Journal of Convex Analysis
Volume 29 (2022), No. 3, 893–920

Fast Convergence of Generalized Forward-Backward Algorithms for Structured Monotone Inclusions

Paul-Emile Maingé
Université des Antilles, Martinique, France
paul-emile.mainge@univ-antilles.fr

Received: July 22, 2021
Accepted: December 10, 2021

We develop rapidly convergent forward-backward algorithms for computing zeroes of the sum of finitely many maximally monotone operators. A modification of the classical forward-backward method for two general operators is first considered, by incorporating an inertial term (close to the acceleration techniques introduced by Nesterov), a constant relaxation factor and a correction term. In a Hilbert space setting, we prove the weak convergence to equilibria of the iterates (x_n), with worst-case rates of $o(n^{-1})$ in terms of both the discrete velocity and the fixed point residual, instead of the classical rates of $\mathcal{O}(n^{-1/2})$ established so far for related algorithms. Our procedure is then adapted to more general monotone inclusions and a fast primal-dual algorithm is proposed for solving convex-concave saddle point problems.

Keywords: Nesterov-type algorithm, inertial-type algorithm, global rate of convergence, fast first-order method, relaxation factors, correction term, accelerated proximal algorithm, fixed point problem.

2010 Mathematics Subject Classification: 90C25, 90C30, 90C60, 68Q25, 49M25.

1. Introduction

Let $\mathcal{H}$ be a real Hilbert space endowed with inner product and induced norm denoted by $\langle .,. \rangle$ and $\| \, . \, \|$, respectively. For any given linear self-adjoint and positive definite mapping $\mathcal{M} : \mathcal{H} \to \mathcal{H}$, and any bounded linear operator $\mathcal{L} : \mathcal{H} \to \mathcal{H}$, we define $\| \, . \, \|_{\mathcal{M}} = \sqrt{\langle \mathcal{M}(.), . \rangle}$ (as an auxiliary metric on $\mathcal{H}$) and $\|\mathcal{L}\| = \sup_{\|x\|=1} \|\mathcal{L}x\|$. Our goal is to propose and study rapidly converging forward-backward methods for solving a wide class of structured monotone inclusions, but it can be applied to a much larger class of problems.

1.1. A new way to speed up forward-backward methods

The simplest case of the considered problems consists of finding a zero of the sum of two operators, especially the following monotone inclusion

$$\text{find } \bar{x} \in \mathcal{H} \text{ such that } 0 \in A(\bar{x}) + B(\bar{x}), \tag{1.1}$$

under the assumptions that $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone and that $B : \mathcal{H} \to \mathcal{H}$ is co-coercive (see, e.g., [6, 15]). This framework finds many important applications in scientific fields such as image processing, computer vision, machine learning, signal processing, optimization, equilibrium theory, economics and game theory, partial

ISSN 0944-6532 / $ 2.50

differential equations, statistics, among other subjects (see, e.g., [5, 6, 13, 28, 30, 41, 46, 50]). In particular, we recall that the above setting encompasses the non-smooth structured convex minimization problem

$$\min_{\mathcal{H}}\{\Theta := f + g\}, \tag{1.2}$$

where $g : \mathcal{H} \to (-\infty, \infty]$ is a proper, convex and lower semi-continuous function and $f : \mathcal{H} \to (-\infty, \infty)$ is convex differentiable with Lipschitz continuous gradient. Indeed, the gradient of a convex and Frechet differentiable function is well-known to be co-coercive provided that it is Lipschitz continuous (see [6]). Thus, (1.2) is nothing but the special instance of (1.1) when $A = \partial f$ (∂f being the Fenchel sub-differential of f) and $B = \nabla g$ (∇g being the gradient of g).

A typical method for solving (1.1), with a β-co-coercive operator B (for some $\beta > 0$), is the forward-backward algorithm (FBA, for short), which operates according to the routine (see, e.g., Lions-Mercier [29], Passty [37])

$$x_{n+1} = J_{\lambda_n A}(x_n - \lambda_n B(x_n)), \tag{1.3}$$

where $(\lambda_n) \subset (0, 2\beta)$ and $J_{\lambda_n A} := (I + \lambda_n A)^{-1}$ is the resolvent operator of A with index λ_n (see Section 1.2.1 for more details on the resolvent operators).

When applied to the minimization problem (1.2), FBA was shown to generate weakly convergent sequences (x_n), with worst-case rates $\Theta(x_n) - \inf_{\mathcal{H}} \Theta = \mathcal{O}(n^{-1})$ (for the function values) and $\|x_{n+1} - x_n\| = \mathcal{O}(n^{-1/2})$ (for the discrete velocity). These rates have been considerably enhanced through an accelerated forward-backward algorithm (AFBA, for short, which will be described in Section 1.2.2), by means of an inertial-type extrapolation process. The latter AFBA generates convergent sequences (x_n) with the improved worst-case rates $\Theta(x_n) - \inf_{\mathcal{H}} \Theta = o(n^{-2})$ and $\|x_{n+1} - x_n\| = o(n^{-1})$. Afterwards, general variants of AFBA, with arbitrary monotone operators $A : \mathcal{H} \to 2^{\mathcal{H}}$ and $B : \mathcal{H} \to \mathcal{H}$, have been extensively studied and many attempts to get convergence rates similar to that reached for the convex minimization case can be found in the literature (see, e.g., [1, 30, 33]). However, most of these works are only concerned with empirical results.

Our purpose here is twofold with regard to the previous observations:

First, we aim at extending, in terms of discrete velocity and fixed point residual, the above convergence properties (obtained for structured convex minimization) to some new variant of (1.3). This process will be investigated with an additional preconditioning strategy, as suggested by Lorenz-Pock [30], under the following general conditions:

$$A : \mathcal{H} \to 2^{\mathcal{H}} \quad \text{is maximally monotone on } \mathcal{H}; \tag{1.4a}$$

$$B : \mathcal{H} \to \mathcal{H} \quad \text{is co-coercive w.r.t a linear, self-adjoint and positive definite map } L, \tag{1.4b}$$

$$S := (A + B)^{-1}(0) \neq \emptyset. \tag{1.4c}$$

An operator $B : \mathcal{H} \to 2^{\mathcal{H}}$ is said to be co-coercive w.r.t. a linear and positive definite map $L : \mathcal{H} \to \mathcal{H}$ if it satisfies

$$\langle Bx - By, x - y\rangle \geq \|Bx - By\|^2_{L^{-1}}, \quad \text{for } (x, y) \in \mathcal{H}^2. \tag{1.5}$$

Note that in the simple case when $L = \beta^{-1}I$ (for some $\beta > 0$), any operator B satisfying (1.5) is β-co-coercive. It can also turn out useful to consider more general map L as discussed in [30].

Specifically, for solving (1.1)-(1.4), we introduce CRIFBA (corrected relaxed inertial forward-backward algorithm) which consists of sequences $\{z_n, x_n\} \subset \mathcal{H}$ generated by the following process:

CRIFBA:

$\triangleright$ **Step 1** (initialization):
Let $M : \mathcal{H} \to \mathcal{H}$ be a linear self-adjoint and positive definite map, let $\{z_{-1}, x_{-1}, x_0\} \subset \mathcal{H}$, $\{s_1, s_0, \nu_0\} \subset [0, \infty)$, $\{e, \lambda, w\} \subset (0, \infty)$, and set

$$\nu_n = s_1 n + \nu_0, \quad \theta_n = 1 - \tfrac{e+s_1}{e+\nu_{n+1}}, \quad \gamma_n = 1 - \tfrac{s_0}{e+\nu_{n+1}}. \tag{1.6}$$

$\triangleright$ **Step 2** (main step):
Given $\{z_{n-1}, x_{n-1}, x_n\} \subset \mathcal{H}$ (with $n \geq 0$), we compute the updates by

$$v_n = z_{n-1} - x_n, \tag{1.7a}$$

$$z_n = x_n + \theta_n(x_n - x_{n-1}) + \gamma_n v_n, \tag{1.7b}$$

$$x_{n+1} = (1 - w)z_n + wJ_{\lambda M^{-1}A}(z_n - \lambda M^{-1}B(z_n)). \tag{1.7c}$$

where $J_{\lambda M^{-1}A} := (I + \lambda M^{-1}A)^{-1}$ denotes the resolvent of $M^{-1}A$ (also referred to as a generalized resolvent of A).

The algorithm under consideration can be regarded as a preconditioned and relaxed variant of the classical forward-backward in which we further incorporate the momentum term "$\theta_n(x_n - x_{n-1})$" (inspired by Güler's acceleration techniques for convex minimization) and the correction term "$\gamma_n(z_{n-1} - x_n)$" (similar to that suggested by Kim [26] in proximal point iterations). Note that (1.7c) can be formulated as

$$x_{n+1} = (1 - w)z_n + w(M + \lambda A)^{-1}(Mz_n - \lambda B(z_n)). \tag{1.8}$$

Then, as illustrated in [30] in the context of convex-concave saddle-point problems, the introduction of the map M can be helpful in some situations to make the iteration feasible, but it can be also interpreted as a left pre-conditioner to the monotone inclusion (1.1). Furthermore, this procedure allows us to investigate our algorithm with respect to the auxiliary metric $\|.\|_M$, which can be also used as a speeding up process.

As a useful tool in our methodology we introduce the fixed point residual operator defined by

$$G_\lambda^M = \tfrac{1}{\lambda}\left(I - J_{\lambda M^{-1}A} \circ (I - \lambda M^{-1}B)\right). \tag{1.9}$$

The term $G_\lambda^M(z)$ (for $z \in \mathcal{H}$) can be regarded as a tool to measure the accuracy of some point z to the solution set S since $G_\lambda(z) = 0$ is equivalent to $z \in S$.

The simple form of CRIFBA allows us to extend the fast convergence rates obtained for AFBA (in the context of potential operators) to the wide framework of the structured monotone inclusion (1.1)-(1.4), especially in terms of the discrete velocity and fixed point residual $\|G_\lambda^M(x_n)\|_M$. In particular, using appropriate parameters $\{\theta_n, \gamma_n, \lambda, w\}$, we establish, among others (see, Theorem 2.1), the weak convergence of (x_n) towards equilibria belonging to S, with the worst-case rates $\|x_{n+1} - x_n\|_M = o(n^{-1})$ (for the discrete velocity) and $\|G_\lambda^M(x_n)\|_M = o(n^{-1})$ (for the fixed point residual).

Secondly, by following the methodology developed by Raguet-Fadili-Peyré [40], we aim at adapting the proposed acceleration techniques to the more general inclusion problem :

$$\text{find } \bar{x} \in \mathcal{H} \text{ s.t. } 0 \in B(\bar{x}) + \textstyle\sum_{k=1}^{p} A_k(\bar{x}), \tag{1.10}$$

where $B : \mathcal{H} \to \mathcal{H}$ is β-co-coercive and $(A_i)_{i=1}^{p} : \mathcal{H} \to 2^{\mathcal{H}}$ is a family of p maximally monotone operators. This gives rise to G-CRIFBA (generalized corrected and regularized forward-backward algorithm), whose structure bears similarities with FBA (1.3). Indeed, G-CRIFBA consists of an explicit forward step, followed by an implicit step in which the resolvent operators of each A_i are computed in parallel. Let us underline that G-CRIFBA provides yet another way for computing the resolvent of the sum of maximally monotone operators at a point $y \in \text{rang}\,(I + \sum_{i=1}^{p} A_i)$. This can be seen when taking B as the particular operator defined for $x \in \mathcal{H}$ by $B(x) = x - y$ (hence B is 1-co-coercive). A particular attention will also be paid to the case of potential operators.

1.2. Motivations and reminders on proximal splitting algorithms

1.2.1. Some splitting variants of the basic proximal algorithm

A classical method to compute zeroes of a maximally monotone operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is the so-called PPA (proximal point algorithm) (see Martinet [32], Rockafellar [43, 44]), which consists of the iteration

$$x_{n+1} = J_{\lambda A}(x_n), \tag{1.11}$$

where $J_{\lambda A} := (I + \lambda A)^{-1}$ (for some positive parameter λ) is the resolvent operator of A, which is well-known to be everywhere defined and single-valued (see, e.g., [9, 24, 27] for more details). In particular, when $A = \partial g$ is the Fenchel sub-differential of a proper convex and lower-semi-continuous function $g : \mathcal{H} \to (-\infty, \infty]$, the resolvent operator of A reduces to the proximal mapping of g given by

$$\text{prox}_{\lambda g}(x) := J_{\lambda \partial g} = \text{argmin}_{y \in \mathcal{H}} \left(g(y) + (2\lambda)^{-1}\|x - y\|^2\right). \tag{1.12}$$

It is well-known that PPA generates weakly convergent sequences (x_n) with a discrete velocity that vanishes at the rate $\|x_{n+1} - x_n\|^2 = \mathcal{O}(n^{-1})$ (or equivalently, $\|x_n - J_{\lambda A}(x_n)\|^2 = \mathcal{O}(n^{-1})$ for the fixed point residual).

In many situations, however, evaluating the resolvent of the sum of two maximally monotone operators A and B turns out to be more complicated that evaluating separately the proximal operators of A and B. This observation gave rise to two main categories of splitting methods of practical interest (for instance, in the context of sparse signal recovery [16, 19], machine learning [22] and image processing [40]):

(c1) The first category of splitting algorithms is composed of those that essentially include backward steps, namely, no evaluations of A and B, but only evaluations of both the resolvent operators of A and B. This kind of methods originates from the Peaceman-Rachford and the Douglas-Rachford splitting algorithms (see [20, 29, 38]). As an example, we mention the following iteration (see Corman-Yuan [18], Eckstein-Bertsekas [23])

$$x_{n+1} = H(x_n), \text{ where } H = \big(J_{\lambda A}(2J_{\lambda B} - I) + I - J_{\lambda B}\big) \text{ (for some } \lambda > 0). \tag{1.13}$$

The operator defined by $D := H^{-1} - I$ was shown to satisfy the remarkable properties (see [23])

$$D \text{ is maximally monotone}, \quad H = J_D, \quad D^{-1}(0) = S := (A+B)^{-1}(0). \tag{1.14}$$

The iterates (x_n) produced by (1.13) were shown to converge weakly to some element of S.

(c2) The splitting methods in the second category combine both backward steps (evaluations of resolvent operators) and forward steps (evaluations of one of the two operators).

– A popular example of such methods, for a single-valued mapping B, is given by so-called forward-backward algorithm (1.3). In specific, when B is β- co-coercive, (1.3) was shown to be a weakly convergent method, provided that $(\lambda_n) \subset (0, 2\beta)$.

– Another example, for a single-valued mapping B, is given by the following forward-backward-forward algorithm proposed by Tseng [47] (see also [8])

$$y_n = J_{\lambda_n A}(x_n - \lambda_n B(x_n)), \qquad x_{n+1} = y_n - \lambda_n (B(y_n) - B(x_n)). \tag{1.15}$$

This method was shown to generate (weakly) convergent sequences, even when B is L-Lipschitz continuous (which is a weaker condition than co-coerciveness), provided that $(\lambda_n) \subset (0, L^{-1})$.

– Later on, generalized variants of the forward-backward and forward-backward-forward algorithms (see, e.eg., [14, 40]) have been adapted to more general inclusion problems such as (1.10).

1.2.2. Proximal splitting algorithms and acceleration processes

It is suitable to accelerate PPA and its splitting variants, in view of their various applications.

Note that, in the context of the convex minimization (1.2), FBA (1.3) reduces to

$$x_{n+1} = \text{prox}_{\lambda_n g}(x_n - \lambda_n \nabla f(x_n)). \tag{1.16}$$

When ∇f is L-Lipschitz continuous, and for values $(\lambda_n) \subset (0, L^{-1})$, (1.16) generates (weakly) convergent sequences (x_n) that satisfy the (sub-linear) convergence rate $\Theta(x_n) - \inf_{\mathcal{H}} \Theta = \mathcal{O}(n^{-1})$ ([10]).

Afterwards, (1.16) was enhanced through the *Fast Iterative Soft Thresholding Algorithm* (FISTA) proposed by Beck-Teboule [7] (see also [48]), based upon the techniques of Güler [25] and Nesterov [34, 35, 36]). FISTA was shown to produce iterates (x_n) that guarantee a rate of convergence $\Theta(x_n) - \inf_{\mathcal{H}} \Theta = \mathcal{O}(n^{-2})$. However, the convergence of these iterates has not been established.

This drawback was overcome by the following variant of FISTA recently introduced by Chambolle-Dossal (see [12]) (see also Attouch-Peypouquet [3]) given by

$$\begin{aligned} z_n &= x_n + \frac{n-1}{n+\alpha-1}(x_n - x_{n-1}), \\ x_{n+1} &= \text{prox}_{\lambda g}(z_n - \lambda \nabla f(z_n)), \end{aligned} \tag{1.17}$$

where $\lambda \in (0, L^{-1})$ and $\alpha > 0$.

It was proved for $\alpha > 3$ (see [3]) that (1.17) generates (weakly) convergent sequences (x_n) that minimize the function values $\Theta(x_n)$ with a complexity result of $o(n^{-2})$, instead of the rates $\mathcal{O}(n^{-1})$ and $\mathcal{O}(n^{-2})$ obtained for (1.16) and FISTA, respectively.

In the case of an arbitrary maximally monotone operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ that accelerated variants of PPA have been proposed and investigated through RIPA (Regularized Inertial Proximal Algorithm) by Attouch-Peypouquet [4], PRINAM (Proximal Regularized Inertial Newton Algorithm) by Attouch-László [2]. These algorithms, despite their interesting asymptotic features, require unbounded proximal indexes for convergence and so cannot be extended to the forward-backward framework. In the same context, an accelerated proximal point method involving constant proximal indexes was proposed by Kim [26], based on the performance estimation problem (PEP) approach of Drori-Teboulle [21]. This yields the worst-case convergence rate of $\mathcal{O}(n^{-2})$ in terms of fixed point residuals. Once again, no convergence of the iterates was established.

To the best of our knowledge, regarding the existing algorithmic solutions to the problems (1.1)–(1.4) with general operators, there are no analogous theoretical convergence results to that obtained for (1.17). Only somewhat empirical accelerations have been proposed, except for the work of Attouch-Cabot [1], via relaxation and inertial extrapolation techniques. Some of these processes are recalled below:

(e1) Inertial variants of (1.3), with a co-coercive operator B, have been discussed by Moudafi-Oliny [33]:

$$z_n = x_n + \alpha_n(x_n - x_{n-1}), \quad x_{n+1} = J_{\lambda_n A}(z_n - \lambda_n B(x_n)), \tag{1.18}$$

and by Lorenz-Pock [30]:

$$z_n = x_n + \alpha_n(x_n - x_{n-1}), \quad x_{n+1} = J_{\lambda_n A}(z_n - \lambda_n B(z_n)), \tag{1.19}$$

where (α_n) and (λ_n) are positive sequences. Note that this second method involves the evaluation of B at z_n instead of x_n (as done in (1.18)).

(e2) A reflected variant of (1.3), with a Lipschitz continuous operator B, was investigated by Cevher-Vu [11]:

$$y_n = 2x_n - x_{n-1}, \quad x_{n+1} = J_{\lambda A}(x_n - \lambda B y_n). \tag{1.20}$$

(e3) An inertial and relaxed variant of (1.3), with a β-co-coercive operator B, has been discussed by Attouch-Cabot [1]:

$$z_n = x_n + \alpha_n(x_n - x_{n-1}), \quad x_{n+1} = (1 - w_n)z_n + w_n J_{\lambda_n A}(z_n - \lambda_n B z_n), \tag{1.21}$$

where $\{\alpha_n, w_n, \lambda_n\}$ are positive and bounded sequences. Under various conditions on the parameters, the authors have established the weak convergence to equilibria of the iterates (x_n), with additional convergence rates in terms of the discrete velocity and the fixed point residual $\|G_\lambda^I(x_n)\|^2$ (where G_λ^I was introduced in (1.9)).

In particular, this work meets the setting of Nesterov's accelerated methods through the choice of $\alpha_n = 1 - \alpha n^{-1}$ with $\alpha > 2$ (for the momentum coefficient), along with $\lambda_n = \lambda \in (0, 2\beta)$ (for the proximal indexes), $w_n = 1 - \rho n^{-2}$ and finally with $0 < \rho < \alpha(\alpha - 2)(1 - \frac{\lambda}{4\beta})$ (for the relaxation factors).

In this framework (see [1, Corollary 4.9], Attouch and Cabot have reached the estimates $\|x_{n+1} - x_n\| = \mathcal{O}(n^{-1})$ and $\sum_n n\|x_{n+1} - x_n\|^2 < \infty$ (for the discrete velocity), along with $\sum_n n^{-1}\|G_\lambda^I(x_n)\|^2 < \infty$ and $\lim_{n\to\infty} \|G_\lambda^I(x_n)\| = 0$ (for the fixed point residuals).

As a consequence of the correction term in CRIFBA we improve these last rates from $\|x_{n+1} - x_n\| = \mathcal{O}(n^{-1})$ and $\|G_\lambda^I(x_n)\| = \mathcal{O}(1)$ to $\|x_{n+1} - x_n\| = o(n^{-1})$ and $\|G_\lambda^I(x_n)\| = o(n^{-1})$ (see Theorem 2.1).

1.3. Organization of the paper

An outline of this paper is as follows. In Section 2, we present our main convergence results (see Theorem 2.1). Some preparatory estimations regarding CRIFBA are also stated. Section 3 is devoted to the convergence analysis of CRIFBA and a proof of the main results is then proposed. In Section 4, we specialize CRIFBA to the setting of the more general monotone inclusion (1.10). Next, in Section 5, we give an application of CRIFBA relative to convex-concave saddle point problems.

Remark 1.1. From now on, so as to simplify the notations, we (often) use the following notation: given any sequence (u_n), we denote $\dot{u}_n = u_n - u_{n-1}$.

2. Main convergence results and preliminary estimations

2.1. Main convergence results

The following result states the convergence of CRIFBA, with an accuracy measured through the operator G_λ^M (introduced in (1.9)).

Theorem 2.1. *Let $L, M : \mathcal{H} \to \mathcal{H}$ be linear self-adjoint and positive definite maps, and let $A : \mathcal{H} \to 2^{\mathcal{H}}$ and $B : \mathcal{H} \to \mathcal{H}$ satisfy (1.4), with $S := (A + B)^{-1}(0) \neq \emptyset$. Suppose that $\{x_n, v_n\} \subset \mathcal{H}$ are generated by CRIFBA with $\{s_1, s_0, \nu_0\} \subset [0, \infty)$ and $\{e, \lambda, w\} \subset (0, \infty)$ such that*

$$2s_1 < s_0 < e, \quad 0 < w < 1. \tag{2.1}$$

Suppose furthermore that one of the following conditions (2.2a) *and* (2.2b) *holds :*

$$\exists \delta > 0 \ \text{ s.t. } \ \lambda\|L\| \leq 4\delta \ \text{ and } \ \bar{M}_1 := M - \frac{\delta}{w(1-w)} I \ \text{ is positive definite}, \tag{2.2a}$$

$$\bar{M}_2 := M - \frac{\lambda}{w(1-w)} L \ \text{ is positive definite.} \tag{2.2b}$$

Then the following properties are reached:

$$\|\dot{x}_{n+1}\|_M = o(n^{-1}), \ \ \textstyle\sum_n n\|\dot{x}_{n+1}\|_M^2 < \infty, \ \ \textstyle\sum_n n^2\|\dot{x}_{n+1} - \dot{x}_n\|_M^2 < \infty, \tag{2.3a}$$

$$\|v_n\|_M = o(n^{-1}), \ \ \textstyle\sum_n n\|v_n\|_M^2 < \infty, \textstyle\sum_n n^2\|\dot{v}_{n+1}\|_M^2 < \infty, \tag{2.3b}$$

$$\|G_\lambda^M(x_n)\|_M = o(n^{-1}), \ \ \textstyle\sum_n n\|G_\lambda^M(x_n)\|_M^2 < \infty. \tag{2.3c}$$

Moreover, denoting $y_n = x_n + \left(1 - \frac{1}{w}\right) v_n$, we have:

$$\|\dot{y}_{n+1}\|_M = o(n^{-1}), \ \ \textstyle\sum_n n\|\dot{y}_{n+1}\|_M^2 < \infty, \tag{2.3d}$$

$$\|G_\lambda^M(y_n)\|_M = o(n^{-1}), \ \ \textstyle\sum_n n\|G_\lambda^M(y_n)\|_M^2 < \infty. \tag{2.3e}$$

If, in addition, M and L are bounded, then:

$$\exists \bar{x} \in S, \text{ s.t. } (x_n, y_n) \rightharpoonup (\bar{x}, \bar{x}) \text{ weakly in } (\mathcal{H}, \|.\|_M)^2, \tag{2.3f}$$

$$\exists y_n^* \in (A+B)(y_n), \text{ s.t. } \|y_n^*\|_M = o(n^{-1}) \text{ and } \textstyle\sum_n n\|y_n^*\|_M^2 < \infty. \tag{2.3g}$$

Theorem 2.1 will be proved in Section 3.3.

Let us give some comments on the above theorem.

Remark 2.2. Recall that a linear self-adjoint operator $\mathcal{M} : \mathcal{H} \to \mathcal{H}$ is called positive definite if it satisfies $\inf_{x\in\mathcal{H}} \|\langle \mathcal{M}x, x\rangle\|/\|x\| > 0$. So, we emphasize, for a bounded operator L, that the two conditions (2.2a) and (2.2b) are always feasible for δ and λ small enough. In specific, regarding the classical setting of (1.1)–(1.4) when $L = \beta^{-1}I$ (with $\beta > 0$) and $M = I$ (as used for instance in [1, 33, 40]), (2.2a) reduces to $0 < \lambda \leq 4\beta\delta$ and $0 < \delta < w(1-w)$, for some $\delta > 0$, which is equivalent to $0 < \lambda < 4\beta w(1-w)$, while (2.2b) becomes the more stringent condition $0 < \lambda < w(1-w)\beta$. Nonetherless, condition (2.2b) (in general) does not require L to be bounded.

2.2. Preliminary estimations on CRIFBA

As standing assumptions regarding this section, we assume that $L : \mathcal{H} \to \mathcal{H}$ and $M : \mathcal{H} \to \mathcal{H}$ are linear self-adjoint and positive definite maps, and that $A : \mathcal{H} \to 2^{\mathcal{H}}$ and $B : \mathcal{H} \to \mathcal{H}$ satisfy condition (1.4).

2.2.1. A useful reformulation of the algorithm

We begin by making the following observation.

Remark 2.3. It is importance to notice that (1.7c) can be reformulated as

$$x_{n+1} = z_n - w\lambda G_\lambda^M(z_n), \tag{2.4}$$

where G_λ^M was introduced in (1.9).

A key result in our analysis is then given by the following proposition.

Proposition 2.4. *The iterates (x_n) and (v_n) generated by CRIFBA satisfy $(n \geq 0)$*

$$v_{n+1} = (\lambda w)G_\lambda^M(z_n), \tag{2.5a}$$

$$\dot{x}_{n+1} + v_{n+1} = \theta_n \dot{x}_n + \gamma_n v_n. \tag{2.5b}$$

Proof. For $n \geq 0$, we have $v_{n+1} = z_n - x_{n+1}$ (by definition of v_n) together with $z_n = x_{n+1} + (\lambda w)G_\lambda^M(z_n)$ (from (2.4)), hence, we obviously infer $v_{n+1} = (\lambda w)G_\lambda^M(z_n)$, which yields (2.5a). Furthermore, by $z_n = x_n + \theta_n \dot{x}_n + \gamma_n v_n$ (from (1.7b)), we immediately obtain

$$v_{n+1} := z_n - x_{n+1} = \theta_n \dot{x}_n + \gamma_n v_n - \dot{x}_{n+1},$$

which entails (2.5b). □

2.2.2. Co-coerciveness of G_λ^M and basic properties on (x_n) and (v_n)

The next result plays a central role in our methodology. For simplification, given any $(x_1, x_2) \in \mathcal{H}^2$, we write:

$$\Delta G_\lambda^M(x_1, x_2) = G_\lambda^M(x_1) - G_\lambda^M(x_2) \text{ and } \Delta B(x_1, x_2) = B(x_1) - B(x_2).$$

Proposition 2.5. *For any $\lambda > 0$ and for any $(x_1, x_2) \in \mathcal{H}^2$, we have*

$$\begin{aligned} &\langle \Delta G_\lambda^M(x_1,x_2), x_1 - x_2\rangle_M \\ &\geq \|\Delta B(x_1,x_2)\|_{L^{-1}}^2 + \lambda \|\Delta G_\lambda^M(x_1,x_2)\|_M^2 - \lambda \langle \Delta G_\lambda^M(x_1,x_2), \Delta B(x_1,x_2)\rangle. \end{aligned} \tag{2.6}$$

Hence, for $i = 1, 2$, the following inequalities hold:

$$\begin{aligned} &\langle \Delta G_\lambda^M(x_1,x_2), x_1 - x_2\rangle_M \\ &\geq \alpha_i \|\Delta B(x_1,x_2)\|_{L^{-1}}^2 + \lambda \langle H_i \Delta G_\lambda^M(x_1,x_2), \Delta G_\lambda^M(x_1,x_2)\rangle, \end{aligned} \tag{2.7}$$

with parameters α_i and H_i defined by

$$\alpha_1 = 1 - \frac{\lambda}{4\delta}\|L\| \quad \text{and} \quad H_1 = M - \delta I \quad \text{(for some given } \delta > 0\text{)}, \tag{2.8a}$$

$$\alpha_2 = \frac{3}{4} \quad \text{and} \quad H_2 = M - \lambda L. \tag{2.8b}$$

The proof of Proposition 2.5 is given in the Appendix A–1 and it makes use of the following observation.

Remark 2.6. Recall that for any linear self-adjoint and positive definite operator $\mathcal{M} : \mathcal{H} \to \mathcal{H}$, there exists a linear self-adjoint and positive definite, map denoted $\mathcal{M}^{\frac{1}{2}}$ such that $\mathcal{M} = \mathcal{M}^{\frac{1}{2}}\mathcal{M}^{\frac{1}{2}}$ (see [45]). The operator $\mathcal{M}^{\frac{1}{2}}$ is called of roof of $\mathcal{M}$ and it also satisfies $\|\mathcal{M}^{\frac{1}{2}}\|^2 = \|\mathcal{M}\|$.

As an immediate consequence of the previous proposition, we provide basic properties concerning the sequences (x_n) and (v_n).

Proposition 2.7. *Suppose that $\{x_n, v_n\} \subset \mathcal{H}$ are generated by CRIFBA with parameters $\{s_0, s_1\} \subset [0, \infty)$, $\{e, \nu_0, \lambda\} \subset (0, \infty)$ and $w \in (0, 1)$. Suppose furthermore, for some integer $i_c \in \{1, 2\}$, that $\bar{M}_{i_c}$ is positive definite, where $\bar{M}_1$ and $\bar{M}_2$ are defined by*

$$\bar{M}_1 := M - \frac{\delta}{w(1-w)} I \quad \text{(for some given } \delta > 0\text{)}, \quad \bar{M}_2 := M - \frac{\lambda}{w(1-w)} L. \tag{2.9}$$

Then, for $n \geq 1$ and for any $q \in S$ we have the following inequalities:

$$\langle v_n, x_n - q\rangle_M \geq (\lambda w)\alpha_{i_c} \|B(z_{n-1}) - B(q)\|_{L^{-1}}^2 + \frac{(1-w)^2}{w}\|v_n\|_M^2, \tag{2.10a}$$

$$\langle \dot{v}_{n+1}, \dot{x}_{n+1}\rangle_M \geq (\lambda w)\alpha_{i_c} \|B(z_n) - B(z_{n-1})\|_{L^{-1}}^2 + \frac{(1-w)^2}{w}\|\dot{v}_{n+1}\|_M^2, \tag{2.10b}$$

where α_{i_c} is given by (2.8).

Proof. For convenience of the reader we set $\zeta_n = G_\lambda^M(z_n)$. Clearly, for $n \geq 1$, we have $v_n = \lambda w \zeta_{n-1}$ (from (2.5a)), or equivalently $x_n = z_{n-1} - \lambda w \zeta_{n-1}$ (as we have $v_n := z_{n-1} - x_n$). It follows immediately for $q \in S$ that

$$\langle v_n, x_n - q\rangle_M = (\lambda w)\langle \zeta_{n-1}, x_n - q\rangle_M = (\lambda w)\langle \zeta_{n-1}, z_{n-1} - q\rangle_M - (\lambda w)^2 \|\zeta_{n-1}\|_M^2.$$

Moreover, by Proposition 2.5 and recalling that $G_\lambda^M(q) = 0$ we have (for $i = 1, 2$)

$$\langle \zeta_{n-1}, z_{n-1} - q\rangle_M \geq \alpha_i \|B(z_{n-1}) - B(q)\|_{L^{-1}}^2 + \lambda \langle H_i \zeta_{n-1}, \zeta_{n-1}\rangle.$$

Consequently, by the previous statements, we deduce that

$$\begin{aligned}&\langle v_n, x_n - q\rangle_M\\&\ge (\lambda w)\alpha_i\|B(z_{n-1}) - B(q)\|_{L^{-1}}^2 + \lambda^2 w\langle H_i\zeta_{n-1}, \zeta_{n-1}\rangle - (\lambda w)^2\|\zeta_{n-1}\|_M^2.\end{aligned} \tag{2.11}$$

In addition, concerning the last two terms in the right side of the previous inequality we have

$$\begin{aligned}&\lambda^2 w\langle H_i\zeta_{n-1}, \zeta_{n-1}\rangle - (\lambda w)^2\|\zeta_{n-1}\|_M^2\\&= \Big\langle \big(\lambda^2 wH_i - (\lambda w)^2 M)\big)\,\zeta_{n-1}, \zeta_{n-1}\Big\rangle = \lambda^2 w\Big\langle\, (H_i - wM)\,\zeta_{n-1}, \zeta_{n-1})\Big\rangle,\\&= \lambda^2 w(1-w)\big\langle \bar{J}_i\zeta_{n-1}, \zeta_{n-1}\big\rangle,\end{aligned} \tag{2.12}$$

where $\bar{J}_i = (1-w)^{-1}(H_i - wM)$ (for $i = 1, 2$). Moreover, by $H_1 = M - \delta I$ and $H_2 = M - \lambda L$ (in light of (2.8)), we get

$$\bar{J}_1 = (1-w)^{-1}(H_1 - wM) = (1-w)^{-1}(M - \delta I - wM) = (1-w)M + w\bar{M}_1,$$

together with

$$\bar{J}_2 = (1-w)^{-1}(H_2 - wM) = (1-w)^{-1}(M - \lambda L - wM) = (1-w)M + w\bar{M}_2.$$

This, combined with (2.11) and (2.12), leads immediately to (2.10a). Next, for $n \ge 1$, by (2.5a) we obviously have that $\dot{v}_{n+1} = \lambda w(\zeta_n - \zeta_{n-1})$, or equivalently $\dot{x}_{n+1} = \dot{z}_n - \lambda w(\zeta_n - \zeta_{n-1})$ (since $v_{n+1} := z_n - x_{n+1}$). Then we immediately see that

$$\langle \dot{v}_{n+1}, \dot{x}_{n+1}\rangle_M = \lambda w\langle \zeta_n - \zeta_{n-1}, \dot{z}_n\rangle_M - (\lambda w)^2\|\zeta_n - \zeta_{n-1}\|_M^2.$$

In addition, by Proposition 2.5, we have (for $i = 1, 2$)

$$\langle \dot{\zeta}_n, \dot{z}_n\rangle_M \ge \alpha_i\|B(z_n) - B(z_{n-1})\|_{L^{-1}}^2 + \lambda\langle H_i\dot{\zeta}_n, \dot{\zeta}_n\rangle.$$

Then by the previous two results we are led to

$$\begin{aligned}&\langle \dot{v}_{n+1}, \dot{x}_{n+1}\rangle_M - (\lambda w)\alpha_i\|B(z_n) - B(z_{n-1})\|_{L^{-1}}^2\\&\ge \langle(\lambda^2 wH_i - (\lambda w)^2 M)\dot{\zeta}_n, \dot{\zeta}_n\rangle = \lambda^2 w\langle(H_i - wM)\dot{\zeta}_n, \dot{\zeta}_n\rangle\\&= \lambda^2 w(1-w)\langle \bar{J}_i\dot{\zeta}_n, \dot{\zeta}_n\rangle,\end{aligned}$$

where $\bar{J}_i$ was previously introduced. This readily amounts to (2.10b). □

2.3. Links between the iterates and the graph of $(A + B)$

Consider the elements y_n and y_n^* defined by

$$y_n = -\tfrac{1}{w}v_n + z_{n-1} \quad\text{and}\quad y_n^* = (\lambda w)^{-1}Mv_n + B(y_n) - B(z_{n-1}). \tag{2.13}$$

The following result makes the connection between y_n^* and $(A + B)(y_n)$.

Proposition 2.8. *Let $\{x_n, v_n\}$ be produced by CRIFBA. Then, for $n \ge 0$, we have*

$$(\lambda w)^{-1}Mv_{n+1} \in B(z_n) + A(y_{n+1}), \tag{2.14a}$$

$$y_{n+1}^* \in B(y_{n+1}) + A(y_{n+1}). \tag{2.14b}$$

If, in addition, M and L are bounded and such that $M-\rho L$ or $M-\rho I$ is positive definite (for some $\rho>0$), then

$$\|y^*_{n+1}\|_M \leq \frac{1}{w}\left(\frac{1}{\lambda}\|M\| + \frac{1}{\rho^{1/2}}(\|M\|.\|L\|)^{\frac{1}{2}}(1+\|L\|^{\frac{1}{2}})\right)\|v_{n+1}\|_M. \tag{2.15}$$

The proof of Proposition 2.8 is given in the Appendix A–2.

3. Convergence analysis of CRIFBA

A series of estimates are obtained here by means of a Lyapunov analysis but also by using the reformulation of CRIFBA given in Proposition 2.4. The main results of Theorem 2.1 will be derived as a combination of these estimations. As standing assumptions we assume that $L:\mathcal{H}\to\mathcal{H}$ and $M:\mathcal{H}\to\mathcal{H}$ are linear self-adjoint and positive definite maps and that $A:\mathcal{H}\to 2^{\mathcal{H}}$ and $B:\mathcal{H}\to\mathcal{H}$ are maximally monotone operators.

3.1. Estimates from an energy-like sequence

With the iterates $\{x_n, v_n\}$ produced by CRIFBA, we associate the sequence $(\mathcal{E}_n(s,q))$ defined for $(s,q)\in(0,\infty)\times S$ and for $n\geq 0$ by

$$\begin{aligned}\mathcal{E}_n(s,q) = \frac{1}{2}\|s(q-x_n)-\nu_n\dot{x}_n\|_M^2 + \frac{1}{2}s(e-s)\|x_n-q\|_M^2 \\ + s(e+\nu_n)\langle v_n, x_n-q\rangle_M.\end{aligned} \tag{3.1}$$

3.1.1. Properties of the energy-like sequence

Our Lyapunov analysis will be based upon the following lemma.

Lemma 3.1. *Suppose that (1.4) holds and that $\{x_n, v_n\}\subset\mathcal{H}$ are generated by* CRIFBA *with parameters $\{s_0,s_1\}\subset[0,\infty)$ and $\{e,\nu_0,w,\lambda\}\subset(0,\infty)$ such that*

$$0\leq s_1<s_0<e, \quad 0<w<1. \tag{3.2}$$

Suppose further that there exists $i_c\in\{1,2\}$ such that the following condition holds:

$$\alpha_{i_c} \text{ is non-negative and } \bar{M}_{i_c} \text{ is positive definite,} \tag{3.3}$$

where α_{i_c} and $\bar{M}_{i_c}$ are given by (2.8) and (2.9), respectively.
Then, for any $(s,q)\in(0,e]\times\mathcal{H}$ and for $n\geq 1$, we have

$$\begin{aligned}&\dot{\mathcal{E}}_{n+1}(s,q) + s(s_0-s_1)\langle v_n, x_n-q\rangle_M \\ &+\frac{(1-w)}{w}(e+\nu_{n+1})(e-s+\nu_{n+1})\|\dot{v}_{n+1}\|_M^2 + \frac{1}{2}(e+\nu_{n+1})\|\dot{x}_{n+1}-\theta_n\dot{x}_n\|_M^2 \\ &+(s_0-s)(e+\nu_{n+1})\langle v_n, \dot{x}_{n+1}\rangle_M + \frac{1}{2}(e-s)(e+2\nu_{n+1})\|\dot{x}_{n+1}\|_M^2 \leq 0.\end{aligned} \tag{3.4}$$

In particular, for $q\in S$, the sequence $(\mathcal{E}_n(s_0,q))_{n\geq1}$ is non-increasing and convergent, and the following estimates are reached:

$$\sup_{n\geq1}\|x_n-q\|_M^2 \leq \frac{2\mathcal{E}_1(s_0,q)}{s_0(e-s_0)}, \tag{3.5a}$$

$$\sup_{n\geq1}\nu_n\langle v_n, x_n-q\rangle_M \leq \frac{\mathcal{E}_1(s_0,q)}{s_0}, \tag{3.5b}$$

$$\sup_n n\|\dot{x}_n\|_M < \infty, \tag{3.5c}$$

$$\sum_{n\geq 1}\nu_{n+1}^2\|\dot{v}_{n+1}\|_M^2 \leq \frac{w\mathcal{E}_1(s_0,q)}{(1-w)^2}, \tag{3.5d}$$

$$\sum_{n\geq 1}\nu_{n+1}\|\dot{x}_{n+1}\|_M^2 \leq \frac{\mathcal{E}_1(s_0,q)}{e-s_0}, \tag{3.5e}$$

$$\sum_{n\geq 1}\langle v_n, x_n - q\rangle_M \leq \frac{\mathcal{E}_1(s_0,q)}{s_0(s_0-s_1)}, \tag{3.5f}$$

$$\sum_n n^2\|\dot{x}_{n+1} - \dot{x}_n\|_M^2 < \infty. \tag{3.5g}$$

3.1.2. Proof of Lemma 3.1

3.1.2.1 Part 1: a useful equality for a Lyapunov analysis

We recall here an equality of independent interest which is related to our method through the wider framework of sequences $\{x_n, d_n\} \subset \mathcal{H}$ and $\{e, \nu_n, \theta_n\} \subset (0,\infty)$ satisfying (for $n \geq 0$)

$$\dot{x}_{n+1} - \theta_n\dot{x}_n + d_n = 0, \tag{3.6a}$$

$$(e + \nu_{n+1})\theta_n = \nu_n. \tag{3.6b}$$

Basic properties regarding these iterates are established through the following result stated in [31, Proposition 2.3].

Proposition 3.2. *Let $\{x_n, d_n\} \subset \mathcal{H}$ and $\{\theta_n, \nu_n, e\} \subset (0,\infty)$ satisfy (3.6), and suppose that $M : \mathcal{H} \to \mathcal{H}$ is a linear self-adjoint and positive definite map. Then for $(s,q) \in (0,e] \times \mathcal{H}$ and for $n \geq 0$ we have*

$$\begin{aligned}&\dot{F}_{n+1}(s,q) + \frac{1}{2}(e+\nu_{n+1})^2\|\dot{x}_{n+1} - \theta_n\dot{x}_n\|_M^2 + s(e+\nu_{n+1})\langle d_n, x_{n+1} - q\rangle_M \\ &+ (e+\nu_{n+1})(e - s + \nu_{n+1})\langle d_n, \dot{x}_{n+1}\rangle_M = -\frac{1}{2}(e-s)(e+2\nu_{n+1})\|\dot{x}_{n+1}\|_M^2,\end{aligned} \tag{3.7}$$

where $F_n(s,q)$ is defined by

$$F_n(s,q) = \frac{1}{2}\|s(q-x_n) - \nu_n\dot{x}_n\|_M^2 + \frac{1}{2}s(e-s)\|x_n - q\|_M^2. \tag{3.8}$$

3.1.2.2 Part 2: a main inequality and related estimates

Let us begin by proving (3.4). It can be observed (from (2.5b)) that the iterates $\{x_n, v_n\}$ generated by CRIFBA enter the special case of the general iterative process (3.6) when taking $d_n = v_{n+1} - \gamma_n v_n$. Hence, for $n \geq 0$, by Proposition 3.2 we obtain

$$\begin{aligned}&\dot{F}_{n+1}(s,q) + \frac{1}{2}\tau_n^2\|W_n\|_M^2 + (s\tau_n)\langle d_n, x_{n+1} - q\rangle_M + (\tau_n^2\vartheta_n)\langle d_n, \dot{x}_{n+1}\rangle_M \\ &= -\frac{1}{2}(e-s)(e+2\nu_{n+1})\|\dot{x}_{n+1}\|_M^2.\end{aligned} \tag{3.9}$$

where $\tau_n = e + \nu_{n+1}$, $\vartheta_n = 1 - s\tau_n^{-1}$ and $W_n = \dot{x}_{n+1} - \theta_n\dot{x}_n$. Let us evaluate the quantity $(s\tau_n)\langle d_n, x_{n+1} - q\rangle_M + (\tau_n^2\vartheta_n)\langle d_n, \dot{x}_{n+1}\rangle_M$. Setting $U_n = \langle v_n, x_n - q\rangle_M$, by $d_n = v_{n+1} - \gamma_n v_n$ we have

$$\langle d_n, x_{n+1} - q\rangle_M = U_{n+1} + \langle -\gamma_n v_n, \dot{x}_{n+1}\rangle_M - \gamma_n U_n, \tag{3.10}$$

$$\langle d_n, \dot{x}_{n+1}\rangle_M = \langle v_{n+1}, \dot{x}_{n+1}\rangle_M + \langle -\gamma_n v_n, \dot{x}_{n+1}\rangle_M. \tag{3.11}$$

This, noticing that $s\tau_n + \tau_n^2\vartheta_n = \tau_n^2$, amounts to

$$(s\tau_n)\langle d_n, x_{n+1} - q\rangle_M + (\tau_n^2\vartheta_n)\langle d_n, \dot{x}_{n+1}\rangle_M$$
$$= (s\tau_n)U_{n+1} - (s\tau_n)\gamma_n U_n + (\tau_n^2\vartheta_n)\langle v_{n+1}, \dot{x}_{n+1}\rangle_M + \tau_n^2\langle -\gamma_n v_n, \dot{x}_{n+1}\rangle_M.$$

In addition, we obviously have

$$\langle v_{n+1}, \dot{x}_{n+1}\rangle_M = \langle \dot{v}_{n+1}, \dot{x}_{n+1}\rangle_M + \langle v_n, \dot{x}_{n+1}\rangle_M.$$

Then, by the previous arguments we obtain

$$\begin{aligned}
&(s\tau_n)\langle d_n, x_{n+1} - q\rangle_M + (\tau_n^2\vartheta_n)\langle d_n, \dot{x}_{n+1}\rangle_M \\
&= (s\tau_n)U_{n+1} - \gamma_n(s\tau_n)U_n + (\tau_n^2\vartheta_n)\langle v_{n+1}, \dot{x}_{n+1}\rangle_M - \gamma_n\tau_n^2\langle v_n, \dot{x}_{n+1}\rangle_M \\
&= (s\tau_n)U_{n+1} - \gamma_n(s\tau_n)U_n + (\tau_n^2\vartheta_n)\langle \dot{v}_{n+1}, \dot{x}_{n+1}\rangle_M + \tau_n^2(\vartheta_n - \gamma_n)\langle v_n, \dot{x}_{n+1}\rangle_M.
\end{aligned}$$

Furthermore, observing that $\gamma_n = 1 - s_0\tau_n^{-1}$ and that $\dot{\tau}_n := \tau_n - \tau_{n-1} = s_1$, an easy computation gives us

$$\gamma_n\tau_n = \tau_n - s_0 = \tau_{n-1} - (s_0 - s_1) \quad \text{and} \quad \tau_n(\vartheta_n - \gamma_n) = (s_0 - s).$$

In addition, for $n \geq 1$, by Proposition 2.7 and setting $\eta = (1-w)^2/w$, we have

$$\langle \dot{v}_{n+1}, x_{n+1}\rangle_M \geq \eta\|\dot{v}_{n+1}\|_M^2.$$

Therefore combining the previous results amounts to

$$\begin{aligned}
(s\tau_n)\langle d_n, x_{n+1} - q\rangle_M + (\tau_n^2\vartheta_n)\langle d_n, \dot{x}_{n+1}\rangle_M &\geq (s\tau_n)U_{n+1} - (s\tau_{n-1})U_n \\
+s(s_0 - s_1)U_n + \eta(\tau_n^2\vartheta_n)\|\dot{v}_{n+1}\|_M^2 &+ \tau_n(s_0 - s)\langle v_n, \dot{x}_{n+1}\rangle_M. \qquad (3.12)
\end{aligned}$$

Consequently, in light of (3.9), we deduce for $n \geq 1$ that

$$\begin{aligned}
&\dot{F}_{n+1}(s,q) + (s\tau_n)U_{n+1} - (s\tau_{n-1})U_n + s(s_0 - s_1)U_n + \eta\tau_n^2\vartheta_n\|\dot{v}_{n+1}\|_M^2 \\
&\quad +\tau_n\,(s_0 - s)\,\langle v_n, \dot{x}_{n+1}\rangle_M \leq -\frac{1}{2}\,(e - s)\,(e + 2\nu_{n+1})\|\dot{x}_{n+1}\|_M^2.
\end{aligned}$$

This, from $\mathcal{E}_n(s,q) = F_n(s,q) + (s\tau_{n-1})U_n$ (in light of (3.1)), can be rewritten as (3.4).

Now we prove the second part of Lemma 3.1. For $q \in S$ and $s = s_0$, inequality (3.4) becomes (for $n \geq 1$)

$$\begin{aligned}
&\dot{\mathcal{E}}_{n+1}(s_0,q) + s_0(s_0 - s_1)\langle v_n, x_n - q\rangle_M + \eta\tau_n^2\gamma_n\|\dot{v}_{n+1}\|_M^2 \\
&\quad +\frac{1}{2}\tau_n^2\|W_n\|_M^2 + \frac{1}{2}\,(e - s_0)\,(e + 2\nu_{n+1})\|\dot{x}_{n+1}\|_M^2 \leq 0. \qquad (3.13)
\end{aligned}$$

Clearly, we know that U_n is non-negative (from Proposition 2.7). It follows immediately that the non-negative sequence $(\mathcal{E}_n(s_0,q))_{n\geq 1}$ is non-increasing, since the constants $s_0 - s_1$ and $e - s_0$ are non-negative (in light of condition (3.2)). Whence it is convergent and bounded. Also recall from (3.1) that

$$\begin{aligned}
\mathcal{E}_n(s_0,q) &= \frac{1}{2}\|s_0(q - x_n) - \nu_n\dot{x}_n\|_M^2 \\
&+\frac{1}{2}s_0(e - s_0)\|x_n - q\|_M^2 + s_0(e + 2\nu_n)\langle v_n, x_n - q\rangle_M. \qquad (3.14)
\end{aligned}$$

Then, for $n \geq 1$, by the inequality $\mathcal{E}_n(s_0, q) \leq \mathcal{E}_1(s_0, q)$ we get

$$\frac{1}{2}s_0(e - s_0)\|x_n - q\|_M^2 \leq \mathcal{E}_1(s_0, q) \tag{3.15}$$

$$s_0(e + 2\nu_n)\langle v_n, x_n - q\rangle_M \leq \mathcal{E}_1(s_0, q), \tag{3.16}$$

$$\|\nu_n \dot{x}_n\|_M - \|s_0(q - x_n)\|_M \leq \sqrt{\mathcal{E}_1(s_0, q)}. \tag{3.17}$$

Estimates (3.5a), (3.5b) and (3.5c) are direct consequences of these last three inequalities. Moreover, given some integer $N \geq 1$, by adding (3.13) from $n = 1$ to $n = N$ we obtain

$$\begin{aligned} &\mathcal{E}_{N+1}(s_0, q) + s_0(s_0 - s_1)\sum_{n=1}^{N}\langle v_n, x_n - q\rangle_M + \eta\sum_{n=1}^{N}\tau_n^2\gamma_n\|\dot{v}_{n+1}\|_M^2 \\ &+\frac{1}{2}\sum_{n=1}^{N}\tau_n^2\|\dot{x}_{n+1} - \theta_n\dot{x}_n\|_M^2 + \frac{1}{2}(e - s_0)\sum_{n=1}^{N}(e + 2\nu_{n+1})\|\dot{x}_{n+1}\|_M^2 \leq \mathcal{E}_1(s_0, q). \end{aligned} \tag{3.18}$$

It follows immediately that

$$\eta\textstyle\sum_{n=1}^{N}\tau_n^2\gamma_n\|\dot{v}_{n+1}\|_M^2 \leq \mathcal{E}_1(s_0, q), \tag{3.19a}$$

$$\frac{1}{2}(e - s_0)\textstyle\sum_{n=1}^{N}(e + 2\nu_{n+1})\|\dot{x}_{n+1}\|_M^2 \leq \mathcal{E}_1(s_0, q), \tag{3.19b}$$

$$s_0(s_0 - s_1)\textstyle\sum_{n=1}^{N}\langle v_n, x_n - q\rangle_M \leq \mathcal{E}_1(s_0, q), \tag{3.19c}$$

$$\frac{1}{2}\textstyle\sum_{n=1}^{N}\tau_n^2\|\dot{x}_{n+1} - \theta_n\dot{x}_n\|_M^2 \leq \mathcal{E}_1(s_0, q). \tag{3.19d}$$

This straightforwardly yields (3.5d), (3.5e) and (3.5f). The last estimate (3.5g) is simply deduced from (3.19d) and (3.5e) (in light of the definition of θ_n) □

3.2. Estimates from the reformulation of the method

Additional estimates are established regarding CRIFBA, especially on the sequence $(v_n + \dot{x}_n)$.

Lemma 3.3. *Let $\{x_n, v_n\} \subset \mathcal{H}$ be given by CRIFBA with $\{s_1, s_0\} \subset [0, \infty)$ and $\{e, \nu_0, w, \lambda\} \subset (0, \infty)$ satisfying $0 < s_0 < e$. Then, for any $(s, q) \in (0, e] \times \mathcal{H}$ and for $n \geq 1$, we have*

$$\begin{aligned} &(e + \nu_{n+1})^2\|v_{n+1} + \dot{x}_{n+1}\|_M^2 - (e + \nu_n)^2\|v_n + \dot{x}_n\|_M^2 \\ &+(s_0 - 2s_1)(e + \nu_{n+1})\|v_n + \dot{x}_n\|_M^2 \leq s_0^{-1}(e - s_0 + s_1)^2(e + \nu_{n+1})\|\dot{x}_n\|_M^2. \end{aligned} \tag{3.20}$$

Suppose, in addition, that condition (3.3) *holds and that the parameters satisfy*

$$0 \leq s_1 < (1/2)s_0 \quad \textit{and} \quad 0 < w < 1. \tag{3.21}$$

Then the following estimates are reached

$$\textstyle\sum_n n\|\dot{x}_n + v_n\|_M^2 < \infty, \tag{3.22a}$$

$$\|\dot{x}_n + v_n\|_M = o(n^{-1}). \tag{3.22b}$$

Proof. For $n \geq 1$, according to Lemma 2.4 and denoting $\tau_n = e + \nu_{n+1}$ we have $v_{n+1} + \dot{x}_{n+1} = \theta_n\dot{x}_n + \gamma_n v_n$, together with $\gamma_n = 1 - s_0\tau_n^{-1}$ and

$$\theta_n = \nu_n\tau_n^{-1} = 1 - (e + s_1)\tau_n^{-1},$$

which yields $v_{n+1}+\dot{x}_{n+1} = \gamma_n(v_n+\dot{x}_n)+(\theta_n-\gamma_n)\dot{x}_n$. Then, setting $H_n = v_n+\dot{x}_n$, we equivalently have $H_{n+1} = \gamma_n H_n+(\theta_n-\gamma_n)\dot{x}_n$, along with $\theta_n-\gamma_n = -(e-s_0+s_1)\tau_n^{-1}$, which amounts to

$$H_{n+1} = \gamma_n H_n + (1-\gamma_n)\,\frac{\theta_n-\gamma_n}{1-\gamma_n}\dot{x}_n.$$

Hence, by convexity of the squared norm we infer that

$$\begin{aligned}\|H_{n+1}\|_M^2 &\leq \gamma_n\|H_n\|_M^2 + (1-\gamma_n)\left(\frac{\gamma_n-\theta_n}{1-\gamma_n}\right)^2\|\dot{x}_n\|_M^2\\ &= \gamma_n\|H_n\|_M^2 + \frac{(e-s_0+s_1)^2}{1-\gamma_n}\tau_n^{-2}\|\dot{x}_n\|_M^2.\end{aligned}$$

Hence we obtain

$$\|H_{n+1}\|_M^2 \leq (1-s_0\tau_n^{-1})\|H_n\|_M^2 + s_0^{-1}(e-s_0+s_1)^2\tau_n^{-1}\|\dot{x}_n\|_M^2.$$

Then multiplying this last inequality by τ_n^2 amounts to

$$\tau_n^2\|H_{n+1}\|_M^2 \leq (\tau_n^2-s_0\tau_n)\|H_n\|_M^2 + s_0^{-1}(e-s_0+s_1)^2\tau_n\|\dot{x}_n\|_M^2,$$

while by $\dot{\tau_n} = s_1$ (from the definitions of τ_n and ν_n) we simply have

$$\tau_n^2-\tau_{n-1}^2 \leq s_1(\tau_n+\tau_{n-1}) \leq 2s_1\tau_n.$$

Combining these last two results amounts to

$$\tau_n^2\|H_{n+1}\|_M^2 \leq \tau_n^2\|H_n\|_M^2 - (s_0\tau_n-2s_1\tau_n)\|H_n\|_M^2 + s_0^{-1}(e-s_0+s_1)^2\tau_n\|\dot{x}_n\|_M^2,$$

which leads to the desired inequality.

Next we prove the second part of the lemma. Clearly, the assumptions of Lemma 3.3 guarantee that $\sum_n \tau_n\|\dot{x}_n\|_M^2 < \infty$ (according to Lemma 3.1). Consequently, by (3.20) under the condition $0 \leq s_1 < (1/2)s_0$, we classically deduce that

$$\textstyle\sum_n \tau_n\|v_n+\dot{x}_n\|_M^2 < \infty$$

(namely (3.22a)) and that the sequence $(\tau_n^2\|v_{n-1}+\dot{x}_{n-1}\|_M^2)$ is convergent.

Thus, there exists $l_1 \geq 0$ such that $\lim_{n\to\infty}\tau_n^2\|v_{n-1}+\dot{x}_{n-1}\|_M^2 = l_1$, hence, we also have $\lim_{n\to\infty}\tau_n^2\|v_n+\dot{x}_n\|_M^2 = l_1$ (since $\frac{\tau_{n-1}}{\tau_n} \to 1$ as $n\to\infty$). So, noticing that $\sum_n \tau_n^{-1} = \infty$, we are led to $l_1 = 0$, which proves (3.22b). □

3.3. Proof of Theorem 2.1

Let us begin by observing that the conditions (2.2a) and (2.2b) correspond to condition (3.3) with $i_c = 1$ and $i_c = 2$, respectively. The rest of the proof will be divided into the following three steps **(B1), (B2)** and **(B3)**:

(B1) In order to reach (2.3a), (2.3b) and (2.3d), we prove the following estimates:

$$\textstyle\sum_n n\|v_n\|_M^2 < \infty, \tag{3.23}$$

$$\textstyle\sum_n n\,|\langle v_n, \dot{x}_{n+1}\rangle_M| < \infty, \tag{3.24}$$

$$\|\dot{x}_n\|_M = o(n^{-1}), \tag{3.25}$$

$$\|v_n\|_M = o(n^{-1}). \tag{3.26}$$

Indeed, from a quick computation, we have

$$n\|v_n\|_M^2 \leq 2n\|v_n+\dot{x}_n\|_M^2 + 2n\|\dot{x}_n\|_M^2. \tag{3.27}$$

So, by $\sum_n n\|\dot{x}_n\|_M^2 < \infty$ (from (3.5e)) and $\sum_n n\|v_n + \dot{x}_n\|_M^2 < \infty$ (from (3.22a)), we immediately obtain (3.23). The estimate (3.24) is an immediate consequence of $\sum_n n\|\dot{x}_n\|_M^2 < \infty$ (from (3.5e)) and $\sum_n n\|v_n\|_M^2 < \infty$ (from (3.23)). Next, passing to the limit as $s \to 0^+$ in (3.4) amounts to

$$\begin{aligned} &\nu_{n+1}^2\|\dot{x}_{n+1}\|_M^2 - \nu_n^2\|\dot{x}_n\|_M^2 \\ &+s_0(e+\nu_{n+1})\langle v_n, \dot{x}_{n+1}\rangle_M + \frac{1}{2}e(e+2\nu_{n+1})\|\dot{x}_{n+1}\|_M^2 \leq 0. \end{aligned} \tag{3.28}$$

Then, in light of $\sum_n n\,|\langle v_n, \dot{x}_{n+1}\rangle_M| < \infty$ (from (3.24)) and $\sum_n n\|\dot{x}_n\|_M^2 < \infty$ (from (3.5e)), we derive, from (3.28), that $(\nu_n^2\|\dot{x}_n\|_M^2)$ is convergent, namely, there exists some $l_2 \geq 0$ such that $\lim_{n\to\infty} \nu_n^2\|\dot{x}_n\|_M^2 = l_2$. Moreover, by $\sum_n \nu_n\|\dot{x}_n\|_M^2 < \infty$, and recalling that $\sum_n \nu_n^{-1} = \infty$, we get $\liminf_{n\to\infty} \nu_n^2\|\dot{x}_n\|_M^2 = 0$. It follows that $l_2 = 0$, that is (3.25). Next, combining this last result with $\|\dot{x}_n + v_n\|_M = o(n^{-1})$ (from (3.22b)) gives us $\lim_{n\to\infty} n\|v_n\|_M = 0$, that is (3.26).

It can be seen that the estimations in (2.3a) are given by (3.25), (3.5e) and (3.5g), respectively. The estimates in (2.3b) follow from (3.26) and (3.23). Furthermore, by $y_n = x_n + \left(1 - \frac{1}{w}\right) v_n$, we obviously have $\dot{y}_n = \dot{x}_n + \left(1 - \frac{1}{w}\right) \dot{v}_n$, which implies that

$$\|\dot{y}_n\|_M^2 \leq 2\|\dot{x}_n\|_M^2 + 2\left(1 - \tfrac{1}{w}\right)^2 \|\dot{v}_n\|_M^2. \tag{3.29}$$

Therefore, by (3.29) in light of $\|\dot{x}_n\|_M^2 = o(n^{-2})$ (from (3.25)) and $\sum_n n^2\|\dot{v}_n\|_M^2 < \infty$ (from (3.5d)), we obtain $\|\dot{y}_n\|_M^2 = o(n^{-2})$, i.e., the first result in (2.3d). Furthermore, by (3.29), along with $\sum_n n\|\dot{x}_n\|_M^2 < \infty$ (from (3.5e)) and $\sum_n n^2\|\dot{v}_n\|_M^2 < \infty$ (from (3.5d)), we are led to $\sum_n n\|\dot{y}_n\|_M^2 < \infty$, that is the second result in (2.3d).

(B2) Let us prove (2.3c) and (2.3e). From (2.3b) and $G_\lambda^M(z_n) = (\lambda w)v_{n+1}$, we readily have

$$\|G_\lambda^M(z_n)\|_M^2 = \mathcal{O}(n^{-2}) \quad\text{and}\quad \textstyle\sum_n n\|G_\lambda^M(z_n)\|_M^2 < \infty. \tag{3.30}$$

In addition, given $\{x, z\} \subset \mathcal{H}$ and setting $\Delta G_\lambda^M(x,z) = G_\lambda^M(x) - G_\lambda^M(z)$, by the simple decomposition $G_\lambda^M(x) = \Delta G_\lambda^M(x,z) + G_\lambda^M(z)$, we readily get

$$\|G_\lambda^M(x)\|_M^2 \leq 2\|\Delta G_\lambda^M(x,z)\|_M^2 + 2\|G_\lambda^M(z)\|_M^2. \tag{3.31}$$

Moreover, from condition (3.3) and (2.7) we have

$$\lambda\langle H_{i_c}\Delta G_\lambda^M(x,z), \Delta G_\lambda^M(x,z)\rangle \leq \langle \Delta G_\lambda^M(x,z), x - z\rangle_M, \tag{3.32}$$

where $H_{i_c} = M - K_{i_c}I$ (with $K_{i_c} = \delta I$ if $i_c = 1$ and $K_{i_c} = \lambda L$ otherwise), hence, by applying Peter-Paul's inequality, we obtain

$$\begin{aligned} &\lambda\langle H_{i_c}\Delta G_\lambda^M(x,z), \Delta G_\lambda^M(x,z)\rangle \\ &\leq (1/2)(\lambda w^2)\|\Delta G_\lambda^M(x,z)\|_M^2 + (1/2)(\lambda w^2)^{-1}\|x-z\|_M^2, \end{aligned}$$

or equivalently

$$\lambda\langle\left(H_{i_c} - (1/2)w^2 M\right)\Delta G_\lambda^M(x,z), \Delta G_\lambda^M(x,z)\rangle \leq (1/2)(\lambda w^2)^{-1}\|x-z\|_M^2. \tag{3.33}$$

Furthermore, we simply have

$$\begin{aligned} H_{i_c} - (1/2)w^2 M &= (1-w)M + (w-w^2)M - K_{i_c} I + (1/2)w^2 M \\ &= (1-w)M + w(1-w)\bar{M}_{i_c} + (1/2)w^2 M \\ &= (1/2)(1+(1-w)^2)M + w(1-w)\bar{M}_{i_c}. \end{aligned}$$

Then, as $\bar{M}_{i_c}$ is positive definite (from condition (3.3)), combining the previous two results yields

$$\lambda \|\Delta G_\lambda^M(x,z)\|_M^2 \le (\lambda w)^{-1} \|x-z\|_M^2, \tag{3.34}$$

which by (3.31) entails that

$$\|G_\lambda^M(x)\|_M^2 \le 2\lambda^{-2} w^{-1} \|x-z\|_M^2 + 2\|G_\lambda^M(z)\|_M^2. \tag{3.35}$$

In particular, using (3.35), by $x_{n+1} - z_n = -\lambda w G_\lambda^M(z_n)$ we obtain

$$\|G_\lambda^M(x_{n+1})\|_M^2 \le 2(w+1)\|G_\lambda^M(z_n)\|_M^2, \tag{3.36}$$

which in light of (3.30) amounts to (2.3c). By $y_{n+1} = x_{n+1} + \left(1 - \frac{1}{w}\right) v_{n+1}$ and using (3.35) again we obtain

$$\|G_\lambda^M(y_{n+1})\|_M^2 \le 2\lambda^{-2} w^{-1} \left(1 - \tfrac{1}{w}\right)^2 \|v_{n+1}\|_M^2 + 2\|G_\lambda^M(x_{n+1})\|_M^2,$$

which in light of (2.3b) and (2.3c) implies (2.3e).

(B3) Let us prove (2.3f) and (2.3g). From Proposition 2.8, we know that there exists a sequence $(y_n^*) \subset \mathcal{H}$ (given by (2.8)) satisfying

$$y_n^* \in (A+B)(y_n), \tag{3.37}$$

where $y_n = x_n + \left(1 - \frac{1}{w}\right) v_n$. It can also be noticed from (2.15) that $\|y_n^*\|_M = \mathcal{O}(n^{-1})$, since $\|v_n\|_M = \mathcal{O}(n^{-1})$ (from (3.26)). This leads to (2.3f). At once, we prove (2.3g), by means of the well-known Opial lemma which guarantees that (x_n) converges to some element of S, provided that the following results hold:

(h1) for any $q \in S$, the sequence $(\|x_n - q\|_M)$ is convergent,

(h2) any weak-cluster point of (x_n), in $(\mathcal{H}, \|.\|_M)$, belongs to S.

Let us prove (h1). Take $q \in S$. Clearly, as a straightforward consequence of the boundedness of (x_n) (given by (3.5a)) along with (3.26) we have

$$\langle v_n, x_n - q\rangle_M = o(n^{-1}). \tag{3.38}$$

Moreover, we know that $(\mathcal{E}_n(s_0,q))$ is convergent (from Lemma 3.1), together with

$$\mathcal{E}_n(s_0,q) = \frac{1}{2}\|s_0(q-x_n) - \nu_n \dot{x}_n\|_M^2 + \frac{1}{2}\beta_{s_0}\|x_n - q\|_M^2 + s_0(e+2\nu_n)\langle v_n, x_n - q\rangle_M, \tag{3.39}$$

where $\beta_{s_0} = s_0(e - s_0)$. Then we conclude from $\nu_n \|\dot{x}_n\|_M \to 0$ (from (3.25)) and $(e+2\nu_n)\langle v_n, x_n - q\rangle_M \to 0$ (according to (3.38)) as $n \to \infty$ that

$$\lim_{n\to\infty} \mathcal{E}_n(s_0,q) = \lim_{n\to\infty} \tfrac{1}{2} s_0 e \|x_n - q\|_M^2. \tag{3.40}$$

This entails (h1). Now, we prove (h2). Let u be a weak cluster point of (x_n) in $(\mathcal{H}, \|.\|_M)$, namely there exists a subsequence (x_{n_k}) that converges weakly to u in

$(\mathcal{H}, \|.\|_M)$, as $k \to \infty$. Observe that, as $n \to \infty$, by $y_n = x_n + \left(1 - \frac{1}{w}\right) v_n$ and $\|v_n\|_M \to 0$ (from (3.26)) we have $\|y_n - x_n\|_M \to 0$, whence, (y_{n_k}) converges weakly to u in $(\mathcal{H}, \|.\|_M)$ (as $k \to \infty$). Moreover, by (2.3f) we know that $\|y_n^*\|_M \to 0$ (as $n \to \infty$), while (3.37) gives us

$$y_{n_k}^* \in (A + B)(y_{n_k}). \tag{3.41}$$

Then passing to the limit as $k \to \infty$ in (3.41) and recalling that the graph of the maximally monotone operator $A+B$ is sequentially closed with respect to the weak-strong topology of the product space $\mathcal{H} \times \mathcal{H}$ (see, for instance, [9]), we deduce that $0 \in (A + B)(u)$, namely $u \in S$. This proves (h2) and completes the proof □

4. From CRIFBA to its generalized variant G-CRIFBA

Our purpose here is to adapt CRIFBA to the general structured monotone inclusion problem

$$\text{find} \quad \bar{x} \in S := \left(B + \sum_{k=1}^{p} A_k\right)^{-1} (0) \neq \emptyset, \tag{4.1}$$

where $B : \mathcal{H} \to \mathcal{H}$ is β-co-coercive on $\mathcal{H}$, while $(A_i)_{i=1}^p : \mathcal{H} \to \mathcal{H}$ is a family of p maximally monotone operators whose resolvent operators can be easily evaluated. For the sake of clarity we do not include pre-conditioning in the proposed method. To deal with this problem, we follow the methodology of Raguet-Fadili-Peyre [40], as described below.

Let $\{\rho_k\}_{k=1}^p \subset (0, 1)$ be such that $\sum_{k=1}^p \rho_k = 1$ and consider the Hilbert space $E = \mathcal{H}^p$ endowed with the scalar product $(.|.)_{\mathcal{H}^p}$ defined, for elements $x = (x_k)_{k=1}^p$ and $y = (y_k)_{k=1}^p$ belonging to E, by $(x|y)_{\mathcal{H}^p} = \sum_{k=1}^p \rho_k \langle x_k, y_k\rangle$. The induced norm of $(.|.)_{\mathcal{H}^p}$ will be denoted by $\|.\|_{\mathcal{H}^p}$.

Consider also the auxiliary problem

$$\text{find} \quad \bar{z} \in S_p := \{z \in \mathcal{H}^p \ ; \sum_{i=1}^{p} \rho_i z_i \in S\}, \tag{4.2}$$

which was shown to have a nonempty solution set S_p (whenever $S \neq \emptyset$). It can also be reformulated as a monotone inclusion that fits the structure (1.1) and (1.4) on E. Introduce indeed the mappings $\bar{A}_G$ and $\tilde{B}_G$ from E onto E defined for $(x_i)_{i=1}^p \in E$ by

$$\bar{A}_G\Big((x_i)_{i=1}^p\Big) = \left(\frac{\lambda}{\rho_i} A_i(x_i)\right)_{i=1}^p, \quad \tilde{B}_G\Big((x_i)_{i=1}^p\Big) = \Big(B(x_i)\Big)_{i=1}^p, \tag{4.3}$$

and let $N_\Gamma : \mathcal{H}^p \to 2^{\mathcal{H}^p}$ be the normal cone to the (nonempty) closed convex set

$$\Gamma = \{\big(x_i\big)_{i=1}^p \in \mathcal{H}^p \ ; \ x_1 = x_2 = ... = x_p\}.$$

It can be checked that $\bar{A}_G$ and $\tilde{B}_G$ are maximally monotone operator on E. So the reflection operators $R_{\bar{A}_G} = 2J_{\bar{A}_G} - I_{\mathcal{H}^p}$ and $R_{N_\Gamma} = 2J_{N_\Gamma} - I_{\mathcal{H}^p}$ are well-defined, which allows us to consider the mappings T_1, T_2 and T, from E onto E and such that

$$T_1 = \frac{1}{2}(R_{\bar{A}_G} \circ R_{N_\Gamma} + I_{\mathcal{H}^p}), \quad T_2 = I_{\mathcal{H}^p} - \lambda \tilde{B}_G \circ J_{N_\Gamma} \quad \text{and} \quad T = T_1 \circ T_2. \tag{4.4}$$

The results given in the following proposition were established in [40].

Proposition 4.1. *The following statements are obtained:*

(1) (see [40, Propositions 4.1, 4.2 and 4.6]) *There exists some maximally monotone operator $\mathcal{A} : E \to 2^E$ such that*

$$S_p = \left(\mathcal{A} + \tilde{B}_G \circ J_{N_\Gamma}\right)^{-1}(0). \tag{4.5}$$

(2) (see the proof of [40, Proposition 4.1]) *The operator $\tilde{B}_G \circ J_{N_\Gamma}$ is β co-coercive.*

(3) *S_p is nothing but the fixed point set of the operator T which can be rewritten as*

$$T = J_{\mathcal{A}} \circ (I_{\mathcal{H}^p} - \lambda \tilde{B}_G \circ J_{N_\Gamma}). \tag{4.6}$$

Consequently, a strategy to solve (4.1) consists first of approaching an element of S_p (namely a fixed point of T) by means of sequences

$$(z_n) = \left((z_{n,k})_{k=1}^p\right) \subset \mathcal{H}^p \text{ and } (\zeta_n) = \left((\zeta_{n,k})_{k=1}^p\right) \subset \mathcal{H}^p$$

generated by CRIFBA, in the context of (4.5)–(4.6), as follows:

$$z_n = \zeta_n + \theta_n(\zeta_n - \zeta_{n-1}) + \gamma_n(z_{n-1} - \zeta_n), \tag{4.7a}$$

$$\zeta_{n+1} = (1-w)z_n + wT(z_n). \tag{4.7b}$$

Next, we derive an element of S as the limit of $(x_n) \subset \mathcal{H}$ given by $x_n = \sum_{k=1}^p \rho_k \zeta_{n,k}$. This leads us (see the proof of Theorem 4.2) to the algorithm (G-CRIFBA) given below:

(G-CRIFBA):

▷ **Step 1** (initialization):
Let $\{z_{-1}, \zeta_{-1}, \zeta_0\} \subset \mathcal{H}^p$, $\{e, s_0, s_1, \nu_0, \lambda, w\} \subset [0, \infty)$, $\{\rho_k\}_{k=1}^p \subset (0,1)$, and set $\nu_n = s_1 n + \nu_0$, $\theta_n = 1 - \frac{e+s_1}{e+\nu_{n+1}}$ and $\gamma_n = 1 - \frac{s_0}{e+\nu_{n+1}}$.

▷ **Step 2** (main step):
Given $\{z_{n-1}, \zeta_{n-1}, \zeta_n\} \subset \mathcal{H}^p$ (with $n \geq 0$), we compute

$$z_n = \zeta_n + \theta_n(\zeta_n - \zeta_{n-1}) + \gamma_n(z_{n-1} - \zeta_n), \tag{4.8a}$$

$$u_n = \textstyle\sum_{k=1}^p \rho_k z_{n,k}, \tag{4.8b}$$

$$\zeta_{n+1} = \left(z_{n,k} + w\left(J_{\frac{\lambda}{\rho_k} A_k}(2u_n - \lambda B(u_n) - z_{n,k}) - u_n\right)\right)_{k=1}^p, \tag{4.8c}$$

$$x_{n+1} = \textstyle\sum_{k=1}^p \rho_k \zeta_{n,k}. \tag{4.8d}$$

The next theorem sets the convergence rates of the iterates (ζ_n) generated by G-CRIFBA in terms of discrete velocity and fixed point residual to $\|\dot{\zeta}_n\|_{\mathcal{H}^p} = o(n^{-1})$ and $\|T(\zeta_n) - \zeta_n\|_{\mathcal{H}^p} = o(n^{-1})$, respectively, instead of the convergence rates $\|\dot{\zeta}_n\|_{\mathcal{H}^p} = \mathcal{O}(n^{-1/2})$ and $\|T(\zeta_n) - \zeta_n\|_{\mathcal{H}^p} = \mathcal{O}(n^{-1/2})$ obtained for classical fixed point iterations of T as in [40] (that is (4.8) with $w = 1$ and $\theta_n = \gamma_n = 0$).

Theorem 4.2. *Let $\{x_n\} \subset \mathcal{H}$ be generated by* G-CRIFBA *with $\{\rho_k\}_{k=1}^p \subset (0,1)$ satisfying $\sum_{k=1}^p \rho_k = 1$, together with the other parameters such that*

$$0 < \lambda < 4w(1-w)\beta, \quad 0 < w < 1, \quad 2s_1 < s_0 < e. \tag{4.9}$$

Then the following results are reached:

$$\|\dot{\zeta}_n\|_{\mathcal{H}^p} = o(n^{-1}), \quad \textstyle\sum_n n\|\dot{\zeta}_n\|^2_{\mathcal{H}^p} < \infty, \quad \sum_n n^2\|\dot{\zeta}_{n+1} - \dot{\zeta}_n\|^2_{\mathcal{H}^p} < \infty, \tag{4.10a}$$

$$\|\zeta_{n+1} - z_n\|_{\mathcal{H}^p} = o(n^{-1}), \quad \textstyle\sum_n n\|\zeta_{n+1} - z_n\|^2_{\mathcal{H}^p} < \infty, \tag{4.10b}$$

$$\|T(\zeta_n) - \zeta_n\|_{\mathcal{H}^p} = o(n^{-1}), \quad \textstyle\sum_n n\|T(\zeta_n) - \zeta_n\|^2_{\mathcal{H}^p} < \infty, \tag{4.10c}$$

$$\exists \bar{\zeta} \in S_p, \text{ s.t. (for } k = 1, .., p) \ \bar{\zeta}_{n,k} \rightharpoonup \bar{\zeta}_k \text{ weakly in } \mathcal{H}, \text{ as } n \to \infty, \tag{4.10d}$$

$$x_n \rightharpoonup \bar{x} = \textstyle\sum_{k=1}^p \rho_k \bar{\zeta}_k \in S \text{ weakly in } \mathcal{H}. \tag{4.10e}$$

Proof. Let us evaluate the operator T on E from its formulation given by (4.4), namely $T = T_1 \circ T_2$. It can be checked (see [40, Lemma 4.1]) that, for $(x_k)_{k=1}^p \in E$, we have

$$J_{N_\Gamma}\big(\{x_i\}_{i=1}^p\big) = \big(\textstyle\sum_{j=1}^p \rho_j x_j\big)_{i=1}^p, \tag{4.11}$$

$$R_{N_\Gamma}\big(\{x_k\}_{k=1}^p\big) = \big(2\textstyle\sum_{i=1}^p \rho_i x_i - x_k\big)_{k=1}^p, \tag{4.12}$$

$$R_{\bar{A}_G}\big(\{x_k\}_{k=1}^p\big) = \big(2J_{\frac{\lambda}{\rho_k}A_k}(x_k) - x_k\big)_{k=1}^p. \tag{4.13}$$

Then, for $\big(z_k\big)_{k=1}^p \in E$, by $T_2 = I_{\mathcal{H}^p} - \lambda\tilde{B}_G \circ J_{N_\Gamma}$ we obtain

$$\begin{aligned} T_2\big((z_k)_{k=1}^p\big) &= (z_k)_{k=1}^p - \lambda(\tilde{B}_G \circ J_{N_\Gamma})\big((z_k)_{k=1}^p\big) \\ &= (z_k)_{k=1}^p - \lambda\tilde{B}_G\Big(\Big(\sum_{i=1}^p \rho_i z_i\Big)_{k=1}^p\Big) = \Big(z_k - \lambda B\Big(\sum_{i=1}^p \rho_i z_i\Big)\Big)_{k=1}^p. \end{aligned}$$

In addition, for $(y_k)_{k=1}^p \in E$ and setting $\bar{y} = \sum_{i=1}^p \rho_i y_i$ we get in consequence of $T_1 = (1/2)(R_{\bar{A}_G} \circ R_{N_\Gamma} + I_{\mathcal{H}^p})$:

$$\begin{aligned} T_1\big((y_k)_{k=1}^p\big) &= \frac{1}{2}\Big(R_{\bar{A}_G}\left((2\bar{y} - y_k)_{k=1}^p\right) + (y_k)_{k=1}^p\Big) \\ &= \frac{1}{2}\Big((2J_{\frac{\lambda}{\rho_k}A_k}(2\bar{y} - y_k) - 2\bar{y} + y_k)_{k=1}^p + (y_k)_{k=1}^p\Big) \\ &= \Big(J_{\frac{\lambda}{\rho_k}A_k}(2\bar{y} - y_k) - \bar{y} + y_k\Big)_{k=1}^p. \end{aligned}$$

Hence, taking $(y_k)_{k=1}^p := T_2\big((z_k)_{k=1}^p\big)$ and setting $\bar{z} = \sum_{i=1}^p \rho_i z_i$, we have

$$T((z_k)_{k=1}^p) = T_1\left((y_k)_{k=1}^p\right),$$

or equivalently

$$T((z_k)_{k=1}^p) = T_1\Big(\big(z_k - \lambda B(\bar{z})\big)_{k=1}^p\Big) = \Big(J_{\frac{\lambda}{\rho_k}A_k}(2\bar{z} - \lambda B(\bar{z}) - z_k) - \bar{z} + z_k\Big)_{k=1}^p.$$

It follows that

$$\begin{aligned} &(1 - w)(z_k)_{k=1}^p + wT\big((z_k)_{k=1}^p\big) \\ &= \Big(z_k + w\Big(J_{\frac{\lambda}{\rho_k}A_k}\left(2\bar{z} - \lambda B\left(\bar{z}\right) - z_k\right) - \bar{z}\Big)\Big)_{k=1}^p. \end{aligned} \tag{4.14}$$

This leads us to the formulation of G-CRIFBA and the results (4.10a) to (4.10d) follow straightforwardly from Theorem 2.1 (see also Remark 2.2), while (4.10e) is immediately deduced from (4.10d). □

5. An application of (CRIFBA) to some convex-concave saddle-point problem.

Here we apply CRIFBA to the problem below (discussed by Lorenz-Pock [30]).

Let X and Y be two Hilbert spaces endowed with scalar products $\langle .,. \rangle_X$ and $\langle .|. \rangle_Y$, respectively, and induced norms denoted by $\|.\|_X$ and $\|.\|_Y$, and consider the following saddle-point problem

$$\min_{x\in X}\max_{y\in Y} \quad G(x)+Q(x)+\langle Kx,y\rangle_Y - F^*(y) - P^*(y), \tag{5.1}$$

where the function $K : X \to Y$ is linear and bounded, $G : X \to (-\infty,\infty]$ and $F^* : Y \to (-\infty,\infty]$ are convex functions, while the functions $Q : X \to (-\infty,\infty]$ and $P^* : Y \to (-\infty,\infty]$ are convex differentiable with Lipschitz continuous gradient (whose respective Lipschitz constants are l_Q and l_{P^*}).

The above problem covers several primal-dual formulation of nonlinear problems encountered for instance in image processing.

We denote by S the solution set of (5.1) and we assume that $S \neq \emptyset$.

Introduce the Hilbert space $E = X\times Y$ endowed with the scalar product $(.|.)$ defined for $\zeta_1 = (x_1, y_1) \in E$ and $\zeta_2 = (x_2, y_2) \in E$ by

$$(\zeta_1|\zeta_2) = \langle x_1, x_2\rangle_X + \langle y_1, y_2\rangle_Y,$$

and let us denote its induced norm by $\|.\|$. So, (5.1) through its optimality condition can be reformulated as

$$\text{find} \quad (x,y)\in E \quad \text{such that} \quad 0 \in (A+B)\begin{pmatrix} x \\ y \end{pmatrix}, \tag{5.2}$$

where A and B are the monotone operators on E defined by

$$A = \begin{pmatrix} \partial G & K^* \\ -K & \partial F^* \end{pmatrix}, \quad B = \begin{pmatrix} \nabla Q & 0 \\ 0 & \nabla P^* \end{pmatrix}. \tag{5.3}$$

It is established in [30] the following result.

Proposition 5.1. (See [30, Proof of Theorem 5]) *The operator B is co-coercive w.r.t. to the mapping*

$$L = \begin{pmatrix} l_Q I_X & 0 \\ 0 & l_{P^*} I_Y \end{pmatrix}, \tag{5.4}$$

where I_X and I_Y denote the identity mappings on X and Y, respectively.

Proof. Given $\{\zeta_1 = (x_1,y_1), \zeta_2 = (x_2,y_2)\} \subset E$, we have

$$\begin{aligned}
&(B(\zeta_1) - B(\zeta_2)|\zeta_1 - \zeta_2)\\
&= \langle \nabla Q(x_1) - \nabla Q(x_2), x_1 - x_2\rangle_X + \langle \nabla P^*(y_1) - \nabla P^*(y_2)|y_1 - y_2\rangle_Y\\
&\geq l_Q^{-1}\|Q(x_1) - \nabla Q(x_2)\|_X + l_{P^*}^{-1}\|\nabla P^*(y_1) - \nabla P^*(y_2)\|_Y\\
&= \big(L^{-1}(B(\zeta_1) - B(\zeta_2))|B(\zeta_1) - B(\zeta_2)\big) = \|B(\zeta_1) - B(\zeta_2)\|^2_{L^{-1}} \qquad \square
\end{aligned}$$

As a consequence, the above monotone inclusion (5.2) enters the setting of (1.1) and (1.4) and so it can be solved by means of the proposed method CRIFBA.

In general, as explained in [30], evaluating the proximal mapping $(I+\lambda A)^{-1}$ may be prohibitively expensive, which would make our algorithm impracticable in the standard case when $M=I$. Fortunately, this drawback can be overcame when choosing the pre-conditioner mapping M in order to cancel out the upper off-diagonal block in the sum $M+A$, as follows

$$M=\begin{pmatrix} \tau^{-1}I_X & -K^* \\ -K & \sigma^{-1}I_Y \end{pmatrix}, \quad M+A=\begin{pmatrix} \tau^{-1}I_X+\partial G & 0 \\ -2K & \sigma^{-1}I_Y+\partial F^* \end{pmatrix}, \tag{5.5}$$

where τ and σ are positive real numbers. Furthermore, an easy computation gives us the following result.

Proposition 5.2. *For any* $(\xi',\chi')\in X\times Y$, *we obtain*

$$(M+A)^{-1}\begin{pmatrix} \xi' \\ \chi' \end{pmatrix}=\begin{pmatrix} \mathrm{prox}_{\tau G}(\tau\xi') \\ \mathrm{prox}_{\sigma F^*}\left(\sigma\chi'+2\sigma K\mathrm{prox}_{\tau G}(\tau\xi')\right) \end{pmatrix}. \tag{5.6}$$

It is clear that a solution to (5.2) can be approximated by means of a sequence $\{(x_n,y_n)\}\subset X\times Y$ generated by CRIFBA (with $\lambda=1$), in the context of (5.3) and (5.5), as follows

$$(\xi_n,\chi_n)=(x_n,y_n)+\theta_n((x_n,y_n)-(x_{n-1},y_{n-1}))+\gamma_n((\xi_{n-1},\chi_{n-1})-(x_n,y_n)), \tag{5.7a}$$

$$(x_{n+1},y_{n+1})=(1-w)(\xi_n,\chi_n)+wT(\xi_n,\chi_n), \tag{5.7b}$$

where $T=(M+A)^{-1}(M-B)$. This in light of (5.6) leads us to the following corrected relaxed inertial primal dual algorithm:

(CRIPDA):

- $\triangleright$ **Step 1** (initialization):
  Let $\{(\xi_{-1},\chi_{-1}),(x_{-1},y_{-1}),(x_0,y_0)\}\subset X\times Y$, $\{e,s_0,s_1,\nu_0,w,\sigma,\tau\}\subset[0,\infty)$, and set $\nu_n=s_1n+\nu_0$, $\theta_n=1-\frac{e+s_1}{e+\nu_{n+1}}$ and $\gamma_n=1-\frac{s_0}{e+\nu_{n+1}}$.
- $\triangleright$ **Step 2** (main step):
  Given $\{(\xi_{n-1},\chi_{n-1}),(x_{n-1},y_{n-1}),(x_n,y_n)\}\subset X\times Y$ (with $n\geq 0$), we compute

$$\xi_n=x_n+\theta_n(x_n-x_{n-1})+\gamma_n(\xi_{n-1}-x_n), \tag{5.8a}$$

$$\chi_n=y_n+\theta_n(y_n-y_{n-1})+\gamma_n(\chi_{n-1}-y_n), \tag{5.8b}$$

$$x_{n+1}=(1-w)\xi_n+(w)\mathrm{prox}_{\tau G}\big(\xi_n-\tau(\nabla Q(\xi_n)+K^*\chi_n)\big), \tag{5.8c}$$

$$\bar{\xi}_n=2w^{-1}\left(x_{n+1}-(1-w)\xi_n\right), \tag{5.8d}$$

$$y_{n+1}=(1-w)\chi_n+(w)\mathrm{prox}_{\sigma F^*}\big(\chi_n-\sigma(\nabla P^*(\chi_n)-K\bar{\xi}_n)\big). \tag{5.8e}$$

In the absence of any correction term and relaxation factor (that is $\gamma_n=0$ and $w=1$) we retrieve the primal-dual algorithm in [30], which was discussed with step-size rules regarding the momentum term θ_n. In the absence of inertial and correction terms (that is $\theta_n=\gamma_n=0$) we retrieve the primal-dual algorithms, proposed by Condat [17] (for $P^*=0$) and Vu [49], which were also investigated with varying relaxation factors. Compared with these methods, a fast convergence rate is proved for CRIPDA.

The next result establishes the convergence of the above algorithm.

Theorem 5.3. *Let* $\{(x_n, y_n)\} \subset X \times Y$ *be generated by CRIPDA with parameters satisfying*

$$0 < w < 1, \quad 2s_1 < s_0 < e. \tag{5.9}$$

Suppose in addition that one of the two conditions (5.10) *and* (5.11) *holds:*

$$\begin{cases} \delta > \frac{1}{4}\max\{l_Q, l_{P^*}\}, \quad \{\tau, \sigma\} \subset \left(0, \frac{w(1-w)}{\delta}\right), \\ \|K\|^2 < \left(\tau^{-1} - \frac{\delta}{w(1-w)}\right)\left(\sigma^{-1} - \frac{\delta}{w(1-w)}\right), \end{cases} \tag{5.10}$$

$$\begin{cases} 0 < \tau < \frac{w(1-w)}{l_Q}, \quad 0 < \sigma < \frac{w(1-w)}{l_{P^*}}, \\ \|K\|^2 < \left(\tau^{-1} - \frac{l_Q}{w(1-w)}\right)\left(\sigma^{-1} - \frac{l_{P^*}}{w(1-w)}\right). \end{cases} \tag{5.11}$$

Then the following results are reached:

$$\|(\dot{x}_n, \dot{y}_n)\|_M = o(n^{-1}), \quad \textstyle\sum_n n\|(\dot{x}_n, \dot{y}_n)\|_M^2 < \infty, \quad \textit{and}$$
$$\textstyle\sum_n n^2\|(\dot{x}_{n+1} - \dot{x}_n, \dot{y}_{n+1} - \dot{y}_n)\|_M^2 < \infty, \tag{5.12a}$$

$$\|T(x_n, y_n) - (x_n, y_n)\|_M = o(n^{-1}), \quad \textstyle\sum_n n\|T(x_n, y_n) - (x_n, y_n)\|_M^2 < \infty, \tag{5.12b}$$

$$\exists(\bar{x}, \bar{y}) \in S, \quad \textit{s.t.} \quad (x_n, y_n) \rightharpoonup (\bar{x}, \bar{y}) \quad \textit{weakly in } X \times Y. \tag{5.12c}$$

Proof. Clearly, we have $\|L\| = \sup\{l_{P^*}, l_Q\}$, while it can be checked that the operator $M - \frac{\delta}{w(1-w)}I$ is positive definite if the last two conditions in (5.10) are fulfilled. Moreover, it can be verified that the operator $M - \frac{1}{w(1-w)}L$ is positive definite if (5.11) is fulfilled. Therefore, conditions (2.2a) and (2.2b) of Theorem 2.1 are satisfied under conditions (5.10) and (5.11), respectively. The rest of the proof is a direct consequence of Theorem 2.1 □

A. Appendix

A–1. Proof of Proposition 2.5

For reasons of simplification we write G instead of G_λ^M and, given any mapping $\Gamma : \mathcal{H} \to \mathcal{H}$ and any $\{x_1, x_2\} \subset \mathcal{H}$, we denote $\Delta\Gamma(x_1, x_2) = \Gamma(x_1) - \Gamma(x_2)$.

Let $\bar{A} = M^{-1}A$, $\bar{B} = M^{-1}B$ and $C = I - \lambda\bar{B}$. Clearly, $\bar{A}$ is monotone in $(\mathcal{H}, |.|_M)$.

It is also obviously seen for $x \in \mathcal{H}$ that $G(x) = \lambda^{-1}\big(x - J_{\lambda\bar{A}}\big(C(x)\big)\big)$, or equivalently

$$G(x) = \bar{B}(x) + \bar{A}_\lambda(C(x)), \tag{A.1}$$

where $\bar{A}_\lambda := \lambda^{-1}(I - J_{\lambda\bar{A}})$ is the Yosida regularization of $\bar{A}$.

Now, given $(x_1, x_2) \in \mathcal{H}^2$, we get

$$\begin{aligned}
&\langle \Delta G(x_1, x_2), x_1 - x_2\rangle_M = \langle \bar{B}(x_1) + \bar{A}_\lambda(C(x_1)) - \bar{B}(x_2) - \bar{A}_\lambda(C(x_2)), x_1 - x_2\rangle_M \\
&= \langle \Delta\bar{B}(x_1, x_2), x_1 - x_2\rangle_M + \langle \Delta(\bar{A}_\lambda \circ C)(x_1, x_2), x_1 - x_2\rangle_M \\
&= \langle \Delta\bar{B}(x_1, x_2), x_1 - x_2\rangle_M + \langle \Delta(\bar{A}_\lambda \circ C)(x_1, x_2), \Delta C(x_1, x_2)\rangle_M \\
&\qquad + \langle \Delta(\bar{A}_\lambda \circ C)(x_1, x_2), \Delta(I - C)(x_1, x_2)\rangle_M,
\end{aligned}$$

hence, by $I-C=\lambda\bar{B}$, we equivalently obtain

$$\langle \Delta G(x_1,x_2), x_1-x_2\rangle_M \geq \langle \Delta\bar{B}(x_1,x_2), x_1-x_2\rangle_M$$
$$+\langle \Delta(\bar{A}_\lambda\circ C)(x_1,x_2), \Delta C(x_1,x_2)\rangle_M + \lambda\langle \Delta(\bar{A}_\lambda\circ C)(x_1,x_2), \Delta\bar{B}(x_1,x_2)\rangle_M. \quad \text{(A.2)}$$

Let us estimate separately the last two terms in the right side of the previous inequality. As a classical result, by the λ-co-coerciveness of $\bar{A}_\lambda$ in $(\mathcal{H}, \|.\|_M)$, we have

$$\langle \Delta(\bar{A}_\lambda\circ C)(x_1,x_2), \Delta C(x_1,x_2)\rangle_M \geq \lambda\|\Delta(\bar{A}_\lambda\circ C)(x_1,x_2)\|_M^2,$$

which by $\bar{A}_\lambda\circ C = G-\bar{B}$ (from (A.1)) can be rewritten as

$$\langle \Delta(\bar{A}_\lambda\circ C)(x_1,x_2), \Delta C(x_1,x_2)\rangle_M \geq \lambda\|\Delta G(x_1,x_2)-\Delta\bar{B}(x_1,x_2)\|_M^2$$
$$= \lambda\|\Delta G(x_1,x_2)|_M^2 + \lambda\|\Delta\bar{B}(x_1,x_2)\|_M^2 - 2\lambda\langle \Delta G(x_1,x_2), \Delta\bar{B}(x_1,x_2)\rangle_M.$$

Moreover, by $\bar{A}_\lambda\circ C = G-\bar{B}$ (from (A.1)), we simply get

$$\begin{aligned}&\lambda\langle \Delta(\bar{A}_\lambda\circ C)(x_1,x_2), \Delta\bar{B}(x_1,x_2)\rangle_M\\ &= \lambda\langle \Delta G(x_1,x_2)-(\Delta\bar{B}(x_1,x_2)), \Delta\bar{B}(x_1,x_2)\rangle_M\\ &= \lambda\langle \Delta G(x_1,x_2)), \Delta\bar{B}(x_1,x_2)\rangle_M - \lambda\|\Delta\bar{B}(x_1,x_2)\|_M^2.\end{aligned}$$

Thus, by (A.2) and the previous arguments, we obtain

$$\begin{aligned}\langle \Delta G(x_1,x_2), x_1-x_2\rangle_M &\geq \langle \Delta\bar{B}(x_1,x_2), x_1-x_2\rangle_M\\ &\quad + \lambda\|\Delta G(x_1,x_2)\|_M^2 + \lambda\|\Delta\bar{B}(x_1,x_2)\|_M^2 - 2\lambda\langle \Delta G(x_1,x_2), \Delta\bar{B}(x_1,x_2)\rangle_M\\ &\quad + \lambda\langle \Delta G(x_1,x_2), \Delta\bar{B}(x_1,x_2)\rangle_M - \lambda\|\Delta\bar{B}(x_1,x_2)\|_M^2\\ &= \langle \Delta\bar{B}(x_1,x_2), x_1-x_2\rangle_M + \lambda\|\Delta G(x_1,x_2)\|_M^2 - \lambda\langle \Delta G(x_1,x_2), \Delta\bar{B}(x_1,x_2)\rangle_M.\end{aligned}$$

Hence, reminding that $\bar{B}=M^{-1}B$, we equivalently obtain

$$\langle \Delta G(x_1,x_2), x_1-x_2\rangle_M \geq \langle \Delta B(x_1,x_2), x_1-x_2\rangle$$
$$+\lambda\|\Delta G(x_1,x_2)\|_M^2 - \lambda\langle \Delta G(x_1,x_2), \Delta B(x_1,x_2)\rangle.$$

Then by the co-coercivity assumption on B we infer that

$$\langle \Delta G(x_1,x_2), x_1-x_2\rangle_M \geq \|\Delta B(x_1,x_2)\|_{L^{-1}}^2 + \lambda\|\Delta G(x_1,x_2)\|_M^2$$
$$-\lambda\langle \Delta G(x_1,x_2), \Delta B(x_1,x_2)\rangle, \quad \text{(A.3)}$$

that is (2.6).

Now, let us prove (2.7) for $i=1$. From an easy computation, we obtain

$$\|\Delta B(x_1,x_2)\|^2 \leq \|L^{\frac{1}{2}}\|^2\|\Delta B(x_1,x_2)\|_{L^{-1}}^2.$$

hence, for $\delta>0$, using successively Peter-Paul's inequality and the previous inequality, gives us

$$\begin{aligned}\langle \Delta G(x_1,x_2), \Delta B(x_1,x_2)\rangle &\leq \delta\|\Delta G(x_1,x_2)\|^2 + \frac{1}{4\delta}\|\Delta B(x_1,x_2)\|^2,\\ &\leq \delta\|\Delta G(x_1,x_2)\|^2 + \frac{1}{4\delta}\|L^{\frac{1}{2}}\|^2\|\Delta B(x_1,x_2)\|_{L^{-1}}^2. \quad \text{(A.4)}\end{aligned}$$

Therefore, combining this last inequality with (A.3) entails

$$\begin{aligned}
&\langle \Delta G(x_1,x_2), x_1-x_2\rangle_M \\
&\geq \left(1-\frac{\lambda}{4\delta}\|L^{\frac{1}{2}}\|^2\right)\|\Delta B(x_1,x_2)\|^2_{L^{-1}} + \lambda\left(\|\Delta G(x_1,x_2)\|^2_M - \delta\|\Delta G(x_1,x_2)\|^2\right) \\
&= \left(1-\frac{\lambda}{4\delta}\|L^{\frac{1}{2}}\|^2\right)\|\Delta B(x_1,x_2)\|^2_{L^{-1}} + \lambda\langle (M-\delta I)\,\Delta G(x_1,x_2), \Delta G(x_1,x_2)\rangle,
\end{aligned}$$

that is (2.7) with $i=1$.

Let us prove (2.6) for $i=2$. Using again Peter-Paul's inequality, we readily have

$$\begin{aligned}
\lambda\langle \Delta G(x_1,x_2), \Delta B(x_1,x_2)\rangle &= \langle \lambda\Delta G(x_1,x_2), L^{\frac{1}{2}}L^{-\frac{1}{2}}\Delta B(x_1,x_2)\rangle \\
&= \langle \lambda L^{\frac{1}{2}}\Delta G(x_1,x_2), L^{-\frac{1}{2}}\Delta B(x_1,x_2)\rangle \\
&\leq \lambda^2\|L^{\frac{1}{2}}\Delta G(x_1,x_2)\|^2 + \frac{1}{4}\|L^{-\frac{1}{2}}\Delta B(x_1,x_2)\|^2 \\
&= \lambda^2\langle L\Delta G(x_1,x_2), \Delta G(x_1,x_2)\rangle + \frac{1}{4}\|\Delta B(x_1,x_2)\|^2_{L^{-1}},
\end{aligned}$$

which, in light of (A.3), gives us

$\langle \Delta G(x_1,x_2), x_1-x_2\rangle_M \geq \frac{3}{4}\|\Delta B(x_1,x_2)\|^2_{L^{-1}} + \lambda\langle (M-\lambda\delta L)\,\Delta G(x_1,x_2), \Delta G(x_1,x_2)\rangle$,

that is (2.7) with $i=2$. □

A–2. Proof of Proposition 2.8

According to (2.5a), we have

$$v_{n+1} = (\lambda w)G^M_\lambda(z_n) = w\left(z_n - J_{\lambda M^{-1}A}(z_n - \lambda M^{-1}B(z_n))\right),$$

namely $$(I+\lambda M^{-1}A)^{-1}(z_n - \lambda M^{-1}B(z_n)) = z_n - \frac{1}{w}v_{n+1},$$

which is equivalent to the inclusion

$$Mz_n - \lambda B(z_n) \in (M+\lambda A)\left(z_n - \tfrac{1}{w}v_{n+1}\right).$$

This, by $y_{n+1} = z_n - \frac{1}{w}v_{n+1}$ (according to (2.13)) can be reduced to

$$Mz_n - \lambda B(z_n) \in Mz_n - \frac{1}{w}Mv_{n+1} + \lambda A(y_{n+1}),$$

namely $$(\lambda w)^{-1}Mv_{n+1} \in (B(z_n) + A(y_{n+1})),$$

that is (2.14a). Hence, by $y^*_{n+1} = (\lambda w)^{-1}Mv_{n+1} + B(y_{n+1}) - B(z_n)$ (according to (2.13)), we get $y^*_{n+1} \in B(y_{n+1}) + A(y_{n+1})$, that is (2.14b). Next, by definition of y^*_{n+1}, we simply have

$$\|y^*_{n+1}\|_M \leq (\lambda w)^{-1}\|Mv_{n+1}\|_M + \|B(y_{n+1}) - B(z_n)\|_M. \tag{A.5}$$

Let us estimate the two terms in the right side of the previous inequality. Concerning the first term, since M and L are assumed to be bounded, by Remark 2.6 we have

$$\|Mv_{n+1}\|^2_M = \langle (M^{\frac{1}{2}})^6 v_{n+1}, v_{n+1}\rangle = \langle M^2M^{1/2}v_{n+1}, M^{1/2}v_{n+1}\rangle \leq \|M\|^2\|v_{n+1}\|^2_M. \tag{A.6}$$

Concerning the second term, we simply get

$$\|B(y_{n+1}) - B(z_n)\|_M \leq \|M^{\frac{1}{2}}L^{\frac{1}{2}}\| \times \|B(y_{n+1}) - B(z_n)\|_{L^{-1}},$$

while the co-coerciveness of B (given by condition (1.4)) yields

$$\|B(y_{n+1}) - B(z_n)\|_{L^{-1}} \leq \|L^{\frac{1}{2}}(y_{n+1} - z_n)\|,$$

whence it comes that

$$\|B(y_{n+1}) - B(z_n)\|_M \leq \|M^{\frac{1}{2}}L^{\frac{1}{2}}\| \times \|L^{\frac{1}{2}}(y_{n+1} - z_n)\|, \tag{A.7}$$

where $y_{n+1} - z_n = -\frac{1}{w}v_{n+1}$ (from (2.13)). Then, by (A.5), (A.6) and (A.7), we get

$$\|y^*_{n+1}\| \leq (\lambda w)^{-1}\|M\| \times \|v_{n+1}\|_M + \frac{1}{w}\|M^{\frac{1}{2}}L^{\frac{1}{2}}\|. \times \|L^{\frac{1}{2}}v_{n+1}\|. \tag{A.8}$$

On the one hand, if $M - \rho L$ is positive definite, we have

$$\|v_{n+1}\|^2_M - \rho\|L^{\frac{1}{2}}v_{n+1}\|^2 = \langle (M - \rho L)\, v_{n+1}, v_{n+1}\rangle \geq 0,$$

hence $\|L^{\frac{1}{2}}v_{n+1}\| \leq \rho^{-1/2}\|Lv_{n+1}\|_M$. However, if $M - \rho I$ is positive definite, we get

$$\|M^{\frac{1}{2}}v_{n+1}\|^2 - \rho\|v_{n+1}\|^2 = \langle (M - \rho I)\, v_{n+1}, v_{n+1}\rangle \geq 0,$$

which yields $\|v_{n+1}\| \leq \rho^{-1/2}\|v_{n+1}\|_M$, hence

$$\|L^{\frac{1}{2}}v_{n+1}\| \leq \rho^{-1/2}\|L^{\frac{1}{2}}\|.\|v_{n+1}\|_M.$$

Consequently, regarding the previous two situations, by (A.8) we are led to

$$\|y^*_{n+1}\|_M \leq (\lambda w)^{-1}\|M\| \times \|v_{n+1}\|_M + \tfrac{1}{w}\|M^{\frac{1}{2}}L^{\frac{1}{2}}\| \times \rho^{-1/2}(1 + \|L^{\frac{1}{2}}\|) \times \|v_{n+1}\|_M,$$

which amounts to (2.15). □

References

[1] H. Attouch, A. Cabot: *Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions*, Appl. Math. Optim. 80 (2019) 547–598.

[2] H. Attouch, S. C. László: *Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators*, SIAM J. Optim. 30/4 (2020) 3252–3283.

[3] H. Attouch, J. Peypouquet: *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than* $1/k^2$, SIAM J. Optimization 26/3 (2016) 1824–1834.

[4] H. Attouch, J. Peypouquet: *Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators*, Math. Programming 174 (2019) 391–432.

[5] H. Attouch, M. Soueycatt: *Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE's and control*, Pac. J. Optimization 5/1 (2009) 17–37.

[6] H. H. Bauschke, P. L. Combettes: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York (2017).

[7] A. Beck, M. Teboulle: *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. 2/1 (2009) 183-202.

[8] R. I. Boţ, M. Sedlmayer, P. T. Vuong: *A relaxed inertial forward-backward-forward algorithm for solving monotone inclusions with application to GANs*, arXiv: 2003.07886 (2020).

[9] H. Brezis: *Opérateurs Maximaux Monotones*, North Holland Mathematics Studies 5, North-Holland, Amsterdam (1973).

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

[11] V. Cevher, B. C. Vu: *A reflected forward-backward splitting method for monotone inclusions involving Lipschitzian operators*, Set-Valued Var. Anal. 29 (2021) 163–174.

[12] A. Chambolle, C. Dossal: *On the convergence of the iterates of Fista*, J. Optimization Theory Appl. 166/3 (2015) 968–982.

[13] P. L. Combettes: *Monotone operator theory in convex optimization*, Math. Programming B 170/1 (2018) 177–206.

[14] P. L. Combettes, J.-C. Pesquet: *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators*, Set-Valued Var. Analysis 20/2 (2012) 307–330.

[15] P. L. Combettes, B. C. Vu: *Variable metric forward-backward splitting with applications to monotone inclusions in duality*, Optimization 63/9 (2014) 1289–1318.

[16] P. L. Combettes, V. Wajs: *Signal recovery by proximal forwardbackward splitting*, SIAM Multiscale Modelling and Simulation 4/4 (2005) 1168–1200.

[17] L. Condat: *A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms*, J. Optimization Theory Appl. 158/2 (2013) 460–479.

[18] E. Corman, X. Yuan: *A generalized proximal point algorithm and its convergence rate*, SIAM J. Optimization 24/4 (2014) 1614–1638.

[19] I. Daubechies, M. Defrise, C. De Mol: *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math. 57 (2004) 1413–1457.

[20] J. Douglas, H. H. Rachford: *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Society 82 (1956) 421–439.

[21] Y. Drori, M. Teboulle: *Performance of first-order methods for smooth convex minimization: A novel approach*, Math. Programming 145/1-2 (2014) 451–482.

[22] J. Duchi, Y. Singer: *Efficient online and batch learning using forward backward splitting*, J. Machine Learning Res. 10 (2009) 2899–2934.

[23] J. Eckstein, D. P. Bertsekas: *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Programming 55/1-3 (1992) 293–318.

[24] O. Güler: *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optimization 29 (1991) 403–419.

[25] O. Güler: *New proximal point algorithms for convex minimization*, SIAM J. Optimization 2/4 (1992) 649–664.

[26] D. Kim: *Accelerated proximal point method for maximally monotone operators*, Math. Programming 190/1-2 (A) (2021) 57–87.

[27] B. Lemaire: *The proximal algorithm*, in: *New Methods in Optimization and their Industrial Uses*, J. P. Penot (ed.), International Series in Numerical Mathematics 87, Birkhäuser, Basel (1989) 73–87.

[28] H. Lin, J. Mairal, Z. Harchaoui: *Catalyst acceleration for first-order convex optimization: from theory to practice*, J. Machine Learning Res. 18/212 (2018) 1–54.

[29] P. L. Lions, B. Mercier: *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Analysis 16/6 (1979) 964–979.

[30] D. A. Lorenz, T. Pock: *An inertial forward-backward algorithm for monotone inclusions*, J. Math. Imaging Vision 51 (2015) 311–325.

[31] P. E. Maingé: *Accelerated proximal algorithms with a correction term for monotone inclusions*, Appl. Math. Optimization 84/2 (2021) 2027–2061.

[32] B. Martinet: *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Fr. Auto. Infor. Rech. Oper. 4 (1970) 154–158.

[33] A. Moudafi, M. Oliny: *Convergence of a splitting inertial proximal method for monotone operators*, J. Comput. Appl. Math. 155/2 (2003) 447–454.

[34] Y. Nesterov: *A method of solving a convex programming problem with convergence rate O(1/k2)*, Soviet Math. Doklady 27 (1983) 372–376.

[35] Y. Nesterov: *A method for unconstrained convex minimization problem with the rate of convergence O(1/k2)*, Dokl. Akad. Nauk. USSR 269/3 (1983) 543–547.

[36] Y. Nesterov: *Gradient methods for minimizing composite objective function*, Math. Programming B 140 (2013) 125–161.

[37] G. B. Passty: *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, J. Math. Analysis Appl. 72 (1979) 383–390.

[38] D. W. Peaceman, H. H. Rachford: *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Indust. Appl. Math. 3/1 (1955) 28–41.

[39] M. J. D. Powell: *A method for nonlinear constraints in minimization problems*, in: *Optimization*, R. Fletcher (ed.), Academic Press, New York (1969) 283–298.

[40] H. Raguet, J. Fadili, G. Peyré: *A generalized forward–backward splitting*, SIAM J. Imaging Sci. 6 (2013) 1199–1226.

[41] R. T. Rockafellar: *Monotone operators associated with saddle functions and minimax problems*, in: *Nonlinear Functional Analysis, Part 1*, F. E. Browder (ed.), Symposia in Pure Math. 18, American Mathematical Society, Providence (1970) 397–407.

[42] R. T. Rockafellar: *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res. 1/2 (1976) 97–116.

[43] R. T. Rockafellar: *Monotone operators and the proximal point algorithm*, SIAM J. Control. Optimization 14/5 (1976) 877–898.

[44] R. T. Rockafellar, J. B. Wets: *Variational Analysis*, Springer, Berlin (1998).

[45] Z. Sebestyen, Z. Tarcsay: *On the square root of a positive selfadjoint operator*, Periodica Math. Hung. 75 (2017) 268–272.

[46] W. Takahashi: *Nonlinear Functional Analysis*, Yokohama Publ., Yokohama (2000).

[47] P. Tseng: *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optimization 38/2 (2000) 431–446.

[48] S. Villa, S. Salzo, L. Baldassarre, A. Verri: *Accelerated and inexact forward-backward algorithms*, SIAM J. Optimization 23/3 (2013) 1607–1633.

[49] B. Vu: *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math. 38/3 (2013) 667–681.

[50] X. Q. Zhang, M. Burger, X. Bresson, S. Osher: *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imag. Sci. 3/3 (2010) 253–276.