Appl. Set-Valued Anal. Optim. 6 (2024), No. 2, pp. 113-135
Available online at http://asvao.biemdas.com
https://doi.org/10.23952/asvao.6.2024.2.01

# FORWARD-BACKWARD SPLITTING WITH DEVIATIONS FOR MONOTONE INCLUSIONS

HAMED SADEGHI, SEBASTIAN BANERT, PONTUS GISELSSON*

*Department of Automatic Control, Lund University, Lund, Sweden*

**Abstract.** We propose and study a weakly convergent variant of the forward-backward algorithm for solving structured monotone inclusion problems. Our algorithm features a per-iteration deviation vector, providing additional degrees of freedom. The only requirement on the deviation vector to guarantee convergence is that its norm is bounded by a quantity that can be computed online. This approach offers great flexibility and paves the way for the design of new forward-backward-based algorithms, while still retaining global convergence guarantees. These guarantees include linear convergence under a metric subregularity assumption. Choosing suitable monotone operators enables the incorporation of deviations into other algorithms, such as the Chambolle-Pock method and Krasnosel'skiĭ–Mann iterations. We propose a novel inertial primal-dual algorithm by selecting the deviations along a momentum direction and deciding their size by using the norm condition. Numerical experiments validate our convergence claims and demonstrate that even this simple choice of a deviation vector can enhance the performance compared to, for instance, the standard Chambolle–Pock algorithm.
**Keywords.** Forward-backward splitting; Global convergence; Inertial primal-dual algorithm; Linear convergence rate; Monotone inclusions.
**2020 Mathematics Subject Classification.** 47H05, 47N10, 49M29, 65K05, 90C25.

## 1. INTRODUCTION

*Forward–backward* (FB) splitting [6, 24, 26] has been extensively used to solve structured monotone inclusion problems of finding $x \in \mathscr{H}$ such that

$$0 \in Ax + Cx, \tag{1.1}$$

where $A\colon \mathscr{H} \to 2^{\mathscr{H}}$ is a maximally monotone operator, $C\colon \mathscr{H} \to \mathscr{H}$ is a cocoercive operator, and $\mathscr{H}$ is a real Hilbert space. The algorithm sequentially performs a forward step with the operator $C$ followed by a backward step with $A$ to arrive at the iteration

$$x_{n+1} = (\mathrm{Id} + \gamma_n A)^{-1} \circ (\mathrm{Id} - \gamma_n C)x_n, \tag{1.2}$$

where $\gamma_n > 0$ is a step-size parameter. The FB method encompasses various well-established methods as special instances. These include the gradient method, the proximal point algorithm

*Corresponding author.
E-mail address: hamed.sadeghi@control.lth.se (H. Sadeghi), sebastian.banert@control.lth.se (S. Banert), pontus.giselsson@control.lth.se (P. Giselsson).
Received 13 July 2023; Accepted 14 March 2024; Published online 27 March 2024.

[31], the proximal-gradient method [11], the Chambolle–Pock method [7], the Douglas–Rachford method [15, 24], and the Krasnosel'skiĭ–Mann iteration [5, Section 5.2].

In this paper, we present a weakly convergent extension to the standard FB splitting method to solve monotone inclusion (1.1). A simplified instance of our algorithm is given by

$$\begin{aligned} p_n &= (\mathrm{Id}+\gamma_n A)^{-1}\circ(\mathrm{Id}-\gamma_n C)(x_n+u_n) \\ x_{n+1} &= p_n - u_n \end{aligned} \tag{1.3}$$

where $u_n$ is a *deviation vector*. By letting $u_n = 0$, a step of (1.3) reduces to the standard FB step in (1.2). The addition of $u_n$ gives added flexibility that can be utilized to improve performance. In order to ensure convergence of this algorithm, $u_n$ has to satisfy the safeguarding *norm condition*

$$\|u_n\|^2 \le (1-\varepsilon)\tfrac{(2-\gamma_{n-1}\beta)(2-\gamma_n\beta)}{4}\left\|p_{n-1}-x_{n-1}+\tfrac{\gamma_{n-1}\beta}{2-\gamma_{n-1}\beta}u_{n-1}\right\|^2, \tag{1.4}$$

where $\varepsilon\in[0,1)$ is arbitrary. The quantity to the right-hand side of the inequality is computable online since the variables are known from previous iterations. This safeguarding condition plays a different role than the ones in [20, 32, 36, 39] that employ different safeguarding conditions to enable selection between a globally convergent method and a locally fast method while maintaining global convergence. The overarching objective, however, is the same: to enable for enhanced algorithm performance.

Our main algorithm (Algorithm 3.1) uses two deviation vectors and a safeguarding norm condition involving both deviations. A similar algorithm with deviation vectors has been proposed in [4] to extend the proximal gradient method for convex minimization. The fact that we consider the more general monotone inclusion setting, allows us to apply our results, e.g., to the Chambolle–Pock [7] and Condat–Vũ [13, 38] methods—that both are preconditioned FB methods [21]—as well as the Douglas–Rachford method [24] and the Krasnosel'skiĭ–Mann iteration. To facilitate the derivation of some of these special cases, we derive our algorithm with explicit preconditioning, such as in [10, 12, 16, 17, 18, 19, 27, 30].

Our algorithm is also related to inexact FB methods, which are studied in the framework of monotone inclusions [29, 34, 35, 38] and in a convex optimization setting [13, 33, 37]. By including error terms in the FB splitting algorithms, these works allow for inaccuracies in the forward and backward step evaluations. The convergence of the algorithm is usually based on a summability assumption on the error sequences and would therefore allow arbitrarily large errors as long as they only happen for a finite number of iterations. The idea behind our method is in stark contrast to these methods, as our method is designed for actively choosing the deviations with the aim to improve performance.

We instantiate our general scheme in three special settings; the standard FB setting, the primal-dual setting of Condat–Vũ, and the Krasnosel'skiĭ–Mann setting. We also propose a further specialization of the primal-dual setting of Condat–Vũ in which we select the deviations in a heavy-ball type [28] momentum direction. The resulting algorithm bears similarities with the inertial FB methods [1, 2, 3, 9, 25] when applied in a primal-dual setting. Numerical experiments show improved performance of our method compared to the Chambolle–Pock method and a primal–dual version of the Lorenz–Pock method [25].

The organization of the paper is as follows. In Section 2, we provide notations and some definitions. In Section 3, the proposed algorithm is introduced. In Section 4, we prove weak convergence of the method and linear and strong convergence under a metric subregularity assumption. In Section 5, some special cases of the proposed algorithm are presented and Section 6 further specializes one of these to arrive at a novel inertial primal–dual algorithm. We conclude the paper by presenting the numerical results in Section 7.

## 2. Preliminaries

Throughout the paper, the set of real numbers is denoted by $\mathbb{R}$; $\mathcal{H}$ and $\mathcal{K}$ denote real Hilbert spaces that are equipped with inner products and induced norms, which are denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$, respectively. A bounded, self-adjoint, linear operator $M\colon \mathcal{H}\to\mathcal{H}$ is said to be *strongly positive* if there exists some $c>0$ such that $\langle x, Mx\rangle \geq c\|x\|^2 > 0$ for all $x\in\mathcal{H}$. We use the notation $\mathcal{M}(\mathcal{H})$ to denote the set of bounded linear, self-adjoint, strongly positive operators on $\mathcal{H}$. For $M\in\mathcal{M}(\mathcal{H})$ and for all $x,y\in\mathcal{H}$, the $M$-induced inner product and norm are denoted by $\langle x,y\rangle_M = \langle x, My\rangle$ and $\|x\|_M = \sqrt{\langle x, Mx\rangle}$, respectively.

Let $M\in\mathcal{M}(\mathcal{H})$, $x\in\mathcal{H}$, and $S\subset\mathcal{H}$ be a nonempty closed convex set. The $M$-induced projection of $x$ onto the set $S$ is defined as $\Pi_S^M x = \arg\min_{y\in S}\|x-y\|_M$, and the $M$-induced distance from $x$ to $S$ is defined by $\operatorname{dist}_M(x,S) = \inf_{y\in S}\|x-y\|_M$ which satisfies $\operatorname{dist}_M(x,S) = \left\|x-\Pi_S^M x\right\|_M$.

The notation $2^{\mathcal{H}}$ denotes the *power set* of $\mathcal{H}$. A map $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ is characterized by its graph $\operatorname{gra}(A) = \{(x,u)\in\mathcal{H}\times\mathcal{H} : u\in Ax\}$. An operator $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ is *monotone* if $\langle u-v, x-y\rangle \geq 0$ for all $(x,u),(y,v)\in\operatorname{gra}(A)$. A monotone operator $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ is *maximally monotone* if there exists no monotone operator $B\colon\mathcal{H}\to 2^{\mathcal{H}}$ such that $\operatorname{gra}(B)$ properly contains $\operatorname{gra}(A)$.

Let $M\in\mathcal{M}(\mathcal{H})$. An operator $T\colon\mathcal{H}\to\mathcal{H}$ is said to be

(1) *L-Lipschitz continuous* ($L\geq 0$) w.r.t. $\|\cdot\|_M$ if

$$\|Tx-Ty\|_{M^{-1}} \leq L\|x-y\|_M \qquad \text{for all } x,y\in\mathcal{H};$$

(2) $\frac{1}{\beta}$*-cocoercive* ($\beta>0$) w.r.t. $\|\cdot\|_M$ if

$$\langle Tx-Ty, x-y\rangle \geq \frac{1}{\beta}\|Tx-Ty\|_{M^{-1}}^2 \qquad \text{for all } x,y\in\mathcal{H};$$

(3) *nonexpansive* if it is 1-Lipschitz continuous w.r.t. $\|\cdot\|$;
(4) *firmly nonexpansive* if

$$\|Tx-Ty\|^2 + \|(\mathrm{Id}-T)x-(\mathrm{Id}-T)y\|^2 \leq \|x-y\|^2 \qquad \text{for all } x,y\in\mathcal{H}.$$

By the Cauchy–Schwarz inequality, a $\frac{1}{\beta}$-cocoercive operator is $\beta$-Lipschitz continuous. The *resolvent* of a maximally monotone operator $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ is denoted by $J_{\gamma A}\colon\mathcal{H}\to\mathcal{H}$ and defined as $J_{\gamma A} := (\mathrm{Id}+\gamma A)^{-1}$. $J_{\gamma A}$ has full domain, is firmly nonexpansive [5, Corollary 23.8], and is single-valued.

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$, which is convergent to $x^\star$. Then the convergence is (i) *Q-linear* if there exists $q\in(0,1)$ such that $\|x_{n+1}-x^\star\|\leq q\|x_n-x^\star\|$ for all $n$ sufficiently large; (ii) *R-linear* if there exists a sequence $(c_n)_{n\in\mathbb{N}}$ of nonnegative scalars such that $\|x_n-x^\star\|\leq c_n$ and $(c_n)_{n\in\mathbb{N}}$ is Q-linearly convergent to zero.

## 3. Forward-Backward Splitting with Deviations

We consider structured monotone inclusion problems of the form

$$0 \in Ax + Cx, \tag{3.1}$$

that satisfy the following assumptions.

**Assumption 3.1.** Assume that

*(i)* $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone.

*(ii)* $C\colon \mathcal{H} \to \mathcal{H}$ is $\frac{1}{\beta}$-cocoercive with respect to $\|\cdot\|_M$ with $M \in \mathcal{M}(\mathcal{H})$.

*(iii)* The solution set $\operatorname{zer}(A+C) := \{x \in \mathcal{H} : 0 \in Ax + Cx\}$ is nonempty.

Observe that, as a cocoercive operator is maximally monotone [5, Corollary 20.28], and since $C$ has a full domain, the operator $A+C$ is maximally monotone [5, Corollary 25.5].

We present and prove convergence for the following extended variant of FB splitting for solving (3.1).

**Algorithm 3.1**

1: **Input:** initial point $x_0 \in \mathcal{H}$, the sequences $(\zeta_n)_{n\in\mathbb{N}}$, $(\lambda_n)_{n\in\mathbb{N}}$, and $(\gamma_n)_{n\in\mathbb{N}}$ as per Assumption 4.1, and the metric $\|\cdot\|_M$ with $M \in \mathcal{M}(\mathcal{H})$.

2: **set:** $u_0 = v_0 = 0$

3: **for** $n = 0, 1, 2, \ldots$ **do**

4: $y_n = x_n + u_n$

5: $z_n = x_n + \frac{(1-\lambda_n)\gamma_n\beta}{2-\lambda_n\gamma_n\beta} u_n + v_n$

6: $p_n = (M + \gamma_n A)^{-1}(Mz_n - \gamma_n C y_n)$

7: $x_{n+1} = x_n + \lambda_n(p_n - z_n)$

8: choose $u_{n+1}$ and $v_{n+1}$ such that

$$\frac{\lambda_{n+1}\gamma_{n+1}\beta}{2-\lambda_{n+1}\gamma_{n+1}\beta}\|u_{n+1}\|_M^2 + \frac{\lambda_{n+1}(2-\lambda_{n+1}\gamma_{n+1}\beta)}{4-2\lambda_{n+1}-\gamma_{n+1}\beta}\|v_{n+1}\|_M^2 \leq \zeta_n \ell_n^2 \tag{3.2}$$

is satisfied, where

$$\ell_n^2 = \frac{\lambda_n(4-2\lambda_n-\gamma_n\beta)}{2}\left\| p_n - x_n + \frac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta} u_n - \frac{2(1-\lambda_n)}{4-2\lambda_n-\gamma_n\beta} v_n \right\|_M^2 \tag{3.3}$$

9: **end for**

The forward–backward step in Step 6 is unconventional in that it allows the points $y_n$ and $z_n$ to be different. Algorithm 3.1 also allows for *deviations* $u_n$ and $v_n$, which can be seen as design parameters of the algorithm. They can in general be chosen from a subset of $\mathcal{H}$ with non-empty interior (if $\ell_n^2 > 0$ in Step 8), effectively equipping the algorithm with great flexibility. Also note that the upper bound $\ell_n^2$ defined in (3.3) is computable at the time of selecting $u_{n+1}$ and $v_{n+1}$.

Next, we present two special cases of our method. We defer a more detailed discussion on special cases to Section 5.

**Example 3.1.** With the trivial choice of $u_{n+1} = v_{n+1} = 0$, condition (3.2) is already satisfied, and Algorithm 3.1 reduces to the relaxed preconditioned FB iteration

$$\begin{aligned} p_n &= (M + \gamma_n A)^{-1}(Mx_n - \gamma_n C x_n), \\ x_{n+1} &= x_n + \lambda_n(p_n - x_n). \end{aligned}$$

With $M = \mathrm{Id}$ and $\lambda_n = 1$ ($n \in \mathbb{N}$), we recover (1.2).

**Example 3.2.** With $M = \mathrm{Id}$, $\lambda_n = 1$, $v_n = u_n$, and $\zeta_n = 1 - \varepsilon$ ($n \in \mathbb{N}$), we recover the simplified version from (1.3) in the introduction.

**Remark 3.1.** Many works exist that allow for error terms in FB algorithms [13, 29, 38, 37]. Convergence is often based on a summability argument so that any summable sequence of errors is allowed. The strength of our condition (3.2) is that it is iteration-wise; hence, arbitrary large errors would not be accepted. A major difference is that our algorithm does not treat the deviations as errors or inaccuracies in the computation. Instead, they are introduced to allow for actively selecting the deviations with the aim to improve performance.

## 4. Convergence Analysis

In this section, we provide a convergence analysis for Algorithm 3.1 that requires the parameter sequences $(\zeta_n)_{n\in\mathbb{N}}$, $(\lambda_n)_{n\in\mathbb{N}}$, and $(\gamma_n)_{n\in\mathbb{N}}$ to satisfy the following assumption.

**Assumption 4.1.** Choose $\varepsilon \in \left(0, \min\left(1, \frac{4}{3+\beta}\right)\right)$, and assume that, for all $n \in \mathbb{N}$, the following conditions hold:

(i) $0 \le \zeta_n \le 1 - \varepsilon$;
(ii) $\varepsilon \le \gamma_n \le \frac{4-3\varepsilon}{\beta}$;
(iii) $\varepsilon \le \lambda_n \le 2 - \frac{\gamma_n\beta}{2} - \frac{\varepsilon}{2}$.

The sequence $(\zeta_n)_{n\in\mathbb{N}}$ relates the norm of the deviation vectors $u_{n+1}$ and $v_{n+1}$ in (3.2) to its maximum permissible value, $(\lambda_n)_{n\in\mathbb{N}}$ can be seen as a sequence of relaxation parameters for $(x_n)_{n\in\mathbb{N}}$ in Step 7 of Algorithm 3.1, and $(\gamma_n)_{n\in\mathbb{N}}$ is a sequence of step-size parameters for the FB step in Step 6 of Algorithm 3.1.

For the convergence analysis, we first introduce a Lyapunov inequality in Lemma 4.1, which is later used to show weak convergence in Theorem 4.1 and strong and linear convergence under a metric subregularity assumption in Theorem 4.2.

**Lemma 4.1.** *Suppose that Assumption 3.1 and Assumption 4.1 hold. Let $(x_n)_{n\in\mathbb{N}}$, $(u_n)_{n\in\mathbb{N}}$, $(v_n)_{n\in\mathbb{N}}$, $(\ell_n^2)_{n\in\mathbb{N}}$ be sequences generated by Algorithm 3.1 and $x^\star$ be an arbitrary point in $\operatorname{zer}(A+C)$. Then,*

$$\|x_{n+1} - x^\star\|_M^2 + \ell_n^2 \le \|x_n - x^\star\|_M^2 + \frac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\|u_n\|_M^2 + \frac{\lambda_n(2-\lambda_n\gamma_n\beta)}{4-2\lambda_n-\gamma_n\beta}\|v_n\|_M^2 \tag{4.1}$$

*and*

$$\|x_{n+1} - x^\star\|_M^2 + \ell_n^2 \le \|x_n - x^\star\|_M^2 + \zeta_{n-1}\ell_{n-1}^2 \tag{4.2}$$

*hold for all $n \in \mathbb{N}$.*

*Proof.* Let $n \in \mathbb{N}$ be arbitrary. Step 6 in Algorithm 3.1 is equivalent to the inclusion

$$\frac{Mz_n - Mp_n}{\gamma_n} - Cy_n \in Ap_n. \tag{4.3}$$

Since $x^\star \in \operatorname{zer}(A+C)$, we also have

$$-Cx^\star \in Ax^\star. \tag{4.4}$$

Using (4.3), (4.4), and the monotonicity of $A$ gives

$$0 \leq \left\langle \frac{Mz_n - Mp_n}{\gamma_n} - Cy_n + Cx^\star, p_n - x^\star \right\rangle. \tag{4.5}$$

By the $1/\beta$-cocoercivity of $C$ w.r.t. $\|\cdot\|_M$ we have

$$\tfrac{1}{\beta}\|Cy_n - Cx^\star\|_{M^{-1}}^2 \leq \langle Cy_n - Cx^\star, y_n - x^\star\rangle. \tag{4.6}$$

Adding (4.5) and (4.6) yields

$$0 \leq \left\langle \frac{Mz_n - Mp_n}{\gamma_n}, p_n - x^\star \right\rangle + \langle Cy_n - Cx^\star, y_n - p_n\rangle - \tfrac{1}{\beta}\|Cy_n - Cx^\star\|_{M^{-1}}^2.$$

Then, from the step 7 in Algorithm 3.1, we substitute $z_n - p_n = \frac{1}{\lambda_n}(x_n - x_{n+1})$ to obtain

$$\begin{aligned}
0 &\leq \tfrac{1}{\gamma_n\lambda_n}\langle x_n - x_{n+1}, p_n - x^\star\rangle_M + \langle Cy_n - Cx^\star, y_n - p_n\rangle - \tfrac{1}{\beta}\|Cy_n - Cx^\star\|_{M^{-1}}^2 \\
&= \tfrac{1}{2\gamma_n\lambda_n}\Big(\|x_n - x^\star\|_M^2 + \|x_{n+1} - p_n\|_M^2 - \|x_n - p_n\|_M^2 - \|x_{n+1} - x^\star\|_M^2\Big) \\
&\qquad + \langle Cy_n - Cx^\star, y_n - p_n\rangle - \tfrac{1}{\beta}\|Cy_n - Cx^\star\|_{M^{-1}}^2 \\
&\leq \tfrac{1}{2\gamma_n\lambda_n}\Big(\|x_n - x^\star\|_M^2 + \|x_{n+1} - p_n\|_M^2 - \|x_n - p_n\|_M^2 - \|x_{n+1} - x^\star\|_M^2\Big) \\
&\qquad + \tfrac{\beta}{4}\|y_n - p_n\|_M^2
\end{aligned}$$

where we use the identity $2\langle a-b, c-d\rangle_M = \|a-d\|_M^2 + \|b-c\|_M^2 - \|a-c\|_M^2 - \|b-d\|_M^2$ for all $a,b,c,d \in \mathscr{H}$ and Young's inequality $\langle s,x\rangle \leq \frac{\beta}{4}\|x\|_M^2 + \frac{1}{\beta}\|s\|_{M^{-1}}^2$ for all $x,s \in \mathscr{H}$. Multiplying both sides of the last inequality by $2\gamma_n\lambda_n$ and reordering the terms yield

$$\begin{aligned}
&\|x_{n+1} - x^\star\|_M^2 - \|x_n - x^\star\|_M^2 \\
&\leq \|x_{n+1} - p_n\|_M^2 - \|x_n - p_n\|_M^2 + \tfrac{\lambda_n\gamma_n\beta}{2}\|y_n - p_n\|_M^2 \\
&= \|x_n - p_n + \lambda_n(p_n - z_n)\|_M^2 - \|x_n - p_n\|_M^2 + \tfrac{\lambda_n\gamma_n\beta}{2}\|y_n - p_n\|_M^2 \\
&= \lambda_n^2\|p_n - z_n\|_M^2 + 2\lambda_n\langle x_n - p_n, p_n - z_n\rangle_M + \tfrac{\lambda_n\gamma_n\beta}{2}\|y_n - p_n\|_M^2 \\
&= -\lambda_n(2-\lambda_n)\|p_n - z_n\|_M^2 + 2\lambda_n\langle p_n - z_n, x_n - z_n\rangle_M + \tfrac{\lambda_n\gamma_n\beta}{2}\|y_n - p_n\|_M^2,
\end{aligned} \tag{4.7}$$

where we, once again, used the step 7 in Algorithm 3.1 to substitute back $x_{n+1} = x_n + \lambda_n(p_n - z_n)$ into the expression to the right-hand side of the inequality. Now, using the definitions of $y_n$ and $z_n$ in the steps 4 and 5 of Algorithm 3.1, we observe that

$$\begin{aligned}
\ell_n^2 &= \Big(\lambda_n(2-\lambda_n) - \tfrac{\lambda_n\gamma_n\beta}{2}\Big)\Big\|p_n - x_n + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(1-\lambda_n)}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\|_M^2 \\
&= \lambda_n(2-\lambda_n)\|p_n - z_n\|_M^2 + \lambda_n(2-\lambda_n)\Big\|\tfrac{\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n - \tfrac{2-\gamma_n\beta}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\|_M^2 \\
&\quad + 2\lambda_n(2-\lambda_n)\Big\langle p_n - z_n, \tfrac{\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n - \tfrac{2-\gamma_n\beta}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\rangle_M \\
&\quad - \tfrac{\lambda_n\gamma_n\beta}{2}\|p_n - y_n\|_M^2 - \tfrac{\lambda_n\gamma_n\beta}{2}\Big\|\tfrac{2}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(1-\lambda_n)}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\|_M^2 \\
&\quad - \lambda_n\gamma_n\beta\Big\langle p_n - y_n, \tfrac{2}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(1-\lambda_n)}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\rangle_M.
\end{aligned} \tag{4.8}$$

We can estimate the left-hand side of (4.1) by adding (4.7) and (4.8). Let us do this step by step. First, let us look at the two inner products with $p_n - z_n$.

$$\begin{aligned}
&2\lambda_n\Big\langle p_n - z_n, x_n - z_n + (2-\lambda_n)\Big(\tfrac{\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n - \tfrac{2-\gamma_n\beta}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big)\Big\rangle_M \\
&= 2\lambda_n\Big\langle p_n - z_n, \Big(\tfrac{\gamma_n\beta(2-\lambda_n)}{2-\lambda_n\gamma_n\beta} - \tfrac{(1-\lambda_n)\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\Big)u_n - \Big(1+\tfrac{(2-\lambda_n)(2-\gamma_n\beta)}{\gamma_n\beta-2(2-\lambda_n)}\Big)v_n\Big\rangle_M \\
&= 2\lambda_n\Big\langle p_n - z_n, \tfrac{\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{(1-\lambda_n)\gamma_n\beta}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\rangle_M.
\end{aligned}$$

This can be combined with the last term in (4.8), so that we get

$$\begin{aligned}
\|x_{n+1}-x^\star\|_M^2 - \|x_n - x^\star\|_M^2 + \ell_n^2 \le{}& 2\lambda_n\gamma_n\beta\Big\langle y_n - z_n, \tfrac{1}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{(1-\lambda_n)}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\rangle_M \\
&+ \lambda_n(2-\lambda_n)\Big\|\tfrac{\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n - \tfrac{2-\gamma_n\beta}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\|_M^2 \\
&- 2\lambda_n\gamma_n\beta\Big\|\tfrac{1}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{(1-\lambda_n)}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\|_M^2.
\end{aligned} \tag{4.9}$$

With $y_n - z_n = \frac{2-\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n - v_n$, the right-hand side of (4.9) is a quadratic expression in $u_n$ and $v_n$ alone:

$$\begin{aligned}
&\|x_{n+1}-x^\star\|_M^2 - \|x_n - x^\star\|_M^2 + \ell_n^2 \\
&\le 2\lambda_n\gamma_n\beta\Big\langle \tfrac{1-\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n - \tfrac{\gamma_n\beta-3+\lambda_n}{\gamma_n\beta-2(2-\lambda_n)}v_n, \tfrac{1}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{(1-\lambda_n)}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\rangle_M \\
&\quad + \lambda_n(2-\lambda_n)\Big\|\tfrac{\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n - \tfrac{2-\gamma_n\beta}{\gamma_n\beta-2(2-\lambda_n)}v_n\Big\|_M^2.
\end{aligned}$$

In order to verify (4.1), it suffices to check the coefficients of $\|u_n\|_M^2$, $\|v_n\|_M^2$, and $\langle u_n, v_n\rangle_M$ on the right-hand side. This results in

$$\begin{aligned}
&\|x_{n+1}-x^\star\|_M^2 - \|x_n - x^\star\|_M^2 + \ell_n^2 \\
&\le \tfrac{2\lambda_n\gamma_n\beta(1-\gamma_n\beta)+\lambda_n\gamma_n^2\beta^2(2-\lambda_n)}{(2-\lambda_n\gamma_n\beta)^2}\|u_n\|_M^2 \\
&\quad + \tfrac{-2\lambda_n\gamma_n\beta(\gamma_n\beta-3+\lambda_n)(1-\lambda_n)+\lambda_n(2-\lambda_n)(2-\gamma_n\beta)^2}{(\gamma_n\beta-2(2-\lambda_n))^2}\|v_n\|_M^2 \\
&\quad + \tfrac{2\lambda_n\gamma_n\beta(1-\gamma_n\beta)(1-\lambda_n)-2\lambda_n\gamma_n\beta(\gamma_n\beta-3+\lambda_n)-2\lambda_n\gamma_n\beta(2-\lambda_n)(2-\gamma_n\beta)}{(2-\lambda_n\gamma_n\beta)(\gamma_n\beta-2(2-\lambda_n))}\langle u_n, v_n\rangle_M \\
&= \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\|u_n\|_M^2 + \tfrac{\lambda_n(-2+\lambda_n\gamma_n\beta)}{(\gamma_n\beta-2(2-\lambda_n))}\|v_n\|_M^2,
\end{aligned}$$

showing (4.1). Finally, (4.2) follows from inserting (3.2). □

The following theorem is the main convergence result of the paper that guarantees weak convergence for the sequence of iterates obtained from Algorithm 3.1.

**Theorem 4.1.** *Suppose that Assumption 3.1 and Assumption 4.1 hold. Let $(x_n)_{n\in\mathbb{N}}$, $(u_n)_{n\in\mathbb{N}}$, $(v_n)_{n\in\mathbb{N}}$, and $(\ell_n^2)_{n\in\mathbb{N}}$ be generated by Algorithm 3.1. Then, the following conclusions hold:*

- *(i) The sequence $(\ell_n^2)_{n\in\mathbb{N}}$ is summable and the sequences $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ are convergent to zero.*
- *(ii) For all $x^\star \in \operatorname{zer}(A+C)$, the sequence $(\|x_n - x^\star\|_M)_{n\in\mathbb{N}}$ converges.*
- *(iii) The sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(A+C)$.*

*Proof.* We start by proving Item Theorem 4.1 (i) via a telescoping argument for (4.2). To this end, let $N \in \mathbb{N}$. We sum (4.2) for $n = 1, 2, \dots, N$ to obtain

$$\|x_{N+1} - x^\star\|_M^2 + \ell_N^2 + \sum_{n=1}^{N-1} (1-\zeta_n)\ell_n^2 \le \|x_1 - x^\star\|_M^2 + \zeta_0 \ell_0^2.$$

Then, rearranging the terms gives

$$\begin{aligned}\sum_{n=1}^{N} (1-\zeta_n)\ell_n^2 &\le \|x_1 - x^\star\|_M^2 - \|x_{N+1} - x^\star\|_M^2 - \zeta_N \ell_N^2 \\ &\le \|x_1 - x^\star\|_M^2 + \zeta_0 \ell_0^2.\end{aligned}$$

Since the right hand side of the last inequality is independent of $N$, we conclude that

$$\sum_{n=0}^{\infty} (1-\zeta_n)\ell_n^2 < \infty,$$

which, along with $\zeta_n \le 1 - \varepsilon$, implies from Assumption 4.1 that $\ell_n^2 \to 0$ as $n \to \infty$. Then, (3.2) implies that $u_n \to 0$ and $v_n \to 0$ as $n \to \infty$. This proves Theorem 4.1 (i).

The proof of Theorem 4.1 (ii) follows from the property that (4.2) defines a Lyapunov function: since $\zeta_n \le 1$, we obtain from (4.2) that

$$\|x_{n+1} - x^\star\|_M^2 + \ell_n^2 \le \|x_n - x^\star\|_M^2 + \ell_{n-1}^2,$$

i.e., the sequence $\left(\|x_n - x^\star\|_M^2 + \ell_{n-1}^2\right)_{n\in\mathbb{N}}$ is non-increasing. As it is also nonnegative, it is convergent, say $\|x_n - x^\star\|_M^2 + \ell_{n-1}^2 \to \ell_{x^\star} \ge 0$ as $n \to \infty$. Moreover, $\ell_n^2 \to 0$ by Theorem 4.1 (i) as $n \to \infty$, so $\|x_n - x^\star\|_M^2 \to \ell_{x^\star}$, proving Theorem 4.1 (ii). For the proof of Theorem 4.1 (iii), let us define

$$\Delta_n := \frac{Mz_n - Mp_n}{\gamma_n} - (Cy_n - Cp_n).$$

By (4.3), we have $\Delta_n \in Ap_n + Cp_n$, meaning that $(p_n, \Delta_n) \in \operatorname{gra}(A + C)$ for all $n \in \mathbb{N}$. Now, by Assumption 4.1 (iii) we have $\frac{\lambda_n(4-2\lambda_n-\gamma_n\beta)}{2} \ge \varepsilon^2/2$ for all $n \in \mathbb{N}$. By this and $\ell_n \to 0$ as $n \to \infty$, we have that

$$p_n - x_n + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta} u_n + \tfrac{2(\lambda_n-1)}{4-2\lambda_n-\gamma_n\beta} v_n \to 0.$$

Next, since $4 - 2\lambda_n - \gamma_n\beta \ge \varepsilon$ and

$$\lambda_n\gamma_n\beta \le \left(2 - \tfrac{1}{2}\gamma_n\beta - \tfrac{1}{2}\varepsilon\right)\gamma_n\beta = 2 - \tfrac{1}{2}(2-\gamma_n\beta)^2 - \tfrac{1}{2}\varepsilon\gamma_n\beta \le 2 - \tfrac{1}{2}\varepsilon^2\beta,$$

from $u_n \to 0$ and $v_n \to 0$, we conclude that $p_n - x_n \to 0$ as $n \to \infty$. Then,

$$\begin{aligned}\|\Delta_n\|_{M^{-1}} &\le \tfrac{1}{\gamma_n}\|z_n - p_n\|_M + \|Cy_n - Cp_n\|_{M^{-1}} \\ &\le \tfrac{1}{\gamma_n}\left(\|x_n - p_n\|_M + \tfrac{(1-\lambda_n)\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\|u_n\|_M + \|v_n\|_M\right) + \tfrac{1}{\beta}(\|x_n - p_n\|_M + \|u_n\|_M).\end{aligned}$$

Hence, $\Delta_n \to 0$ as $n \to \infty$.

Now, from Theorem 4.1 (ii), we know that $\left(\|x_n - x^\star\|_M^2\right)_{n\in\mathbb{N}}$ is convergent, which implies that $(x_n)_{n\in\mathbb{N}}$ is bounded. Therefore, the latter has at least one weakly convergent subsequence $(x_{k_n})_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x^\star_{\mathrm{wc}} \in \mathcal{H}$ as $n \to \infty$. By the arguments above, we have $p_{k_n} \rightharpoonup x^\star_{\mathrm{wc}}$ and $\Delta_{k_n} \to 0$. Therefore, $(x^\star_{\mathrm{wc}}, 0) \in \operatorname{gra}(A + C)$ by the weak–strong closedness of $\operatorname{gra}(A + C)$ [5, Proposition 20.38]. Then, Theorem 4.1 (iii) follows from [5, Lemma 2.47], and the proof is complete.

□

4.1. **Linear convergence.** Next, we show the linear convergence of Algorithm 3.1 under the following metric subregularity assumption.

**Definition 4.1** (Metric Subregularity)**.** A mapping $T:\mathscr{H}\to 2^{\mathscr{H}}$ is called metrically subregular at $\bar{x}$ for $\bar{y}$ if $(\bar{x},\bar{y})\in \operatorname{gra}(T)$ and there exists a $\kappa\geq 0$ along with neighborhoods $\mathscr{U}$ of $\bar{x}$ and $\mathscr{V}$ of $\bar{y}$ such that

$$\operatorname{dist}_M(x,T^{-1}(\bar{y}))\leq \kappa\operatorname{dist}_{M^{-1}}(\bar{y},T(x)\cap\mathscr{V}) \tag{4.10}$$

for all $x\in\mathscr{U}$ and some $M\in\mathscr{M}(\mathscr{H})$.

The definition above is equivalent to that in [14], but uses the $M$- and $M^{-1}$-induced norm distances instead of the standard canonical norm distance. Using this definition simplifies the notation in the linear convergence analysis. Metric subregularity is an important notion in numerical analysis. For a set-valued operator $T$ and an input vector $\bar{y}$, it simply provides an upper bound of how far a point $x$ is from being a solution to inclusion problem $\bar{y}\in T(x)$. This upper bound is given by (4.10) in terms of the distance of $T(x)$ from the input vector $\bar{y}$. For a detailed discussion on this subject, see [14].

Prior to proceeding with the proof of the linear convergence result for Algorithm 3.1, we provide a brief outline on how this is done. In Lemma 4.2, we show that the $M$-induced squared distance of the iterate $x_n$ to the the solution set $\operatorname{zer}(A+C)$ can be bounded from above by a linear combination of the quantities $\ell_n^2$ and $\ell_{n-1}^2$. Lemma 4.3 shows that the sequence $\|x_{n+1}-x_n\|_M^2$ is also bounded from above by a linear combination of $\ell_n^2$ and $\ell_{n-1}^2$. These lemmas are then combined with Lemma 4.1 to prove local linear convergence results as well as strong convergence of the sequence of iterates to a solution.

**Lemma 4.2.** *Consider the monotone inclusion* (3.1) *along with Assumption 3.1. Let* $(x_n)_{n\in\mathbb{N}}$ *be the sequence of iterates obtained from Algorithm 3.1, and let* $(p_n,\Delta_n)\in\operatorname{gra}(A+C)$ *with*

$$\Delta_n:=\frac{Mz_n-Mp_n}{\gamma_n}-(Cy_n-Cp_n).$$

*Suppose that the following is satisfied for all* $n\in\mathbb{N}$

$$\operatorname{dist}_M^2(p_n,\operatorname{zer}(A+C))\leq\kappa^2\|\Delta_n\|_{M^{-1}}^2. \tag{4.11}$$

*Then, for all* $n\in\mathbb{N}$*, the following holds*

$$\operatorname{dist}_M^2(x_n,\operatorname{zer}(A+C))\leq\tfrac{1}{\psi_n}\ell_n^2+\eta_n\zeta_{n-1}\ell_{n-1}^2, \tag{4.12}$$

*where* $\ell_n^2$ *is given in* (3.3)*, and*

$$\begin{aligned}\psi_n&:=\tfrac{\lambda_n\gamma_n^2(4-2\lambda_n-\gamma_n\beta)}{12\gamma_n^2+24\kappa^2\left(1+\gamma_n^2\beta^2\right)},\\ \eta_n&:=\max\left(\tfrac{(6\lambda_n^2\gamma_n^2+60\kappa^2)\beta}{(2-\lambda_n\gamma_n\beta)\lambda_n\gamma_n},\tfrac{24(\lambda_n-1)^2\gamma_n^2+12\kappa^2\left((2-\gamma_n\beta)^2+4\gamma_n^2\beta^2(\lambda_n-1)^2\right)}{(4-2\lambda_n-\gamma_n\beta)\gamma_n^2(2\lambda_n-\lambda_n^2\gamma_n\beta)}\right).\end{aligned} \tag{4.13}$$

*are positive quantities for all* $n\in\mathbb{N}$.

*Proof.* For all $n \in \mathbb{N}$, we have

$$\begin{aligned}\operatorname{dist}_M^2(x_n, \operatorname{zer}(A+C)) &= \left\|x_n - \Pi^M_{\operatorname{zer}(A+C)} x_n\right\|_M^2 \le \left\|x_n - \Pi^M_{\operatorname{zer}(A+C)} p_n\right\|_M^2 \\ &\le 2\|x_n - p_n\|_M^2 + 2\left\|p_n - \Pi^M_{\operatorname{zer}(A+C)} p_n\right\|_M^2 \\ &= 2\|x_n - p_n\|_M^2 + 2\operatorname{dist}_M^2(p_n, \operatorname{zer}(A+C)) \\ &\le 2\|x_n - p_n\|_M^2 + 2\kappa^2\|\Delta_n\|_{M^{-1}}^2, \end{aligned} \tag{4.14}$$

where the second inequality is implied by Young's inequality, and the last inequality is given by (4.11). For the first term on the right hand side of (4.14), we have

$$\begin{aligned}\|x_n - p_n\|_M^2 &= \left\|p_n - x_n + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n - \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n - \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n\right\|_M^2 \\ &\le 3\left\|p_n - x_n + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n\right\|_M^2 \\ &\quad + \tfrac{3\lambda_n^2\gamma_n^2\beta^2}{(2-\lambda_n\gamma_n\beta)^2}\|u_n\|_M^2 + \tfrac{12(\lambda_n-1)^2}{(2(2-\lambda_n)-\gamma_n\beta)^2}\|v_n\|_M^2. \end{aligned} \tag{4.15}$$

For the second term in the right hand side of (4.14), using the definition of $\Delta_n$, we obtain

$$\begin{aligned}&\|\Delta_n\|_{M^{-1}}^2 \\ &\le \tfrac{2}{\gamma_n^2}\|Mp_n - Mz_n\|_{M^{-1}}^2 + 2\|Cp_n - Cy_n\|_{M^{-1}}^2 \\ &\le \tfrac{2}{\gamma_n^2}\|p_n - z_n\|_M^2 + 2\beta^2\|p_n - y_n\|_M^2 \\ &= \tfrac{2}{\gamma_n^2}\left\|p_n - x_n - \tfrac{(1-\lambda_n)\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n - v_n\right\|_M^2 + 2\beta^2\|p_n - x_n - u_n\|_M^2 \\ &= \tfrac{2}{\gamma_n^2}\Big\|p_n - x_n + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n \\ &\quad - \left(\tfrac{(1-\lambda_n)\gamma_n\beta}{2-\lambda_n\gamma_n\beta} + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\right)u_n - \left(1 + \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}\right)v_n\Big\|_M^2 \\ &\quad + 2\beta^2\left\|p_n - x_n + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n - \left(1 + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\right)u_n - \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n\right\|_M^2 \\ &= \tfrac{2}{\gamma_n^2}\left\|p_n - x_n + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n - \tfrac{\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n - \tfrac{2-\gamma_n\beta}{2(2-\lambda_n)-\gamma_n\beta}v_n\right\|_M^2 \\ &\quad + 2\beta^2\left\|p_n - x_n + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n - \tfrac{2}{2-\lambda_n\gamma_n\beta}u_n - \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n\right\|_M^2 \\ &\le \tfrac{6+6\gamma_n^2\beta^2}{\gamma_n^2}\left\|p_n - x_n + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n\right\|_M^2 \\ &\quad + \tfrac{30\beta^2}{(2-\lambda_n\gamma_n\beta)^2}\|u_n\|_M^2 + \tfrac{6(2-\gamma_n\beta)^2+24\gamma_n^2\beta^2(\lambda_n-1)^2}{(2(2-\lambda_n)-\gamma_n\beta)^2\gamma_n^2}\|v_n\|_M^2, \end{aligned} \tag{4.16}$$

where in the first equality $y_n$ and $z_n$ are replaced by their equivalences from Algorithm 3.1, in the second equality some suitable coefficients of $u_n$ and $v_n$ are added and subtracted, and Young's inequality is used to obtain the last inequality. Then, from (4.15) and (4.16), we have

$$\begin{aligned}&2\|x_n - p_n\|_M^2 + 2\kappa^2\|\Delta_n\|_{M^{-1}}^2 \\ &\le \tfrac{6\gamma_n^2+12\kappa^2\left(1+\gamma_n^2\beta^2\right)}{\gamma_n^2}\left\|p_n - x_n + \tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n + \tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n\right\|_M^2\end{aligned}$$

$$
\begin{aligned}
&+\tfrac{(6\lambda_n^2\gamma_n^2+60\kappa^2)\beta^2}{(2-\lambda_n\gamma_n\beta)^2}\|u_n\|_M^2+\tfrac{24(\lambda_n-1)^2\gamma_n^2+12\kappa^2\left((2-\gamma_n\beta)^2+4\gamma_n^2\beta^2(\lambda_n-1)^2\right)}{(2(2-\lambda_n)-\gamma_n\beta)^2\gamma_n^2}\|v_n\|_M^2\\
=&\tfrac{6\gamma_n^2+12\kappa^2\left(1+\gamma_n^2\beta^2\right)}{\gamma_n^2}\left\|p_n-x_n+\tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n+\tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n\right\|_M^2\\
&+\tfrac{(6\lambda_n^2\gamma_n^2+60\kappa^2)\beta^2(2-\lambda_n\gamma_n\beta)}{(2-\lambda_n\gamma_n\beta)^2\lambda_n\gamma_n\beta}\tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\|u_n\|_M^2\\
&+\tfrac{\left(24(\lambda_n-1)^2\gamma_n^2+12\kappa^2\left((2-\gamma_n\beta)^2+4\gamma_n^2\beta^2(\lambda_n-1)^2\right)\right)(4-2\lambda_n-\gamma_n\beta)}{(2(2-\lambda_n)-\gamma_n\beta)^2\gamma_n^2(2\lambda_n-\lambda_n^2\gamma_n\beta)}\tfrac{2\lambda_n-\lambda_n^2\gamma_n\beta}{4-2\lambda_n-\gamma_n\beta}\|v_n\|_M^2\\
\leq&\tfrac{1}{\psi_n}\ell_n^2+\eta_n\left(\tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\|u_n\|_M^2+\tfrac{2\lambda_n-\lambda_n^2\gamma_n\beta}{4-2\lambda_n-\gamma_n\beta}\|v_n\|_M^2\right)\\
\leq&\tfrac{1}{\psi_n}\ell_n^2+\eta_n\zeta_{n-1}\ell_{n-1}^2,
\end{aligned}
$$

where in the equality above the terms involving $\|u_n\|_M$ and $\|v_n\|_M$ are multiplied and divided by some suitable coefficients, in the second to the last inequality $\eta_n$ and $\psi_n$ from (4.13) and $\ell_n^2$ from (3.3) are substituted, and the last inequality follows from (3.2). Therefore, from the last inequality and (4.14), we see that

$$\operatorname{dist}_M^2(x_n,\operatorname{zer}(A+C))\leq\tfrac{1}{\psi_n}\ell_n^2+\eta_n\zeta_{n-1}\ell_{n-1}^2$$

holds for all $n\in\mathbb{N}$.

Now, using Assumption 4.1, it can be verified that, for all $n\in\mathbb{N}$, $\psi_n$ and $\eta_n$ are positive quantities. For $\psi_n$, using $\lambda_n\geq\varepsilon$, $\gamma_n\geq\varepsilon$, and $4-2\lambda_n-\gamma_n\beta\geq\varepsilon$ (which is attained from Assumption 4.1 (iii)), we have

$$\psi_n=\tfrac{\lambda_n(4-2\lambda_n-\gamma_n\beta)}{12+24\kappa^2\left(1/\gamma_n^2+\beta^2\right)}\geq\tfrac{\varepsilon^2}{12+24\kappa^2(1/\varepsilon^2+\beta^2)}>0,$$

for all $n\in\mathbb{N}$, and for $\eta_n$, using $2-\lambda_n\gamma_n\beta<2$, $\varepsilon\leq\lambda_n<2$, and $\varepsilon\leq\gamma_n<4/\beta$, for the first input argument to $\max(\cdot,\cdot)$ in the definition of $\eta_n$, (4.13), we have

$$\tfrac{(6\lambda_n^2\gamma_n^2+60\kappa^2)\beta}{(2-\lambda_n\gamma_n\beta)\lambda_n\gamma_n}>\tfrac{(6\varepsilon^4+60\kappa^2)\beta}{2\times2\times4/\beta}>0,$$

and thus, it is ensured that $\eta_n$ is positive for all $n\in\mathbb{N}$. Observe that the lower bounds above are not the tight ones, as to attain them, we divided a lower bound of the numerator by an upper bound of the denominator, which results into conservative lower bounds. By this, the proof is complete.

□

**Lemma 4.3.** *Consider Algorithm 3.1 under Assumption 3.1 and suppose that Assumption 4.1 holds. Let $(x_n)_{n\in\mathbb{N}}$ be the sequence generated by Algorithm 3.1. Then, for all $n\in\mathbb{N}$, the following holds*

$$\|x_{n+1}-x_n\|_M^2\leq\vartheta_n\ell_n^2+\theta_n\zeta_{n-1}\ell_{n-1}^2,\tag{4.17}$$

*where $\ell_n^2$ is specified in (3.3), and*

$$
\begin{aligned}
\vartheta_n&:=\tfrac{6\lambda_n}{4-2\lambda_n-\gamma_n\beta_n},\\
\theta_n&:=\max\left(\tfrac{3\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta},\tfrac{3\lambda_n^2(2-\gamma_n\beta)^2}{(4-2\lambda_n-\gamma_n\beta)(2-\lambda_n\gamma_n\beta)\lambda_n}\right).
\end{aligned}\tag{4.18}
$$

*are positive and bounded quantities for all $n\in\mathbb{N}$.*

*Proof.* Starting from the term in the left hand side of (4.17) and substituting $x_{n+1}$ from Algorithm 3.1, we have

$$\begin{aligned}\|x_{n+1}-x_n\|_M^2 &= \lambda_n^2\|p_n-z_n\|_M^2\\ &= \lambda_n^2\left\|p_n-x_n-\tfrac{(1-\lambda_n)\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n-v_n\right\|_M^2\\ &= \lambda_n^2\left\|p_n-x_n+\tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n+\tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n-\tfrac{\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n-\tfrac{2-\gamma_n\beta}{2(2-\lambda_n)-\gamma_n\beta}v_n\right\|_M^2\\ &\le 3\lambda_n^2\left\|p_n-x_n+\tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}u_n+\tfrac{2(\lambda_n-1)}{2(2-\lambda_n)-\gamma_n\beta}v_n\right\|_M^2\\ &\quad+\tfrac{3\lambda_n^2\gamma_n^2\beta^2}{(2-\lambda_n\gamma_n\beta)^2}\|u_n\|_M^2+\tfrac{3\lambda_n^2(2-\gamma_n\beta)^2}{(2(2-\lambda_n)-\gamma_n\beta)^2}\|v_n\|_M^2,\end{aligned}$$

where in the second equality above, $z_n$ is substituted by its update equation in Algorithm 3.1, in the third equality coefficients of $u_n$ and $v_n$ are added and subtracted, and the inequality follows from Young's inequality. Now, by substituting $\ell_n^2$ from (3.3), we have

$$\begin{aligned}\|x_{n+1}-x_n\|_M^2 &\le \tfrac{6\lambda_n}{4-2\lambda_n-\gamma_n\beta_n}\ell_n^2+\tfrac{3\lambda_n^2\gamma_n^2\beta^2}{(2-\lambda_n\gamma_n\beta)^2}\|u_n\|_M^2+\tfrac{3\lambda_n^2(2-\gamma_n\beta)^2}{(2(2-\lambda_n)-\gamma_n\beta)^2}\|v_n\|_M^2\\ &= \tfrac{6\lambda_n}{4-2\lambda_n-\gamma_n\beta_n}\ell_n^2+\tfrac{3\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\|u_n\|_M^2\\ &\quad+\tfrac{3\lambda_n^2(2-\gamma_n\beta)^2}{(4-2\lambda_n-\gamma_n\beta)(2-\lambda_n\gamma_n\beta)\lambda_n}\tfrac{2\lambda_n-\lambda_n^2\gamma_n\beta}{4-2\lambda_n-\gamma_n\beta}\|v_n\|_M^2\\ &\le \vartheta_n\ell_n^2+\theta_n\left(\tfrac{\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta}\|u_n\|_M^2+\tfrac{2\lambda_n-\lambda_n^2\gamma_n\beta}{4-2\lambda_n-\gamma_n\beta}\|v_n\|_M^2\right)\\ &\le \vartheta_n\ell_n^2+\theta_n\zeta_{n-1}\ell_{n-1}^2,\end{aligned}$$

the equality above is attained by multiplying and dividing $\|u_n\|_M^2$ and $\|v_n\|_M^2$ by some appropriate coefficients, the second inequality follows from substitution of $\vartheta_n$ and $\theta_n$ from (4.18), and the last inequality follows from (3.2). For the last part of the proof, using Assumption 4.1, we have the following inequalities,

$$\begin{aligned}\varepsilon &\le \gamma_n \le \tfrac{4-3\varepsilon}{\beta} < \tfrac{4}{\beta},\\ \varepsilon &\le \lambda_n \le 2-\tfrac{\gamma_n\beta}{2}-\varepsilon/2 < 2,\\ \varepsilon &\le 4-2\lambda_n-\gamma_n\beta < 4,\\ \tfrac{1}{2}\beta\varepsilon^2 &\le 2-\lambda_n\gamma_n\beta < 2,\end{aligned}\tag{4.19}$$

where the last one is given by

$$2 > 2-\lambda_n\gamma_n\beta \ge 2-(2-\gamma_n\beta/2-\varepsilon/2)\gamma_n\beta = \tfrac{1}{2}(2-\gamma_n\beta)^2+\tfrac{1}{2}\varepsilon\gamma_n\beta \ge \tfrac{1}{2}\beta\varepsilon^2.$$

Therefore, from (4.18) and by the inequalities above, we have $\vartheta_n \ge \frac{6\varepsilon}{4}$ for all $n\in\mathbb{N}$, and for $\theta_n$, since the first input argument to $\max(\cdot,\cdot)$ in the definition of $\theta_n$ is

$$\tfrac{3\lambda_n\gamma_n\beta}{2-\lambda_n\gamma_n\beta} \ge \tfrac{3\beta\varepsilon^2}{2} > 0,$$

it is ensured that $\theta_n > 0$ for all $n\in\mathbb{N}$. For boundedness of $\vartheta_n$ and $\theta_n$, again, from the inequalities in (4.19), for all $n\in\mathbb{N}$, we have

$$\vartheta_n \le \tfrac{6\times 2}{\varepsilon} \text{ and } \theta_\text{n} \le \max\left(\tfrac{3\times 2}{\varepsilon}, \tfrac{3\times 2^2\times 4}{\varepsilon\times\frac{1}{2}\beta\varepsilon^2\times\varepsilon}\right),$$

where, for the term $(2-\gamma_n\beta)^2$ in the second input argument of $\max(\cdot,\cdot)$ in the definition of $\theta_n$, we used $\beta\varepsilon \le \gamma_n\beta < 4$, to obtain

$$-2 < 2-\gamma_n\beta \le 2-\beta\varepsilon \quad \Rightarrow \quad (2-\gamma_n\beta)^2 < 4.$$

Therefore, both $\vartheta_n$ and $\theta_n$ are bounded quantities for all $n\in\mathbb{N}$. By this, the proof is complete. □

The following result shows linear convergence of the sequence of iterates obtained by Algorithm 3.1, under a metric subregularity assumption. Similar results on linear convergence under a metric subregularity assumption were given in [16, 23].

**Theorem 4.2** (Linear Convergence)**.** *Consider the monotone inclusion problem (3.1) and suppose that Assumption 3.1 holds, that $A+C$ is metrically subregular at all $x^\star \in \operatorname{zer}(A+C)$ for 0, and that either $\mathscr{H}$ is finite-dimensional or that in Definition 4.1 the neighborhood $\mathscr{U}$ at all $x^\star\in\operatorname{zer}(A+C)$ is the whole space $\mathscr{H}$. Let $\delta \in (0, \min(\frac{1-\overline{\zeta}}{1+\overline{\chi}\overline{\zeta}}, \frac{1}{\underline{\psi}}))$ where $\overline{\zeta} = \sup_{n>N}\{\zeta_n\}$, $\overline{\chi} = \sup_{n>N}\{\psi_n\eta_n\}$, and $\underline{\psi} = \inf_{n>N}\{\psi_n\}$ for a sufficiently large $N\in\mathbb{N}$. Then,*

- *(i) the sequence $(\operatorname{dist}^2_M(x_n, \operatorname{zer}(A+C)) + (1-\delta)\ell^2_{n-1})_{n\in\mathbb{N}}$ converges to zero with a local Q-linear rate of convergence.*
- *(ii) the sequence $(x_n)_{n\in\mathbb{N}}$, converges strongly to a point $\bar{x}\in\operatorname{zer}(A+C)$ with an R-linear convergence rate.*

*Proof.* We start by proving (i). From the metric subregularity of $A+C$ at all $x^\star\in\operatorname{zer}(A+C)$ for 0, we obtain

$$\operatorname{dist}_M(x, \operatorname{zer}(A+C)) \le \kappa \operatorname{dist}_{M^{-1}}(0, (A+C)(x)\cap\mathscr{V}_{x^\star}) \le \kappa\|v\|_{M^{-1}} \tag{4.20}$$

for all $x$ in a neighborhood $\mathscr{U}$ of $\operatorname{zer}(A+C)$ and $(x,v)\in\operatorname{gra}(A+C)$ with $\|v\|_{M^{-1}}\le c$ for some $c>0$ and $\kappa>0$.

Now, we recall from the proof of Theorem 4.1 that $(p_n,\Delta_n)\in\operatorname{gra}(A+C)$, where $\Delta_n$ is given by

$$\Delta_n = \frac{Mz_n - Mp_n}{\gamma_n} - (Cy_n - Cp_n).$$

Using Young's inequality along with linearity of $M$ and $\beta$-Lipschitz continuity of $C$ w.r.t. $\|\cdot\|_M$, we obtain

$$\begin{aligned}\|\Delta_n\|^2_{M^{-1}} &= \left\|\tfrac{1}{\gamma_n}(Mz_n - Mp_n) - (Cy_n - Cp_n)\right\|^2_{M^{-1}}\\ &\le \tfrac{2}{\gamma_n^2}\|M(z_n-p_n)\|^2_{M^{-1}} + 2\|Cy_n - Cp_n\|^2_{M^{-1}}\\ &\le \tfrac{2}{\gamma_n^2}\|p_n - z_n\|^2_M + 2\beta^2\|p_n-y_n\|^2_M.\end{aligned} \tag{4.21}$$

From the proof of Theorem 4.1, we have $p_n - z_n\to 0$ and $p_n - y_n \to 0$ as $n\to\infty$. Thus, it follows from (4.21) that $\|\Delta_n\|_{M^{-1}}\to 0$ as $n\to\infty$. Now, if $\mathscr{H}$ is finite-dimensional, since weak convergence implies strong convergence in finite-dimensional Hilbert spaces, it follows from Theorem 4.1 that there exists an $N\in\mathbb{N}$ such that, for all $n>N$, $p_n\in\mathscr{U}$. Otherwise, in infinite-dimensional setting, we have $\mathscr{U}=\mathscr{H}$. Thus, $p_n\in\mathscr{U}$ holds true for all $n\in\mathbb{N}$. Therefore, from $(p_n,\Delta_n)\in\operatorname{gra}(A+C)$ and (4.20), we conclude that

$$\operatorname{dist}^2_M(p_n, \operatorname{zer}(A+C)) \le \kappa^2\|\Delta_n\|^2_{M^{-1}}. \tag{4.22}$$

for all $n > N$. Then, we have from Lemma 4.1 that

$$\|x_{n+1} - x^\star\|_M^2 + \ell_n^2 \le \|x_n - x^\star\|_M^2 + \zeta_{n-1}\ell_{n-1}^2. \tag{4.23}$$

Now, define $x_n^\star$ as the projection of $x_n$ onto the solution set of the problem with respect to $\|\cdot\|_M$, namely, $x_n^\star := \Pi^M_{\operatorname{zer}(A+C)}(x_n)$. Then, from the inequality above, we obtain

$$\begin{aligned}\operatorname{dist}_M^2(x_{n+1}, \operatorname{zer}(A+C)) + \ell_n^2 &\le \|x_{n+1} - x_n^\star\|_M^2 + \ell_n^2 \\ &\le \|x_n - x_n^\star\|_M^2 + \zeta_{n-1}\ell_{n-1}^2 \\ &= \operatorname{dist}_M^2(x_n, \operatorname{zer}(A+C)) + \zeta_{n-1}\ell_{n-1}^2. \end{aligned} \tag{4.24}$$

Next, let $\delta \in (0,1)$ and subtract $\delta\ell_n^2$ from both sides of the inequality above to obtain

$$\operatorname{dist}_M^2(x_{n+1}, \operatorname{zer}(A+C)) + (1-\delta)\ell_n^2 \le \operatorname{dist}_M^2(x_n, \operatorname{zer}(A+C)) + \zeta_{n-1}\ell_{n-1}^2 - \delta\ell_n^2.$$

Now, by (4.22), we have that the requirements of Lemma 4.2 hold for all $n > N$. By using (4.12) in the last inequality for all $n > N$, we have

$$\begin{aligned}&\operatorname{dist}_M^2(x_{n+1}, \operatorname{zer}(A+C)) + (1-\delta)\ell_n^2 \\ &\qquad \le (1-\delta\psi_n)\operatorname{dist}_M^2(x_n, \operatorname{zer}(A+C)) + (1+\delta\psi_n\eta_n)\zeta_{n-1}\ell_{n-1}^2 \\ &\qquad \le (1-\delta\underline{\psi})\operatorname{dist}_M^2(x_n, \operatorname{zer}(A+C)) + \tfrac{(1+\delta\overline{\chi})\overline{\zeta}}{1-\delta}(1-\delta)\ell_{n-1}^2\end{aligned}$$

where $\overline{\zeta} = \sup_{n>N}\{\zeta_n\}$, $\overline{\chi} = \sup_{n>N}\{\psi_n\eta_n\}$, and $\underline{\psi} = \inf_{n>N}\{\psi_n\}$ (we have $\underline{\psi} > 0$ by Lemma 4.2). Thus, from the inequalities above, we obtain

$$\operatorname{dist}_M^2(x_{n+1}, \operatorname{zer}(A+C)) + (1-\delta)\ell_n^2 \le \rho\left(\operatorname{dist}_M^2(x_n, \operatorname{zer}(A+C)) + (1-\delta)\ell_{n-1}^2\right), \tag{4.25}$$

which shows a $Q$-linear rate of convergence with the convergence factor

$$\rho = \max(1-\delta\underline{\psi}, \tfrac{(1+\delta\overline{\chi})\overline{\zeta}}{1-\delta}).$$

To ensure that $\rho \in [0,1)$, we have the following requirement on $\delta$

$$0 < \delta < \min(\tfrac{1-\overline{\zeta}}{1+\overline{\chi}\,\overline{\zeta}}, \tfrac{1}{\underline{\psi}})$$

which is implied by (4.25). It follows from Assumption 4.1 and Lemma 4.2 that $\overline{\chi}\,\overline{\zeta} > 0$. Therefore, the upper bound on $\delta$ is some positive value less than one. Thus, the existence of such a $\delta$ and the local linear convergence factor, $\rho \in [0,1)$, is guaranteed, proving (i). For the proof of (ii), define $V_0 := \rho^{-N}(\operatorname{dist}_M^2(x_N, \operatorname{zer}(A+C)) + (1-\delta)\ell_{N-1}^2)$. For all $n > N$, from (4.25), we obtain

$$\operatorname{dist}_M^2(x_{n+1}, \operatorname{zer}(A+C)) + (1-\delta)\ell_n^2 \le \rho^{n+1}V_0.$$

Then, it follows from the inequality above and Lemma 4.3 that

$$\begin{aligned}\|x_{n+1} - x_n\|_M^2 &\le \vartheta_n\ell_n^2 + \theta_n\zeta_{n-1}\ell_{n-1}^2 \\ &\le \tfrac{\rho\vartheta_n + \theta_n\zeta_{n-1}}{1-\delta}V_0\rho^n \le V_0 \sup_{n>N}\left\{\tfrac{\rho\vartheta_n + \theta_n\zeta_{n-1}}{1-\delta}\right\}\rho^n.\end{aligned}$$

As $(\vartheta_n)_{n\in\mathbb{N}}$ and $(\theta_n)_{n\in\mathbb{N}}$ are bounded by Lemma 4.3 and $(\zeta_n)_{n\in\mathbb{N}}$ is bounded by Assumption 4.1, the supremum above exists and is finite. Therefore, since $\rho < 1$, as $n \to \infty$ the sequence $(\|x_{n+1} - x_n\|_M^2)_{n\in\mathbb{N}}$ and consequently $(\|x_{n+1} - x_n\|_M)_{n\in\mathbb{N}}$, converge $R$-linearly to zero. This entails that there are constants $r \in (0,1)$ and $G > 0$, such that $\|x_{n+1} - x_n\|_M \leq r^n G$ for all $n > N$. For arbitrary indices $m$ and $k$ such that $N < m < k$, we have

$$\begin{aligned} \|x_k - x_m\|_M &\leq \sum_{i=m}^{k-1} \|x_{i+1} - x_i\|_M \leq \sum_{i=m}^{k-1} Gr^i \\ &\leq Gr^m \sum_{i=0}^{k-m-1} r^i = G\tfrac{1-r^{k-m}}{1-r} r^m \end{aligned} \tag{4.26}$$

which leads to the sequence $(\|x_k - x_m\|_M)_{n\in\mathbb{N}}$ converging to zero as $m \to \infty$. Therefore, $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, and thus, it converges strongly to a point $\bar{x} \in \mathscr{H}$. From Theorem 4.1, $(x_n)_{n\in\mathbb{N}}$ is weakly convergent and by uniqueness of the weak limit, we have $\bar{x} = x^\star \in \operatorname{zer}(A+C)$. The linear rate of convergence is obtained by letting $k \to \infty$ in (4.26), which concludes the proof of Theorem 4.2. □

## 5. Special Cases

In this section, we present some special cases of our algorithm.

5.1. **Primal-dual splitting with deviations.** We are concerned with the primal inclusion problem of finding $x \in \mathscr{H}$ such that

$$0 \in Ax + L^* B(Lx) + Cx \tag{5.1}$$

under the following assumption.

**Assumption 5.1.** We assume that

*(i)* $A : \mathscr{H} \to 2^{\mathscr{H}}$ is a maximally monotone operator;
*(ii)* $B : \mathscr{K} \to 2^{\mathscr{K}}$ is a maximally monotone operator;
*(iii)* $L : \mathscr{H} \to \mathscr{K}$ is a bounded linear operator;
*(iv)* $C : \mathscr{H} \to \mathscr{H}$ is a $\frac{1}{\beta}$-cocoercive operator with respect to $\|\cdot\|$;
*(v)* the solution set $\operatorname{zer}(A + L^*BL + C) := \{x \in \mathscr{H} : 0 \in Ax + L^*B(Lx) + Cx\}$ is nonempty.

Problem (5.1) can be translated to a primal–dual problem [21]: $x \in \mathscr{H}$ is a solution to (5.1) if and only if there exists $\mu \in B(Lx)$ (the *dual variable*) such that

$$\begin{aligned} &0 \in Ax + L^*\mu + Cx, \\ &0 \in -Lx + B^{-1}\mu. \end{aligned} \tag{5.2}$$

Define the primal–dual pair $w := (x,\mu) \in \mathscr{H} \times \mathscr{K}$. Then, (5.2) can be restated as

$$0 \in \mathscr{A}w + \mathscr{C}w, \tag{5.3}$$

where (with slight abuse of notation)

$$\mathscr{A} = \begin{bmatrix} A & L^* \\ -L & B^{-1} \end{bmatrix}, \qquad \mathscr{C} = \begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix}. \tag{5.4}$$

The operator $\mathscr{A}$ is maximally monotone by [5, Proposition 26.32] and $\mathscr{C}$ is $1/\beta$-cocoercive with respect to the metric $\|\cdot\|_M$, with

$$M = \begin{bmatrix} I & -\tau L^* \\ -\tau L & \tau\sigma^{-1} I \end{bmatrix}, \tag{5.5}$$

where $\sigma, \tau > 0$ such that $\sigma\tau\|L\|^2 < 1$.

The translation of (5.1) to (5.3) via the two operators $\mathscr{A}$ and $\mathscr{C}$ shows that Algorithm 3.1 using the metric $M$ can be used to solve problem (5.1). We present this special case in Algorithm 5.1, along with the subsequent result on its convergence.

**Algorithm 5.1**

1: **Input:** $(x_0, \mu_0) \in \mathscr{H} \times \mathscr{K}$, the sequences $(\lambda_n)_{n\in\mathbb{N}}$ and $(\zeta_n)_{n\in\mathbb{N}}$ as defined in Assumption 4.1, and $\sigma, \tau > 0$ such that $\sigma\tau\|L\|^2 < 1$.

2: **set:** $u_{x,0} = v_{x,0} = 0$, $v_{\mu,0} = 0$.

3: **for** $n = 0, 1, 2, \ldots$ **do**

4: $\tilde{x}_n = x_n + u_{x,n}$

5: $\begin{bmatrix} \hat{x}_n \\ \hat{\mu}_n \end{bmatrix} = \begin{bmatrix} x_n \\ \mu_n \end{bmatrix} + \begin{bmatrix} \frac{(1-\lambda_n)\tau\beta}{2-\lambda_n\tau\beta} u_{x,n} + v_{x,n} \\ v_{\mu,n} \end{bmatrix}$

6: $\begin{bmatrix} p_{x,n} \\ p_{\mu,n} \end{bmatrix} = \begin{bmatrix} J_{\tau A}\left(\hat{x}_n - \tau L^* \hat{\mu}_n - \tau C\tilde{x}_n\right) \\ J_{\sigma B^{-1}}\left(\hat{\mu}_n + \sigma L(2p_{x,n} - \hat{x}_n)\right) \end{bmatrix}$

7: $\begin{bmatrix} x_{n+1} \\ \mu_{n+1} \end{bmatrix} = \begin{bmatrix} x_n \\ \mu_n \end{bmatrix} + \lambda_n \left( \begin{bmatrix} p_{x,n} \\ p_{\mu,n} \end{bmatrix} - \begin{bmatrix} \hat{x}_n \\ \hat{\mu}_n \end{bmatrix} \right)$

8: choose $u_{n+1} = (u_{x,n+1}, u_{\mu,n+1})$ and $v_{n+1} = (v_{x,n+1}, v_{\mu,n+1})$ such that

$$\begin{aligned} &\frac{\lambda_{n+1}\tau\beta}{2-\lambda_{n+1}\tau\beta}\left\|u_{x,n+1}\right\|^2 + \frac{\lambda_{n+1}(2-\lambda_{n+1}\tau\beta)}{4-2\lambda_{n+1}-\tau\beta}\left\| \begin{bmatrix} v_{x,n+1} \\ v_{\mu,n+1} \end{bmatrix} \right\|_M^2 \\ &\leq \zeta_n \frac{\lambda_n(4-2\lambda_n-\tau\beta)}{2} \left\| \begin{bmatrix} p_{x,n} \\ p_{\mu,n} \end{bmatrix} - \begin{bmatrix} x_n \\ \mu_n \end{bmatrix} + \frac{\lambda_n\tau\beta}{2-\lambda_n\tau\beta} \begin{bmatrix} u_{x,n} \\ 0 \end{bmatrix} - \frac{2(1-\lambda_n)}{4-2\lambda_n-\tau\beta} \begin{bmatrix} v_{x,n} \\ v_{\mu,n} \end{bmatrix} \right\|_M^2 \end{aligned} \tag{5.6}$$

9: **end for**

**Corollary 5.1.** *Consider monotone inclusions (5.3) and suppose that Assumption 5.1 holds. Let $(x_n)_{n\in\mathbb{N}}$ and $(\mu_n)_{n\in\mathbb{N}}$ denote the primal and the dual sequences, respectively, that are obtained from Algorithm 5.1. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in* $\operatorname{zer}(A + L^*BL + C)$*, and $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to a point in the solution set of the dual problem.*

*Proof.* In Algorithm 3.1, replace $A$ by $\mathscr{A}$ and $C$ by $\mathscr{C}$ as devised by (5.4), and substitute $(x_n, \mu_n)$ in place of $x_n$, and also set $p_n = (p_{x,n}, p_{\mu,n})$, $y_n = (\tilde{x}_n, \mu_n)$, $z_n = (\hat{x}_n, \hat{\mu}_n)$, $u_n = (u_{x,n}, 0)$, $v_n = (v_{x,n}, v_{\mu,n})$, $M$ as is in (5.5), and $\gamma_n = \tau$ $(n \in \mathbb{N})$. These changes, along with the update formula

$$\begin{aligned} p_n = (p_{x,n}, p_{\mu,n}) &= (M + \tau\mathscr{A})^{-1}(Mz_n - \tau\mathscr{C}y_n) \\ &= \begin{bmatrix} I + \tau A & 0 \\ -2\tau L & \tau\sigma^{-1}I + \tau B^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \hat{x}_n - \tau L^*\hat{\mu}_n - \tau C\tilde{x}_n \\ -\tau L\hat{x}_n + \tau\sigma^{-1}\hat{\mu}_n \end{bmatrix} \\ &= \begin{bmatrix} (I + \tau A)^{-1}(\hat{x}_n - \tau L^*\hat{\mu}_n - \tau C\tilde{x}_n) \\ (I + \sigma B^{-1})^{-1}(\hat{\mu}_n + \sigma L(2p_{x,n} - \hat{x}_n)) \end{bmatrix} \\ &= \begin{bmatrix} J_{\tau A}\left(\hat{x}_n - \tau L^*\hat{\mu}_n - \tau C\tilde{x}_n\right) \\ J_{\sigma B^{-1}}\left(\hat{\mu}_n + \sigma L(2p_{x,n} - \hat{x}_n)\right) \end{bmatrix}, \end{aligned}$$

result in Algorithm 5.1. Therefore, Algorithm 5.1 is a special instance of Algorithm 3.1; and the corollary is an immediate consequence of Theorem 4.1. □

**Remark 5.1.** In Algorithm 5.1, it might be expected that we get $\tilde{\mu}_n = \mu_n + u_{\mu,n}$, which is the dual counterpart of $\tilde{x}_n = x_n + u_{x,n}$, but we do not. That is because the corresponding part of $\tilde{\mu}_n$ of the operator $\mathscr{C}$ in (5.4), i.e. its second column, is zero, and thus, there is no need to define the dual counterpart of $\tilde{x}_n$.

**Remark 5.2.** In Algorithm 5.1, letting all deviations $u_{x,n}$, $v_{x,n}$, $v_{\mu_n}$ $(n \in \mathbb{N})$ be zero and $\lambda_n = 1$ give

$$\begin{aligned} x_{n+1} &= J_{\tau A}\left(x_n - \tau L^* \mu_n - \tau C x_n\right), \\ \mu_{n+1} &= J_{\sigma B^{-1}}\left(\mu_n + \sigma L(2x_{n+1} - x_n)\right). \end{aligned}$$

This is the Condat–Vũ algorithm in its basic form [13, 38], which, together with $C = 0$, reduces to the basic form of the Chambolle–Pock primal–dual method [7].

**Remark 5.3.** By letting $C = 0$, $\beta = 0$, and $u_{x,n} = 0$ for all $n \in \mathbb{N}$ in Algorithm 5.1, we arrive at a Chambolle–Pock method with deviations and the condition (5.6) reduces to

$$\left\| \begin{bmatrix} v_{x,n+1} \\ v_{\mu,n+1} \end{bmatrix} \right\|_M^2 \leq \zeta_n \tfrac{(2-\lambda_{n+1})(2-\lambda_n)\lambda_n}{\lambda_{n+1}} \left\| \begin{bmatrix} p_{x,n} \\ p_{\mu,n} \end{bmatrix} - \begin{bmatrix} x_n \\ \mu_n \end{bmatrix} - \tfrac{1-\lambda_n}{2-\lambda_n} \begin{bmatrix} v_{x,n} \\ v_{\mu,n} \end{bmatrix} \right\|_M^2.$$

5.2. **Krasnosel'skiĭ–Mann iteration with deviations.** Consider the fixed-point problem

$$x = Tx, \tag{5.7}$$

where $T : \mathscr{H} \to \mathscr{H}$ is a nonexpansive operator. Then, by [5, Remark 4.34, Corollary 23.9], there is a maximally monotone operator $A : \mathscr{H} \to 2^{\mathscr{H}}$ for which $J_{\gamma A} = \frac{1}{2}\mathrm{Id} + \frac{1}{2}T$, with $\gamma > 0$. This correspondence suggests that Algorithm 3.1 can be used to solve (5.7). Letting $C = 0$, $\beta = 0$, $M = \mathrm{Id}$, and $u_n = 0$ for all $n \in \mathbb{N}$ in Algorithm 3.1 results in Algorithm 5.2, which can be used to solve problem (5.7). Weak convergence of Algorithm 5.2 is shown in Corollary 5.2.

**Corollary 5.2.** *Consider the fixed-point problem (5.7) and suppose that its solution set is nonempty and let $J_{\gamma A} = \frac{1}{2}\mathrm{Id} + \frac{1}{2}T$. Then, the sequence $(x_n)_{n\in\mathbb{N}}$, which is generated by Algorithm 5.2, converges weakly to a point in the solution set of the problem.*

**Algorithm 5.2**

**Input:** $x_0 \in \mathscr{H}$, and the sequences $(\lambda_n)_{n\in\mathbb{N}}$, $(\gamma_n)_{n\in\mathbb{N}}$, and $(\zeta_n)_{n\in\mathbb{N}}$ according to Assumption 4.1.
**set:** $v_0 = 0$
**for** $n = 0, 1, \ldots$ **do**
  $z_n = x_n + v_n$
  $p_n = \frac{1}{2}(\mathrm{Id} + T)(x_n + v_n)$
  $x_{n+1} = (1 - \lambda_n)x_n + \lambda_n(p_n - v_n)$
  choose $v_{n+1}$ such that

$$\|v_{n+1}\|^2 \leq \zeta_n \tfrac{\lambda_n(2-\lambda_n)(2-\lambda_{n+1})}{\lambda_{n+1}} \left\| p_n - x_n + \tfrac{\lambda_n - 1}{2-\lambda_n} v_n \right\|^2 \tag{5.8}$$

**end for**

Setting $v_n = 0$ for all $n \in \mathbb{N}$ in Algorithm 5.2 results in

$$x_{n+1} = (1 - \tfrac{\lambda_n}{2})x_n + \tfrac{\lambda_n}{2} T(x_n),$$

which is the standard Krasnosel'skiĭ–Mann iteration [5, Corollary 5.17].

## 6. A Novel Inertial Primal–Dual Ssplitting Algorithm

In this section, we present a novel inertial primal–dual method to solve the problem (5.1) with $C = 0$. We construct this algorithm from Algorithm 5.1 by considering a special structure for the deviation vector. We preset the deviation vector direction at the $n$-th iteration to be aligned with the momentum direction, i.e., $v_n = a_n(x_n - x_{n-1}, \mu_n - \mu_{n-1})$, and use the bound on the norm of deviations to compute $a_n$. Since this algorithm is an instance of Algorithm 5.1, its convergence is guaranteed by Corollary 5.1.

**Algorithm 6.1**

1: **Input:** $(x_0, \mu_0) \in \mathscr{H} \times \mathscr{K}$, and the sequences $(\lambda_n)_{n\in\mathbb{N}}$ and $(\zeta_n)_{n\in\mathbb{N}}$ as stated in Assumption 4.1.
2: **set:** $a_0 = 0$
3: **for** $n = 0, 1, 2, \dots$ **do**
4: $\begin{bmatrix} \hat{x}_n \\ \hat{\mu}_n \end{bmatrix} = \begin{bmatrix} x_n \\ \mu_n \end{bmatrix} + a_n \begin{bmatrix} x_n - x_{n-1} \\ \mu_n - \mu_{n-1} \end{bmatrix}$
5: $\begin{bmatrix} p_{x,n} \\ p_{\mu,n} \end{bmatrix} = \begin{bmatrix} J_{\tau A}(\hat{x}_n - \tau L^* \hat{\mu}_n) \\ J_{\sigma B^{-1}}(\hat{\mu}_n + \sigma L(2p_{x,n} - \hat{x}_n)) \end{bmatrix}$
6: $\begin{bmatrix} x_{n+1} \\ \mu_{n+1} \end{bmatrix} = \begin{bmatrix} x_n \\ \mu_n \end{bmatrix} + \lambda_n \left( \begin{bmatrix} p_{x,n} \\ p_{\mu,n} \end{bmatrix} - \begin{bmatrix} \hat{x}_n \\ \hat{\mu}_n \end{bmatrix} \right)$
7: choose $a_{n+1}$ such that

$$a_{n+1}^2 \left\| \begin{bmatrix} x_{n+1} - x_n \\ \mu_{n+1} - \mu_n \end{bmatrix} \right\|_M^2 \le \zeta_n \tfrac{\lambda_n(2-\lambda_n)(2-\lambda_{n+1})}{\lambda_{n+1}} \left\| \begin{bmatrix} p_{x,n} - x_n \\ p_{\mu,n} - \mu_n \end{bmatrix} + \tfrac{\lambda_n - 1}{2-\lambda_n} a_n \begin{bmatrix} x_n - x_{n-1} \\ \mu_n - \mu_{n-1} \end{bmatrix} \right\|_M^2 \quad (6.1)$$

8: **end for**

**Remark 6.1.** Even though Algorithm 6.1 has similarities with translations of the algorithms of [1, 2, 3, 9, 25] to a primal–dual framework, to the best of our knowledge, the former and the latter cannot be derived from each other, and thus, are essentially different.

6.1. **Efficient evaluation of the norm condition.** In order to compute the bound on the coefficients $a_n$ using (6.1), one needs to compute some $M$-induced norms, which involves evaluating $L$ and $L^*$. Depending on the complexity of evaluating $L$ and $L^*$, these evaluations may be computationally expensive. However, by scrutinizing Algorithm 6.1, it is observed that some of the previous evaluations can be reused to keep the additional computational cost low compared to the standard Chambolle–Pock algorithm. In what follows, we provide more details on how to compute the required scaled norm of the vector quantities in a computationally efficient manner.

As seen in the line 7 of Algorithm 6.1, at each iteration one of each $L$ and $L^*$ evaluations are performed. Similar operations take place at each iteration of, e.g., the Chambolle–Pock algorithm. However, in our algorithm, we have other operations involving evaluations of $L$ and $L^*$. Those are due to verification of the norm condition in line 8 of Algorithm 6.1. More specifically, since the kernel $M$ is given by (5.5) for each evaluation of $\|\cdot\|_M$, we have one more evaluation each of $L$ and $L^*$. This can lead to a substantially higher computational cost. However, except for the first iteration, the extra $L$ and $L^*$ evaluations can be computed from the computations which are already available from previous iterations. That is possible due to the relations

$$\begin{aligned} L\hat{x}_n &= Lx_n + b_n(Lx_n - Lx_{n-1}), \\ L^*\hat{\mu}_n &= L^*\mu_n + b_n(L^*\mu_n - L^*\mu_{n-1}), \\ Lx_{n+1} &= Lx_n + \lambda_n(Lp_{x,n} - L\hat{x}_n), \\ L^*\mu_{n+1} &= L^*\mu_n + \lambda_n(L^*p_{\mu,n} - L^*\hat{\mu}_n), \end{aligned} \tag{6.2}$$

which are derived from the lines 5 and 7 of Algorithm 6.1. In the relations above, for $n > 0$, all quantities to the right hand side are already computed and can be reused, except for $Lp_{x,n}$ and $L^*p_{\mu,n}$ that need to be computed via direct evaluation.

Table 1 provides the list of evaluations involving $L$ and $L^*$ that we need to perform at the first three iterations. It reveals that at the first iteration, we need to perform six different evaluations involving $L$ or $L^*$, of which four might be computationally heavy and two can be done cheaply. After that, i.e. for $n > 0$, we only need to perform two such heavy evaluations per iteration; namely, $Lp_{x,n}$ and $L^*p_{\mu,n}$. The rest of the $L$ and $L^*$ evaluations can be done efficiently by exploiting previously computed quantities and (6.2). This keeps the computational per-iteration cost of our algorithm basically the same as that of the Chambolle–Pock algorithm.

| $n$ | Expensive evaluations | Cheap evaluations |
|---|---|---|
| 0 | $Lx_0$, $L^*\mu_0$, $Lp_{x,0}$, $L^*p_{\mu,0}$ | $Lx_1$, $L^*\mu_1$ |
| 1 | $Lp_{x,1}$, $L^*p_{\mu,1}$ | $L\hat{x}_2$, $Lx_2$, $L^*\hat{\mu}_2$, $L^*\mu_2$ |
| 2 | $Lp_{x,2}$, $L^*p_{\mu,2}$ | $L\hat{x}_3$, $Lx_3$, $L^*\hat{\mu}_3$, $L^*\mu_3$ |

TABLE 1. List of evaluations that involve $L$ and $L^*$ for the first three iterations. The second column shows direct and potentially expensive evaluations and the third column shows evaluations that can be done cheaply via the relations in (6.2).

## 7. Numerical Experiments

We solve an $l_1$-norm regularized SVM problem for classification of the form

$$\underset{x}{\text{minimize}}\ f(Lx) + g(x), \tag{7.1}$$

given a labeled training data set $\{\theta_i, \phi_i\}_{i=1}^N$, where $\theta_i \in \mathbb{R}^d$ and $\phi_i \in \{-1, 1\}$ are training data and labels, respectively, and with

$$f(Lx) = \mathbf{1}^T \max(\mathbf{0}, \mathbf{1} - Lx), \qquad g(x) = \xi\|\omega\|_1, \qquad L = \begin{bmatrix} \phi_1\theta_1^T & \phi_1 \\ \vdots & \vdots \\ \phi_N\theta_N^T & \phi_N \end{bmatrix},$$

where $\mathbf{0} = (0, \ldots, 0)^T$, $\mathbf{1} = (1, \ldots, 1)^T$, $x = (\omega, b)$ is the decision variable with $b \in \mathbb{R}$ and $\omega \in \mathbb{R}^d$, $\max(\cdot,\cdot)$ acts element-wise, and $\xi \geq 0$ is the regularization parameter.

A point $x^\star$ is a solution to (7.1) if and only if it satisfies

$$0 \in L^*\partial f(Lx^\star) + \partial g(x^\star).$$

This holds, since $f$ and $g$ are proper, closed, and convex functions with full domains, and thus, $\partial f$ and $\partial g$ are maximally monotone and $L$ is a linear operator [5, Proposition 16.42]. This monotone inclusion problem is an instance of (5.1) with $A = \partial g$, $B = \partial f$, and $C = 0$. As in Section 5.1, we transform the problem into a primal–dual problem and solve it with primal–dual algorithms.

We compare our inertial primal–dual method, Algorithm 6.1, to the standard Chambolle–Pock (CP) [7], and to the inertial primal–dual algorithm of Lorenz–Pock (LP) [25]. In all experiments, we set the primal and the dual step-sizes to $\tau = \sigma = 0.99/\|L\|$, the regularization parameter of problem (7.1) to $\xi = 0.1$, and $\zeta_n$ is, for each $n \in \mathbb{N}$, sampled from a uniform distribution on $[0, 1-10^{-6}]$. The experiments are done using the *liver disorders* data-set [8] which has 145 samples and 5 features. The solution $(x^\star, \mu^\star)$ is found by running the standard Chambolle–Pock algorithm until the residual gets smaller than $10^{-15}$.

For the $l_1$-norm regularized SVM problem, since $f$ and $g$ are piece-wise linear, the resulting (primal–dual) monotone operator

$$\mathscr{A} = \begin{bmatrix} \partial g & L^* \\ -L & \partial f^* \end{bmatrix}$$

is metrically subregular at any point in the solution set of the problem for 0, see [22, Lemma IV.4]. It therefore follows from Theorem 4.2 that the algorithm exhibits local linear convergence, see Fig. 1 and Fig. 3. The figures reveal that our method needs about half the number of iterations to reach the same accuracy as the other two methods. This improvement comes at essentially no extra computational cost.

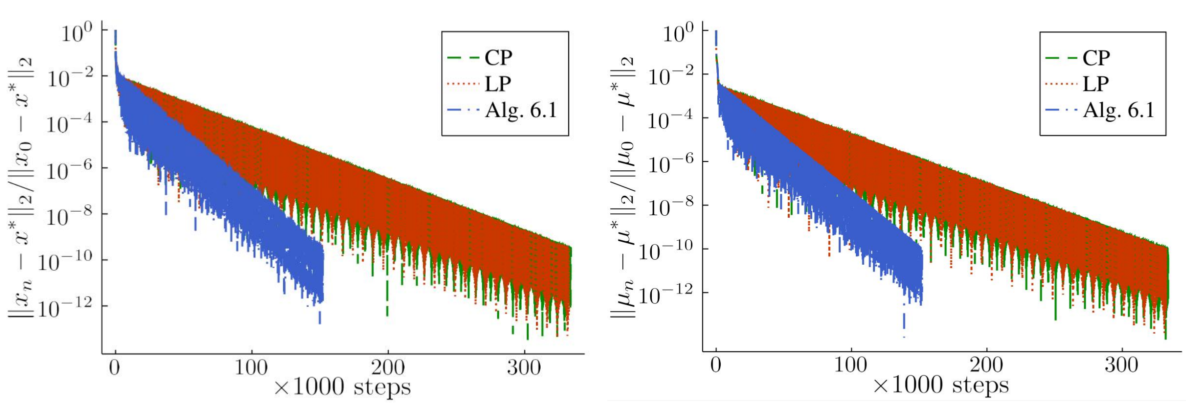
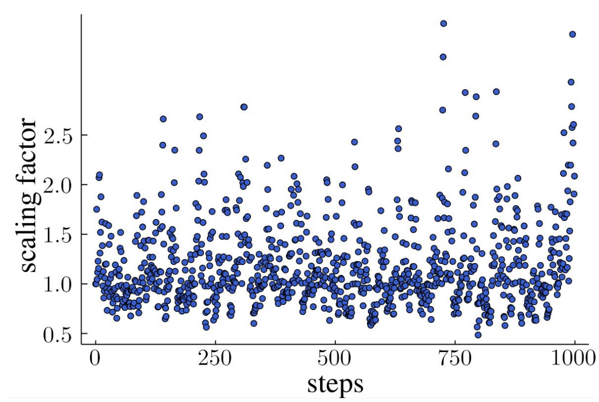



FIGURE 1. Distance to the solution vs. iteration number for the $l_1$-norm regularized SVM (7.1) with $\xi = 0.1$, on the *liver disorders* data-set [8] with 145 samples and 5 features. Solved using Chambolle–Pock primal–dual algorithm (CP), Lorenz–Pock inertial primal–dual method (LP), and Algorithm 6.1 with $\lambda = 1.0$. The primal and dual step-sizes are set to $\tau = \sigma = 0.99/\|L\|$ for all algorithms.

Figure 2 shows the first one thousand scaling factors $a_n$ of Algorithm 6.1 for the same implementation as in Fig. 1. The correction factor attains mostly values close to one.

In Fig. 3, the impact of the relaxation parameter $\lambda$ is investigated. In the sense of convergence rate, it interestingly seems that $\lambda = 1.0$ yields the best performance in this example.




;

FIGURE 2. Scaling factor $a_n$ of Algorithm 6.1 in the experiment shown in Fig. 1 vs. iteration number for the first 1000 iterations.

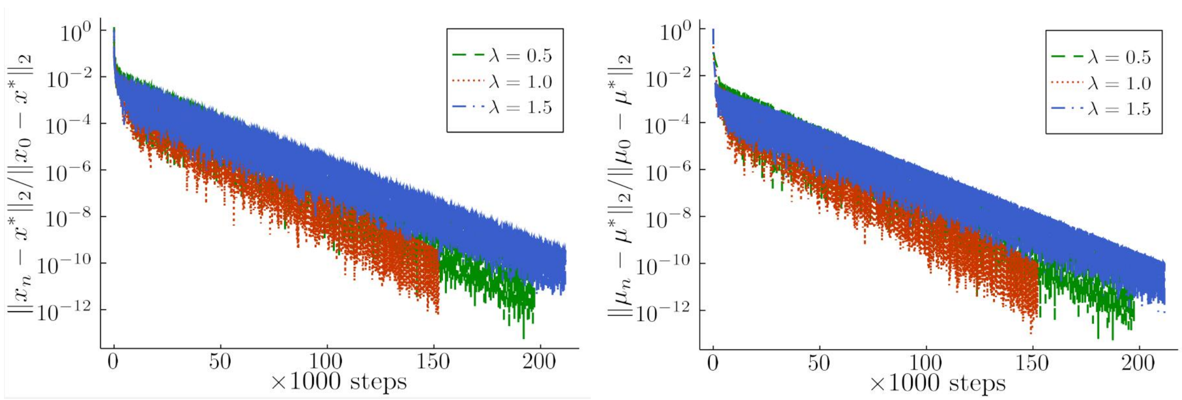



FIGURE 3. Distance to the solution vs. iteration number for the $l_1$-norm regularized SVM (7.1) with $\xi = 0.1$, on the *liver disorders* data-set [8] with 145 samples and 5 features. Solved using Algorithm 6.1 for some values of $\lambda$ with $\tau = \sigma = 0.99/\|L\|$.

## Acknowledgments

The authors would like to thank Bo Bernhardsson for his valuable feedback on this work. This research was partially supported by Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. Sebastian Banert was partially supported by ELLIIT.

## REFERENCES

[1] F. Alvarez, On the minimizing property of a second order dissipative system in Hilbert spaces, SIAM J. Control Optim. 38 (2000), 1102-1119.

[2] F. Alvarez, H. Attouch, An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping, Set-Valed Anal. 9 (2001), 3-11.

[3] H. Attouch, A. Cabot, Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions, Appl. Math. Optim. 8 (2019), 547-598.

[4] S. Banert, J. Rudzusika, O. Oktem, J. Adler, Accelerated forward-backward optimization using deep learning, arXiv:2105.05210, 2021.

[5] H.H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books in Mathematics, Springer, 2017. DOI: 10.1007/978-3-319-48311-5.

[6] R.E. Bruck, An iterative solution of a variational inequality for certain monotone operators in Hilbert space, Bull. Amer. Math. Soc. 81 (1975), 890-892.

[7] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vision 40 (2011), 120-145.

[8] C.C. Chang, C.J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (2011), 1-27.

[9] W. Cholamjiak, P. Cholamjiak, S. Suantai, An inertial forward-backward splitting method for solving inclusion problems in Hilbert spaces, J. Fixed Point Theory Appl. 20 (2018), 42.

[10] E. Chouzenoux, J.C. Pesquet, A. Repetti, Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function, J. Optim. Theory Appl. 162 (2013), 107-132.

[11] P.L. Combettes, J.C. Pesquet, Proximal splitting methods in signal processing, In: Fixed-point Algorithms for Inverse Problems in Science and Engineering, H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, D.R. Luke, H. Wolkowicz (ed.), pp. 185-212, Springer, New York, 2011. DOI: 10.1007/978-1-4419-9569-8_10

[12] P.L. Combettes, B.C. Vũ, Variable metric forward-backward splitting with applications to monotone inclusions in duality, Optimization 63 (2012), 1289-1318.

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

[14] A.L. Dontchev, R.T. Rockafellar, Implicit Functions and Solution Mappings. A View from Variational Analysis, Springer, 2009. DOI: 10.1007/978-0-387-87821-8

[15] J. Eckstein, Splitting methods for monotone operators with applications to parallel optimization, PhD Thesis, Massachusetts Insitute of Technology, 1989. http://hdl.handle.net/1721.1/14356

[16] P. Giselsson, Nonlinear forward-backward splitting with projection correction, SIAM J. Optim. 31 (2021), 2199-2226.

[17] P. Giselsson, S. Boyd, Metric selection in fast dual forward–backward splitting, Automatica, 62 (2015), 1-10.

[18] P. Giselsson, S.P. Boyd, Diagonal scaling in Douglas-Rachford splitting and ADMM, In: 53rd IEEE Conference on Decision and Control, pp. 5033-5039, 2014. DOI: 10.1109/CDC.2014.7040175

[19] P. Giselsson, S.P. Boyd, Preconditioning in fast dual gradient methods, In: 53rd IEEE Conference on Decision and Control, pp. 5040-5045, 2014. DOI: 10.1109/CDC.2014.7040176

[20] P. Giselsson, M. Fält, S. Boyd, Line search for averaged operator iteration, In: 2016 IEEE 55th Conference on Decision and Control (CDC), pp. 1015-1022, 2016. DOI: 10.1109/CDC.2016.7798401

[21] B. He, X. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective, SIAM J. Imaging Sci. 5 (2012), 119-149.

[22] P. Latafat, N. Freris, P. Patrinos, A new randomized block-coordinate primal-dual proximal algorithm for distributed optimization, IEEE Trans. Automat. Control 64 (2019), 4050-4065.

[23] P. Latafat, N. Freris, P. Patrinos, New primal-dual proximal algorithm for distributed optimization, In: 2016 IEEE 55th Conference on Decision and Control (CDC), pp. 1959-1964, 2016. DOI: 10.1109/CDC.2016.7798551

[24] P.L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal. 16 (1979), 964-979.

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

[26] G.B. Passty, Ergodic convergence to a zero of the sum of monotone operators in Hilbert space, J. Math. Anal. Appl. 72 (1979), 383-390.

[27] T. Pock, A. Chambolle, Diagonal preconditioning for first order primal–dual algorithms in convex optimization, In: 2011 International Conference on Computer Vision, pp. 1762-1769, 2011. DOI: 10.1109/ICCV.2011.6126441

[28] B.T. Polyak, Some methods of speeding up the convergence of iteration methods, USSR Comput. Math. Math. Phys. 4 (1964), 1-17.

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

[30] H. Raguet, L. Landrieu, Preconditioning of a generalized forward–backward splitting and application to optimization on graphs, SIAM J. Imaging Sci. 8 (2015), 706-2739.

[31] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim. 14 (1976), 877-898.

[32] H. Sadeghi, P. Giselsson, Hybrid acceleration scheme for variance reduced stochastic optimization algorithms, arXiv:2111.06791, 2021.

[33] M. Schmidt, N. Roux, F. Bach, Convergence rates of inexact proximal-gradient methods for convex optimization, In: J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, K.Q. Weinberger (ed.) Advances in Neural Information Processing Systems (NIPS 2011), vol. 24, pp. 1458-1466, 2011.

[34] M.V. Solodov, B.F. Svaiter, An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, Math. Oper. Res. 25 (2000), 214-230.

[35] M.V. Solodov, B.F. Svaiter, A unified framework for some inexact proximal point algorithms, Numer. Funct. Anal. Optim. 22 (2001), 1013-1035.

[36] A. Themelis, P. Patrinos, SuperMann: A superlinearly convergent algorithm for finding fixed points of nonexpansive operators, IEEE Trans. Automat. Control, 64 (2019), 4875-4890.

[37] S. Villa, S. Salzo, L. Baldassarre, A. Verri, Accelerated and inexact forward-backward algorithms, SIAM J. Optim. 23 (2013), 1607-1633.

[38] B.C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, Adv. Comut. Math. 38 (2013), 667-681.

[39] J. Zhang, B. O'Donoghue, S. Boyd, Globally convergent type-I Anderson acceleration for nonsmooth fixed-point iterations, SIAM J. Optim. 30 (2020), 3170-3197.