Open Journal of Mathematical Optimization

Ernesto Oré, Philippe Mahey & Eladio Ocaña

Inertial-relaxed splitting for composite monotone inclusions

Volume 4 (2023), article no. 3 (20 pages)

https://doi.org/10.5802/ojmo.22

Article submitted on April 21, 2022, revised on November 7, 2022, accepted on January 17, 2023.

Open Journal of Mathematical Optimization is a member of the Centre Mersenne for Open Scientific Publishing

http://www.centre-mersenne.org/

e-ISSN: 2777-5860

Inertial-relaxed splitting for composite monotone inclusions

Ernesto Oré
IMCA, Instituto de Matemática y Ciencias Afines, Universidad Nacional de Ingeniería, Lima, Perú
eore@imca.edu.pe

Philippe Mahey
LIMOS, CNRS, Université Clermont Auvergne, France
philippe.mahey@isima.fr

Eladio Ocaña
IMCA, Instituto de Matemática y Ciencias Afines, Universidad Nacional de Ingeniería, Lima, Perú
eocana@imca.edu.pe

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 flavor, 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.

Digital Object Identifier 10.5802/ojmo.22

Keywords Operator splitting methods, Convex composite optimization.

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 other ones coupled by linear operators:

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

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 β positive.

The above model can be seen as a functional version of the general inclusion problem of finding a zero of the sum of three monotone operators defined on a Hilbert space, one of them being co-coercive (see Davis and Yin [12]). The role of the linear operators A and B in the coupling subspace intends to cover a broad scope of current applications, justifying the reference to "composite inclusions" in the title.

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 $\|a\|_M^2 = a^T Ma$ 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, including unified variants of *ADMM* which 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 η^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, $V_1 - A^tMA$ and $V_2 - B^tMB$ 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 $h = 0$):

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

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

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

The convergence of the previous algorithm is obtained from its relationship with the fixed point formulation applied to particular 1/2-averaged maps[1] (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 a 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.

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

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

Both algorithms include the relaxation parameter ρ^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 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 λ^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 (ρ^k) and inertial (λ^k) parameters in the primal updates, thus inspired by Alves et al. [2], where the authors consider a Relative-error Inertial-Relaxed variant of the Proximal Point Algorithm 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 $(\widetilde{x}^k, \widetilde{z}^k, \widetilde{y}^k) = (x^{k+1}, z^k, y^k)$, we can reobtain Condat–Vũ algorithm, form II. Analogously, interchanging the order of f and g in the model, we obtain the switched version, i.e. Condat–Vũ algorithm, form I.

Alternately, 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.

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

Observe that in the case $h = 0$ and after the change of variables $(\widetilde{x}^k, \widetilde{y}^k) = (x^k, y^{k+1})$, *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 [17] 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 unify *PD3O* (getting switched parallel versions) 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 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

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

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 $G_D^{\mathcal{L}} = D(\mathcal{L} + D^tD)^{-1}D^t$ which is $\frac{1}{2}$−averaged, does not separate the map $\nabla h(x)$. Alternatively we can obtain an equivalent problem to ($V_{\mathcal{L}}$) which is amenable for the application of the Davis–Yin α−averaged map [11], where $\alpha \in (\frac{1}{2}, 1)$, building formally two distinct α−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 $\overline{Q}$ and $\widehat{Q}$ defined below in (11). On the other hand and regardless of the Davis–Yin map, we can obtain another class of α−averaged map with a parallel splitting structure, developed below in (16).

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 $V_1 + A^tMA$ is positive definite. Using the notation $w = \left[\begin{smallmatrix} x \\ z \\ y \end{smallmatrix}\right]$ for the primal-dual space variables belonging to $\mathbb{R}^{n+p+m}$, problem ($V_{\mathcal{L}}$) can be written as*

$$0 \in \left(\bar{A}S^{-1}\bar{A}^t\right)^{-1}(w) + \left[(-\bar{B})T^{-1}(-\bar{B}^t)\right]^{-1}(w) + \bar{A}\left(\bar{A}^t\bar{A}\right)^{-1} C \left(\bar{A}^t\bar{A}\right)^{-1}\bar{A}^t(w), \tag{3}$$

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

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

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, ξ, χ, 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 $w^* \in \mathbb{R}^{n+p+m}$,

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

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 $\bar{A}Q\bar{A}^{\,t}$, and it gives (now in $\mathbb{R}^{n+p}$):

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

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

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

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

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

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

Observe now that the inclusion (3) obtained in Proposition 1 can be seen as a 3-operator setting, amenable to the application of Davis–Yin's framework [11], resumed below with $\bar{S}, \bar{T}, \bar{C}$ being the three operators by

$$0 \in \bar{S}(w) + \bar{T}(w) + \bar{C}(w). \tag{4}$$

The corresponding Davis–Yin's operator, with parameter $\gamma > 0$, associated with the above inclusion, is defined as

$$\mathcal{D}_\gamma := I - J^{\gamma\bar{T}} + J^{\gamma\bar{S}}(2J^{\gamma\bar{T}} - I - \gamma\bar{C}(J^{\gamma\bar{T}})), \tag{5}$$

where $J^{\gamma\bar{T}}$ denotes the classical resolvent $(I + \gamma\bar{T})^{-1}$. In that sense, problem (4) is equivalent to finding a fixed point of operator $\mathcal{D}_\gamma$. When $\bar{S}$ and $\bar{T}$ are maximal monotone operators, and $\bar{C}$ is $\beta-$cocoercive[2], that map $\mathcal{D}_\gamma$ has nice properties: indeed, it is $\frac{2\beta}{4\beta-\gamma}-$averaged (provided $\gamma < 2\beta$) and of full domain, properties ensuring the convergence of fixed point algorithms applied to this map.

Back to the inclusion (3) obtained in Proposition 1, the Davis–Yin map with scalar parameter $\gamma = 1$ associated with that inclusion generates two distinct maps denoted by G_1 and G_2 (where G_2 is obtained switching the position of $\left(\bar{A}S^{-1}\bar{A}^{\,t}\right)^{-1}$ and $\left[(-\bar{B})T^{-1}(-\bar{B}^{\,t})\right]^{-1}$):

$$G_1 := I - J^{\bar{S}} + J^{\bar{T}}(2J^{\bar{S}} - I - \bar{C}(J^{\bar{S}})) \text{ and } G_2 := I - J^{\bar{T}} + J^{\bar{S}}(2J^{\bar{T}} - I - \bar{C}(J^{\bar{T}})), \tag{6}$$

where $\bar{S} := (\bar{A}S^{-1}\bar{A}^{\,t})^{-1}$, $\bar{T} = ((-\bar{B})T^{-1}(-\bar{B}^{\,t}))^{-1}$ and $\bar{C} := \bar{A}\left(\bar{A}^{\,t}\bar{A}\right)^{-1}C\left(\bar{A}^{\,t}\bar{A}\right)^{-1}\bar{A}^{\,t}$.

On the other hand, the positive definiteness assumption on $V_1 + A^tMA$ in Proposition 1 (which is equivalent to $\bar{A}$ being injective) implies that $\bar{S}$ is maximal monotone and $\bar{C}$ is a $\frac{\beta}{\|(V_1+A^tMA)^{-1}\|}-$cocoercive map (since h is convex and $(\frac{1}{\beta})$-Lipschitz-differentiable). Similarly, the map $\bar{T}$ is too maximal monotone under the additional assumption that $V_2 + B^tMB$ is positive definite (equivalent to $\bar{B}$ injective). So, under both conditions (which will be denoted by *Hypo-*$\bar{M}$):

$V_1 + A^tMA$ and $V_2 + B^tMB$ are positive definite,

we get an alternative way to write both G_1 and G_2 as shown in the next proposition:

▶ **Proposition 2.** *Under Hypo-*$\bar{M}$*, the maps* G_1 *and* G_2 *are single valued, they apply* $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ *into itself, and can be rewritten as:*

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

where for G_1

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

$$\widetilde{r} = (V_1 + A^tMA)^{-1}\nabla h(\widetilde{x})$$

$$\widetilde{z} = (\partial g + V_2 + B^tMB)^{-1}\left(V_2^{\frac{1}{2}}z + B^tMA(\widetilde{r} - 2\widetilde{x}) + B^tM^{\frac{1}{2}}y\right).$$

---

[2] An operator C on a Hilbert space H is β-cocoercive (or β-inverse-strongly monotone),$\beta > 0$, if $\langle Cx-Cy, x-y\rangle \geq \beta\|Cx-Cy\|^2, \forall x,y \in H$.

and for G_2

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

Proof. For a given operator Γ and an injective matrix K, the next expression can be easily deduced (see Proposition 23.25 (ii) in [4])

$$J^{(K\Gamma^{-1}K^t)^{-1}} = K(\Gamma + K^tK)^{-1}K^t. \tag{7}$$

Under assumption *Hypo-*$\overline{M}$, both matrices $\overline{A}$ and $\overline{B}$ are injective, and then considering (7) we can rewrite the resolvent of $\overline{S}$ and $\overline{T}$ involved in (6), as $J^{\overline{S}} = \overline{A}(S + \overline{A}^{\,t}\overline{A})^{-1}(\overline{A}^{\,t})$ and $J^{\overline{T}} = -\overline{B}(T + \overline{B}^{\,t}\overline{B})^{-1}(-\overline{B}^{\,t})$, and using that

$$\begin{aligned}
(T + \overline{B}^t\overline{B})^{-1}(p_1, p_2) &= (p_1, (\partial g + V_2 + B^tMB)^{-1}(p_2))\\
(S + \overline{A}^t\overline{A})^{-1}(p_1, p_2) &= ((\partial f + V_1 + A^tMA)^{-1}(p_1), p_2)
\end{aligned}$$

we get

$$J^{\overline{T}}(x, z, y) = \begin{pmatrix} x \\ V_2^{\frac{1}{2}}\overline{z} \\ -M^{\frac{1}{2}}B\overline{z} \end{pmatrix}, \quad J^{\overline{S}}(x, z, y) = \begin{pmatrix} V_1^{\frac{1}{2}}\overline{x} \\ z \\ M^{\frac{1}{2}}A\overline{x} \end{pmatrix} \tag{8}$$

where $\overline{z} = (\partial g + V_2 + B^tMB)^{-1}(V_2^{1/2}z - B^tM^{1/2}y)$ and $\overline{x} = (\partial f + V_1 + A^tMA)^{-1}(V_1^{1/2}x + A^tM^{1/2}y)$. On the other hand, we get also

$$\overline{C}(x, z, y) = \begin{pmatrix} V_1^{\frac{1}{2}}(V_1 + A^tMA)^{-1}\nabla h[(V_1 + A^tMA)^{-1}(V_1^{1/2}x + A^tM^{1/2}y)] \\ 0 \\ M^{\frac{1}{2}}A(V_1 + A^tMA)^{-1}\nabla h[(V_1 + A^tMA)^{-1}(V_1^{1/2}x + A^tM^{1/2}y)] \end{pmatrix} \tag{9}$$

Then combining (8) and (9) in (6), we deduce the result. ◀ ◀

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 $\mathrm{sol}(V_{\mathcal{L}})$, are

$$\left\{\overline{Q}\begin{pmatrix} x^* \\ z^* \\ y^* \end{pmatrix} - \begin{pmatrix} V_1^{\frac{1}{2}}(V_1 + A^tMA)^{-1}\nabla h(x^*) \\ 0 \\ M^{\frac{1}{2}}A(V_1 + A^tMA)^{-1}\nabla h(x^*) \end{pmatrix} : (x^*, z^*, y^*) \in \mathrm{sol}(V_{\mathcal{L}})\right\}, \tag{10}$$

and

$$\widehat{Q}(\mathrm{sol}(V_{\mathcal{L}})) := \{(V_1^{\frac{1}{2}}x^*, V_2^{\frac{1}{2}}z^*,\ M^{\frac{1}{2}}Ax^* + M^{-\frac{1}{2}}y^*) : (x^*, z^*, y^*) \in \mathrm{sol}(V_{\mathcal{L}})\}, \tag{11}$$

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

$$\overline{Q} = \begin{pmatrix} V_1^{\frac{1}{2}} & 0 & 0 \\ 0 & V_2^{\frac{1}{2}} & 0 \\ 0 & -M^{\frac{1}{2}}B & -M^{-\frac{1}{2}} \end{pmatrix} \text{ and } \widehat{Q} = \begin{pmatrix} V_1^{\frac{1}{2}} & 0 & 0 \\ 0 & V_2^{\frac{1}{2}} & 0 \\ M^{\frac{1}{2}}A & 0 & M^{-\frac{1}{2}} \end{pmatrix}. \tag{12}$$

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.

The next proposition resumes the averageness properties of G_1 and G_2.

▶ **Proposition 3.** *Under Hypo-$\overline{M}$. Considering that $\|(V_1 + A^tMA)^{-1}\| \in]0, 2\beta[$. Then, G_1 and G_2 are $\alpha-$averaged with full domain, where*

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

Proof. The fullness of the domains of G_1 and G_2 is deduced from the maximality of ∂f and ∂g and the fullness of the domain of ∇h.

We now show the $\alpha-$averagedness of G_1 and G_2. As said before, since $V_1 + A^tMA$ and $V_2 + B^tMB$ are positive definite, it holds that $\overline{A}$ and $\overline{B}$ are injective, following the maximal monotonicity of $(\overline{A}S^{-1}\overline{A}^{\,t})^{-1}$ and $((-\overline{B})T^{-1}(-\overline{B}^{\,t}))^{-1}$. On the other hand, the cocoercivity property of $\overline{C}$ follows from the $\beta-$cocoercivity of C:

$$\begin{aligned}\langle C\left(\overline{A}^{\,t}\overline{A}\right)^{-1}\overline{A}^{\,t}u - C\left(\overline{A}^{\,t}\overline{A}\right)^{-1}\overline{A}^{\,t}v, \left(\overline{A}^{\,t}\overline{A}\right)^{-1}\overline{A}^{\,t}u - \left(\overline{A}^{\,t}\overline{A}\right)^{-1}\overline{A}^{\,t}v\rangle \\ \geq \beta\|C\left(\overline{A}^{\,t}\overline{A}\right)^{-1}\overline{A}^{\,t}u - C\left(\overline{A}^{\,t}\overline{A}\right)^{-1}\overline{A}^{\,t}v\|^2 = \beta\|\nabla h(D^tD)^{-1}D^tu - \nabla h(D^tD)^{-1}D^tv\|^2 \\ \geq \widehat{\beta}\|D(D^tD)^{-1}\nabla h(D^tD)^{-1}D^tu - D(D^tD)^{-1}\nabla h(D^tD)^{-1}D^tv\|^2 = \widehat{\beta}\|\overline{C}u - \overline{C}v\|^2,\end{aligned}$$

where $D^t = \begin{pmatrix} V_1^{1/2} & A^tM^{1/2}\end{pmatrix}$ and $\widehat{\beta} := \frac{\beta}{\|(D^tD)^{-1}\|} = \frac{\beta}{\|(V_1 + A^tMA)^{-1}\|}$.

From these properties and the definition in (6), the relation above shows that G_1 and G_2 hold the averaged properties of Davis–Yin 3−operator, provided $1 < 2\widehat{\beta}$ (which is satisfied by hypothesis). ◀

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:

$$\text{Minimize}_{(x,z)}(f(x) + g(z)) + \delta_{\{(x,z):Ax+Bz=0\}}(x, z) + h(x) \tag{13}$$

where δ_Ω denotes the characteristic function of a set Ω, i.e. $\delta_\Omega(x) = 0$, if $x \in \Omega$ and $\delta_\Omega(x) = +\infty$, else.

Then, applying Davis–Yin algorithm to the functional setting (13), one obtains 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 $\{(x, z) : Ax + Bz = 0\}$.

Instead of using the methodology resumed in the former section, 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.

Unlike G_1 and G_2 which were obtained from Davis–Yin's 3 operator splitting, the construction of the new required average map is deduced from $\mathcal{P} := \widehat{S}(L + \widehat{S}^t\widehat{S})^{-1}\widehat{S}^t$, an $\frac{1}{2}-$averaged map as developed below, associated with *PPDS* for the Jacobi version (see [19]) with $h = 0$ (recall that L, defined in (2), is the maximal monotone operator associated to the subdifferential of the Lagrangian of the initial problem). In Proposition 4 below, one can notice the adaptation of $\mathcal{P}$ to our general problem (1).

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 $R_1 + 2A^tMA$ and $R_2 + 2B^tMB$ 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

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

where

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

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

$$\widehat{S}(\text{sol}(V_{\mathcal{L}})) = \{(R_1^{\frac{1}{2}}x^*, R_2^{\frac{1}{2}}z^*, M^{\frac{1}{2}}Ax^* - M^{\frac{1}{2}}Bz^*, M^{-\frac{1}{2}}y^*) : (x^*, z^*, y^*) \in \text{sol}(V_{\mathcal{L}})\} \tag{14}$$

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

$$\widehat{S} = \begin{pmatrix} R_1^{\frac{1}{2}} & 0 & 0 \\ 0 & R_2^{\frac{1}{2}} & 0 \\ M^{\frac{1}{2}}A & -M^{\frac{1}{2}}B & 0 \\ 0 & 0 & M^{-\frac{1}{2}} \end{pmatrix}. \tag{15}$$

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

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

where

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

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

▶ **Proposition 4.** *Under the Hypo-M hypotheses, it holds that*

$$G_3 = \widehat{S}\mathcal{G}\widehat{S}^t. \tag{16}$$

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

$$L\left(\mathcal{G}u\right) + \widehat{S}^t\widehat{S}\left(\mathcal{G}u\right) \ni u + \begin{pmatrix} -\nabla h(\eta) \\ 0 \\ 0 \end{pmatrix} \tag{17}$$

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

Proof. The equivalence $G_3 = \widehat{S}\mathcal{G}\widehat{S}^t$ is easily found using the definitions of these maps. To show the second part, given $u := (\widehat{x}, \widehat{z}, \widehat{y})$, considering $(x, z, \nu) = \mathcal{G}u$, it holds that

$$\partial f(x) + A^t\nu + (R_1 + A^tMA)x - A^tMBz \ni \widehat{x} - \nabla h\left((R_1 + 2A^tMA)^{-1}(\widehat{x})\right) \tag{18}$$

$$\partial g(z) + B^t\nu - B^tMAx + (R_2 + B^tMB)z \ni \widehat{z} \tag{19}$$

$$-Ax - Bz + M^{-1}\nu = \widehat{y}$$

Then the last relations show the desirable result. ◀

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

▶ **Proposition 5.** *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 make the formulation easier, 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 $\mathcal{G}\widehat{S}^t\widehat{S}$ using μ_1 and μ_2 are, for $i = 1, 2$:

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

where

$$x_i := (\partial f + R_1 + 2A^tMA)^{-1}\left((R_1 + A^tMA)\widehat{x}_i - A^tMB\widehat{z}_i - A^t\widehat{y}_i - \nabla h(\eta_i)\right)$$
$$\eta_i := (R_1 + 2A^tMA)^{-1}\left((R_1 + A^tMA)\widehat{x}_i - A^tMB\widehat{z}_i\right)$$
$$z_i := (\partial g + R_2 + 2B^tMB)^{-1}\left((R_2 + B^tMB)\widehat{z}_i - B^tMA\widehat{x}_i - B^t\widehat{y}_i\right).$$

From the inclusion (17) and considering u equal to $\widehat{S}^t\widehat{S}\mu_1$ or $\widehat{S}^t\widehat{S}\mu_2$, then using the monotonicity of operator L i.e.

$$\left\langle \mathcal{G}\widehat{S}^t\widehat{S}\mu_1 - \mathcal{G}\widehat{S}^t\widehat{S}\mu_2, \widehat{S}^t\widehat{S}(\mu_1 - \mathcal{G}\widehat{S}^t\widehat{S}\mu_1) - \widehat{S}^t\widehat{S}(\mu_2 - \mathcal{G}\widehat{S}^t\widehat{S}\mu_2)\right\rangle + \left\langle \mathcal{G}\widehat{S}^t\widehat{S}\mu_1 - \mathcal{G}\widehat{S}^t\widehat{S}\mu_2, \begin{pmatrix} -\nabla h(\eta_1) + \nabla h(\eta_2) \\ 0 \\ 0 \end{pmatrix}\right\rangle \geq 0$$

and then using expression (16), we have:

$$\left\langle G_3\widehat{S}\mu_1 - G_3\widehat{S}\mu_2, (\widehat{S}\mu_1 - G_3\widehat{S}\mu_1) - (\widehat{S}\mu_2 - G_3\widehat{S}\mu_2)\right\rangle + \langle x_1 - x_2, \nabla h(\eta_2) - \nabla h(\eta_1)\rangle \geq 0 \tag{20}$$

so that, rewriting the first term, we obtain the following inequality:

$$\|\widehat{S}\mu_1 - \widehat{S}\mu_2\|^2 - \|G_3\widehat{S}\mu_1 - G_3\widehat{S}\mu_2\|^2 - \|\widehat{S}\mu_1 - G_3\widehat{S}\mu_1 - \widehat{S}\mu_2 + G_3\widehat{S}\mu_2\|^2 + 2\langle x_1 - x_2, \nabla h(\eta_2) - \nabla h(\eta_1)\rangle \geq 0 \tag{21}$$

Now we find an appropriate upper bound for $\langle x_1 - x_2, \nabla h(\eta_2) - \nabla h(\eta_1)\rangle$ considering $\|\widehat{S}\mu_1 - G_3\widehat{S}\mu_1 - \widehat{S}\mu_2 + G_3\widehat{S}\mu_2\|$. 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 ∇h is co-coercive and using Cauchy inequality, it holds that

$$\begin{aligned}\langle x_1 - x_2, \nabla h(\eta_2) - \nabla h(\eta_1)\rangle &\leq -\beta\|\nabla h(\eta_2) - \nabla h(\eta_1)\|^2 + \frac{1}{4\beta}\|x_1 - x_2 + \eta_2 - \eta_1\|^2 + \beta\|\nabla h(\eta_2) - \nabla h(\eta_1)\|^2 \\ &= \frac{1}{4\beta}\|x_1 - x_2 + \eta_2 - \eta_1\|^2\end{aligned} \tag{22}$$

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

$$\left\|\widehat{S}\left[\mu_1 - \mathcal{G}\widehat{S}^t\widehat{S}\mu_1 - \mu_2 + \mathcal{G}\widehat{S}^t\widehat{S}\mu_2\right]\right\|^2 = \|(p_1, p_2, p_3, p_4)\|^2 \tag{23}$$

where

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

Denoting $K = \begin{pmatrix} R_1^{\frac{1}{2}} & A^tM^{\frac{1}{2}} & A^tM^{\frac{1}{2}} \end{pmatrix}$, 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 (23), we obtain that

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

Therefore from (22), we obtain the desired upper bound

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

Finally using this upper bound in (20) we obtain that

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

where

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

Then, since for any $s, s' \in \mathbb{R}^q$ there exists $u_1, u_2 \in \mathbb{R}^r$ such that $\widehat{S}^t\widehat{S}u_1 = \widehat{S}^t s$ and $\widehat{S}^t\widehat{S}u_2 = \widehat{S}^t s'$, from the last inequality and (16) we get that G_3 is $\alpha-$averaged. ◀

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 of a 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)*; it includes three parameters (θ driving the relative error measure, λ the inertial parameter, and ρ the relaxation parameter:

▶ 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 $k = 0, 1, \ldots$ **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

$$v^k \in T(\tilde{z}^k), \quad \|c^k v^k + \tilde{z}^k - w^k\|^2 \leq \theta^2 \left(\|\tilde{z}^k - w^k\|^2 + \|c^k v^k\|^2\right) \tag{24}$$

- 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

Observe first that, without relative error ($\theta = 0$), it can be checked easily that $\tilde{z}^k = J_{c^kT}(w^k)$. In the general case, too, *RIRHPP* can be rewritten in terms of the resolvent of T:

Let $J^{c^kT} = (I_{r\times r} + c^kT)^{-1}$, then subproblem (24) is equivalent to solve (with $c^kv^k = \varsigma^k - \tilde{z}^k + w^k$).

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

$$\tilde{z}^k = J^{c^kT}(w^k + \varsigma^k) \quad \|\varsigma^k\|^2 \leq \theta^2 \left(\|\tilde{z}^k - w^k\|^2 + \|w^k + \varsigma^k - \tilde{z}^k\|^2\right). \tag{25}$$

so that this algorithm can be interpreted as a variant of a fixed point method applied to the resolvent of T. Then 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.

This fixed point variant can also be used applied to an $\alpha-$average map. The following lemma shows that, to find a fixed point of an $\alpha-$averaged map, it is equivalent to find a fixed point of a $1-$co-coercive map which is constructed easily from the $\alpha-$averaged one, assuming that we know the parameter α.

▶ **Lemma 6.** *Set $\alpha \in]0, 1[$. It holds that F is an $\alpha-$averaged map if and only if $(1-\frac{1}{2\alpha})I + \frac{1}{2\alpha}F$ is a $1-$co-coercive map. Moreover these maps have the same fixed points.*

Proof. Let N be a nonexpansive map associated with F.

$$F = (1-\alpha)I + \alpha N \;\text{ if only if }\; \left(1 - \frac{1}{2\alpha}\right)I + \frac{1}{2\alpha}F = \frac{1}{2}(I + N)$$

which means that the transformed map is 1/2-averaged (or firmly nonexpansive) and equivalently 1-co-coercive.

Also $x^* = F(x^*)$ if and only if $x^* = (1 - \frac{1}{2\alpha})x^* + \frac{1}{2\alpha}Fx^*$. ◀

Thus, given the problem of finding a fixed point of an $\alpha-$averaged map F with full domain, we consider alternately $(1-\frac{1}{2\alpha})I+\frac{1}{2\alpha}F$ a $1-$co-coercive map with full domain, which has the same fixed points. Thus, one can extend algorithm *RIRHPP* for an $\alpha-$averaged map with full domain. In summary, one obtains the following algorithm where, for sake of 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 $\widehat{\rho} := \frac{\bar{\rho}}{2\alpha}$, where α is the averagedness parameter of F.

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

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

$$w^k = z^k + \lambda(z^k - z^{k-1}) \tag{26}$$

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

$$z^{k+1} = (1-\rho)w^k + \rho F(w^k). \tag{27}$$

end for

We consider similar conditions on the bounds $\bar{\lambda}$ and $\bar{\rho}$ as those 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

$$\bar{\rho} = \frac{2(\bar{\lambda}-1)^2}{2(\bar{\lambda}-1)^2 + 3\bar{\lambda} - 1}. \tag{28}$$

In the case that no inertial term is used ($\lambda = 0$), we bound the relaxation parameter by $\bar{\rho} < 2$ so that $\rho < \frac{1}{\alpha}$ (which is known to guarantee convergence in that case).

▶ **Proposition 7.** *Set $\bar{\lambda}$ and $\bar{\rho}$ satisfying H1. Given F an $\alpha-$averaged 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.*

The convergence of *IRFP* can be directly derived from [2] since the algorithm can be seen as an application of *RIRHPP* to a special $1-$co-coercive map as shown in Lemma 3.1 above, but we give in the annex an equivalent direct proof.

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 $V_1 + A^tMA$ and

$V_2 + B^tMB$ 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

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

$$(x_w^k, z_w^k, y_w^k, r_w^k) = (x^k, z^k, y^k, r^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}, r^k - r^{k-1}) \tag{29}$$

$$\widetilde{x}^{k+1} = (\partial f + V_1 + A^tMA)^{-1}(V_1 x_w^k - A^tMBz_w^k - A^t y_w^k - V_1 r_w^k) \tag{30}$$

$$\widetilde{y}^{k+1} = y_w^k + MA\widetilde{x}^{k+1} + MBz_w^k \tag{31}$$

$$\widetilde{r}^{k+1} = (V_1 + A^tMA)^{-1}\nabla h(\widetilde{x}^{k+1}) \tag{32}$$

$$\widetilde{z}^{k+1} = (\partial g + V_2 + B^tMB)^{-1}(V_2 z_w^k - B^tMA\widetilde{x}^{k+1} - B^t\widetilde{y}^{k+1} + B^tMA\widetilde{r}^{k+1}) \tag{33}$$

$$(x^{k+1}, z^{k+1}, y^{k+1}, r^{k+1}) \;=\; \rho(\widetilde{x}^{k+1}, \widetilde{z}^{k+1}, \widetilde{y}^{k+1}, \widetilde{r}^{k+1}) + (1-\rho)(x_w^k, z_w^k, y_w^k, r_w^k) \tag{34}$$

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

$$\bar{Q}\xi_w^k = \bar{Q}\xi^k + \lambda(\bar{Q}\xi^k - \bar{Q}\xi^{k-1}) \quad \text{and} \quad \bar{Q}\xi^{k+1} = (1-\rho)\bar{Q}\xi_w^k + \rho G_1(\bar{Q}\xi_w^k). \tag{35}$$

This relation allows us to obtain the following convergence result.

▶ **Proposition 8.** *Set $\bar{\lambda}$ and $\bar{\rho}$ satisfying H1. and all scaling matrices satisfying the same conditions as in Proposition 2. If* $\mathrm{sol}(V_{\mathcal{L}})$ *is nonempty, then building the sequences* $(\widetilde{x}^k, \widetilde{r}^k, \widetilde{z}^k, \widetilde{y}^k)$ *in* (29)–(34)*, it holds that* $(\widetilde{x}^k, \widetilde{z}^k, \widetilde{y}^k - MA\widetilde{r}^k)$ *converge to some element of* $\mathrm{sol}(V_{\mathcal{L}})$.

Proof. From relation (35), we observe that this algorithm is related to *IRFP* applied to operator G_1 which, from Proposition 1, is $\alpha-$averaged. Then by Proposition 7 and relation (10), 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 $(x^*, z^*, y^*) \in \mathrm{sol}(V_{\mathcal{L}})$ and $W = (V_1 + A^tMA)^{-1}$. Since $(\partial f + V_1 + A^tMA)^{-1}$ is single-valued and continuous, from (30) we have that $\{\widetilde{x}^k\}$ converges to $(\partial f + V_1 + A^tMA)^{-1}(V_1(x^* - W\nabla h(x^*) + A^tMA(x^* - W\nabla h(x^*)) - A^t y^*) = x^*$; thus, from (31) and (32), we obtain that $\{\widetilde{y}^k\}$ and $\{\widetilde{r}^k\}$ 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 (33), we have that $\{z^k\}$ converges 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

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

$$(x_w^k, z_w^k, y_w^k) \;=\; (x^k, z^k, y^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}) \tag{36}$$

$$\widetilde{z}^{k+1} = (\partial g + V_2 + B^tMB)^{-1}(V_2 z_w^k - B^tMAx_w^k - B^t y_w^k) \tag{37}$$

$$\widetilde{y}^{k+1} = y_w^k + MAx_w^k + MB\widetilde{z}^{k+1} \tag{38}$$

$$r^{k+1} = \nabla h((V_1 + A^tMA)^{-1}(V_1 x_w^k - A^tMB\widetilde{z}^{k+1})) \tag{39}$$

$$\widetilde{x}^{k+1} = (\partial f + V_1 + A^tMA)^{-1}\left(V_1 x_w^k - A^tMB\widetilde{z}^{k+1} - A^t\widetilde{y}^{k+1} - r^{k+1}\right) \tag{40}$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) \;=\; \rho(\widetilde{x}^{k+1}, \widetilde{z}^{k+1}, \widetilde{y}^{k+1}) + (1-\rho)(x_w^k, z_w^k, y_w^k). \tag{41}$$

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

$$\widehat{Q}\zeta_w^k = \widehat{Q}\zeta^k + \lambda(\widehat{Q}\zeta^k - \widehat{Q}\zeta^{k-1}) \quad \text{and} \quad \widehat{Q}\zeta^{k+1} = (1-\rho)\widehat{Q}\zeta_w^k + \rho G_2(\widehat{Q}\zeta_w^k). \tag{42}$$

▶ **Proposition 9.** *Set* $\bar{\lambda}$ *and* $\bar{\rho}$ *satisfying H1, and keep the same hypotheses of Proposition 2. If* $\mathrm{sol}(V_{\mathcal{L}})$ *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* $(\widetilde{x}^k, \widetilde{z}^k, \widetilde{y}^k)$ *in* (36)–(41) *converges to some element of* $\mathrm{sol}(V_{\mathcal{L}})$*.*

Proof. From relations (42), (11) and Proposition 7, 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 $(x^*, z^*, y^*) \in \mathrm{sol}(V_{\mathcal{L}})$, since $(\partial g + V_2 + B^t M B)^{-1}$ is a single-valued continuous map. From (37), we obtain that $\{\widetilde{z}^k\}$ 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 (38) and (39), it holds that $\{\widetilde{y}^k\}$ 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 $\{\widetilde{x}^k\}$ 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

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

$$(x_w^k, z_w^k, y_w^k) = (x^k, z^k, y^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}) \tag{43}$$

$$r^k = \nabla h\left((R_1 + 2A^t M A)^{-1}((R_1 + A^t M A)x_w^k - A^t M B z_w^k)\right) \tag{44}$$

$$\widetilde{x}^{k+1} = (\partial f + R_1 + 2A^t M A)^{-1}\left((R_1 + A^t M A)x_w^k - A^t M B z_w^k - A^t y_w^k - r^k\right) \tag{45}$$

$$\widetilde{z}^{k+1} = (\partial g + R_2 + 2B^t M B)^{-1}\left((R_2 + B^t M B)z_w^k - B^t M A x_w^k - B^t y_w^k\right) \tag{46}$$

$$\widetilde{y}^{k+1} = y_w^k + M A \widetilde{x}^{k+1} + M B \widetilde{z}^{k+1} \tag{47}$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\widetilde{x}^{k+1}, \widetilde{z}^{k+1}, \widetilde{y}^{k+1}) + (1-\rho)(x_w^k, z_w^k, y_w^k) \tag{48}$$

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

$$\widehat{S}\varsigma_w^k = \widehat{S}\varsigma^k + \lambda(\widehat{S}\varsigma^k - \widehat{S}\varsigma^{k-1}) \quad \text{and} \quad \widehat{S}\varsigma^{k+1} = (1-\rho)\widehat{S}\varsigma_w^k + \rho G_3(\widehat{S}\varsigma_w^k). \tag{49}$$

▶ **Proposition 10.** *Set* $\bar{\lambda}$ *and* $\bar{\rho}$ *satisfying H1. and with the same hypotheses as in Proposition 5. If* $\mathrm{sol}(V_{\mathcal{L}})$ *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* $(\widetilde{x}^k, \widetilde{y}^k, \widetilde{y}^k)$ *in* (43)–(48) *converges to some element of* $\mathrm{sol}(V_{\mathcal{L}})$*.*

Proof. From relation (49), we observe that this algorithm is related to IRFP applied to operator G_3 which, from the hypotheses, is $\alpha-$averaged. Then by Proposition 7 and relation (14), 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 $(x^*, z^*, y^*) \in \operatorname{sol}(V_{\mathcal{L}})$. From the continuity of ∇h and (44), we have that $\{r^k\}$ converges to $\nabla h(x^*)$; then since $(\partial f + R_1 + 2A^tMA)^{-1}$ and $(\partial g + R_2 + 2B^tMB)^{-1}$ are single-valued continuous maps from (45) and (46), we obtain that $\{\widetilde{x}^k\}$ and $\{\widetilde{z}^k\}$ converges to x^* and z^* respectively. It holds that $\{y_w^k\}$ converges to y^*, then from (47) we have $\{\widetilde{y}^k\}$ converges to y^*. ◀

▶ Remark 11. 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 $(\bar{x}^k, \bar{z}^k, \bar{y}^k) := (\widetilde{x}^k, \widetilde{z}^{k+1}, \widetilde{y}^{k+1})$ and $(\bar{x}^k, \bar{z}^k, \bar{y}^k) := (\widetilde{x}^k, \widetilde{z}^k, \widetilde{y}^k)$ in the above algorithms, we obtain the following new version of *PPDS* applied to model (1).

▶ Proximal-Gradient primal-dual Algorithm (PGPDS).

$$\left\{\begin{array}{lcl} \bar{r}^k & = & (V_1 + A^tMA)^{-1}\nabla h\left((V_1 + A^tMA)^{-1}(V_1\bar{x}^k - A^tMB\bar{z}^k)\right) \\ \bar{x}^{k+1} & \in & \operatorname{argmin}\left\{f(x) + \frac{1}{2}\|A(x+\bar{r}^k) + B\bar{z}^k + M^{-1}\bar{y}^k\|_M^2 + \frac{1}{2}\|x + \bar{r}^k - \bar{x}^k\|_{V_1}^2\right\} \\ \bar{z}^{k+1} & \in & \operatorname{argmin}\left\{g(z) + \frac{1}{2}\|A\bar{\eta}^k + Bz + M^{-1}\bar{y}^k\|_M^2 + \frac{1}{2}\|z - \bar{z}^k\|_{V_2}^2\right\} \\ \bar{y}^{k+1} & = & \bar{y}^k + M(A\bar{x}^{k+1} + B\bar{z}^{k+1}) \end{array}\right.$$

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 10 and 9 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 $V_1 = R_1 + A^tMA$ and $V_2 = R_2 + B^tMB$, 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 [17]. Concerning *MIRPD, Form I*, we show in Section 5 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 follows:

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

$$M = \sigma I_{m\times m}, \quad V_1 = \tau^{-1}I_{n\times n} - \sigma A^tA \quad \text{and} \quad V_2 = \mu^{-1}I_{p\times p} - \sigma B^tB \tag{50}$$

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

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

$$M = \widetilde{\sigma} I_{m\times m}, \quad R_1 = \widetilde{\tau}^{-1}I_{n\times n} - 2\widetilde{\sigma}A^tA \quad \text{and} \quad R_2 = \widetilde{\mu}^{-1}I_{p\times p} - 2\widetilde{\sigma}B^tB \tag{51}$$

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 (50) and (51), obtaining the following specialized algorithms.

Corresponding to *MIRPD, Form I*, with matrix parameters satisfying (50) 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 $\eta^k = \sigma Ax^k + y^k - \sigma z^k - \sigma Ar^k$, $\eta_w^k = \sigma Ax_w^k + y_w^k - \sigma z_w^k - \sigma Ar_w^k$ and $\widetilde{\eta}^k = \sigma A\widetilde{x}^k + \widetilde{y}^k - \sigma\widetilde{z}^k - \sigma A\widetilde{r}^k$, we get the following algorithm:

▶ Inertial-Relaxed Primal-Dual Three Operator, Form I (IRPD3O, Form I). Choose $(x^0, \eta^0, r^0) = (x^{-1}, \eta^{-1}, r^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$, σ and τ positive reals such that $\sigma\tau\|A\|^2 \le 1$ 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 $\alpha := \frac{2\beta}{4\beta - \tau}$.

$$(x_w^k, \eta_w^k, r_w^k) = (x^k, \eta^k, r^k) + \lambda(x^k - x^{k-1}, \eta^k - \eta^{k-1}, r^k - r^{k-1}) \tag{52}$$

$$\widetilde{x}^{k+1} = (\tau\partial f + I_{n\times n})^{-1}(x_w^k - \tau A^t\eta_w^k - r_w^k) \tag{53}$$

$$\widetilde{r}^{k+1} = \tau\nabla h(\widetilde{x}^{k+1}) \tag{54}$$

$$\widetilde{\eta}^{k+1} = (\sigma\partial g^* + I_{m\times m})^{-1}(\eta_w^k + \sigma A(2\widetilde{x}^{k+1} - x_w^k) + \sigma Ar_w^k - \sigma A\widetilde{r}^{k+1}) \tag{55}$$

$$(x^{k+1}, \eta^{k+1}, r^{k+1}) = \rho(\widetilde{x}^{k+1}, \widetilde{\eta}^{k+1}, \widetilde{r}^{k+1}) + (1-\rho)(x_w^k, \eta_w^k, r_w^k). \tag{56}$$

Without inertial 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 goes back to *PD3O*. Therefore this algorithm can be seen as resulting of the inclusion of inertial and relaxed terms into *PD3O*.

Analogously, considering *MIRPD, Form II*, with matrix parameters satisfying (50) and $V_2 = 0$, we obtain the switched version of *Inertial-Relaxed PD3O*.

▶ Inertial-Relaxed Primal-Dual Three Operator, Form II (IRPD3O, Form II). Choose $(x^0, y^0) = (x^{-1}, y^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m$, σ and τ positive real parameters such that $\sigma\tau\|A\|^2 \le 1$ 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 $\alpha := \frac{2\beta}{4\beta - \tau}$.

$$
\begin{aligned}
(x_w^k, y_w^k) &= (x^k, y^k) + \lambda(x^k - x^{k-1}, y^k - y^{k-1}) && (57)\\
\widetilde{y}^{k+1} &= (\sigma\partial g^* + I_{m\times m})^{-1}(y_w^k + \sigma A x_w^k) && (58)\\
r^{k+1} &= \nabla h(x_w^k - \tau A^t(\widetilde{y}^{k+1} - y_w^k)) && (59)\\
\widetilde{x}^{k+1} &= (\tau\partial f + I_{n\times n})^{-1}(x_w^k - \tau A^t(2\widetilde{y}^{k+1} - y_w^k) - \tau r^{k+1}) && (60)\\
(x^{k+1}, y^{k+1}) &= \rho(\widetilde{x}^{k+1}, \widetilde{y}^{k+1}) + (1-\rho)(x_w^k, y_w^k). && (61)
\end{aligned}
$$

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–Vũ, 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.

Alternatively, considering *MIRPPD*, with matrix parameters satisfying (51) 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 $(x^0, y^0) = (x^{-1}, y^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m$, $\widetilde{\sigma}$ and $\widetilde{\tau}$ positive real parameters such that $2\widetilde{\sigma}\widetilde{\tau}\|A\|^2 \le 1$ and $\widetilde{\tau} < 2\beta$, 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 $\alpha := \frac{2\beta}{4\beta - \widetilde{\tau}}$.

$$
\begin{aligned}
(x_w^k, z_w^k, y_w^k) &= (x^k, z^k, y^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}) && (62)\\
r^k &= \nabla h(x_w^k - \widetilde{\tau}\widetilde{\sigma}A^t A x_w^k + \widetilde{\sigma}\widetilde{\tau}A^t z_w^k) && (63)\\
\widetilde{x}^{k+1} &= (\widetilde{\tau}\partial f + I_{n\times n})^{-1}(x_w^k - \widetilde{\tau}A^t(\widetilde{\sigma}A x_w^k - \widetilde{\sigma} z_w^k + y_w^k) - \widetilde{\tau} r^k) && (64)\\
\widetilde{z}^{k+1} &= (\partial g + 2\widetilde{\sigma} I_{m\times m})^{-1}(\widetilde{\sigma} z_w^k + \widetilde{\sigma} A x_w^k + y_w^k) && (65)\\
\widetilde{y}^{k+1} &= y_w^k + \widetilde{\sigma} A \widetilde{x}^{k+1} - \widetilde{\sigma}\widetilde{z}^{k+1} && (66)\\
(x^{k+1}, z^{k+1}, y^{k+1}) &= \rho(\widetilde{x}^{k+1}, \widetilde{z}^{k+1}, \widetilde{y}^{k+1}) + (1-\rho)(x_w^k, z_w^k, y_w^k). && (67)
\end{aligned}
$$

The convergence of the last algorithms follows from Propositions 8, 9 and 10 respectively.

6 Numerical Results

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

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

where Q is a $q \times n$ matrix, $b \in \mathbb{R}^q$ and A is a $n-1 \times n$ matrix defined by

$$
A = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \cdots & \cdots & \\ & & & -1 & 1 \end{bmatrix}
$$

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 $B = -\mathbb{I}$), 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 τ and σ, 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 (τ, σ).

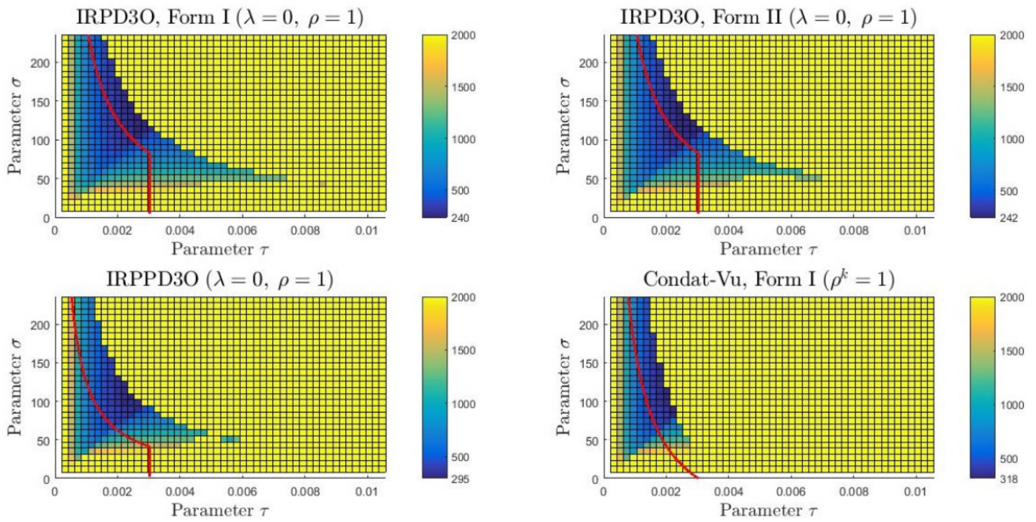

Figure 1 Number of iterations for an error of 10^{-6}

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 τ and the relaxation parameter ρ without inertial term ($\lambda = 0$). In Figure 3, we change the parameter τ and the inertial term λ without relaxation ($\rho = 1$).

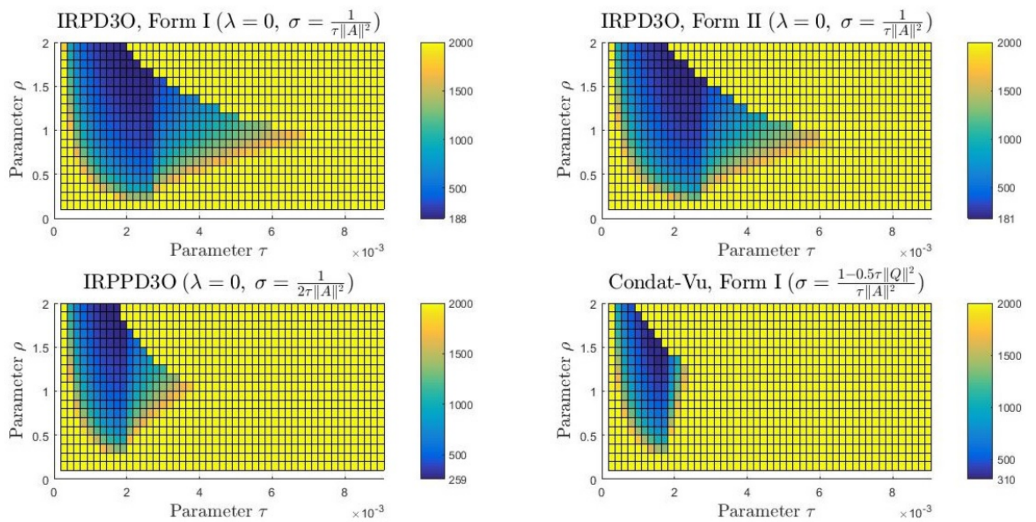

Figure 2 Varying the relaxation parameter

Finally, we change the inertial and relaxation parameters, forcing the values to belong to the curved red line of Figure 1

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

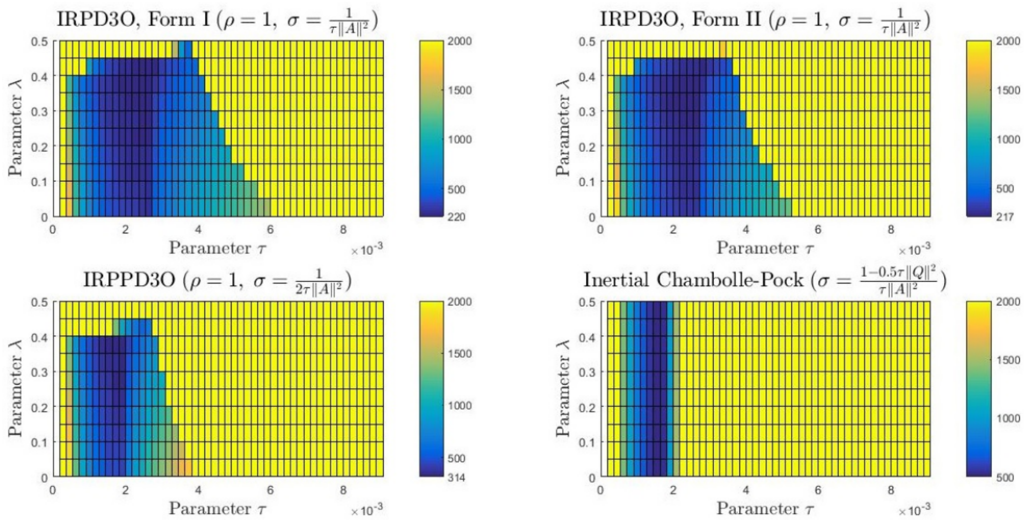

Figure 3 Varying the inertial parameter

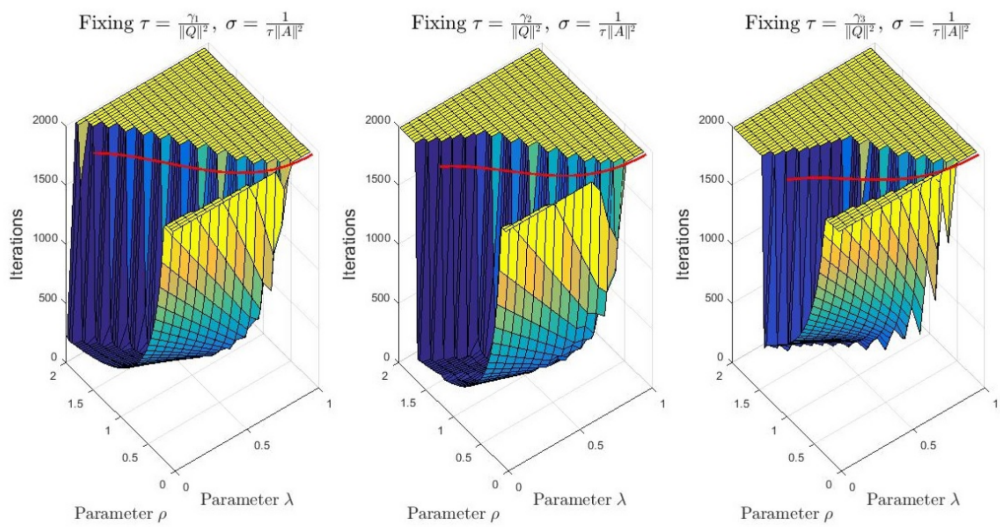

Figure 4 Varying λ and ρ

with $\gamma_1 = 1,\ \gamma_2 = 1.5$ and $\gamma_3 = 1.99$. We evaluate the number of iterations (bounded by 2000) needed to obtain a primal error less than 10^{-6}; again, a red line is plotted on the 3-D surface for the theoretical bounds.

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 [16] when the model does not present 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.

Appendix

We give here a direct proof of Proposition 7. We will need first the following technical result:

▶ **Lemma 12.** *Let $\{r_k\}$ and $\{v_k\}$ be two sequences in $[0,+\infty[$ and $\gamma > 0$ such that*

$$r_k + \gamma v_k \geq v_{k+1}, \quad \forall\, k \geq 0.$$

a. *For all $n \in \mathbb{N}$, it holds $\sum_{j=0}^{n} \gamma^j r_{n-j} + \gamma^{n+1} v_0 \geq v_{n+1}$.*
b. *If $\gamma = 1$ and $\sum r_k < +\infty$ then $\{v_k\}$ converges.*

Proof.

a. Multiplying by $\gamma^j > 0$, it holds

$$\gamma^j r_{n-j} \geq \gamma^j v_{n+1-j} - \gamma^{j+1} v_{n-j}, \quad \forall\, j = 0, \dots, n$$

then the result is obtained summing.

b. From a it holds $\sum r_k + v_0 \geq \sum_{j=0}^{n} r_j + v_0 \geq v_{n+1}$, then $\{v_n\}$ is bounded. Therefore the sequence $\sum_{k=0}^{n}(r_k + v_k - v_{k+1}) = \sum_{k=0}^{n} r_k + v_0 - v_{n+1}$ converges since it has nonnegative terms and it is bounded, deducing the convergence of $\{v_n\}$. Observe that, in this case, $\{v_k\}$ is a quasi-Fejer monotone sequence (see Lemma 5.31 in [4]). ◀

We can now present the proof of Proposition 7:

Proof of Proposition 7. Since F is $\alpha-$averaged, then $(1-\rho)I + \rho F$ is $\alpha\rho-$averaged with the same fixed point. Then given x^* any fixed point of F, and using (27), we have

$$\|w^k - x^*\|^2 \geq \|z^{k+1} - x^*\|^2 + \frac{1-\alpha\rho}{\alpha\rho}\|w^k - z^{k+1}\|^2. \tag{68}$$

From (26), 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 (68) and $\rho \leq \widehat{\rho}$, it holds

$$\lambda(\varphi_k - \varphi_{k-1}) + \lambda(1+\lambda)\|z^k - z^{k-1}\|^2 \geq \varphi_{k+1} - \varphi_k + \frac{1-\alpha\widehat{\rho}}{\alpha\widehat{\rho}}\|w^k - z^{k+1}\|^2. \tag{69}$$

Since $1-\alpha\widehat{\rho} = 1-\bar{\rho}/2 > 0$ and $\varphi_0 = \varphi_{-1}$, applying Lemma 12 a with $\gamma = \lambda$, considering $r_k = \lambda(1+\lambda)\|z^k - z^{k-1}\|^2$ and $v_k = \varphi_k - \varphi_{k-1}$, it holds

$$\sum_{j=0}^{n} \lambda^j r_{n-j} + \varphi_n \geq \varphi_{n+1}$$

We will show later that $\sum \|z^k - z^{k-1}\|^2$ (equal to $\frac{\sum r_k}{\lambda(1+\lambda)}$) indeed converges. Now, since $\lambda \leq \bar{\lambda} < 1$, applying Lemma 12 b with $\gamma = 1$ and considering $\widetilde{r}_n = \sum_{j=0}^{n} \lambda^j r_{n-j}$ and $\widetilde{v}_n = \varphi_n$, it holds that $\{\varphi_n\}$ converges, which from (69) also implies $\sum \|w^k - z^{k+1}\|^2 < +\infty$. Therefore $\{z^k\}$ is bounded, $\{z^k - z^{k-1}\}$ and $\{w^k - z^{k+1}\}$ both converge to zero. Later given z' a cluster point of $\{z^k\}$, from (26) and (27), we have $z' = (1-\rho)z' + \rho F(z')$ then z' is a fixed point of F, then considering $z^* = z'$, we have that $\{\|z^k - z'\|^2\}$ converges to zero which implies that $\{w^k\}$ and $\{z^k\}$ both converge to z'.

So it just remains to prove that $\sum \|z^k - z^{k-1}\|^2$ converges, when $\bar{\lambda}$ and $\bar{\rho}$ satisfy H1. From (26), 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 $\|\cdot\|^2$, it is true that

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

Then, replacing in (69), and denoting $\eta = \lambda(\alpha\widehat{\rho})^{-1} + \lambda^2(2 - (\alpha\widehat{\rho})^{-1})$, it holds

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

The last relation give us $\varphi^k - \varphi^{k-1} \geq \varphi^{k+1} - \varphi^k - \eta\|z^k - z^{k-1}\|^2$, so if $\eta < 0$, since $\varphi^1 = \varphi^0$, then $\{\varphi^k - \varphi^{k-1}\}$ is a decreasing sequence of nonnegative terms, deducing that $\sum \|z^k - z^{k-1}\|^2$ converges. Now we consider that $\eta \geq 0$, defining $\mu^{k-1} := \varphi^k - \lambda\varphi^{k-1} + \eta\|z^k - z^{k-1}\|^2$, the last inequality is rewritten as

$$\mu^{k-1} \geq \mu^k + q(\lambda)\|z^k - z^{k+1}\|^2 \tag{70}$$

where q is a quadratic function defined by

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

Since $\bar{\rho} < 2$, then $q(0) > 0$, it holds that $q(\lambda)$ is strictly positive when $\lambda \in [0, \widehat{\lambda}(\widehat{\rho})[$, where $\widehat{\lambda}(\widehat{\rho})$ is the smallest or largest positive root of q depending of the sign of the principal coefficient of q, which is equal to $\widehat{\lambda}(\widehat{\rho}) = \bar{\lambda}(\bar{\rho})$ (recall that $2\alpha\widehat{\rho} = \bar{\rho}$), with $\bar{\lambda} : (0,2) \to (0,1)$ 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 $\bar{\rho} : (0,1) \to (0,2)$ given by

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

Therefore if we choose $(\bar{\lambda}, \bar{\rho}) \in (0,1) \times (0,2)$ satisfying H1, since $\lambda < \bar{\lambda}$, we have that $q(\lambda)$ is strictly positive. Then from (70), in order to prove that $\sum \|z^k - z^{k-1}\|^2$ converges, it is sufficient to show that $\{\mu^k\}$ is bounded from below. So, from the definition of μ^k, using the fact that $\eta \geq 0$ (otherwise $\{\varphi^k\}$ is decreasing, deducing that μ^k is bounded), it holds that $\mu^k + \lambda\varphi^k \geq \varphi^{k+1}$, applying Lemma 12 a with $\gamma = \lambda$ and using the fact that μ^k is decreasing (from relation (70)), it finally holds that

$$\left(\sum_{j=0}^{n} \lambda^j\right)\mu^0 + \lambda^{n+1}\varphi_0 \geq \varphi_{n+1}.$$

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

References

1 Felipe Alvarez and Hedy Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.*, 9:3–11, 2001.

2 M. Marques Alves, Jonathan Eckstein, Marina Geremia, and Jefferson G. Melo. Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms. *Comput. Optim. Appl.*, 75:389–422, 2020.

3 M. Marques Alves and Marina 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(1):263–295, 2019.

4 Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, 2011.

5 Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning with the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, 2011.

6 Luis Briceño-Arias, Julio Deride, Sergio López-Rivera, and Francisco J. Silva. Primal-Dual Partial Inverse Splitting for Constrained Monotone Inclusions. `https://arxiv.org/abs/2007.01983v1`, 2020.

7 Luis Briceño-Arias and Fernando Roldán. Primal-dual splittings as fixed-point iterations in the range of linear operators. `https://arxiv.org/abs/1910.02329v1`, 2019.

8 Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex programs with applications to imaging. *J. Math. Imaging Vis.*, 40:1–26, 2011.

9 Laurent Condat. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, 158:460–479, 2013.

10 Laurent Condat, Daichi Kitahara, Andrés Contreras, and Akira Hirabayashi. Proximal splitting algorithms : a tour of recent advances with new twists. `https://arxiv.org/abs/1912.00137`, 2020.

11 Damek Davis and Wotao 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*, Scientific Computation, pages 115–163. Springer, 2016.

12 Damek Davis and Wotao Yin. A three-operator splitting scheme and its optimization applications. *Set-Valued Var. Anal.*, 25(4):829–858, 2017.

13 Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators. *Math. Program.*, 55(3):293–318, 1992.

14 Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. *Comput. Math. Appl.*, 2:17–40, 1976.

15 Osman Güler. New proximal point algorithms for convex optimization. *SIAM J. Optim.*, 2:649–664, 1992.

16 Bingsheng He, Feng Ma, and Xiaoming Yuan. Optimally linearizing the alternating direction method of multipliers for convex programming. *Comput. Optim. Appl.*, 75(2):361–388, 2020.

17 Min Li, Defeng Sun, and Kim-Chuan Toh. A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. *SIAM J. Optim.*, 26(2):922–950, 2016.

18 Yurii 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 Ernesto Oré-Albornoz, Philippe Mahey, and Eladio Ocana-Anaya. A unified splitting algorithm for composite monotone inclusions. *J. Convex Anal.*, 27(3):893–922, 2020.

20 Fabian Pedregosa and Gauthier Gidel. Adaptive three-operator splitting. In *Proceedings of 35th Int. Conf. on Machine Learning*, pages 4085–4094, 2018.

21 R. Tyrell Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1:97–116, 1976.

22 Ron Shefi and Marc Teboulle. Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. *SIAM J. Optim.*, 24(1):269–297, 2014.

23 Băng Công Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667–681, 2013.

24 Ming Yan. A new primal-dual algorithm for minimizing the sum of three functions with a linear operator. *J. Sci. Comput.*, 76(3):1698–1717, 2018.