Article

A Relaxed Inertial Method for Solving Monotone Inclusion Problems with Applications

Chunxiang Zong [1,†], Yuchao Tang [2,†] and Guofeng Zhang [1,*,†]

[1] School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China; zongchx19@lzu.edu.cn
[2] School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China; yctang@gzhu.edu.cn
[*] Correspondence: gf_zhang@lzu.edu.cn
[†] These authors contributed equally to this work.

Abstract: We study a relaxed inertial forward–backward–half-forward splitting approach with variable step size to solve a monotone inclusion problem involving a maximal monotone operator, a cocoercive operator, and a monotone Lipschitz operator. The convergence of the sequence of iterations generated by the discretisations of a continuous-time dynamical system is established under suitable conditions. Given the challenges associated with computing the resolvent of the composite operator, the proposed method is employed to tackle the composite monotone inclusion problem. Additionally, a convergence analysis is conducted under certain conditions. To demonstrate the effectiveness of the algorithm, numerical experiments are performed on the image deblurring problem.

Keywords: forward–backward–half-forward splitting method; relaxed parameters; inertial skill; primal–dual idea

1. Introduction

In a (real) Hilbert space $\mathcal{H}$, we focus on resolving the monotone inclusion problem, which entails the sum of three operators, as follows:

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + Bx + Cx, \tag{1}$$

where $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximal monotone, $B : \mathcal{H} \to \mathcal{H}$ is monotone L-Lipschitz continuous with $L > 0$, and $C : \mathcal{H} \to \mathcal{H}$ is β-cocoercive, for some $\beta > 0$. Moreover, let $A + B$ be maximal monotone and assume that it has a solution, namely,

$$\text{zer}(A + B + C) \neq \varnothing.$$

Problem (1) captures numerous significant challenges in convex optimization problems, signal and image processing, saddle point problems, variational inequalities, partial differential equations, and similar problems. For example, see [1–4].

In recent years, many popular algorithms dealing with monotone inclusion problems involving the sum of three or more operators have been covered in the literature. Although traditional splitting algorithms [5–8] play an indispensable part in addressing monotone inclusions that include the sum of two operators, they cannot be directly applied to solve problems beyond the sum of two operators. A generalized forward–backward splitting (GFBS) method [3] is designed to address the monotone inclusion problem:

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in \sum_{i=1}^{n} A_i x + Cx, \tag{2}$$

where $n \geq 1$, $\{A_i\}_i^n : \mathcal{H} \to 2^{\mathcal{H}}$ indicate maximal monotone operators, and $C : \mathcal{H} \to \mathcal{H}$ is the same as (1). A subsequent work by Raguet and Landrieu [9] addressed (2) using a

preconditioned generalized forward–backward splitting algorithm. An equivalent form of equation (2) can be expressed as the following monotone inclusion formulated in the product space:

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + N_V x + Cx, \tag{3}$$

where $A : \mathcal{H} \to 2^{\mathcal{H}}$ and $C : \mathcal{H} \to \mathcal{H}$ are the same as (1), and N_V denotes the normal cone of a closed vector subspace V. Two novel approaches for solving (3) are presented in [10], where the two methods are equivalent under appropriate conditions. Interestingly, N_V can be extended to a maximal monotone operator. However, the resolvent in this case is no longer linear, necessitating more complicated work to establish convergence. In [11], Davis and Yin finished this work by a three-operator splitting method. In a sense, it extends the GFBS method. It is a remarkable fact that the links between the above methods were precisely studied by Raguet [12], who also derived a new approach to solve (2) along with an extra maximal monotone operator M. Note that the case of (3) occurs when N_V is generalized as a maximal monotone operator and C is relaxed to monotone Lipschitz continuous. A new approach [13] has been established to deal with this case within the computer-assisted skill. In contrast to [13], two new schemes [14] were discovered for tackling the same problem through discretising a continuous-time dynamical system. If N_V is replaced by a monotone Lipschitz operator B, (3) can be translated into (1). Concerning (1), a forward–backward–half-forward (FBHF) splitting method was derived by Briceño-Arias and Davis [15], who exploited the cocoercivity of C by only utilizing it once in every iteration with great ingenuity. See also [16–18] for recent advances in four-operator methods.

Designed as an acceleration method, the inertial scheme is a powerful approach that leverages the characteristic of each new iteration being defined by fully using the previous two iterations. The basic idea was first considered in [19] as a heavy ball method, which was further developed in [20]. Later, Güler [21] and Alvarez et al. [22] generalized the accelerated scheme in [20] for addressing the proximal point scheme and maximal monotone problem, respectively. After that, numerous works involving inertial features were discussed and studied in [23–29].

Relaxation approaches, a principal part of resolving monotone inclusions, offer greater flexibility to the iterative schemes (see [4,30]). In particular, it makes sense to unite inertial and relaxation parameters in a way that enjoys their advantages. Motivated by the inertial proximal algorithm [31], a relaxed inertial proximal algorithm (RIPA) was reported to find the zero of a maximal monotone operator by Attouch and Cabot [32], who also exploited RIPA to address non-smooth convex minimization problems and studied convergence rates in [33]. Further research was made on the more general structure of approaching the solution of the sum of two operators with one being the cocoercive operator [34]. Meanwhile, the idea of combining inertial effect and relaxation method has also been used in the context of the Douglas–Rachford algorithm [35], Tseng's forward–backward–forward algorithm [36], and alternating minimization algorithm [37].

This paper aims to develop a relaxed inertial forward–backward–half-forward (RIF-BHF) scheme that serves as an extension of the FBHF method [15] by combining inertial effects and relaxation parameters to solve (1). It is noteworthy that the FBHF method [15] was considered in a set constraint ($x \in X$) of the monotone inclusion (1). For simplicity, we only study (1). Specifically, the relaxed inertial algorithm is derived from the perspective of discretisations of the continuous-time dynamical system [38], and its convergence is analysed under certain assumptions. We also discuss the relationship between the relaxed parameters and inertial effects. Since estimation of the resolvent of L^*BL is generally challenging, solving the composite monotone inclusion is not straightforward. By drawing upon the primal–dual idea introduced in [39], the composite monotone inclusion can be reformulated equivalently as presented in (1), which can be addressed by our scheme. Similarly, a convex minimization problem is also solved accordingly. At last, numerical tests are designed to validate the effectiveness of the proposed algorithm.

The structure of the paper is outlined as follows. Section 2 provides an introduction to fundamental definitions and lemmas. Section 3 presents the development of a relaxed inertial forward–backward–half-forward splitting algorithm through discretisations of a continuous-time dynamical system, accompanied by a comprehensive convergence analysis. In Section 4, we apply the proposed algorithm to solve the composite monotone inclusion and the convex minimization problem. Section 5 presents several numerical experiments to demonstrate the effectiveness of the proposed algorithm. Finally, conclusions are given in the last section.

2. Preliminaries

In the following discussion, $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces equipped with an inner product $\langle \cdot, \cdot \rangle$ and corresponding norms $\| \cdot \|$, and $\mathbb{N}$ represents the set of nonnegative integers. The $\rightharpoonup$ and $\rightarrow$ signify weak and strong convergence, respectively. $\mathcal{H} \oplus \mathcal{G}$ denotes the Hilbert direct sum of $\mathcal{H}$ and $\mathcal{G}$. The set of proper lower semicontinuous convex functions from $\mathcal{H}$ to $(-\infty, +\infty]$ is denoted by $\Gamma_0(\mathcal{H})$. The followings are denoted:

- The set of zeros of A is zer $A := \{x \in \mathcal{H} \mid 0 \in Ax\}$.
- The domain of A is dom $A := \{x \in \mathcal{H} \mid Ax \neq \varnothing\}$.
- The range of A is ran $A := \{y \in \mathcal{H} \mid \exists x \in \mathcal{H} : y \in Ax\}$.
- The graph set A is gra $A := \{(x, y) \in \mathcal{H}^2 \mid y \in Ax\}$.

The definitions and lemmas that follow are among the most commonly encountered, as documented in the monograph referenced as [4].

Let an operator $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a set-valued map; then,

- A is characterized as monotone if it satisfies the inequality $\langle x - y, u - v \rangle \geq 0$ for all (x, u) and (y, v) belonging to the graph of A.
- A is called maximal monotone if no other monotone operator $B : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ exists for which its graph strictly encompasses the graph of A.
- A is called β-strongly monotone with $\beta \in (0, +\infty)$ if for all $(x, u) \in$ gra A and $(y, v) \in$ gra A, there holds that $\langle x - y, u - v \rangle \geq \beta \|x - y\|^2$.
- A is said to be β-cocoercive with $\beta \in (0, +\infty)$ if $\beta \|Ax - Ay\|^2 \leq \langle Ax - Ay, x - y \rangle$ for all $x, y \in \mathcal{H}$.
- The resolvent of A is defined by

$$J_{\lambda A} = (Id + \lambda A)^{-1},$$

where Id is identity mapping, and $\lambda > 0$. The mapping $A : \mathcal{H} \rightarrow \mathcal{H}$ is L-Lipschitz continuous with $L > 0$ if for every pair of points x and y in $\mathcal{H}$, the inequality $\|Ax - Ay\| \leq L \|x - y\|$ holds. Specifically, A is referred to as nonexpansive when $L = 1$.

Let $f \in \Gamma_0(\mathcal{H})$; the subdifferential of f, denoted by ∂f, is defined as $\partial f : \mathcal{H} \rightarrow 2^{\mathcal{H}} : x \mapsto u \in \mathcal{H} : \forall y \in \mathcal{H}, \langle y - x, u \rangle + f(x) \leq f(y)$. It is well-known that ∂f is maximal monotone. The proximity operator of $f \in \Gamma_0(\mathcal{H})$ is then defined as:

$$\text{prox}_f(u) = \arg\min_x \left\{ \frac{1}{2} \|x - u\|^2 + f(x) \right\}.$$

The well-established relationship $\text{prox}_f = J_{\partial f}$ holds. According to the Baillon–Haddad theorem, if $f : \mathcal{H} \rightarrow \mathbb{R}$ is a convex and differentiable function with a gradient that is Lipschitz continuous with constant $\frac{1}{\beta}$ for some $\beta \in (0, +\infty)$, then ∇f is said to be β-cocoercive. The following equation will be employed later:

$$\|\alpha x + (1 - \alpha)y\|^2 = \alpha \|x\|^2 + (1 - \alpha)\|y\|^2 - \alpha(1 - \alpha)\|x - y\|^2,$$

for all $x \in \mathcal{H}$, $y \in \mathcal{H}$, and $\alpha \in \mathbb{R}$.

The subsequent two lemmas will play a crucial role in the convergence analysis of the proposed algorithm.

Lemma 1 ([23], Lemma 2.3). *Assume* $\{\varphi_k\}_{k\geq 0}$, $\{\delta_k\}_{k\geq 1}$, *and* $\{\alpha_k\}_{k\geq 1}$ *are the sequences in* $[0, +\infty)$ *such that for each* $k \geq 1$,

$$\varphi_{k+1} \leq \varphi_k + \alpha_k(\varphi_k - \varphi_{k-1}) + \delta_k, \quad \sum_{k=1}^{\infty} \delta_k < +\infty$$

and there exists a real number α *satisfying* $0 \leq \alpha_k \leq \alpha < 1$ *for all* $k \geq 1$. *Thus, the following assertions hold:*

(i) $\sum_{k=1}^{\infty} [\varphi_k - \varphi_{k-1}]_+ < +\infty$, *where* $[t]_+ = max\{t, 0\}$;
(ii) *There exists* $\varphi^* \in [0, +\infty)$ *such that* $\lim_{k \to +\infty} \varphi_k = \varphi^*$.

Lemma 2 ((Opial) [4]). *Let* $\mathcal{C}$ *be a nonempty subset of* $\mathcal{H}$, *and let* $\{x_k\}_{k\geq 1}$ *be a sequence in* $\mathcal{H}$ *satisfying the following conditions:*

(1) *For all* $x^* \in \mathcal{C}$, $\lim_{k \to \infty} \|x_k - x^*\|$ *exists;*
(2) *Every weak sequential cluster point of* $\{x_k\}_{k\geq 1}$ *belongs to* $\mathcal{C}$.

Then $\{x_k\}_{k\geq 1}$ *converges weakly to a point in* $\mathcal{C}$.

3. The RIFBHF Algorithm

We establish the RIFBHF algorithm from the perspective of discretisations of continuous-time dynamical systems. First, we pay attention to the second-order dynamical system of the FBHF method studied in [15]:

$$\begin{cases} x(t) = J_{\gamma A}(I - \gamma(B + C))z(t), \\ \ddot{z}(t) + \delta(t)\dot{z}(t) + \tau(t)[z(t) - x(t) - \gamma(Bz(t) - Bx(t))] = 0, \\ z(0) = z_0, \quad \dot{z}(0) = v_0, \end{cases} \quad (4)$$

where $\delta, \tau : [0, +\infty) \to [0, +\infty)$ indicate Lebesgue measurable functions, $0 < \gamma < \frac{\beta}{\beta L + 1}$ and β and L are as in (1), and $z_0, v_0 \in \mathcal{H}$. Let

$$Tz = z - J_{\gamma A}(I - \gamma(B + C))z - \gamma[Bz - BJ_{\gamma A}(I - \gamma(B + C))z].$$

Thereby, (4) is equal to

$$\begin{cases} \ddot{z}(t) + \delta(t)\dot{z}(t) + \tau(t)Tz(t) = 0, \\ z(0) = z_0, \quad \dot{z}(0) = v_0. \end{cases} \quad (5)$$

Note that the cocoercivity of an operator implies its Lipschitz continuity, which implies, in turn, that $B + C$ is Lipschitz continuous with the Lipschitz constant $L + \frac{1}{\beta}$. One can find that T is Lipschitz continuous by Proposition 1 in [36]. Therefore, by the Cauchy–Lipschitz theorem for absolutely continuous trajectories, it can be deduced that the trajectory of (4) exists and is unique.

Next, the trajectories of (5) are approximated at the time point $(kh_k)_{k\in\mathbb{N}}$ using discrete trajectories $(z_k)_{k\in\mathbb{N}}$. Specifically, we employ the central discretisation $\ddot{z}(t) \approx \frac{z_{k+1} - 2z_k + z_{k-1}}{h_k^2}$ and the backward discretisation $\dot{z}(t) \approx \frac{z_k - z_{k-1}}{h_k}$. Let w_k be an extrapolation of z_k and z_{k-1}; one gets

$$\frac{1}{h_k^2}(z_{k+1} - 2z_k + z_{k-1}) + \frac{\delta_k}{h_k}(z_k - z_{k-1}) + \tau_k Tw_k = 0,$$

which implies

$$z_{k+1} = z_k + \alpha_k(z_k - z_{k-1}) - \lambda_k Tw_k,$$

where $\alpha_k = 1 - \delta_k h_k$ and $\lambda_k = h_k^2 \tau_k$. Define that $w_k = z_k + \alpha_k(z_k - z_{k-1})$ for all $k \geq 1$; then, one gets the following RIFBHF iterative for all $k \geq 1$:

$$\begin{cases} w_k = z_k + \alpha_k(z_k - z_{k-1}), \\ x_k = J_{\gamma_k A}(w_k - \gamma_k(B+C)w_k), \\ t_k = x_k + \gamma_k(Bw_k - Bx_k), \\ z_{k+1} = (1 - \lambda_k)w_k + \lambda_k t_k. \end{cases} \tag{6}$$

Remark 1. *The subsequent iterative algorithms can be regarded as specific instances of the proposed algorithm:*

(i) *FBHF method [15]: assume $\alpha_k = 0$ and $\lambda_k = 1$ when $k \geq 1$,*

$$\begin{cases} x_k = J_{\gamma_k A}(z_k - \gamma_k(B+C)z_k), \\ z_{k+1} = x_k - \gamma_k(Bx_k - Bz_k). \end{cases}$$

(ii) *Inertial forward–backward–half-forward scheme [40]: assume $\lambda_k = 1$ when $k \geq 1$,*

$$\begin{cases} w_k = z_k + \alpha_k(z_k - z_{k-1}), \\ x_k = J_{\gamma_k A}(w_k - \gamma_k(B+C)w_k), \\ z_{k+1} = x_k + \gamma_k(Bw_k - Bx_k). \end{cases}$$

(iii) *Relaxed forward–backward–half-forward method: assume $\alpha_k = 0$ when $k \geq 1$,*

$$\begin{cases} x_k = J_{\gamma_k A}(z_k - \gamma_k(B+C)z_k), \\ t_k = x_k - \gamma_k(Bx_k - Bz_k), \\ z_{k+1} = (1 - \lambda_k)z_k + \lambda_k t_k. \end{cases}$$

Furthermore, the convergence results of the proposed algorithm will be established. The convergence analysis relies heavily on the following properties.

Proposition 1. *Consider the problem (1). Suppose $\{\gamma_k\}_{k\geq 1}$ is a sequence of positive numbers, and $\{t_k\}_{k\geq 1}$, $\{w_k\}_{k\geq 1}$, and $\{x_k\}_{k\geq 1}$ are the sequences generated by (6). Assume that $y_k = w_k - \gamma_k(B+C)w_k$ in (6) for all $k \geq 1$. For all $z^* \in \mathrm{zer}(A+B+C)$, then*

$$\|t_k - z^*\|^2 \leq \|w_k - z^*\|^2 - L^2(\chi^2 - \gamma_k^2)\|w_k - x_k\|^2$$
$$- \frac{2\beta\gamma_k}{\chi}(\chi - \gamma_k)\|Cw_k - Cz^*\|^2 - \frac{\chi}{2\beta}\left\|w_k - x_k - \frac{2\beta\gamma_k}{\chi}(Cw_k - Cz^*)\right\|^2,$$

where $\gamma_k \in (0, \chi)$, and

$$\chi = \frac{4\beta}{1 + \sqrt{1 + 16\beta^2 L^2}} \leq \min\{2\beta, L^{-1}\}.$$

Proof. By definition, we get $x_k = J_{\gamma_k A}(y_k)$ and $0 \in (A+B+C)z^*$ such that $\frac{1}{\gamma_k}(y_k - x_k) \in Ax_k$ and $-(B+C)z^* \in Az^*$. Therefore, in view of the monotonicity of A and B, one has

$$\left\langle x_k - z^*, \frac{1}{\gamma_k}(y_k - x_k) + (B+C)z^* \right\rangle \geq 0, \tag{7}$$

and

$$\langle x_k - z^*, Bx_k - Bz^* \rangle \geq 0. \tag{8}$$

Using (7) and (8), we yield

$$
\begin{aligned}
\langle x_k - z^*, x_k &- y_k - \gamma_k B x_k \rangle \\
&= \langle x_k - z^*, \gamma_k C z^* \rangle + \langle x_k - z^*, x_k - y_k - \gamma_k(B + C)z^* \rangle \\
&\quad + \langle x_k - z^*, \gamma_k(Bz^* - Bx_k) \rangle \\
&\leq \langle x_k - z^*, \gamma_k C z^* \rangle.
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
2\gamma_k &\langle x_k - z^*, Bw_k - Bx_k \rangle \\
&= 2\langle x_k - z^*, \gamma_k Bw_k + y_k - x_k \rangle + 2\langle x_k - z^*, x_k - y_k - \gamma_k Bx_k \rangle \\
&\leq 2\langle x_k - z^*, \gamma_k(B + C)w_k + y_k - x_k \rangle + 2\langle x_k - z^*, \gamma_k Cz^* - \gamma_k Cw_k \rangle \\
&= 2\langle x_k - z^*, w_k - x_k \rangle + 2\langle x_k - z^*, \gamma_k Cz^* - \gamma_k Cw_k \rangle \\
&= \|w_k - z^*\|^2 - \|x_k - z^*\|^2 - \|w_k - x_k\|^2 + 2\langle x_k - z^*, \gamma_k Cz^* - \gamma_k Cw_k \rangle.
\end{aligned} \tag{9}
$$

Since C is cocoercive, one gets for all $\varepsilon > 0$:

$$
\begin{aligned}
2\langle x_k - z^*, &\gamma_k Cz^* - \gamma_k Cw_k \rangle \\
&= 2\langle w_k - z^*, \gamma_k Cz^* - \gamma_k Cw_k \rangle + 2\langle x_k - w_k, \gamma_k Cz^* - \gamma_k Cw_k \rangle \\
&\leq -2\gamma_k \beta \|Cw_k - Cz^*\|^2 + 2\langle x_k - w_k, \gamma_k Cz^* - \gamma_k Cw_k \rangle \\
&= -2\gamma_k \beta \|Cw_k - Cz^*\|^2 + \varepsilon\|w_k - x_k\|^2 + \frac{\gamma_k^2}{\varepsilon}\|Cw_k - Cz^*\|^2 \\
&\quad - \varepsilon\left\|w_k - x_k - \frac{\gamma_k}{\varepsilon}(Cw_k - Cz^*)\right\|^2 \\
&= \varepsilon\|w_k - x_k\|^2 - \gamma_k\left(2\beta - \frac{\gamma_k}{\varepsilon}\right)\|Cw_k - Cz^*\|^2 \\
&\quad - \varepsilon\left\|w_k - x_k - \frac{\gamma_k}{\varepsilon}(Cw_k - Cz^*)\right\|^2.
\end{aligned} \tag{10}
$$

Thus, in view of (9), (10), and the Lipschitz continuity of B, then

$$
\begin{aligned}
\|t_k - z^*\|^2 &= \|x_k + \gamma_k(Bw_k - Bx_k) - z^*\|^2 \\
&= \|x_k - z^*\|^2 + 2\gamma_k\langle x_k - z^*, Bw_k - Bx_k \rangle + \gamma_k^2\|Bw_k - Bx_k\|^2 \\
&\leq \|x_k - z^*\|^2 + \|w_k - z^*\|^2 - \|x_k - z^*\|^2 - \|w_k - x_k\|^2 \\
&\quad + \gamma_k^2\|Bw_k - Bx_k\|^2 + 2\langle x_k - z^*, \gamma_k Cz^* - \gamma_k Cw_k \rangle \\
&\leq \|w_k - z^*\|^2 - \|w_k - x_k\|^2 + \gamma_k^2\|Bw_k - Bx_k\|^2 + \varepsilon\|w_k - x_k\|^2 \\
&\quad - \gamma_k\left(2\beta - \frac{\gamma_k}{\varepsilon}\right)\|Cw_k - Cz^*\|^2 - \varepsilon\left\|w_k - x_k - \frac{\gamma_k}{\varepsilon}(Cw_k - Cz^*)\right\|^2 \\
&\leq \|w_k - z^*\|^2 - \|w_k - x_k\|^2 + \gamma_k^2 L^2\|w_k - x_k\|^2 + \varepsilon\|w_k - x_k\|^2 \\
&\quad - \gamma_k\left(2\beta - \frac{\gamma_k}{\varepsilon}\right)\|Cw_k - Cz^*\|^2 - \varepsilon\left\|w_k - x_k - \frac{\gamma_k}{\varepsilon}(Cw_k - Cz^*)\right\|^2 \\
&= \|w_k - z^*\|^2 - \left(1 - \varepsilon - \gamma_k^2 L^2\right)\|w_k - x_k\|^2 \\
&\quad - \frac{\gamma_k}{\varepsilon}(2\beta\varepsilon - \gamma_k)\|Cw_k - Cz^*\|^2 - \varepsilon\left\|w_k - x_k - \frac{\gamma_k}{\varepsilon}(Cw_k - Cz^*)\right\|^2.
\end{aligned} \tag{11}
$$

Similar to [15], let $\chi = \frac{\sqrt{1-\varepsilon}}{L} = 2\beta\varepsilon$ for $0 < \varepsilon < 1$, allowing us to determine the widest interval for γ_k such that the second and third terms on the right-hand side of (11) are negative. □

Proposition 2. *Consider the problem (1) and assume that $z^* \in \mathrm{zer}(A + B + C)$. Suppose that $\{\lambda_k\}_{k \geq 1} > 0, 0 < \gamma_k < \chi$, and χ is defined as in Proposition 1. Assume $\{\alpha_k\}_{k \geq 1}$ is nondecreasing and satisfies $0 \leq \alpha_k \leq \alpha < 1$. Let $\{z_k\}_{k \geq 1}$ denote the sequence generated by (6). Then*

(i) $\|z_{k+1} - z^*\|^2 \leq \|w_k - z^*\|^2 - \varphi_k \|z_{k+1} - w_k\|^2,$

where $\varphi_k = \left(\dfrac{L^2(\chi^2 - \gamma_k^2)}{\lambda_k(1 + \gamma_k L)^2} + \dfrac{1 - \lambda_k}{\lambda_k} \right).$

(ii) *Define that*

$$\Pi_k := \|z_k - z^*\|^2 - \alpha_k \|z_{k-1} - z^*\|^2 + \left(\alpha_k(1 + \alpha_k) - \varphi_k(\alpha_k^2 - \alpha_k) \right) \|z_k - z_{k-1}\|^2.$$

Then

$$\Pi_{k+1} - \Pi_k \leq -\zeta_k \|z_{k+1} - z_k\|^2, \tag{12}$$

where $\zeta_k = \varphi_k(1 - \alpha_{k+1}) - \alpha_{k+1}(1 + \alpha_{k+1}) + \varphi_{k+1}(\alpha_{k+1}^2 - \alpha_{k+1}).$

Proof.

(i) Proposition 1 leads to

$$
\begin{aligned}
&\|z_{k+1} - z^*\|^2 \\
&= \|(1 - \lambda_k)w_k + \lambda_k t_k - z^*\|^2 \\
&= \|(1 - \lambda_k)(w_k - z^*) + \lambda_k(t_k - z^*)\|^2 \\
&= (1 - \lambda_k)\|w_k - z^*\|^2 + \lambda_k \|t_k - z^*\|^2 - \lambda_k(1 - \lambda_k)\|t_k - w_k\|^2 \\
&\leq (1 - \lambda_k)\|w_k - z^*\|^2 + \lambda_k \left(\|w_k - z^*\|^2 - L^2(\chi^2 - \gamma_k^2)\|w_k - x_k\|^2 \right) \\
&\quad - \lambda_k(1 - \lambda_k)\|t_k - w_k\|^2 \\
&= \|w_k - z^*\|^2 - \lambda_k L^2(\chi^2 - \gamma_k^2)\|w_k - x_k\|^2 - \lambda_k(1 - \lambda_k)\|t_k - w_k\|^2. \tag{13}
\end{aligned}
$$

According to the Lipschitz continuity of B,

$$
\begin{aligned}
\frac{1}{\lambda_k}\|z_{k+1} - w_k\| = \|t_k - w_k\| &\leq \|t_k - x_k\| + \|x_k - w_k\| \\
&= \gamma_k \|Bw_k - Bx_k\| + \|w_k - x_k\| \\
&\leq (1 + \gamma_k L)\|w_k - x_k\|,
\end{aligned}
$$

which implies that

$$\frac{L^2(\chi^2 - \gamma_k^2)}{\lambda_k(1 + \gamma_k L)^2}\|z_{k+1} - w_k\|^2 \leq \lambda_k L^2(\chi^2 - \gamma_k^2)\|w_k - x_k\|^2. \tag{14}$$

Combining (13) and (14), we have

$$\|z_{k+1} - z^*\|^2 \leq \|w_k - z^*\|^2 - \left(\frac{L^2(\chi^2 - \gamma_k^2)}{\lambda_k(1 + \gamma_k L)^2} + \frac{1 - \lambda_k}{\lambda_k} \right)\|z_{k+1} - w_k\|^2.$$

(ii) It follows from the definition of w_k and the Cauchy–Schwartz inequality that

$$
\begin{aligned}
\|z_{k+1} - w_k\|^2 &= \|z_{k+1} - z_k - \alpha_k(z_k - z_{k-1})\|^2 \\
&= \|z_{k+1} - z_k\|^2 + \alpha_k^2 \|z_k - z_{k-1}\|^2 - 2\alpha_k \langle z_{k+1} - z_k, z_k - z_{k-1} \rangle \\
&\geq \|z_{k+1} - z_k\|^2 + \alpha_k^2 \|z_k - z_{k-1}\|^2 - 2\alpha_k \|z_{k+1} - z_k\| \|z_k - z_{k-1}\| \\
&\geq (1 - \alpha_k)\|z_{k+1} - z_k\|^2 + (\alpha_k^2 - \alpha_k)\|z_k - z_{k-1}\|^2. \tag{15}
\end{aligned}
$$

Simultaneously, we have

$$
\begin{aligned}
\|w_k - z^*\|^2 &= \|z_k + \alpha_k(z_k - z_{k-1}) - z^*\|^2 \\
&= \|(1 + \alpha_k)(z_k - z^*) - \alpha_k(z_{k-1} - z^*)\|^2 \\
&= (1 + \alpha_k)\|z_k - z^*\|^2 - \alpha_k\|z_{k-1} - z^*\|^2 + \alpha_k(1 + \alpha_k)\|z_k - z_{k-1}\|^2. \quad (16)
\end{aligned}
$$

By Propositions 2(i), (15), and (16), we have

$$
\begin{aligned}
\|z_{k+1} - z^*\|^2 &\le \|w_k - z^*\|^2 - \varphi_k\|z_{k+1} - w_k\|^2 \\
&\le (1 + \alpha_k)\|z_k - z^*\|^2 - \alpha_k\|z_{k-1} - z^*\|^2 + \alpha_k(1 + \alpha_k)\|z_k - z_{k-1}\|^2 \\
&\quad - \varphi_k\big((1 - \alpha_k)\|z_{k+1} - z_k\|^2 + (\alpha_k^2 - \alpha_k)\|z_k - z_{k-1}\|^2\big) \\
&= (1 + \alpha_k)\|z_k - z^*\|^2 - \alpha_k\|z_{k-1} - z^*\|^2 - \varphi_k(1 - \alpha_k)\|z_{k+1} - z_k\|^2 \\
&\quad + \big(\alpha_k(1 + \alpha_k) - \varphi_k(\alpha_k^2 - \alpha_k)\big)\|z_k - z_{k-1}\|^2 \\
&= (1 + \alpha_k)\|z_k - z^*\|^2 - \alpha_k\|z_{k-1} - z^*\|^2 - \pi_k\|z_{k+1} - z_k\|^2 \\
&\quad + \eta_k\|z_k - z_{k-1}\|^2, \quad (17)
\end{aligned}
$$

where $\pi_k := \varphi_k(1 - \alpha_k)$ and $\eta_k := \alpha_k(1 + \alpha_k) - \varphi_k(\alpha_k^2 - \alpha_k)$. Furthermore, we define that

$$
\Pi_k := \|z_k - z^*\|^2 - \alpha_k\|z_{k-1} - z^*\|^2 + \eta_k\|z_k - z_{k-1}\|^2.
$$

Now, by (17) and $\alpha_k \le \alpha_{k+1}$, we obtain

$$
\begin{aligned}
\Pi_{k+1} - \Pi_k &= \|z_{k+1} - z^*\|^2 - \alpha_{k+1}\|z_k - z^*\|^2 + \eta_{k+1}\|z_{k+1} - z_k\|^2 \\
&\quad - \|z_k - z^*\|^2 + \alpha_k\|z_{k-1} - z^*\|^2 - \eta_k\|z_k - z_{k-1}\|^2 \\
&\le \|z_{k+1} - z^*\|^2 - (1 + \alpha_k)\|z_k - z^*\|^2 + \alpha_k\|z_{k-1} - z^*\|^2 \\
&\quad + \eta_{k+1}\|z_{k+1} - z_k\|^2 - \eta_k\|z_k - z_{k-1}\|^2 \\
&\le (1 + \alpha_k)\|z_k - z^*\|^2 - \alpha_k\|z_{k-1} - z^*\|^2 - \pi_k\|z_{k+1} - z_k\|^2 \\
&\quad + \eta_k\|z_k - z_{k-1}\|^2 - (1 + \alpha_k)\|z_k - z^*\|^2 + \alpha_k\|z_{k-1} - z^*\|^2 \\
&\quad + \eta_{k+1}\|z_{k+1} - z_k\|^2 - \eta_k\|z_k - z_{k-1}\|^2 \\
&= -(\pi_k - \eta_{k+1})\|z_{k+1} - z_k\|^2. \quad (18)
\end{aligned}
$$

If follows from $0 < \alpha_k \le \alpha_{k+1} \le \alpha$ that

$$
\begin{aligned}
\pi_k - \eta_{k+1} &= \varphi_k(1 - \alpha_k) - \alpha_{k+1}(1 + \alpha_{k+1}) + \varphi_{k+1}(\alpha_{k+1}^2 - \alpha_{k+1}) \\
&\ge \varphi_k(1 - \alpha_{k+1}) - \alpha_{k+1}(1 + \alpha_{k+1}) + \varphi_{k+1}(\alpha_{k+1}^2 - \alpha_{k+1}).
\end{aligned}
$$

Let $\zeta_k = \varphi_k(1 - \alpha_{k+1}) - \alpha_{k+1}(1 + \alpha_{k+1}) + \varphi_{k+1}(\alpha_{k+1}^2 - \alpha_{k+1})$; we obtain

$$
\Pi_{k+1} - \Pi_k \le -\zeta_k\|z_{k+1} - z_k\|^2.
$$

The proof is completed.

$\square$

Furthermore, seeking to ensure the convergence of (6), let $\lim_{k \to +\infty} \alpha_k = \alpha \ge 0$, $\lim_{k \to +\infty} \gamma_k = \gamma > 0$ and $\lim_{k \to +\infty} \lambda_k = \lambda > 0$ by the idea of Boţ et al. [36]. Proposition 2(ii) implies that

$$
\lim_{k \to +\infty} \zeta_k = \frac{L^2\chi^2 + 1 + 2\gamma L}{\lambda(1 + \gamma L)^2}(1 - \alpha)^2 - 1 + \alpha - 2\alpha^2.
$$

Since $\chi = \frac{\sqrt{1-\varepsilon}}{L}$, we have

$$\lim_{k \to +\infty} \zeta_k = \frac{2(1+\gamma L) - \varepsilon}{\lambda(1+\gamma L)^2}(1-\alpha)^2 - 1 + \alpha - 2\alpha^2.$$

Then, to ensure $\lim_{k \to +\infty} \zeta_k > 0$, the following holds:

$$0 < \lambda < \frac{2(1+\gamma L) - \varepsilon}{(1+\gamma L)^2}\frac{(1-\alpha)^2}{(2\alpha^2 - \alpha + 1)}.$$

Next, we establish the principal convergence theorem.

Theorem 1. *In problem (1), we assume that $z^* \in \mathrm{zer}(A + B + C)$. Let a nondecreasing sequence $\{\alpha_k\}_{k \geq 1}$ satisfy $0 \leq \alpha_k \leq \alpha < 1$. Assume that $\eta \leq \gamma_k \leq \chi - \eta$, $0 < \eta < \frac{\chi}{2}$, and χ is defined as in Proposition 1. In addition, let $\{\lambda_k\}_{k \geq 1}$ be nonnegative, and*

$$0 < \lim_{k \to +\infty} \lambda_k = \lambda < \frac{2(1+\gamma L) - \varepsilon}{(1+\gamma L)^2}\frac{(1-\alpha)^2}{(2\alpha^2 - \alpha + 1)}, \tag{19}$$

where $\varepsilon = \frac{2}{1+\sqrt{1+16\beta^2 L^2}}$. Then, the sequence $\{z_k\}_{k \geq 1}$ obtained by (6) converges weakly to a solution of $\mathrm{zer}(A + B + C)$.

Proof. For any $z^* \in \mathrm{zer}(A + B + C)$, by (19), we have $\lim_{k \to +\infty} \zeta_k > 0$. This implies the existence of $k_0 \geq 1$ such that for every $k \geq k_0$,

$$\Pi_{k+1} - \Pi_k \leq -\zeta_k \|z_{k+1} - z_k\|^2 \leq 0. \tag{20}$$

As a result, the sequence $\{\Pi_k\}_{k \geq k_0}$ is nonincreasing, and the bound for $\{\alpha_k\}_{k \geq 1}$ yields

$$-\alpha\|z_{k-1} - z^*\|^2 \leq \|z_k - z^*\|^2 - \alpha\|z_{k-1} - z^*\|^2 \leq \Pi_k \leq \Pi_{k_0}, \tag{21}$$

which indicates that

$$\|z_k - z^*\|^2 \leq \alpha\|z_{k-1} - z^*\|^2 + \Pi_{k_0}$$
$$\leq \alpha^{k-k_0}\|z_{k_0} - z^*\|^2 + \Pi_{k_0}\frac{1-\alpha^{k-k_0}}{1-\alpha}. \tag{22}$$

Combining (20)–(22) and $\alpha \in [0, 1)$, we have

$$\sum_{i=k_0}^{k} \zeta_i \|z_{i+1} - z_i\|^2 \leq \Pi_{k_0} - \Pi_{k+1}$$
$$\leq \Pi_{k_0} + \alpha\|z_k - z^*\|^2$$
$$\leq \alpha^{k-k_0+1}\|z_{k_0} - z^*\|^2 + \Pi_{k_0}\frac{1-\alpha^{k-k_0+1}}{1-\alpha}$$
$$\leq \|z_{k_0} - z^*\|^2 + \frac{\Pi_{k_0}}{1-\alpha},$$

which indicates that $\lim_{k \to +\infty} \zeta_k \|z_{k+1} - z_k\| = 0$. Since $\liminf_{k \to +\infty} \zeta_k > 0$, this yields $\lim_{k \to +\infty} \|z_{k+1} - z_k\| = 0$. Let us take account of (17) and Lemma 1 and observe that $\lim_{k \to +\infty} \|z_k - z^*\|$ exists.

Meanwhile,

$$\|t_k - w_k\| = \frac{1}{\lambda_k}\|z_{k+1} - w_k\| \leq \frac{1}{\lambda_k}\|z_{k+1} - z_k\| + \frac{\alpha_k}{\lambda_k}\|z_k - z_{k-1}\|, \tag{23}$$

which implies that $\lim_{k \to +\infty} \|t_k - w_k\| = 0$. In addition, for every $k \geq 1$, we have

$$
\begin{aligned}
\|t_k - w_k\| &= \|x_k + \gamma_k(Bw_k - Bx_k) - w_k\| \\
&\geq \|x_k - w_k\| - \gamma_k\|Bw_k - Bx_k\| \\
&\geq (1 - \gamma_k L)\|x_k - w_k\|.
\end{aligned}
$$

Due to $\gamma_k < \frac{1}{L}$, we deduce that

$$
\lim_{k \to +\infty} \|x_k - w_k\| = 0. \tag{24}
$$

Assume $\bar{z}$ is a weak limit point of the sequence $\{z_k\}_{k \geq 1}$. Since $\{z_k\}_{k \geq 1}$ is bounded, there exists a subsequence $\{z_{k_i}\}_{i \geq 0}$ that converges weakly to $\bar{z}$. In view of (23) and (24), $\{w_{k_i}\}_{i \geq 0}$ and $\{x_{k_i}\}_{i \geq 0}$ also converge weakly to $\bar{z}$. Next, since $x_{k_i} = J_{\gamma_{k_i} A}\left(w_{k_i} - \gamma_{k_i}(B + C)w_{k_i}\right)$, we have $u_{k_i} := \gamma_{k_i}^{-1}(w_{k_i} - x_{k_i}) - (B + C)w_{k_i} + (B + C)x_{k_i} \in (A + B + C)x_{k_i}$. Therefore, utilizing (24), the fact that $\gamma_k \in [\eta, \chi - \eta]$, and combined with the Lipschitz continuity of $B + C$, we conclude that $u_{k_i} \to 0$. Due to the maximal monotonicity of $A + B$ and the cocoercivity of C, it follows that $A + B + C$ is maximal monotone, and its graph is closed in the weak–strong topology in $\mathcal{H} \times \mathcal{H}$ (see Proposition 20.37(ii) in [4]). As a result, $\bar{z} \in \text{zer}(A + B + C)$. Following Lemma 2, we conclude that the sequence $\{z_k\}_{k \geq 1}$ weakly converges to an element of $\text{zer}(A + B + C)$. This completes the proof. $\square$

Remark 2 (Inertia versus relaxation). *The inequation (19) establishes a relationship between inertial and relaxation parameters. Figure 1 displays the relationship between α_k and λ_k by a graphical representation for some given values of γ and ε, which has a similar graphical representation to Figure 1 in [36]. It is noteworthy that the expression of the upper bound for λ_k resembles that in ([36], Remark 2) if $\varepsilon \searrow 0$. Assume $\frac{2(1+\gamma L) - \varepsilon}{(1+\gamma L)^2}$ is constant; it follows from (19) that the upper bound of λ_k takes the form of $\lambda_{\max}(\alpha, \gamma) = \frac{2(1+\gamma L) - \varepsilon}{(1+\gamma L)^2} \frac{(1-\alpha)^2}{(2\alpha^2 - \alpha + 1)}$ with $0 \leq \alpha < 1$. Further, note that the relaxation parameter $\lambda_{\max}(\alpha, \gamma)$ is a decreasing function with respect to inertial parameter α on the interval $[0, 1]$: for example, when $\alpha \nearrow 1$, then $\lambda_{\max}(\alpha, \gamma) \searrow 0$. Of course, we can also get $\lambda_{\max}(\gamma) = \frac{2(1+\gamma L) - \varepsilon}{(1+\gamma L)^2}$ when $\alpha = 0$, and $\lambda_{\max}(\gamma)$ is also decreasing on $(0, \frac{1}{L})$ because of limiting values $2 - \varepsilon$ as $\gamma \searrow 0$ and $1 - \frac{\varepsilon}{4}$ as $\gamma \nearrow \frac{1}{L}$.*
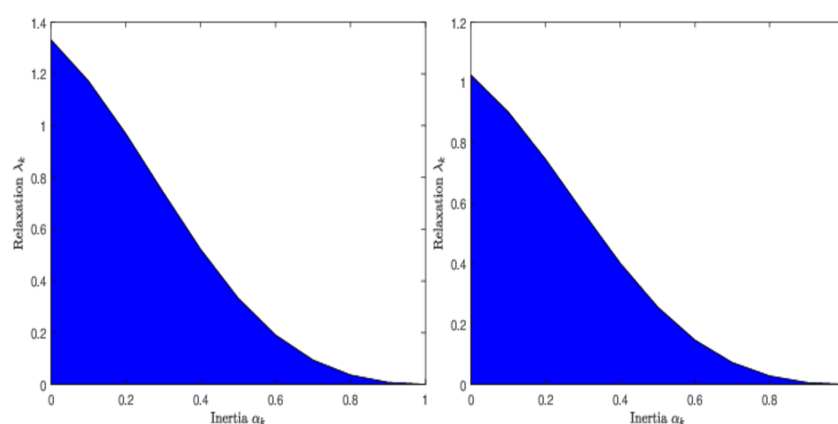
Figure 1. Balance between α_k and λ_k with $\varepsilon = 0.1$, $\gamma = \frac{0.45}{L}$ (**left**), and $\gamma = \frac{0.9}{L}$ (**right**).

Remark 3. *The parameters for FBHF [15] and RIFBHF are given in Table 1, which shows that the two schemes have the same range of step sizes. Different from FBHF [15], RIFBHF introduces relaxation parameter λ_k and inertial parameter α_k. Note that RIFBHF can reduce to FBHF [15] if $\lambda_k = 1$ and $\alpha_k = 0$.*

Table 1. The parameter selection range for FBHF [15] and RIFBHF.

Algorithms	γ_k	α_k	$\lim_{k\to+\infty} \lambda_k = \lambda$
FBHF [15]	$[\eta, \chi - \eta]$	0	1
RIFBHF	$[\eta, \chi - \eta]$	$[0,1)$	$\left(0, \frac{2(1+\gamma L)-\varepsilon}{(1+\gamma L)^2} \frac{(1-\alpha)^2}{(2\alpha^2 - \alpha + 1)}\right)$

Remark 4. *The existence of the over-relaxation* $(\lambda_k > 1)$ *for RIFBHF deserves further discussion. If* $0 \le \alpha < \frac{\sqrt{\theta^2 + 4(2\gamma^2 L^2 + 2\gamma L + \varepsilon)(1-\varepsilon-\gamma^2 L^2)} - \theta}{2(2\gamma^2 L^2 + 2\gamma L + \varepsilon)}$ *with* $\theta = 2\gamma L - \gamma^2 L^2 + 3 - 2\varepsilon$ *for* $k \ge k_0$, *we conclude that* λ_k *has over-relaxation. In addition, observe that the over-relaxation in [36] exists when* $0 \le \alpha < \frac{\sqrt{(3-\gamma L)^2 + 8\gamma L(1-\gamma L)} - 3 + \gamma L}{4\gamma L}$ *for* $k \ge k_0$. *Although the upper bounds of* α *for the two approaches are different, the over-relaxation* $(\lambda_k > 1)$ *for the two methods is possible when* α *is in a small range.*

4. Composite Monotone Inclusion Problem

The aim of this section is to use the proposed algorithm to solve a more generalized inclusion problem, which is outlined as follows:

$$0 \in Ax + L^*BLx + Cx, \tag{25}$$

where $A : \mathcal{H} \to 2^{\mathcal{H}}$ and $B : \mathcal{G} \to 2^{\mathcal{G}}$ represent two maximal monotone operators, $C : \mathcal{H} \to \mathcal{H}$ is a β-cocoercive operator with $\beta > 0$, and $L : \mathcal{H} \to \mathcal{G}$ is a bounded linear operator. In addition, the following assumption is given:

$$\text{zer}(A + L^*BL + C) \ne \emptyset.$$

The key to solving (25) is to know the exact resolvent J_{L^*BL}. As we know, J_{L^*BL} can be estimated exactly using only the resolvent of the operator B, the linear operator L, and its adjoint L^* when $L^*L = \nu I$ for some $\nu > 0$. However, this condition usually does not hold in our interesting problems, such as total variation regularization. To address this challenge, we introduce an efficient iterative algorithm to tackle (25) by combining the primal–dual approach [39] and (6). Specifically, drawing inspiration from the fully splitting primal–dual algorithm studied by Briceño-Arias and Combettes [39], we naturally rewrite (25) as the following problem, letting $\mathcal{K} = \mathcal{H} \oplus \mathcal{G}$:

$$\text{find} \quad x \in \mathcal{K} \quad \text{such that} \quad 0 \in Mx + Sx + Nx, \tag{26}$$

where

$$\begin{cases} M : \mathcal{K} \to 2^{\mathcal{K}} : (x,y) \mapsto (Ax) \times (B^{-1}y), \\ S : \mathcal{K} \to \mathcal{K} : (x,y) \mapsto (L^*y, -Lx), \\ N : \mathcal{K} \to \mathcal{K} : (x,y) \mapsto (Cx, 0). \end{cases} \tag{27}$$

Notice that M is maximal monotone and S is monotone Lipschitz continuous within a constant $\|L\|$, as stated in Proposition 2.7 of [39]. This implies that $M + S$ is also maximal monotone since S is skew-symmetric. A result yields the cocoercivity of N by the cocoercivity of C. Therefore, it follows from (6) and (27) that the following convergence analysis can deal with (26), which implies that (25) is also solved.

Corollary 1. *Suppose that* $A : \mathcal{H} \to 2^{\mathcal{H}}$ *is maximal monotone; let* $B : \mathcal{G} \to 2^{\mathcal{G}}$ *be maximal monotone, and assume that* $C : \mathcal{H} \to \mathcal{H}$ *is* β-cocoercive with $\beta > 0$. *Let* $L : \mathcal{H} \to \mathcal{G}$ *be a nonzero bounded linear operator. Given initial data* $x_0, x_1 \in \mathcal{H}$ *and* $y_0, y_1 \in \mathcal{G}$, *the iteration sequences are defined:*

$$\left|\begin{array}{l} w_{1,k} = x_k + \alpha_k(x_k - x_{k-1}), \\ w_{2,k} = y_k + \alpha_k(y_k - y_{k-1}), \\ z_{1,k} = J_{\gamma_k A}(w_{1,k} - \gamma_k(L^* w_{2,k} + C w_{1,k})), \\ z_{2,k} = J_{\gamma_k B^{-1}}(w_{2,k} + \gamma_k L w_{1,k}), \\ t_{1,k} = z_{1,k} + \gamma_k(L^* w_{2,k} - L^* z_{2,k}), \\ t_{2,k} = z_{2,k} + \gamma_k(-L w_{1,k} + L z_{1,k}), \\ x^{k+1} = (1 - \lambda_k) w_{1,k} + \lambda_k t_{1,k}, \\ y^{k+1} = (1 - \lambda_k) w_{2,k} + \lambda_k t_{2,k}, \end{array}\right. \qquad (28)$$

where $\eta \leq \gamma_k \leq \chi - \eta$, $0 < \eta < \frac{\chi}{2}$, χ *is defined in Proposition 1, and* $\{\alpha_k\}_{k \geq 1}$ *is non-decreasing such that* $0 \leq \alpha_k \leq \alpha < 1$. *Assume that the sequence* $\{\lambda_k\}_{k \geq 1}$ *fulfils the condition*

$$0 < \lim_{k \to +\infty} \lambda_k = \lambda < \frac{2(1 + \gamma\|L\|) - \varepsilon}{(1 + \gamma\|L\|)^2} \frac{(1 - \alpha)^2}{(2\alpha^2 - \alpha + 1)},$$

where $\varepsilon = \frac{2}{1 + \sqrt{1 + 16\beta^2\|L\|^2}}$. *Therefore, the iterative sequence* $\{x_k\}_{k \geq 1}$ *generated by (28) weakly converges to a solution of* $\mathrm{zer}(A + L^* BL + C)$.

Proof. Using Proposition 2.7 in [39], we observe that M is maximal monotone and S is Lipschitz continuous together with $\|L\|$. Considering the β-cocoercivity of C, it follows that N is also β-cocoercive. Additionally, for arbitrary $k \geq 1$, let

$$\begin{cases} x_k = (x_k, y_k), \\ w_k = (w_{1,k}, w_{2,k}), \\ z_k = (z_{1,k}, z_{2,k}), \\ t_k = (t_{1,k}, t_{2,k}). \end{cases}$$

Therefore, using (27) and Proposition 2.7 (iv) in [39], we can rewrite (28) in the following form:

$$\begin{cases} w_k = x_k + \alpha_k(x_k - x_{k-1}), \\ z_k = J_{\gamma_k M}(w_k - \gamma_k(S + N)w_k), \\ t_k = z_k + \gamma_k(S w_k - S z_k), \\ x_{k+1} = (1 - \lambda_k)w_k + \lambda_k t_k, \end{cases}$$

which has the same structure as (6). Meanwhile, our assumptions guarantee that the conditions of Theorem 1 are held. Hence, according to Theorem 1, the sequence $\{x_k\}_{k \geq 1}$ generated by (28) weakly converges to an element of $\mathrm{zer}(A + L^* BL + C)$. □

In the following, we apply the results of Corollary 1 to tackle the convex minimization problem.

Corollary 2. *Consider the convex optimization problem as follows:*

$$\min_{x \in \mathcal{H}} f(x) + g(Lx) + h(x), \qquad (29)$$

where $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, $h : \mathcal{H} \to \mathbb{R}$ *is convex differentiable with* $\frac{1}{\beta}$-*Lipschitz continuous gradient for some* $\beta > 0$, *and* $L : \mathcal{H} \to \mathcal{G}$ *is a bounded linear operator. For (29), given initial data* $x_0, x_1 \in \mathcal{H}$ *and* $y_0, y_1 \in \mathcal{G}$, *iteration sequences are presented for* $k \geq 1$:

$$\left|\begin{aligned}
&w_{1,k} = x_k + \alpha_k(x_k - x_{k-1}), \\
&w_{2,k} = y_k + \alpha_k(y_k - y_{k-1}), \\
&z_{1,k} = \text{prox}_{\gamma_k f}(w_{1,k} - \gamma_k(L^* w_{2,k} + \nabla h(w_{1,k}))), \\
&z_{2,k} = \text{prox}_{\gamma_k g^*}(w_{2,k} + \gamma_k L w_{1,k}), \\
&t_{1,k} = z_{1,k} + \gamma_k(L^* w_{2,k} - L^* z_{2,k}), \\
&t_{2,k} = z_{2,k} + \gamma_k(-L w_{1,k} + L z_{1,k}), \\
&x_{k+1} = (1 - \lambda_k)w_{1,k} + \lambda_k t_{1,k}, \\
&y_{k+1} = (1 - \lambda_k)w_{2,k} + \lambda_k t_{2,k},
\end{aligned}\right. \tag{30}$$

where $\eta \le \gamma_k \le \chi - \eta$, $0 < \eta < \frac{\chi}{2}$, χ is defined in Proposition 1, and $\{\alpha_k\}_{k\ge 1}$ is nondecreasing such that $0 \le \alpha_k \le \alpha < 1$. Assume that the sequence $\{\lambda_k\}_{k\ge 1}$ satisfies that

$$0 < \lim_{k \to +\infty} \lambda_k = \lambda < \frac{2(1 + \gamma\|L\|) - \varepsilon}{(1 + \gamma\|L\|)^2} \frac{(1 - \alpha)^2}{(2\alpha^2 - \alpha + 1)},$$

where $\varepsilon = \frac{2}{1 + \sqrt{1 + 16\beta^2\|L\|^2}}$. If $\text{zer}(\partial f + L^* \circ \partial g \circ L + \nabla h)$ is nonempty, then the sequence $\{x_k\}_{k\ge 1}$ weakly converges to a minimizer of (29).

Proof. According to [4], ∂f and ∂g are maximal monotone. In view of the Baillon–Haddad theorem, ∇h is indicated to be β-cocoercive. Thus, solving (29) with our algorithm is equivalent to (25) under suitable qualification conditions, which gives

$$A = \partial f, \quad B = \partial g \quad \text{and} \quad C = \nabla h.$$

Therefore, it follows from the same arguments as the proof of Corollary 1 that we arrive at the conclusions of Corollary 2. □

5. Numerical Experiments

This section reports the feasibility and efficiency of (6). In particular, we discuss the impact of parameters on (6). All experiments were conducted using MATLAB on a standard Lenovo machine equipped with an Intel(R) Core(TM) i5-8265U CPU @ 1.60 GHz with 1.80 GHz boost. Our objective is to address the following constrained total variation (TV) minimization problem:

$$\min_{z \in D} \frac{1}{2}\|Az - d\|_2^2 + \mu\|z\|_{TV}, \tag{31}$$

in which $A \in \mathbb{R}^{m \times n}$ represents the blurring matrix, z signifies the unknown original image in $\mathbb{R}^{n^2}$, and D indicates a nonempty closed convex set and represents the prior information regarding the deblurred images. Specifically, D is selected as a nonnegative constraint set, $\mu > 0$ indicates the regularization parameter, $\|z\|_{TV}$ denotes the total variation term, and d stands for the recorded blurred image data.

Notice that (31) can be equivalently written with the following structure:

$$\min_z \frac{1}{2}\|Az - d\|_2^2 + \mu\|z\|_{TV} + \delta_D(z), \tag{32}$$

where $\delta_D(z)$ represents the indicator function, which equals 0 when $z \in D$ and $+\infty$ otherwise. The term $\|z\|_{TV}$ can be expressed as a combination of a convex function φ (either using $\|\cdot\|_1$ for the anisotropic total-variation or $\|\cdot\|_2$ for the isotropic total-variation) with a first-order difference matrix H, denoted as $\|z\|_{TV} = \varphi(Hz)$ (refer to Section 4 in [41]), where H represents a $2n^2 \times n^2$ matrix written as:

$$H := \begin{bmatrix} I_n \otimes E \\ E \otimes I_n \end{bmatrix} \quad \text{and} \quad E := \begin{bmatrix} 0 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}_{n \times n},$$

where I_n denotes the $n \times n$ identity matrix, and $\otimes$ is the Kronecker product. Consequently, it is evident that (32) constitutes a special instance of (29).

In order to assess the quality of the deblurred images, we employ the signal-to-noise ratio (SNR) [40], which is defined by

$$SNR = 10 \log \frac{\|x\|^2}{\|x - x_r\|^2},$$

where x is the original image, and x_r is the deblurred image. The following stopping criterion is utilized to terminate iterative algorithms by the relative error between adjacent iterative steps. We choose test image "Barbara" with a size of 512×512 and use four typical image blurring scenarios as in Table 2. In addition, the range of the values in the original images is $[0, 255]$, and the norm of operator A in (31) is equal to 1 for the selected experiments. The cocoercive coefficient β is 1, and $\|H\|^2 = 8$ for the total variation, as estimated in [41], where H is the linear operator. We terminate the iterative algorithms if the relative error is less than 5×10^{-4} or if the maximum number of iterations reaches 1000.

Table 2. Description of image blurring scenarios.

Scenario	Blur Kernel	Gaussian Noise
1	9×9 box average kernel	$\sigma = 1.5$
2	9×9 box average kernel	$\sigma = 3$
3	7×7 Gaussian kernel with $\sigma_a = 10$	$\sigma = 1.5$
4	7×7 Gaussian kernel with $\sigma_a = 10$	$\sigma = 3$

Prior to the comparisons, we do a test to display how the performance of RIFBHF is affected by different parameters. For simplicity, we set $\gamma_k = \gamma$ for all $k \geq 1$. In view of (19), the relationship between inertial parameter α and relaxation parameter λ is presented as follows:

$$0 < \lim_{k \to +\infty} \lambda_k = \lambda < \frac{2(1 + \gamma \|L\|) - \varepsilon}{(1 + \gamma \|L\|)^2} \frac{(1 - \alpha)^2}{(2\alpha^2 - \alpha + 1)},$$

where $0 \leq \alpha < 1$, and $\varepsilon = \frac{2}{1 + \sqrt{1 + 16\beta^2 \|L\|^2}}$. As we know, the upper boundedness of λ is similar to the one of ([36], Remark 2) if $\varepsilon \searrow 0$. Firstly, we assume $\gamma = \frac{4\beta}{1 + \sqrt{1 + 16\beta^2 L^2}}$, $\alpha = 0$, and $\lambda = 1$, or $\alpha = 0.3$ and $\lambda = 0.6$ for RIFBHF. The SNR values and the numbers of iterations used with various μ for RIFBHF are recorded in Tables 3 and 4. Observe that the SNR values in image blurring scenarios 1 and 2 are the best when $\mu = 1$, while the SNR values in image blurring scenarios 3 and 4 are the best when $\mu = 0.1$. Therefore, we choose $\mu = 1$ for image blurring scenarios 1 and 2 and $\mu = 0.1$ for image blurring scenarios 3 and 4. For the case of $\mu = 1$ in image blurring scenario 1, we further study the effect of the other parameters. The development of the SNR value and the normalized error of the objective function $\log(f - f^*)/f^*$ along the running time are considered; here, f represents the present objective value, while f^* signifies the optimal objective value. To obtain an approximation of the optimal objective value, we set the function value given by running experimental algorithms for 5000 iterations as an estimate of the optimal value. One can observe that a larger γ results in better error when $\alpha = 0.2$ and $\lambda = 0.7$, and a larger α also brings better error when $\gamma = \frac{0.2}{\sqrt{8}}$ and $\lambda = 0.7$ in Figure 2. Of course, α and γ also affect the value of λ. Meanwhile, a conclusion that a larger λ allows for better error can be given. It is worth noting that over-relaxation ($\lambda > 1$) exists, and it enjoys a better

effect. Figure 3 shows the development of SNR for different parameters; the experiment results are similar to those in Figure 2.

Table 3. SNR values and iterations when $\gamma = \frac{4}{1+\sqrt{1+16*8}}$, $\alpha = 0$, and $\lambda = 1$ with different parameter μ for Barbara image in four blurred scenarios.

Scenario	1		2		3		4	
μ	SNR (dB)	Iter	SNR (dB)	Iter	SNR (dB)	Iter	SNR (dB)	Iter
0.1	17.5358	50	17.5144	52	17.9741	45	17.9510	48
0.5	17.5414	55	17.5196	56	17.9385	49	17.9148	51
1	17.5515	61	17.5304	62	17.9071	54	17.8799	55
2	17.4859	65	17.4658	65	17.8012	56	17.7788	57
3	17.4001	65	17.3831	65	17.7126	56	17.6937	57
4	17.3251	65	17.3106	66	17.6368	57	17.6203	57
5	17.2584	67	17.2443	68	17.5713	58	17.5553	59
6	17.1948	69	17.1823	69	17.5117	60	17.4972	60
7	17.1344	71	17.1230	71	17.4563	61	17.4422	62
8	17.0779	73	17.0663	74	17.4033	63	17.3895	64
9	17.0237	75	17.0123	76	17.3530	65	17.3413	65
10	16.9718	77	16.9608	78	17.3049	67	17.2940	67
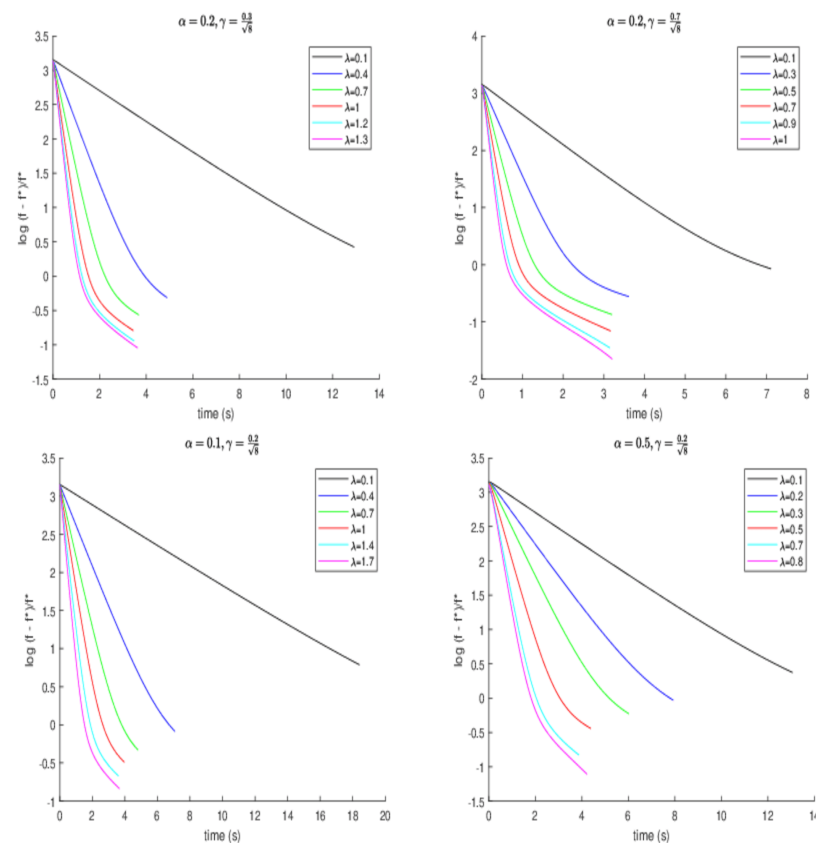
Figure 2. Behaviour of the error of objective value $\log(f - f^*)/f^*$ against running time for different parameters α, γ, and λ.

Table 4. SNR values and iterations when $\gamma = \frac{4}{1+\sqrt{1+16*8}}$, $\alpha = 0.3$, and $\lambda = 0.6$ with different parameter μ for Barbara image in four blurred scenarios.

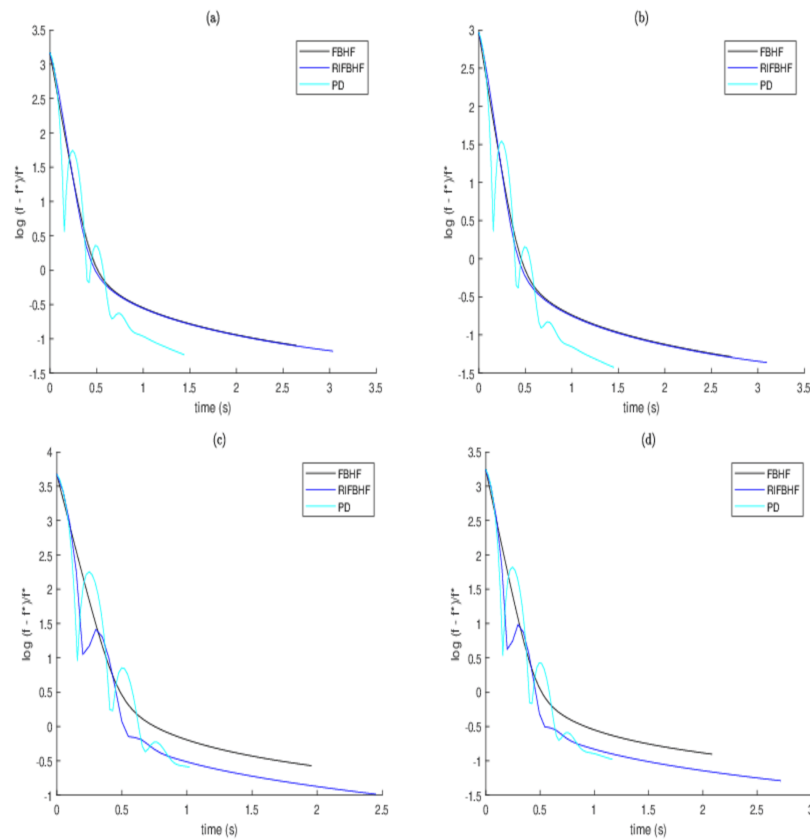| μ	Scenario	1		2		3		4	
		SNR (dB)	*Iter*	*SNR* (dB)	*Iter*	*SNR* (dB)	*Iter*	*SNR* (dB)	*Iter*
	0.1	17.4757	52	17.4513	53	17.9109	46	17.8957	49
	0.5	17.4740	55	17.4573	56	17.8890	50	17.8643	51
	1	17.5060	62	17.4884	63	17.8783	56	17.8545	57
	2	17.4687	68	17.4502	68	17.7913	59	17.7707	60
	3	17.3929	69	17.3785	70	17.7093	60	17.6920	61
	4	17.3214	70	17.3072	70	17.6367	61	17.6212	61
	5	17.2577	72	17.2446	72	17.5723	62	17.5582	63
	6	17.1970	74	17.1850	75	17.5144	64	17.5007	65
	7	17.1384	76	17.1277	76	17.4600	66	17.4480	66
	8	17.0832	78	17.0725	79	17.4078	68	17.3953	69
	9	17.0298	81	17.0203	81	17.3580	70	17.3463	71
	10	16.9787	83	16.9686	84	17.3103	72	17.3005	72
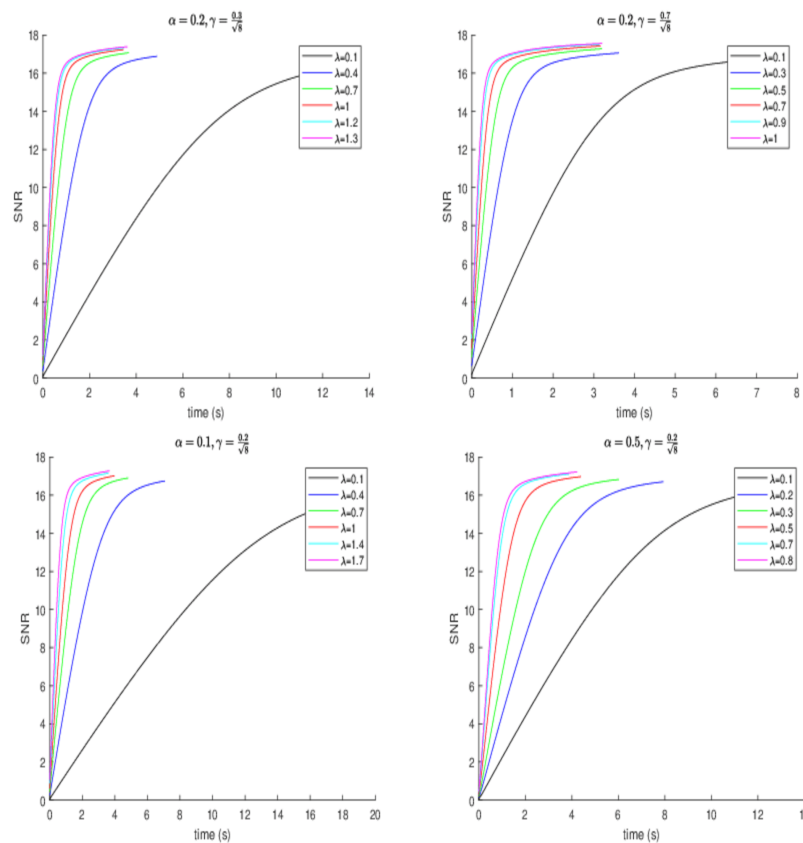
Figure 3. Behaviour of SNR against running time for different parameters α, γ, and λ.

To further validate the rationality and efficiency of (30), the following algorithms and parameter settings are utilized:

- FBHF [15]: $\gamma = \frac{4\beta}{1+\sqrt{1+16\beta^2 L^2}}$ for four image blurring scenarios.
- PD: the first-order primal–dual splitting algorithm [42] with $\tau = \frac{1}{3}$, $\sigma = \frac{1}{3}$, and $\theta = 1$ for four image blurring scenarios.
- RIFBHF: $\gamma = \frac{4\beta}{1+\sqrt{1+16\beta^2 L^2}}$ using $\alpha = 0.2$ and $\lambda = 0.9$ for image blurring scenarios 1 and 2 and $\alpha = 0.6$ and $\lambda = 0.8$ for image blurring scenarios 3 and 4.

Figure 4 plots the normalized error of the objective function of FBHF, PD, and RIFBHF along the running time. Note that PD appears to be the fastest algorithm. FBHF and RIFBHF are almost the same in image blurring scenarios 1 and 2, while in image blurring scenarios 3 and 4, the effect of RIFBHF is better than that of FBHF, which shows that our algorithm is acceptable. Meanwhile, to succinctly illustrate the deblurring impact of the proposed algorithm, the deblurred results of image blurring scenario 4 are shown in Figure 5. One can observe visually the better deblurred images generated by RIFBHF.

Figure 4. Behaviour of the error of objective value $\log(f - f^*)/f^*$ against running time for different image burring scenarios, i.e., (**a**) scenario 1, (**b**) scenario 2, (**c**) scenario 3, and (**d**) scenario 4.

Figure 5. (**a**) The original image of Barbara. (**b**) The blurred image of Barbara. (**c**) The deblurred image by FBHF. (**d**) The deblurred image by RIFBHF.

6. Conclusions

In this paper, we proposed an RIFBHF algorithm to solve (1). On the way, the proposed approach was deduced by discretising a continuous-time dynamical system, and a variable stepsize was introduced into the proposed algorithm. Additionally, we studied the theoretical convergence properties of (6) under reasonable parameter conditions. Inspired by the primal–dual scheme, our approach tackled both the composite monotone inclusion problem (25) and analysed the composite convex optimization problem (29). Subsequently, we conducted numerical experiments focused on image deblurring to illustrate the effectiveness of our proposed technique.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Daubechies, I.; Defrise, M.; De Mol, C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.* **2004**, *57*, 1413—1457.
2. Combettes, P.L.; Wajs, V.R. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **2005**, *4*, 1168—1200.
3. Raguet, H.; Fadili, J.; Peyré, G. A generalized forward-backward splitting. *SIAM J. Imaging Sci.* **2013**, *6*, 1199–1226.
4. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*; Springer: London, UK, 2017.
5. Goldstein, A.A. Convex programming in Hilbert space. *Bull. Am. Math. Soc.* **1964**, *70*, 709–710.
6. Lions, P.-L.; Mercier, B. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **1979**, *16*, 964–979.
7. Combettes, P.L.; Pesquet, J.-C. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE J. Sel. Top. Signal Process.* **2007**, *1*, 564–574.
8. Tseng, P. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.* **2000**, *38*, 431–446.
9. Raguet, H.; Landrieu, L. Preconditioning of a generalized forward-backward splitting and application to optimization on graphs. *SIAM J. Imaging Sci.* **2015**, *8*, 2706–2739.
10. Briceño-Arias, L.M. Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions. *Optimization* **2015**, *64*, 1239–1261.
11. Davis, D.; Yin, W.T. A three-operator splitting scheme and its optimization applications. *Set-Valued Var. Anal.* **2017**, *25*, 829–858.
12. Raguet, H. A note on the forward-Douglas-Rachford splitting for monotone inclusion and convex optimization. *Optim. Lett.* **2019**, *13*, 717–740.
13. Ryu, E.K.; Vũ, B.C. Finding the forward-Douglas-Rachford-forward method. *J. Optim. Theory Appl.* **2020**, *184*, 858–876.
14. Rieger, J.; Tam, M.K. Backward-forward-reflected-backward splitting for three operator monotone inclusions. *Appl. Math. Comput.* **2020**, *381*, 125248.
15. Briceño-Arias, L.M.; Davis, D. Forward-backward-half forward algorithm for solving monotone inclusions. *SIAM J. Optim.* **2018**, *28*, 2839–2871.
16. Alves, M.M.; Geremia, M. Iteration complexity of an inexact Douglas-Rachford method and of a Douglas-Rachford-Tseng's F-B four-operator splitting method for solving monotone inclusions. *Numer. Algorithms* **2019**, *82*, 263–295.
17. Giselsson, P. Nonlinear forward-backward splitting with projection correction. *SIAM J. Optim.* **2021**, *31*, 2199–2226.
18. Briceño-Arias, L.; Chen, J.; Roldán, F.; Tang, Y. Forward-partial inverse-half-forward splitting algorithm for solving monotone inclusions. *Set-Valued Var. Anal.* **2022**, *30*, 1485–1502.
19. Polyak, B.T. Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.* **1964**, *4*, 1–17.
20. Nesterov, Y. A method for unconstrained convex minimization problem with the rate of convergence $O(\frac{1}{k^2})$. *Doklady AN SSSR* **1983**, *269*, 543–547.
21. Güler, O. New proximal point algorithms for convex minimization. *SIAM J. Optim.* **1992**, *2*, 649–664.
22. Alvarez, F.; Attouch, H. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.* **2001**, *9*, 3–11.
23. Alvarez, F. Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space. *SIAM J. Optim.* **2003**, *14*, 773–782.
24. Ochs, P.; Chen, Y.; Brox, T.; Pock, T. iPiano: Inertial proximal algorithm for nonconvex optimization. *SIAM J. Imaging Sci.* **2014**, *7*, 1388–1419.
25. Boţ, R.I.; Csetnek, E.R. A hybrid proximal-extragradient algorithm with inertial effects. *Numer. Funct. Anal. Optim.* **2015**, *36*, 951–963.
26. Chen, C.; Chan, R.H.; Ma, S.; Yang, J. Inertial proximal ADMM for linearly constrained separable convex optimization. *SIAM J. Imaging Sci.* **2015** *8*, 2239–2267.
27. Dong, Q.; Lu, Y.; Yang, J. The extragradient algorithm with inertial effects for solving the variational inequality. *Optimization* **2016**, *65*, 2217–2226
28. 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.* **2017**, *27*, 2356–2380.
29. Attouch, H.; Cabot, A. Convergence rates of inertial forward-backward algorithms. *SIAM J. Optim.* **2018**, *28*, 849–874.
30. Eckstein, J.; Bertsekas, D.P. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* **1992**, *55*, 293–318.
31. Attouch, H.; Peypouquet, J. Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. *Math. Program.* **2019**, *174*, 391–432.
32. Attouch, H.; Cabot, A. Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. *Math. Program.* **2020**, *184*, 243–287.
33. Attouch, H.; Cabot, A. Convergence rate of a relaxed inertial proximal algorithm for convex minimization. *Optimization* **2020**, *69*, 1281–1312.

34. Attouch, H.; Cabot, A. Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions. *Appl. Math. Optim.* **2019**, *80*, 547–598.

35. Boţ, R.I.; Csetnek, E.R.; Hendrich, C. Inertial Douglas-Rachford splitting for monotone inclusion problems. *Appl. Math. Comput.* **2015**, *256*, 472–487.

36. 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.* **2023**, 1–37.

37. Tang, Y.; Yang, Y.; Peng, J. Convergence analysis of a relaxed inertial alternating minimization algorithm with applications. In *Advanced Mathematical Analysis and Its Applications*, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2023; 27p.

38. Boţ R.I.; Csetnek, E.R. Second order forward-backward dynamical systems for monotone inclusion problems. *SIAM J. Control Optim.* **2016**, *54*, 1423–1443.

39. Briceño-Arias, L.M.; Combettes, P.L. A monotone+skew splitting model for composite monotone inclusions in duality. *SIAM J. Optim.* **2011**, *21*, 1230–1250.

40. Zong, C.; Tang, Y.; Zhang, G. An accelerated forward-backward-half forward splitting algorithm for monotone inclusion with applications to image restoration. *Optimization* **2024**, *73*, 401–428.

41. Micchelli, C.A.; Shen, L.; Xu, Y. Proximity algorithms for image models: Denoising. *Inverse Problems* **2011**, *27*, 045009.

42. Chambolle, A.; Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision* **2011**, *40*, 120–145.