Unified Acceleration Methods for Solving Monotone Inclusions

Chris Junchi Li[◇]

Department of Electrical Engineering and Computer Sciences[◇]
University of California, Berkeley

September 27, 2024

Abstract

This paper presents an advanced analysis of iterative methods for solving monotone inclusions, a fundamental problem in optimization and variational inequality theory. We explore a range of classical and modern techniques, including the proximal-point method, forward-backward splitting, and Douglas-Rachford splitting, providing enhanced convergence guarantees through acceleration. By establishing a unified framework, we demonstrate the mathematical equivalence between Halpern's fixed-point iteration and Nesterov's accelerated gradient methods, showing how both can be adapted to solve maximally monotone operator problems. Our results extend the applicability of these methods, improving convergence rates for monotone operator problems in finite-dimensional spaces. Additionally, we provide new insights into the solution structure of monotone inclusions, offering practical guidelines for implementation in various optimization contexts.

Keywords: Monotone Inclusions, Maximally Monotone Operators, Proximal-Point Method, Forward-Backward Splitting, Halpern Iteration, Nesterov Acceleration, Variational Inequalities

1 Introduction

Monotone inclusions are a central problem in optimization, variational inequalities, nonlinear analysis, and machine learning. They involve finding a point in the zero set of a maximally monotone operator, which arises in diverse applications such as equilibrium modeling, signal processing, and game theory. The theory of monotone operators provides a powerful framework for addressing these challenges, particularly when dealing with large-scale or structured datasets. Classical methods for solving monotone inclusions, including the proximal-point method, forward-backward splitting, and Douglas-Rachford splitting, rely on operator splitting techniques to approximate solutions. While these methods are widely used for their simplicity and robustness, their convergence can be slow in practice.

Recent advances in iterative algorithms, such as Halpern's fixed-point iteration and Nesterov's accelerated gradient method, have demonstrated that acceleration techniques can substantially improve convergence rates. In this paper, we present a unified framework that bridges these two prominent methods, showing that they share underlying mechanics despite originating from different contexts. By analyzing the specific parameter settings that link Halpern's and Nesterov's methods, we establish their mathematical equivalence.

This unified perspective not only elucidates the common foundations of these acceleration techniques but also extends their applicability to a broader class of optimization schemes, including the proximal-point method, forward-backward splitting, and Douglas-Rachford splitting. Our framework offers a practical tool for enhancing the efficiency of solving monotone operator problems across various domains.

Background. Approximating a solution of a maximally monotone inclusion is a fundamental problem in optimization, nonlinear analysis, mechanics, and machine learning, among many other areas, see, e.g., [5, 9, 19, 37, 40, 41, 42]. This problem lies at the heart of monotone operator theory, and has been intensively studied in the literature for several decades, see, e.g., [5, 27, 41, 42] as a few examples. Various numerical methods, including gradient/forward, extragradient, past-extragradient, proximal-point, and their variants have been proposed to solve this problem, its extensions, and its special cases [5, 19]. When the underlying operator is the sum of two or multipe maximally monotone operators, forward-backward splitting, forward-backward-forward splitting, Douglas-Rachford splitting, projective splitting methods, and their variants have been extensively developed for approximating its solutions under different assumptions and context, see, e.g., [5, 15, 16, 27, 30, 39, 46].

Contribution. The main contributions of this paper are two-fold: (i) We extend our analysis to classical optimization schemes, including proximal-point, forward-backward splitting, and Douglas-Rachford splitting, introducing acceleration techniques that significantly improve their convergence rates. (ii) We establish a novel connection between Halpern's fixed-point iteration and Nesterov's accelerated method, demonstrating their equivalence in the context of monotone inclusions, thus offering a new perspective on the acceleration of iterative algorithms for these problems.

The main contributions of this paper are two-folded:

(i) We extend our analysis to classical optimization schemes, including proximal-point, forward-backward splitting, Douglas-Rachford splitting, and three-operator splitting methods. We introduce acceleration techniques that significantly improve their convergence rates, making these accelerated versions more efficient than their classical counterparts.

(ii) We demonstrate that Halpern's fixed-point iteration and Nesterov's accelerated method are mathematically equivalent under specific parameter configurations, establishing a novel connection between these two acceleration techniques. Our results show that the iterates generated by both schemes are identical under certain parameter choices, though different parameters lead to varying convergence guarantees. This unification enables us to derive the convergence of one scheme from the other.

Organization. The rest of the paper is organized as follows. Section 2 provides the necessary preliminaries on monotone operators and the mathematical background required for our results. Section 3 presents the main results on accelerating classical algorithms for monotone inclusions. In Section 4, we discuss the equivalence between Halpern's and Nesterov's methods and their application to monotone inclusions. Section 5 concludes the paper with a discussion of potential future research directions.

2 Preliminaries

This section recalls some necessary notation and concepts which will be used in the sequel. We work with finite dimensional spaces $\mathbb{R}^p$ and $\mathbb{R}^n$ equipped with the standard inner product $\langle \cdot, \cdot \rangle$ and Euclidean norm $\| \cdot \|$. For a set-valued mapping $\mathbf{F} : \mathbb{R}^p \rightrightarrows 2^{\mathbb{R}^p}$, $\mathrm{dom}(\mathbf{F}) = \{x \in \mathbb{R}^p : \mathbf{F}x \neq \emptyset\}$ denotes its domain, $\mathrm{graph}(\mathbf{F}) = \{(x, y) \in \mathbb{R}^p \times \mathbb{R}^p : y \in \mathbf{F}x\}$ denotes its graph, where $2^{\mathbb{R}^p}$ is the set of all subsets of $\mathbb{R}^p$. The inverse of $\mathbf{F}$ is defined as $\mathbf{F}^{-1}y := \{x \in \mathbb{R}^p : y \in \mathbf{F}x\}$. Throughout this paper, we work with finite dimensional spaces; however, we believe that our result can be easily extended to Hilbert spaces.

Monotonicity. For a set-valued mapping $\mathbf{F} : \mathbb{R}^p \rightrightarrows 2^{\mathbb{R}^p}$, we say that $\mathbf{F}$ is monotone if $\langle u - v, x - y\rangle \geq 0$ for all $x, y \in \mathrm{dom}(\mathbf{F})$, $u \in \mathbf{F}x$, and $v \in \mathbf{F}y$. $\mathbf{F}$ is said to be $\mu_{\mathbf{F}}$-strongly monotone (or sometimes called coercive) if $\langle u - v, x - y\rangle \geq \mu_{\mathbf{F}}\|x - y\|^2$ for all $x, y \in \mathrm{dom}(\mathbf{F})$, $u \in \mathbf{F}x$, and $v \in \mathbf{F}y$, where $\mu_{\mathbf{F}} > 0$ is called a strong monotonicity parameter. If $\mu_{\mathbf{F}} < 0$, then we say that $\mathbf{F}$ is weakly monotone (also known as $-\mu_{\mathbf{F}}$-hypomonotone). If $\mathbf{F}$ is single-valued, then these conditions reduce to $\langle \mathbf{F}x - \mathbf{F}y, x - y\rangle \geq 0$ and $\langle \mathbf{F}x - \mathbf{F}y, x - y\rangle \geq \mu_{\mathbf{F}}\|x - y\|^2$ for all $x, y \in \mathrm{dom}(\mathbf{F})$, respectively. We say that $\mathbf{F}$ is maximally monotone if $\mathrm{graph}(\mathbf{F})$ is not properly contained in the graph of any other monotone operator. Note that $\mathbf{F}$ is maximally monotone, then $\alpha\mathbf{F}$ is also maximally monotone for any $\alpha > 0$, and if $\mathbf{F}$ and $\mathbf{H}$ are maximally monotone, and $\mathrm{dom}(F) \cap \mathrm{int}\,(\mathrm{dom}(H)) \neq \emptyset$, then $\mathbf{F} + \mathbf{H}$ is maximally monotone.

Lipschitz continuity and co-coerciveness. A single-valued operator $\mathbf{F}$ is said to be L-Lipschitz continuous if $\|\mathbf{F}x - \mathbf{F}y\| \leq L\|x - y\|$ for all $x, y \in \mathrm{dom}(\mathbf{F})$, where $L \geq 0$ is a Lipschitz constant. If $L = 1$, then we say that $\mathbf{F}$ is nonexpansive, while if $L \in [0, 1)$, then we say that $\mathbf{F}$ is L-contractive, and L is its contraction factor. We say that $\mathbf{F}$ is $\frac{1}{L}$-co-coercive if $\langle \mathbf{F}x - \mathbf{F}y, x - y\rangle \geq \frac{1}{L}\|\mathbf{F}x - \mathbf{F}y\|^2$ for all $x, y \in \mathrm{dom}(\mathbf{F})$. If $L = 1$, then we say that $\mathbf{F}$ is firmly nonexpansive. Note that if $\mathbf{F}$ is $\frac{1}{L}$-cocoercive, then it is also monotone and L-Lipschitz continuous (by using the Cauchy-Schwarz inequality), but the reverse statement is not true in general. If $L < 0$, then we say that $\mathbf{F}$ is $\frac{1}{L}$-co-monotone [6] (also known as $-\frac{1}{L}$-cohypomonotone).

Resolvent operator. The operator $J_{\mathbf{F}}x := \{y \in \mathbb{R}^p : x \in y + \mathbf{F}y\}$ is called the resolvent of $\mathbf{F}$, often denoted by $J_{\mathbf{F}}x = (\mathbf{I} + \mathbf{F})^{-1}x$, where $\mathbf{I}$ is the identity mapping. Clearly, evaluating $J_{\mathbf{F}}$ requires solving a strongly monotone inclusion $0 \in y - x + \mathbf{F}y$. If $\mathbf{F}$ is monotone, then $J_{\mathbf{F}}$ is singled-valued, and if $\mathbf{F}$ is maximally monotone then $J_{\mathbf{F}}$ is singled-valued and $\mathrm{dom}(J_{\mathbf{F}}) = \mathbb{R}^p$. If $\mathbf{F}$ is monotone, then $J_{\mathbf{F}}$ is firmly nonexpansive [5, Proposition 23.10].

3 Monotone Inclusions and Solution Characterization, with Applications

We are interested in the following monotone inclusion and its special cases:

$$0 \in \mathbf{A}y^\star + \mathbf{B}y^\star + \mathbf{C}y^\star \tag{MI}$$

where

- $\mathbf{A}, \mathbf{B} : \mathbb{R}^p \rightrightarrows 2^{\mathbb{R}^p}$ are multivalued and maximally monotone operators, and
- $\mathbf{C} : \mathbb{R}^p \to \mathbb{R}^p$ is a $\frac{1}{L}$-co-coercive operator.

Let $\mathbf{M} := \mathbf{A} + \mathbf{B} + \mathbf{C}$ and we assume that $\mathrm{zer}(\mathbf{M}) := \{y^\star \in \mathbb{R}^p : 0 \in \mathbf{A}y^\star + \mathbf{B}y^\star + \mathbf{C}y^\star\} = \mathbf{M}^{-1}(0)$ is nonempty. We first presents some

Preliminary results In order to characterize solutions of (MI), we recall the following two operators. The first operator is the forward-backward residual mapping with $\mathbf{C} = 0$, in which case (MI) reduces to $0 \in \mathbf{A}y^\star + \mathbf{B}y^\star$:

$$G_{\lambda\mathbf{M}}y := \frac{1}{\lambda}\left(y - J_{\lambda\mathbf{B}}(y - \lambda\mathbf{A}y)\right) \tag{1}$$

where $\mathbf{A}$ is single-valued, $\mathbf{M} := \mathbf{A} + \mathbf{B}$, and $J_{\lambda\mathbf{B}}$ is the resolvent of $\lambda\mathbf{B}$ for any $\lambda > 0$. Notice that if $0 < \lambda < \frac{4}{L}$, then $G_{\lambda\mathbf{M}}$ is $\frac{\lambda(4-\lambda L)}{4}$-cocoercive, see, e.g., [5, Proposition 26.1] or a short proof in Proposition 1 below. It is obvious that $y^\star$ is a solution of (MI) if and only if $G_{\lambda\mathbf{M}}y^\star = 0$. Hence, solving (MI) is equivalent to solving the co-coercive equation $G_{\lambda\mathbf{M}}y^\star = 0$ as in (CoCo).

Proposition 1. *Let* $\mathbf{B}$ *and* $\mathbf{A}$ *in* (MI) *be maximally monotone,* $\mathbf{A}$ *be single-valued, and* $\mathbf{C} = 0$*. Let* $G_{\lambda\mathbf{M}}$ *be the forward-backward residual mapping defined by* (1)*. Then, we have*

$$\langle G_{\lambda\mathbf{M}}x - G_{\lambda\mathbf{M}}y, x - y + \lambda(\mathbf{A}x - \mathbf{A}y)\rangle \geq \lambda\|G_{\lambda\mathbf{M}}x - G_{\lambda\mathbf{M}}y\|^2 + \langle \mathbf{A}x - \mathbf{A}y, x - y\rangle \tag{2}$$

Moreover, $G_{\lambda\mathbf{M}}y^\star = 0$ *if and only if* $y^\star \in \mathrm{zer}(\mathbf{A} + \mathbf{B})$*. If, additionally,* $\mathbf{A}$ *is* $\frac{1}{L}$*-co-coercive, then* $G_{\lambda\mathbf{M}}$ *is* $\frac{\lambda(4-\lambda L)}{4}$*-co-coercive provided that* $0 < \lambda < \frac{4}{L}$*.*

The second operator is the following residual mapping of a three-operator splitting scheme:

$$E_{\lambda\mathbf{M}}y := \frac{1}{\lambda}(J_{\lambda\mathbf{A}}y - J_{\lambda\mathbf{B}}(2J_{\lambda\mathbf{A}}y - y - \lambda\mathbf{C} \circ J_{\lambda\mathbf{A}}y)) \tag{3}$$

where $J_{\lambda\mathbf{B}}$ and $J_{\lambda\mathbf{A}}$ are the resolvents of $\lambda\mathbf{B}$ and $\lambda\mathbf{A}$, respectively, and $\circ$ is a composition operator. As proved in Proposition 2 below (see also [5, 16, 21]) that $y^\star$ is a solution of (MI) if and only if $E_{\lambda\mathbf{M}}y^\star = 0$. Moreover, if $0 < \lambda < \frac{4}{L}$, then $E_{\lambda\mathbf{M}}$ is $\frac{\lambda(4-L\lambda)}{4}$-co-coercive. Hence, solving (MI) is equivalent to solving the co-coercive equation $E_{\lambda\mathbf{M}}y^\star = 0$ as stated in (CoCo) below.

Proposition 2. *Let* $\mathbf{B}$ *and* $\mathbf{A}$ *in* (MI) *be maximally monotone, and* $\mathbf{C}$ *be* $\frac{1}{L}$*-co-coercive. Let* $E_{\lambda\mathbf{M}}$ *be the residual mapping defined by* (3)*. Then,* $E_{\lambda\mathbf{M}}u^\star = 0$ *iff* $y^\star \in \mathrm{zer}(\mathbf{A} + \mathbf{B} + \mathbf{C})$*, where* $y^\star = J_{\lambda\mathbf{A}}u^\star$*. Moreover, we have* $E_{\lambda\mathbf{M}}$ *satisfies the following property for all* u *and* v*:*

$$\langle E_{\lambda\mathbf{M}}u - E_{\lambda\mathbf{M}}v, u - v\rangle \geq \frac{\lambda(4 - L\lambda)}{4}\|E_{\lambda\mathbf{M}}u - E_{\lambda\mathbf{M}}v\|^2 \tag{4}$$

If $\mathbf{A}$ *is single-valued and* $\mathbf{C} = 0$*, and* $G_{\lambda\mathbf{M}}$ *is defined by* (1)*, then we have* $E_{\lambda\mathbf{M}}u = G_{\gamma Q}y$*, where* $y = J_{\lambda\mathbf{A}}u$ *or equivalently,* $u = y + \lambda\mathbf{A}y$*.*

By means of Proposition 1 and Proposition 2, one can transform (MI) and its special cases equivalently to a co-coercive equation (CoCo).

3.1 The Halpern Fixed-Point Scheme and Its Convergence

To present our analysis framework, we consider the following co-coercive equation:

$$\text{Find } y^\star \in \mathbb{R}^p \text{ such that: } \mathbf{F}y^\star = 0 \tag{CoCo}$$

where $\mathbf{F} : \mathbb{R}^p \to \mathbb{R}^p$ is a single-valued and $\frac{1}{L}$-co-coercive operator. We denote by $\mathrm{zer}(\mathbf{F}) := \mathbf{F}^{-1}(0) = \{y^\star \in \mathbb{R}^p : \mathbf{F}y^\star = 0\}$ the solution set of (CoCo), and assume that $\mathrm{zer}(\mathbf{F})$ is nonempty.

The Halpern fixed-point scheme for solving (CoCo) can be written as follows (see [17, 20, 26]):

$$y_{k+1} := \beta_k y_0 + (1 - \beta_k)y_k - \eta_k \mathbf{F}y_k \tag{5}$$

where $\beta_k \in (0, 1)$ and $\eta_k > 0$ are appropriately chosen.

The convergence rate of (5) has been established in [17, 26] using different tools. While [26] provides a direct proof and uses a performance estimation problem approach to establish convergence of (5), [17] exploits a Lyapunov technique to analyze convergence rate. Let us summarize the result in [17] in our context.

The standard Lyapunov function to study (5) is

$$\mathcal{L}_k := \frac{p_k}{L}\|\mathbf{F}y_k\|^2 + q_k\langle \mathbf{F}y_k, y_k - y_0\rangle \tag{6}$$

where $p_k := q_0 k(k + 1)$ and $q_k := q_0(k + 1)$ (for some $q_0 > 0$) are given parameters. The following theorem states the convergence of (5).

Theorem 1 ([17, 26])**.** *Assume that* $\mathbf{F}$ *in* (CoCo) *is* $\frac{1}{L}$*-co-coercive with* $L \in (0, +\infty)$*, and* $y^\star \in \mathrm{zer}(\mathbf{F})$*. Let* $\{y_k\}$ *be generated by* (5) *using* $\beta_k := \frac{1}{k+2}$ *and* $\eta_k := \frac{2(1-\beta_k)}{L} = \frac{2(k+1)}{(k+2)L}$*. Then*

$$\|\mathbf{F}y_k\| \le \frac{L\|y_0 - y^\star\|}{k + 1} \tag{7}$$

If we choose $\beta_k := \frac{1}{k+2}$ *and* $\eta_k := \frac{(1-\beta_k)}{L}$*, then we have* $\|\mathbf{F}y_k\|^2 \le \frac{4L^2\|y_0-y^\star\|^2}{(k+1)(k+3)}$ *and* $\sum_{k=0}^{\infty}(k + 1)(k + 2)\|\mathbf{F}y_{k+1} - \mathbf{F}y_k\|^2 \le 2L^2\|y_0 - y^\star\|^2$*.*

The last statement of Theorem 1 was not proved in [17, 26], but requires a few elementary justification, and hence we omit it here. Note that if $\beta_k := \frac{1}{k+2}$, then we can rewrite (5) into the Halpern fixed-point iteration as in [26]:

$$y_{k+1} := \frac{1}{(k + 2)}y_0 + \left(1 - \frac{1}{(k + 2)}\right)\mathbf{T}y_k, \quad \text{where} \quad \mathbf{T}y_k := y_k - \frac{2}{L}\mathbf{F}y_k \tag{8}$$

Since $\mathbf{F}$ is $\frac{1}{L}$-co-coercive, $\mathbf{T} = \mathbf{I} - \frac{2}{L}\mathbf{F}$ is nonexpansive, see [5, Proposition 4.11]. Therefore, (5) is equivalent to the scheme studied in [26], and Theorem 1 can be obtained from the results in [26]. The choice of β_k and η_k are tight and the bound (7) is unimprovable since there exists an instance that achieves this rate as the lower bound, see, e.g., [26] for such an example.

We will be proving the following results in the next section:

Corollary 2. *Assume that* $\mathbf{F}$ *in* (CoCo) *is* $\frac{1}{L}$*-co-coercive and* $\mathrm{zer}(\mathbf{F}) \neq \emptyset$. *Let* $\{(x_k, y_k)\}$ *be generated by* (15) *using* $\gamma_k := \frac{1}{L}$, $\beta_k := \frac{1}{k+2}$, *and* $\eta_k := \frac{(1-\beta_k)}{L}$. *Then, we obtain* $\theta_k := \frac{k}{k+2}$, $\nu_k := \frac{k+1}{k+2}$, *and* $\kappa_k := 0$. *Moreover,* (15) *reduces to* (Nes) (*or equivalently* (16)), *and the following guarantee holds:*

$$\|\mathbf{F}y_k\|^2 \leq \frac{4L^2\|y_0 - y^\star\|^2}{(k+1)(k+3)}, \quad \textit{and} \quad \sum_{k=0}^{\infty}(k+1)(k+2)\|\mathbf{F}y_{k+1} - \mathbf{F}y_k\|^2 \leq 2L^2\|y_0 - y^\star\|^2 \tag{18}$$

If we use $\gamma_k := \frac{1}{L}$, $\beta_k := \frac{1}{k+2}$, *and* $\eta_k := \frac{2(1-\beta_k)}{L}$, *then we obtain* $\theta_k := \frac{k}{k+2}$, $\nu_k := 0$, *and* $\kappa_k := \frac{k}{k+2}$, *and* (15) *reduces to* (17). *Moreover, the following guarantee holds:* $\|\mathbf{F}y_k\| \leq \frac{L\|y_0 - y^\star\|}{(k+1)}$.

Theorem 3. *Assume that* $\mathbf{F}$ *in* (CoCo) *is* $\frac{1}{L}$*-co-coercive and* $\mathrm{zer}(\mathbf{F}) \neq \emptyset$. *Let* $\{(x_k, y_k)\}$ *be generated by* (Nes) *to solve* (CoCo) *using* $\gamma_k := \gamma \in \left(0, \frac{1}{L}\right)$, $\theta_k := \frac{k+1}{k+2\omega+2}$, *and* $\nu_k := \frac{k+\omega+2}{k+2\omega+2} \in (0, 1)$ *for a given constant* $\omega > 2$. *Then, the following estimates hold:*

$$\begin{cases} \sum_{k=0}^{\infty}(k+\omega+1)\|x_{k+1} - x_k\|^2 < +\infty \quad \textit{and} \quad \|x_{k+1} - x_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+2\omega+1)\|y_k - x_k\|^2 < +\infty \quad \textit{and} \quad \|y_k - x_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega+1)\|\mathbf{F}y_{k-1}\|^2 < +\infty \quad \textit{and} \quad \|\mathbf{F}y_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega+1)\|\mathbf{F}x_k\|^2 < +\infty \quad \textit{and} \quad \|\mathbf{F}x_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega)\|y_{k+1} - y_k\|^2 < +\infty \quad \textit{and} \quad \|y_{k+1} - y_k\|^2 = o\left(\frac{1}{k^2}\right) \end{cases} \tag{21}$$

Consequently, both $\{x_k\}$ *and* $\{y_k\}$ *converge to* $y^\star \in \mathrm{zer}(\mathbf{F})$.

Theorem 4. *Let* $\{y_k\}$ *be generated by the Halpern fixed-point iteration* (5) *using* $\beta_k := \frac{\omega+1}{k+2\omega+2}$ *and* $\eta_k := \gamma(1-\beta_k)$ *for a fixed* $\gamma \in \left(0, \frac{1}{L}\right)$ *and* $\omega > 2$. *Then, the following statements hold:*

$$\begin{cases} \sum_{k=0}^{\infty}(k+\omega+1)\|\mathbf{F}y_{k-1}\|^2 < +\infty \quad \textit{and} \quad \|\mathbf{F}y_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega)\|y_{k+1} - y_k\|^2 < +\infty \quad \textit{and} \quad \|y_{k+1} - y_k\|^2 = o\left(\frac{1}{k^2}\right) \end{cases} \tag{22}$$

Consequently, $\{y_k\}$ *converges to* $y^\star \in \mathrm{zer}(\mathbf{F})$.

In the next three sections, we present some applications of these results to proximal-point, forward-backward splitting, and three-operator splitting methods. The main idea is to transform (MI) and its special cases to a co-coercive equation of the form (CoCo), and then apply the results to this equation.

3.2 Application to Proximal-Point Method

Here we consider the case of $\mathbf{A} = 0$ and $\mathbf{C} = 0$, where (MI) reduces to finding $y^\star \in \mathbb{R}^p$ such that $0 \in \mathbf{B}y^\star$. We will investigate the convergence of an accelerated proximal-point algorithm and the interplay between the Halpern iteration and Nesterov's accelerated interpretations.

Let $J_{\lambda\mathbf{B}}y := (\mathbf{I} + \lambda\mathbf{B})y$ be the resolvent of $\lambda\mathbf{B}$ for any $\lambda > 0$ and $G_{\lambda\mathbf{B}}y = \frac{1}{\lambda}(\mathbf{I} - J_{\lambda\mathbf{B}})y = \frac{1}{\lambda}(y - J_{\lambda\mathbf{B}}y)$ be the Yosida approximation of $\mathbf{B}$ with index $\lambda > 0$. Then, it is well-known that $G_{\lambda\mathbf{B}}$ is λ-co-coercive [5, Corollary 23.11]. Moreover, $y^\star$ solves $0 \in \mathbf{B}y^\star$ if and only if $G_{\lambda\mathbf{B}}y^\star = 0$. Hence, solving $0 \in \mathbf{B}y^\star$ is equivalent to solving the co-coercive equation $G_{\lambda\mathbf{B}}y^\star = 0$ with λ-co-coercive $G_{\lambda\mathbf{B}}$.

In this case, the Halpern-type fixed-point scheme (5) applying to $G_{\lambda\mathbf{B}}y^\star = 0$, or equivalently, to solving $0 \in \mathbf{B}y^\star$, can be written as

$$y_{k+1} := \beta_k y_0 + (1-\beta_k)y_k - \eta_k G_{\lambda\mathbf{B}}y_k = \beta_k y_0 + \left(1 - \beta_k - \frac{\eta_k}{\lambda}\right) y_k + \frac{\eta_k}{\lambda} J_{\lambda\mathbf{B}}y_k \tag{9}$$

where β_k and η_k can be chosen either in Theorem 1 or Theorem 4 to guarantee convergence of (9). If $\beta_k := \frac{1}{k+2}$ and $\eta_k := 2\lambda(1-\beta_k)$ as in Theorem 1, then we have

$$\begin{array}{lcl} y_{k+1} & := & \beta_k y_0 + (1-\beta_k)y_k - 2(1-\beta_k)(y_k - J_{\lambda\mathbf{B}}y_k) \\ & = & \beta_k y_0 + (1-\beta_k)(2J_{\lambda\mathbf{B}}y_k - y_k) \\ & = & \beta_k y_0 + (1-\beta_k)R_{\lambda\mathbf{B}}y_k \end{array}$$

where $R_{\lambda\mathbf{B}} := 2J_{\lambda\mathbf{B}} - \mathbf{I}$ is the reflected resolvent of $\lambda\mathbf{B}$. Moreover, under this choice of parameters, we have the following result from Theorem 1:

$$\|G_{\lambda\mathbf{B}}y_k\| \le \frac{\|y_0 - y^\star\|}{\lambda(k+1)}$$

If we choose $\beta_k := \frac{\omega+1}{k+2\omega+2}$ and $\eta_k := \gamma(1-\beta_k)$ as in Theorem 4, then (9) becomes

$$y_{k+1} := \frac{\omega+1}{k+2\omega+2}\cdot y_0 + \frac{k+\omega+1}{k+2\omega+2}\cdot\left[\left(1-\frac{\gamma}{\lambda}\right)y_k + \frac{\gamma}{\lambda}J_{\lambda\mathbf{B}}y_k\right]$$

This expression can be viewed as a new variant of the Halpern fixed-point iteration applied to the averaged mapping $P_{\rho A}y = (1-\rho)y + \rho J_{\lambda\mathbf{B}}y$ with $\rho := \frac{\gamma}{\lambda}$ provided that $\gamma \in (0, \lambda]$. In this case, we obtain a convergence result as in (22).

Alternatively, if we apply (Nes) to solve $G_{\lambda\mathbf{B}}y^\star = 0$, then we obtain a Nesterov's accelerated interpretation of (9) as

$$\left\{\begin{array}{lcl} x_{k+1} & := & y_k - \gamma_k G_{\lambda\mathbf{B}}y_k = (1-\rho_k)\,y_k + \rho_k J_{\lambda\mathbf{B}}y_k \quad \text{with} \quad \rho_k := \frac{\gamma_k}{\lambda} \\ y_{k+1} & := & x_{k+1} + \theta_k(x_{k+1} - x_k) + \nu_k(y_k - x_{k+1}) \end{array}\right. \tag{10}$$

This method was studied in [28]. Nevertheless, our analysis in Theorem 3 is simpler than that of [28] when it applies to (10). In particular, if we choose $\gamma_k := \lambda$, then the first line of (10) reduces to $x_{k+1} = J_{\lambda\mathbf{B}}y_k$. The convergence rate guarantees of (10) can be obtained as results of Corollary 2 and Theorem 3, respectively.

Finally, if we apply (15) to solve $G_{\lambda\mathbf{B}}y^\star = 0$ and choose $\gamma_k := \lambda$, $\eta_k := \lambda\left(\frac{\beta_k}{\beta_{k-1}} + 1 - \beta_k\right)$ such that $\nu_k = 0$ and $\kappa_k = \frac{\beta_k}{\beta_{k-2}}$, then (15) reduces to

$$\left\{\begin{array}{lcl} x_{k+1} & := & J_{\lambda\mathbf{B}}y_k \\ y_{k+1} & := & x_{k+1} + \theta_k(x_{k+1} - x_k) + \kappa_k(y_{k-1} - x_k) \end{array}\right.$$

Clearly, if we choose $\beta_k := \frac{1}{k+2}$, then $\theta_k = \frac{k}{k+2}$ and $\kappa_k = \frac{k}{k+2}$. In this case, the last scheme reduces to the accelerated proximal-point algorithm studied in [22]. In addition, we have $\eta_k = \frac{2\lambda(k+1)}{k+2}$ as in Theorem 1. Hence, the result of Corollary 2 is still applicable to this scheme to obtain a convergence rate guarantee $\|G_{\lambda\mathbf{M}}y_k\| \le \frac{\|y_0 - y^\star\|}{\lambda(k+1)}$ as in [22, Theorem 4.1].

3.3 Application to Forward-Backward Splitting Method

We consider the case $\mathbf{C} = 0$ where (MI) reduces to finding $y^\star \in \mathbb{R}^p$ such that $0 \in \mathbf{A}y^\star + \mathbf{B}y^\star$. In this case, we will investigate the convergence of an accelerated forward-backward splitting scheme in §3.3. Clearly, this case covers variational inequality problems (VIP) as special cases when $\mathbf{A} = \mathcal{N}_{\mathcal{X}}$, the normal cone of a closed and convex set $\mathcal{X}$.

By Proposition 1, $y^\star \in \mathrm{zer}(\mathbf{A} + \mathbf{B})$ if and only if $G_{\lambda\mathbf{M}}y^\star = 0$, where $\mathbf{M} := \mathbf{A} + \mathbf{B}$ and $G_{\lambda\mathbf{M}}$ is defined by (1). Moreover, $G_{\lambda\mathbf{M}}$ is $\frac{\lambda(4-\lambda L)}{4}$-co-coercive, provided that $0 < \lambda < \frac{4}{L}$.

If we apply (5) to solve $G_{\lambda\mathbf{M}}y^\star = 0$, then its iterate can be written explicitly as

$$y_{k+1} := \beta_k y_0 + (1-\beta_k)y_k - \eta_k G_{\lambda\mathbf{M}}y_k = \beta_k y_0 + (1-\beta_k)\big[(1-\rho)y_k + \rho J_{\lambda\mathbf{B}}(y_k - \lambda\mathbf{A}y_k)\big] \qquad (11)$$

where we have set $\rho := \frac{4-\lambda L}{2}$. In particular, if we choose $\lambda := \frac{2}{L}$, then $\rho = 1$ and (11) reduces to $y_{k+1} := \beta_k y_0 + (1-\beta_k)J_{\lambda\mathbf{B}}(y_k - \lambda\mathbf{A}y_k)$, which can be viewed as the Halpern fixed-point iteration applied to approximate a fixed-point of $J_{\lambda\mathbf{B}}(y_k - \lambda\mathbf{A}y_k)$.

Depending on the choice of β_k and ρ as in Theorem 1 or Theorem 4, we obtain

$$\|G_{\lambda\mathbf{M}}y_k\| \le \frac{4\|y_0 - y^\star\|}{\lambda(4-\lambda L)(k+1)}, \quad \text{or} \quad \|G_{\lambda\mathbf{M}}y_k\| = o\left(\frac{1}{k}\right)$$

respectively, provided that $0 < \lambda < \frac{4}{L}$.

Now, we consider Nesterov's accelerated variant of (11) by applying (Nes) to $G_{\lambda\mathbf{M}}y^\star = 0$:

$$\begin{cases} x_{k+1} & := & (1-\rho_k)y_k + \rho_k J_{\lambda\mathbf{B}}(y_k - \lambda\mathbf{A}y_k) \\ y_{k+1} & := & x_{k+1} + \theta_k(x_{k+1} - x_k) + \nu_k(y_k - x_{k+1}) \end{cases} \qquad (12)$$

where $\rho_k := \frac{\gamma_k}{\lambda}$, $\theta_k := \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}$, and $\nu_k := \frac{\beta_k}{\beta_{k-1}}$. This scheme is similar to the one studied in [29]. Again, the convergence of (12) can be guaranteed by either Corollary 2 or Theorem 3 depending on the choice of γ_k, θ_k, and ν_k. However, we omit the details here.

3.4 Application to Three-Operator Splitting Method

Finally, we consider the general case where we will investigate the convergence of an accelerated three-operator splitting scheme for solving (MI), and its special case: the accelerated Douglas-Rachford splitting scheme in §3.4.

As stated in Proposition 2, $y^\star \in \mathrm{zer}(\mathbf{A}+\mathbf{B}+\mathbf{C})$ if and only if $E_{\lambda\mathbf{M}}y^\star = 0$, where $\mathbf{M} := \mathbf{A}+\mathbf{B}+\mathbf{C}$ and $E_{\lambda\mathbf{M}}$ is defined by (3). Let us apply (5) to $E_{\lambda\mathbf{M}}y^\star = 0$ and arrive at the following scheme:

$$\begin{array}{lcl} y_{k+1} & := & \beta_k y_0 + (1-\beta_k)y_k - \eta_k E_{\lambda\mathbf{M}}y_k \\ & = & \beta_k y_0 + (1-\beta_k)y_k - \frac{\eta_k}{\lambda}\left(J_{\lambda\mathbf{A}}y_k - J_{\lambda\mathbf{B}}(2J_{\lambda\mathbf{A}}y_k - y_k - \lambda\mathbf{C}\circ J_{\lambda\mathbf{A}}y_k)\right) \end{array}$$

Unfolding this scheme by using intermediate variables z_k and w_k, we obtain

$$\begin{cases} z_k & := & J_{\lambda\mathbf{A}}y_k \\ w_k & := & J_{\lambda\mathbf{B}}(2z_k - y_k - \lambda\mathbf{C}z_k) \\ y_{k+1} & := & \beta_k y_0 + (1-\beta_k)y_k - \frac{\eta_k}{\lambda}(z_k - w_k) \end{cases} \qquad (13)$$

This is called a Halpern-type three-operator splitting scheme for solving (MI). If $\mathbf{C} = 0$, then it reduces to Halpern-type Douglas-Rachford splitting scheme where (MI) reduces to finding $y^\star \in \mathbb{R}^p$ such that $0 \in \mathbf{A}y^\star + \mathbf{B}y^\star$. The latter case was proposed in [45] with a direct convergence proof for both dynamic and constant stepsizes, but the convergence is given on $G_{\lambda\mathbf{M}}$ instead of $E_{\lambda\mathbf{M}}$. Note that the convergence results of Theorem 1 and Theorem 4 can be applied to (13) to obtain convergence rates on $\|E_{\lambda\mathbf{M}}y_k\|$. Such rates can be transformed to the ones on $\|G_{\lambda\mathbf{M}}z_k\|$ when $\mathbf{C} = 0$ and $\mathbf{A}$ is single valued.

Next, we can also derive Nesterov's accelerated variant of (13) by applying (Nes) to solve $E_{\lambda\mathbf{M}}y^\star = 0$. In this case, (Nes) becomes

$$\begin{cases} z_k & := & J_{\lambda\mathbf{A}}y_k \\ w_k & := & J_{\lambda\mathbf{B}}(2z_k - y_k - \lambda\mathbf{C}z_k) \\ x_{k+1} & := & y_k + \frac{1}{\lambda}(w_k - z_k) \\ y_{k+1} & := & x_{k+1} + \theta_k(x_{k+1} - x_k) + \nu_k(y_k - x_{k+1}) \end{cases} \tag{14}$$

Here, the parameters θ_k and ν_k can be chosen as in either Corollary 2 or Theorem 3. This scheme essentially has the same per-iteration complexity as the standard three-operator splitting scheme in the literature [16]. However, its convergence rate is much faster than the standard one by applying either Corollary 2 or Theorem 3. If $\mathbf{C} = 0$, then (14) reduces to accelerated Douglas-Rachford splitting scheme, where its fast convergence rate guarantee can be obtained as a special case of either Corollary 2 or Theorem 3.

4 Equivalence Between Halpern and Nesterov Acceleration

Accelerated methods have become indispensable in both optimization and fixed-point theory due to their significant improvements in convergence rates over traditional iterative techniques. Two prominent examples of such methods are Nesterov's accelerated gradient method, well-known in the optimization community, and Halpern's fixed-point iteration, which stems from fixed-point theory. While these two methods were developed in different mathematical frameworks, both share the overarching goal of accelerating convergence for various classes of problems. Nesterov's method, originally designed for convex optimization, has optimal convergence rates for first-order optimization methods. In contrast, Halpern's iteration, traditionally used for non-expansive operators, has only recently been shown to achieve comparable acceleration. This raises the natural question of whether a deeper connection exists between these two seemingly distinct methods.

Both Nesterov's accelerated and Halpern iteration schemes show significant improvement on convergence rates over classical methods for solving (MI). However, they are derived from different perspectives, and it is unclear if they are closely related to each other. In this section, we show that these schemes are actually equivalent, though they may use different sets of parameters.

4.1 The Relation Between Halpern's and Nesterov's Accelerations

Our next step is to show that the Halpern fixed-point iteration (5) can be transformed into a Nesterov's accelerated interpretation and vice versa.

Theorem 2. *Let $\{x_k\}$ and $\{y_k\}$ be generated by the following scheme starting from $y_0 \in \mathbb{R}^p$ and $x_0 = x_{-1} = y_{-1} := y_0$ and $\beta_{-1} = \eta_{-1} = 0$:*

$$\begin{cases} x_{k+1} & := & y_k - \gamma_k \mathbf{F} y_k \\ y_{k+1} & := & x_{k+1} + \theta_k(x_{k+1} - x_k) + \nu_k(y_k - x_{k+1}) + \kappa_k(y_{k-1} - x_k) \end{cases} \tag{15}$$

where $\theta_k := \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}$, $\nu_k := \frac{\beta_k}{\beta_{k-1}} + 1 - \beta_k - \frac{\eta_k}{\gamma_k}$, and $\kappa_k := \frac{\beta_k}{\beta_{k-1}}\left(\frac{\eta_{k-1}}{\gamma_{k-1}} - 1 + \beta_{k-1}\right)$. Then, the sequence $\{y_k\}$ is identical to the one generated by (5) *for solving* (CoCo).

In particular, we have

Corollary 1. *If we choose $\gamma_k := \frac{\eta_k}{1-\beta_k}$, then $\nu_k = \frac{\beta_k}{\beta_{k-1}}$, $\kappa_k = 0$, and* (15) *reduces to*

$$\begin{cases} x_{k+1} & := & y_k - \gamma_k \mathbf{F} y_k \\ y_{k+1} & := & x_{k+1} + \theta_k(x_{k+1} - x_k) + \nu_k(y_k - x_{k+1}) \end{cases} \tag{Nes}$$

Remark. Both (15) and (Nes) can be viewed as Nesterov's accelerated variants with correction terms. While (15) has two correction terms $\nu_k(y_k - x_{k+1})$ and $\kappa_k(y_{k-1} - x_k)$, (Nes) has only one term $\nu_k(y_k - x_{k+1})$. In fact, (Nes) covers the proximal-point scheme in [28] as a special case. As discussed in [2], (Nes) can be viewed as Ravine's method if convergence is given in y_k instead of x_k.

In particular, if we choose $\gamma_k := \frac{1}{L}$, $\beta_k := \frac{1}{k+2}$, and $\eta_k := \frac{(1-\beta_k)}{L} = \frac{(k+1)}{L(k+2)}$, then (15) reduces to the following one:

$$\begin{cases} x_{k+1} & := & y_k - \frac{1}{L}\mathbf{F} y_k \\ y_{k+1} & := & x_{k+1} + \frac{k}{k+2}(x_{k+1} - x_k) + \frac{k+1}{k+2}(y_k - x_{k+1}) \end{cases} \tag{16}$$

Alternatively, if we choose $\gamma_k := \frac{1}{L}$, $\beta_k := \frac{1}{k+2}$, and $\eta_k := \frac{2(1-\beta_k)}{L}$, then (15) reduces to

$$\begin{cases} x_{k+1} & := & y_k - \frac{1}{L}\mathbf{F} y_k \\ y_{k+1} & := & x_{k+1} + \frac{k}{k+2}(x_{k+1} - x_k) + \frac{k}{k+2}(y_{k-1} - x_k) \end{cases} \tag{17}$$

The convergence on $\|\mathbf{F}y_k\|$ of both (16) and (17) is guaranteed by Theorem 1. The scheme (17) covers [22] as a special case when $\mathbf{F}y = J_{\lambda\mathbf{B}}y$, the resolvent of a maximally monotone operator $\lambda\mathbf{B}$. We state this result in the following corollary. However, it is not clear how to derive convergence of $\|\mathbf{F}x_k\|$ as well as $o\left(1/k^2\right)$-rates as in Theorem 1. We summarize this result in the following lemma.

Corollary 2. *Assume that $\mathbf{F}$ in* (CoCo) *is $\frac{1}{L}$-co-coercive and $\mathrm{zer}(\mathbf{F}) \neq \emptyset$. Let $\{(x_k, y_k)\}$ be generated by* (15) *using $\gamma_k := \frac{1}{L}$, $\beta_k := \frac{1}{k+2}$, and $\eta_k := \frac{(1-\beta_k)}{L}$. Then, we obtain $\theta_k := \frac{k}{k+2}$, $\nu_k := \frac{k+1}{k+2}$, and $\kappa_k := 0$. Moreover,* (15) *reduces to* (Nes) (*or equivalently* (16)), *and the following guarantee holds:*

$$\|\mathbf{F}y_k\|^2 \le \frac{4L^2\|y_0 - y^\star\|^2}{(k+1)(k+3)}, \quad \textit{and} \quad \sum_{k=0}^{\infty}(k+1)(k+2)\|\mathbf{F}y_{k+1} - \mathbf{F}y_k\|^2 \le 2L^2\|y_0 - y^\star\|^2 \tag{18}$$

If we use $\gamma_k := \frac{1}{L}$, $\beta_k := \frac{1}{k+2}$, and $\eta_k := \frac{2(1-\beta_k)}{L}$, then we obtain $\theta_k := \frac{k}{k+2}$, $\nu_k := 0$, and $\kappa_k := \frac{k}{k+2}$, and (15) *reduces to* (17). *Moreover, the following guarantee holds: $\|\mathbf{F}y_k\| \le \frac{L\|y_0 - y^\star\|}{(k+1)}$.*

The constant factor in the bound (18) is slightly worse than the one in $\|\mathbf{F}y_k\| \le \frac{L\|y_0 - y^\star\|}{(k+1)}$. In fact, the latter one is exactly optimal since there exists an instance showing it matches the lower bound complexity, see, e.g., [17, 26].

4.2 Convergence Analysis of Nesterov's Accelerated Scheme (Nes)

In this subsection, we provide a direct convergence analysis of (Nes) without using Theorem 1. For simplicity, we will analyze the convergence of (Nes) with only one correction term. However, our analysis can be easily extended to (15) when $\kappa_k \neq 0$ with some simple modifications.

Our analysis relies on the following Lyapunov function:

$$\mathcal{V}_k := a_k\|\mathbf{F}y_{k-1}\|^2 + b_k\langle \mathbf{F}y_{k-1}, x_k - y_k\rangle + \|x_k + t_k(y_k - x_k) - y^\star\|^2 + \mu\|x_k - y^\star\|^2 \tag{19}$$

where a_k, b_k, and $t_k > 0$ are given parameters, which will be determined later, and $\mu \geq 0$ is an optimal parameter. This Lyapunov is slightly different from $\mathcal{L}_k$ defined by (6), but it is closely related to standard Nesterov's potential function (see, e.g., [3]). To see the connection between $\mathcal{V}_k$ and $\mathcal{L}_k$, we prove the following lemma.

Proposition 3. *Let $\mathcal{L}_k$ be defined by* (6) *and $\mathcal{V}_k$ be defined by* (19). *Assume that $a_{k+1} := \frac{4p_k^2}{Lq_k^2} + \frac{4p_k\eta_k}{Lq_k(1-\beta_k)}$, $b_{k+1} := \frac{4p_k}{Lq_k\beta_k}$, $t_{k+1} := \frac{1}{\beta_k}$, and $\gamma_k := \frac{\eta_k}{1-\beta_k}$. Then, we have*

$$\mathcal{V}_{k+1} = \frac{4p_k}{Lq_k^2}\mathcal{L}_k + \|y_0 - y^\star\|^2 + \mu\|x_{k+1} - y^\star\|^2 \tag{20}$$

Remark. (i) For $p_k = q_0k(k+1)$ and $q_k = q_0(k+1)$, we have $\frac{4p_k}{Lq_k^2} = \frac{4k}{Lq_0(k+1)} \approx \frac{4}{Lq_0}$. Hence, we have $\mathcal{V}_{k+1} = \frac{4k}{Lq_0(k+1)}\mathcal{L}_k + \|y_0 - y^\star\|^2 + \mu\|x_{k+1} - y^\star\|^2$.

(ii) If we choose $p_k = cq_k^2$ for some $c > 0$, then $\mathcal{V}_{k+1} = \frac{4c}{L}\mathcal{L}_k + \|y_0 - y^\star\|^2 + \mu\|x_{k+1} - y^\star\|^2$. Clearly, if $\mu = 0$, then $\mathcal{V}_{k+1} = \frac{4c}{L}\mathcal{L}_k + \|y_0 - y^\star\|^2$.[1]

The following theorem proves convergence of Nesterov's accelerated scheme (Nes), but using a different set of parameters compared to Theorem 1.

Theorem 3. *Assume that $\mathbf{F}$ in* (CoCo) *is $\frac{1}{L}$-co-coercive and* $\mathrm{zer}(\mathbf{F}) \neq \emptyset$. *Let $\{(x_k, y_k)\}$ be generated by* (Nes) *to solve* (CoCo) *using $\gamma_k := \gamma \in \left(0, \frac{1}{L}\right)$, $\theta_k := \frac{k+1}{k+2\omega+2}$, and $\nu_k := \frac{k+\omega+2}{k+2\omega+2} \in (0,1)$ for a given constant $\omega > 2$. Then, the following estimates hold:*

$$\begin{cases} \sum_{k=0}^{\infty}(k+\omega+1)\|x_{k+1} - x_k\|^2 < +\infty \quad \textit{and} \quad \|x_{k+1} - x_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+2\omega+1)\|y_k - x_k\|^2 < +\infty \quad \textit{and} \quad \|y_k - x_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega+1)\|\mathbf{F}y_{k-1}\|^2 < +\infty \quad \textit{and} \quad \|\mathbf{F}y_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega+1)\|\mathbf{F}x_k\|^2 < +\infty \quad \textit{and} \quad \|\mathbf{F}x_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega)\|y_{k+1} - y_k\|^2 < +\infty \quad \textit{and} \quad \|y_{k+1} - y_k\|^2 = o\left(\frac{1}{k^2}\right) \end{cases} \tag{21}$$

Consequently, both $\{x_k\}$ and $\{y_k\}$ converge to $y^\star \in \mathrm{zer}(\mathbf{F})$.

Remark. (i) If we choose $\gamma = \frac{1}{L}$, then we only obtain the first result of (21) and $\|x_{k+1} - x_k\|^2 = o\left(\frac{1}{k^2}\right)$. This rate is slightly better than the $\mathcal{O}\left(1/k^2\right)$ rate in [22] when k is sufficiently large.

[1] The term $\mu\|x_k - y^\star\|^2$ allows us to get the tail $\|x_{k+1} - x_k\|^2$ in (25b) of Lemma 2 later in §4.4, which is a key to prove convergence in Theorem 3, especially $o\left(1/k^2\right)$-convergence rates. It remains unclear to us how to prove such a convergence rate without the term $\mu\|x_k - y^\star\|^2$.

(ii) If $\gamma \in \left(0, \frac{1}{L}\right)$, then we can prove $o\left(\frac{1}{k^2}\right)$ convergence rates of $\|\mathbf{F}y_k\|^2$, $\|\mathbf{F}x_k\|^2$, $\|y_k - x_k\|^2$, and $\|y_{k+1} - y_k\|^2$. Note that we can simply choose $\omega = 3$ to further simplify the results. In this case, we obtain $\theta_k = \frac{k+1}{k+8}$, which is different from $\theta_k = \frac{k}{k+2}$ in (16) obtained by Theorem 1. We emphasize that $o(\cdot)$ convergence rates have recently studied in a number of works such as [1, 3, 28, 29].

From the result of Theorem 3, we can derive the convergence of the Halpern fixed-point iteration (5), but under different choice of parameters.

Theorem 4. *Let $\{y_k\}$ be generated by the Halpern fixed-point iteration* (5) *using $\beta_k := \frac{\omega+1}{k+2\omega+2}$ and $\eta_k := \gamma(1-\beta_k)$ for a fixed $\gamma \in \left(0, \frac{1}{L}\right)$ and $\omega > 2$. Then, the following statements hold:*

$$\begin{cases} \sum_{k=0}^{\infty}(k+\omega+1)\|\mathbf{F}y_{k-1}\|^2 < +\infty \quad \textit{and} \quad \|\mathbf{F}y_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega)\|y_{k+1}-y_k\|^2 < +\infty \quad \textit{and} \quad \|y_{k+1}-y_k\|^2 = o\left(\frac{1}{k^2}\right) \end{cases} \tag{22}$$

Consequently, $\{y_k\}$ converges to $y^\star \in \mathrm{zer}(\mathbf{F})$.

If we set $\omega = 0$, then we obtain $\beta_k = \frac{1}{k+2}$ as in Theorem 1. In this case, we have to set $\mu = 0$ in $\mathcal{V}_k$ from (19), and hence only obtain $\|\mathbf{F}y_k\|^2 = \mathcal{O}\left(1/k^2\right)$ convergence rate. Note that other choices of parameters in Theorem 3 are possible, e.g., by changing μ and ω. Here, we have not tried to optimize the choice of these parameters. As shown in [5, Proposition 4.11] that $\mathbf{T}$ is a non-expansive mapping if and only if $\mathbf{F} := \mathbf{I} - \mathbf{T}$ is $\frac{1}{2}$-co-coercive. Therefore, we can obtain new convergence results on the residual norm $\|y_k - \mathbf{T}y_k\|$ from Theorem 4 for a Halpern fixed-point iteration scheme to approximate a fixed-point $y^\star$ of $\mathbf{T}$.

4.3 Proof of Theorem 2

Proof of Theorem 2.

(15) $\Rightarrow$ (5) Substituting θ_k, ν_k, and κ_k into (15), and simplifying the result, we get

$$\begin{aligned} y_{k+1} &= \left(\tfrac{\beta_k}{\beta_{k-1}} - \beta_k + 1\right)x_{k+1} - \tfrac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}x_k + \left(\tfrac{\beta_k}{\beta_{k-1}} + 1 - \beta_k - \tfrac{\eta_k}{\gamma_k}\right)(y_k - x_{k+1}) \\ &\quad + \tfrac{\beta_k}{\beta_{k-1}}\left(\tfrac{\eta_{k-1}}{\gamma_{k-1}} - 1 + \beta_{k-1}\right)(y_{k-1} - x_k) \end{aligned}$$

Now, using the first line of (15) into this expression, we get

$$\begin{aligned} y_{k+1} &= \left(\tfrac{\beta_k}{\beta_{k-1}} - \beta_k + 1\right)(y_k - \gamma_k\mathbf{F}y_k) - \tfrac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}(y_{k-1} - \gamma_{k-1}\mathbf{F}y_{k-1}) \\ &\quad + \left(\tfrac{\beta_k}{\beta_{k-1}} + 1 - \beta_k - \tfrac{\eta_k}{\gamma_k}\right)\gamma_k\mathbf{F}y_k + \tfrac{\beta_k}{\beta_{k-1}}\left(\tfrac{\eta_{k-1}}{\gamma_{k-1}} - 1 + \beta_{k-1}\right)\gamma_{k-1}\mathbf{F}y_{k-1} \end{aligned}$$

Rearranging this expression, we arrive at

$$\frac{1}{\beta_k}y_{k+1} - \left(\frac{1}{\beta_k} - 1\right)y_k + \frac{\eta_k}{\beta_k}\mathbf{F}y_k = \frac{1}{\beta_{k-1}}y_k - \left(\frac{1}{\beta_{k-1}} - 1\right)y_{k-1} - \frac{\eta_{k-1}}{\beta_{k-1}}\mathbf{F}y_{k-1}$$

By induction, and noticing that $y_{-1} = y_0$, and $\eta_{-1} = 0$, this expression leads to

$$\frac{1}{\beta_k}y_{k+1} - \left(\frac{1}{\beta_k} - 1\right)y_k + \frac{\eta_k}{\beta_k}\mathbf{F}y_k = y_0$$

This is indeed equivalent to (5).

(5)⇒(15) Firstly, shifting the index from k to $k-1$ in (5), we have $y_k = \beta_{k-1}y_0 + (1-\beta_{k-1})y_{k-1} - \eta_{k-1}\mathbf{F}y_{k-1}$. Here, we assume that $y_{-1} = y_0$. Multiplying this expression by β_k and (5) by $-\beta_{k-1}$ and adding the results, we obtain

$$\beta_{k-1}y_{k+1} - \beta_k y_k = \beta_{k-1}(1-\beta_k)y_k - \beta_k(1-\beta_{k-1})y_{k-1} - \beta_{k-1}\eta_k\mathbf{F}y_k + \beta_k\eta_{k-1}\mathbf{F}y_{k-1}$$

This expression leads to

$$y_{k+1} = \left(\frac{\beta_k}{\beta_{k-1}} + 1 - \beta_k\right)y_k - \eta_k\mathbf{F}y_k - \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}y_{k-1} + \frac{\beta_k\eta_{k-1}}{\beta_{k-1}}\mathbf{F}y_{k-1} \tag{23}$$

Next, let us introduce $x_{k+1} := y_k - \gamma_k\mathbf{F}y_k$ for some $\gamma_k > 0$. Then, we have $\mathbf{F}y_k = \frac{1}{\gamma_k}(y_k - x_{k+1})$. Substituting this relation into (23), we obtain

$$\begin{aligned} y_{k+1} &= \left(\tfrac{\beta_k}{\beta_{k-1}} + 1 - \beta_k\right)y_k - \tfrac{\eta_k}{\gamma_k}(y_k - x_{k+1}) - \tfrac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}y_{k-1} + \tfrac{\beta_k\eta_{k-1}}{\beta_{k-1}\gamma_{k-1}}(y_{k-1} - x_k) \\ &= x_{k+1} + \tfrac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}(x_{k+1} - x_k) + \left(\tfrac{\beta_k}{\beta_{k-1}} + 1 - \beta_k - \tfrac{\eta_k}{\gamma_k}\right)(y_k - x_{k+1}) \\ &\quad + \tfrac{\beta_k}{\beta_{k-1}}\left(\tfrac{\eta_{k-1}}{\gamma_{k-1}} - 1 + \beta_{k-1}\right)(y_{k-1} - x_k) \end{aligned}$$

If we define $\theta_k := \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}$, $\nu_k := \frac{\beta_k}{\beta_{k-1}} + 1 - \beta_k - \frac{\eta_k}{\gamma_k}$ and $\kappa_k := \frac{\beta_k}{\beta_{k-1}}\left(\frac{\eta_{k-1}}{\gamma_{k-1}} - 1 + \beta_{k-1}\right)$, then this expression can be rewritten as

$$y_{k+1} = x_{k+1} + \theta_k(x_{k+1} - x_k) + \nu_k(y_k - x_{k+1}) + \kappa_k(y_{k-1} - x_k)$$

Combining this line and $x_{k+1} = y_k - \gamma_k\mathbf{F}y_k$, we get (15).

Finally, if we choose $\gamma_k := \frac{\eta_k}{1-\beta_k}$, then it is obvious that $\kappa_k = 0$, and $\nu_k = \frac{\beta_k}{\beta_{k-1}}$. Hence, (15) reduces to (Nes).

□

4.4 Proof of Theorem 3

Firts, we prove the following key lemma for our convergence analysis.

Lemma 1. *Let $\{(x_k, y_k)\}$ be generated by* (Nes) *using $\gamma_k := \gamma > 0$, and $\mathcal{V}_k$ be defined by* (19)*. Then, if $b_{k+1}\theta_k + 2\gamma t_k(t_k-1) - 2\gamma\nu_k\theta_k t_{k+1}^2 \geq 0$, then we have*

$$\begin{aligned} \mathcal{V}_k - \mathcal{V}_{k+1} \geq\ & \left(\gamma b_{k+1}\nu_k + \gamma^2 t_k^2 - \gamma^2 t_{k+1}^2\nu_k^2 - a_{k+1} - \tfrac{\gamma^2 b_k^2}{4a_k}\right)\|\mathbf{F}y_k\|^2 \\ &+ \left[b_{k+1}\theta_k + 2\gamma t_k(t_k-1) - 2\gamma\nu_k\theta_k t_{k+1}^2 - b_k\right]\langle\mathbf{F}y_{k-1}, x_{k+1} - x_k\rangle \\ &+ \left(t_k^2 - 2t_k + 1 + \mu - t_{k+1}^2\theta_k^2\right)\|x_{k+1} - x_k\|^2 + a_k\left\|\mathbf{F}y_{k-1} - \tfrac{\gamma b_k}{2a_k}\mathbf{F}y_k\right\|^2 \\ &+ \left(\tfrac{1}{L} - \gamma\right)\left[b_{k+1}\theta_k + 2\gamma t_k(t_k-1) - 2\gamma\nu_k\theta_k t_{k+1}^2\right]\|\mathbf{F}y_k - \mathbf{F}y_{k-1}\|^2 \\ &+ 2(t_k - t_{k+1}\theta_{k+1} - 1 - \mu)\langle x_{k+1} - x_k, x_{k+1} - y^\star\rangle \\ &+ \gamma\left(t_k - t_{k+1}\nu_k\right)\langle\mathbf{F}y_k, x_{k+1} - y^\star\rangle \end{aligned} \tag{24}$$

Our next lemma is to provide a particular choice of parameters such that $\mathcal{V}_k - \mathcal{V}_{k+1} \geq 0$.

Lemma 2. *Let $0 < \gamma \le \frac{1}{L}$, $\mu \ge 0$, and $\omega \ge 1$ be given. Let $\{(x_k, y_k)\}$ be generated by* (Nes) *and $\mathcal{V}_k$ be defined by* (19)*. Assume that t_k, θ_k, ν_k, a_k, and b_k in* (Nes) *and* (19) *are chosen as follows:*

$$\begin{array}{l} t_k := \frac{k+2\omega+1}{\omega}, \quad \theta_k := \frac{t_k - 1 - \mu}{t_{k+1}}, \quad \nu_k := 1 - \frac{1}{t_{k+1}} \\ b_k := 2\gamma t_k(t_k - 1), \quad \textit{and} \quad a_k := \gamma^2 t_k(t_k - 1) \end{array} \tag{25a}$$

Then, it holds that

$$\begin{array}{lcl} \mathcal{V}_k - \mathcal{V}_{k+1} & \ge & \mu(2t_k - 1 - \mu)\|x_{k+1} - x_k\|^2 + \gamma^2 t_k(t_k-1)\|\mathbf{F}y_k - \nu_{k-1}\mathbf{F}y_{k-1}\|^2 \\ & & + 2\gamma\left(\frac{1}{L} - \gamma\right) t_k(t_k-1)\|\mathbf{F}y_k - \mathbf{F}y_{k-1}\|^2 + \frac{\gamma(\omega-1)}{L\omega}\|\mathbf{F}y_k\|^2 \ge 0 \end{array} \tag{25b}$$

Moreover, we have $\mathcal{V}_k \ge \mu\|x_k - y^\star\|^2 + \frac{b_k}{t_k}\left(\frac{1}{L} - \gamma\right)\|\mathbf{F}y_{k-1}\|^2 \ge 0$. Consequently, we obtain

$$\begin{cases} \sum_{k=0}^{\infty} \mu(2t_k - 1 - \mu)\|x_{k+1} - x_k\|^2 & \le \mathcal{V}_0 \\ \frac{\gamma(\omega-1)}{L\omega}\sum_{k=0}^{\infty}\|\mathbf{F}y_k\|^2 & \le \mathcal{V}_0 \\ \frac{2\gamma(1-L\gamma)}{L}\sum_{k=0}^{\infty} t_k(t_k-1)\|\mathbf{F}y_k - \mathbf{F}y_{k-1}\|^2 & \le \mathcal{V}_0 \\ \gamma^2\sum_{k=0}^{\infty} t_k(t_k-1)\|x_{k+1} - x_k - \theta_{k-1}(x_k - x_{k-1})\|^2 & \le \mathcal{V}_0 \end{cases} \tag{25c}$$

We are ready to present

Proof of Theorem 3. The first claim in the first line of (21) is directly from (25c) by noticing that $t_k - 1 = \frac{k+\omega+1}{\omega}$. Now, we prove the second line of (21). We first choose $\mu = 1$ in (19). Then, from (Nes), we have

$$\begin{array}{lcl} y_{k+1} - x_{k+1} & = & \theta_k(x_{k+1} - x_k) + \gamma\nu_k\mathbf{F}y_k = \nu_k(x_{k+1} + \gamma\mathbf{F}y_k - x_k) + (\theta_k - \nu_k)(x_{k+1} - x_k) \\ & = & \nu_k(y_k - x_k) + (\theta_k - \nu_k)(x_{k+1} - x_k) \end{array}$$

Hence, by Young's inequality, $\nu_k \in (0,1)$, and this expression, we can show that

$$\begin{array}{lcl} t_{k+1}^2\|y_{k+1} - x_{k+1}\| & = & t_{k+1}^2\|\nu_k(y_k - x_k) + (\theta_k - \nu_k)(x_{k+1} - x_k)\|^2 \\ & \le & t_{k+1}^2\nu_k\|y_k - x_k\|^2 + \frac{t_{k+1}^2(\theta_k - \nu_k)^2}{1-\nu_k}\|x_{k+1} - x_k\|^2 \end{array}$$

Notice from (25a) that $t_{k+1}^2\nu_k = t_k^2 - \frac{\omega(\omega-2)t_k + \omega - 1}{\omega^2}$ and $\frac{t_{k+1}^2(\theta_k-\nu_k)^2}{1-\nu_k} = \frac{(\omega-1)^2(k+2\omega+2)}{\omega^3}$. Utilizing these expressions into the last inequality, we obtain

$$\frac{\omega(\omega-2)t_k + \omega - 1}{\omega^2}\|y_k - x_k\|^2 \le t_k^2\|y_k - x_k\|^2 - t_{k+1}^2\|y_{k+1} - x_{k+1}\|^2 + \frac{(\omega-1)^2 t_{k+1}}{\omega^2}\|x_{k+1} - x_k\|^2 \tag{26}$$

Summing up this estimate from $k = 0$ to $k = K$, we get

$$\sum_{k=0}^{K}\frac{\omega(\omega-2)t_k + \omega - 1}{\omega^2}\|y_k - x_k\|^2 \le t_0^2\|y_0 - x_0\|^2 + \frac{(\omega-1)^2}{\omega^3}\sum_{k=0}^{K}(k + 2\omega + 2)\|x_{k+1} - x_k\|^2$$

Using the first line of (21) into this inequality and $\omega > 2$, we obtain $\sum_{k=0}^{\infty}\left[(\omega-2)(k+2\omega+1) + \omega - 1\right]\|y_k - x_k\|^2 < +\infty$, which implies the first claim in the second line of (21). Moreover, (26) also shows that $\lim_{k\to\infty} t_k^2\|x_k - y_k\|^2$ exists. Combining this fact and $\sum_{k=0}^{\infty}(k+2\omega+1)\|y_k - x_k\|^2 < +\infty$, we obtain $\lim_{k\to\infty} t_k^2\|x_k - y_k\|^2 = 0$, which shows that $\|x_k - y_k\|^2 = o\left(1/k^2\right)$.

To prove the third line of (21), we note that $\gamma\nu_k\mathbf{F}y_k = (y_{k+1} - x_{k+1}) - \theta_k(x_{k+1} - x_{k+1})$. Hence, $\gamma^2\nu_k^2(t_k-1)\|\mathbf{F}y_k\|^2 \le 2(t_k-1)\|y_{k+1} - x_{k+1}\|^2 + 2\theta_k^2(t_k-1)\|x_{k+1} - x_k\|^2$. Exploiting the finite sum of the last two terms from (21), we obtain $\sum_{k=0}^{\infty}(k+\omega+1)\|\mathbf{F}y_k\|^2 < +\infty$.

To prove the second part in the first line of (21), utilizing (Nes), we can show that

$$
\begin{array}{lcl}
\mathcal{T}_{[1]} & := & \theta_{k-1}^2 t_k^2\|x_k - x_{k-1}\|^2 - \theta_k^2 t_{k+1}^2\|x_{k+1} - x_k\|^2 \\
& = & t_k^2\|y_k - x_k - \gamma\nu_{k-1}\mathbf{F}y_{k-1}\|^2 - t_k^2\|y_k - x_k - \gamma\mathbf{F}y_k\|^2 + (t_k^2 - \theta_k^2 t_{k+1}^2)\|x_{k+1} - x_k\|^2 \\
& = & \gamma^2 t_k^2\|\mathbf{F}y_k - \nu_{k-1}\mathbf{F}y_{k-1}\|^2 + 2\gamma t_k^2\langle \mathbf{F}y_k - \nu_{k-1}\mathbf{F}y_{k-1}, x_{k+1} - x_k\rangle \\
& & + \ (t_k^2 - \theta_k^2 t_{k+1}^2)\|x_{k+1} - x_k\|^2 \\
& = & \gamma^2 t_k^2\|\mathbf{F}y_k - \nu_{k-1}\mathbf{F}y_{k-1}\|^2 + 2\gamma t_k^2\langle \mathbf{F}y_k - \mathbf{F}y_{k-1}, x_{k+1} - x_k\rangle \\
& & + \ 2\gamma t_k^2(1-\nu_k)\langle \mathbf{F}y_{k-1}, x_{k+1} - x_k\rangle + (t_k^2 - \theta_k^2 t_{k+1}^2)\|x_{k+1} - x_k\|^2 \\
& \ge & \gamma^2 t_k^2\|\mathbf{F}y_k - \nu_{k-1}\mathbf{F}y_{k-1}\|^2 + 2\gamma t_k^2\left(\frac{1}{L} - \gamma\right)\|\mathbf{F}y_k - \mathbf{F}y_{k-1}\|^2 \\
& & + \ 2\gamma t_k^2(1-\nu_k)\langle \mathbf{F}y_{k-1}, x_{k+1} - x_k\rangle + (t_k^2 - \theta_k^2 t_{k+1}^2)\|x_{k+1} - x_k\|^2
\end{array}
$$

Employing the update rule (25a) and Young's inequality, this inequality leads to

$$
\begin{array}{lcl}
\theta_{k-1}^2 t_k^2\|x_k - x_{k-1}\|^2 - \theta_k^2 t_{k+1}^2\|x_{k+1} - x_k\|^2 & \ge & 2\gamma t_k^2(1-\nu_k)\langle \mathbf{F}y_{k-1}, x_{k+1} - x_k\rangle \\
& \ge & -\frac{\gamma t_k^2}{t_{k+1}}\left[\|\mathbf{F}y_{k-1}\|^2 + \|x_{k+1} - x_k\|^2\right]
\end{array}
$$

Following the same argument as in the proof of $\|x_k - y_k\|^2$, we can show that $\lim_{k\to\infty} t_k^2\|x_{k+1} - x_k\| = 0$, and hence, $\|x_{k+1} - x_k\|^2 = o\left(1/k^2\right)$, which proves the second part in the first line of (21). Since $\gamma^2\|\mathbf{F}y_k\|^2 = \|x_{k+1} - y_k\|^2 \le 2\|x_{k+1} - x_k\|^2 + 2\|y_k - x_k\|^2$, we also obtain $\|\mathbf{F}y_k\|^2 = o\left(1/k^2\right)$.

Since $\|\mathbf{F}x_k\|^2 \le 2\|\mathbf{F}x_k - \mathbf{F}y_k\| + 2\|\mathbf{F}y_k\|^2 \le 2L^2\|x_k - y_k\|^2 + 2\|\mathbf{F}y_k\|^2$, we obtain the fourth line of (21) from the previous lines.

Now, to prove the last line of (21). Since $y_{k+1} - y_k = x_{k+1} - x_k + \theta_k(x_{k+1} - x_k) - \theta_{k-1}(x_k - x_{k-1}) - \gamma\left(\nu_k\mathbf{F}y_k - \nu_{k-1}\mathbf{F}y_{k-1}\right]$, we can bound

$$
\begin{array}{lcl}
\|y_{k+1} - y_k\|^2 & \le & 4(1+\theta_k)^2\|x_{k+1} - x_k\|^2 + 4\theta_{k-1}^2\|x_k - x_{k-1}\|^2 \\
& & + \ 4\gamma^2\nu_k^2\|\mathbf{F}y_k - \mathbf{F}y_{k-1}\|^2 + 4\gamma(\nu_k - \nu_{k-1})^2\|\mathbf{F}y_{k-1}\|^2 \\
& \le & 16\|x_{k+1} - x_k\|^2 + 4\|x_k - x_{k-1}\|^2 + 4\gamma^2\|\mathbf{F}y_k - \mathbf{F}y_{k-1}\|^2 + 4\gamma^2\|\mathbf{F}y_{k-1}\|^2
\end{array}
$$

Here, we have used the facts that $\theta_k, \theta_{k-1}, \nu_k, \nu_{k-1} \in (0,1)$ and $(\nu_k - \nu_{k-1})^2 < 1$. Combining this estimate and the first and second lines of (21), we obtain the third line of (21).

Finally, to prove the convergence of $\{x_k\}$ and $\{y_k\}$, we note that $\|x_k - y^\star\|^2 \le \mathcal{V}_k \le \mathcal{V}_0$. Hence $\{x_k\}$ is bounded, which has a limit point. Moreover, since $\lim_{k\to\infty}\|x_k - x_{k-1}\| = 0$, $\{x_k\}$ is convergent to $y^\star$. We have $\|\mathbf{F}x_k\| \le \|\mathbf{F}y_k\| + \|\mathbf{F}y_k - \mathbf{F}x_k\| \le \|\mathbf{F}y_k\| + L\|x_k - y_k\| \to 0$ as $k \to \infty$. Since $\mathbf{F}$ is $\frac{1}{L}$-co-cocercive, it is L-Lipschitz continuous. Passing the limit through $\mathbf{F}$ and using the continuity of $\mathbf{F}$, we obtain $\mathbf{F}y^\star = 0$. Since $\|x_k - y_k\| \to \infty$, we also have $\lim_{k\to\infty} y^k = y^\star$. $\square$

4.5 Proof of Theorem 4

Proof of Theorem 4. As proved in Theorem 2, (5) is equivalent to (Nes) provided that $\theta_k = \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}$, $\nu_k = \frac{\beta_k}{\beta_{k-1}}$, and $\gamma_k := \frac{\eta_k}{1-\beta_k}$. Using the choice of β_k, ν_k, and γ_k in Theorem 3, we can show that

$\theta_k = \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}} = \frac{k+1}{k+2\omega+2}$ and $\nu_k = \frac{\beta_k}{\beta_{k-1}} = \frac{k+\omega+2}{k+2\omega+2}$. These relations lead to $\beta_k = \frac{\omega+1}{k+2\omega+2}$. Moreover, since $\gamma_k = \frac{\eta_k}{1-\beta_k} = \gamma \in \left(0, \frac{1}{L}\right)$, we have $\eta_k = \gamma(1-\beta_k)$. Consequently, (22) follows from (21). $\square$

5 Conclusion

In this paper, we developed a unified framework for solving monotone inclusions by bridging two prominent acceleration techniques: Halpern's fixed-point iteration and Nesterov's accelerated gradient method. Despite their distinct origins in fixed-point theory and optimization, we demonstrated that these methods share deep mathematical foundations and are equivalent under specific parameter choices. This unification not only enhances our understanding of their mechanics but also extends their applicability to a broad range of optimization problems, including the proximal-point method, forward-backward splitting, and Douglas-Rachford splitting.

Our results show significant improvements in convergence rates, particularly in solving monotone inclusions and variational inequalities, where accelerated methods offer a substantial practical advantage. By unifying these techniques, we have laid the foundation for more efficient and scalable solutions, especially in large-scale and structured problems.

Future research can focus on extending these acceleration techniques to more complex settings, such as non-smooth or stochastic optimization problems. Further exploration into continuous-time interpretations and optimized parameter choices could provide additional theoretical insights and lead to even greater advances in solving monotone inclusions. Ultimately, this work sets the stage for enhanced methods capable of addressing a wide spectrum of computational challenges in optimization and beyond.

References

[1] H. Attouch and A. Cabot. Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. *Math. Program.*, 184(1):243–287, 2020.

[2] H. Attouch and J. Fadili. From the Ravine method to the Nesterov method and vice versa: a dynamical system perspective. *arXiv preprint arXiv:2201.11643*, 2022.

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

[4] H. Attouch and J. Peypouquet. Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. *Math. Program.*, 174(1-2):391–432, 2019.

[5] H. H. Bauschke and P. Combettes. *Convex analysis and monotone operators theory in Hilbert spaces.* Springer-Verlag, 2nd edition, 2017.

[6] H. H. Bauschke, W. M. Moursi, and X. Wang. Generalized monotone operators and their averaged resolvents. *Math. Program.*, pages 1–20, 2020.

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

[8] S. Bubeck, Y. T. Lee, and M. Singh. A geometric alternative to Nesterov's accelerated gradient descent. *arXiv preprint arXiv:1506.08187*, 2015.

[9] R. S. Burachik and A. Iusem. *Set-Valued Mappings and Enlargements of Monotone Operators.* New York: Springer, 2008.

[10] X. Cai, C. Song, C. Guzmán, and J. Diakonikolas. Stochastic halpern iteration with variance reduction for stochastic monotone inclusions. *Advances in Neural Information Processing Systems*, 35:24766–24779, 2022.

[11] Y. Cai, A. Oikonomou, and W. Zheng. Accelerated algorithms for constrained nonconvex-nonconcave min-max optimization and comonotone inclusion. In *Forty-first International Conference on Machine Learning*, 2022.

[12] Y. Cai, A. Oikonomou, and W. Zheng. Tight last-iterate convergence of the extragradient and the optimistic gradient descent-ascent algorithm for constrained monotone variational inequalities. *arXiv preprint arXiv:2204.09228*, 2022.

[13] Y. Cai and W. Zheng. Accelerated single-call methods for constrained min-max optimization. *arXiv preprint arXiv:2210.03096*, 2022.

[14] A. Chambolle and C. Dossal. On the convergence of the iterates of the "Fast iterative shrinkage/thresholding algorithm". *J. Optim. Theory Appl.*, 166(3):968–982, 2015.

[15] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4:1168–1200, 2005.

[16] D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set-valued and Variational Analysis*, 25(4):829–858, 2017.

[17] J. Diakonikolas. Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. In *Conference on Learning Theory*, pages 1428–1451. PMLR, 2020.

[18] J. Diakonikolas, C. Daskalakis, and M. Jordan. Efficient methods for structured nonconvex-nonconcave min-max optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 2746–2754. PMLR, 2021.

[19] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems*, volume 1-2. Springer-Verlag, 2003.

[20] B. Halpern. Fixed points of nonexpanding maps. *Bull. Am. Math. Soc.*, 73(6):957–961, 1967.

[21] B. He and X. Yuan. On the convergence rate of Douglas–Rachford operator splitting method. *Math. Program.*, 153(2):715–722, 2015.

[22] D. Kim. Accelerated proximal point method for maximally monotone operators. *Math. Program.*, pages 1–31, 2021.

[23] D. Kim and J. A. Fessler. Optimized first-order methods for smooth convex minimization. *Math. Program.*, 159(1-2):81–107, 2016.

[24] G. M. Korpelevic. An extragradient method for finding saddle-points and for other problems. *Èkonom. i Mat. Metody.*, 12(4):747–756, 1976.

[25] S. Lee and D. Kim. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. *Thirty-fifth Conference on Neural Information Processing Systems (NeurIPs2021)*, 2021.

[26] F. Lieder. On the convergence rate of the halpern-iteration. *Optimization Letters*, 15(2):405–418, 2021.

[27] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Num. Anal.*, 16:964–979, 1979.

[28] P.-E. Maingé. Accelerated proximal algorithms with a correction term for monotone inclusions. *Applied Mathematics & Optimization*, 84(2):2027–2061, 2021.

[29] P. E. Maingé. Fast convergence of generalized forward-backward algorithms for structured monotone inclusions. *arXiv preprint arXiv:2107.10107*, 2021.

[30] Y. Malitsky. Projected reflected gradient methods for monotone variational inequalities. *SIAM Journal on Optimization*, 25(1):502–520, 2015.

[31] A. Nemirovskii and D. Yudin. *Problem Complexity and Method Efficiency in Optimization.* Wiley Interscience, 1983.

[32] Y. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $\mathcal{O}(1/k^2)$. *Doklady AN SSSR*, 269:543–547, 1983. Translated as Soviet Math. Dokl.

[33] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87 of *Applied Optimization.* Kluwer Academic Publishers, 2004.

[34] Y. Nesterov. Smooth minimization of non-smooth functions. *Math. Program.*, 103(1):127–152, 2005.

[35] Y. Ouyang and Y. Xu. Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems. *Math. Program.*, online first:1–35, 2019.

[36] J. Park and E. K. Ryu. Exact optimal accelerated complexity for fixed-point iterations. *arXiv preprint arXiv:2201.11413*, 2022.

[37] R. R. Phelps. *Convex functions, monotone operators and differentiability*, volume 1364. Springer, 2009.

[38] B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

[39] L. D. Popov. A modification of the Arrow-Hurwicz method for search of saddle points. *Mathematical notes of the Academy of Sciences of the USSR*, 28(5):845–848, 1980.

[40] R. Rockafellar and R. Wets. *Variational Analysis*, volume 317. Springer, 2004.

[41] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14:877–898, 1976.

[42] E. K. Ryu and S. Boyd. Primer on monotone operator methods. *Appl. Comput. Math*, 15(1):3–43, 2016.

[43] B. Shi, S. S. Du, M. I. Jordan, and W. Su. Understanding the acceleration phenomenon via high-resolution differential equations. *Math. Program.*, pages 1–70, 2021.

[44] W. Su, S. Boyd, and E. Candes. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2510–2518, 2014.

[45] Q. Tran-Dinh and Y. Luo. Halpern-type accelerated and splitting algorithms for monotone inclusions. *arXiv preprint arXiv:2110.08150*, 2021.

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

[47] T. Yoon and E. K. Ryu. Accelerated algorithms for smooth convex-concave minimax problems with $\mathcal{O}(1/k^2)$ rate on squared gradient norm. In *International Conference on Machine Learning*, pages 12098–12109. PMLR, 2021.