), we have that  $q(\lambda)$  is strictly positive. Then from (20), it is sufficient to show that  $u^k$  is bounded from below. So, from the definition of  $\mu^k$ , it holds that  $\mu^k \geq \varphi^k - \lambda\varphi^{k-1}$ , and we get

$$\lambda^j u^{k-j} \geq \lambda^j \varphi^{k-j} - \lambda^{j+1} \varphi^{k-j-1}, \quad \forall j = 0, \dots, k-1,$$

then summing and using that  $u^k$  is decreasing (from relation (20)), it finally holds that

$$\left( \sum_{j=0}^{k-1} \lambda^j \right) u^1 + \lambda^k \varphi^0 \geq \varphi^k.$$

Since  $\lambda < 1$ , it holds that  $\{\varphi^k\}$  is bounded which implies the boundedness of  $\{u^k\}$ . ■

## 4 Inertial-relaxed splitting algorithms

Based on the averaged maps constructed in section 2 and the variants of the fixed point algorithm *IRFP* developed in the previous section, we obtain generalized splitting algorithms, including the Inertial and Relaxation parameters, to solve the composite model (1).

### 4.1 Splitting algorithms in the Gauss-Seidel version

Considering first the map  $G_1$ , we will obtain a splitting algorithm which, without inertial nor relaxation tuning parameters, goes back to *PD3O* (as we will see in section 5). We obtain thus a different algorithm compared to Condat-Vũ Algorithm, Form I, when  $h \neq 0$ . This new algorithm is termed “Multi-scaling Inertial-Relaxed primal-dual algorithm, Form I”:

#### Multi-scaling Inertial-Relaxed primal-dual algorithm, Form I (MIRPD, Form I)

Choose  $(x^0, z^0, y^0, r^0) = (x^{-1}, z^{-1}, y^{-1}, r^{-1}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^n$ ,  $V_1 \in \mathbb{R}^{n \times n}$ ,  $V_2 \in \mathbb{R}^{p \times p}$  and  $M \in \mathbb{R}^{m \times m}$  such that  $V_1 + A^t M A$  and  $V_2 + B^t M B$  are positive definite, and parameters  $(\bar{\lambda}, \bar{\rho}) \in ]0, 1[ \times ]0, 2[$  such that  $\lambda \in [0, \bar{\lambda}]$  and  $\rho \in ]0, \frac{\bar{\rho}}{2\alpha}]$ , where

$$\alpha := \frac{2\beta}{4\beta - \|(V_1 + A^t M A)^{-1}\|}.$$

$$(x_w^k, z_w^k, y_w^k, r_w^k) = (x^k, z^k, y^k, r^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}, r^k - r^{k-1}) \quad (21)$$

$$\tilde{x}^{k+1} = (\partial f + V_1 + A^t M A)^{-1}(V_1 x_w^k - A^t M B z_w^k - A^t y_w^k - V_1 r_w^k) \quad (22)$$

$$\tilde{y}^{k+1} = y_w^k + M A \tilde{x}^{k+1} + M B z_w^k \quad (23)$$

$$\tilde{r}^{k+1} = (V_1 + A^t M A)^{-1} \nabla h(\tilde{x}^{k+1}) \quad (24)$$

$$\tilde{z}^{k+1} = (\partial g + V_2 + B^t M B)^{-1}(V_2 z_w^k - B^t M A \tilde{x}^{k+1} - B^t \tilde{y}^{k+1} + B^t M A \tilde{r}^{k+1}) \quad (25)$$

$$(x^{k+1}, z^{k+1}, y^{k+1}, r^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}, \tilde{r}^{k+1}) + (1 - \rho)(x_w^k, z_w^k, y_w^k, r_w^k) \quad (26)$$

Considering  $\xi^k = (x^k - r^k, z^k, y^k)$  and  $\xi_w^k = (x_w^k - r_w^k, z_w^k, y_w^k)$ , the relation of this algorithm with *IRFP* applied to the averaged map  $G_1$  is

$$\bar{Q}\xi_w^k = \bar{Q}\xi^k + \lambda(\bar{Q}\xi^k - \bar{Q}\xi^{k-1}) \quad \text{and} \quad \bar{Q}\xi^{k+1} = (1 - \rho)\bar{Q}\xi_w^k + \rho G_2(\bar{Q}\xi_w^k). \quad (27)$$

This relation allows us to obtain the following convergence result.

**Proposition 6** *Set  $\bar{\lambda}$  and  $\bar{\rho}$  satisfying (H1). With the same hypothesis of the matrix as in Proposition 2. If  $\text{sol}(V_{\hat{L}})$  is nonempty, then building the sequences  $(\tilde{x}^k, \tilde{r}^k, \tilde{z}^k, \tilde{y}^k)$  in (21)-(26), it holds that  $(\tilde{x}^k, \tilde{z}^k, \tilde{y}^k - M A \tilde{r}^k)$  converge to some element of  $\text{sol}(V_{\hat{L}})$ .*



**Proof.** From relation (27), we observe that this algorithm is related to *IRFP* applied to operator  $G_1$  which, from Proposition 1, is  $\alpha$ -averaged. Then by Proposition 5 and relation (3), we deduce that

$$\begin{pmatrix} V_1^{\frac{1}{2}}(x_w^k - r_w^k) \\ V_2^{\frac{1}{2}}z_w^k \\ -M^{\frac{1}{2}}Bz_w^k - M^{-\frac{1}{2}}y_w^k \end{pmatrix} \text{ converge to } \begin{pmatrix} V_1^{\frac{1}{2}}[x^* - W\nabla h(x^*)] \\ V_2^{\frac{1}{2}}z^* \\ M^{\frac{1}{2}}A[x^* - W\nabla h(x^*)] - M^{-\frac{1}{2}}y^* \end{pmatrix},$$

where  $(x^*, z^*, y^*) \in \text{sol}(V_{\mathcal{L}})$  and  $W = (V_1 + A^tMA)^{-1}$ . Since  $(\partial f + V_1 + A^tMA)^{-1}$  is uni-valued and continuous, from (22) we have that  $\{\tilde{x}^k\}$  converges to  $(\partial f + V_1 + A^tMA)^{-1}(V_1(x^* - W\nabla h(x^*) + A^tMA(x^* - W\nabla h(x^*)) - A^ty^*) = x^*$ , and then from (23) and (24) we obtain that  $\{\tilde{y}^k\}$  and  $\{\tilde{r}^k\}$  converge to  $MA(A^tMA + V)^{-1}\nabla h(x^*) + y^*$  and  $(V_1 + A^tMA)^{-1}\nabla h(x^*)$  respectively, and then from continuity of  $(\partial g + V_2 + B^tMB)^{-1}$  and (25) we have that  $\{z^k\}$  converge to  $z^*$ . ■

Back now to the switched operator  $G_2$ , we obtain the second form of the last splitting algorithm, switching the order of action of the proximal steps, in the same manner as Condat-Vũ's algorithm forms. This new algorithm, without inertial nor relaxation tuning parameters, goes back to the Gauss-Seidel version of *PPDS* when  $h = 0$ .

### Multi-scaling Inertial-Relaxed primal-dual algorithm, Form II (MIRPD, Form II)

Choose  $(x^0, z^0, y^0) = (x^{-1}, z^{-1}, y^{-1}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ ,  $V_1 \in \mathbb{R}^{n \times n}$ ,  $V_2 \in \mathbb{R}^{p \times p}$  and  $M \in \mathbb{R}^{m \times m}$  such that  $V_1 + A^tMA$  and  $V_2 + B^tMB$  are positive definite matrices, and parameters  $(\bar{\lambda}, \bar{\rho}) \in ]0, 1[ \times ]0, 2[$  such that  $\lambda \in [0, \bar{\lambda}]$  and  $\rho \in ]0, \frac{\bar{\rho}}{2\alpha}]$ , where

$$\alpha := \frac{2\beta}{4\beta - \|(V_1 + A^tMA)^{-1}\|}.$$

$$(x_w^k, z_w^k, y_w^k) = (x^k, z^k, y^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}) \quad (28)$$

$$\tilde{z}^{k+1} = (\partial g + V_2 + B^tMB)^{-1}(V_2z_w^k - B^tMAx_w^k - B^ty_w^k) \quad (29)$$

$$\tilde{y}^{k+1} = y_w^k + MAx_w^k + MB\tilde{z}^{k+1} \quad (30)$$

$$r^{k+1} = \nabla h((V_1 + A^tMA)^{-1}(V_1x_w^k - A^tMB\tilde{z}^{k+1})) \quad (31)$$

$$\tilde{x}^{k+1} = (\partial f + V_1 + A^tMA)^{-1}(V_1x_w^k - A^tMB\tilde{z}^{k+1} - A^t\tilde{y}^{k+1} - r^{k+1}) \quad (32)$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1 - \rho)(x_w^k, z_w^k, y_w^k). \quad (33)$$

Analogous to the last algorithm, we have a relation with *IRFP* applied to  $G_2$ . Considering  $\zeta^k := (x^k, z^k, y^k)$  and  $\zeta_w^k := (x_w^k, z_w^k, y_w^k)$  it holds

$$\widehat{Q}\zeta_w^k = \widehat{Q}\zeta^k + \lambda(\widehat{Q}\zeta^k - \widehat{Q}\zeta^{k-1}) \quad \text{and} \quad \widehat{Q}\zeta^{k+1} = (1 - \rho)\widehat{Q}\zeta_w^k + \rho G_1(\widehat{Q}\zeta_w^k). \quad (34)$$

**Proposition 7** Set  $\bar{\lambda}$  and  $\bar{\rho}$  satisfying (H1), and keep the same hypotheses of Proposition 2. If  $\text{sol}(V_{\bar{L}})$  is nonempty, then for an arbitrary  $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ , the sequence  $(\tilde{x}^k, \tilde{z}^k, \tilde{y}^k)$  in (28)-(33) converges to some element of  $\text{sol}(V_L)$ .

**Proof.** From relations (34), (4) and Proposition 5, it holds that  $\widehat{Q}\xi_w^k$  converge to a fixed point of  $G_2$ , then we have

$$\widehat{Q} \begin{pmatrix} x_w^k \\ z_w^k \\ y_w^k \end{pmatrix} = \begin{pmatrix} V_1^{\frac{1}{2}} x_w^k \\ V_2^{\frac{1}{2}} z_w^k \\ M^{\frac{1}{2}} A x_w^k + M^{-\frac{1}{2}} y_w^k \end{pmatrix} \text{ converge to } \begin{pmatrix} V_1^{\frac{1}{2}} x^* \\ V_2^{\frac{1}{2}} z^* \\ M^{\frac{1}{2}} A x^* + M^{-\frac{1}{2}} y^* \end{pmatrix},$$

where  $(x^*, z^*, y^*) \in \text{sol}(V_{\mathcal{L}})$ , since  $(\partial g + V_2 + B^t M B)^{-1}$  is a single-valued continuous map. From (29), we obtain that  $\{\tilde{z}^k\}$  converges to  $(\partial g + V_2 + B^t M B)^{-1}(V_2 z^* - B^t M A x^* - B^t y^*) = z^*$ . Then from (30) and (31), it holds that  $\{\tilde{y}^k\}$  and  $\{r^k\}$  converge to  $y^*$  and  $\nabla h(x^*)$ , respectively. In the same way, from the continuity of  $(\partial f + V + A^t M A)^{-1}$ , we deduce that  $\{\tilde{x}^k\}$  converges to  $x^*$ . ■

## 4.2 A splitting algorithm in the Jacobi version

Now considering  $G_3$ , we will obtain a new splitting algorithm which goes back to the Jacobi version of proximal primal-dual algorithms (PPDS) when  $h = 0$  and with the scaling matrices  $V_1 = R_1 + A^t M A$ ,  $V_2 = R_2 + B^t M B$ . We call it “Multi-scaling Inertial-Relaxed parallel primal-dual algorithm”:

### Multi-scaling Inertial-Relaxed primal-dual algorithm, parallel version (MIRPD, Jacobi version)

Choose  $(x^0, z^0, y^0) = (x^{-1}, z^{-1}, y^{-1}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ ,  $R_1 \in \mathbb{R}^{n \times n}$ ,  $R_2 \in \mathbb{R}^{p \times p}$  and  $M \in \mathbb{R}^{m \times m}$  such that  $R_1 + 2A^t M A$  and  $R_2 + 2B^t M B$  are positive definite, and parameters  $(\bar{\lambda}, \bar{\rho}) \in ]0, 1[ \times ]0, 2[$  such that  $\lambda \in [0, \bar{\lambda}]$  and  $\rho \in ]0, \frac{\bar{\rho}}{2\alpha}]$ , where

$$\alpha := \frac{2\beta}{4\beta - \|(R_1 + 2A^t M A)^{-1}\|}.$$

$$(x_w^k, z_w^k, y_w^k) = (x^k, z^k, y^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}) \quad (35)$$

$$r^k = \nabla h((R_1 + 2A^t M A)^{-1}((R_1 + A^t M A)x_w^k - A^t M B z_w^k)) \quad (36)$$

$$\tilde{x}^{k+1} = (\partial f + R_1 + 2A^t M A)^{-1}((R_1 + A^t M A)x_w^k - A^t M B z_w^k - A^t y_w^k - r^k) \quad (37)$$

$$\tilde{z}^{k+1} = (\partial g + R_2 + 2B^t M B)^{-1}((R_2 + B^t M B)z_w^k - B^t M A x_w^k - B^t y_w^k) \quad (38)$$

$$\tilde{y}^{k+1} = y_w^k + M A \tilde{x}^{k+1} + M B \tilde{z}^{k+1} \quad (39)$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1 - \rho)(x_w^k, z_w^k, y_w^k) \quad (40)$$

Again this algorithm is related to *IRFP* applied to  $G_3$ . Considering  $\varsigma^k = (x^k, z^k, y^k)$  and  $\varsigma_w^k = (x_w^k, z_w^k, y_w^k)$  it holds

$$\widehat{S}_{\varsigma_w^k} = \widehat{S}_{\varsigma^k} + \lambda(\widehat{S}_{\varsigma^k} - \widehat{S}_{\varsigma^{k-1}}) \text{ and } \widehat{S}_{\varsigma^{k+1}} = (1 - \rho)\widehat{S}_{\varsigma_w^k} + \rho G_3(\widehat{S}_{\varsigma_w^k}). \quad (41)$$

**Proposition 8** *Set  $\bar{\lambda}$  and  $\bar{\rho}$  satisfying (H1). With the same hypotheses as in Proposition 4 and if  $\text{sol}(V_{\bar{L}})$  is nonempty, then for an arbitrary  $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ , the sequence  $(\tilde{x}^k, \tilde{y}^k, \tilde{y}^k)$  in (35)-(40) converges to some element of  $\text{sol}(V_L)$ .*

**Proof.** From relation (41), we observe that this algorithm is related to *IRFP* applied to operator  $G_3$  which, from the hypotheses, is  $\alpha$ -average. Then by Proposition 5 and relation (7), we deduce that

$$\widehat{S} \begin{pmatrix} x_w^k \\ z_w^k \\ y_w^k \end{pmatrix} = \begin{pmatrix} R_1^{\frac{1}{2}} x_w^k \\ R_2^{\frac{1}{2}} z_w^k \\ M^{\frac{1}{2}} A x_w^k - M^{\frac{1}{2}} B z_w^k \\ M^{-\frac{1}{2}} y_w^k \end{pmatrix} \text{ converge to } \begin{pmatrix} R_1^{\frac{1}{2}} x^* \\ R_2^{\frac{1}{2}} z^* \\ M^{\frac{1}{2}} A x^* - M^{\frac{1}{2}} B z^* \\ M^{-\frac{1}{2}} y^* \end{pmatrix},$$

where  $(x^*, z^*, y^*) \in \text{sol}(V_{\bar{L}})$ . From continuity of  $\nabla h$  and (36), we have that  $\{r^k\}$  converge to  $\nabla h(x^*)$ , then since  $(\partial f + R_1 + 2A^t M A)^{-1}$  and  $(\partial g + R_2 + 2B^t M B)^{-1}$  are uni-valued continuous maps from (37) and (38), we obtain that  $\{\tilde{x}^k\}$  and  $\{\tilde{z}^k\}$  converges to  $x^*$  and  $z^*$  respectively. It holds that  $\{y_w^k\}$  converge to  $y^*$ , then from (39) we have  $\{\tilde{y}^k\}$  converge to  $y^*$ .  $\blacksquare$

**Remark 4.1** *When neither inertial nor relaxed parameters are considered ( $\lambda = 0$  and  $\rho = 1$ ) in MIRPD, Form II and MIRPD, Jacobi version, these algorithms yield a clear extension of PPDS. Defining  $(\bar{x}^k, \bar{z}^k, \bar{y}^k) := (\tilde{x}^k, \tilde{z}^{k+1}, \tilde{y}^{k+1})$  and  $(\bar{x}^k, \bar{z}^k, \bar{y}^k) := (\tilde{x}^k, \tilde{z}^k, \tilde{y}^k)$  in the above algorithms, we obtain the following new version of PPDS applied to model (1).*

*Proximal-Gradient primal-dual Algorithm (PGPDS)*

$$\begin{cases} \bar{r}^k &= (V_1 + A^t M A)^{-1} \nabla h((V_1 + A^t M A)^{-1} (V_1 \bar{x}^k - A^t M B \bar{z}^k)) \\ \bar{x}^{k+1} &\in \text{argmin} \left\{ f(x) + \frac{1}{2} \|A(x + \bar{r}^k) + B \bar{z}^k + M^{-1} \bar{y}^k\|_M^2 + \frac{1}{2} \|x + \bar{r}^k - \bar{x}^k\|_{V_1}^2 \right\} \\ \bar{z}^{k+1} &\in \text{argmin} \left\{ g(z) + \frac{1}{2} \|A \bar{\eta}^k + B z + M^{-1} \bar{y}^k\|_M^2 + \frac{1}{2} \|z - \bar{z}^k\|_{V_2}^2 \right\} \\ \bar{y}^{k+1} &= \bar{y}^k + M(A \bar{x}^{k+1} + B \bar{z}^{k+1}) \end{cases}$$

where choosing  $\bar{\eta}^k$  as below gives us two algorithm versions:

$$\bar{\eta}^k := \begin{cases} \bar{x}^k & \text{for Jacobi version algorithm} \\ \bar{x}^{k+1} & \text{for Gauss-Seidel version algorithm} \end{cases}$$

From Proposition 8 and 7 some conditions on matrices  $V_1$  and  $V_2$  need to be imposed in order to obtain convergence: they are positive semi-definite for the Gauss-Seidel version and, for the Jacobi version, they are of the form  $V_1 = R_1 + A^t M A$  and  $V_2 = R_2 + B^t M B$ , for some  $R_1$  and  $R_2$  positive semi-definite.

Reformulating MIRPD, Form II as PGPDS for Gauss-Seidel version algorithm allows us to obtain a clearer comparison with *iPADMM* algorithm described in [16]. Concerning MIRPD, Form I, in Section 5, we show that it is related to PD3O.

## 5 Resulting variants of Condat-Vũ and PD3O algorithms

Independently of the structure of matrices  $A$  and  $B$ , a practical tuning of the scaling parameters of the last algorithms which are still inside the theoretical bounds for convergence can be defined as follow:

Let  $\sigma, \tau, \mu$  positive such that  $\sigma\tau\|A\|^2 \leq 1$ ,  $\sigma\mu\|B\|^2 \leq 1$  and  $\tau < 2\beta$ , consider

$$M = \sigma I_{m \times m}, \quad V_1 = \tau^{-1} I_{n \times n} - \sigma A^t A \quad \text{and} \quad V_2 = \mu^{-1} I_{p \times p} - \sigma B^t B \quad (42)$$

for algorithm *MIRPD*, *Form I* and *Form II*.

Let  $\tilde{\sigma}, \tilde{\tau}, \tilde{\mu}$  positive such that  $2\tilde{\sigma}\tilde{\tau}\|A\|^2 \leq 1$ ,  $2\tilde{\sigma}\tilde{\mu}\|B\|^2 \leq 1$  and  $\tilde{\tau} < 2\beta$ , consider

$$M = \tilde{\sigma} I_{m \times m}, \quad R_1 = \tilde{\tau}^{-1} I_{n \times n} - 2\tilde{\sigma} A^t A \quad \text{y} \quad R_2 = \tilde{\mu}^{-1} I_{p \times p} - 2\tilde{\sigma} B^t B \quad (43)$$

for algorithm *MIRPD*, *Jacobi version*.

In order to compare Condat-Vũ and PD3O algorithms with the new algorithms of the last section, we consider the case  $B = -I_{p \times p}$ , and the special matrix parameters (42) and (43), obtaining the following specialized algorithms.

Corresponding to *MIRPD*, *Form I*, with matrix parameters satisfying (42) and  $V_2 = 0$ ; using that  $(\sigma T^{-1} + I_{m \times m})^{-1} = I - \sigma(T + \sigma I_{m \times m})^{-1}$ , and considering  $\eta^k = \sigma A x^k + y^k - \sigma z^k - \sigma A r^k$ ,  $\eta_w^k = \sigma A x_w^k + y_w^k - \sigma z_w^k - \sigma A r_w^k$  and  $\tilde{\eta}^k = \sigma A \tilde{x}^k + \tilde{y}^k - \sigma \tilde{z}^k - \sigma A \tilde{r}^k$ , we get the following algorithm:

### Inertial-Relaxed Primal-Dual Three Operator, Form I (IRPD3O, Form I)

Choose  $(x^0, \eta^0, r^0) = (x^{-1}, \eta^{-1}, r^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m$ ,  $\sigma$  y  $\tau$  positive reals such that  $\sigma\tau\|A\|^2 \leq 1$  and  $\tau < 2\beta$ , and reals  $\bar{\lambda}$  and  $\bar{\rho}$  such that  $\lambda \in [0, \bar{\lambda}]$  and  $\rho \in ]0, \frac{\bar{\rho}}{2\alpha}]$ , where  $\alpha := \frac{2\beta}{4\beta - \tau}$ .

$$\begin{aligned} (x_w^k, \eta_w^k, r_w^k) &= (x^k, \eta^k, r^k) + \lambda(x^k - x^{k-1}, \eta^k - \eta^{k-1}, r^k - r^{k-1}) \\ \tilde{x}^{k+1} &= (\tau \partial f + I_{n \times n})^{-1}(x_w^k - \tau A^t \eta_w^k - r_w^k) \\ \tilde{r}^{k+1} &= \tau \nabla h(\tilde{x}^{k+1}) \\ \tilde{\eta}^{k+1} &= (\sigma \partial g^* + I_{m \times m})^{-1}(\eta_w^k + \sigma A(2\tilde{x}^{k+1} - x_w^k) + \sigma A r_w^k - \sigma A \tilde{r}^{k+1}) \\ (x^{k+1}, \eta^{k+1}, r^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{\eta}^{k+1}, \tilde{r}^{k+1}) + (1 - \rho)(x_w^k, \eta_w^k, r_w^k). \end{aligned}$$

Without inertia and relaxed terms and after the change of variables  $(\bar{x}^k, \bar{y}^k, \bar{z}^k) = (x^k, \eta^{k-1}, x^{k-1} - \tau A^t \eta^{k-1} - r^{k-1})$ , the last algorithm gets back to *PD3O*. Therefore

this algorithm can be seen as resulting of the inclusion of inertial and relaxed terms into *PD3O*.

Analogous, considering *MIRPD*, Form II, with matrix parameters satisfying (42) and  $V_2 = 0$ , we obtain the switching version of Inertial-Relaxed *PD3O*.

**Inertial-Relaxed Primal-Dual Three Operator, Form II**  
**(IRPD3O, Form II)**

Choose  $(x^0, y^0) = (x^{-1}, y^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m$ ,  $\sigma$  and  $\tau$  positive real parameters such that  $\sigma\tau\|A\|^2 \leq 1$  and  $\tau < 2\beta$ , and  $\bar{\lambda}$  and  $\bar{\rho}$  positive real parameters such that  $\lambda \in [0, \bar{\lambda}]$  and  $\rho \in ]0, \frac{\bar{\rho}}{2\alpha}]$ , where  $\alpha := \frac{2\beta}{4\beta - \tau}$ .

$$\begin{aligned} (x_w^k, y_w^k) &= (x^k, y^k) + \lambda(x^k - x^{k-1}, y^k - y^{k-1}) \\ \tilde{y}^{k+1} &= (\sigma\partial g^* + I_{m \times m})^{-1}(y_w^k + \sigma Ax_w^k) \\ r^{k+1} &= \nabla h(x_w^k - \tau A^t(\tilde{y}^{k+1} - y_w^k)) \\ \tilde{x}^{k+1} &= (\tau\partial f + I_{n \times n})^{-1}(x_w^k - \tau A^t(2\tilde{y}^{k+1} - y_w^k) - \tau r^{k+1}) \\ (x^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1 - \rho)(x_w^k, y_w^k). \end{aligned}$$

The two forms of the last algorithm *IRPD3O*, avoid the use of variable  $z$  compared to its source algorithm *MIRPD*, and we can clearly notice the distinction with Condat-Vü, form I and form II respectively, when  $h \neq 0$ .

Comparing now the two forms of the last algorithm *IRPD3O*, Form II avoids using the previous values  $r^{k-1}$  and  $r^k$ .

On the other hand, considering *MIRPPD*, with matrix parameters satisfying (43) and  $R_2 = 0$ ; we obtain an algorithm that can be considered as the parallel version of *PD3O*, since *MIRPPD* is the Gauss-Seidel version of *MIRPD* and the Form I of the last one is related to *PD3O*.

**Inertial-Relaxed parallel Primal-Dual Three Operator**  
**(IRPPD3O)**

Choose  $(x^0, y^0) = (x^{-1}, y^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m$ ,  $\tilde{\sigma}$  and  $\tilde{\tau}$  positive real parameters such that  $2\tilde{\sigma}\tilde{\tau}\|A\|^2 \leq 1$  and  $\tilde{\tau} < 2\beta$ , and  $\bar{\lambda}$  and  $\bar{\rho}$  positive real parameters such that  $\lambda \in [0, \bar{\lambda}]$

and  $\rho \in ]0, \frac{\bar{\rho}}{2\alpha}]$ , where  $\alpha := \frac{2\beta}{4\beta - \tilde{\tau}}$ .

$$\begin{aligned}
(x_w^k, z_w^k, y_w^k) &= (x^k, z^k, y^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}) \\
r^k &= \nabla h(x_w^k - \tilde{\tau}\tilde{\sigma}A^tAx_w^k + \tilde{\sigma}\tilde{\tau}A^tz_w^k) \\
\tilde{x}^{k+1} &= (\tilde{\tau}\partial f + I_{n \times n})^{-1}(x_w^k - \tilde{\tau}A^t(\tilde{\sigma}Ax_w^k - \tilde{\sigma}z_w^k + y_w^k) - \tilde{\tau}r^k) \\
\tilde{z}^{k+1} &= (\partial g + 2\tilde{\sigma}I_{m \times m})^{-1}(\tilde{\sigma}z_w^k + \tilde{\sigma}Ax_w^k + y_w^k) \\
\tilde{y}^{k+1} &= y_w^k + \tilde{\sigma}A\tilde{x}^{k+1} - \tilde{\sigma}\tilde{z}^{k+1} \\
(x^{k+1}, z^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1 - \rho)(x_w^k, z_w^k, y_w^k).
\end{aligned}$$

The convergence of the last algorithms follows from Propositions 7, 6 and 8 respectively.

## 6 Numerical Results

We consider the problem (commonly referred as *fused lasso*) with the least squares loss as in [24]

$$\min_x \frac{1}{2}\|Qx - b\|_2^2 + \mu_1\|x\|_1 + \mu_2\|Ax\|_1 \quad (P_{fl})$$

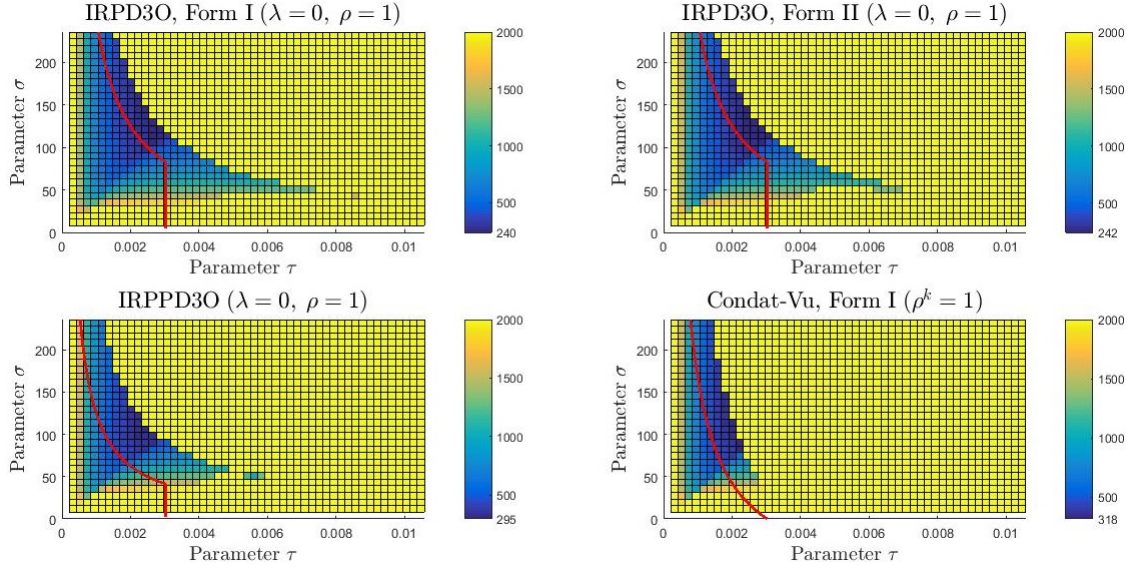
where  $A$  is a  $n \times p$  matrix defined by

$$A = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \dots & \dots & \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(p-1) \times p}.$$

We take the values  $\mu_1 = 20$  and  $\mu_2 = 200$  for the weights in the objective function. Moreover, the matrix  $Q$  and vector  $b$  are randomly generated, following Yan's paper as described in [24]. We just consider the dimension  $n = 400$  and  $q = 40$ .

Since the problem  $(P_{fl})$  has the structure of problem 1 (case  $B = -I$ ), we apply algorithm *IRPD3O* and *IRPPD3O* and we compare them with Condat-Vũ's and Chambolle-Pock's algorithms. For the evaluation, we consider the primal error, i.e.  $\|x^k - x^*\|$ , where  $x^*$  is approximated as the primal value of the 18000-iteration of *PD3O* algorithm.

We show first on Figure 1 the number of iterations needed to obtain a primal error less than  $10^{-6}$  for different values of the parameters  $\tau$  and  $\sigma$ , without inertial-relaxed parameters ( $\alpha = 0$  and  $\rho = 1$ ). We have plotted a red line to show the theoretical limits of convergence ( $\sigma\tau\|A\|^2 \leq 1$ ,  $\tau < \frac{2}{\|Q\|^2}$ , for *IRPD3O*, and  $2\sigma\tau\|A\|^2 \leq 1$ ,  $\tau < \frac{2}{\|Q\|^2}$ , for *IRPPD3O*). We consider a maximum number of 2000 iterations for each fixed pair  $(\tau, \sigma)$ .

Figure 1: Number of iterations for an error of  $10^{-6}$ 

The next figure shows the number of iterations (2000 as maximum) needed to obtain a primal error less than  $10^{-6}$ , for each  $(\sigma(\tau), \tau)$  in the curved red line ( $\sigma(\tau) = \frac{1}{\tau\|A\|^2}$ , for *IRPD3O*, and  $\sigma(\tau) = \frac{1}{2\tau\|A\|^2}$ , for *IRPPD3O*). In figure 2, we change the parameter  $\tau$  and the relaxation parameter  $\rho$  without inertial term ( $\lambda = 0$ ). In figure 3, we change the parameter  $\tau$  and the inertial term  $\lambda$  without relaxation ( $\rho = 1$ ).

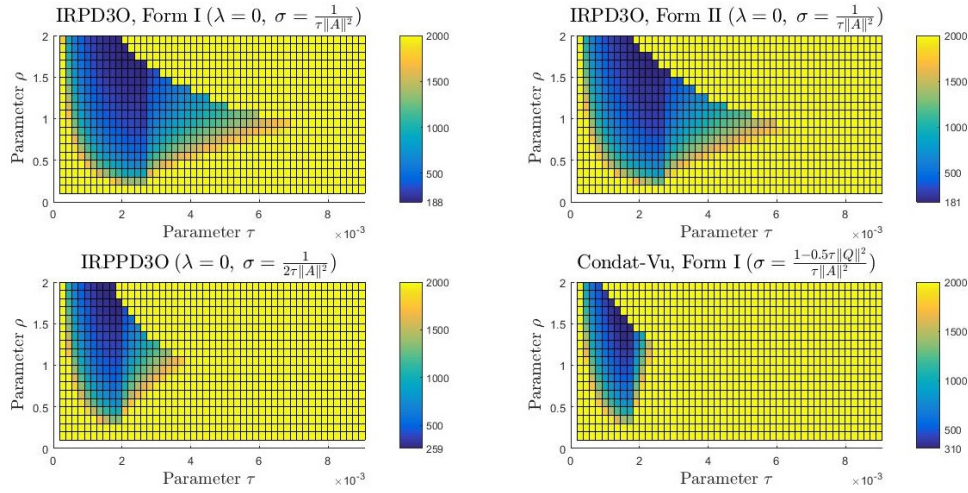


Figure 2: Varying the relaxation parameter

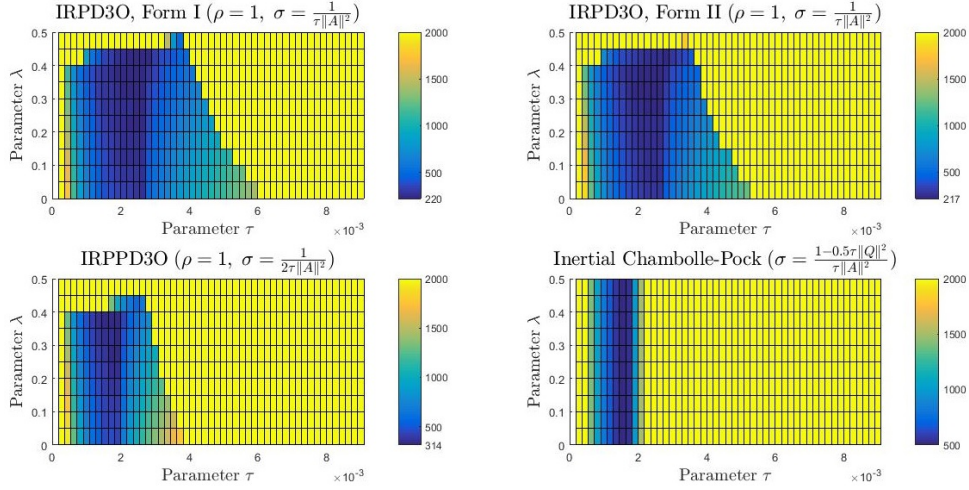
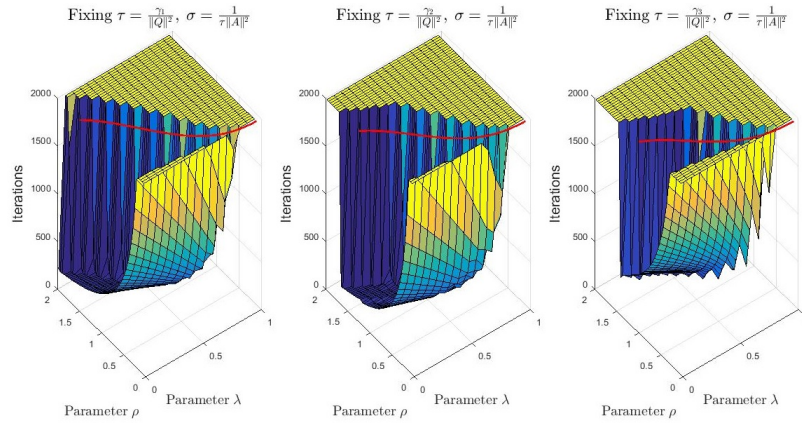


Figure 3: Varying the inertial parameter

Finally we change the inertial and relaxation parameters, considering the parameter belonging to the curved red line of figure 1

$$\tau = \frac{\gamma_i}{\|Q\|^2} \quad y \quad \sigma = \frac{1}{\tau \|A\|^2}$$

choosing  $\gamma_1 = 1$ ,  $\gamma_2 = 1.5$  y  $\gamma_3 = 1.99$ . We consider the number of iterations (2000 as maximum) needed to obtain a primal error less than  $10^{-6}$ , also we plot a red line for the theoretical bounds.

Figure 4: Varying  $\lambda$  and  $\rho$ 

We observe that the three variants of *PD3O* have a larger theoretical area of convergence. Even the parallel version has a light advantage with respect to Condat–Vũ Algorithms. Also notice that the tuning of the Relaxation parameter yields more impact on the convergence of the algorithms than the Inertial parameter. Finally the numerical results show the necessity to investigate the possibility of extending



that theoretical area of convergence, similarly to the study given in [17] when the model does not have a smooth part.

## 7 Conclusions

Associated with the composite model based on three types of functions, we have obtained two new averaged splitting maps : the Gauss-Seidel type that generalizes the Davis-Yin averaged map, and the Jacobi type that is a generalized parallel version of Davis-Yin averaged map. Then, similarly to the construction of ADMM from the Douglas-Rachford map, but also considering a variant of the fixed point algorithm, we have obtained three new splitting algorithms from these averaged maps, including in all of them Inertial-Relaxed parameters. Choosing special scaling matrices parameters allows us to obtain algorithmic variants of *PD3O*, which we have compared numerically, showing the high sensitivity of the rate of convergence with respect to the relaxation parameters, and also noticing the advantage of the variants of *PD3O* compared to Condat-Vũ Algorithms.

Observe that parallel implementations of a few algorithms have been mentioned but it remains to confirm their respective speedups on real-life cases. Moreover, the numerical experimentation has revealed the high sensitivity of the performance in terms of number of iterations with respect to the tuning of the different parameters. Corresponding adaptive strategies to allow dynamic changes for the proximal and inertial parameters are currently on study.

## References

- [1] F. Alvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-valued analysis*, 9:3–11, 2001.
- [2] M. Alves, J.E. Eckstein, M. Geremia, and J.G. Melo. Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms. *Comput. Optimization and Appl.*, 75:389–422, 2020.
- [3] M. Alves and M. Geremia. Iteration complexity of an inexact Douglas-Rachford and a Douglas-Rachford-Tseng Forward-Backward four-operator splitting method for solving monotone inclusions. *Numer. Algorithms*, 82:263–295, 2019.
- [4] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer V., 2011.
- [5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning with the alternating direction method of multipliers. In M. Jordan, editor, *Foundations and Trends in Machine Learning*, volume 3, pages 1–122. 2011.

- [6] L.M. Briceño Arias, J. Deride, and S. Lopez-Rivera. Primal-dual partial inverse splitting for constrained monotone inclusions. arXiv:2007.01983[Math.OC], 2020.
- [7] L.M. Briceño Arias and F. Roldan. Primal-dual splittings as fixed-point iterations in the range of linear operators. arXiv:1910.02329v1[math.OC], 2019.
- [8] A. Chambolle and T. Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.
- [9] L. Condat. A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *J. Optimization Theory and Applications*, 158:460–479, 2013.
- [10] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi. Proximal splitting algorithms : a tour of recent advances with new twists. arXiv:1912.00137[math.OC], 2020.
- [11] D. Davis and W. Yin. Convergence rate analysis of several splitting schemes. In R. Glowinski, S.J. Osher, and W. Yin, editors, *Splitting Methods in Communication, Imaging, Science and Engineering*, pages 115–163. Springer International, 2016.
- [12] D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set Valued Var. Anal.*, 25(4):829–858, 2017.
- [13] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming*, 55:293–318, 1992.
- [14] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. *Computer Math. Appl.*, 2:17–40, 1976.
- [15] O. Güler. New proximal point algorithms for convex optimization. *SIAM J. on Optimization*, 2:649–664, 1992.
- [16] M. Li, D. Sun, and K.C. Toh. A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. *SIAM J. on Optimization*, 26:922–950, 2016.
- [17] F. Ma, B. He, and X. Yuan. Optimally linearizing the alternating direction method of multipliers for convex programming. *Computational Optimization and Applications*, 75:361–388, 2020.
- [18] Y. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence  $o(1/k^2)$ . *Dokl. Akad. Nauk SSSR*, 269:543–547, 1983.
- [19] E. Ore Albornoz, P. Mahey, and E. Ocaña Anaya. A unified splitting algorithm for composite monotone inclusions. *J. of Convex Analysis*, 27(2):893–922, 2020.

- [20] F. Pedregosa and G. Gidel. Adaptive three-operator splitting. In *Proceedings of 35th Int. Conf. on Machine Learning*, pages 4085–4094, 2018.
- [21] R. T. Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.
- [22] R. Shefi and M. Teboulle. Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. *SIAM J. on Optimization*, 24:269–297, 2014.
- [23] B.C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667–681, 2013.
- [24] M. Yan. A new primal-dual algorithm for minimizing the sum of three functions with a linear operator. *J. Sci. Comput.*, 76:1698–1717, 2018.