UNIVERSIDAD DE CONCEPCIÓN

Centro de Investigación en Ingeniería Matemática (CI²MA)

Relaxed and inertial nonlinear Forward-Backward algorithm

Juan José Maulén, Fernando Roldán,
Cristian Vega

PREPRINT 2026-06

SERIE DE PRE-PUBLICACIONES

RELAXED AND INERTIAL NONLINEAR FORWARD-BACKWARD ALGORITHM

JUAN JOSÉ MAULÉN[1,2], FERNANDO ROLDÁN[3], AND CRISTIAN VEGA[4]

ABSTRACT. The Nonlinear Forward-Backward (NFB) algorithm, also known as *warped resolvent* iterations, is a splitting method for finding zeros of sums of monotone operators. In particular cases, NFB reduces to well-known algorithms such as Forward-Backward, Forward-Backward-Forward, Chambolle–Pock, and Condat–Vũ. Therefore, NFB can be used to solve monotone inclusions involving sums of maximally monotone, cocoercive, monotone and Lipschitz operators as well as linear compositions terms. In this article, we study the weak and strong (linear) convergence of NFB with inertial and relaxation steps. Our results recover known convergence guarantees for the aforementioned methods when extended with inertial and relaxation terms. Additionally, we establish the convergence of inertial and relaxed variants of the Forward-Backward-Half-Forward and Forward-Primal–Dual-Half-Forward algorithms, which, to the best of our knowledge, are new contributions. We consider both nondecreasing and decreasing sequences of inertial parameters, the latter being a novel approach in the context of inertial algorithms. To evaluate the performance of these strategies, we present numerical experiments on optimization problems with affine constraints and on image restoration tasks. Our results show that decreasing inertial sequences can accelerate the numerical convergence of the algorithms.

Keywords. *Operator splitting, monotone operators, monotone inclusion, inertial methods, convex optimization*

2020 *Mathematics Subject Classification.* 47H05, 65K05, 65K15, 90C25

1. INTRODUCTION

In this article, we study an inertial and relaxed version of the Nonlinear Forward-Backward algorithm (NFB) [26, 46] (see also [35]). This method can be applied to numerically solve a broad class of monotone inclusion problems, such as the following:

Problem 1.1. *Let $\boldsymbol{\mathcal{H}}$ be a real Hilbert space, let $\boldsymbol{A} \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}$ be a maximally monotone operator, and let $\boldsymbol{C} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ be a β-cocoercive operator with $\beta \in \,]0, +\infty[$. The problem is to*

$$\text{find } \boldsymbol{x} \in \boldsymbol{\mathcal{H}} \text{ such that } 0 \in \boldsymbol{A}\boldsymbol{x} + \boldsymbol{C}\boldsymbol{x}, \tag{1.1}$$

under the assumption that its solution set is nonempty.

This problem encompasses several applications such as mechanical problems [44, 47, 48], differential inclusions [9, 60], convex programming [34], game theory [20, 24], data science [38], image processing [16, 21], traffic theory [19, 43, 45], among other disciplines.

---

[1] Instituto de Ciencias de la Ingeniería, Universidad de O'Higgins, Rancagua, Chile. *E-mail address:* juan.maulen@postdoc.uoh.cl.

[2] Centro de Modelamiento Matemático (CMM), Universidad de Chile, Chile.

[3] Departamento de Ingeniería Matemática and CI²MA, Universidad de Concepción, Concepción, Chile. *E-mail address:* fernandoroldan@udec.cl.

[4] Instituto de Alta investigación (IAI), Universidad de Tarapacá, Arica, Chile. *E-mail address:* cristianvegacereno6@gmail.com.

In the context of Problem 1.1, given $\boldsymbol{z}_0 \in \boldsymbol{\mathcal{H}}$, NFB iterates as follows:

$$(\forall n \in \mathbb{N}) \quad \left| \begin{array}{l} \boldsymbol{x}_n = (\boldsymbol{M}_n + \boldsymbol{A})^{-1}(\boldsymbol{M}_n - \boldsymbol{C})\boldsymbol{z}_n, \\ \boldsymbol{z}_{n+1} = \boldsymbol{z}_n - \theta_n \mu_n \boldsymbol{S}^{-1}(\boldsymbol{M}_n \boldsymbol{z}_n - \boldsymbol{M}_n \boldsymbol{x}_n), \end{array} \right. \tag{1.2}$$

where $\boldsymbol{S} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ is a linear bounded definite positive operator, for every $n \in \mathbb{N}$, $\boldsymbol{M}_n \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ is a monotone operator, $(\theta_n) \in\ ]0, 2[$, and $(\mu_n)_{n\in\mathbb{N}}$ is an adequate sequence in $]0, +\infty[$ (see (3.55)). For guaranteeing the convergence of NFB to a solution to Problem 1.1, the authors in [46] assume that, for every $n \in \mathbb{N}$, $\boldsymbol{M}_n \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ is Lipschitz continuous and 1-strongly monotone with respect to a linear, bounded and definite positive operator $\boldsymbol{P} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$. In particular instances, the iterations defined by (1.2) reduce to well known optimization methods.

- **Forward-Backward**: If $\boldsymbol{S} = \mathbf{Id}$ and, for every $n \in \mathbb{N}$, $\theta_n \mu_n = \gamma$, $\boldsymbol{M}_n = \mathbf{Id}/\gamma$, for $\gamma \in\ ]0, +\infty[$, NFB reduces to the classic Forward-Backward algorithm (FB) [48, 50, 56]. Moreover, since the Chambolle–Pock algorithm [31] and the Condat–Vũ algorithm [39, 65] are preconditioned versions of FB applied to primal-dual inclusions, NFB also covers these algorithms.

- **Forward-Backward-Forward**: Given a maximally monotone operator $A \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}$ and a Lipschitz operator $B \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$, by defining $\boldsymbol{A} = A + B$, $\boldsymbol{S} = \mathbf{Id}$, $\theta_n \mu_n = \gamma$, for every $n \in \mathbb{N}$, $\boldsymbol{M}_n = \mathbf{Id}/\gamma - B$ and assuming that $\boldsymbol{C} = 0$, the Problem (1.1) reduces to

$$\text{find } \boldsymbol{x} \in \boldsymbol{\mathcal{H}} \text{ such that } 0 \in A\boldsymbol{x} + B\boldsymbol{x}, \tag{1.3}$$

and NFB reduces to the Forward-Backward-Forward method (FBF) [64]. Given $z_0 \in \boldsymbol{\mathcal{H}}$, FBF iterates as follows:

$$(\forall n \in \mathbb{N}) \quad \left| \begin{array}{l} x_n = J_{\tau A}(z_n - \tau D z_n)) \\ z_{n+1} = x_n - \tau(D x_n - D z_n). \end{array} \right. \tag{1.4}$$

- **Forward-Backward-Half-Forward**: considering the same setting as before but with $\boldsymbol{C} \neq 0$, the Forward-Backward-Half-Forward algorithm (FBHF) [22] is recovered. Note that, although the structure of FBHF can be recovered naturally, this is not the case for the condition on the steps guaranteeing convergence for FBHF, which deduction is laborious (see [46, Appendix A.3]). Given $z_0 \in \boldsymbol{\mathcal{H}}$, FBHF iterates as follows:

$$(\forall n \in \mathbb{N}) \quad \left| \begin{array}{l} x_n = J_{\tau A}(z_n - \tau(D z_n + \boldsymbol{C} z_n)) \\ z_{n+1} = x_n - \tau(D x_n - D z_n). \end{array} \right. \tag{1.5}$$

These specific instances highlight the versatility of NFB, as it generalizes several methods from the literature and enables the solution of problems involving Lipschitz operators, primal-dual inclusions, linear compositions, among others. In [54], the authors proposed a nonlinear forward-backward with momentum correction term (NFBM), also used for solving Problem 1.1. Moreover, NFBM also allows recovering and extend existing methods in the literature, such as, forward-half-reflected-backward (FHRB) [52], Chambolle–Pock, Condat–Vũ, Douglas–Rachford [42, 50], among others. However, NFBM cannot recover Tseng-type algorithms such as FBF, FBHF, and forward-primal-dual-half-forward (FPDHF) [58]. Note that NFBM also considers the *warped resolvent* $(\boldsymbol{M}_n + \boldsymbol{A})^{-1}$ on its iterates. The authors in [54], derived the convergence of NFBM under the assumption that, for every $n \in \mathbb{N}$, $\gamma_n \boldsymbol{M}_n - \boldsymbol{S}$ is ζ_n-Lipschitz with respect to $\boldsymbol{S}$, for $\zeta_n \in\ ]0, 1[$ and $\gamma_n \in\ ]0, +\infty[$. In particular, if this condition holds, it is possible to choose $\boldsymbol{P}$ so that $\boldsymbol{M}_n$ is Lipschitz continuous and 1-strongly monotone with respect to $\boldsymbol{P}$ [54, Proposition 2.1], thereby recovering the assumptions on $\boldsymbol{M}_n$ for NFB made in [46].

Several works have incorporated inertial and relaxation steps into splitting algorithms for solving monotone inclusions and it has been shown that inertial and relaxation steps can improve the performance of the algorithm. For instance, inertial proximal algorithms were introduced

in [2] and further developed in works such as [5, 7, 55, 61, 57]. The inertial forward-backward algorithm was first studied in [1] and later has been studied, for example, in [5, 8, 6, 11, 30, 51, 13, 62]. Inertial and relaxed variants of the forward-backward-forward algorithm have been analyzed in [13, 14, 18, 63]. Moreover, inertial versions of the Douglas–Rachford and Chambolle–Pock algorithms have been studied, for example, in [3, 15] and [49, 66], respectively. The Krasnosel'skiĭ–Mann algorithm, which generalize some methods mentioned above, has been extended to incorporate inertial and relaxation steps in [37, 40, 41, 53]. The inertial version of the reflected-forward-backward [28] has been studied in [17]. In the recent paper [59], the authors presented an inertial and relaxed version of NFBM, and consequently, of all the methods generalized by NFBM.

In this article, we analyze the convergence of NFB, including inertial and relaxation steps, under the assumption that $\gamma_n \boldsymbol{M}_n - \boldsymbol{S}$ is ζ_n-Lipschitz with respect to $\boldsymbol{S}$. This assumption simplifies the derivation of convergence conditions on the step-size, relaxation, and inertial parameters, making it easier to recover the corresponding conditions for particular cases of NFB (see, for instance, Remark 4.42 and [46, Appendix A.3]). We also derive the linear convergence of NFB with inertial and relaxation steps in the case where $\boldsymbol{A}$ is strongly monotone. As a result, we recover and extend the relaxed and inertial versions of FB, FBF, FBHF, Chambolle–Pock, Condat–Vũ, and FPDHF. In particular, we consider the case where the sequence of inertial parameters is decreasing, which, to the best of our knowledge, is the first approach that considers this type of inertial strategy, whereas they are usually assumed to be nondecreasing, for instance in [12, 13, 15, 18, 51, 53]. To facilitate implementation and the choice of step-sizes, relaxation parameters, and inertial parameters, we propose an initialization procedure for running these algorithms. We also discuss how the original assumptions made in [46] for including inertial and relaxation steps in NFB can be incorporated into our analysis. Finally, we present numerical experiments testing the benefits of inertial and relaxation steps in NFB, applied to optimization problems with affine constraints and image restoration tasks, both in the monotone and in the strongly monotone case.

The main contributions of this article are summarized as follows:

- We incorporate inertial and relaxation steps into NFB.
- We recover convergence results for inertial and relaxed versions of FB, FBF, Chambolle–Pock, and Condat–Vũ, and extend them to the case of decreasing inertial parameters.
- We propose relaxed and inertial versions of FBHF and FPDHF. In particular, we show that FPDHF correspond to a NFB iteration.
- We provide numerical evidence showing that decreasing inertial parameters can accelerate convergence.

The article is structured as follows. In Section 2, we introduce our notation and preliminaries. Section 3 presents our main convergence results for inertial and relaxed NFB. This section also discusses how the assumptions in [46] relate to our framework and analyzes the conditions on the inertial and relaxation parameters to ensure convergence. In Section 4, we derive relaxed and inertial versions of FB, FBF, FBHF, Chambolle–Pock, Condat–Vũ, and FPDHF as particular instances of NFB. Section 5 presents numerical experiments in optimization with affine constraints and image restoration. Finally, conclusions are given in Section 6.

2. Preliminaries and Notation

Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot \mid \cdot \rangle$, we denote the induced norm by the inner product by $\|\cdot\|$. The symbols $\rightharpoonup$ and $\to$ denote the weak and strong convergence, respectively. Let $S \colon \mathcal{H} \to \mathcal{H}$ be a self adjoint positive definite linear bounded operator, we denote $\langle \cdot \mid \cdot \rangle_S = \langle \cdot \mid S(\cdot) \rangle$ and $\|\cdot\|_S = \sqrt{\langle \cdot \mid \cdot \rangle_S}$. Note that $(\mathcal{H}, \langle \cdot \mid \cdot \rangle_S)$ is a real Hilbert space

and we have

$$(\forall x \in \mathcal{H}) \quad \|S^{-1}x\|_S = \|x\|_{S^{-1}} \tag{2.1}$$

and

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad 2\langle x \mid y \rangle = 2\langle x \mid S^{-1}y \rangle_S = \|Sx - y\|_{S^{-1}}^2 - \|x\|_S^2 - \|y\|_{S^{-1}}^2. \tag{2.2}$$

Moreover, it follows from [10, Corollary 2.15] that

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})(\forall \alpha \in \mathbb{R}) \quad \|\alpha x + (1-\alpha)y\|_S^2 = \alpha\|x\|_S^2 + (1-\alpha)\|y\|_S^2 - \alpha(1-\alpha)\|x-y\|_S^2. \tag{2.3}$$

Let $T\colon \mathcal{H} \to \mathcal{H}$, and $\beta \in]0, +\infty[$. The operator T is β-cocoercive with respect to S if

$$(\forall (x,y) \in \mathcal{H}^2) \quad \langle x - y \mid Tx - Ty \rangle \geq \beta\|Tx - Ty\|_{S^{-1}}^2. \tag{2.4}$$

The operator T is β-Lipschitzian with respect to S if

$$\|Tx - Ty\|_{S^{-1}} \leq \beta\|x - y\|_S. \tag{2.5}$$

Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator, we define the graph of A and the set of zeros of A, respectively, by

$$\mathrm{gra}\, A = \{(x,u) \in \mathcal{H} \times \mathcal{H} \mid u \in Ax\}, \tag{2.6}$$

$$\mathrm{zer} A = \{x \in \mathcal{H} \mid 0 \in Ax\}. \tag{2.7}$$

The inverse of A is $A^{-1}\colon \mathcal{H} \to 2^{\mathcal{H}}\colon u \mapsto \{x \in \mathcal{H} \mid u \in Ax\}$ and the resolvent of A is $J_A = (\mathrm{Id} + A)^{-1}$. Let $s \in \mathbb{R}$, the operator A is s-monotone with respect to S if

$$(\forall (x,u) \in \mathrm{gra}\, A)(\forall (y,v) \in \mathrm{gra}\, A) \quad \langle x - y \mid u - v \rangle \geq s\|u - v\|_S^2. \tag{2.8}$$

If $s = 0$, A is called monotone and, if $s > 0$, A is called s-strongly monotone. Moreover, A is maximally monotone if it is monotone and there exists no monotone operator $B : \mathcal{H} \to 2^{\mathcal{H}}$ such that $\mathrm{gra}\, B$ properly contains $\mathrm{gra}\, A$. If A is maximally monotone, then J_A is single valued and has full domain. We say that A is β-strongly monotone with respect to S, for $\beta > 0$, if

$$(\forall (x,u) \in \mathrm{gra}\, A)(\forall (y,v) \in \mathrm{gra}\, A) \quad \langle x - y \mid u - v \rangle \geq \beta\|x - y\|_S^2. \tag{2.9}$$

We denote by $\Gamma_0(\mathcal{H})$ the class of proper lower semicontinuous convex functions $f\colon \mathcal{H} \to]-\infty, +\infty]$. Let $f \in \Gamma_0(\mathcal{H})$. The Fenchel conjugate of f is given by

$$f^*\colon u \mapsto \sup_{x \in \mathcal{H}}(\langle x \mid u \rangle - f(x)) \tag{2.10}$$

and we have $f^* \in \Gamma_0(\mathcal{H})$. The subdifferential of f is the set-valued operator defined by

$$\partial f\colon x \mapsto \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ f(x) + \langle y - x \mid u \rangle \leq f(y)\}.$$

It follows that ∂f is a maximally monotone operator, that $(\partial f)^{-1} = \partial f^*$, and that $\mathrm{zer}\partial f$ is the set of minimizers of f, which is denoted by $\arg\min_{x \in \mathcal{H}} f$. The proximity operator of f is defined by

$$\mathrm{prox}_f\colon x \mapsto \arg\min_{y \in \mathcal{H}} \left(f(y) + \frac{1}{2}\|x - y\|^2\right). \tag{2.11}$$

Note that $\mathrm{prox}_f = J_{\partial f}$. Given a nonempty closed convex set $C \subset \mathcal{H}$, we denote by P_C the projection onto C and by $\iota_C \in \Gamma_0(\mathcal{H})$ the indicator function of C, which takes the value 0 in C and $+\infty$ otherwise. We denote by $\mathcal{N}_C = \partial\iota_C$ the normal cone to C. For further properties of monotone operators, nonexpansive mappings, and convex analysis, the reader is referred to [10].

We denote by $\ell_+^1(\mathbb{N})$ the set of nonnegative summable sequences. The following lemmas will be used throughout the article.

Lemma 2.1 (Lemma 11 [18]). *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $[0,+\infty[$, let $(b_n)_{n\in\mathbb{N}}$ be a sequence in $\ell^1_+(\mathbb{N})$, and let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $[0,\alpha]$ for $\alpha \in [0,1[$. Suppose that*

$$(\forall n \geq 1) \quad a_{n+1} \leq a_n + \alpha_n(a_n - a_{n-1}) + b_n.$$

Then $(a_n)_{n\in\mathbb{N}}$ converges.

Lemma 2.2 (Lemma 5.31 [10]). *Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences in $[0,+\infty[$, and let $(t_n)_{n\in\mathbb{N}}$ and $(r_n)_{n\in\mathbb{N}}$ be sequences in $\ell^1_+(\mathbb{N})$ such that*

$$(\forall n \in \mathbb{N}) \quad a_{n+1} \leq (1+t_n)a_n - b_n + r_n.$$

Then $(a_n)_{n\in\mathbb{N}}$ converges and $\sum_{n\in\mathbb{N}} b_n < +\infty$.

Lemma 2.3 (Lemma 3.80 [32]). *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $[0,+\infty[$, let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $[0,+\infty[$ such that $\alpha_n \to 0$, and let $r \in]0,1[$. Suppose that*

$$(\forall n \in \mathbb{N}) \quad a_{n+1} \leq (1 - r + \alpha_n)a_n.$$

Then, $(a_n)_{n\in\mathbb{N}}$ converges linearly to 0.

3. Inertial Nonlinear Forward-Backward

In this section we consider the following problem.

Problem 3.1. *Let $\mathcal{H}$ be a real Hilbert space, let $\boldsymbol{S}\colon \mathcal{H} \to \mathcal{H}$ be a self adjoint strongly monotone linear bounded operator, let $\boldsymbol{A}\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $\boldsymbol{C}\colon \mathcal{H} \to \mathcal{H}$ be a β-cocoercive operator with respect to $\boldsymbol{S}$ for $\beta \in]0,+\infty[$. The problem is to*

$$\text{find } \boldsymbol{x} \in \mathcal{H} \text{ such that } 0 \in \boldsymbol{A}\boldsymbol{x} + \boldsymbol{C}\boldsymbol{x},$$

under the assumption that its solution set is nonempty.

The following assumption plays a central role in the convergence analysis of the inertial-relaxed nonlinear forward-backward algorithm studied in this article.

Assumption 3.2. *In the context of Problem 3.1, let $\gamma \in]0,+\infty[$ and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\gamma,+\infty[$. Moreover, let $\boldsymbol{M}_n\colon \mathcal{H} \to \mathcal{H}$ be a singled valued operator such that $\gamma_n\boldsymbol{M}_n - \boldsymbol{S}$ is a ζ_n-Lipschitz operator with respect to $\boldsymbol{S}$ for $\zeta_n \in [\underline{\zeta},\zeta]$ where $\underline{\zeta} \in [0,\zeta]$ and $\zeta \in]0,1[$.*

The following algorithm incorporates inertial and relaxation steps into NFB.

Algorithm 3.3. *In the context of Problem 3.1 and Assumption 3.2, let $(\boldsymbol{z}_0, \boldsymbol{z}_{-1}) \in \mathcal{H}^2$, let $(\alpha_n)_{n\in\mathbb{N}}$ be a nonnegative sequence such that $\alpha_n \to \alpha \in [0,1[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\underline{\lambda}, \overline{\lambda}[$ for $(\underline{\lambda}, \overline{\lambda}) \in]0,2[^2$, and consider the recurrence defined by*

$$\boldsymbol{y}_n = \boldsymbol{z}_n + \alpha_n(\boldsymbol{z}_n - \boldsymbol{z}_{n-1}), \tag{3.1a}$$

$$\boldsymbol{x}_n = (\boldsymbol{M}_n + \boldsymbol{A})^{-1}(\boldsymbol{M}_n - \boldsymbol{C})\boldsymbol{y}_n, \tag{3.1b}$$

$$\boldsymbol{w}_{n+1} = \boldsymbol{y}_n - \gamma_n\boldsymbol{S}^{-1}(\boldsymbol{M}_n\boldsymbol{y}_n - \boldsymbol{M}_n\boldsymbol{x}_n), \tag{3.1c}$$

$$\boldsymbol{z}_{n+1} = \lambda_n\boldsymbol{w}_{n+1} + (1 - \lambda_n)\boldsymbol{y}_n. \tag{3.1d}$$

Remark 3.4. *Note that, Algorithm 3.3 is well defined in view of Assumption 3.2 [54, Proposition 3.1].*

The following proposition provides essential estimates for establishing the convergence of Algorithm 3.3.

Proposition 3.5. *In the context of Problem 3.1 and Assumption 3.2, let $(\boldsymbol{z}_0, \boldsymbol{z}_1) \in \mathcal{H}^2$ and consider the sequence $(\boldsymbol{z}_n)_{n \in \mathbb{N}}$ generated by Algorithm 3.3. Suppose that $\boldsymbol{A}$ is s-monotone with respect to $\boldsymbol{S}$ for $s \geq 0$. Let $\varepsilon > 0$, let $\theta \in\,]0, +\infty[$ be such that $s\theta < 1$, and, for every $n \in \mathbb{N}$, define*

$$r = \frac{s\lambda}{2} \min\{(1 - \zeta^2 - \varepsilon)\,\theta, 2\gamma\}, \tag{3.2}$$

$$\nu_n = \begin{cases} 0, & \text{if } -(\gamma_n \boldsymbol{M}_n - \boldsymbol{S}) \text{ is monotone,} \\ 2\zeta_n, & \text{otherwise.} \end{cases} \tag{3.3}$$

$$\psi_n = \frac{2 - \varepsilon + \nu_n - s\theta(1 - \zeta_n^2 - \varepsilon)}{1 + \zeta_n^2 + \nu_n}, \tag{3.4}$$

$$\rho_n = \frac{\psi_n}{\lambda_n} - 1, \tag{3.5}$$

$$\delta_n = (1 - r)(1 - \alpha_n)\rho_n - \alpha_{n+1}(1 - \alpha_{n+1})\rho_{n+1} - (1 - r)\alpha_{n+1}(1 + \alpha_{n+1}). \tag{3.6}$$

Moreover, suppose that, for every $n \geq N \in \mathbb{N}$, $\rho_n \geq 0$ and let $\varepsilon > 0$ be such that

$$(\forall n \geq N) \quad \varepsilon < 1 - \zeta_n^2. \tag{3.7}$$

Therefore, for every $\boldsymbol{z} \in \operatorname{zer}(\boldsymbol{A} + \boldsymbol{C})$ and every $n \geq N$,

$$\|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha_n \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + (1 - \alpha_n)\rho_n \|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2$$

$$\leq (1 - r)(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha_n \|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + (1 - \alpha_{n-1})\rho_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2)$$

$$- \delta_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2 - \frac{\lambda_n \gamma_n}{\varepsilon}(2\beta\varepsilon - \gamma_n)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2. \tag{3.8}$$

Proof. Fix $n \in \mathbb{N}$. Let $\boldsymbol{z} \in \operatorname{zer}(\boldsymbol{A} + \boldsymbol{C})$, thus, $-\boldsymbol{C}\boldsymbol{z} \in \boldsymbol{A}\boldsymbol{z}$. It follows from (3.1b) that

$$\boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n - \boldsymbol{C}\boldsymbol{y}_n \in \boldsymbol{A}\boldsymbol{x}_n.$$

Hence, by the s-strong monotonicity of $\boldsymbol{A}$ we have

$$2\gamma_n \langle \boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n \mid \boldsymbol{x}_n - \boldsymbol{z} \rangle - 2\gamma_n \langle \boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z} \mid \boldsymbol{x}_n - \boldsymbol{z} \rangle \geq 2s\gamma_n \|\boldsymbol{x}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2. \tag{3.9}$$

Furthermore, by the β-cocoercivity of $\boldsymbol{C}$ with respect to $\boldsymbol{S}$ and Young's inequality, we have

$$-2\gamma_n \langle \boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z} \mid \boldsymbol{x}_n - \boldsymbol{z} \rangle = -2\gamma_n \langle \boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z} \mid \boldsymbol{y}_n - \boldsymbol{z} \rangle - 2\gamma_n \langle \boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z} \mid \boldsymbol{x}_n - \boldsymbol{y}_n \rangle$$

$$\leq -\frac{\gamma_n}{\varepsilon}(2\beta\varepsilon - \gamma_n)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2 + \varepsilon\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2. \tag{3.10}$$

Moreover, it follows from (3.1c) that

$$\|\boldsymbol{w}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 = \|\boldsymbol{y}_n - \gamma_n \boldsymbol{S}^{-1}(\boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n) - \boldsymbol{z}\|_{\boldsymbol{S}}^2$$

$$= \|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - 2\gamma_n \langle \boldsymbol{y}_n - \boldsymbol{z} \mid \boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n \rangle + \gamma_n^2 \|\boldsymbol{S}^{-1}(\boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n)\|_{\boldsymbol{S}}^2. \tag{3.11}$$

Note that, by summing the last two terms in (3.11) and the first term in (3.9), by the Lipschitzian property of $\gamma_n \boldsymbol{M}_n - \boldsymbol{S}$, we obtain

$$2\gamma_n \langle \boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n \mid \boldsymbol{x}_n - \boldsymbol{z} \rangle - 2\gamma_n \langle \boldsymbol{y}_n - \boldsymbol{z} \mid \boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n \rangle + \gamma_n^2 \|\boldsymbol{S}^{-1}(\boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n)\|_{\boldsymbol{S}}^2$$

$$= 2\gamma_n \langle \boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n \mid \boldsymbol{x}_n - \boldsymbol{y}_n \rangle + \gamma_n^2 \|\boldsymbol{S}^{-1}(\boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n)\|_{\boldsymbol{S}}^2$$

$$\overset{(2.2)}{=} -\|\boldsymbol{x}_n - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2 + \|(\gamma_n \boldsymbol{M}_n - \boldsymbol{S})\boldsymbol{y}_n - (\gamma_n \boldsymbol{M}_n - \boldsymbol{S})\boldsymbol{x}_n)\|_{\boldsymbol{S}^{-1}}^2$$

$$\leq -(1 - \zeta_n^2)\|\boldsymbol{x}_n - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2. \tag{3.12}$$

By combining (3.9)-(3.12), we deduce

$$\|\boldsymbol{w}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 \leq \|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \left(1 - \zeta_n^2 - \varepsilon\right)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2 - \frac{\gamma_n}{\varepsilon}\left(2\beta\varepsilon - \gamma_n\right)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2$$
$$- 2s\gamma_n\|\boldsymbol{x}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2. \tag{3.13}$$

Now, it follows from (3.1d) and (3.13) that

$$\|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 = \|\lambda_n \boldsymbol{w}_{n+1} + (1 - \lambda_n)\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2$$
$$= \|(1 - \lambda_n)(\boldsymbol{y}_n - \boldsymbol{z}) + \lambda_n(\boldsymbol{w}_{n+1} - \boldsymbol{z})\|_{\boldsymbol{S}}^2$$
$$\overset{(2.3)}{=} (1 - \lambda_n)\|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + \lambda_n\|\boldsymbol{w}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \lambda_n(1 - \lambda_n)\|\boldsymbol{w}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2$$
$$\leq (1 - \lambda_n)\|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + \lambda_n\|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \lambda_n\left(1 - \zeta_n^2 - \varepsilon\right)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2$$
$$- \frac{\lambda_n\gamma_n}{\varepsilon}\left(2\beta\varepsilon - \gamma_n\right)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2 - 2s\lambda_n\gamma_n\|\boldsymbol{x}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \lambda_n(1 - \lambda_n)\|\boldsymbol{w}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2$$
$$= \|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \lambda_n\left(1 - \zeta_n^2 - \varepsilon\right)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2 - \frac{\lambda_n\gamma_n}{\varepsilon}\left(2\beta\varepsilon - \gamma_n\right)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2$$
$$- 2s\lambda_n\gamma_n\|\boldsymbol{x}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \frac{1 - \lambda_n}{\lambda_n}\|\boldsymbol{z}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2$$
$$= \|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \lambda_n(1 - s\theta)\left(1 - \zeta_n^2 - \varepsilon\right)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2 - \frac{\lambda_n\gamma_n}{\varepsilon}\left(2\beta\varepsilon - \gamma_n\right)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2$$
$$- \lambda_n s\theta\left(1 - \zeta_n^2 - \varepsilon\right)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2 - 2s\lambda_n\gamma_n\|\boldsymbol{x}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \frac{1 - \lambda_n}{\lambda_n}\|\boldsymbol{z}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2$$
$$\leq \|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \lambda_n(1 - s\theta)\left(1 - \zeta_n^2 - \varepsilon\right)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2 - \frac{\lambda_n\gamma_n}{\varepsilon}\left(2\beta\varepsilon - \gamma_n\right)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2$$
$$- \frac{s\lambda_n}{2}\min\{\theta\left(1 - \zeta_n^2 - \varepsilon\right), 2\gamma_n\}\|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \frac{1 - \lambda_n}{\lambda_n}\|\boldsymbol{z}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2$$
$$\leq (1 - r)\|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \lambda_n(1 - s\theta)\left(1 - \zeta_n^2 - \varepsilon\right)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2 - \frac{\lambda_n\gamma_n}{\varepsilon}\left(2\beta\varepsilon - \gamma_n\right)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2$$
$$- \frac{1 - \lambda_n}{\lambda_n}\|\boldsymbol{z}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2, \tag{3.14}$$

where the last inequality follows by noticing that $r \leq \frac{s\lambda_n}{2}\min\{\theta\left(1 - \zeta_n^2 - \varepsilon\right), 2\gamma_n\}$. Note that if $-(\gamma_n \boldsymbol{M_n} - \boldsymbol{S})$ is monotone then

$$2\langle(\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{y}_n - (\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{x}_n \mid \boldsymbol{y}_n - \boldsymbol{x}_n\rangle \leq 0.$$

Otherwise, by the Cauchy–Schwarz inequality and the ζ_n-Lipschitz property of $\gamma_n \boldsymbol{M}_n - \boldsymbol{S}$

$$2\langle(\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{y}_n - (\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{x}_n \mid \boldsymbol{y}_n - \boldsymbol{x}_n\rangle \leq 2\zeta_n\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2.$$

Hence, from (3.3) we conclude that

$$2\langle(\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{y}_n - (\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{x}_n \mid \boldsymbol{y}_n - \boldsymbol{x}_n\rangle \leq \nu_n\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2. \tag{3.15}$$

Moreover, from (3.1d), (3.1c), the ζ_n-Lipschitz property of $\gamma_n \boldsymbol{M}_n - \boldsymbol{S}$, and (3.15) we have

$$
\begin{aligned}
\frac{1}{\lambda_n^2}\|\boldsymbol{z}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2 &= \|\boldsymbol{w}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2 \\
&= \|\gamma_n \boldsymbol{S}^{-1}(\boldsymbol{M_n}\boldsymbol{y}_n - \boldsymbol{M_n}\boldsymbol{x}_n)\|_{\boldsymbol{S}}^2 \\
&= \|\boldsymbol{S}^{-1}((\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{y}_n - (\gamma_n \boldsymbol{M}_n - \boldsymbol{S})\boldsymbol{x}_n) + (\boldsymbol{y}_n - \boldsymbol{x}_n)\|_{\boldsymbol{S}}^2 \\
&= \|(\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{y}_n - (\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}}^2 + \|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2 \\
&\quad + 2\langle (\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{y}_n - (\gamma_n \boldsymbol{M_n} - \boldsymbol{S})\boldsymbol{x}_n \mid \boldsymbol{y}_n - \boldsymbol{x}_n \rangle \\
&\leq (1 + \zeta_n^2 + \nu_n)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}}^2. \tag{3.16}
\end{aligned}
$$

Hence, by combining (3.14) with (3.16), by condition (3.7), and by noticing that

$$
\frac{(1 - s\theta)(1 - \zeta_n^2 - \varepsilon)}{\lambda_n(1 + \zeta_n^2 + \nu_n)} + \frac{(1 - \lambda_n)}{\lambda_n} = \frac{\psi_n}{\lambda_n} - 1 = \rho_n,
$$

we deduce

$$
\|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 \leq (1 - r)\|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \rho_n\|\boldsymbol{z}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2 - \frac{\lambda_n \gamma_n}{\varepsilon}(2\beta\varepsilon - \gamma_n)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2. \tag{3.17}
$$

Now, (3.1a) yields

$$
\begin{aligned}
\|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 &= \|\boldsymbol{z}_n + \alpha_n(\boldsymbol{z}_n - \boldsymbol{z}_{n-1}) - \boldsymbol{z}\|_{\boldsymbol{S}}^2 \\
&= \|(1 + \alpha_n)(\boldsymbol{z}_n - \boldsymbol{z}) - \alpha_n(\boldsymbol{z}_{n-1} - \boldsymbol{z})\|_{\boldsymbol{S}}^2 \\
&\overset{(2.3)}{=} (1 + \alpha_n)\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha_n\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + \alpha_n(1 + \alpha_n)\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2. \tag{3.18}
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
-\|\boldsymbol{z}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}}^2 &= -\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n - \alpha_n(\boldsymbol{z}_n - \boldsymbol{z}_{n-1})\|_{\boldsymbol{S}}^2 \\
&= -\|(1 - \alpha_n)(\boldsymbol{z}_{n+1} - \boldsymbol{z}_n) + \alpha_n(\boldsymbol{z}_{n+1} - 2\boldsymbol{z}_n + \boldsymbol{z}_{n-1})\|_{\boldsymbol{S}}^2 \\
&\overset{(2.3)}{=} -(1 - \alpha_n)\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 + \alpha_n(1 - \alpha_n)\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2 - \alpha_n\|\boldsymbol{z}_{n+1} - 2\boldsymbol{z}_n + \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2 \\
&\leq -(1 - \alpha_n)\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 + \alpha_n(1 - \alpha_n)\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2. \tag{3.19}
\end{aligned}
$$

Therefore, in view of (3.17)-(3.19), we deduce that

$$
\begin{aligned}
&\|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - (1 - r)\alpha_n\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + (1 - \alpha_n)\rho_n\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 \\
&\leq (1 - r)(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha_n\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + (1 - \alpha_{n-1})\rho_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2) \\
&\quad - ((1 - r)(1 - \alpha_{n-1})\rho_{n-1} - \alpha_n(1 - \alpha_n)\rho_n - (1 - r)\alpha_n(1 + \alpha_n))\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2 \\
&\quad - \frac{\lambda_n \gamma_n}{\varepsilon}(2\beta\varepsilon - \gamma_n)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2
\end{aligned}
$$

and the result follows. $\qquad\square$

Remark 3.6. *Let $(\rho_n)_{n \in \mathbb{N}}$ and $(\delta_n)_{n \in \mathbb{N}}$ be the sequences defined in (3.5), (3.6), respectively. We will derive the weak convergence of Algorithm 3.3 under the assumption $\liminf \delta_n = \delta > 0$. Note that this assumption ensures the hypothesis on $(\rho_n)_{n \in \mathbb{N}}$ stated in Proposition 3.5, that is, there exists $N \in \mathbb{N}$ such that, $\rho_n \geq 0$, for $n \geq N$. Indeed, since $\alpha_n \to \alpha$, we have*

$$
(1 - r)(1 - \alpha)\liminf \rho_n = \liminf(1 - r)(1 - \alpha_n)\rho_n \geq \liminf \delta_n \geq \delta > 0.
$$

The following theorem proves the convergence of Algorithm 3.3 under nondecreasing inertial sequences.

Theorem 3.7. *In the context of Problem 3.1 and Assumption 3.2, consider the sequence $(\boldsymbol{z}_n)_{n \in \mathbb{N}}$ defined recursively by Algorithm 3.3 with initialization points $(\boldsymbol{z}_0, \boldsymbol{z}_{-1}) \in \mathcal{H}^2$. Let $(\nu_n)_{n \in \mathbb{N}}$, $(\rho_n)_{n \in \mathbb{N}}$, and $(\delta_n)_{n \in \mathbb{N}}$ be the sequences defined in (3.3), and (3.5), (3.6), respectively, for $r = 0$. Suppose that, there exist $(\varepsilon, \delta) \in]0, +\infty[^2$ and $N \in \mathbb{N}$ such that*

$$(\forall n \geq N) \quad 1 - \zeta_n^2 - \varepsilon > \delta \text{ and } 2\beta\varepsilon - \gamma_n \geq 0. \tag{3.20}$$

Additionally, suppose that $(\alpha_n)_{n \geq N}$ is nondecreasing and that

$$\liminf \delta_n = \delta > 0. \tag{3.21}$$

Then, the following assertions hold:

(1) $\sum_{n \in \mathbb{N}} \|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$ and, for every $\boldsymbol{z} \in \mathrm{zer}(\boldsymbol{A} + \boldsymbol{C})$, $(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}})_{n \in \mathbb{N}}$ is convergent.
(2) $(\boldsymbol{z}_n)_{n \in \mathbb{N}}$ converges weakly to some solution to Problem 3.1.

Proof. (1) Define, for each $n \in \mathbb{N}$,

$$\boldsymbol{H}_n = \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha_{n-1}\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + \alpha_{n-1}(1 - \alpha_{n-1})\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2. \tag{3.22}$$

Hence, it follows from Proposition 3.5, Remark 3.6, and the nondecreasing property of $(\alpha_n)_{n \in \mathbb{N}}$, that for every $n \geq N$, we have

$$\boldsymbol{H}_{n+1} \leq \boldsymbol{H}_n - \delta_n\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 - \frac{\lambda_n \gamma_n}{\varepsilon}(2\beta\varepsilon - \gamma_n)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{S}^{-1}}^2. \tag{3.23}$$

Therefore, in view of (3.20) and (3.21), we conclude that $(\boldsymbol{H}_n)_{n \geq N}$ is nonincreasing. Hence, since $(\alpha_n)_{n \in \mathbb{N}}$ is nondecreasing, we deduce

$$(\forall n \geq N) \quad \boldsymbol{H}_N \geq \boldsymbol{H}_n \geq \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha_n\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 \geq \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2. \tag{3.24}$$

Therefore, by an inductive procedure, we deduce

$$\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 \leq \alpha\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + \boldsymbol{H}_N$$

$$\leq \alpha^{n-N}\|\boldsymbol{z}_N - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + \boldsymbol{H}_N \sum_{k=0}^{n-N-1} \alpha^k$$

$$= \alpha^{n-N}\|\boldsymbol{z}_N - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + \boldsymbol{H}_N \frac{1 - \alpha^{n-N}}{1 - \alpha}.$$

Hence, $(\boldsymbol{z}_n)_{n \in \mathbb{N}}$ and $(\boldsymbol{H}_n)_{n \in \mathbb{N}}$ are bounded sequences. Moreover, since $(\boldsymbol{H}_n)_{n \geq N}$ is nonincreasing, it is convergent. Now, by telescoping (3.23) we have,

$$(\forall n \geq N) \quad \sum_{k=N}^{n} \delta_k\|\boldsymbol{z}_{k+1} - \boldsymbol{z}_k\|_{\boldsymbol{S}}^2 \leq \boldsymbol{H}_N - \boldsymbol{H}_n. \tag{3.25}$$

Then, by taking $n \to \infty$, we conclude that $\sum_{n \in \mathbb{N}} \delta_n\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$. Moreover, since $\liminf \delta_n = \delta > 0$, we conclude that $\sum_{n \in \mathbb{N}} \|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$. By defining, for each $n \in \mathbb{N}$, $a_n = \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2$ and $b_n = (\epsilon_{n+1} - \delta_n)\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 + \epsilon_n\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$, since $(\epsilon_n)_{n \in \mathbb{N}}$ is bounded, it follows from Proposition 3.5 and Lemma 2.1 that, for every $\boldsymbol{z} \in \mathrm{zer}(\boldsymbol{A} + \boldsymbol{C})$, $(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}})_{n \in \mathbb{N}}$ is convergent.

(2) Note that, since $\gamma_n \geq \gamma$, by (3.1c)

$$
\begin{aligned}
\gamma \|\boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}} &\leq \|\gamma_n \boldsymbol{M}_n \boldsymbol{y}_n - \gamma_n \boldsymbol{M}_n \boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}} \\
&= \|\boldsymbol{w}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}} \\
&= \frac{1}{\lambda_n} \|\boldsymbol{z}_{n+1} - \boldsymbol{y}_n\|_{\boldsymbol{S}} \\
&\leq \frac{1}{\underline{\lambda}} (\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}} + \alpha_n \|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}) \to 0,
\end{aligned}
$$

thus, we conclude that

$$
\|\boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}} \to 0 \text{ and } \|\gamma_n \boldsymbol{M}_n \boldsymbol{y}_n - \gamma_n \boldsymbol{M}_n \boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}} \to 0. \tag{3.26}
$$

On the other hand,

$$
\begin{aligned}
\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}} &\leq \|(\gamma_n \boldsymbol{M}_n - \boldsymbol{S})\boldsymbol{y}_n - (\gamma_n \boldsymbol{M}_n - \boldsymbol{S})\boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}} + \|\gamma_n \boldsymbol{M}_n \boldsymbol{y}_n - \gamma_n \boldsymbol{M}_n \boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}} \\
&\leq \zeta_n \|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}} + \|\gamma_n \boldsymbol{M}_n \boldsymbol{y}_n - \gamma_n \boldsymbol{M}_n \boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}}.
\end{aligned}
$$

Therefore, since $\zeta_n < \zeta < 1$,

$$
(1 - \zeta)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}} \leq \|\gamma_n \boldsymbol{M}_n \boldsymbol{y}_n - \gamma_n \boldsymbol{M}_n \boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}} \to 0. \tag{3.27}
$$

Furthermore, by the β-cocoercivity of $\boldsymbol{C}$, we conclude

$$
\|\boldsymbol{C}_n \boldsymbol{y}_n - \boldsymbol{C}_n \boldsymbol{x}_n\|_{S^{-1}} \leq \frac{1}{\beta} \|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{S}} \to 0. \tag{3.28}
$$

Now, let $\boldsymbol{z}^*$ be a weak cluster point of $(\boldsymbol{z}_n)_{n \in \mathbb{N}}$ and let $(\boldsymbol{z}_{n_k})_{k \in \mathbb{N}}$ be a subsequence such that $\boldsymbol{z}_{n_k} \rightharpoonup \boldsymbol{z}^*$. We have, $\boldsymbol{y}_{n_k} = \boldsymbol{z}_{n_k} + \alpha_{n_k}(\boldsymbol{z}_{n_k} - \boldsymbol{z}_{n_k+1}) \rightharpoonup \boldsymbol{z}^*$, thus, by (3.27), $\boldsymbol{x}_{n_k} \rightharpoonup \boldsymbol{z}^*$. Therefore, it follows from (3.1b) that

$$
(\boldsymbol{M}_{n_k} \boldsymbol{y}_{n_k} - \boldsymbol{M}_{n_k} \boldsymbol{x}_{n_k}) - (\boldsymbol{C} \boldsymbol{y}_{n_k} - \boldsymbol{C} \boldsymbol{x}_{n_k}) \in (\boldsymbol{A} + \boldsymbol{C})\boldsymbol{x}_{n_k}. \tag{3.29}
$$

Since $\boldsymbol{C}$ is cocoercive, $\boldsymbol{A} + \boldsymbol{C}$ is a maximally monotone operator [10, Corollary 25.5]. Hence, from the weak-strong closedness of $\mathrm{gra}\,(\boldsymbol{A} + \boldsymbol{C})$ [10, Proposition 20.38], (3.26), (3.28), and (3.29), we conclude that every weak cluster point of $(\boldsymbol{z}_n)_{n \in \mathbb{N}}$ belongs to $\mathrm{zer}(\boldsymbol{A} + \boldsymbol{C})$. The result follows by applying [10, Lemma 2.47] in the space $(\mathcal{H}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{S}})$.
□

Remark 3.8.   *(1) By consider $\boldsymbol{M}_n = \frac{1}{\gamma_n} \boldsymbol{S}$, according to (3.3), $\zeta_n \equiv \nu_n \equiv 0$ and Algorithm 3.3 reduces to the inertial and relaxed version of the classical forward-backward algorithm. If $\gamma_n \equiv \gamma \in \,]0, +\infty[$, $\lambda_n \to \lambda \in \,]0, 2[$, $r = 0$, and we set $\varepsilon = \gamma/2\beta$, (3.21) is equivalent to*

$$
\frac{(1 - \alpha)^2}{\lambda} \left(2 - \lambda - \frac{\gamma}{2\beta} \right) - \alpha(1 + \alpha) > 0
$$

which corresponds with the condition in [5, Corollary 3.12]. Furthermore, if $\lambda = 1$, it reduces to

$$
1 - 3\alpha - \frac{\gamma(1 - \alpha)^2}{2\beta} > 0, \tag{3.30}
$$

which is the condition proposed in [51, Remark 3].
(2) The possibility of adding inertia in NFB in the case where $\boldsymbol{C} = 0$ and $\boldsymbol{S} = \mathrm{Id}$ was discussed in [26, Remark 4.4(ii)]. However, the conditions that guarantee convergence were not derived.

The nondecreasing property of $(\alpha_n)_{n\in\mathbb{N}}$ is crucial for proving that $(\boldsymbol{H}_n)_{n\in\mathbb{N}}$, defined in (3.22), is nonincreasing, which in turn allows us to establish the convergence of Algorithm 3.3.

The next theorem ensures the convergence of Algorithm 3.3 for a decreasing sequence of inertial parameters $(\alpha_n)_{n\in\mathbb{N}}$ satisfying $\sum_{n\in\mathbb{N}}(\alpha_n - \alpha) < +\infty$, for an adequate $\alpha \in]0,1[$. Hence, the inertial step $\boldsymbol{y}_n$ can be interpreted as an ℓ^1-type approximation of the inertial step generated with an admissible parameter α. In this case, the term $\boldsymbol{K}_n$, defined in (3.33), is not Fejér monotone but quasi Fejér monotone, i.e., admits an ℓ^1-type error (see [10, Section 5.4]). Although this result may appear weaker than Theorem 3.7, it allows a practical acceleration of convergence (see Section 5), as it permits larger inertial parameters during the initial iterations. To the best of our knowledge, this is the first approach that considers a decreasing sequence of inertial parameters.

Theorem 3.9. *In the context of Problem 3.1 and Assumption 3.2, consider the sequence $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ defined recursively by Algorithm 3.3 with initialization points $(\boldsymbol{z}_0, \boldsymbol{z}_{-1}) \in \mathcal{H}^2$. Let $(\nu_n)_{n\in\mathbb{N}}$, $(\rho_n)_{n\in\mathbb{N}}$, and $(\delta_n)_{n\in\mathbb{N}}$ be the sequences defined in (3.3), (3.5), and (3.6), respectively, for $r = 0$. Suppose that, there exist $(\varepsilon, \delta) \in]0, +\infty[^2$ and $N \in \mathbb{N}$ such that*

$$(\forall n \geq N) \quad 1 - \zeta_n^2 - \varepsilon > \delta \text{ and } 2\beta\varepsilon - \gamma_n \geq 0. \tag{3.31}$$

Additionally, suppose that $(\alpha_n)_{n\geq N}$ is decreasing, that $\sum_{n\in\mathbb{N}}(\alpha_n - \alpha) < +\infty$, and that

$$\liminf \delta_n = \delta > 0. \tag{3.32}$$

Then, the following assertions hold:

(1) $\sum_{n\in\mathbb{N}} \|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$ and, for every $\boldsymbol{z} \in \mathrm{zer}(\boldsymbol{A} + \boldsymbol{C})$, $(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}})_{n\in\mathbb{N}}$ is convergent.
(2) $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ converges weakly to some solution to Problem 3.1.

Proof. (1) For every $n \in \mathbb{N}$, define

$$\boldsymbol{K}_n = \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + (1 - \alpha_{n-1})\rho_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2. \tag{3.33}$$

First, suppose that $\alpha = 0$, thus $\sum_{n\in\mathbb{N}} \alpha_n < +\infty$ and $\inf_{n\in\mathbb{N}} \boldsymbol{K}_n \geq 0$. Then Proposition 3.5, Remark 3.6, and (3.31) yield

$$(\forall n \geq N) \quad \boldsymbol{K}_{n+1} \leq \boldsymbol{K}_n + \alpha_n\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha_n\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \delta_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$$

$$\leq (1 + \alpha_n)\boldsymbol{K}_n - \delta_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2.$$

Hence, in view of Lemma 2.2 applied to $a_n = \boldsymbol{K}_n$, $t_n = \alpha_n$, and $b_n = \delta_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$, we conclude that $(\boldsymbol{K}_n)_{n\in\mathbb{N}}$ is convergent and $\sum_{n\in\mathbb{N}} \delta_n\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$. Since $\liminf \delta_n = \delta > 0$, we deduce that $\sum_{n\in\mathbb{N}} \|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$. Then, since $((1 - \alpha_n)\rho_n)_{n\in\mathbb{N}}$ is bounded and $\alpha = 0$, by the convergence of $(\boldsymbol{K}_n)_{n\in\mathbb{N}}$ and (3.33), we have that $(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}})_{n\in\mathbb{N}}$ is convergent for every $\boldsymbol{z} \in \mathrm{zer}(\boldsymbol{A} + \boldsymbol{C})$.

Now, suppose that $\alpha > 0$,. Since $\alpha < 1$, it follows from Proposition 3.5 and (3.31) that, for every $n \geq N$,

$$\boldsymbol{K}_{n+1} \leq \boldsymbol{K}_n + (\alpha_n - \alpha)\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - (\alpha_n - \alpha)\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \delta_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$$

$$\leq \boldsymbol{K}_n + \frac{(\alpha_n - \alpha)}{\alpha}\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha\frac{(\alpha_n - \alpha)}{\alpha}\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \delta_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$$

$$\leq \left(1 + \frac{\alpha_n - \alpha}{\alpha}\right)\boldsymbol{K}_n - \delta_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2. \tag{3.34}$$

If there exist $n_0 \geq N$ such that $\boldsymbol{K}_{n_0} \leq 0$, it follows from (3.34) that $\boldsymbol{K}_n \leq 0$, for every $n \geq n_0$. Therefore, we have

$$(\forall n \geq n_0) \quad \boldsymbol{K}_n \leq 0 \Leftrightarrow (\forall n \geq n_0) \quad \|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 \leq \alpha\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - (1 - \alpha_{n-1})\rho_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$$

$$\leq \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - (1 - \alpha_{n-1})\rho_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2.$$

Hence, in view of Lemma 2.2 applied to $a_n = \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2$, $t_n \equiv 0$, and $b_n = (1 - \alpha_{n-1})\rho_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$, we conclude that, for every $\boldsymbol{z} \in \mathrm{zer}(\boldsymbol{A} + \boldsymbol{C})$, $(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}})_{n\in\mathbb{N}}$ is convergent and $\sum_{n\in\mathbb{N}}(1-\alpha_n)\rho_n\|\boldsymbol{z}_{n+1}-\boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$. Since $\liminf(1-\alpha_n)\rho_n \geq \delta > 0$, we deduce that $\sum_{n\in\mathbb{N}}\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$.

On the other hand, if $\boldsymbol{K}_n \geq 0$, for every $n \geq N$, by (3.34) and invoking again Lemma 2.2 applied to $a_n = \boldsymbol{K}_n$, $t_n = (\alpha_n - \alpha)\alpha^{-1}$, and $b_n = \delta_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$, we conclude that $(\boldsymbol{K}_n)_{n\in\mathbb{N}}$ is convergent and $\sum_{n\in\mathbb{N}}\delta_n\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$. Since $\liminf \delta_n = \delta > 0$, we deduce that $\sum_{n\in\mathbb{N}}\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$. By defining $b_n = \frac{\alpha_n - \alpha}{\alpha}\boldsymbol{K}_n + (1 - \alpha_{n-1})\rho_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$, it follows from (3.33) and (3.34) that, for every $n \geq N$,

$$\|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 \leq \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + \alpha(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2) + ((1-\alpha_{n-1})\rho_{n-1} - \delta_{n-1})\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$$

$$+ \frac{\alpha_n - \alpha}{\alpha}\boldsymbol{K}_n - (1 - \alpha_n)\rho_n\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2$$

$$\leq \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + \alpha(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2) + b_n. \tag{3.35}$$

Now, since $\sum_{n\in\mathbb{N}}(\alpha_n - \alpha) < +\infty$, $\sum_{n\in\mathbb{N}}\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$, $(\boldsymbol{K}_n)_{n\in\mathbb{N}}$ is convergent (therefore bounded), and $((1 - \alpha_n)\rho_n)_{n\in\mathbb{N}}$ is bounded, we deduce $\sum_{n\in\mathbb{N}} b_n < +\infty$. Therefore, by Lemma 2.1 applied to $a_n = \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2$, we conclude that $(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2)_{n\in\mathbb{N}}$ is convergent for every $\boldsymbol{z} \in \mathrm{zer}(\boldsymbol{A} + \boldsymbol{C})$.

(2) By the convergence of $(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}})_{n\in\mathbb{N}}$, for every $\boldsymbol{z} \in \mathrm{zer}(\boldsymbol{A}+\boldsymbol{C})$, and since $\sum_{n\in\mathbb{N}}\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 < +\infty$, the proof of this assertion is analogous to the proof of Theorem 3.7(2). $\qquad\square$

3.1. **Linear Convergence.** In this section, we prove the linear convergence of Algorithm 3.3 in the case where $\boldsymbol{A}$ is s strongly monotone for $s > 0$. In this analysis, for simplicity, we suppose that the sequence $(\rho_n)_{n\in\mathbb{N}}$ converges to some $\rho \in]0, +\infty[$. The following proposition yields a fundamental estimate for proving the strong convergence of our algorithm.

Proposition 3.10. *In the context of Problem 3.1 and Assumption 3.2, consider the sequence $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ defined recursively by Algorithm 3.3 with initialization points $(\boldsymbol{z}_0, \boldsymbol{z}_{-1}) \in \mathcal{H}^2$. Let $(\nu_n)_{n\in\mathbb{N}}$, $(\rho_n)_{n\in\mathbb{N}}$, and $(\delta_n)_{n\in\mathbb{N}}$ be the sequences defined in (3.3), (3.5), and (3.6), respectively. Suppose that $\rho_n \to \rho \in]0, +\infty[$, that $\alpha_n \to \alpha \in]0, +\infty[$, and that*

$$\delta := (1 - r)(1 - \alpha)\rho - \alpha(1 - \alpha)\rho - \alpha(1 + \alpha) > 0. \tag{3.36}$$

Let $(\boldsymbol{H}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{K}_n)_{n\in\mathbb{N}}$ be the sequences defined in (3.22) and (3.33), respectively. Then, there exists $N \in \mathbb{N}$ such that, for every $n \geq N$, the following statements hold:

(1) There exists $\epsilon_1 \in]0, +\infty[$ such that

$$\delta + r(1-\alpha)\rho + \alpha(1-\alpha)\rho + \alpha^2 + \epsilon_1 \geq (1-\alpha_n)\rho_n - \alpha \geq \delta + r(1-\alpha)\rho + \alpha(1-\alpha)\rho + \alpha^2 - \epsilon_1 > 0.$$

(2) There exists $\epsilon_2 \in]0, +\infty[$ such that

$$\rho_n(1 - \alpha_n)(1 - \alpha) - \alpha \geq \delta + \alpha^2 - \epsilon_2 > 0.$$

(3) $\liminf \delta_n \geq \delta$.

(4) $\boldsymbol{K}_{n+1} \geq \left(\dfrac{\delta + \alpha^2 - \epsilon_2}{\delta + r(1-\alpha)\rho + \alpha(1-\alpha)\rho + \alpha^2 + \epsilon_1}\right) \|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2$.

(5) $\boldsymbol{H}_{n+1} \geq \left(\dfrac{\delta + \alpha^2 - \epsilon_2}{\delta + r(1-\alpha)\rho + \alpha(1-\alpha)\rho + \alpha^2 + \epsilon_1}\right) \|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2$.

Proof. (1) Since $\rho_n \to \rho$ and $\alpha_n \to \alpha$, the result follows by noticing that

$$(1 - \alpha_n)\rho_n - \alpha \to (1 - \alpha)\rho - \alpha = \delta + r(1-\alpha)\rho + \alpha(1-\alpha)\rho + \alpha^2 > 0.$$

(2) Analogously, we have

$$\rho_n(1-\alpha_n)(1-\alpha) - \alpha \to \rho(1-\alpha)^2 - \alpha = (1-\alpha)\rho - \alpha(1-\alpha)\rho - \alpha$$
$$\geq (1-r)(1-\alpha)\rho - \alpha(1-\alpha)\rho - \alpha$$
$$= \delta + \alpha^2.$$

The result follows.

(3) It follows directly from

$$\liminf \delta_n = (1-r)(1-\alpha)\rho - \alpha(1-\alpha)\rho - (1-r)\alpha(1+\alpha) \geq \delta.$$

(4) Fix $n \in \mathbb{N}$, let $\kappa_n \in \,]0,+\infty[$. By (3.33) and Young's inequality, we have

$$\boldsymbol{K}_{n+1} = \|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + (1-\alpha_n)\rho_n\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2$$

$$\geq \|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha\left(1 + \frac{1}{\kappa_n}\right)\|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \alpha(1+\kappa_n)\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2 + (1-\alpha_n)\rho_n\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2$$

$$\geq \left(1 - \alpha\left(1 + \frac{1}{\kappa_n}\right)\right)\|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 + ((1-\alpha_n)\rho_n - \alpha(1+\kappa_n))\|\boldsymbol{z}_{n+1} - \boldsymbol{z}_n\|_{\boldsymbol{S}}^2.$$

Choosing κ_n such that $((1-\alpha_n)\rho_n - \alpha(1+\kappa_n)) = 0$, we obtain $\kappa_n = \frac{(1-\alpha_n)\rho_n - \alpha}{\alpha}$. Note that, in view of 1, for $n \geq N$, we have $\kappa_n \geq \frac{\delta + r(1-\alpha)\rho + \alpha(1-\alpha)\rho + \alpha^2 - \epsilon_1}{\alpha} > 0$, thus, it is well defined. Hence, by 1 and 2

$$\boldsymbol{K}_{n+1} \geq \left(1 - \alpha\left(1 + \frac{1}{\kappa_n}\right)\right)\|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2$$
$$= \left(\frac{\rho_n(1-\alpha_n)(1-\alpha) - \alpha}{\rho_n(1-\alpha_n) - \alpha}\right)\|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 \qquad (3.37)$$
$$\geq \left(\frac{\delta + \alpha^2 - \epsilon_2}{\delta + r(1-\alpha)\rho + \alpha(1-\alpha)\rho + \alpha^2 + \epsilon_1}\right)\|\boldsymbol{z}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2.$$

(5) By the nondecreasing property of $(\alpha_n)_{n\in\mathbb{N}}$, for every $n \in \mathbb{N}$, we have $\alpha_n \leq \alpha$. Hence, the result follows directly from 4 noticing that $\boldsymbol{H}_n \geq \boldsymbol{K}_n$.

$\square$

Theorem 3.11. *In the context of Problem 3.1 and Assumption 3.2, consider the sequence $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ defined recursively by Algorithm 3.3 with initialization points $(\boldsymbol{z}_0, \boldsymbol{z}_{-1}) \in \mathcal{H}^2$. Let $(\nu_n)_{n\in\mathbb{N}}$, $(\rho_n)_{n\in\mathbb{N}}$, and $(\delta_n)_{n\in\mathbb{N}}$ be the sequences defined in (3.3), (3.5), and (3.6), respectively. Suppose that, there exists $(\varepsilon, \delta) \in \,]0,+\infty[^2$ and $N \in \mathbb{N}$ such that*

$$(\forall n \geq N) \quad 1 - \zeta_n^2 - \varepsilon > \delta \text{ and } 2\beta\varepsilon - \gamma_n \geq 0. \qquad (3.38)$$

Additionally, suppose that $\rho_n \to \rho \in \,]0,+\infty[$, that $\alpha_n \to \alpha \in \,]0,+\infty[$, that

$$\delta := (1-r)(1-\alpha)\rho - \alpha(1-\alpha)\rho - \alpha(1+\alpha) > 0, \qquad (3.39)$$

and that one of the following assertions holds

(1) *$(\alpha_n)_{n\in\mathbb{N}}$ is nondecreasing.*
(2) *$(\alpha_n)_{n\in\mathbb{N}}$ is decreasing and $\sum_{n\in\mathbb{N}}(\alpha_n - \alpha) < +\infty$.*

Then, if $\boldsymbol{A}$ is s-strongly monotone with respect to $\boldsymbol{S}$, for some $s \in \,]0,+\infty[$, $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ converges strongly, at linear rate, to the unique solution of Problem 3.1.

Proof. Let us assume that $\boldsymbol{A}$ is s-strongly monotone with respect to $\boldsymbol{S}$ for $s \in \,]0,+\infty[$.

(1) First, suppose that $(\alpha_n)_{n\in\mathbb{N}}$ is nondecreasing. It follows from Proposition 3.5 that for $(\boldsymbol{H}_n)_{n\in\mathbb{N}}$ defined in (3.22) and all $n \geq N$, we have

$$\boldsymbol{H}_{n+1} \leq (1-r)\boldsymbol{H}_n, \tag{3.40}$$

where $r > 0$ is defined in (3.2). Therefore, $(\boldsymbol{H}_n)_{n\in\mathbb{N}}$ converges to 0 at linear rate. If $\alpha = 0$, since $(\alpha_n)_{n\in\mathbb{N}}$ is nondecreasing, we have $\alpha_n \equiv 0$. Hence, $\boldsymbol{H}_n \geq \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2$ and we conclude that $(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}})_{n\in\mathbb{N}}$ converges to 0 at linear rate. Otherwise, if $\alpha > 0$, $(\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}})_{n\in\mathbb{N}}$ converges to 0 at linear rate in view of Proposition 3.10(5).

(2) Suppose now that $(\alpha_n)_{n\in\mathbb{N}}$ is decreasing and $\sum_{n\in\mathbb{N}}(\alpha_n - \alpha) < +\infty$. If $\alpha = 0$, from Proposition 3.5 it follows that, for $(\boldsymbol{K}_n)_{n\in\mathbb{N}}$ defined in (3.33) and every $n \geq N$,

$$\boldsymbol{K}_{n+1} \leq (1-r)\boldsymbol{K}_n + \alpha_n\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - (1-r)\alpha_n\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \delta_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$$
$$\leq (1 - r + \alpha_n)\boldsymbol{K}_n.$$

By Lemma 2.3 we conclude that $(\boldsymbol{K}_n)_{n\in\mathbb{N}}$ converges linearly to 0 and the result follows by noticing that $\boldsymbol{K}_n \geq \|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2$. On the other hand, if $\alpha > 0$, by Proposition 3.5, we deduce

$$\boldsymbol{K}_{n+1} \leq (1-r)\boldsymbol{K}_n + (\alpha_n - \alpha)\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - (1-r)(\alpha_n - \alpha)\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \delta_{n-1}\|\boldsymbol{z}_n - \boldsymbol{z}_{n-1}\|_{\boldsymbol{S}}^2$$

$$\leq (1-r)\boldsymbol{K}_n + \frac{(\alpha_n - \alpha)}{\alpha}\|\boldsymbol{z}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - (\alpha_n - \alpha)\|\boldsymbol{z}_{n-1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2$$

$$\leq \left(1 - r + \frac{\alpha_n - \alpha}{\alpha}\right)\boldsymbol{K}_n.$$

The result follows by combining Proposition 3.10(3) and Lemma 2.3.

$\square$

Remark 3.12. *For simplicity, we assume condition* (3.39) *on the parameters* $(\alpha_n)_{n\in\mathbb{N}}$ *and* $(\rho_n)_{n\in\mathbb{N}}$ *where* $\rho_n \to \rho$. *However, in view of* (3.37), *this requirement can be relaxed to* $\liminf \delta_n = \delta$ *and*

$$\liminf\left(\frac{\rho_n(1-\alpha_n)(1-\alpha) - \alpha}{\rho_n(1-\alpha_n) - \alpha}\right) > 0.$$

3.2. **Analyzing the inertial and relaxation parameters.** In this section, we provide conditions in order to the sequences of inertial and relaxation parameters to fulfill the hypotheses required to ensure the convergence of Algorithm 3.3. First, let us recall that the sequences $(\alpha_n)_{n\in\mathbb{N}}$, $(\lambda_n)_{n\in\mathbb{N}}$, and $(\gamma_n)_{n\in\mathbb{N}}$ are design parameters of the algorithm and can therefore be tuned by the user. In contrast, $(\rho_n)_{n\in\mathbb{N}}$ and $(\delta_n)_{n\in\mathbb{N}}$ depend on $(\zeta_n)_{n\in\mathbb{N}}$, which are the Lipschitz constants of the operators $\gamma_n\boldsymbol{M}_n - \boldsymbol{S}$ and on $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$. The key parameter conditions ensuring convergence of the algorithm, as stated in Proposition 3.5, Theorem 3.7, and Theorem 3.9, are

$$(\exists N \in \mathbb{N})(\forall n \geq N) \quad \rho_n \geq 0 \quad \text{and} \quad \liminf_{n\to+\infty} \delta_n = \delta > 0. \tag{3.41}$$

Now we will determine suitable $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ to ensure that these requirements are satisfied. We split our analysis in two cases: when $\boldsymbol{A}$ is merely monotone ($s = 0$) and when it is strongly monotone ($s > 0$).

3.2.1. $\boldsymbol{A}$ *is a monotone operator.*

Claim 3.13. Condition $\rho_n \geq 0$ is satisfied, if $\varepsilon < 1 - \zeta_n^2$ and $\lambda_n \in [\underline{\lambda}, \psi_n]$, where $\psi_n \in]1, 2[$ for all $n \in \mathbb{N}$.

Proof. According to (3.5), $\rho_n = \psi_n/\lambda_n - 1$. Since λ_n is a positive number, $\rho_n \geq 0$ is fulfilled, if and only if, $\psi_n \geq \lambda_n$. Note that (3.7) imposes $\varepsilon < 1 - \zeta_n^2$, thus, since $\nu_n \geq 0$ (see (3.3)) and by (3.4), we deduce

$$\varepsilon < 1 - \zeta_n^2 \iff \psi_n = \frac{2 - \varepsilon + \nu_n}{1 + \zeta_n^2 + \nu_n} > 1. \tag{3.42}$$

According to Algorithm 3.3, there exists $\underline{\lambda} \in]0, 2[$ such that $\lambda_n \geq \underline{\lambda}$. Hence,

$$\lambda_n \in [\underline{\lambda}, \psi_n] = \left[\underline{\lambda}, \frac{2 - \varepsilon + \nu_n}{1 + \zeta_n^2 + \nu_n}\right]. \tag{3.43}$$

In view of (3.3) $\nu_n = 0$, or $\nu_n = 2\zeta_n$. When $\nu_n = 0$, as $\zeta_n \in [0, 1[$, we have

$$2 - \frac{\varepsilon}{2} > \frac{2}{1 + \zeta_n^2} - \frac{\varepsilon}{1 + \zeta_n^2} = \psi_n. \tag{3.44}$$

On the other hand, if $\nu_n = 2\zeta_n$, we deduce

$$2 - \frac{\varepsilon}{4} > \frac{2(1 + \zeta_n)}{(1 + \zeta_n)^2} - \frac{\varepsilon}{(1 + \zeta_n)^2} = \psi_n. \tag{3.45}$$

Thus, (3.42), (3.44), and (3.44) gives $\psi_n \in]1, 2[$. $\qquad\square$

Remark 3.14. *Notice that the case $\nu_n = 0$ includes, the choice $\boldsymbol{M}_n = \frac{1}{\gamma_n}\boldsymbol{S}$ presented in Remark 3.8. Additional examples where $\nu_n = 0$ will be presented in Section 4.1.*

We study the following condition in a simplified framework. In the context of Problem 3.1, Assumption 3.2, and Algorithm 3.3, suppose that $\lambda_n \to \lambda \in]0, 2[$, $\zeta_n \to \zeta \in [0, 1[$, and $\nu_n \to \nu \in \{0, 2\zeta\}$. Then, in this case, $\rho_n \to \psi/\lambda - 1$ where

$$\psi = \frac{2 - \varepsilon + \nu}{1 + \zeta^2 + \nu}. \tag{3.46}$$

Claim 3.15. Assume that $1 - \zeta^2 - \varepsilon > 0$ and $2\beta\varepsilon \geq \gamma$. Condition $\liminf \delta_n > 0$ is satisfied in each of the following cases:

(1) For $\alpha \in [0, 1[$, we have

$$\lambda \in]0, \psi \cdot \phi(\alpha)[, \quad \text{where} \quad \phi(\alpha) = \frac{(1 - \alpha)^2}{2\alpha^2 - \alpha + 1}. \tag{3.47}$$

In particular, $\phi(\alpha) \in [0, 1[$ and $\lambda \in]0, 2[$.

(2) For $\lambda \in]0, \psi[$, we have

$$\alpha \in [0, \overline{\alpha}[, \quad \text{where} \quad \overline{\alpha} = \frac{2\left(\frac{\psi}{\lambda} - 1\right)}{\left(\frac{2\psi}{\lambda} - 1\right) + \sqrt{\frac{8\psi}{\lambda} - 7}}. \tag{3.48}$$

Proof. (1) In our setting, condition $\liminf \delta_n > 0$ reduces to

$$\liminf \delta_n = (1 - \alpha)^2 \rho - \alpha(1 + \alpha) > 0 \iff (1 - \alpha)^2 \frac{\psi}{\lambda} - 1 + \alpha - 2\alpha^2 > 0 \tag{3.49}$$

$$\iff \frac{(1 - \alpha)^2}{2\alpha^2 - \alpha + 1} \cdot \psi > \lambda \tag{3.50}$$

$$\iff \phi(\alpha) \cdot \psi > \lambda.$$

It is easy to check, that $\phi(0) = 1$, $\lim_{\alpha \nearrow 1} \phi(\alpha) = 0$, and ϕ is a decreasing function over $[0, 1[$, as seen in Figure 1. Using the same arguments as in Claim 3.13, $\psi \in]0, 2[$ and hence, $\lambda \in]0, 2[$.
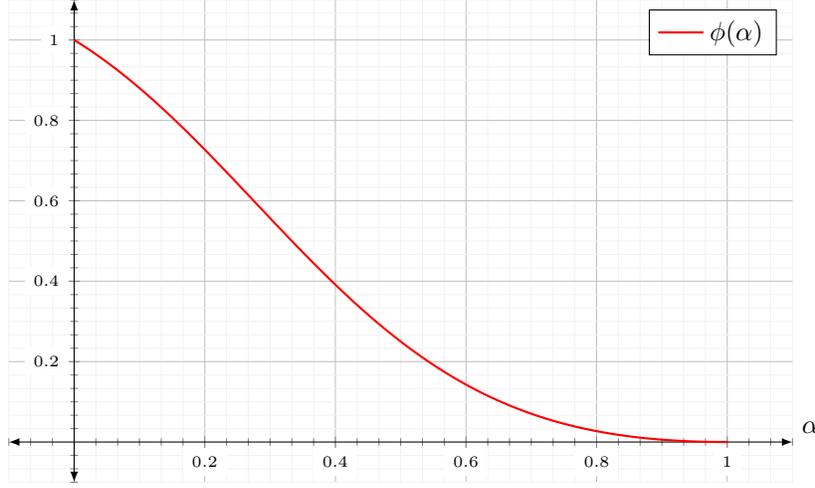
FIGURE 1. Plot of the function $\alpha \mapsto \phi(\alpha) = \dfrac{(1-\alpha)^2}{2\alpha^2 - \alpha + 1}$.

(2) On the other hand, consider $\lambda \in \,]0, \psi[$, and let us define $a = \left(\frac{\psi}{\lambda} - 2\right)$. Then, (3.49) is equivalent to

$$a\alpha^2 - (2a + 3)\alpha + a + 1 > 0. \tag{3.51}$$

Since $\lambda \in]0, \psi[$, we have that $a > -1$. Notice that in the case $a = 0$ (3.51) reduces to $\alpha < \frac{1}{3}$. Now, if $a \neq 0$, the solutions of the quadratic equation corresponding to (3.51), are given by

$$\alpha_1 = \frac{2a + 3 - \sqrt{8a + 9}}{2a} = \frac{2(a + 1)}{(2a + 3 + \sqrt{8a + 9})} \tag{3.52}$$

and

$$\alpha_2 = \frac{2a + 3 + \sqrt{8a + 9}}{2a} = \frac{2(a + 1)}{(2a + 3 - \sqrt{8a + 9})}. \tag{3.53}$$

If $-1 < a < 0$, then $\alpha_1 > 0$ and, since $2a + 3 + \sqrt{8a + 9} > 0$, $\alpha_2 < 0$. Therefore, we conclude that (3.49) holds is $\alpha \in [0, \alpha_1[$. Conversely, if $a > 0$, we have $\frac{1}{3} < \alpha_1 < 1$ and $\alpha_2 > 1$, then (3.49) holds if $\alpha \in [0, \alpha_1[$ and the conclusion follows noticing that $\alpha_1 = \overline{\alpha}$.
□

Remark 3.16. The following particular cases are deduced from Claim 3.15.

(1) If $\lambda \to \psi$, then $\overline{\alpha} \to 0$.
(2) If $\lambda = \psi/2$, then $\overline{\alpha} = 1/3$, which is usually a bound for inertial parameters (see, for instance, [2, 18, 51, 53]).
(3) If $\lambda = 1$, from (3.48) and (3.46) we obtain

$$\alpha \in \left[0, \frac{2\left(\frac{2-\varepsilon+\nu(\zeta)}{1+\zeta^2+\nu(\zeta)} - 1\right)}{\left(2\frac{2-\varepsilon+\nu(\zeta)}{1+\zeta^2+\nu(\zeta)} - 1\right) + \sqrt{8\frac{2-\varepsilon+\nu(\zeta)}{1+\zeta^2+\nu(\zeta)} - 7}}\right[.$$

3.2.2. **A** *is s-strongly monotone.*

Let us consider the case $\lambda_n \to \lambda \in]0,2[$, $\zeta_n \to \zeta \in [0,1[$, and $\nu_n \to \nu \in \{0,2\zeta\}$ and assume that $1 - \zeta^2 - \varepsilon > 0$ and $2\beta\varepsilon \geq \gamma$. Under this setting, the condition guaranteeing convergence in Theorem 3.11 is:

$$\delta := (1-r)(1-\alpha)\rho - \alpha(1-\alpha)\rho - \alpha(1+\alpha) > 0. \tag{3.54}$$

The following claim shows that, for θ sufficiently small, the conditions in (3.43) and (3.49) studied for the case $r = 0$ also guarantee that $\rho > 0$ and that (3.54) holds when $r > 0$.

Claim 3.17.   (1) Suppose that (see (3.43))

$$\lambda \in \left[\underline{\lambda}, \frac{2 - \varepsilon + \nu}{1 + \zeta^2 + \nu} \right[.$$

Then there exists $\theta \in]0,+\infty[$ sufficiently small such that $\rho > 0$.
    (2) Suppose that (see (3.49))

$$(1-\alpha)^2\rho - \alpha(1+\alpha) > 0.$$

Then, there exists $\theta \in]0,+\infty[$ sufficiently small such that $\delta > 0$.

Proof.   (1) Consider $\lambda \in \left[\underline{\lambda}, \frac{2 - \varepsilon + \nu}{1 + \zeta^2 + \nu} \right[$. Let

$$\delta_1 := \frac{2 - \varepsilon + \nu}{\lambda(1 + \zeta + \nu)} - 1 > 0.$$

By picking $\theta > 0$ such that $\theta \leq \delta_1\lambda(1 + \zeta^2 + \nu)/(2s(1 - \zeta^2 - \varepsilon))$, then, by (3.5),

$$\rho = \frac{2 - \varepsilon + \nu - s\theta(1 - \zeta^2 - \varepsilon)}{\lambda(1 + \zeta^2 + \nu)} - 1$$

$$= \delta_1 - \frac{s\theta(1 - \zeta^2 - \varepsilon)}{\lambda(1 + \zeta^2 + \nu)}$$

$$\geq \delta_1 - \frac{\delta_1}{2} > 0.$$

(2) Let $\delta_2 := (1-\alpha)^2\rho - \alpha(1+\alpha) > 0$. Hence, choosing θ such that $r \leq \delta_2/(2\rho(1-\alpha))$ (see (3.2)), we have

$$(1-r)(1-\alpha)\rho - \alpha(1-\alpha)\rho - \alpha(1+\alpha) = (1-\alpha)^2\rho - \alpha(1+\alpha) - r\rho(1-\alpha)$$

$$\geq \delta_2 - \frac{\delta_2}{2} = \frac{\delta_2}{2},$$

which shows that (3.54) holds.

$\square$

The next Claim provides conditions guaranteeing (3.54) in the cases where $\lambda = 1$ and $\alpha = 0$.

Claim 3.18.   (1) Suppose that $\lambda = 1$, then $\rho > 0$ and (3.54) holds if

$$\alpha \in \left[0, \frac{(2-r)\rho + 1 - \sqrt{((2-r)\rho + 1)^2 - 4(1-r)\rho(\rho-1)}}{2(\rho-1)} \right[.$$

(2) Suppose that $\alpha = 0$, then (3.54) holds if

$$\frac{2 - \varepsilon + \nu - s\theta(1 - \zeta^2 - \varepsilon)}{1 + \zeta^2 + \nu} > \lambda,$$

which is always true if $\lambda < 1$.

Proof.   (1) Suppose that $\lambda = 1$, we have

$$\rho = \frac{2 - \varepsilon + \nu - s\theta(1 - \zeta^2 - \varepsilon)}{1 + \zeta^2 + \nu} - 1 = \frac{(1 - s\theta)(1 - \zeta^2 - \varepsilon)}{1 + \zeta^2 + \nu}.$$

Hence, since $\theta \in {]}0, 1/r[$ and $1 - \zeta^2 - \varepsilon > 0$, we have $\rho > 0$. Moreover, it is easy to verify that $\rho - 1 < 0$ and $1 - r > 0$, thus,

$$(1 - r)(1 - \alpha)\rho - \alpha(1 - \alpha)\rho - \alpha(1 + \alpha) = (1 - r)\rho - ((2 - r)\rho + 1)\alpha + (\rho - 1)\alpha^2 > 0$$

$$\Leftrightarrow \alpha \in \left[0, \frac{(2 - r)\rho + 1 - \sqrt{((2 - r)\rho + 1)^2 - 4(1 - r)\rho(\rho - 1)}}{2(\rho - 1)}\right[.$$

The result follows.

(2) When $\alpha = 0$, we have $\delta = (1 - r)\rho$. Hence, since $r < 1$, (3.54) reduces to

$$\rho = \frac{2 - \varepsilon + \nu - s\theta(1 - \zeta^2 - \varepsilon)}{\lambda(1 + \zeta^2 + \nu)} - 1 > 0 \Leftrightarrow \frac{2 - \varepsilon + \nu - s\theta(1 - \zeta^2 - \varepsilon)}{1 + \zeta^2 + \nu} > \lambda.$$

$\square$

3.3. **Relaxing the hypothesis on $\gamma_n \boldsymbol{M}_n - \boldsymbol{S}$.** In this section we discuss how to consider the assumptions made in [46] for the convergence of Algorithm 3.3. Consider the following assumption.

Assumption 3.19. [46, Assumptions 3.1 & 3.2] *Let $\boldsymbol{P}$ be a self adjoint strongly monotone linear operator, let $\boldsymbol{A} \colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $\boldsymbol{C} \colon \mathcal{H} \to \mathcal{H}$ be β-cocoercive with respect to $\boldsymbol{P}$ for $\beta \in {]}0, 1/4[$, and let, for every $n \in \mathbb{N}$, $\boldsymbol{M}_n \colon \mathcal{H} \to \mathcal{H}$ be a 1-strongly monotone with respect to $\boldsymbol{P}$ and ζ-Lipschitz operator for $\zeta > 0$.*

Weak convergence of (1.2) under Assumption 3.19 was established in [46, Theorem 5.1] for $(\theta_n)_{n \in \mathbb{N}}$ in ${]}0, 2[$ and $(\mu_n)_{n \in \mathbb{N}}$ satisfying

$$(\forall n \in \mathbb{N}) \quad \mu_n \le \frac{\langle \boldsymbol{M}_n \boldsymbol{z}_n - \boldsymbol{M}_n \boldsymbol{x}_n \mid \boldsymbol{z}_n - \boldsymbol{x}_n \rangle - \frac{1}{4\beta} \|\boldsymbol{z}_n - \boldsymbol{x}_n\|_{\boldsymbol{P}}^2}{\|\boldsymbol{M}_n \boldsymbol{z}_n - \boldsymbol{M}_n \boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}}^2}. \tag{3.55}$$

In view of [54, Proposition 2.1], with an appropriate choice of $\boldsymbol{P}$, Assumption 3.19 holds in the context of Problem 3.1 and Assumption 3.2. Therefore, Assumption 3.2 can be seen as a particular case of Assumption 3.19.

For simplicity, our previous convergence analysis was restricted to the setting of Assumption 3.2. Nonetheless, Assumption 3.2 allowed us to establish the convergence of the proposed algorithm under simpler step-size conditions, thereby recovering and extending previous results in the literature in a more direct manner (see Section 4.1). In particular, all the examples considered in that section satisfy Assumption 3.2.

Even so, our analysis can be adapted to incorporate Assumption 3.19. Indeed, consider the sequence $(\boldsymbol{z}_n)_{n \in \mathbb{N}}$ generated by Algorithm 3.3 under Assumption 3.19, and let $(\mu_n)_{n \in \mathbb{N}}$ be such that

$$(\forall n \in \mathbb{N}) \quad \mu_n \le \frac{\langle \boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n \mid \boldsymbol{y}_n - \boldsymbol{x}_n \rangle - \frac{1}{4\beta} \|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{P}}^2}{\|\boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n\|_{\boldsymbol{S}^{-1}}^2}. \tag{3.56}$$

Note that, when $\gamma_n \equiv \theta_n \mu_n$, Algorithm 3.3 yields a relaxed and inertial version of the recurrence in (1.2).

We now proceed to modify the proof of Proposition 3.5 accordingly. In particular, equations (3.9) and (3.12) remain valid. On the other hand, by the cocoercivity of $\boldsymbol{C}$ with respect

to $\boldsymbol{P}$, (3.11) is modified to

$$-2\gamma_n\langle \boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z} \mid \boldsymbol{x}_n - \boldsymbol{z}\rangle = -2\gamma_n\langle \boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z} \mid \boldsymbol{y}_n - \boldsymbol{z}\rangle - 2\gamma_n\langle \boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z} \mid \boldsymbol{x}_n - \boldsymbol{y}_n\rangle$$

$$\leq -\frac{\gamma_n}{\varepsilon}\left(2\beta\varepsilon - \gamma_n\right)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{P}^{-1}}^2 + \varepsilon\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{P}}^2. \qquad (3.57)$$

Moreover, in view of (3.56), (3.12) is modified to

$$2\gamma_n\langle \boldsymbol{M}_n\boldsymbol{y}_n - \boldsymbol{M}_n\boldsymbol{x}_n \mid \boldsymbol{x}_n - \boldsymbol{z}\rangle - 2\gamma_n\langle \boldsymbol{y}_n - \boldsymbol{z} \mid \boldsymbol{M}_n\boldsymbol{y}_n - \boldsymbol{M}_n\boldsymbol{x}_n\rangle + \gamma_n^2\|\boldsymbol{S}^{-1}(\boldsymbol{M}_n\boldsymbol{y}_n - \boldsymbol{M}_n\boldsymbol{x}_n)\|_{\boldsymbol{S}}^2$$

$$= 2\gamma_n\langle \boldsymbol{M}_n\boldsymbol{y}_n - \boldsymbol{M}_n\boldsymbol{x}_n \mid \boldsymbol{x}_n - \boldsymbol{y}_n\rangle + \gamma_n^2\|\boldsymbol{S}^{-1}(\boldsymbol{M}_n\boldsymbol{y}_n - \boldsymbol{M}_n\boldsymbol{x}_n)\|_{\boldsymbol{S}}^2$$

$$\leq \left(\frac{\gamma_n^2}{\mu_n} - 2\gamma_n\right)\langle \boldsymbol{M}_n\boldsymbol{y}_n - \boldsymbol{M}_n\boldsymbol{x}_n \mid \boldsymbol{y}_n - \boldsymbol{x}_n\rangle - \frac{\gamma_n^2}{4\beta\mu_n}\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{P}}^2. \qquad (3.58)$$

Then, by combining (3.9), (3.12), (3.57), and (3.58), we deduce

$$\|\boldsymbol{w}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 \leq \|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \left(\frac{\gamma_n^2}{4\beta\mu_n} - \varepsilon\right)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{P}}^2 - \frac{\gamma_n}{\varepsilon}\left(2\beta\varepsilon - \gamma_n\right)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{P}^{-1}}^2$$

$$+ \left(\frac{\gamma_n^2}{\mu_n} - 2\gamma_n\right)\langle \boldsymbol{M}_n\boldsymbol{y}_n - \boldsymbol{M}_n\boldsymbol{x}_n \mid \boldsymbol{y}_n - \boldsymbol{x}_n\rangle. \qquad (3.59)$$

Furthermore, if

$$\frac{\gamma_n^2}{\mu_n} - 2\gamma_n \leq 0, \qquad (3.60)$$

it follows from the 1-strongly monotonicity of $\boldsymbol{M}_n$ that

$$\|\boldsymbol{w}_{n+1} - \boldsymbol{z}\|_{\boldsymbol{S}}^2 \leq \|\boldsymbol{y}_n - \boldsymbol{z}\|_{\boldsymbol{S}}^2 - \left(\frac{\gamma_n^2}{4\beta\mu_n} - \varepsilon - \frac{\gamma_n^2}{\mu_n} + 2\gamma_n\right)\|\boldsymbol{y}_n - \boldsymbol{x}_n\|_{\boldsymbol{P}}^2$$

$$- \frac{\gamma_n}{\varepsilon}\left(2\beta\varepsilon - \gamma_n\right)\|\boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{C}\boldsymbol{z}\|_{\boldsymbol{P}^{-1}}^2. \qquad (3.61)$$

Hence, in view of (3.61), if

$$(\forall n \in \mathbb{N}) \quad \left(\frac{\gamma_n^2}{4\beta\mu_n} - \varepsilon - \frac{\gamma_n^2}{\mu_n} + 2\gamma_n\right) \geq \epsilon \quad \text{and} \quad 2\beta\varepsilon - \gamma_n \geq 0, \qquad (3.62)$$

for some $\epsilon > 0$, it is possible to derive a similar result to the one presented in Proposition 3.5 and consequently establish the convergence of the Algorithm 3.3 under Assumption 3.19 for adequate relaxation and inertial parameters. Note that (3.60) is equivalent to $\gamma_n \leq 2\mu_n$ and it holds when $\gamma_n = \theta_n\mu_n$ and $\theta_n \in]0, 2[$, which are the conditions guaranteeing convergence in [46].

4. Particular cases of the inertial and relaxed Nonlinear-Forward-Backward

In this section, we derive several algorithms with relaxation and inertial steps as specific instances of Algorithm 3.3. In particular, we introduce a new version of FPDHF incorporating both features, which is a novel contribution to the literature. As a consequence, we derive the convergence of relaxed and inertial version of FB, FBF, and FBHF. To the best of the authors' knowledge, there are no existing inertial and relaxed versions of FBHF, being also a new contribution. To this end, we introduce the following problem.

Problem 4.1. *Let $(\mathcal{H}, \langle \cdot \mid \cdot \rangle)$ and $(\mathcal{G}, \langle \cdot \mid \cdot \rangle)$ be real Hilbert spaces, let $A : \mathcal{H} \to 2^{\mathcal{H}}$ and $B : \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone operators, let $L : \mathcal{H} \to \mathcal{G}$ be a linear bounded operator, let*

$D : \mathcal{H} \to \mathcal{H}$ be a ζ-Lipschitzian and monotone operator for $\zeta \in \,]0, +\infty[$, and let $C : \mathcal{H} \to \mathcal{H}$ be a β-cocoercive operator for $\beta \in \,]0, +\infty[$. The problem is to

$$\text{find } (x, u) \in \mathcal{H} \times \mathcal{G} \text{ such that } \begin{cases} 0 & \in (A + C + D)x + L^*u \\ 0 & \in B^{-1}u - Lx, \end{cases} \tag{4.1}$$

under the hypothesis that its solution set, denoted by $\boldsymbol{Z}$, is nonempty.

Note that, given $(\hat{x}, \hat{u}) \in \boldsymbol{Z}$, $\hat{x}$ is a solution to the primal monotone inclusion

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in (A + L^*BL + C + D)x \tag{4.2}$$

and $\hat{u}$ is a solution to the dual monotone inclusion

$$\text{find } u \in \mathcal{G} \text{ such that } 0 \in B^{-1}u - L(A + C + D)^{-1}(-L^*u). \tag{4.3}$$

This inclusion problem has also been studied in multivariate settings and in the context of parallel sums, for example, in [4, 27, 33, 36]. We propose the following inertial and relaxed version of FPDHF, presented in [58, Algorithm 2.3] for solving Problem 4.1.

Algorithm 4.2. *In the context of Problem 4.1, let $(z_0, u_0) \in \mathcal{H} \times \mathcal{G}$, let $(z_{-1}, u_{-1}) \in \mathcal{H} \times \mathcal{G}$, let $(\sigma, \tau) \in \,]0, +\infty[^2$, let $(\alpha_n)_{n \in \mathbb{N}}$ be a nonnegative sequence such that $\alpha_n \to \alpha \in [0, 1[$, let $\lambda_n \in \,]0, +\infty[$, and consider the iteration:*

$$(\forall n \in \mathbb{N}) \quad \left| \begin{array}{l} (p_n, q_n) = (z_n, u_n) + \alpha_n(z_n - z_{n-1}, u_n - u_{n-1}) \\ x_n = J_{\tau A}(p_n - \tau(L^*q_n + Dp_n + Cp_n)) \\ w_{n+1} = x_n - \tau(Dx_n - Dp_n) \\ v_{n+1} = J_{\sigma B^{-1}}(q_n + \sigma L(x_n + w_{n+1} - p_n)) \\ (z_{n+1}, u_{n+1}) = \lambda_n(w_{n+1}, v_{n+1}) + (1 - \lambda_n)(p_n, q_n). \end{array} \right. \tag{4.4}$$

Theorem 4.3. *In the context of Problem 4.1, let $(\sigma, \tau) \in \,]0, +\infty[^2$, let $\varepsilon \in \,]0, 1[$, let $(z_0, u_0) \in \mathcal{H} \times \mathcal{G}$, let $(z_{-1}, u_{-1}) \in \mathcal{H} \times \mathcal{G}$, let $(\alpha_n)_{n \in \mathbb{N}}$ be a nonnegative sequence such that $\alpha_n \to \alpha \in [0, 1[$, let $(\lambda_n)_{n \in \mathbb{N}}$ be a nonnegative sequence, and let $(z_n)_{n \in \mathbb{N}}$ and $(u_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 4.2. Assume that one of the following assertions hold*

(1) $(\alpha_n)_{n \in \mathbb{N}}$ is nondecreasing.
(2) $(\alpha_n)_{n \in \mathbb{N}}$ is decreasing and $\sum_{n \in \mathbb{N}}(\alpha_n - \alpha) < +\infty$.
Define

$$\widetilde{\beta} = \beta(1 - \sigma\tau\|L\|^2), \tag{4.5}$$

$$\widetilde{\zeta} = \tau\zeta/(\sqrt{1 - \sigma\tau\|L\|^2}), \tag{4.6}$$

$$\nu = \begin{cases} 0, & \text{if } T : \mathcal{H} \times \mathcal{G} \to \mathcal{H} \times \mathcal{G} : (x, u) \mapsto \tau(Dx, -\tau LDx) \text{ is monotone,} \\ 2\widetilde{\zeta}, & \text{otherwise,} \end{cases} \tag{4.7}$$

$$\psi = \frac{2 - \varepsilon + \nu}{1 + (\widetilde{\zeta})^2 + \nu}, \tag{4.8}$$

$$(\forall n \in \mathbb{N}) \quad \rho_n = \frac{\psi}{\lambda_n} - 1, \tag{4.9}$$

$$(\forall n \in \mathbb{N}) \quad \delta_n = (1 - \alpha_n)\rho_n - \alpha_{n+1}(1 - \alpha_{n+1})\rho_{n+1} - \alpha_{n+1}(1 + \alpha_{n+1}). \tag{4.10}$$

Assume that $1 - \sigma\tau\|L\|^2 > 0$, $\widetilde{\zeta} < 1$,

$$1 - (\widetilde{\zeta})^2 - \varepsilon > 0, \tag{4.11}$$

$$\tau \le 2\widetilde{\beta}\varepsilon, \tag{4.12}$$

for every $n \in \mathbb{N}$, $\rho_n \geq 0$, and that

$$\liminf \delta_n = \delta > 0. \qquad (4.13)$$

Then, there exists $(x, u) \in \mathbf{Z}$ such that $(z_n, u_n) \rightharpoonup (x, u)$.

Proof. Let $\boldsymbol{\mathcal{H}} = \mathcal{H} \times \mathcal{G}$ and consider the operators:

$$\begin{aligned}
\boldsymbol{A} &: \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} : (x, u) \mapsto (Ax + Dx + L^*u) \times (B^{-1}u - Lx) \\
\boldsymbol{M} &: \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} : (x, u) \mapsto (x/\tau - Dx - L^*u, -Lx + \tau LDx + u/\sigma) \\
\boldsymbol{C} &: \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} : (x, u) \mapsto (Cx, 0) \\
\boldsymbol{S} &: \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} : (x, u) \mapsto (x - \tau L^*u, -\tau Lx + \tau u/\sigma) \\
\boldsymbol{T} &: \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} : (x, u) \mapsto (x/\tau - Dx, u/\tau)
\end{aligned} \qquad (4.14)$$

For every $n \in \mathbb{N}$, let

$$\boldsymbol{M}_n = \boldsymbol{M}, \quad \gamma_n = \tau, \quad \boldsymbol{y}_n = (p_n, q_n), \quad \boldsymbol{x}_n = (x_n, v_{n+1}), \quad \boldsymbol{w}_n = (w_n, v_n), \text{ and } \boldsymbol{z}_n = (z_n, u_n). \qquad (4.15)$$

Note that, $\boldsymbol{M} = \boldsymbol{S} \circ \boldsymbol{T}$, thus $\boldsymbol{S}^{-1} \circ \boldsymbol{M} = \boldsymbol{T}$. Moreover, for every $(x, u) \in \boldsymbol{\mathcal{H}}$ we have,

$$(\boldsymbol{M}_n - \boldsymbol{C})(x, u) = (x/\tau - Dx - Cx - L^*u, \tau LDx - Lx + u/\sigma), \qquad (4.16)$$

and

$$(\boldsymbol{M}_n + \boldsymbol{A})(x, u) = (Ax + x/\tau) \times (B^{-1}u - 2Lx + \tau LDx + u/\sigma). \qquad (4.17)$$

Now, set $(r_n, s_n) = \boldsymbol{r}_n = (\boldsymbol{M}_n - \boldsymbol{C})\boldsymbol{y}_n$, then, in view of (4.16)

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{r}_n = (\boldsymbol{M}_n - \boldsymbol{C})\boldsymbol{y}_n \Leftrightarrow \begin{cases} r_n = p_n/\tau - Dp_n - Cp_n - L^*q_n, \\ s_n = \tau LDp_n - Lp_n + q_n/\sigma. \end{cases} \qquad (4.18)$$

Moreover, it follows from (4.4) that, for every $n \in \mathbb{N}$,

$$\begin{aligned}
\boldsymbol{w}_{n+1} = (w_{n+1}, v_{n+1}) &= (x_n - \tau(Dx_n - Dp_n), v_{n+1}) \\
&= (p_n, q_n) - (p_n - \tau Dp_n, q_n) + (x_n - \tau Dx_n, v_{n+1}) \\
&= \boldsymbol{y}_n - \tau(\boldsymbol{T}\boldsymbol{y}_n - \boldsymbol{T}\boldsymbol{x}_n) \\
&= \boldsymbol{y}_n - \tau(\boldsymbol{S}^{-1}\boldsymbol{M}_n\boldsymbol{y}_n - \boldsymbol{S}^{-1}\boldsymbol{M}_n\boldsymbol{x}_n) \\
&= \boldsymbol{y}_n - \gamma_n \boldsymbol{S}^{-1}(\boldsymbol{M}_n\boldsymbol{y}_n - \boldsymbol{M}_n\boldsymbol{x}_n). \qquad (4.19)
\end{aligned}$$

Hence, by (4.4), (4.18), (4.19), and (4.17), we deduce

$$\begin{cases} x_n = J_{\tau A}(p_n - \tau(L^*q_n + Dp_n + Cp_n)) \\ v_{n+1} = J_{\sigma B^{-1}}(q_n + \sigma L(x_n + w_{n+1} - p_n)) \end{cases}$$

$$\Leftrightarrow \begin{cases} x_n = J_{\tau A}(p_n - \tau(L^*q_n + Dp_n + Cp_n)) \\ v_{n+1} = J_{\sigma B^{-1}}(q_n + \sigma L(2x_n - \tau Dx_n + \tau Dp_n - p_n)) \end{cases}$$

$$\Leftrightarrow \begin{cases} x_n = J_{\tau A}(\tau r_n) \\ v_{n+1} = J_{\sigma B^{-1}}(\sigma(s_n + 2Lx_n - \tau LDx_n)) \end{cases}$$

$$\Leftrightarrow \begin{cases} r_n \in Ax_n + x_n/\tau \\ s_n \in B^{-1}v_{n+1} - 2Lx_n + \tau LDx_n + v_{n+1}/\sigma \end{cases}$$

$$\Leftrightarrow \boldsymbol{r}_n \in (\boldsymbol{M}_n + \boldsymbol{A})\boldsymbol{x}_n$$

$$\Leftrightarrow \boldsymbol{x}_n = (\boldsymbol{M}_n + \boldsymbol{A})^{-1}\boldsymbol{r}_n.$$

Therefore, we deduce that, (4.4), can be written as

$$
\begin{aligned}
\boldsymbol{y}_n &= \boldsymbol{z}_n + \alpha_n(\boldsymbol{z}_n - \boldsymbol{z}_{n-1}), \\
\boldsymbol{x}_n &= (\boldsymbol{M}_n + \boldsymbol{A})^{-1}(\boldsymbol{M}_n - \boldsymbol{C})\boldsymbol{y}_n \\
\boldsymbol{w}_{n+1} &= \boldsymbol{y}_n - \gamma_n \boldsymbol{S}^{-1}(\boldsymbol{M}_n \boldsymbol{y}_n - \boldsymbol{M}_n \boldsymbol{x}_n), \\
\boldsymbol{z}_{n+1} &= \lambda_n \boldsymbol{w}_{n+1} + (1 - \lambda_n)\boldsymbol{y}_n.
\end{aligned}
$$

Thus, (4.4) is a particular instance of Algorithm 3.3. Now, note that, in view of the Lipschitzian property of D and [54, Lemma 6.1], we have for every $\boldsymbol{z} = (x, u) \in \mathcal{H}$ and every $\boldsymbol{x} = (y, v) \in \mathcal{H}$

$$
\begin{aligned}
\|(\gamma_n \boldsymbol{M}_n - \boldsymbol{S})\boldsymbol{z} - (\gamma_n \boldsymbol{M}_n - \boldsymbol{S})\boldsymbol{x}\|_{\boldsymbol{S}^{-1}}^2 &= \|\boldsymbol{S}^{-1}(\gamma_n \boldsymbol{M}_n - \boldsymbol{S})\boldsymbol{z} - \boldsymbol{S}^{-1}(\gamma_n \boldsymbol{M}_n - \boldsymbol{S})\boldsymbol{x}\|_{\boldsymbol{S}}^2 \\
&= \|(\gamma_n \boldsymbol{T} - \mathrm{Id})\boldsymbol{z} - (\gamma_n \boldsymbol{T} - \mathrm{Id})\boldsymbol{x}\|_{\boldsymbol{S}}^2 \\
&= \langle \tau(Dx - Dy, 0) \mid \tau \boldsymbol{S}(Dx - Dy, 0) \rangle \\
&= \langle \tau(Dx - Dy, 0) \mid \tau(Dx - Dy, \tau L(Dy - Dx)) \rangle \\
&= \tau^2 \|Dx - Dy\|^2 \\
&\leq \tau^2 \zeta^2 \|x - y\|^2 \\
&\leq \frac{\tau^2 \zeta^2}{1 - \tau\sigma\|L\|^2} \|(x - y, u - v))\|_{\boldsymbol{S}}^2 \\
&= \frac{\tau^2 \zeta^2}{1 - \tau\sigma\|L\|^2} \|\boldsymbol{z} - \boldsymbol{x}\|_{\boldsymbol{S}}^2.
\end{aligned}
$$

Then, it follows that $\gamma_n \boldsymbol{M}_n - \boldsymbol{S}$ is $\widetilde{\zeta}$-Lipschitz with respect to $\boldsymbol{S}$. Moreover, $\boldsymbol{C}$ is $\widetilde{\beta}$-cocoercive with respect to $\boldsymbol{S}$ [54, Corollary 6.1]. Note that, in this case $-(\gamma_n \boldsymbol{M}_n - \boldsymbol{S})\boldsymbol{z} = \tau(Dx, -\tau LDx) = T(x, u)$, then $\nu_n \equiv \nu$. Hence, δ_n defined in (4.10) coincide with (3.6). Altogether, the result follows by invoking Theorem 3.7. □

Remark 4.4.   *(1)* In view of Theorem 3.11, if the operator $\boldsymbol{A}$, defined in (4.14), is strongly monotone, Algorithm 4.2 converges linearly to the unique solution of Problem 4.1. This occurs, for instance, when A is strongly monotone and B is cocoercive.

(2) *In the case where $\alpha_n \equiv 0$ and $\lambda_n \equiv 1$, Algorithm 4.2 reduces to the FPDHF algorithm [58, Algorithm 2.3]. In this case, (4.13) reduces to*

$$
\delta = \rho = \psi - 1 = \frac{2 - \varepsilon + \nu}{1 + \nu + (\widetilde{\zeta})^2} - 1 > 0 \Leftrightarrow 1 - \varepsilon - (\widetilde{\zeta})^2 > 0.
$$

Moreover, by considering $\varepsilon = \dfrac{\epsilon}{1 - \sigma\tau\|L\|^2}$, with $\epsilon > 0$, we have

$$
1 - \varepsilon - (\widetilde{\zeta})^2 > 0 \Leftrightarrow 1 - \epsilon > \sigma\tau\|L\|^2 + \tau^2\zeta^2, \tag{4.20}
$$

and (4.12) reduces to

$$
\tau \leq 2\widetilde{\beta}\varepsilon \Leftrightarrow \tau \leq 2\beta\epsilon. \tag{4.21}
$$

Note that (4.20) and (4.21) are the conditions on the step-size for guaranteeing the convergence of FPDHF in [58, Theorem 4.2]. This shows that FPDHF is a particular case of the Nonlinear Forward-Backward algorithm, which has not been previously established.

(3) *Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, $L\colon \mathcal{H} \to \mathcal{G}$ be a linear bounded operator, $h\colon \mathcal{H} \to \mathbb{R}$ be a differentiable convex function with a ζ-Lipschitz gradient for some $\zeta \in \,]0, +\infty[$, and let*

$d\colon \mathcal{G} \to \mathbb{R}$ *be a differentiable convex function with a* $1/\beta$-*Lipschitz gradient for some* $\beta \in]0, +\infty[$. *Consider the optimization problem,*

$$\min_{x \in \mathcal{H}} f(x) + g(Lx) + h(x) + d(x), \tag{4.22}$$

under the assumption that its solution set is nonempty. Under standard qualification conditions (see [10, *Proposition 6.19]) by considering the Fermat's rule* [10, *Theorem 16.3] and subdifferential calculus rules* [10, *Theorem 16.47], and setting* $A = \partial f$, $B = \partial g$, $C = \nabla d$, *and* $D = \nabla h$, *the problem in* (4.22) *can be solved by Algorithm 4.2, which, in this setting, iterates as follows:*

$$(\forall n \in \mathbb{N}) \quad \left| \begin{array}{l} (p_n, q_n) = (z_n, u_n) + \alpha_n(z_n - z_{n-1}, u_n - u_{n-1}) \\ x_n = \operatorname{prox}_{\tau f}(p_n - \tau(L^* q_n + \nabla h(p_n) + \nabla d(p_n))) \\ w_{n+1} = x_n - \tau(\nabla h(x_n) - \nabla h(p_n)) \\ v_{n+1} = \operatorname{prox}_{\sigma g^*}(q_n + \sigma L(x_n + w_{n+1} - p_n)) \\ (z_{n+1}, u_{n+1}) = \lambda_n(w_{n+1}, v_{n+1}) + (1 - \lambda_n)(p_n, q_n). \end{array} \right. \tag{4.23}$$

To initialize Algorithm 4.2, the parameters ε, σ, and τ and the sequences $(\alpha_n)_{n \in \mathbb{N}}$ and $(\lambda_n)_{n \in \mathbb{N}}$ must be chosen to satisfy conditions (4.11)–(4.13). Since this selection may not be straightforward, we propose an initialization strategy. This scheme depends on $\overline{\varepsilon} \in]0, +\infty[$ and $\chi \in]0, +\infty[$, which are specified in (4.24) and (4.30), respectively. To implement the algorithm, the user only needs to choose $(t, \kappa_1, \kappa_2) \in]0, 1[^3$. The constant t provides some flexibility in the step-size and the relaxation-inertial parameters. The constants κ_1 and κ_2 regulate the step-sizes τ and σ, respectively. A detailed explanation of each step of the initialization, can be found below.

Initialization 4.5. In the context of Problem 4.1, let $(\alpha_n)_{n \in \mathbb{N}}$ and $(\lambda_n)_{n \in \mathbb{N}}$ be nonnegative sequences. Suppose that $\alpha_n \to \alpha$ and $\lambda_n \to \lambda$ and consider the following initialization procedure for Algorithm 4.2.

(1) Choose $t \in]0, 1]$, and set $\varepsilon = t \cdot \overline{\varepsilon}$, where

$$\overline{\varepsilon} = \frac{1 - \sqrt{1 + 16\beta^2 \zeta^2}}{8\beta^2 \zeta^2} = \frac{2}{1 + \sqrt{1 + 16\beta^2 \zeta^2}}. \tag{4.24}$$

Compute $\chi \in]0, +\infty[$ as

$$\chi = \frac{4\beta}{1 + \sqrt{1 + 16\beta^2 \zeta^2}}.$$

(2) Choose $\kappa_1 \in]0, 1[$ and set $\tau = \kappa_1 \chi$.
(3) Choose $\kappa_2 \in]0, 1[$ and set $\sigma = \frac{\kappa_2}{\tau \|L\|^2}\left(1 - \frac{\tau}{\chi}\right)$.
(4) Let $\widetilde{\zeta}$, ν, ψ be defined as in (4.6), (4.7), and (4.8), respectively.
(5) Select one of the following scenarios:
   • Choose $\alpha \in [0, 1[$ and λ according (3.47), thus

$$\lambda \in \left]0, \frac{(1 - \alpha)^2}{2\alpha^2 - \alpha + 1} \cdot \psi\right[. \tag{4.25}$$

   • Choose $\lambda \in]0, \psi[$ and α according (3.48), thus

$$\alpha \in \left[0, \overline{\alpha}(\widetilde{\zeta}, \varepsilon, \lambda)\right[. \tag{4.26}$$

Now, we provide explicit values for $\overline{\varepsilon}$ and χ to be considered in step (1) in Initialization 4.5 and an explanation of each step on it.

First, note that (4.11) is equivalent to

$$1 - \varepsilon - \frac{\tau^2 \zeta^2}{1 - \sigma\tau\|L\|^2} > 0 \Leftrightarrow \frac{1 - \varepsilon - \tau^2\zeta^2}{(1-\varepsilon)\tau\|L\|^2} > \sigma \tag{4.27}$$

and (4.12) is equivalent to

$$\tau \le 2\beta(1 - \sigma\tau\|L\|^2)\varepsilon \Leftrightarrow \frac{1}{\tau\|L\|^2}\left(1 - \frac{\tau}{2\beta\varepsilon}\right) \ge \sigma. \tag{4.28}$$

Hence, in order to guarantee the existence of $\sigma > 0$ satisfying both (4.27) and (4.28), it is necessary that

$$1 - \varepsilon - \tau^2\zeta^2 > 0 \text{ and } 1 - \tau/(2\beta\varepsilon) > 0 \Leftrightarrow \frac{\sqrt{1-\varepsilon}}{\zeta} \ge \tau \text{ and } 2\beta\varepsilon > \tau. \tag{4.29}$$

To allow larger values of τ, following a similar approach to that in [22, Proposition 2.1.3], we choose ε such that $\sqrt{1-\varepsilon}/\zeta = 2\beta\varepsilon$ which yields (4.24). Therefore, (4.29) holds if

$$\tau \in]0, \chi[, \quad \text{where} \quad \chi = \frac{\sqrt{1-\overline{\varepsilon}}}{\zeta} = 2\beta\overline{\varepsilon} = \frac{4\beta}{1 + \sqrt{1 + 16\beta^2\zeta^2}}, \tag{4.30}$$

which yield the constants on (1) in Initialization 4.5. Hence, we have $\tau \in]0, \chi[$, which yields the step (2) in Initialization 4.5. In order to chose σ such that (4.27) and (4.28) hold, note that $1 - \overline{\varepsilon} = \chi^2\zeta^2$, thus, (4.27) is equivalent to $(\chi^2 - \tau^2)(\chi^2\tau\|L\|^2)^{-1} > \sigma$. Furthermore, since $2\beta\overline{\varepsilon} = \chi$, (4.28) is equivalent to $(1 - \tau/\chi)(\tau\|L\|^2)^{-1} \ge \sigma$. Hence, since $\tau < \chi$, we have

$$\frac{\chi^2 - \tau^2}{\chi^2}\frac{1}{\tau\|L\|^2} > \left(1 - \frac{\tau}{\chi}\right)\frac{1}{\tau\|L\|^2}. \tag{4.31}$$

Then, both (4.27) and (4.28) hold if

$$\sigma \in \left]0, \frac{1}{\tau\|L\|^2}\left(1 - \frac{\tau}{\chi}\right)\right[, \tag{4.32}$$

which yields the step (3) in Initialization 4.5. Thus, we have determined $\overline{\varepsilon}$ and χ giving the largest interval for τ and σ such that (4.11) and (4.12) are feasible. Step (4) and (5) in Initialization 4.5 guarantee that (4.13) holds in view of (3.47) and (3.48).

4.1. **Particular instances of Algorithm 4.2.** In this section, we present particular instances of Algorithm 4.2 that are of interest in their own right, as they recover and extend inertial and relaxed versions of relevant methods from the literature. In addition, we provide the values of $\overline{\varepsilon}$ and χ required to apply Initialization 4.5 to these methods.

In the context of Problem 4.1, let $(\alpha_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ be nonnegative sequences such that $\alpha_n \to \alpha$ and $\lambda_n \to \lambda$. Consider the following cases:

- **Condat–Vũ**: In the case where $D = 0$, Algorithm 4.2 reduces to the relaxed and inertial Condat–Vũ algorithm. In this case $\zeta = \nu = \widetilde{\zeta} = 0$, $\rho = \frac{2-\varepsilon}{\lambda} - 1$. Then, by considering $\varepsilon = \tau/(2\widetilde{\beta})$ we have that (4.12) holds and (4.11) reduces to

$$\sigma\tau\|L\|^2 + \frac{\tau}{2\beta} < 1, \tag{4.33}$$

  which corresponds with the condition in [39, Theorem 3.1(i)]. By taking $\chi = 2\beta$, $\tau \in]0, 2\beta[$, and $\sigma \in \left]0, \frac{1}{\tau\|L\|^2}\left(1 - \frac{\tau}{2\beta}\right)\right[$, it follows that (4.29) holds directly, thus,

it can be considered $\bar{\varepsilon} = 1$ in Initialization 4.5. Moreover, according (4.8) we have $\psi = 2 - \tau(2\beta(1 - \tau\sigma\|L\|^2))^{-1}$. Then, if $\alpha = 0$, (4.25) holds if

$$\lambda \in \left]0, 2 - \frac{\tau}{2\beta(1 - \sigma\tau\|L\|^2)}\right[, \tag{4.34}$$

which corresponds to the condition in [39, Theorem 3.1(ii)].

- **Chambolle–Pock**: If $C = D = 0$, Algorithm 4.2 reduces to the relaxed and inertial Chambolle–Pock algorithm (see [53, Equation (38)]). In this case, by taking $\zeta = 0$ and $\beta \to +\infty$, and according to (4.29), we can choose $\bar{\varepsilon} = 0$ in Initialization 4.5. Moreover, χ can be chosen arbitrarily in $]0, +\infty[$, and the only condition to be met is $\sigma\tau\|L\|^2 < 1$. Additionally, from (4.8), we have $\psi \to 2$. Hence, (4.25) is equivalent to

$$\frac{2(1 - \alpha)^2}{(1 - \alpha)^2 + \alpha(1 + \alpha)} > \lambda. \tag{4.35}$$

Notice that, if $\alpha = 0$, (4.35) holds for $\lambda \in]0, 2[$, and if $\lambda = 1$, it holds for $\alpha \in]0, 1/3[$.

- **Chambolle–Pock combined with FBF**: In the case where $C = 0$, Algorithm 4.2 is a relaxed and inertial algorithm combining Chambolle–Pock and FBF (see [58, Eq. (44)]). If we take $\beta \to +\infty$, in view of (4.12), we have $\varepsilon \to 0$. Hence, it can be considered $\bar{\varepsilon} = 0$ and $\chi = 1/\zeta$ in Initialization 4.5 (see (4.30)).

- **Forward-Backward-Half-Forward**: In the case where $B = 0$ and $L = 0$ Algorithm 4.2 iterates as follows:

$$(\forall n \in \mathbb{N}) \quad \left|\begin{array}{l} p_n = z_n + \alpha_n(z_n - z_{n-1}) \\ x_n = J_{\tau A}(p_n - \tau(Dp_n + Cp_n)) \\ w_{n+1} = x_n - \tau(Dx_n - Dp_n) \\ z_{n+1} = \lambda w_{n+1} + (1 - \lambda)p_n. \end{array}\right. \tag{4.36}$$

This recurrence is a relaxed and inertial version of FBHF. Initialization 4.5 can be considered in this case with $\sigma = 0$ ($\kappa_2 = 0$) and $\bar{\varepsilon}$ and χ as in (4.24) and (4.30), respectively. Moreover, by (4.30), the step-size τ must satisfy the condition

$$\tau \in \left]0, \frac{4\beta}{1 + \sqrt{1 + 16\beta^2\zeta^2}}\right[, \tag{4.37}$$

which corresponds to the step-size condition for FBHF proposed in [22, Theorem 2.3].

Note that the step-size requirements for FBHF were more easily derived, even with inertial and relaxation steps, than those obtained in [46, Appendix A.3]. This highlights that Assumption 3.2 simplifies the process of obtaining step-size constraints and, consequently, the practical implementation of the algorithm.

To our knowledge, no previous work in the existing literature has presented inertial or relaxed versions of FBHF.

Note that, in this setting, the operator $-(\gamma_n \boldsymbol{M}_n - \boldsymbol{S}) = \tau(D, 0)$ is monotone, thus, we can consider $\nu = 0$.

- **Forward-Backward-Forward**: In the case where $B = 0$, $L = 0$, and $C = 0$, Algorithm 4.4 reduces to the inertial and relaxed FBF with constant step-sizes studied in [18]. If we take $\beta \to +\infty$, in view of (4.24) and (4.30), we have $\bar{\varepsilon} \to 0$ and $\chi \to 1/\zeta$, which values can be considered in Initialization 4.5 together with $\sigma = 0$ ($\kappa_2 = 0$). In this setting we have $-(\gamma_n \boldsymbol{M}_n - \boldsymbol{S}) = \tau D$, which is a monotone operator. Then $\nu = 0$ and $\psi = 2/(1 + (\tau\zeta)^2)$ (see (4.8)). Therefore (4.25) is equivalent to

$$\frac{2(1 - \alpha)^2}{((1 - \alpha)^2 + \alpha(1 + \alpha))(1 + \zeta^2\tau^2)} > \lambda. \tag{4.38}$$

Note that this condition is slightly less restrictive than the condition in [18, Equation (25)], where ν was taken as $2\zeta\tau$. Indeed, since $\zeta\tau < 1$,

$$\frac{2(1-\alpha)^2}{((1-\alpha)^2 + \alpha(1+\alpha))(1+\zeta^2\tau^2)} > \frac{2(1-\alpha)^2}{((1-\alpha)^2 + \alpha(1+\alpha))(1+\zeta\tau)}.$$

5. Numerical simulations.

In this section, we present three numerical experiments for testing the inclusion of inertial and relaxation steps in NFB[1]. In particular, we implement FBHF and FPDHF for solving optimization problems with affine constraints and image restoration tasks.

5.1. **Optimization with affine constraints.** Let $(N,m,p) \in \mathbb{N}^3$, $M \in \mathbb{R}^{m \times N}$, $b \in \mathbb{R}^m$, $s_1, \ldots, s_p \in \mathbb{R}^N$, and $\mathcal{S} = \{x \in \mathbb{R}^p \mid (\forall i \in \{1,\ldots,p\}) \quad s_i^\top x \leq 0\}$. Consider the following optimization problem presented in [22, Section 7.1]

$$\min_{x \in \mathcal{S}} \iota_{[0,1]^N}(x) + \frac{1}{2}\|Mx - b\|^2. \tag{5.1}$$

This problem is equivalent to (see [22, Section 6.3])

$$\text{find} \quad (x,u) \in \mathbb{R}^N \times \mathbb{R}^p, \quad \text{such that } (0,0) \in A(x,u) + C(x,u) + D(x,u), \tag{5.2}$$

where $A = \mathcal{N}_{[0,1]^N} \times \mathcal{N}_{[0,+\infty[^p}$, $C(x,u) = (M^\top(Mx-b),0)$, $D(x,u) = (S^\top u, -Sx)$, and $S = [s_1,\ldots,s_p]^\top$. Note that C is $(\|M\|^{-2})$-cocoercive and D is $\|S\|$-Lipschitz. Hence, (5.2) is a particular instance of Problem 4.1 by considering $B = 0$ and $L = 0$ and it can be solved by FBHF. In view of [10, Proposition 23.18], for $(x,u) \in \mathbb{R}^N \times \mathbb{R}^p$, we have $J_{\tau A}(x,u) = ((\min(\max(x_i,0),1))_{1 \leq i \leq N}, (\max(u_j,0)_{1 \leq j \leq p}))$. We solve this problem repeatedly using FBHF with and without inertial and relaxation steps. We test different values of the relaxation parameters as well as both constant and decreasing inertial sequences. In particular, we consider nine instances of (N,m,p), as presented in Table 2. The implemented inertial and relaxation parameters are listed in Table 1 and Table 2. Note that these parameters depend on $\overline{\alpha} = \overline{\alpha}(\tau\|R\|,\varepsilon,\lambda)$ and $\psi = \psi(\tau\|R\|,\varepsilon)$, defined in (3.48) and (3.46), respectively. The step-size is set to $\tau = 2\varepsilon/\|M\|^2$ for $\varepsilon = t(1 + \sqrt{1+16\|S\|^2/\|M\|^4})^{-1}$ and $t \in [0,1[$. The specific value of t used in each case is given in Table 2. The stopping criterion is a relative error with tolerance 10^{-6} and a maximum of 10^6 iterations. For each of the nine instances of (N,m,p) described in Table 2, we generate 20 random realizations of the matrices M and S using the MATLAB function *randn*. The results expressed as the average number of iteration and average CPU time in seconds, over each group of 20 realizations, are reported in Table 2. From this table, we observe that, in all nine cases, FBHF outperforms both IFBHF (inertial FBHF) and RIFBHF (relaxed and inertial FBHF) in terms of the average number of iterations (IN) and average computation time (T). This indicates that, in general, selecting larger step-sizes yields better numerical performance than prioritizing larger relaxation parameters at the expense of reducing step-sizes. On the other hand, the best instance of DIFBHF (decreasing inertial FBHF) provides better results in eight out of the nine dimension settings. It is worth to mention that the single dimension, namely $(N,m,k) = (2000,500,100)$, where it does not lead to acceleration, only 2 out of 20 random realizations for M and S failed to converge, resulting in a high average value. However, if we consider a sequence of smaller inertial parameters the average results are better for this dimension, as we have shown in Table 3. This highlights the benefit of using decreasing inertial parameters, which allows for larger step-sizes and larger relaxation parameters in the initial iterations. In Figure 2, we plot the residual value, given by

---

[1]All numerical experiments were implemented in MATLAB on a desktop computer equipped with an Intel Core i7-14700K processor (3.4/5.6 GHz), 64 GB of RAM, and running Windows 11 Pro 64-bit. The code is available in this repository
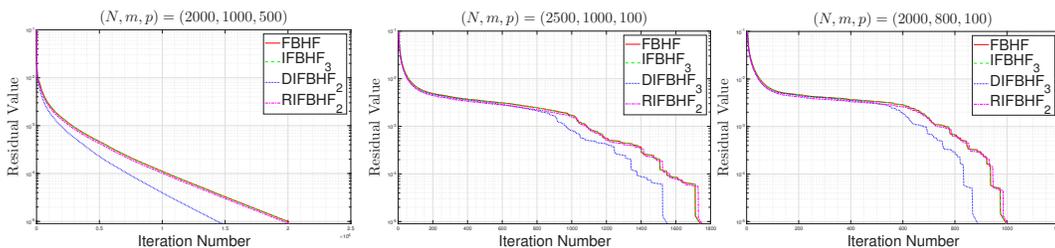
(A) $(N, m, k) = (2000, 1000, 500)$ (B) $(N, m, k) = (2500, 1000, 100)$ (C) $(N, m, k) = (2000, 800, 100)$

FIGURE 2. Residual value ($\|p_n - x_n - \tau(Dp_n + Cp_n) + \tau(Dx_n + Cx_n)\|$) for FBHF and for the best performing instances of IFBHF, DIFBHF, and RIFBHF. The graphs display the results for the first of the 20 random realizations of M and S.

$\|p_n - x_n - \tau(Dp_n + Cp_n) + \tau(Dx_n + Cx_n)\|$ (see (4.36)), versus the iteration number for the best cases of IFBHF, DIFBHF, and RIFBHF, considering three different values of (N, m, p). For each (N, m, p) case, the plot corresponds to the first of the 20 random realizations of M and S. From this figure, we observe that the residual value decreases faster for the decreasing inertial parameters. A drawback of this decreasing scheme is the lack of a general rule to select the sequence of inertial parameters. For instance, in five cases the best performance is obtained with the sequence $(\alpha_n^2)_n$ while in three cases it is achieved with $(\alpha_n^3)_{n \in \mathbb{N}}$. In addition, the sequence $(\alpha_n^1)_n$ consistently yields worse convergence results.

Decreasing Inertia	$\alpha_n^1 = (1 + 0.001 \cdot n \cdot (\log(n))^{1.001})^{-1}$
	$\alpha_n^2 = (3 + 0.00001 \cdot n \cdot (\log(n))^{1.00001})^{-1}$
	$\alpha_n^3 = (9 + 0.00001 \cdot n \cdot (\log(n))^{1.00001})^{-1}$

TABLE 1. Decreasing and summable inertial parameters.

5.2. **Image restoration.** To compare the inertial and relaxed version of FPDHF with its classical version, we consider the image restoration problem proposed in [58, Section 6].

Let $(N, M) \in \mathbb{N}^2$ and let $x \in \mathbb{R}^{N \times N}$ be an image to be recovered from an observation

$$z = Tx + \epsilon, \tag{5.3}$$

where $T \colon \mathbb{R}^{N \times N} \times \mathbb{R}^{M \times M}$ is an operator representing a blur process and ϵ is an additive noise perturbing the observation. The original image x can be approximated by solving the following optimization problem.

$$\min_{x \in [0,1]^{N \times N}} \frac{1}{2}\|Tx - z\|^2 + \mu_1\|\nabla x\|_1 + \mu_2 H_\delta(Wx), \tag{5.4}$$

where $(\mu_1, \mu_2) \in {]0, +\infty[}^2$ are regularization parameters, $\|\cdot\|_1$ is the ℓ^1 norm, and $H_\delta \colon \mathbb{R}^{N \times N} \to \mathbb{R}$ is the Huber function defined for $\delta > 0$ (see [23] for details on the Huber function). The linear operator $\nabla \colon \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times N} \times \mathbb{R}^{N \times N}$ is the discrete gradient with Neumann boundary conditions. Its adjoint ∇^* is the discrete divergence [29]. The operator $W : \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times N}$ is an orthogonal basis wavelet transform (we assume that N is a power of 2 and we choose a Haar basis of level 3). For further details on the model see [58, Section 6]. By setting $f = \iota_{[0,1]^{N \times N}}$, $g = \mu_1\|\cdot\|_1$, $L = \nabla$, $d = \|T(\cdot) - z\|_2^2/2$, and $h = \mu_2(H_\delta \circ W)$, the optimization problem in equation (5.4) is a particular instance of the problem in (4.22) and it can be solved by FPDHF in its

Algorithm	α_n	λ	t	(N,m,p)		(N,m,p)		(N,m,p)	
				IN	T	IN	T	IN	T
				$(2000,1000,100)$		$(2000,1000,500)$		$(2000,1000,800)$	
FBHF	0	1	0.999	16955	2.95	17975	8.02	36452	27.25
IFBHF	$\epsilon \cdot \overline{\alpha}$	1	0.8	17965	3.31	19127	8.50	38634	28.87
		1	0.9	17237	3.16	18292	8.20	37054	27.26
		1	0.999	16943	3.04	17961	8.08	36426	26.67
DIFBHF	α_n^1	1	0.999	26638	4.69	30625	13.65	60074	44.93
	α_n^2	1	0.999	**13252**	**2.34**	**13680**	**5.83**	**29134**	**21.94**
	α_n^3	1	0.999	15667	2.76	16477	7.16	33745	25.18
RIFBHF	$\epsilon \cdot \overline{\alpha}$	$0.95 \cdot \psi$	0.9	17132	3.06	18179	7.98	36839	27.45
		$\epsilon \cdot \psi$	0.9	17078	3.05	18116	8.18	36721	27.35
				$(1500,1000,100)$		$(2500,1000,100)$		$(3500,1000,100)$	
FBHF	0	1	0.999	10853	0.95	1199	0.61	701	1.11
IFBHF	$\epsilon \cdot \overline{\alpha}$	1	0.8	11574	1.00	1270	0.64	782	1.22
		1	0.9	11051	0.95	1205	0.60	732	1.16
		1	0.999	10844	0.93	1198	0.60	700	1.11
DIFBHF	α_n^1	1	0.999	16946	1.44	44556	22.75	98885	153.93
	α_n^2	1	0.999	**8042**	**0.69**	80181	40.96	100000	158.33
	α_n^3	1	0.999	9898	0.86	**1061**	**0.55**	**629**	**0.98**
RIFBHF	$\epsilon \cdot \overline{\alpha}$	$0.95 \cdot \psi$	0.9	10980	0.95	1216	0.61	711	1.13
		$\epsilon \cdot \psi$	0.9	10941	0.95	1211	0.61	708	1.12
				$(2000,500,100)$		$(2000,800,100)$		$(2000,1500,100)$	
FBHF	0	1	0.999	**530**	**0.11**	1212	0.21	16586	3.39
IFBHF	$\epsilon \cdot \overline{\alpha}$	1	0.8	574	0.12	1315	0.24	17679	3.65
		1	0.9	536	0.11	1243	0.22	16887	3.49
		1	0.999	530	0.11	1210	0.21	16573	3.40
DIFBHF	α_n^1	1	0.999	80474	15.98	28818	4.65	24484	5.00
	α_n^2	1	0.999	100000	22.19	80198	14.03	**12511**	**2.52**
	α_n^3	1	0.999	10421	1.68	**1094**	**0.18**	15168	3.10
RIFBHF	$\epsilon \cdot \overline{\alpha}$	$0.95 \cdot \psi$	0.9	538	0.11	1229	0.21	16779	3.42
		$\epsilon \cdot \psi$	0.9	536	0.12	1224	0.21	16720	3.34

TABLE 2. Comparison of FBHF with and without inertial and relaxation steps. The step-size implemented is $\tau = 2\varepsilon/\|M\|^2$ for $\varepsilon = t(1 + \sqrt{1+16\|R\|^2/\|M\|^4})^{-1}$ and $t \in [0,1[$. We set $\epsilon = 0.9999$. The parameters $\overline{\alpha} = \overline{\alpha}(\tau\|R\|, \varepsilon, \lambda)$ and $\psi = \psi(\tau\|R\|, \varepsilon)$ are defined in (3.48) and (3.46), respectively. The decreasing inertial parameters $\alpha_n^1, \alpha_n^2, \alpha_n^3$ are defined in Table 1. The best results obtained in terms of average number of iterations (IN) and average CPU time (T) are highlighted in bold.

variational version (4.23). For $x \in \mathbb{R}^{N \times N}$, we have $\text{prox}_{\tau f}(x) = ((\min(\max(x_{i,j}, 0), 1))_{1 \le i,j \le N}$. Closed formulas for $\text{prox}_{\sigma g^*}$ and ∇h, can be found in [10, Example 24.20] and [32, Section 6.1], respectively. Note that, ∇d is β-cocoercive and ∇h is ζ-Lipschitz, where $\beta = \|T\|^{-2}$ and

Algorithm	α_n	λ	t	IN	T
DIFBHF	$(10 + 0.00001 \cdot n \cdot (\log(n))^{1.00001})^{-1}$	1	0.999	477	0.08

TABLE 3. Result of DIFBHF for a sequence of smaller inertial parameters and $(N, m, k) = (2000, 500, 100)$ in terms of average number of iterations (IN) and average CPU time (T).

$\zeta = \mu_2/\delta$. In the numerical experiments, we consider three scenarios for the operator T being a blur model performed by a 3×3 averaging kernel, 9×9 averaging kernel, and 3×3 with Gaussian kernel. Particularly, T was implemented using the imfilter function in MATLAB with symmetric boundary conditions (we have $\|T\| = 1$ in all cases). The observation z is obtained through (5.3) where ϵ is an additive zero-mean white Gaussian noise with standard deviation 10^{-3} (implemented by the imnoise function in MATLAB). We generate 20 random observations trough T and ϵ for $N \in \{128, 256, 512\}$. We set $\mu_1 = 10^{-2}$, $\mu_2 = 10^{-3}$ and $\delta = 10^{-2}$ in our experiments (see [58, Section 6] for a discussion on these parameters). We consider Initialization 4.5 for running our experiments with $\varepsilon = 2t(1 + \sqrt{1 + \zeta^2\beta^2})^{-1}$, for $t \in]0, 1[$, and χ defined in (4.30). Furthermore, we first tune up the parameter κ_1 to obtain the best result of convergence in classical FPDHF. The implemented parameters κ_1 and κ_2, for each case, are described in Table 4. From this table we observe that the constant inertial version (IFPDHF) and the relaxed-inertial version (RIFBHF) are comparable in terms of iteration number and CPU time to the classical version of FPDHF. Note that, despite RIFBHF present lower average iteration number, it needs more time to complete the iterations, this is attributable to the additional inertial and relaxation step on its iterations. On the other hand, from Table 4, we observe that the decreasing inertial version (DIFPDHF), produce numerical advantages in terms of both, iteration number and CPU time. In particular, the sequence $(\alpha_n^2)_{n \in \mathbb{N}}$ has the best performance reducing over 20% of average iteration number and CPU time in comparison to FPDHF, for the 3×3 averaging kernel and 9×9 averaging kernel. On the other hand, the sequence $(\alpha_n^1)_{n \in \mathbb{N}}$ provides better results for the 3×3 Gaussian Kernel reducing almost 50% of average iteration number and CPU time. However, the lacking of a rule for choosing the inertial parameter is a drawback of this method as we mentioned in previous section.

From Table 4, we observe that the constant inertial version (IFPDHF) and the relaxed-inertial version (RIFPDHF) are comparable to the classical FPDHF in terms of both iteration number and CPU time. Although RIFPDHF achieves slightly lower average iteration numbers, it requires more time to complete the iterations. This increase in computational time can be attributed to the additional inertial and relaxation steps involved in each iteration. On the other hand, Table 4 clearly shows that using decreasing inertial sequences provides numerical advantages in both iteration number and CPU time. In particular, the sequence $(\alpha_n^2)_{n \in \mathbb{N}}$ yields the best overall performance, achieving more than a 20% reduction in average iteration number and CPU time compared to FPDHF for both the 3×3 and 9×9 averaging kernels. Moreover, the sequence $(\alpha_n^1)_{n \in \mathbb{N}}$ performs best in the case of the 3×3 Gaussian kernel, with reductions of nearly 50% in both metrics. In Figure 2, we plot the residual value, given by $\|(p_n - w_{n+1} - \tau(L^*(q_n - v_{n+1}) + Cp_n - Cx_n), q_n - v_{n+1} + \sigma(L(w_{n+1} - p_n)))\|$ (see (4.4)), versus the iteration number for the best cases of IFPDHF, DIFPDHF, and RIFPDHF, considering three different blur operators and dimensions. In each case, the plot corresponds to the first of the 20 random realizations of z. From Figures 3a and 3b, we observe that the residual value decreases faster for the decreasing inertial parameters. In Figure 3c, we observe that the decreasing inertial parameters produce a slower decay of the residual value during the first iterations; however, the convergence accelerates in the final iterations, reaching the stopping criterion earlier. As discussed in the previous section, a drawback of using decreasing inertial
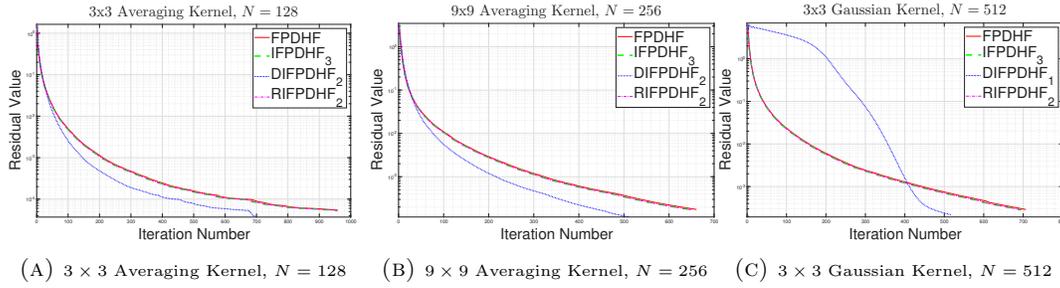
(A) 3×3 Averaging Kernel, $N = 128$ (B) 9×9 Averaging Kernel, $N = 256$ (C) 3×3 Gaussian Kernel, $N = 512$

FIGURE 3. Residual value ($\|(p_n - w_{n+1} - \tau(L^*(q_n - v_{n+1}) + Cp_n - Cx_n), q_n - v_{n+1} + \sigma(L(w_{n+1} - p_n)))\|$) for FPDHF and for the best performing instances of IFPDHF, DIFPDHF, and RIF-PDHF. The graphs display the results for the first of the 20 random realizations of z.
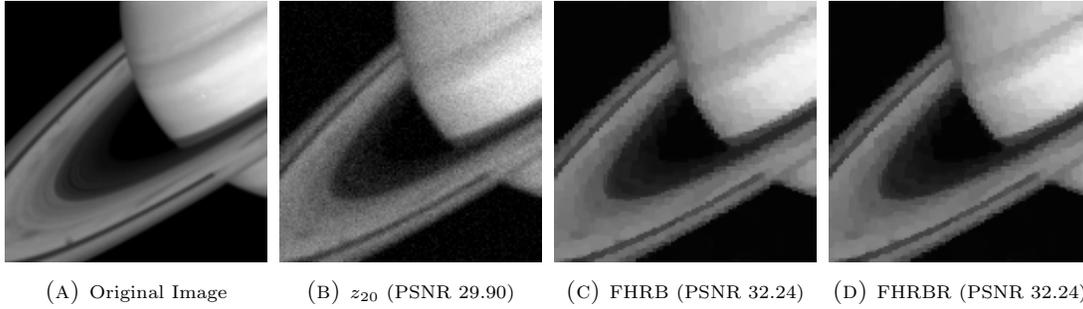

(A) Original Image   (B) z_{20} (PSNR 29.90)   (C) FHRB (PSNR 32.24)   (D) FHRBR (PSNR 32.24)

FIGURE 4. Original image, blurred and noisy observation # 20 with a blur of 3×3 averaging kernel, and recovered images for FPDHF and DIFPDHF with $(\alpha_n^2)_{n \in \mathbb{N}}$, with their respective PSNR (dB), for $N = 128$.

(A) Original Image   (B) z_{20} (PSNR 24.56)   (C) FHRB (PSNR 27.94)   (D) FHRBR (PSNR 27.94)
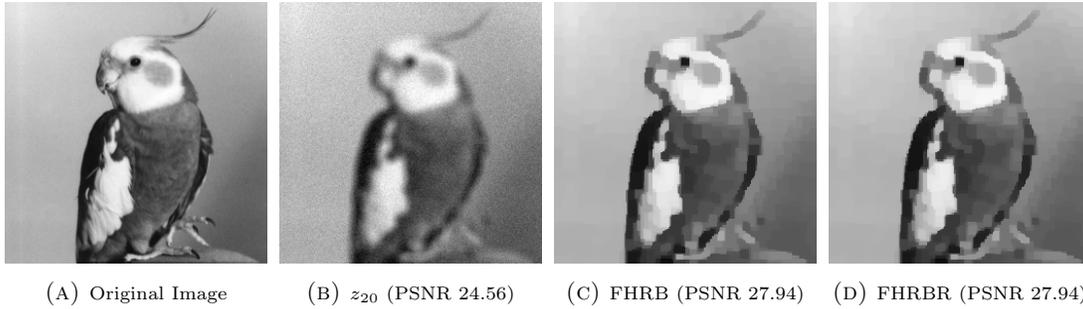
FIGURE 5. Original image, blurred and noisy observation # 20 with a blur of 9×9 averaging kernel, and recovered images for FPDHF and DIFPDHF with $(\alpha_n^2)_{n \in \mathbb{N}}$, with their respective PSNR (dB), for $N = 256$.

parameters is the lack of a general rule for choosing the sequence. The original, blurred, and noisy images are shown in Figures 4–6, along with the recovered images for specific instances described in the corresponding captions.

Algorithm	α_n	λ	t	(N,κ_1,κ_2) IN	T	(N,κ_1,κ_2) IN	T	(N,κ_1,κ_2) IN	T
3×3 Averaging Kernel				(128,0.17,0.99)		(256,0.24,0.99)		(512,0.31,0.99)	
FPDHF	0	1	0.999	865	0.87	481	1.51	494	14.81
IFPDHF	$\epsilon \cdot \overline{\alpha}$	1	0.8	901	0.91	614	1.96	592	17.55
		1	0.9	797	0.80	502	1.60	493	14.63
		1	0.999	852	0.86	474	1.51	488	14.52
DIFPDHF	α_n^1	1	0.999	818	0.82	469	1.50	507	15.08
	α_n^2	1	0.999	**691**	**0.70**	**368**	**1.18**	**385**	**11.44**
	α_n^3	1	0.999	808	0.81	445	1.43	459	13.64
RIFPDHF	$\epsilon \cdot \overline{\alpha}$	$0.95 \cdot \psi$	0.95	855	0.89	476	1.57	490	15.43
		$\epsilon \cdot \psi$	0.999	852	0.88	474	1.56	488	15.36
9×9 Averaging Kernel				(128,0.29,0.99)		(256,0.52,0.99)		(512,0.59,0.99)	
FPDHF	0	1	0.999	2400	2.58	1395	4.45	1522	45.54
IFPDHF	$\epsilon \cdot \overline{\alpha}$	1	0.8	2991	3.24	1705	5.54	1834	54.41
		1	0.9	2502	2.72	1442	4.68	1558	46.18
		1	0.999	2369	2.57	1384	4.49	1511	44.82
DIFPDHF	α_n^1	1	0.999	1956	2.12	1276	4.15	1399	41.51
	α_n^2	1	0.999	**1871**	**2.03**	**1097**	**3.56**	**1189**	**35.26**
	α_n^3	1	0.999	2237	2.43	1302	4.23	1417	42.03
RIFPDHF	$\epsilon \cdot \overline{\alpha}$	$0.95 \cdot \psi$	0.95	2375	2.68	1387	4.65	1515	47.66
		$\epsilon \cdot \psi$	0.999	2368	2.67	1384	4.64	1511	47.56
3×3 Gaussian Kernel				(128,0.05,0.99)		(256,0.1,0.99)		(512,0.1,0.99)	
FPDHF	0	1	0.999	1972	2.04	1109	3.53	1265	37.75
IFPDHF	$\epsilon \cdot \overline{\alpha}$	1	0.8	2533	2.65	1435	4.64	1647	48.68
		1	0.9	2109	2.20	1192	3.85	1362	40.23
		1	0.999	1935	2.02	1090	3.52	1244	36.73
DIFPDHF	α_n^1	1	0.999	**1113**	**1.16**	**563**	**1.82**	**674**	**19.89**
	α_n^2	1	0.999	1555	1.63	857	2.77	984	29.04
	α_n^3	1	0.999	1839	1.92	1030	3.33	1176	34.72
RIFPDHF	$\epsilon \cdot \overline{\alpha}$	$0.95 \cdot \psi$	0.95	1946	2.13	1094	3.65	1249	39.12
		$\epsilon \cdot \psi$	0.999	1935	2.11	1089	3.63	1243	38.98

TABLE 4. Comparison of FPDHF with and without inertial and relaxation steps. The step-size (τ, σ) is generated by Initialization 4.5 for $\varepsilon = t(1 + \sqrt{1 + 16\beta^2/\zeta^2})^{-1}$, where $\beta = \|T\|^{-2}$, $\zeta = \mu_2 \delta$, $t \in [0,1[$, the parameter χ is defined in (4.30), and κ_1, and κ_2 are specified on this table. We set $\epsilon = 0.9999$. The parameters $\overline{\alpha} = \overline{\alpha}(\tau\|R\|, \varepsilon, \lambda)$ and $\psi = \psi(\tau\|R\|, \varepsilon)$ are defined in (3.48) and (3.46), respectively. The decreasing inertial parameters α_n^1, α_n^2, and α_n^3 are defined in Table 1. The best results obtained in terms of average number of iterations (IN) and average CPU time (T) are highlighted in bold.

(A) Original Image   (B) z_{20} (PSNR 28.17)   (C) FHRB (PSNR 31.43)   (D) FHRBR (PSNR 31.43)
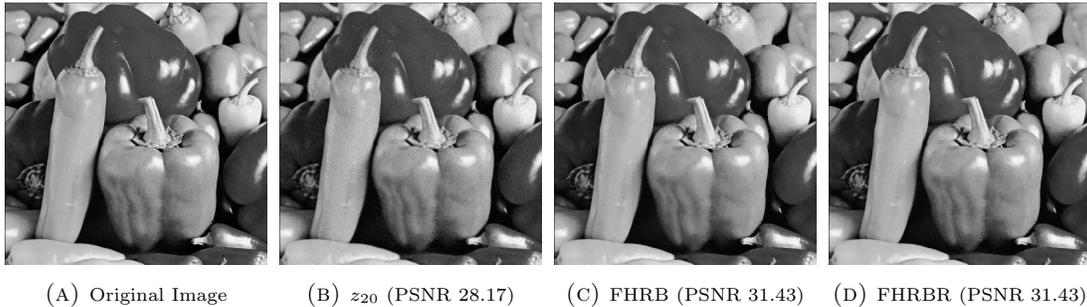
FIGURE 6. Original image, blurred and noisy observation # 20 with a blur of 3×3 averaging kernel, and recovered images for FPDHF and DIFPDHF with $(\alpha_n^2)_{n \in \mathbb{N}}$, with their respective PSNR (dB), for $N = 512$.

5.3. **Linear convergence in denoising problems.** In this section, we analyze the behavior of the FBF method, with and without the inertial step, in the case where the sum of the operators involved in the monotone inclusion is strongly monotone. In this setting, the algorithm converges linearly to the unique solution of the problem. To study this case, we consider the following denoising model. Let $N \in \mathbb{N}$ and let $x \in \mathbb{R}^{N \times N}$ be an image to be recovered from an observation

$$z = x + \epsilon, \tag{5.5}$$

where ϵ is an additive noise perturbing the observation. In this case, we aim to recover the original image x by solving the following optimization problem.

$$\min_{x \in \mathbb{R}^{N \times N}} \frac{1}{2} \|x - z\|^2 + \mu H_\delta(Wx), \tag{5.6}$$

where $\mu \in]0, +\infty[$ is a regularization parameter, $H_\delta : \mathbb{R}^{N \times N} \to \mathbb{R}$ is the Huber function and $W : \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times N}$ is the orthogonal basis wavelet transform defined in Section 5.2. By defining $A = \mathrm{Id} - z$ and $D = \mu W \nabla H_\delta W$, this problem can be solved by FBF. Note that A is 1-strongly monotone and D is μ/δ-Lipschitz. In our experiments we consider $N \in \{128, 256, 512\}$, a Gaussian noise with zero mean and variance 0.004, $\mu = 0.07$, $\delta = 0.01$, and 20 random realizations of the observation z, for each dimension. The stopping criterion is a relative error with tolerance 10^{-9} and a maximum of 5000 iterations. In this setting, we test the algorithm in (4.36) for $C = 0$, $\lambda = 1$, $\tau = 0.9\delta/\mu$, and the values of $(\alpha_n)_{n \in \mathbb{N}}$ described in Table 5. This table shows the results obtained in terms of the average number of iterations (IN) and the average CPU time (T). From this table, we observe the advantage of using a decreasing inertial sequence, which allows reaching the stopping criterion with fewer iterations and less computation time. The inertial sequence $\alpha_n = (\sqrt{\mu/\delta + 1} - 1)(\sqrt{\mu/\delta + 1} + 1 + 0.0001n)^{-1}$, is motivated by [25, Corollary 4.1]. Although the analysis in [25, Corollary 4.1] is carried out for the FB method and strongly convex functions, the numerical experiments show good performance for this particular decreasing inertial sequence in FBF. This motivates further research aimed at finding optimal inertial sequences and convergence rates for NFB with inertia.

6. CONCLUSIONS

In this article, we introduced an inertial and relaxed version of the Nonlinear Forward-Backward (NFB) algorithm and established its weak and strong convergence for both non-decreasing and decreasing sequences of inertial parameters. The analysis for decreasing sequences offers a novel perspective in the study of inertial methods. As particular cases, we

Algorithm	α_n	$N = 128$		$N = 256$		$N = 512$	
		IN	T	IN	T	IN	T
FBF	0	151	0.11	148	0.36	149	3.21
IFBF	$0.99 \cdot 2\left(\psi - 1\right)\left(2\psi - 1 + \sqrt{8\psi - 7}\right)^{-1}$	138	0.11	134	0.32	136	3.00
DIFBF	$\left(9 + 0.00001 \cdot n \cdot (\log(n))^{1.00001}\right)^{-1}$	133	0.10	129	0.31	131	2.90
	$(\sqrt{\mu/\delta + 1} - 1)(\sqrt{\mu/\delta + 1} + 1 + 0.0001n)^{-1}$	**71**	**0.05**	**69**	**0.17**	**74**	**1.65**

TABLE 5. Comparison of FBF with and without inertial in the strongly monotone case. We consider the recurrence in (4.36) for $C = 0$, $\lambda = 1$, and $\tau = 0.9\delta/\mu$. In this case $\psi = 2/(1 + \tau^2\zeta^2)$. The best results obtained in terms of average number of iterations (IN) and average CPU time (T) are highlighted in bold.

(A) $N = 128$ (B) $N = 256$ (C) $N = 512$
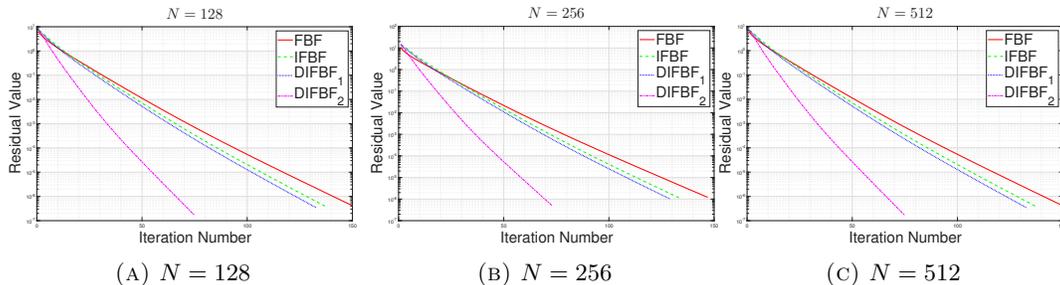
FIGURE 7. Residual value ($\|x_n - z + D(x_n)\|$) for FBF, IFBF, and DIFBF. The graphs display the results for the first of the 20 random realizations of z.

recovered the convergence of the inertial and relaxed versions of algorithms such as Forward-Backward, Forward-Backward-Forward, Chambolle–Pock, and Condat–Vũ, extending previous results to the decreasing case. Furthermore, we derived new convergence results for the Forward-Backward-Half-Forward and Forward-Primal-Dual-Half-Forward algorithms incorporating inertial and relaxation steps. Finally, numerical experiments on optimization problems with affine constraints and image restoration problems demonstrated that using decreasing inertial sequences can enhance the practical performance of these methods by accelerating convergence. A drawback of this approach is the need to choose an appropriate decreasing inertial sequence for each problem. Nevertheless, this acceleration is particularly useful in scenarios where one needs to solve multiple instances of a problem with fixed operators and parameters but varying observations, as in image reconstruction or restoration. Hence, these results, and in particular the *singular* choice of the inertial sequence that yields the favorable numerical results observed in Section 5.3 for strongly monotone operators, motivate further research aimed at determining the rate of convergence in both the monotone and strongly monotone cases, as well as identifying the optimal inertial and relaxation parameters to enhance convergence.

Acknowledgments

References

[1] Alvarez, F.: Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in hilbert space. SIAM J. Optim. **14**(3), 773–782 (2004)

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

[3] Alves, M.M., Eckstein, J., Geremia, M., Melo, J.G.: Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms. Comput. Optim. Appl. **75**(2), 389–422 (2020). DOI 10.1007/s10589-019-00165-y. URL https://doi.org/10.1007/s10589-019-00165-y

[4] Attouch, H., Briceño-Arias, L.M., Combettes, P.L.: A parallel splitting method for coupled monotone inclusions. SIAM J. Control Optim. **48**(5), 3246–3270 (2009/10). DOI 10.1137/090754297. URL https://doi.org/10.1137/090754297

[5] Attouch, H., Cabot, A.: Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions. Appl. Math. Optim. **80**(3), 547–598 (2019). DOI 10.1007/s00245-019-09584-z

[6] Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. SIAM J. Optim. **26**(3), 1824–1834 (2016). DOI 10.1137/15M1046095. URL https://doi.org/10.1137/15M1046095

[7] Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. Math. Program. **174**(1-2), 391–432 (2019). DOI 10.1007/s10107-018-1252-x. URL https://doi.org/10.1007/s10107-018-1252-x

[8] Attouch, H., Peypouquet, J., Redont, P.: A dynamical approach to an inertial forward-backward algorithm for convex minimization. SIAM J. Optim. **24**(1), 232–256 (2014). DOI 10.1137/130910294. URL https://doi.org/10.1137/130910294

[9] Aubin, J.P., Frankowska, H.: Set-valued analysis. Modern Birkhäuser Classics. Birkhäuser Boston, Inc., Boston, MA (2009). DOI 10.1007/978-0-8176-4848-0

[10] Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces, second edn. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham (2017). DOI 10.1007/978-3-319-48311-5

[11] Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans. Image Process. **18**(11), 2419–2434 (2009). DOI 10.1109/TIP.2009.2028250. URL https://doi.org/10.1109/TIP.2009.2028250

[12] Boţ, R., Csetnek, E.: An inertial alternating direction method of multipliers. Minimax Theory Appl. **1**, 29–49 (2016)

[13] Boţ, R.I., Csetnek, E.R.: An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems. Numer. Algorithms **71**(3), 519–540 (2016). DOI 10.1007/s11075-015-0007-5

[14] Boţ, R.I., Csetnek, E.R.: An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems. J. Optim. Theory Appl. **171**(2), 600–616 (2016). DOI 10.1007/s10957-015-0730-z. URL https://doi.org/10.1007/s10957-015-0730-z

[15] Boţ, R.I., Csetnek, E.R., Hendrich, C.: Inertial Douglas-Rachford splitting for monotone inclusion problems. Appl. Math. Comput. **256**, 472–487 (2015). DOI 10.1016/j.amc.2015.01.017

[16] Boţ, R.I., Hendrich, C.: Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization. J. Math. Imaging Vision **49**(3), 551–568 (2014). DOI 10.1007/s10851-013-0486-8. URL https://doi.org/10.1007/s10851-013-0486-8

[17] Boţ, R.I., Nguyen, D.K., Zong, C.: Fast Reflected Forward-Backward algorithm: achieving fast convergence rates for convex optimization with linear cone constraints. J. Sci. Comput. **105**(3), Paper No. 73 (2025). DOI 10.1007/s10915-025-03103-9. URL https://doi.org/10.1007/s10915-025-03103-9

[18] Boţ, R.I., Sedlmayer, M., Vuong, P.T.: A relaxed inertial forward-backward-forward algorithm for solving monotone inclusions with application to GANs. J. Mach. Learn. Res. **24**, Paper No. [8], 37 (2023). DOI 10.4995/agt.2023.12424. URL https://doi.org/10.4995/agt.2023.12424

[19] Briceño, L., Cominetti, R., Cortés, C.E., Martínez, F.: An integrated behavioral model of land use and transport system: a hyper-network equilibrium approach. Netw. Spat. Econ. **8**(2-3), 201–224 (2008). DOI 10.1007/s11067-007-9052-5

[20] Briceño-Arias, L.M., Combettes, P.L.: Monotone operator methods for Nash equilibria in non-potential games. In: Computational and analytical mathematics, *Springer Proc. Math. Stat.*, vol. 50, pp. 143–159. Springer, New York (2013). DOI 10.1007/978-1-4614-7621-4_9

[21] Briceño-Arias, L.M., Combettes, P.L., Pesquet, J.C., Pustelnik, N.: Proximal algorithms for multi-component image recovery problems. J. Math. Imaging Vision **41**(1-2), 3–22 (2011). DOI 10.1007/s10851-010-0243-1. URL https://doi.org/10.1007/s10851-010-0243-1

[22] Briceño-Arias, L.M., Davis, D.: Forward-backward-half forward algorithm for solving monotone inclusions. SIAM J. Optim. **28**(4), 2839–2871 (2018). DOI 10.1137/17M1120099

[23] Briceño-Arias, L.M., Pustelnik, N.: Theoretical and numerical comparison of first order algorithms for cocoercive equations and smooth convex optimization. Signal Processing **206**, 108,900 (2023). DOI https://doi.org/10.1016/j.sigpro.2022.108900. URL https://www.sciencedirect.com/science/article/pii/S016516842200439X

[24] Briceño-Ariass, L., López Rivera, S.: A projected primal-dual method for solving constrained monotone inclusions. J. Optim. Theory Appl. **180**(3), 907–924 (2019). DOI 10.1007/s10957-018-1430-2

[25] Briceño-Arias, L.M.: Lyapunov analysis for FISTA under strong convexity (2025). URL https://arxiv.org/abs/2506.11785

[26] Bùi, M.N., Combettes, P.L.: Warped proximal iterations for monotone inclusions. J. Math. Anal. Appl. **491**(1), 124,315, 21 (2020). DOI 10.1016/j.jmaa.2020.124315. URL https://doi.org/10.1016/j.jmaa.2020.124315

[27] Bùi, M.N., Combettes, P.L.: Multivariate monotone inclusions in saddle form. Math. Oper. Res. **47**(2), 1082–1109 (2022). DOI 10.1287/moor.2021.1161. URL https://doi.org/10.1287/moor.2021.1161

[28] Cevher, V., Vũ, B.: A reflected forward-backward splitting method for monotone inclusions involving lipschitzian operators. Set-Valued Var. Anal. (2020). DOI 10.1007/s11228-020-00542-4

[29] Chambolle, A., Caselles, V., Cremers, D., Novaga, M., Pock, T.: An introduction to total variation for image analysis. In: Theoretical Foundations and Numerical Methods for Sparse Recovery, *Radon Ser. Comput. Appl. Math.*, vol. 9, pp. 263–340. Walter de Gruyter, Berlin (2010). DOI 10.1515/9783110226157.263

[30] Chambolle, A., Dossal, C.: On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". J. Optim. Theory Appl. **166**(3), 968–982 (2015). DOI 10.1007/s10957-015-0746-4. URL https://doi.org/10.1007/s10957-015-0746-4

[31] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vision **40**(1), 120–145 (2011). DOI 10.1007/s10851-010-0251-1

[32] Chouzenoux, E., Pesquet, J.C., Roldán, F.: Solution of mismatched monotone+Lipschitz inclusion problems. SIAM J. Optim. **34**(4), 3564–3591 (2024). DOI 10.1137/23M1609166. URL https://doi.org/10.1137/23M1609166

[33] Combettes, P.L.: Systems of structured monotone inclusions: duality, algorithms, and applications. SIAM J. Optim. **23**(4), 2420–2447 (2013). DOI 10.1137/130904160

[34] Combettes, P.L.: Monotone operator theory in convex optimization. Math. Program. **170**(1), 177–206 (2018). DOI 10.1007/s10107-018-1303-3

[35] Combettes, P.L.: The geometry of monotone operator splitting methods. Acta Numer. **33**, 487–632 (2024). DOI 10.1017/S0962492923000065. URL https://doi.org/10.1017/S0962492923000065

[36] Combettes, P.L., Eckstein, J.: Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions. Math. Program. **168**(1-2, Ser. B), 645–672 (2018). DOI 10.1007/s10107-016-1044-0

[37] Combettes, P.L., Glaudin, L.E.: Quasi-nonexpansive iterations on the affine hull of orbits: from Mann's mean value algorithm to inertial methods. SIAM J. Optim. **27**(4), 2356–2380 (2017). DOI 10.1137/17M112806X. URL https://doi.org/10.1137/17M112806X

[38] Combettes, P.L., Pesquet, J.C.: Fixed point strategies in data science. IEEE Transactions on Signal Processing **69**, 3878–3905 (2021). DOI 10.1109/TSP.2021.3069677

[39] Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J. Optim. Theory Appl. **158**(2), 460–479 (2013). DOI 10.1007/s10957-012-0245-9

[40] Cortild, D., Peypouquet, J.: Krasnoselskii–Mann iterations: Inertia, perturbations and approximation. Journal of Optimization Theory and Applications **204**(2), 35 (2025)

[41] Dong, Y., Sun, M.: New acceleration factors of the Krasnosel'skii-Mann iteration. Results in Mathematics **77**(5), 194 (2022)

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

[43] Fukushima, M.: The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem. Math. Program. **72**(1, Ser. A), 1–15 (1996). DOI 10.1016/0025-5610(95)00012-7

[44] Gabay, D.: Chapter IX applications of the method of multipliers to variational inequalities. In: M. Fortin, R. Glowinski (eds.) Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, *Studies in Mathematics and Its Applications*, vol. 15, pp. 299 – 331. Elsevier, New York (1983). DOI 10.1016/S0168-2024(08)70034-1

[45] Gafni, E.M., Bertsekas, D.P.: Two-metric projection methods for constrained optimization. SIAM J. Control Optim. **22**(6), 936–964 (1984). DOI 10.1137/0322061

[46] Giselsson, P.: Nonlinear forward-backward splitting with projection correction. SIAM J. Optim. **31**(3), 2199–2226 (2021). DOI 10.1137/20M1345062. URL https://doi.org/10.1137/20M1345062

[47] Glowinski, R., Marrocco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér. **9**(R-2), 41–76 (1975)

[48] Goldstein, A.A.: Convex programming in Hilbert space. Bulletin of the American Mathematical Society **70**(5), 709 – 710 (1964). DOI bams/1183526263. URL https://doi.org/

[49] He, X., Hu, R., Fang, Y.P.: Inertial accelerated primal-dual methods for linear equality constrained convex optimization problems. Numer. Algorithms **90**(4), 1669–1690 (2022). DOI 10.1007/s11075-021-01246-y. URL https://doi.org/10.1007/s11075-021-01246-y

[50] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979). DOI 10.1137/0716071

[51] Lorenz, D.A., Pock, T.: An inertial forward-backward algorithm for monotone inclusions. J. Math. Imaging Vision **51**(2), 311–325 (2015). DOI 10.1007/s10851-014-0523-2

[52] Malitsky, Y., Tam, M.K.: A forward-backward splitting method for monotone inclusions without cocoercivity. SIAM J. Optim. **30**(2), 1451–1472 (2020). DOI 10.1137/18M1207260

[53] Maulén, J.J., Fierro, I., Peypouquet, J.: Inertial Krasnoselskii-Mann iterations. Set-Valued and Variational Analysis **32**(2), 10 (2024)

[54] Morin, M., Banert, S., Giselsson, P.: Nonlinear Forward-Backward Splitting with Momentum Correction. Set-Valued Var. Anal. **31**(4), 37 (2023). DOI 10.1007/s11228-023-00700-4. URL https://doi.org/10.1007/s11228-023-00700-4

[55] Moudafi, A., Oliny, M.: Convergence of a splitting inertial proximal method for monotone operators. J. Comput. Appl. Math. **155**(2), 447–454 (2003). DOI 10.1016/S0377-0427(02)00906-8. URL https://doi.org/10.1016/S0377-0427(02)00906-8

[56] Passty, G.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J. Math. Anal. Appl. **72**(2), 383–390 (1979)

[57] Pesquet, J.C., Pustelnik, N.: A parallel inertial proximal optimization method. Pac. J. Optim. **8**(2), 273–306 (2012)

[58] Roldán, F.: Forward-primal-dual-half-forward algorithm for splitting four operators. J. Optim. Theory Appl. **204**(1), Paper No. 11, 26 (2025). DOI 10.1007/s10957-024-02560-w. URL https://doi.org/10.1007/s10957-024-02560-w

[59] Roldán, F., Vega, C.: Relaxed and Inertial Nonlinear Forward–Backward with Momentum. J. Optim. Theory Appl. **206**(2), 23 (2025). DOI 10.1007/s10957-025-02694-5. URL https://doi.org/10.1007/s10957-025-02694-5

[60] Showalter, R.E.: Monotone Operators in Banach Space and Nonlinear Partial Differential Equations, *Mathematical Surveys and Monographs*, vol. 49. American Mathematical Society, Providence, RI (1997). DOI 10.1090/surv/049

[61] Thang, T.V., Hien, N.D., Thach, H.T.C., Anh, P.N.: Weak convergence of inertial proximal algorithms with self adaptive stepsize for solving multivalued variational inequalities. Optimization **73**(4), 995–1023 (2024). DOI 10.1080/02331934.2022.2135966. URL https://doi.org/10.1080/02331934.2022.2135966

[62] Thang, T.V., Tien, H.M.: Parallel inertial forward-backward splitting methods for solving variational inequality problems with variational inclusion constraints. Math. Methods Appl. Sci. **48**(1), 748–764 (2025). DOI 10.1002/mma.10356. URL https://doi.org/10.1002/mma.10356

[63] Tran-Dinh, Q., Nguyen-Trung, N.: Accelerated extragradient-type methods – part 2: Generalization and sublinear convergence rates under co-hypomonotonicity (2025). URL https://arxiv.org/abs/2501.04585

[64] Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM J. Control Optim. **38**(2), 431–446 (2000). DOI 10.1137/S0363012998338806

[65] Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv. Comput. Math. **38**(3), 667–681 (2013). DOI 10.1007/s10444-011-9254-8

[66] Valkonen, T.: Inertial, corrected, primal-dual proximal splitting. SIAM J. Optim. **30**(2), 1391–1420 (2020). DOI 10.1137/18M1182851. URL https://doi.org/10.1137/18M1182851

Centro de Investigación en Ingeniería Matemática (CI²MA)

PRE-PUBLICACIONES 2025 - 2026

2025-24 JULIO ARACENA, FLORIAN BRIDOUX, MAXIMILIEN GADOULEAU, PIERRE GUILLON, KEVIN PERROT, ADRIEN RICHARD, GUILLAUME THEYSSIER: *On the Dynamics of Bounded-Degree Automata Networks*

2025-25 DANIEL CAJAS GUIJARRO, JOHN CAJAS GUIJARRO: *Resonant and non-resonant double Hopf bifurcation in a 4D Goodwin model with wage inequality*

2025-26 ALONSO J. BUSTOS, SERGIO CAUCAO: *A Banach space mixed formulation for the unsteady Brinkman problem with spatially varying porosity*

2025-27 JULIO ARACENA, KATERIN DE LA HOZ, ALEXIS POINDRON, LILIAN SALINAS: *An exploratory approach to the compatibility and inference of unate functions*

2025-28 JESSIKA CAMAÑO, SERGIO CAUCAO, CRISTIAN INZUNZA: *A five-field mixed formulation for the fully dynamic Biot–Brinkman model*

2025-29 JUAN BARAJAS-CALONGE, MAURICIO SEPÚLVEDA, NICOLÁS TORRES, LUIS M. VILLADA: *Dynamically consistent finite volume scheme for a bimonomeric simplified model with inflammation processes for Alzheimer's disease*

2026-01 ALONSO J. BUSTOS, GABRIEL N. GATICA, RICARDO RUIZ-BAIER, BENJAMÍN N. VENEGAS: *A perturbed threefold saddle-point formulation yielding new mixed finite element methods for poroelasticity with reduced symmetry*

2026-02 JORGE AGUAYO, RODOLFO ARAYA: *A Nitsche-type finite element method for the linear elasticity equation with discontinuity jumps*

2026-03 SERGIO CAUCAO, GABRIEL N. GATICA, ADRIAN SUAREZ, IVAN YOTOV: *A skew-symmetry-based mixed formulation for an Oseen-type Kelvin–Voigt–Brinkman–Forchheimer model*

2026-04 AKBAR DAVOODI, DIANA PIGUET, HANKA RADA, NICOLÁS SANHUEZA-MATAMALA: *The asymptotic version of the Erdős–Sós conjecture and beyond*

2026-05 RODRIGO ABARCA DEL RIO, FERNANDO CAMPOS, CRISTÓBAL CARO-RAMÍREZ, JEAN FRANÇOIS CRETAUX, DANIEL MOREIRA, ALFREDO RIBEIRO NETO, JONAS FELIPE SANTOS DE SOUZA, MAURICIO SEPÚLVEDA: *First insights into the performance of the SWOT Level 2 River Single-Pass Vector Data Product in rivers with complex morphology: application to the Bíobío River basin, Chile*

2026-06 JUAN JOSÉ MAULÉN, FERNANDO ROLDÁN, CRISTIAN VEGA: *Relaxed and inertial nonlinear Forward-Backward algorithm*