Unifying Halpern's Fixed-Point Iteration and Nesterov's Acceleration

Chris Junchi Li[◇]

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

October 5, 2024

Abstract

This paper explores the connection between two prominent techniques for solving monotone inclusions and root-finding problems: Halpern's fixed-point iteration and Nesterov's accelerated methods. Both methods have been widely studied for their efficiency in optimization, yet their relationship has not been fully understood. We demonstrate that these schemes are not only related but can be transformed into one another under specific parameter choices. We provide a unified framework for analyzing their convergence rates and extend our results to several popular algorithms, including proximal-point, forward-backward, and Douglas-Rachford splitting methods. This work offers new insights into the theoretical underpinnings of acceleration techniques in nonconvex optimization and monotone inclusion problems.

Keywords: Halpern's Fixed-Point Iteration, Nesterov's Acceleration, Monotone Inclusions, Convergence Rates, Root-Finding Problems

1 Introduction

Approximating solutions to maximally monotone inclusions is a fundamental problem in optimization, with applications spanning nonlinear analysis, mechanics, and machine learning. Among the various methods proposed to address this challenge, Halpern's fixed-point iteration and Nesterov's accelerated schemes have emerged as two leading approaches. However, despite their independent development and the growing body of literature on each, the relationship between these methods has remained unclear.

In recent years, accelerated methods have gained significant attention due to their effectiveness in large-scale optimization problems. Nesterov's acceleration, in particular, has been widely applied to problems involving convex optimization, minimax formulations, and variational inequalities. Its success in improving convergence rates from $O(1/k)$ to $O(1/k^2)$ has made it a cornerstone in optimization research. On the other hand, Halpern's method, although developed earlier, has recently been recognized for its potential in improving the convergence rate of fixed-point iterations. This raises an important question: what is the underlying relationship between these two seemingly distinct approaches?

In this paper, we bridge this gap by showing that Halpern's fixed-point iteration can be transformed into Nesterov's accelerated interpretation, and vice versa. By analyzing these schemes within a unified framework, we derive convergence guarantees that shed light on their similarities and differences. Our results extend beyond the theoretical comparison of these two methods, offering new insights into the acceleration of various classical algorithms such as proximal-point and forward-backward splitting methods.

Backgrounds 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., [6, 12, 19, 40, 43, 44, 45]. This problem lies at the heart of monotone operator theory, and has been intensively studied in the literature for many decades, see, e.g., [6, 29, 44, 45] as a few examples. Various numerical methods, including proximal-point-type, gradient/forward, extragradient, past-extragradient, and their variants have been proposed to solve this problem, and its extensions as well as special cases [6, 15, 19, 33, 34, 42, 44]. When the underlying operator is the sum of two or multiple 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 solutions of this problem under different assumptions and context, see, e.g., [6, 14, 15, 29, 32, 42, 51] as a few references.

Motivation and related work In the last decades, accelerated first-order methods have become an extremely popular and attractive research topic in optimization and related fields due to their applications to large-scale optimization problems in machine learning, statistics, signal and image processing, and engineering, see, e.g., [8,35,36,37]. In this research theme, Nesterov's accelerated approach [35] presents as a leading research topic for many years, and remains emerging in optimization community. This well-known technique has been extended to different directions, including minimax problems, variational inequalities (VIPs), and monotone inclusions [1, 10, 22, 31]. Convergence rates of these methods have been intensively studied, which show significant improvements from $\mathcal{O}(1/k)$ to $\mathcal{O}(1/k^2)$ rates, where k is the iteration counter. The latter rate matches the lower bound rates in different settings using different criteria, see, e.g., [23, 36, 38]. In recent years, many papers, including [1, 22, 31], have focused on developing Nesterov's accelerated schemes for monotone inclusions. They have proven $\mathcal{O}(1/k^2)$-convergence rates, and also $o(1/k^2)$ convergence rates on the squared norm of the residual operator associated with the problem, while obtaining asymptotic convergence on iterate sequences, see [3, 13, 23]. Note that the problem of approximating a solution of a maximally monotone inclusion can be reformulated equivalently to approximating a fixed-point of a nonexpansive operator [6]. Therefore, theory and solution methods from one field can be applied to the other and vice versa.

Alternatively, Halpern's fixed-point iteration is a classical method in fixed-point theory rooted from [20] to approximate a fixed-point of a nonexpansive operator, see [5, 24, 53]. This method has recently attracted great attention due to its ability to accelerate convergence rate in terms of operator residual norm. Lieder specifically proved an $\mathcal{O}(1/k^2)$ rate on the squared norm of the operator residual for Halpern's fixed-point iteration in [28], but [46] appears to be the first one achieving this rate for a variant of Halpern's fixed-point method. Unlike Nesterov's accelerated method which is originally developed for solving convex optimization problems and its convergence rate is given in terms of the objective residual in the general convex case, Halpern's fixed-point iteration is proposed to approximate a fixed-point of a nonexpansive operator, which is much more general than convex optimization, and hence, very convenient to extend to maximally monotone inclusions, and, in particular, minimax problems, game theory, robust optimization, online learning, and reinforcement learning, see, e.g., [16, 27, 55].

A natural question is therefore arising: ***What is the relation between the two accelerated techniques?*** Such a type of questions was mentioned in [55]. Since both schemes come from different roots, at first glance, it is unclear to see a close relation between Nesterov's accelerated and the Halpern fixed-point schemes. Nesterov's accelerated method is perhaps rooted from the gradient descent scheme in convex optimization with an additional momentum term [41]. Its acceleration behavior has been explained through different view points, including geometric interpretation [11] and continuous views via ordinary differential equations (ODEs) and variational perspectives [2, 4, 47, 48, 52]. In addition, Nesterov's accelerated method has various variants [4, 22, 23, 31, 47]. For instance, [23] derived an "optimized" Nesterov's accelerated variant to solve composite convex optimization problems which is slightly different from the original one in [36]. As other examples, both [22, 31] proposed Nesterov's accelerated schemes using different "correction" terms to solve monotone inclusions. Alternatively, Halpern's fixed-point method was proposed in [20] since 1967, existing convergence guarantees are essentially asymptotic or slow convergence rates [53]. Its accelerated rate has just recently been established in [28, 46] and followed up by a number of works, including [16, 27, 55]. Interestingly, the analysis of both schemes is quite different, but still relies on appropriate Lyapunov or potential functions, and variable parameters (e.g., stepsizes, extrapolation, and momentum parameters).

Our contribution In this paper, we show that Halpern's fixed-point method can be transformed into Nesterov's accelerated interpretation and vice versa. We first present our results on approximating a solution of a co-coercive equation, and then extend them to other schemes. In the first case, we establish that the iterate sequences generated by both schemes are identical, but the choice of parameters in these schemes could be different, leading to different convergence guarantees. In fact, we can obtain the convergence guarantee of one scheme from another and vice versa. Then, by utilizing our analysis, we prove that a number of methods, including proximal-point, forward-backward splitting, Douglas-Rachford splitting, and three-operator splitting schemes can be easily accelerated and achieve faster rates compared to their classical counterparts.

- Note that, in convex optimization, we often use the objective residual as a potential energy term to form a Lyapunov or an energy function for establishing convergence rates. Moreover, this objective residual presents as a main metric to measure the approximate optimality of the current iterate. However, such an objective function does not exist in root-finding problems. Therefore, establishing $\mathcal{O}(1/k^2)$ and $o(1/k^2)$ convergence rates for a root-finding problem requires different metric as well as different techniques than those in convex optimization.

- Our next contribution is to show that the recent extra-anchored gradient (EAG) proposed in [27, 55] can be transformed into Nesterov's accelerated interpretation. We provide an alternative analysis using a slightly different Lyapunov function and still obtain the same convergence rates as in [27, 55]. One important fact is that this accelerated method works with a monotone and Lipschitz continuous operator instead of a co-coercive one as in [1, 22, 31]. This approach is expected to provide an initial step toward understanding Nesterov's accelerated behavior on non-co-coercive operators, and possibly their continuous views.

- Finally, we derive Nesterov's accelerated variant of the past-extra anchored gradient method in [50] and provide a different convergence rate analysis than that of [50] by using two different stepsizes for the extra-gradient step. Interestingly, such a new scheme is very different from existing ones, e.g., [1, 22, 31], due to the use of two consecutive past iterates in the momentum/correction terms. To the best of our knowledge, this algorithm is the first one illustrating that we can build Nesterov's accelerated scheme for monotone equations without co-coerciveness.

Paper organization The rest of this paper is organized as follows. Section 2 provides some necessary background of monotone operators, which will be used in the sequel. Section 3 presents the equivalence between Halpern's fixed-point iteration and Nesterov's accelerated scheme for solving a co-coercive equation. Next, we discuss its application to other methods in Sect. 4. Then, Sect. 5 investigates the connection between the extra-anchored gradient method and its variants and the corresponding Nesterov's interpretations.

2 Background of Monotone operators

We work with finite-dimensional spaces $\mathbb{R}^p$ and $\mathbb{R}^n$ endowed with the standard inner product $\langle \cdot, \cdot \rangle$ and Euclidean norm $\| \cdot \|$. For a set-valued mapping $G : \mathbb{R}^p \rightrightarrows 2^{\mathbb{R}^p}$, $\mathrm{dom}(G) = \{x \in \mathbb{R}^p : Gx \neq \emptyset\}$ denotes its domain, $\mathrm{gra}(G) = \{(x, y) \in \mathbb{R}^p \times \mathbb{R}^p : y \in Gx\}$ denotes its graph, where $2^{\mathbb{R}^p}$ is the set of all subsets of $\mathbb{R}^p$. The inverse of G is defined as $G^{-1}y := \{x \in \mathbb{R}^p : y \in Gx\}$.

Monotonicity For a set-valued mapping $G : \mathbb{R}^p \rightrightarrows 2^{\mathbb{R}^p}$, we say that G is monotone if $\langle u - v, x - y \rangle \geq 0$ for all $(x, u), (y, v) \in \mathrm{gra}(G)$.$G$ is said to be μ_G-strongly monotone (also called coercive) if $\langle u - v, x - y \rangle \geq \mu_G \|x - y\|^2$ for all $(x, u), (y, v) \in \mathrm{gra}(G)$, where $\mu_G > 0$ is called the strong monotonicity parameter. If G is single-valued, then these conditions reduce to $\langle Gx - Gy, x - y \rangle \geq 0$ and $\langle Gx - Gy, x - y \rangle \geq \mu_G \|x - y\|^2$ for all $x, y \in \mathrm{dom}(G)$, respectively. We say that G is maximally monotone if $\mathrm{gra}(G)$ is not properly contained in the graph of any other monotone operator. Further details can be found, e.g., in [6, 45].

Lipschitz continuity and co-coerciveness A single-valued operator G is said to be L-Lipschitz continuous if $\|Gx - Gy\| \leq L\|x - y\|$ for all $x, y \in \mathrm{dom}(G)$, where $L \geq 0$ is the Lipschitz constant. If $L = 1$, then we say that G is nonexpansive, while if $L \in [0, 1)$, then we say that G is L-contractive, and L is its contraction factor. We say that G is $\frac{1}{L}$-co-coercive if $\langle Gx - Gy, x - y \rangle \geq \frac{1}{L} \|Gx - Gy\|^2$ for all $x, y \in \mathrm{dom}(G)$. If $L = 1$, then we say that G is firmly nonexpansive. Note that if G is $\frac{1}{L}$-co-coercive, then it is also monotone and L-Lipschitz continuous, but the reverse is not true in general. If $L < 0$, then we say that G is $\frac{1}{L}$-co-monotone [7] (also known as $-\frac{1}{L}$-cohypomonotone).

Resolvent operator The operator $J_G x := \{y \in \mathbb{R}^p : x \in y + Gy\}$ is called the resolvent of G, often denoted by $J_G x = (\mathbb{I} + G)^{-1} x$, where $\mathbb{I}$ is the identity mapping. Clearly, evaluating J_G requires solving a strongly monotone inclusion $0 \in y - x + Gy$ in y for given x. If G is monotone, then J_G is singled-valued, and if G is maximally monotone, then J_G is singled-valued and $\mathrm{dom}\,(J_G) = \mathbb{R}^p$. If G is monotone, then J_G is firmly nonexpansive [6, Proposition 23.10].

3 The equivalence of Halpern's and Nesterov's accelerated schemes

Both Nesterov's accelerated and Halpern's fixed-point iteration schemes show significant improvement on convergence rates over classical methods for solving (29). 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.

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

$$\text{Find } y^\star \in \mathrm{dom}(G) \text{ such that: } Gy^\star = 0 \tag{1}$$

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

3.1 The Halpern fixed-point scheme and its convergence

The Halpern fixed-point scheme [16, 20, 28] for solving (1) is written as follows:

$$y_{k+1} := \beta_k y_0 + (1-\beta_k)y_k - \eta_k G y_k \tag{2}$$

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

The convergence rate of (2) has been established in [16,28] using different tools. While [28] provides a direct proof and uses a performance estimation problem approach to establish convergence of (2), [16] exploits a Lyapunov's technique to analyze its convergence rate. Let us summarize the result in [16] in our context.

The standard Lyapunov function to study (2) is

$$\mathcal{L}_k := \frac{p_k}{L}\|Gy_k\|^2 + q_k \langle Gy_k, y_k - y_0\rangle \tag{3}$$

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 is from [16, 28] and states the convergence rate of (2).

Theorem 1 ([16, 28]**.** *] Assume that G in (1) is $\frac{1}{L}$-co-coercive with $L \in (0, +\infty)$, and $y^\star \in \mathrm{zer}(G)$. Let $\{y_k\}$ be generated by (2) using $\beta_k := \frac{1}{k+2}$ and $\eta_k := \frac{2(1-\beta_k)}{L}$. Then*

$$\|Gy_k\| \le \frac{L\|y_0 - y^\star\|}{k+1} \tag{4}$$

Remark 1. *If we choose $\beta_k := \frac{1}{k+2}$ and $\eta_k := \frac{1-\beta_k}{L}$, then using a similar proof as in Theorem 1 from* [16, 28]*, we can show that*

$$\|Gy_k\|^2 \le \frac{4L^2\|y_0 - y^\star\|^2}{(k+1)(k+3)} \quad \text{and} \quad \sum_{k=0}^{\infty}(k+1)(k+2)\,\|Gy_{k+1} - Gy_k\|^2 \le 2L^2\|y_0 - y^\star\|^2$$

However, if we choose $\eta_k := \frac{2(1-\beta_k)}{L}$ as in Theorem 1, then we do not obtain the last summable inequality.

Note that if $\beta_k := \frac{1}{k+2}$, then we can rewrite (2) into the Halpern fixed-point iteration as in [28]:

$$y_{k+1} := \frac{1}{k+2}y_0 + \left(1 - \frac{1}{k+2}\right)Ty_k, \qquad \text{where } Ty_k := y_k - \frac{2}{L}Gy_k \tag{5}$$

Since G is $\frac{1}{L}$-co-coercive, $T = \mathbb{I} - \frac{2}{L}G$ is nonexpansive, see [6, Proposition 4.11]. Therefore, (2) is equivalent to the scheme studied in [28], and Theorem 1 can be obtained from the results in [28]. The choice of β_k and η_k in Theorem 1 are tight and the bound (4) is unimprovable since there exists an example that achieves this rate as the lower bound, see, e.g., [28] for such an example.

3.2 The equivalence between Halpern's and Nesterov's accelerated schemes

Our next step is to show that Halpern's fixed-point iteration (2) 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 $x_0 = y_{-1} := y_0, \eta_{-1} = \gamma_{-1} = 0$, and $\beta_{-1} > 0$:*

$$\begin{cases} x_{k+1} := y_k - \gamma_k G y_k \\ y_{k+1} := x_{k+1} + \theta_k (x_{k+1} - x_k) + v_k (y_k - x_{k+1}) + \kappa_k (y_{k-1} - x_k) \end{cases} \tag{6}$$

where $\theta_k := \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}$, $v_k := \frac{\beta_k}{\beta_{k-1}} + 1 - \beta_k - \frac{\eta_k}{\gamma_k}$, 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 (2) for solving (1).

In particular, if $\gamma_k := \frac{\eta_k}{1-\beta_k}$, then $v_k = \frac{\beta_k}{\beta_{k-1}}, \kappa_k = 0$, and (6) reduces to

$$\begin{cases} x_{k+1} := y_k - \gamma_k G y_k \\ y_{k+1} := x_{k+1} + \theta_k (x_{k+1} - x_k) + v_k (y_k - x_{k+1}) \end{cases} \tag{7}$$

Both (6) and (7) can be viewed as Nesterov's accelerated variants with correction terms. Here, (6) has two correction terms $v_k (y_k - x_{k+1})$ and $\kappa_k (y_{k-1} - x_k)$, while (7) has only one term $v_k (y_k - x_{k+1})$. In fact, (6) shows that (2) is equivalent to Nesterov's accelerated scheme with gradient correction in [47] as shown in Remark 2 below. Alternatively, compared to the "optimized gradient method" (OGM1) in [23] for convex optimization, our coefficient v_k in (7) is positive in contrast to a negative value in OGM1. Note that (7) covers the proximal-point scheme in [30] as a special case. As discussed in [2], (7) can be viewed as a variant of Ravine's method if the convergence rate is given in y_k instead of x_k.

Proof of Theorem 2. **[(6) $\Rightarrow$(2)]** Substituting θ_k, v_k, and κ_k into (6), and simplifying the result, we get

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

Now, using the first line of (6) into this expression, we get

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

Rearranging this expression, we arrive at

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

By induction, and noticing that $y_{-1} = y_0, \beta_{-1} > 0$, and $\eta_{-1} = 0$, this expression leads to $\frac{1}{\beta_k} y_{k+1} - \left(\frac{1}{\beta_k} - 1\right) y_k + \frac{\eta_k}{\beta_k} G y_k = y_0$, which is indeed equivalent to (2).

[(2) $\Rightarrow$(6)] First, shifting the index from k to $k-1$ in (2), we have $y_k = \beta_{k-1} y_0 + (1-\beta_{k-1}) y_{k-1} - \eta_{k-1} G y_{k-1}$. Here, we assume that $y_{-1} = y_0$. Multiplying this expression by $-\beta_k$ and (2) by β_{k-1} and adding the results, we obtain

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

This expression leads to

$$y_{k+1} = \left(\frac{\beta_k}{\beta_{k-1}} + 1 - \beta_k\right) y_k - \eta_k G y_k - \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}} y_{k-1} + \frac{\beta_k \eta_{k-1}}{\beta_{k-1}} G y_{k-1} \tag{8}$$

Next, let us introduce $x_{k+1} := y_k - \gamma_k G y_k$ for some $\gamma_k > 0$. Then, we have $Gy_k = \frac{1}{\gamma_k}(y_k - x_{k+1})$. Substituting this relation into (8), we obtain

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

If we let $\theta_k := \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}, \upsilon_k := \frac{\beta_k}{\beta_{k-1}} + 1 - \beta_k - \frac{\eta_k}{\gamma_k}$ 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 $y_{k+1} = x_{k+1} + \theta_k(x_{k+1} - x_k) + \upsilon_k(y_k - x_{k+1}) + \kappa_k(y_{k-1} - x_k)$. Combining this line and $x_{k+1} = y_k - \gamma_k G y_k$, we get (6).

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, (6) reduces to (7). □

Remark 2. *Using (8), we can rewrite the Halpern-type scheme (2) equivalently to*

$$y_{k+1} := y_k + \theta_k(y_k - y_{k-1}) - r_k G y_k - s_k(Gy_k - Gy_{k-1}),$$

where $\theta_k := \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}, r_k := \eta_k - \frac{\beta_k\eta_{k-1}}{\beta_{k-1}}$, and $s_k := \frac{\beta_k\eta_{k-1}}{\beta_{k-1}} > 0$. This expression shows that, under an appropriate choice of parameters, (2) is equivalent to Nesterov's accelerated scheme with gradient correction studied in [47], when $Gy = \nabla f(y)$, the gradient of a convex function f. In particular, if $\beta_k = \frac{1}{k+2}$ and $\eta_k := \frac{2(1-\beta_k)}{L}$ as in Theorem 1, then r_k reduces to $r_k = \frac{2}{(k+2)L} > 0$.

Remark 3. *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 (6) reduces to the following one:*

$$\begin{cases} x_{k+1} := y_k - \frac{1}{L}Gy_k \\ y_{k+1} := x_{k+1} + \frac{k}{k+2}(x_{k+1} - x_k) + \frac{k+1}{k+2}(y_k - x_{k+1}) \end{cases} \tag{9}$$

Alternatively, if $\gamma_k := \frac{1}{L}, \beta_k := \frac{1}{k+2}$, and $\eta_k := \frac{2(1-\beta_k)}{L}$, then (6) becomes

$$\begin{cases} x_{k+1} := y_k - \frac{1}{L}Gy_k \\ y_{k+1} := x_{k+1} + \frac{k}{k+2}(x_{k+1} - x_k) + \frac{k}{k+2}(y_{k-1} - x_k) \end{cases} \tag{10}$$

The scheme (10) covers [22] as a special case when $Gy = J_{\lambda A}y$, the resolvent of a maximally monotone operator λA, which is firmly nonexpansive.

Utilizing Theorems 1 and 2, we obtain the following corollary showing the convergence rate of (6) under a particular choice of parameters. However, this choice of parameters is not necessarily unique as shown in Remark 1.

Corollary 1. *Assume that G in (1) is $\frac{1}{L}$-co-coercive and $\mathrm{zer}(G) \neq \emptyset$. Let $\{(x_k, y_k)\}$ be generated by (6) with $\gamma_k := \frac{1}{L}$, $\beta_k := \frac{1}{k+2}$, and $\eta_k := \frac{1-\beta_k}{L}$. Then, we obtain $\theta_k := \frac{k}{k+2}$, $\upsilon_k := \frac{k+1}{k+2}$, and $\kappa_k := 0$. Moreover, (6) reduces to (7) (or equivalently (9)), and the following guarantee holds:*

$$\|Gy_k\|^2 \leq \frac{4L^2\|y_0 - y^\star\|^2}{(k+1)(k+3)}$$
$$\sum_{k=0}^{\infty}(k+1)(k+2)\,\|Gy_{k+1} - Gy_k\|^2 \leq 2L^2\|y_0 - y^\star\|^2 \tag{11}$$

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 $\theta_k := \frac{k}{k+2}, \upsilon_k := 0$, and $\kappa_k := \frac{k}{k+2}$, and (6) reduces to (10). Moreover, one has $\|Gy_k\| \leq \frac{L\|y_0 - y^\star\|}{k+1}$.

The constant factor in the bound (11) is slightly worse than the one in $\|Gy_k\| \leq \frac{L\|y_0-y^\star\|}{k+1}$. In fact, the latter one is exactly optimal since there exists an example showing that it matches the lower bound complexity, see, e.g. [16, 28].

3.3 Convergence analysis of Nesterov's accelerated scheme (7)

Now, we provide a direct analysis of (7) without using Theorem 1. For simplicity, we will analyze the convergence of (7) with only one correction term. However, our analysis can be easily extended to (6) when $\kappa_k \neq 0$ with some simple modifications.

Our analysis relies on the following Lyapunov function:

$$\mathcal{V}_k := a_k\|Gy_{k-1}\|^2 + b_k\,\langle Gy_{k-1}, x_k - y_k\rangle + \|x_k + t_k(y_k - x_k) - y^\star\|^2 + \mu\|x_k - y^\star\|^2 \tag{12}$$

where $a_k, b_k, t_k > 0$, and $\mu \geq 0$ are given parameters, which will be determined later. This Lyapunov is slightly different from $\mathcal{L}_k$ defined by (3), but it is closely related to standard Nesterov's potential function (see, e.g., [3, 18, 47]). To see the connection between $\mathcal{V}_k$ and $\mathcal{L}_k$, we prove the following lemma.

Lemma 1. *Let $\mathcal{L}_k$ be defined by (3) and $\mathcal{V}_k$ be defined by (12). 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*

$$\mathcal{V}_{k+1} = \frac{4p_k}{Lq_k^2}\mathcal{L}_k + \|y_0 - y^\star\|^2 + \mu\|x_{k+1} - y^\star\|^2 \tag{13}$$

Before proving this lemma, we make the following remarks.

- Note that the proof of Theorem 1 uses $\mathcal{L}_k$ defined by (3) with $p_k = q_0k(k+1)$ and $q_k = q_0(k+1)$. In this case, we have $\frac{4p_k}{Lq_k^2} = \frac{4k}{Lq_0(k+1)} \approx \frac{4}{Lq_0}$. Hence, we can show that $\mathcal{V}_{k+1} = \frac{4k}{Lq_0(k+1)}\mathcal{L}_k + \|y_0 - y^\star\|^2 + \mu\|x_{k+1} - y^\star\|^2$.
- 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 $\mathcal{V}_{k+1} = \frac{4c}{L}\mathcal{L}_k + \|y_0 - y^\star\|^2$, showing that $\mathcal{V}_k$ is equivalent to $\mathcal{L}_k$.
- The term $\mu\|x_k - y^\star\|^2$ allows us to get the tail $\|x_{k+1} - x_k\|^2$ in (22) of Lemma 3, which is a key to prove convergence in Theorem 3, especially the summable results and the $o(1/k^2)$-convergence rates. It remains unclear to us how to prove such a convergence rate without using the term $\mu\|x_k - y^\star\|^2$.

Proof of Lemma 1. First, from $\mathcal{L}_k = \frac{p_k}{L}\|Gy_k\|^2 + q_k \langle Gy_k, y_k - y_0\rangle$ in (3), we can write it as

$$\mathcal{L}_k = \frac{p_k}{L}\left\|Gy_k - \frac{Lq_k}{2p_k}(y_0 - y^\star)\right\|^2 + q_k \langle Gy_k, y_k - y^\star\rangle - \frac{Lq_k^2}{4p_k}\|y_0 - y^\star\|^2$$

Second, from (2), we have $y_0 - y^\star = \frac{1}{\beta_k}(y_{k+1} + \eta_k Gy_k - (1-\beta_k)y_k) - y^\star = y_k - y^\star + \frac{1}{\beta_k}(y_{k+1} - y_k) + \frac{\eta_k}{\beta_k}Gy_k$. Using $x_{k+1} = y_k - \frac{\eta_k}{1-\beta_k}Gy_k$, we have $y_k = x_{k+1} + \frac{\eta_k}{1-\beta_k}Gy_k$. Combining these lines, we get

$$Gy_k - \frac{Lq_k}{2p_k}(y_0 - y^\star) = Gy_k - \frac{Lq_k}{2p_k}\left[x_{k+1} + \frac{1}{\beta_k}(y_{k+1} - x_{k+1}) - y^\star\right]$$

Now, let $z_{k+1} := x_{k+1} + \frac{1}{\beta_k}(y_{k+1} - x_{k+1})$. Then, we can rewrite $\mathcal{L}_k$ as

$$\mathcal{L}_k = \frac{p_k}{L}\left\|Gy_k - \frac{Lq_k}{2p_k}(z_{k+1} - y^\star)\right\|^2 + q_k \langle Gy_k, y_k - y^\star\rangle - \frac{Lq_k^2}{4p_k}\|y_0 - y^\star\|^2$$

Since $y_k = x_{k+1} + \frac{\eta_k}{1-\beta_k}Gy_k$, using the definition of z_{k+1}, we have $y_k - z_{k+1} = x_{k+1} - z_{k+1} + \frac{\eta_k}{1-\beta_k}Gy_k = \frac{1}{\beta_k}(x_{k+1} - y_{k+1}) + \frac{\eta_k}{1-\beta_k}Gy_k$. Further expanding the first term of $\mathcal{L}_k$ and using this relation, we can easily show that

$$\begin{aligned}
\mathcal{L}_k =& \frac{p_k}{L}\|Gy_k\|^2 + q_k\langle Gy_k, y_k - z_{k+1}\rangle + \frac{Lq_k^2}{4p_k}\|z_{k+1} - y^\star\|^2 - \frac{Lq_k^2}{4p_k}\|y_0 - y^\star\|^2 \\
=& \frac{Lq_k^2}{4p_k}\left[\left(\frac{4p_k^2}{L^2q_k^2} + \frac{4p_k\eta_k}{Lq_k(1-\beta_k)}\right)\|Gy_k\|^2\right. \\
&\left. + \frac{4p_k}{Lq_k\beta_k}\langle Gy_k, x_{k+1} - y_{k+1}\rangle + \|z_{k+1} - y^\star\|^2\right] \\
&- \frac{Lq_k^2}{4p_k}\|y_0 - y^\star\|^2
\end{aligned}$$

Finally, this expression together with (12) imply (13). □

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

Lemma 2. *Let $\{(x_k, y_k)\}$ be generated by (7) using $\gamma_k := \gamma > 0$, and $\mathcal{V}_k$ be defined by (12). Then, if $b_{k+1}\theta_k + 2\gamma t_k(t_k - 1) - 2\gamma v_k\theta_k t_{k+1}^2 \geq 0$, then we have*

$$\begin{aligned}
\mathcal{V}_k - \mathcal{V}_{k+1} \geq & \left(\gamma b_{k+1}v_k + \gamma^2 t_k^2 - \gamma^2 t_{k+1}^2 v_k^2 - a_{k+1} - \frac{\gamma^2 b_k^2}{4a_k}\right)\|Gy_k\|^2 \\
& + \left[b_{k+1}\theta_k + 2\gamma t_k(t_k-1) - 2\gamma v_k\theta_k t_{k+1}^2 - b_k\right]\langle Gy_{k-1}, x_{k+1} - x_k\rangle \\
& + \left(t_k^2 - 2t_k + 1 + \mu - t_{k+1}^2\theta_k^2\right)\|x_{k+1} - x_k\|^2 \\
& + \left(\frac{1}{L} - \gamma\right)\left[b_{k+1}\theta_k + 2\gamma t_k(t_k-1) - 2\gamma v_k\theta_k t_{k+1}^2\right]\|Gy_k - Gy_{k-1}\|^2 \\
& + 2(t_k - t_{k+1}\theta_k - 1 - \mu)\langle x_{k+1} - x_k, x_{k+1} - y^\star\rangle \\
& + 2\gamma(t_k - t_{k+1}v_k)\langle Gy_k, x_{k+1} - y^\star\rangle \\
& + a_k\left\|Gy_{k-1} - \frac{\gamma b_k}{2a_k}Gy_k\right\|^2
\end{aligned} \tag{14}$$

Proof. Firstly, from (12), we have

$$\begin{aligned}\mathcal{V}_k - \mathcal{V}_{k+1} =& a_k\|Gy_{k-1}\|^2 - a_{k+1}\|Gy_k\|^2 + b_k \langle Gy_{k-1}, x_k - y_k\rangle \\ &- b_{k+1}\langle Gy_k, x_{k+1} - y_{k+1}\rangle + \mu\|x_k - y^\star\|^2 - \mu\|x_{k+1} - y^\star\|^2 \\ &+ \|x_k - y^\star + t_k(y_k - x_k)\|^2 - \|x_{k+1} - y^\star + t_{k+1}(y_{k+1} - x_{k+1})\|^2 \end{aligned} \tag{15}$$

Next, since $y_k = x_{k+1} + \gamma G y_k$ from (7), it is easy to show that

$$\langle Gy_{k-1}, x_k - y_k\rangle = -\langle Gy_{k-1}, x_{k+1} - x_k\rangle - \gamma\langle Gy_{k-1}, Gy_k\rangle \tag{16}$$

Similarly, from (7) we have $x_{k+1} - y_{k+1} = -\theta_k(x_{k+1} - x_k) - \gamma v_k G y_k$, leading to

$$\langle Gy_k, x_{k+1} - y_{k+1}\rangle = -\theta_k\langle Gy_k, x_{k+1} - x_k\rangle - \gamma v_k\|Gy_k\|^2. \tag{17}$$

Then, using again $y_k = x_{k+1} + \gamma G y_k$ from (7), we can derive

$$\begin{aligned}\|x_k - y^\star + t_k(y_k - x_k)\|^2 =& \|x_{k+1} - y^\star + (t_k - 1)(x_{k+1} - x_k) + \gamma t_k G y_k\|^2 \\ =& \|x_{k+1} - y^\star\|^2 + (t_k - 1)^2\|x_{k+1} - x_k\|^2 \\ &+ \gamma^2 t_k^2\|Gy_k\|^2 + 2(t_k - 1)\langle x_{k+1} - x_k, x_{k+1} - y^\star\rangle \\ &+ 2\gamma t_k(t_k - 1)\langle Gy_k, x_{k+1} - x_k\rangle + 2\gamma t_k\langle Gy_k, x_{k+1} - y^\star\rangle. \end{aligned} \tag{18}$$

Similarly, using $y_{k+1} - x_{k+1} = \theta_k(x_{k+1} - x_k) + \gamma v_k G y_k$, we can show that

$$\begin{aligned}\|x_{k+1} - y^\star + t_{k+1}(y_{k+1} - x_{k+1})\|^2 =& \|x_{k+1} - y^\star + t_{k+1}\theta_k(x_{k+1} - x_k) + \gamma t_{k+1} v_k G y_k\|^2 \\ =& \|x_{k+1} - y^\star\|^2 + t_{k+1}^2\theta_k^2\|x_{k+1} - x_k\|^2 \\ &+ 2\gamma v_k\theta_k t_{k+1}^2\langle Gy_k, x_{k+1} - x_k\rangle \\ &+ 2t_{k+1}\theta_k\langle x_{k+1} - x_k, x_{k+1} - y^\star\rangle \\ &+ 2\gamma t_{k+1} v_k\langle Gy_k, x_{k+1} - y^\star\rangle + \gamma^2 t_{k+1}^2 v_k^2\|Gy_k\|^2 \end{aligned} \tag{19}$$

Substituting (16), (17), (18), and (19) into (15), and using $\mu\|x_k - y^\star\|^2 - \mu\|x_{k+1} - y^\star\|^2 = \mu\|x_{k+1} - x_k\|^2 - 2\mu\langle x_{k+1} - x_k, x_{k+1} - y^\star\rangle$, we can show that

$$\begin{aligned}\mathcal{V}_k - \mathcal{V}_{k+1} =& a_k\|Gy_{k-1}\|^2 + \left[\gamma b_{k+1} v_k + \gamma^2 t_k^2 - \gamma^2 t_{k+1}^2 v_k^2 - a_{k+1}\right]\|Gy_k\|^2 \\ &- \gamma b_k\langle Gy_{k-1}, Gy_k\rangle - b_k\langle Gy_{k-1}, x_{k+1} - x_k\rangle \\ &+ \left[b_{k+1}\theta_k + 2\gamma t_k(t_k - 1) - 2\gamma v_k\theta_k t_{k+1}^2\right]\langle Gy_k, x_{k+1} - x_k\rangle \\ &+ \left[\mu + (t_k - 1)^2 - t_{k+1}^2\theta_k^2\right]\|x_{k+1} - x_k\|^2 \\ &+ 2(t_k - t_{k+1}\theta_k - 1 - \mu)\langle x_{k+1} - x_k, x_{k+1} - y^\star\rangle \\ &+ 2\gamma(t_k - t_{k+1} v_k)\langle Gy_k, x_{k+1} - y^\star\rangle \end{aligned} \tag{20}$$

By the $\frac{1}{L}$-co-coerciveness of G and $x_{k+1} = y_k - \gamma G y_k$ from (7), we can derive

$$\langle Gy_k - Gy_{k-1}, x_{k+1} - x_k\rangle \geq \left(\frac{1}{L} - \gamma\right)\|Gy_k - Gy_{k-1}\|^2$$

This inequality implies that $\langle Gy_k, x_{k+1} - x_k\rangle \geq \langle Gy_{k-1}, x_{k+1} - x_k\rangle + \left(\frac{1}{L} - \gamma\right)\|Gy_k - Gy_{k-1}\|^2$. Finally, if we assume that $b_{k+1}\theta_k + 2\gamma t_k(t_k - 1) - 2\gamma v_k\theta_k t_{k+1}^2 \geq 0$, then by substituting the last inequality into (20) and using $a_k\|Gy_{k-1}\|^2 - \gamma b_k\langle Gy_k, Gy_{k-1}\rangle = a_k\left\|Gy_{k-1} - \frac{\gamma b_k}{2a_k} G y_k\right\|^2 - \frac{\gamma^2 b_k^2}{4a_k}\|Gy_k\|^2$, we obtain (14). $\square$

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

Lemma 3. *Let* $0 < \gamma \leq \frac{2}{L}, \mu \geq 0$, *and* $\omega \geq 1$ *be given. Let* $\{(x_k, y_k)\}$ *be generated by (7) and* $\mathcal{V}_k$ *be defined by (12). Assume that* t_k, θ_k, v_k, a_k, *and* b_k *in (7) and (12) are chosen as follows:*

$$\begin{aligned} & t_k := \frac{k + 2\omega + 1}{\omega}, \quad \theta_k := \frac{t_k - 1 - \mu}{t_{k+1}}, \quad v_k := 1 - \frac{1}{t_{k+1}} \\ & b_k := 2\gamma t_k(t_k - 1), \qquad \text{and } a_k := \gamma^2 t_k(t_k - 1) \end{aligned} \tag{21}$$

Then, it holds that

$$\begin{aligned} \mathcal{V}_k - \mathcal{V}_{k+1} \geq & \mu\left(2t_k - 1 - \mu\right)\|x_{k+1} - x_k\|^2 + \frac{\gamma(\omega - 1)}{\omega}\left(\frac{2}{L} - \gamma\right)\|Gy_k\|^2 \\ & + \gamma\left(\frac{2}{L} - \gamma\right) t_k(t_k - 1)\|Gy_k - Gy_{k-1}\|^2 \geq 0 \end{aligned} \tag{22}$$

Moreover, we have $\mathcal{V}_k \geq \frac{\gamma(2-\gamma L)(t_k-1)}{L}\|Gy_{k-1}\|^2 + \mu\|x_k - y^\star\|^2 \geq 0$ *and*

$$\begin{cases} \sum_{k=0}^{\infty} \mu\left(2t_k - 1 - \mu\right)\|x_{k+1} - x_k\|^2 & \leq \mathcal{V}_0 \\ \frac{\gamma(2-L\gamma)(\omega-1)}{L\omega}\sum_{k=0}^{\infty}\|Gy_k\|^2 & \leq \mathcal{V}_0 \\ \frac{\gamma(2-L\psi)}{L}\sum_{k=0}^{\infty} t_k(t_k-1)\|Gy_k - Gy_{k-1}\|^2 & \leq \mathcal{V}_0 \\ \frac{(\omega+1)(2-L\gamma)}{L\gamma(2\omega-1)}\sum_{k=0}^{\infty} t_k^2\left\|x_{k+1} - x_k - \theta_{k-1}\left(x_k - x_{k-1}\right)\right\|^2 & \leq \mathcal{V}_0 \end{cases} \tag{23}$$

Proof. Let us show that t_k, θ_k, v_k, and b_k chosen by (21) satisfy

$$t_k - t_{k+1}\theta_k - 1 - \mu = 0 \text{ and } b_{k+1}\theta_k + 2\gamma t_k(t_k - 1) - 2\gamma v_k\theta_k t_{k+1}^2 - b_k = 0 \tag{24}$$

First, since $\theta_k = \frac{t_k - 1 - \mu}{t_{k+1}}$, the first condition of (24) holds. Next, using $b_k = 2\gamma t_k(t_k - 1)$ and $v_k = 1 - \frac{1}{t_{k+1}}$, we can easily verify the second condition of (24).

Now, using (21), we can directly compute the following coefficients of (14):

$$\begin{cases} t_k^2 - 2t_k + 1 + \mu - t_{k+1}^2\theta_k^2 & = \mu\left(2t_k - 1 - \mu\right) \\ t_k - t_{k+1}v_k & = \frac{\omega-1}{\omega} \\ b_{k+1}\theta_k + 2\gamma t_k(t_k - 1) - 2\gamma v_k\theta_k t_{k+1}^2 & = 2\gamma t_k(t_k - 1) \\ \gamma b_{k+1}v_k + \gamma^2 t_k^2 - \gamma^2 t_{k+1}^2 v_k^2 - a_{k+1} - \frac{\gamma^2 b_k^2}{4a_k} & = \frac{\gamma^2(\omega-1)}{\omega} \end{cases}$$

Substituting (24) and these expressions into (14), and then using $a_k = \gamma^2 t_k(t_k - 1)$ and $2a_k = \gamma b_k$ from (21) into the resulting inequality, we can simplify it as

$$\begin{aligned} \mathcal{V}_k - \mathcal{V}_{k+1} \geq & \frac{\gamma^2(\omega - 1)}{\omega}\|Gy_k\|^2 + \mu\left(2t_k - 1 - \mu\right)\|x_{k+1} - x_k\|^2 \\ & + \gamma\left(\frac{2}{L} - \gamma\right) t_k(t_k - 1)\|Gy_k - Gy_{k-1}\|^2 + \frac{2\gamma(\omega - 1)}{\omega}\langle Gy_k, x_{k+1} - y^\star\rangle \end{aligned}$$

Moreover, since $x_{k+1} = y_k - \gamma Gy_k$, using $Gy^\star = 0$ and the $\frac{1}{L}$-co-coerciveness of G, we have $\langle Gy_k, x_{k+1} - y^\star\rangle = \langle Gy_k, y_k - y^\star\rangle - \gamma\|Gy_k\|^2 \geq \left(\frac{1}{L} - \gamma\right)\|Gy_k\|^2$. Substituting this inequality (24) into the last expression, we get (22).

Next, since $Gy^\star = 0$, using the $\frac{1}{L}$-co-coerciveness of G, we get

$$\begin{aligned}\langle Gy_{k-1}, x_k - y^\star \rangle &= \langle Gy_{k-1} - Gy^\star, y_{k-1} - y^\star - \gamma\left(Gy_{k-1} - Gy^\star\right)\rangle \\ &\geq \left(\frac{1}{L} - \gamma\right)\|Gy_{k-1}\|^2\end{aligned}$$

Utilizing this bound and the Cauchy-Schwarz inequality, we can show that

$$\begin{aligned}b_k\left\langle Gy_{k-1}, x_k - y_k\right\rangle =& \frac{b_k}{t_k}\left\langle Gy_{k-1}, x_k - y^\star\right\rangle - \frac{b_k}{t_k}\left\langle Gy_{k-1}, x_k - y^\star + t_k(y_k - x_k)\right\rangle \\ \geq& -\frac{b_k^2}{4t_k^2}\|Gy_{k-1}\|^2 - \|x_k - y^\star + t_k(y_k - x_k)\|^2 \\ &+ \frac{b_k}{t_k}\left(\frac{1}{L} - \gamma\right)\|Gy_{k-1}\|^2.\end{aligned}$$

Substituting this bound into the definition (12) of $\mathcal{V}_k$ and noticing that $a_k - \frac{b_k^2}{4t_k^2} = \gamma^2(t_k - 1)$, we get $\mathcal{V}_k \geq \left(a_k - \frac{b_k^2}{4t_k^2}\right)\|Gy_{k-1}\|^2 + \frac{b_k}{t_k}\left(\frac{1}{L} - \gamma\right)\|Gy_{k-1}\|^2 + \mu\|x_k - y^\star\|^2 = \frac{\gamma(2-\gamma L)(t_k-1)}{L}\|Gy_{k-1}\|^2 + \mu\|x_k - y^\star\|^2$, which proves that $\mathcal{V}_k \geq 0$.

Summing up (22) from $k := 0$ to $k := K$ and using $\mathcal{V}_{K+1} \geq 0$, we get

$$\begin{aligned}\sum_{k=0}^{K}\Bigg[&\mu\left(2t_k - 1 - \mu\right)\|x_{k+1} - x_k\|^2 + \frac{\gamma(\omega - 1)}{\omega}\left(\frac{2}{L} - \gamma\right)\|Gy_k\|^2 \\ &+\gamma\left(\frac{2}{L} - \gamma\right)t_k(t_k - 1)\|Gy_k - Gy_{k-1}\|^2\Bigg] \leq \mathcal{V}_0 - \mathcal{V}_{K+1} \leq \mathcal{V}_0\end{aligned} \tag{25}$$

Letting $K \to \infty$ in this inequality, we obtain the first three expressions of (23).

Finally, since $x_{k+1} - x_k - \theta_{k-1}\left(x_k - x_{k-1}\right) = \gamma\left(Gy_k - v_{k-1}Gy_{k-1}\right)$, using Young's inequality and then $c_k := \frac{\omega t_k}{(\omega-1)(t_k-1)}$, we can derive that

$$\begin{aligned}&\|x_{k+1} - x_k - \theta_{k-1}\left(x_k - x_{k-1}\right)\|^2 = \gamma^2\left\|Gy_k - v_{k-1}Gy_{k-1}\right\|^2 \\ &\quad\leq \gamma^2\left(1 + c_k\right)\left[v_{k-1}^2\|Gy_k - Gy_{k-1}\|^2 + \frac{\left(v_{k-1} - 1\right)^2}{c_k}\|Gy_k\|^2\right] \\ &\quad\leq \gamma^2\left(1 + c_k\right)\left[\frac{\left(t_k - 1\right)^2}{t_k^2}\|Gy_k - Gy_{k-1}\|^2 + \frac{1}{c_k t_k^2}\|Gy_k\|^2\right] \\ &\quad= \frac{\gamma^2\left(1 + c_k\right)\left(t_k - 1\right)}{t_k^3}\left[(t_k - 1)t_k\|Gy_k - Gy_{k-1}\|^2 + \frac{t_k}{c_k(t_k - 1)}\|Gy_k\|^2\right] \\ &\quad\leq \frac{\gamma^2(2\omega - 1)}{(\omega - 1)t_k^2}\left[t_k(t_k - 1)\|Gy_k - Gy_{k-1}\|^2 + \frac{\omega - 1}{\omega}\|Gy_k\|^2\right]\end{aligned}$$

Combining this inequality and (25) (after dropping its first term), and then letting $K \to \infty$, we obtain the last line of (23). □

The following theorem proves the convergence of Nesterov's accelerated scheme (7), but using a different set of parameters compared to Theorem 1.

Theorem 3. *Assume that G in (1) is $\frac{1}{L}$-co-coercive and $\mathrm{zer}(G) \neq \emptyset$. Let $\{(x_k, y_k)\}$ be generated by (7) to solve (1) using $\gamma_k := \gamma \in \left(0, \frac{1}{L}\right), \theta_k := \frac{k+1}{k+2\omega+2}$, and $v_k :=$ $\frac{k+\omega+2}{k+2\omega+2} \in (0,1)$ for a given constant $\omega > 2$. Then, we have*

$$\begin{cases} \sum_{k=0}^{\infty}(k+\omega+1)\|x_{k+1}-x_k\|^2 < +\infty \ \text{and} \ \|x_{k+1}-x_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+2\omega+1)\|y_k-x_k\|^2 < +\infty \ \text{and} \ \|y_k-x_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega+1)\|Gy_k\|^2 < +\infty \ \text{and} \ \|Gy_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega+1)\|Gx_k\|^2 < +\infty \ \text{and} \ \|Gx_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega)\|y_{k+1}-y_k\|^2 < +\infty \ \text{and} \ \|y_{k+1}-y_k\|^2 = o\left(\frac{1}{k^2}\right) \end{cases} \tag{26}$$

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

Before proving Theorem 3, we make the following remarks.

Remark 4. *First, if we choose $\gamma = \frac{2}{L}$, then we only obtain the first result of (26) and $\|x_{k+1}-x_k\|^2 = o\left(\frac{1}{k^2}\right)$. This rate is theoretically better than the $\mathcal{O}(1/k^2)$ rate in [22] when k is sufficiently large. Second, although we only state the $o(\cdot)$ rates of the four different quantities in Theorem 3, the corresponding $\mathcal{O}(\cdot)$ rates of these quantities can also be achieved through our proof below. Moreover, the upper bound of these rates can be expressed explicitly. For instance, by the first line of (23) and (27) below, we can easily prove a $\mathcal{O}(1/k^2)$-rate of $\|y_k - x_k\|^2$ as*

$$\|y_k - x_k\|^2 \leq \frac{(\omega+1)^2\mathcal{V}_0}{(k+2\omega+1)^2}, \qquad \text{where } \mathcal{V}_0 \text{ is given by } (12)$$

Third, if $\gamma \in \left(0, \frac{1}{L}\right)$, then we can prove $o\left(\frac{1}{k^2}\right)$ convergence rates of $\|Gy_k\|^2, \|Gx_k\|^2$, $\|y_k - x_k\|^2$, and $\|y_{k+1}-y_k\|^2$. Finally, 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 (9) obtained by Theorem 1.

The key step to prove Theorem 3 is Lemma 2. We believe that this lemma is new and its proof is relatively elementary. This proof technique can be further extended to study other methods. For instance, it has been recently exploited to study accelerated randomized coordinate methods for solving (1) in [49]. Moreover, it is worth to emphasize that our results in this paper (see, e.g., Corollary 2 below) show that Halpern's fixed-point methods for solving (1) can achieve both $\mathcal{O}(\cdot)$ and $o(\cdot)$ rates by choosing different parameters. Note that $o(\cdot)$ convergence rates for Nesterov's accelerated methods have been recently studied in a number of works such as [1, 3, 26, 30, 31], but we are not aware of any $o(1/k^2)$ rate for Halpern's fixed-point methods a prior to this work. After the first version of this paper is completed and available online, we find a recent work [9] that also studies $o(\cdot)$ convergence rates of different Nesterov's accelerated schemes for (1). However, [9] relies on the discretizations of dynamical systems as in [1], and thus is different from our approach. Another related work is [56], which shows that the iterate sequences generated by several Halpern-type schemes are actually close to each other and eventually converge to a solution of (1). Though this work did not specifically study $o(\cdot)$, but several summable bounds were also obtained.

Proof of Theorem 3. For simplicity of our analysis, we fix $\mu := 1$ in $\mathcal{V}_k$ defined by (12). The first claim in the first line of (26) comes directly from (23) by noticing that $t_k - 1 = \frac{k+\omega+1}{\omega}$. Now, we prove the second line of (26). Indeed, from (7), we have

$$\begin{aligned} y_{k+1} - x_{k+1} &= v_k\left(x_{k+1} + \gamma G y_k - x_k\right) + \left(\theta_k - v_k\right)\left(x_{k+1} - x_k\right) \\ &= v_k(y_k - x_k) + \left(\theta_k - v_k\right)\left(x_{k+1} - x_k\right) \end{aligned}$$

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

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

Notice from (21) that $t_{k+1}^2 v_k = t_k^2 - \frac{\omega(\omega-2)t_k+\omega-1}{\omega^2}$ and $\frac{t_{k+1}^2(\theta_k - v_k)^2}{1-v_k} = \frac{(\omega+1)^2(k+2\omega+2)}{\omega^3}$. Utilizing these expressions into the last inequality, we obtain

$$\begin{aligned} \frac{\omega(\omega-2)t_k + \omega - 1}{\omega^2}\|y_k - x_k\|^2 \le & t_k^2 \|y_k - x_k\|^2 - t_{k+1}^2 \|y_{k+1} - x_{k+1}\|^2 \\ & + \frac{(\omega+1)^2 t_{k+1}}{\omega^2}\|x_{k+1} - x_k\|^2 \end{aligned} \tag{27}$$

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

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

Using the first line of (26) into this inequality and $\omega > 2$, we obtain $\sum_{k=0}^{\infty}[(\omega-2)(k+2\omega+1) + \omega - 1]\|y_k - x_k\|^2 < +\infty$, which implies the first claim in the second line of (26). Moreover, (27) also shows that $\lim_{k\to\infty} t_k^2 \|x_k - y_k\|^2$ exists (Here, we use [54, Lemma 2.5]). 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 $\|x_k - y_k\|^2 = o(1/k^2)$.

To prove the third line of (26), we note that $\gamma v_k G y_k = (y_{k+1} - x_{k+1}) - \theta_k(x_{k+1} - x_k)$. Hence, we have $\gamma^2 v_k^2 (t_k - 1)\|Gy_k\|^2 \le 2(t_k - 1)\|y_{k+1} - x_{k+1}\|^2 + 2\theta_k^2 (t_k - 1)\|x_{k+1} - x_k\|^2$. Exploiting the last two terms from (26), we obtain $\sum_{k=0}^{\infty}(k + \omega + 1)\|Gy_k\|^2 < +\infty$

To prove the second part in the first line of (26), utilizing both lines of (7), we have $\theta_{k-1}^2\|x_k - x_{k-1}\|^2 = \|y_k - x_k - v_{k-1}(y_{k-1} - x_k)\|^2 = \|x_{k+1} - x_k + \gamma(Gy_k - v_{k-1}Gy_{k-1})\|^2$. Therefore, expanding this expression and using $\langle Gy_k - Gy_{k-1}, x_{k+1} - x_k\rangle \ge \left(\frac{1}{L} - \gamma\right)\|Gy_k - Gy_{k-1}\|^2$ from the proof of Lemma 2,

we can derive that

$$\begin{aligned} \mathcal{T}_{[1]} :=& \theta_{k-1}^2 t_k^2 \|x_k - x_{k-1}\|^2 - \theta_k^2 t_{k+1}^2 \|x_{k+1} - x_k\|^2 \\ =& \gamma^2 t_k^2 \|Gy_k - v_{k-1}Gy_{k-1}\|^2 + 2\gamma t_k^2 \langle Gy_k - Gy_{k-1}, x_{k+1} - x_k\rangle \\ & + 2\gamma t_k^2 (1 - v_k)\langle Gy_{k-1}, x_{k+1} - x_k\rangle + \left(t_k^2 - \theta_k^2 t_{k+1}^2\right)\|x_{k+1} - x_k\|^2 \\ \ge& \gamma^2 t_k^2 \|Gy_k - v_{k-1}Gy_{k-1}\|^2 + 2\gamma t_k^2\left(\frac{1}{L} - \gamma\right)\|Gy_k - Gy_{k-1}\|^2 \\ & + 2\gamma t_k^2 (1 - v_k)\langle Gy_{k-1}, x_{k+1} - x_k\rangle + \left(t_k^2 - \theta_k^2 t_{k+1}^2\right)\|x_{k+1} - x_k\|^2 \end{aligned}$$

Note that by the choice of t_k, θ_k, and v_k as in Theorem 3, we have $2\gamma t_k^2(1 - v_k) \ge 0$ and $t_k^2 - \theta_k^2 t_{k+1}^2 \ge 0$. Employing the update rule (21) and Young's inequality, the last inequality leads to

$$\begin{aligned} \theta_{k-1}^2 t_k^2 \|x_k - x_{k-1}\|^2 - \theta_k^2 t_{k+1}^2 \|x_{k+1} - x_k\|^2 &\ge 2\gamma t_k^2 (1 - v_k)\langle Gy_{k-1}, x_{k+1} - x_k\rangle \\ &\ge -\frac{\gamma t_k^2}{t_{k+1}}\left[\|Gy_{k-1}\|^2 + \|x_{k+1} - x_k\|^2\right] \end{aligned}$$

Following the same argument as in the proof of $\|x_k - y_k\|^2$, we can show that $\lim_{k\to\infty} t_k^2\|x_{k+1} - x_k\|^2 = 0$, and hence, $\|x_{k+1} - x_k\|^2 = o(1/k^2)$, which proves the second part in the first line of (26). Since $\gamma^2\|Gy_k\|^2 = \|x_{k+1} - y_k\|^2 \le 2\|x_{k+1} - x_k\|^2 + 2\|y_k - x_k\|^2$, we also obtain $\|Gy_k\|^2 = o(1/k^2)$.

Since $\|Gx_k\|^2 \le 2\|Gx_k - Gy_k\|^2 + 2\|Gy_k\|^2 \le 2L^2\|x_k - y_k\|^2 + 2\|Gy_k\|^2$, we obtain the fourth line of (26) from the previous lines.

Now, we prove the last line of (26). Since $y_{k+1} - y_k = x_{k+1} - x_k + \theta_k(x_{k+1} - x_k) - \theta_{k-1}(x_k - x_{k-1}) - \gamma(v_k Gy_k - v_{k-1}Gy_{k-1})$, we can bound

$$\begin{aligned}\|y_{k+1} - y_k\|^2 \le{} & 16\|x_{k+1} - x_k\|^2 + 4\|x_k - x_{k-1}\|^2 + 4\gamma^2\|Gy_k - Gy_{k-1}\|^2 \\ & + 4\gamma^2\|Gy_{k-1}\|^2\end{aligned}$$

Here, we have used the facts that $\theta_k, \theta_{k-1}, v_k, v_{k-1} \in (0,1)$ and $(v_k - v_{k-1})^2 < 1$. Combining this inequality and the first and second lines of (26), we obtain the last line of (26).

Finally, to prove the convergence of $\{x_k\}$ and $\{y_k\}$, we note that $\mathcal{V}_k$ is nonnegative and nonincreasing, it converges. Combining this fact and $\lim_{k\to\infty} t_k\|Gy_{k-1}\| = \lim_{k\to\infty} t_k\|x_k - y_k\| = 0$, and noting that $\mu = 1$, we conclude that $\lim_{k\to\infty}\|x_k - y^\star\|$ exists. Moreover, since G is $\frac{1}{L}$-co-coercive, it is L-Lipschitz continuous. Therefore, any limit point $y^\star$ of $\{x_k\}$ is in $\mathrm{zer}(G)$, and thus $\lim_{k\to\infty}\|x_k - y^\star\| = 0$. We conclude that $\{x_k\}$ is convergent to $y^\star$. Since $\|x_k - y_k\| \to 0$, combining this fact and $\lim_{k\to\infty}\|x_k - y^\star\| = 0$, we also have $\lim_{k\to\infty} y_k = y^\star$. □

From the result of Theorem 3, we can derive the convergence of Halpern's fixedpoint iteration (2), but under a different choice of parameters.

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

$$\left\{\begin{array}{lll} \sum_{k=0}^{\infty}(k+\omega+1)\|Gy_k\|^2 < +\infty & \text{and} & \|Gy_k\|^2 = o\left(\frac{1}{k^2}\right) \\ \sum_{k=0}^{\infty}(k+\omega)\|y_{k+1} - y_k\|^2 < +\infty & \text{and} & \|y_{k+1} - y_k\|^2 = o\left(\frac{1}{k^2}\right)\end{array}\right.$$

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

Proof. As proved in Theorem 2, (2) is equivalent to (7) 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 θ_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 $v_k = \frac{\beta_k}{\beta_{k-1}} = \frac{k+\omega+2}{k+2\omega+2}$. These relations lead to $\beta_k = \frac{\omega+1}{k+2\omega+2}$. Moreover, since $\gamma_k = \frac{\eta_k}{1-\beta_k} = \gamma \in \left(0, \frac{2}{L}\right)$, we have $\eta_k = \gamma(1-\beta_k)$. Consequently, (28) follows from (26). □

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 (12), and hence only obtain $\|Gy_k\|^2 = \mathcal{O}(1/k^2)$ 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 [6, Proposition 4.11] that T is a non-expansive mapping if and only if $G := \mathbb{I} - T$ is $\frac{1}{2}$-co-coercive. Therefore, we can obtain new convergence results on the residual norm $\|y_k - Ty_k\|$ from Corollary 2 for a Halpern's fixed-point iteration scheme to approximate a fixed-point $y^\star$ of T.

4 Application to Monotone inclusions

In this section, we present three applications of Theorems 1 and 3 to proximal-point, forward-backward splitting, and three-operator splitting methods.

4.1 Monotone inclusions and solution characterization

We consider the following monotone inclusion and its special cases:

$$0 \in Ay^\star + By^\star + Cy^\star, \tag{29}$$

where $A, B : \mathbb{R}^p \rightrightarrows 2^{\mathbb{R}^p}$ are multivalued and maximally monotone operators, and $C : \mathbb{R}^p \to \mathbb{R}^p$ is a $\frac{1}{L}$-co-coercive operator. Let $Q := A + B + C$ and we assume that $\mathrm{zer}(Q) := Q^{-1}(0) = \{y^\star \in \mathbb{R}^p : 0 \in Ay^\star + By^\star + Cy^\star\}$ is nonempty. We will consider the following cases in this paper.

- **Case 1.** If $A = 0$ and $B = 0$, then by overloading $G = C$, (29) reduces to the co-coercive equation (1) studied in Sect. 3.
- **Case 2.** If $B = 0$ and $C = 0$, then (29) reduces to $0 \in Ay^\star$. Then, we will investigate the convergence of an accelerated proximal-point algorithm and the interplay between Halpern's fixed-point iteration and Nesterov's accelerated interpretations in Sect. 4.2.
- **Case 3.** If $C = 0$, then (29) reduces to $0 \in Ay^\star + By^\star$, also covers monotone VIPs. We will investigate the convergence of an accelerated forward-backward splitting scheme in Sect. 4.3 using our results in Sect. 3.
- **Case 4.** Finally, we will also investigate the convergence of an accelerated threeoperator splitting scheme for solving (29), and its special case: the accelerated Douglas-Rachford splitting scheme in Sect. 4.4.

In order to characterize solutions of (29), we recall the following two operators. The first operator is the forward-backward residual mapping associated with Case 3 of (29) (i.e. $0 \in Ay^\star + By^\star$) :

$$G_{\lambda Q} y := \frac{1}{\lambda}\left(y - J_{\lambda A}(y - \lambda By)\right) \tag{30}$$

where B is single-valued, $Q := A + B$, and $J_{\lambda A}$ is the resolvent of λA for any $\lambda > 0$. The following result is proved similarly to [6, Proposition 26.1] and [50].

Lemma 4. *Let A and B in (29) be maximally monotone, B be single-valued, and $C = 0$. Let $G_{\lambda Q}$ be defined by (30). Then*

$$\langle G_{\lambda Q}x - G_{\lambda Q}y, x - y + \lambda(Bx - By)\rangle \ge \lambda \left\|G_{\lambda Q}x - G_{\lambda Q}y\right\|^2 + \langle Bx - By, x - y\rangle. \tag{31}$$

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

Since $G_{\lambda Q}$ is co-coercive, Lemma 4 shows that solving (29) is equivalent to solving the co-coercive equation $G_{\lambda Q}y^\star = 0$ as a special case of (1).

The second operator is the residual mapping of a three-operator splitting scheme, which is defined as

$$E_{\lambda Q} y := \frac{1}{\lambda}\left(J_{\lambda B}y - J_{\lambda A}\left(2J_{\lambda B}y - y - \lambda C \circ J_{\lambda B}y\right)\right) \tag{32}$$

where $J_{\lambda A}$ and $J_{\lambda B}$ are the resolvents of λA and λB, respectively, and $\circ$ is a composition operator. The following result is similar to the one in, e.g., [6, 15, 21] and we omit its proof here.

Lemma 5. *Let A and B in (29) be maximally monotone, and C be $\frac{1}{L}$-co-coercive. Let $E_{\lambda Q}$ be defined by (32). Then, $E_{\lambda Q}u^\star = 0$ iff $y^\star \in \mathrm{zer}(A+B+C)$, where $y^\star = J_{\lambda B}u^\star$. Moreover, $E_{\lambda Q}$ satisfies the following property for all u and v :*

$$\langle E_{\lambda Q}u - E_{\lambda Q}v, u - v\rangle \geq \frac{\lambda(4-L\lambda)}{4}\|E_{\lambda Q}u - E_{\lambda Q}v\|^2 \tag{33}$$

If B is single-valued and $C = 0$, then we have $E_{\lambda Q}u = G_{\lambda Q}y$, where $y = J_{\lambda B}u$ (or equivalently, $u = y + \lambda By$).

4.2 Application to proximal-point method

We consider **Case 2** where (29) reduces to finding $y^\star \in \mathbb{R}^p$ such that $0 \in Ay^\star$. Let $J_{\lambda A}y := (\mathbb{I}+\lambda A)^{-1}y$ be the resolvent of λA for any $\lambda > 0$ and $G_{\lambda A}y = \frac{1}{\lambda}(\mathbb{I} - J_{\lambda A})\,y = \frac{1}{\lambda}(y - J_{\lambda A}y)$ be the Yosida approximation of A with index $\lambda > 0$. Then, by [6, Corollary 23.11], $G_{\lambda A}$ is λ-co-coercive. Moreover, $y^\star$ solves $0 \in Ay^\star$ if and only if $G_{\lambda A}y^\star = 0$. Hence, solving $0 \in Ay^\star$ is equivalent to solving the λ-co-coercive equation $G_{\lambda A}y^\star = 0$.

In this case, the Halpern-type fixed-point scheme (2) applying to $G_{\lambda A}y^\star = 0$, or equivalently, to solving $0 \in Ay^\star$, can be written as

$$y_{k+1} := \beta_k y_0 + (1-\beta_k)y_k - \eta_k G_{\lambda A}y_k = \beta_k y_0 + \left(1 - \beta_k - \frac{\eta_k}{\lambda}\right)y_k + \frac{\eta_k}{\lambda}J_{\lambda A}y_k \tag{34}$$

where β_k and η_k can be chosen either in Theorem 1 or Corollary 2 to guarantee convergence of (34). If $\beta_k := \frac{1}{k+2}$ and $\eta_k := 2\lambda(1-\beta_k)$ as in Theorem 1, then

$$y_{k+1} := \beta_k y_0 + (1-\beta_k)y_k - 2(1-\beta_k)\,(y_k - J_{\lambda A}y_k) = \beta_k y_0 + (1-\beta_k)R_{\lambda A}y_k$$

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

$$\|G_{\lambda A}y_k\| \leq \frac{\|y_0 - y^\star\|}{\lambda(k+1)}$$

If we choose $\beta_k := \frac{\omega+1}{k+2\omega+2}$ and $\eta_k := \gamma(1-\beta_k)$ as in Corollary 2, then (34) becomes

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

This expression can be viewed as a new variant of Halpern's fixed-point iteration applied to the averaged mapping $\mathcal{T}_{\rho A}y = (1-\rho)y + \rho J_{\lambda A}y$ with $\rho := \frac{\gamma}{\lambda}$ provided that $\gamma \in (0, \lambda]$. In this case, we obtain a convergence result as in (28).

Alternatively, if we apply (7) to solve $G_{\lambda A}y^\star = 0$, then we obtain a Nesterov's accelerated interpretation of (34) as

$$\begin{cases} x_{k+1} := y_k - \gamma_k G_{\lambda A}y_k = (1-\rho_k)\,y_k + \rho_k J_{\lambda A}y_k \quad \text{with} \quad \rho_k := \frac{\gamma_k}{\lambda} \\ y_{k+1} := x_{k+1} + \theta_k(x_{k+1} - x_k) + v_k\,(y_k - x_{k+1}) \end{cases} \tag{35}$$

This method was studied in [30]. Nevertheless, our analysis in Theorem 3 is simpler than that of [30] when it applies to (35). In particular, if we choose $\gamma_k := \lambda$, then the first line of (35) reduces to $x_{k+1} = J_{\lambda A}y_k$. The convergence rate guarantees of (35) can be obtained as results of Corollary 1 and Theorem 3, respectively.

Finally, if we apply (6) to solve $G_{\lambda A}y^\star = 0$ and choose $\eta_k := \lambda\left(\frac{\beta_k}{\beta_{k-1}} + 1 - \beta_k\right)$ and $\gamma_k := \lambda$ such that $\nu_k = 0$ and $\kappa_k = \frac{\beta_k}{\beta_{k-2}}$, then (6) reduces to

$$\begin{cases} x_{k+1} := J_{\lambda A}y_k \\ y_{k+1} := x_{k+1} + \theta_k(x_{k+1} - x_k) + \kappa_k\left(y_{k-1} - x_k\right) \end{cases}$$

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

4.3 Application to forward-backward splitting method

Let us consider Case 3 when (29) reduces to finding $y^\star \in \mathbb{R}^p$ such that $0 \in Ay^\star + By^\star$. By Lemma 4, $y^\star \in \mathrm{zer}(A+B)$ if and only if $G_{\lambda Q}y^\star = 0$, where $Q := A + B$ and $G_{\lambda Q}$ is defined by (30). Moreover, $G_{\lambda Q}$ is $\frac{\lambda(4-\lambda L)}{4}$-co-coercive, provided that $0 < \lambda < \frac{4}{L}$.

If we apply (2) to solve $G_{\lambda Q}y^\star = 0$, then its iterate can be written as

$$y_{k+1} := \beta_k y_0 + (1-\beta_k)\left[(1-\rho)y_k + \rho J_{\lambda A}\left(y_k - \lambda B y_k\right)\right] \tag{36}$$

where we have set $\rho := \frac{4-\lambda L}{2}$. In particular, if we choose $\lambda := \frac{2}{L}$, then $\rho = 1$ and (36) reduces to $y_{k+1} := \beta_k y_0 + (1-\beta_k)J_{\lambda A}\left(y_k - \lambda B y_k\right)$, which can be viewed as Halpern's fixed-point iteration applied to approximate a fixed-point of $J_{\lambda_A}((\cdot) - \lambda B(\cdot))$.

Depending on the choice of β_k and ρ as in Theorem 1 or Corollary 2, we obtain

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

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

Now, we consider Nesterov's accelerated variant of (36) by applying (7) to $G_{\lambda Q}y^\star = 0$ to obtain the following one:

$$\begin{cases} x_{k+1} := (1-\rho_k)\,y_k + \rho_k J_{\lambda A}\left(y_k - \lambda B y_k\right) \\ y_{k+1} := x_{k+1} + \theta_k(x_{k+1} - x_k) + \upsilon_k\left(y_k - x_{k+1}\right) \end{cases} \tag{37}$$

where $\rho_k := \frac{\gamma_k}{\lambda}, \theta_k := \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}$, and $\upsilon_k := \frac{\beta_k}{\beta_{k-1}}$. This scheme is similar to the one studied in [31]. Again, the convergence of (37) can be guaranteed by either Corollary 1 or Theorem 3 depending on the choice of γ_k, θ_k, and υ_k. However, we omit the details here.

4.4 Application to three-operator splitting method

Finally, we consider the general case, **Case 4**. As stated in Lemma 5, $y^\star \in \mathrm{zer}(A+B+C)$ if and only if $E_{\lambda Q}y^\star = 0$, where $Q := A + B + C$ and $E_{\lambda Q}$ is defined by (32). Let us apply (2) to $E_{\lambda Q}y^\star = 0$ and arrive at the following scheme:

$$y_{k+1} := \beta_k y_0 + (1-\beta_k)y_k - \frac{\eta_k}{\lambda}\left(J_{\lambda B}y_k - J_{\lambda A}\left(2J_{\lambda B}y_k - y_k - \lambda C \circ J_{\lambda B}y_k\right)\right)$$

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

$$\begin{cases} z_k & := J_{\lambda B} y_k \\ w_k & := J_{\lambda A}\left(2z_k - y_k - \lambda C z_k\right) \\ y_{k+1} & := \beta_k y_0 + (1-\beta_k) y_k - \frac{\eta_k}{\lambda}\left(z_k - w_k\right) \end{cases} \tag{38}$$

This is called a Halpern-type three-operator splitting scheme for solving (29). If $C = 0$, then it reduces to a Halpern-type Douglas-Rachford splitting scheme for solving **Case 3** of (29) derived from (2). The latter case was proposed in [50] with a direct convergence proof for both dynamic and constant stepsizes, but the convergence is given on $G_{\lambda Q}$ instead of $E_{\lambda Q}$. Note that the convergence results of Theorem 1 and Corollary 2 can be applied to (38) to obtain convergence rates on $\|E_{\lambda Q} y_k\|$. Such rates can be transformed into the ones on $\|G_{\lambda Q} z_k\|$ when $C = 0$ and B is single-valued.

Next, we can also derive Nesterov's accelerated variant of (38) by applying (7) to solve $E_{\lambda Q} y^\star = 0$. In this case, (7) becomes

$$\begin{cases} z_k & := J_{\lambda B} y_k \\ w_k & := J_{\lambda A}\left(2z_k - y_k - \lambda C z_k\right) \\ x_{k+1} & := y_k + \frac{1}{\lambda}\left(w_k - z_k\right) \\ y_{k+1} & := x_{k+1} + \theta_k (x_{k+1} - x_k) + v_k\left(y_k - x_{k+1}\right) \end{cases} \tag{39}$$

Here, the parameters θ_k and v_k can be chosen as in either Corollary 1 or Theorem 3. This scheme essentially has the same per-iteration complexity as the standard threeoperator splitting scheme in the literature, including [15]. However, its convergence rate is much faster than the standard one by applying either Corollary 1 or Theorem 3. If $C = 0$, then (39) reduces to an accelerated Douglas-Rachford splitting scheme, where its fast convergence rate can be obtained as a special case of either Corollary 1 or Theorem 3.

5 Extra-anchored gradient method and its variants

Motivation. While the gradient of a convex and L-smooth function is co-coercive, monotone and Lipschitz continuous operators are not co-coercive in general. As a simple example, one can take $Gx = \big(Av, -A^\top u\big)$ as the gradient of the saddle objective function in a bilinear game, where A in $\mathbb{R}^{m\times n}$ is given and $x = (u, v)$. In order to solve (1) when G is only monotone and L-Lipschitz continuous, the extragradient method (EG) appears to be one of the most suitable candidates [25]. This method has recently been extended to weak Minty VIP, i.e. $\langle Gy, y - y^\star\rangle \geq -\rho\|Gy\|^2$ for all $y \in \mathbb{R}^p$ and $y^\star \in \mathrm{zer}(G)$ in our context, see, e.g. [17]. In [55], Yoon and Ryu applied Halpern's fixed-point iteration to EG and obtained a new algorithm called extra-anchored gradient method (EAG). This algorithm achieves optimal convergence rate on $\|Gy_k\|$. Recently, [27] extended EAG to a co-monotone setting of (1) and still achieved the same rate $\|Gy_k\| = \mathcal{O}(1/k)$ as in [55]. This is perhaps surprising since G is nonmonotone. An extension to Popov's scheme (also called past-extra-gradient, or reflected forward methods) can be found in [50]. Our goal in this section is to derive a corresponding Nesterov's accelerated interpretation of these schemes and possibly provide an alternative convergence rate analysis for the existing results in [27, 50, 55].

5.1 The extra-anchored gradient method and its convergence

The extra-anchored gradient method (EAG) was proposed in [55] to solve (1) under the monotonicity and L-Lipschitz continuity of G, which can be written as

$$\begin{cases} z_{k+1} := \beta_k y_0 + (1-\beta_k)y_k - \eta_k G y_k \\ y_{k+1} := \beta_k y_0 + (1-\beta_k)y_k - \widehat{\eta}_k G z_{k+1} \end{cases} \tag{40}$$

where $\beta_k \in (0,1), y_0$ is an initial point, and η_k and $\widehat{\eta}_k$ are two given step-sizes. Here, we use two different step-sizes η_k and $\widehat{\eta}_k$ compared to the original EAG in [55] by adopting the idea of EG+ from [17], see also [27].

As proven in [55], if we update $\beta_k := \frac{1}{k+2}, \eta_{k+1} := \left(1 - \frac{L^2\eta_k^2}{(1-L^2\eta_k^2)(k+1)(k+3)}\right)\eta_k$, and $\widehat{\eta}_{k+1} := \eta_{k+1}$ with some $0 < \eta_0 < \frac{1}{L}$, then we obtain

$$\|Gy_k\|^2 \le \frac{C_*\|y_0 - y^\star\|^2}{(k+1)(k+2)}, \quad \text{where } \eta_* = \lim_{k\to\infty}\eta_k > 0 \text{ and } C_* := \frac{4\left(1+\eta_0\eta_* L^2\right)}{\eta_*^2} \tag{41}$$

Alternatively, one can also fix the step-size $\widehat{\eta}_k = \eta_k = \eta \in \left(0, \frac{1}{8L}\right]$, and the following convergence guarantee is established:

$$\|Gy_k\|^2 \le \frac{C_*\|y_0 - y^\star\|^2}{(k+1)^2}, \quad \text{where} \quad C_* := \frac{4\left(1+\eta L + \eta^2 L^2\right)}{\eta^2(1+\eta L)} \tag{42}$$

In particular, if $\eta := \frac{1}{8L}$, then $C_* = 260$. Both (41) and (42) were proven in [55].

5.2 Nesterov's accelerated interpretation of EAG

Now, let us derive Nesterov's accelerated interpretation of (40) by proving the following result.

Theorem 4. *Let $\{(x_k, y_k, z_k)\}$ be generated by the following scheme:*

$$\begin{cases} x_{k+1} := y_k - \gamma_k G y_k \\ z_{k+1} := x_{k+1} + \theta_k(x_{k+1} - x_k) + v_k(z_k - x_{k+1}) \\ y_{k+1} := z_{k+1} - \widehat{\eta}_k G z_{k+1} + \eta_k G y_k \end{cases} \tag{43}$$

starting from $z_0 = x_0 := y_0$, where $\gamma_k := \frac{\eta_k}{1-\beta_k}, \theta_k := \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}$, and $\nu_k := \frac{\beta_k}{\beta_{k-1}}$ for $\beta_{-1} > 0, \beta_k, \eta_k$, and $\widehat{\eta}_k$ given in (40). Then, $\{(y_k, z_k)\}$ is identical to the one generated by the EAG scheme (40).

Clearly, (43) is new compared to any Nesterov's accelerated scheme in the literature. To see a relation to existing methods, we rewrite (43) equivalently to

$$\begin{cases} x_{k+1} := z_k - \widehat{\eta}_{k-1}\widehat{G}z_k \\ z_{k+1} := x_{k+1} + \theta_k(x_{k+1} - x_k) + v_k(z_k - x_{k+1}) \end{cases} \tag{44}$$

where $\widehat{G}z_k := Gz_k + \frac{\gamma_k}{\widehat{\eta}_{k-1}}Gy_k - \frac{\eta_{k-1}}{\widehat{\eta}_{k-1}}Gy_{k-1}$ and $y_{k+1} := z_{k+1} - \widehat{\eta}_k G z_{k+1} + \eta_k G y_k$. Clearly, the first two lines of (44) are similar to (7), but using an approximate operator $\widehat{G}z_k$ instead of the exact evaluation Gz_k as in (7). Therefore, (44) can be viewed as an inexact variant of Nesterov's accelerated method (7).

Proof of Theorem 4. We only prove that (40) leads to (43). The opposite direction from (43) to (40) is obtained by reverting back the derivations below.

Firstly, multiplying the first line of (40) by β_{k-1}, we have

$$\beta_{k-1}z_{k+1} = \beta_k\beta_{k-1}y_0 + \beta_{k-1}(1-\beta_k)y_k - \beta_{k-1}\eta_k Gy_k$$

Shifting the index from k to $k-1$ of the first line of (40), and then multiplying the result by $-\beta_k$, we get

$$-\beta_k z_k = -\beta_k\beta_{k-1}y_0 - \beta_k(1-\beta_{k-1})y_{k-1} + \beta_k\eta_{k-1}Gy_{k-1}$$

Summing up both expressions, we arrive at

$$\beta_{k-1}z_{k+1} - \beta_k z_k = \beta_{k-1}(1-\beta_k)y_k - \beta_{k-1}\eta_k Gy_k - \beta_k(1-\beta_{k-1})y_{k-1} + \beta_k\eta_{k-1}Gy_{k-1}$$

This expression leads to

$$z_{k+1} = \frac{\beta_k}{\beta_{k-1}}z_k + (1-\beta_k)y_k - \eta_k Gy_k - \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}y_{k-1} + \frac{\beta_k\eta_{k-1}}{\beta_{k-1}}Gy_{k-1}$$

Next, subtracting the first line from the second one of (40), we have $y_{k+1} - z_{k+1} = -\widehat{\eta}_k Gz_{k+1} + \eta_k Gy_k$, leading to $y_{k+1} = z_{k+1} - \widehat{\eta}_k Gz_{k+1} + \eta_k Gy_k$. Combining this expression and the last line above, we obtain

$$\begin{cases} z_{k+1} = \frac{\beta_k}{\beta_{k-1}}z_k + (1-\beta_k)y_k - \eta_k Gy_k - \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}y_{k-1} + \frac{\beta_k\eta_{k-1}}{\beta_{k-1}}Gy_{k-1} \\ y_{k+1} = z_{k+1} - \widehat{\eta}_k Gz_{k+1} + \eta_k Gy_k \end{cases} \tag{45}$$

Let us introduce $x_{k+1} := y_k - \gamma_k Gy_k$. Then, we have $Gy_k = \frac{1}{\gamma_k}(y_k - x_{k+1})$. Substituting this expression into the first line of (45), we get

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

If we choose γ_k such that $1 - \beta_k - \frac{\eta_k}{\gamma_k} = 0$ (or equivalently, $\gamma_k = \frac{\eta_k}{1-\beta_k}$), then we have

$$z_{k+1} = \frac{\beta_k}{\beta_{k-1}}z_k + \left(1 - \frac{\beta_k}{\beta_{k-1}}\right)x_{k+1} + \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}(x_{k+1} - x_k)$$

Finally, putting the above derivations together, we eventually get (43). □

In order to analyze the convergence of (43), following the same approach as in Sect. 3, we consider the following Lyapunov function:

$$\mathcal{Q}_k := a_k\|Gy_{k-1}\|^2 + b_k\langle Gy_{k-1}, x_k - z_k\rangle + \|x_k + t_k(z_k - x_k) - y^\star\|^2 + \mu\|x_k - y^\star\|^2 \tag{46}$$

where $a_k > 0, b_k > 0, t_k > 0$, and $\mu \geq 0$ are given, determined later.

Recall that the Lyapunov function used in [55] is $\mathcal{L}_k := p_k\|Gy_k\|^2 + q_k \langle Gy_k, y_k - y_0\rangle$, where p_k and q_k are given. Similar to Lemma 1, we can show that if $\mu = 0$, $a_k := \frac{4p_k(p_k+\gamma_k q_k)}{q_k^2}$, and $b_k := \frac{4p_k}{\beta_k q_k}$, then $\mathcal{L}_k = \frac{q_k^2}{4p_k}\left[\mathcal{Q}_{k+1} - \|y_0 - y^\star\|^2\right]$. This relation allows one to adopt the analysis in [55] to prove convergence of (43). However, if $\frac{q_k^2}{4p_k}$ is not a constant, then $\mathcal{Q}_{k+1}$ different from $\mathcal{L}_k$.

The following lemma proves a monotone property of $\mathcal{Q}_k$, which plays a key role to establish convergence of (43).

Lemma 6. *Suppose that G in (1) is monotone and L-Lipschitz continuous. Let $\{(x_k, y_k, z_k)\}$ be generated by (43) and $\mathcal{Q}_k$ be defined by (46) with $\mu := 0$. For $t_k > 1$, and $t_{-1} > 0$, assume that $\gamma_k, \widehat{\eta}_k, \eta_k, \theta_k, v_k, a_k$, and b_k are updated by*

$$\begin{aligned} &\gamma_k := \gamma \in \left(0, \frac{1}{L}\right], \quad \widehat{\eta}_k := \gamma, \quad \eta_k := \frac{\gamma\,(t_{k+1}-1)}{t_{k+1}}, \quad \theta_k := \frac{t_k - 1}{t_{k+1}}, \quad v_k := \frac{t_k}{t_{k+1}} \\ &a_k := \frac{\gamma b_k\,(t_{k-1}+2)}{2t_k}, \quad \text{and} \quad b_{k+1} := \frac{b_k t_{k+1}}{t_k - 1} \end{aligned} \tag{47}$$

Then, for all $k \geq 0$, the following estimates hold:

$$\begin{aligned} \mathcal{Q}_k - \mathcal{Q}_{k+1} &\geq \frac{b_k(1-L^2\gamma^2)t_k}{2L^2\gamma(t_k-1)}\|Gy_k - Gz_k\|^2 + \frac{\gamma b_k(t_{k-1}-t_k+1)}{2t_k}\|Gy_{k-1}\|^2 \\ \mathcal{Q}_{k+1} &\geq \frac{\gamma b_k}{2t_k}\left(t_{k-1} - \frac{b_k}{2\gamma t_k}\right)\|Gy_{k-1}\|^2 \end{aligned} \tag{48}$$

Proof. Similar to the proof of (20), using (43) and (46), we can prove that

$$\begin{aligned} \mathcal{Q}_k - \mathcal{Q}_{k+1} =\,& a_k\|Gy_{k-1}\|^2 - a_{k+1}\|Gy_k\|^2 - b_k\langle Gy_{k-1}, x_{k+1} - x_k\rangle \\ &+ b_{k+1}\theta_k\langle Gy_k, x_{k+1} - x_k\rangle - b_k\langle Gy_{k-1}, z_k - x_{k+1}\rangle \\ &+ b_{k+1}v_k\langle Gy_k, z_k - x_{k+1}\rangle + (t_k^2 - v_k^2 t_{k+1}^2)\,\|z_k - x_{k+1}\|^2 \\ &+ \left[(t_k-1)^2 - t_{k+1}^2\theta_k^2 + \mu\right]\|x_{k+1} - x_k\|^2 \\ &+ 2\,(t_k - 1 - t_{k+1}\theta_k - \mu)\,\langle x_{k+1} - x_k, x_{k+1} - y^\star\rangle \\ &+ 2\,(t_k - t_{k+1}v_k)\,\langle z_k - x_{k+1}, x_{k+1} - y^\star\rangle \\ &+ 2\left[t_k(t_k-1) - v_k\theta_k t_{k+1}^2\right]\langle z_k - x_{k+1}, x_{k+1} - x_k\rangle \end{aligned} \tag{49}$$

Now, let us choose the parameters t_k, b_k, θ_k, and v_k such that

$$\begin{aligned} &t_k - t_{k+1}v_k = 0, t_k(t_k-1) - v_k\theta_k t_{k+1}^2 = 0 \\ &t_k - 1 - t_{k+1}\theta_k - \mu = 0, \text{ and } b_k = b_{k+1}\theta_k \end{aligned} \tag{50}$$

The third condition leads to $\theta_k := \frac{t_k-1-\mu}{t_{k+1}}$, while the first one gives us $v_k := \frac{t_k}{t_{k+1}}$. The second condition becomes $t_k(t_k-1) = t_k(t_k-1-\mu)$, which is satisfied if $\mu = 0$. The last condition holds if $b_{k+1} := \frac{b_k}{\theta_k}$. These updates are exactly (47).

Noticing from (43) that $z_k - x_{k+1} = \gamma_k Gy_k + \widehat{\eta}_{k-1}Gz_k - \eta_{k-1}Gy_{k-1}$. Using this relation and our choice $b_{k+1} = \frac{b_k}{\theta_k}$, we can easily show that

$$\begin{aligned} \mathcal{T}_{[2]} :=\,& b_{k+1}v_k\langle Gy_k, z_k - x_{k+1}\rangle - b_k\langle Gy_{k-1}, z_k - x_{k+1}\rangle \\ =\,& \frac{b_k\gamma_k v_k}{\theta_k}\|Gy_k\|^2 - b_k\left(\gamma_k + \frac{\eta_{k-1}v_k}{\theta_k}\right)\langle Gy_k, Gy_{k-1}\rangle + \frac{b_k\widehat{\eta}_{k-1}v_k}{\theta_k}\langle Gy_k, Gz_k\rangle \\ &- b_k\widehat{\eta}_{k-1}\,\langle Gy_{k-1}, Gz_k\rangle + b_k\eta_{k-1}\|Gy_{k-1}\|^2. \end{aligned}$$

Utilizing the monotonicity of G, $b_k = b_{k+1}\theta_k$, and $x_{k+1} = y_k - \gamma_k G y_k$, we have

$$\begin{aligned}\mathcal{T}_{[3]} &:= b_{k+1}\theta_k\langle Gy_k, x_{k+1} - x_k\rangle - b_k\langle Gy_{k-1}, x_{k+1} - x_k\rangle \\ &\geq -b_k\gamma_k\|Gy_k\|^2 + b_k(\gamma_k + \gamma_{k-1})\langle Gy_k, Gy_{k-1}\rangle - b_k\gamma_{k-1}\|Gy_{k-1}\|^2\end{aligned}$$

Substituting (50), $\mathcal{T}_{[2]}$, and $\mathcal{T}_{[3]}$ into (49) and using $\frac{v_k}{\theta_k} = \frac{t_k}{t_k-1-\mu} = \frac{t_k}{t_k-1}$, we can further lower bound

$$\begin{aligned}\mathcal{Q}_k - \mathcal{Q}_{k+1} \geq & [a_k + b_k(\eta_{k-1} - \gamma_{k-1})]\|Gy_{k-1}\|^2 + \left(\frac{b_k\gamma_k}{t_k - 1} - a_{k+1}\right)\|Gy_k\|^2 \\ & - b_k\left(\frac{\eta_{k-1}t_k}{t_k - 1} - \gamma_{k-1}\right)\langle Gy_k, Gy_{k-1}\rangle \\ & + \frac{b_k\widehat{\eta}_{k-1}}{t_k - 1}\langle Gy_k, Gz_k\rangle + b_k\widehat{\eta}_{k-1}\langle Gy_k - Gy_{k-1}, Gz_k\rangle \end{aligned} \tag{51}$$

Now, using the L-Lipschitz continuity of G, we have $\|Gz_k - Gy_k\|^2 \leq L^2\|z_k - y_k\|^2 = L^2\|\widehat{\eta}_{k-1}Gz_k - \eta_{k-1}Gy_{k-1}\|^2$, which leads to

$$\begin{aligned}&\|Gy_k\|^2 + (1 - L^2\widehat{\eta}_{k-1}^2)\|Gz_k\|^2 - 2(1 - L^2\eta_{k-1}\widehat{\eta}_{k-1})\langle Gy_k, Gz_k\rangle \\ &\quad - 2L^2\eta_{k-1}\widehat{\eta}_{k-1}\langle Gz_k, Gy_k - Gy_{k-1}\rangle - L^2\eta_{k-1}^2\|Gy_{k-1}\|^2 \leq 0\end{aligned}$$

Multiplying this inequality by $\frac{b_k}{2L^2\eta_{k-1}}$ and adding the result to (51), we get

$$\begin{aligned}\mathcal{Q}_k - \mathcal{Q}_{k+1} \geq & \frac{b_k(1 - L^2\widehat{\eta}_{k-1}^2)}{2L^2\eta_{k-1}}\|Gz_k\|^2 + \left(\frac{b_k\gamma_k}{t_k - 1} + \frac{b_k}{2L^2\eta_{k-1}} - a_{k+1}\right)\|Gy_k\|^2 \\ & - b_k\left(\frac{1}{L^2\eta_{k-1}} - \frac{\widehat{\eta}_{k-1}t_k}{t_k - 1}\right)\langle Gy_k, Gz_k\rangle \\ & + \left[a_k + \frac{b_k}{2}(\eta_{k-1} - 2\gamma_{k-1})\right]\|Gy_{k-1}\|^2 \\ & - b_k\left(\frac{\eta_{k-1}t_k}{t_k - 1} - \gamma_{k-1}\right)\langle Gy_k, Gy_{k-1}\rangle \end{aligned} \tag{52}$$

Let us choose γ_{k-1}, $\widehat{\eta}_{k-1}$, and η_{k-1} as in (47), i.e.:

$$\gamma_{k-1} = \widehat{\eta}_{k-1} := \gamma \in \left(0, \frac{1}{L}\right], \quad \text{and} \quad \eta_{k-1} := \frac{\gamma(t_k - 1)}{t_k} \tag{53}$$

From (50), we have $b_k = b_{k+1}\theta_k = \frac{b_{k+1}(t_k-1)}{t_{k+1}}$, leading to $b_{k+1} = \frac{b_k t_{k+1}}{t_k - 1}$ as in (47). If we choose $a_{k+1} := \frac{\gamma b_k(t_k+2)}{2(t_k-1)}$, then since $b_{k-1} = \frac{b_k(t_{k-1}-1)}{t_k}$, we get $a_k = \frac{\gamma b_k(t_{k-1}+2)}{2t_k}$ as given in (47). Next, utilizing (47), we can show that

$$\begin{cases}\frac{b_k(1-L^2\widehat{\eta}_{k-1}^2)}{2L^2\eta_{k-1}} & = \frac{b_k(1-L^2\gamma^2)t_k}{2L^2\gamma(t_k-1)} \\ b_k\left(\frac{1}{L^2\eta_{k-1}} - \frac{\widehat{\eta}_{k-1}t_k}{t_k-1}\right) & = \frac{b_k(1-L^2\gamma^2)t_k}{L^2\gamma(t_k-1)} \\ \frac{b_k-1}{t_k-1} + \frac{b_k}{2L^2\eta_{k-1}} - a_{k+1} & = \frac{b_k(1-L^2\gamma^2)t_k}{2L^2\gamma_{(t_k}-1)} \\ a_k + \frac{b_k}{2}(\eta_{k-1} - 2\gamma_{k-1}) & = \frac{\gamma b_k(t_{k-1} - t_k + 1)}{2t_k} \\ b_k\left(\frac{\eta_{k-1}t_k}{t_k-1} - \gamma_{k-1}\right) & = 0\end{cases}$$

Using these expressions, we can simplify (52) as

$$\mathcal{Q}_k - \mathcal{Q}_{k+1} \geq \frac{b_k(1-L^2\gamma^2)t_k}{2L^2\gamma(t_k-1)}\|Gy_k - Gz_k\|^2 + \frac{\gamma b_k(t_{k-1}-t_k+1)}{2t_k}\|Gy_{k-1}\|^2$$

which proves the first estimate of (48).

Finally, using the definition (46) of $\mathcal{Q}_k, x_k = y_{k-1} - \gamma_{k-1}Gy_{k-1}$, the monotonicity of G, and (47), with a similar argument as in Lemma 3, we can show that

$$\mathcal{Q}_k \geq \left(a_k - \frac{b_k^2}{4t_k^2} - \frac{b_k\gamma_{k-1}}{t_k}\right)\|Gy_{k-1}\|^2 = \frac{\gamma b_k}{2t_k}\left(t_{k-1} - \frac{b_k}{2\gamma t_k}\right)\|Gy_{k-1}\|^2$$

which proves the second estimate of (48). $\square$

Now, we can state the main convergence result of (43) in the following theorem.

Theorem 5. *Suppose that G in (1) is monotone and L-Lipschitz continuous. Let $\{(x_k, y_k, z_k)\}$ be generated by (43) using $t_k := k + \omega$ and (47) for a given $\omega > 1$. Then, for all $k \geq 0$, we have*

$$\begin{aligned} \|Gy_k\| &\leq \frac{2\sqrt{2\omega^2-1}}{\gamma(k+\omega)}\|y_0 - y^\star\| \\ \sum_{l=0}^{k}(l+\omega)^2\|Gy_l - Gz_l\|^2 &\leq \frac{2L^2\left(2\omega^2-1\right)}{1-L^2\gamma^2}\|y_0 - y^\star\|^2 \end{aligned} \tag{54}$$

Proof. Since $t_k := k + \omega$ for any $\omega > 1$, we have $b_{k+1} = \frac{b_k t_{k+1}}{t_k - 1} = \frac{b_k(k+\omega+1)}{k+\omega-1}$. By induction, we get $b_k = \frac{b_1(k+\omega)(k+\omega-1)}{\omega(\omega+1)}$. Using this expression and choosing $b_1 := \gamma\omega(\omega+1)$, we obtain from the second line of (48) the following bound:

$$\mathcal{Q}_{k+1} \geq \frac{\gamma b_1(k+\omega)^2}{2\omega(\omega+1)}\left(1 - \frac{b_1}{2\gamma\omega(\omega+1)}\right)\|Gy_k\|^2 = \frac{\gamma^2(k+\omega)^2}{4}\|Gy_k\|^2 \tag{55}$$

Next, since $t_k := k + \omega$, we also have $t_{k-1} - t_k + 1 = 0$. Moreover, since $0 < \gamma \leq \frac{1}{L}$ and $b_k = \gamma(k+\omega)(k+\omega-1)$, the first line of (48) leads to

$$\mathcal{Q}_k - \mathcal{Q}_{k+1} \geq \frac{(1-L^2\gamma^2)(k+\omega)^2}{2L^2}\|Gy_k - Gz_k\|^2 \geq 0 \tag{56}$$

However, since $x_0 = z_0 := y_0$ and $\gamma Gy_{-1} = x_0 - y_{-1}$ from the first line of (43), we have $\mathcal{Q}_0 = \frac{a_0}{\gamma^2}\|x_0 - y_{-1}\|^2 + \|x_0 - y^\star\|^2$. If we choose $y_{-1} := J_{\gamma G}(x_0)$, then we get $\mathcal{Q}_0 \leq \left(2\omega^2 - 1\right)\|y_0 - y^\star\|^2$. From (56), by induction, we obtain $\mathcal{Q}_{k+1} \leq \mathcal{Q}_0 \leq \left(2\omega^2-1\right)\|y_0 - y^\star\|^2$. Combining this inequality, and (55), we get the first line of (54). Finally, summing up (56) from $l := 0$ to $l := k$, we can deduce the second line of (54). $\square$

5.3 EAG for co-monotone case and its Nesterov's acceleration

We consider the variant of EAG in [27] for the co-monotone operator G. Recall that the operator G in (1) is said to be ρ-comonotone if $\langle Gx - Gy, x - y\rangle \geq \rho\|Gx - Gy\|^2$ for all $x, y \in \mathbb{R}^p$, where $\rho < 0$. In this subsection, we consider the case that G is also L-Lipschitz continuous and ρ satisfies the condition $-\frac{1}{2L} < \rho \leq \frac{1}{L}$. Hence, it covers three cases: co-coerciveness when $\rho > 0$, monotonicity when $\rho = 0$, and co-monotonicity when $\rho < 0$. The second case has been studied in Sect. 5.2.

More specifically, [27] proposes a variant of (40) to solve (1) as follows:

$$\begin{cases} z_{k+1} := \beta_k y_0 + (1-\beta_k) y_k - (1-\beta_k)(2\rho + \eta_k) G y_k \\ y_{k+1} := \beta_k y_0 + (1-\beta_k) y_k - 2\rho(1-\beta_k) G y_k - \eta_k G z_{k+1} \end{cases} \tag{57}$$

where $\beta_k := \frac{1}{k+1}, \eta_k := \frac{1}{L}$. If $\rho := 0$, then (57) reduces to (43).

As proved in [27], the following convergence guarantee is obtained for (57):

$$\|Gy_k\|^2 \leq \frac{4L^2\|y_0 - y^\star\|^2}{(1+2\rho L)k^2}, \quad \forall k \geq 1 \tag{58}$$

Here, the key condition is $\rho > -\frac{1}{2L}$, which allows one to handle a class of nonmonotone operators G.

Let us rewrite (57) in a different form. First, we have $y_{k+1} - z_{k+1} = -\eta_k G z_{k+1} + \eta_k(1-\beta_k) G y_k$. Hence, we get $2\rho(1-\beta_k) G y_k = \frac{2\rho}{\eta_k}(y_{k+1} - z_{k+1}) + 2\rho G z_{k+1}$. Then, we have $y_{k+1} = \beta_k y_0 + (1-\beta_k) y_k - \eta_k G z_{k+1} - \frac{2\rho}{\eta_k}(y_{k+1} - z_{k+1}) - 2\rho G z_{k+1}$. This expression implies that $(\eta_k + 2\rho) y_{k+1} = 2\rho z_{k+1} + \eta_k [\beta_k y_0 + (1-\beta_k) y_k] - \eta_k(\eta_k + 2\rho) G z_{k+1}$. Therefore, we eventually arrive at

$$y_{k+1} = (1-\tau_k) z_{k+1} + \tau_k \left[\beta_k y_0 + (1-\beta_k) y_k - \frac{\eta_k}{\tau_k} G z_{k+1}\right] \quad \text{where} \quad \tau_k := \frac{\eta_k}{\eta_k + 2\rho} > 0$$

Overall, the scheme (57) can be rewritten as

$$\begin{cases} z_{k+1} := \beta_k y_0 + (1-\beta_k) y_k - \frac{(1-\beta_k)\eta_k}{\tau_k} G y_k \\ w_{k+1} := \beta_k y_0 + (1-\beta_k) y_k - \frac{\eta_k}{\tau_k} G z_{k+1} \\ y_{k+1} := (1-\tau_k) z_{k+1} + \tau_k w_{k+1} \end{cases} \tag{59}$$

The last line is a linear combination of the first two lines of the accelerated variant of EG+ from [17]. If $\tau_k = 1$ (i.e. G is monotone), then (59) reduces to (40).

Now, we derive Nesterov's accelerated interpretation of (57). Following the same derivation as of (43), we can show that (57) is equivalent to

$$\begin{cases} x_{k+1} := y_k - (\eta_k + 2\rho) G y_k \\ z_{k+1} := x_{k+1} + \theta_k(x_{k+1} - x_k) + \upsilon_k(z_k - x_{k+1}) \\ y_{k+1} := z_{k+1} - \eta_k (G z_{k+1} - (1-\beta_k) G y_k) \end{cases} \tag{60}$$

This Nesterov's accelerated interpretation reduces to (43) when $\rho = 0$. Moreover, it still achieves $\mathcal{O}(1/k^2)$ rate even when G is co-monotone (i.e. $-\frac{1}{2L} < \rho < 0$), which is non-monotone. The analysis in Theorem 5 can be applied to (60), but we omit it here.

5.4 Nesterov's interpretation of the Halpern-type PEG method

Our final step is to consider the Halpern-type past-extragradient (PEG) scheme for solving (1) studied in [50], which can be written as

$$\begin{cases} z_{k+1} := \beta_k y_0 + (1-\beta_k) y_k - \eta_k G z_k \\ y_{k+1} := \beta_k y_0 + (1-\beta_k) y_k - \widehat{\eta}_k G z_{k+1} \end{cases} \tag{61}$$

Here, $z_0 := y_0, \beta_k \in (0,1)$, and $\eta_k, \widehat{\eta}_k > 0$ are given parameters, which will be determined in Theorem 6. The convergence of (61) has been proven in [50]. However, we provide a slightly different variant in Theorem 6 with $\eta_k \neq \widehat{\eta}_k$ and with a simple parameter update for η_k and $\widehat{\eta}_k$ compared to [50]. We also provide a range of $\widehat{\eta}_k$ instead of fixing it at $\widehat{\eta}_k = \frac{1}{2L}$.

Theorem 6. *Assume that G in (1) is monotone and L-Lipschitz continuous and $\mathrm{zer}(G) \neq \emptyset$. Let $\{(y_k, z_k)\}$ be generated by (61) to solve (1) using $\beta_k := \frac{1}{k+\omega}$, $\widehat{\eta}_k := \widehat{\eta} \in \left(0, \frac{1}{2L}\right]$, and $\eta_k := \widehat{\eta}(1-\beta_k)$, where $\omega > 1$ is given. Then, we have*

$$\begin{cases} \|Gy_k\|^2 + 2L^2\|z_k - y_k\|^2 & \leq \frac{C_0\|y_0 - y^\star\|^2}{(k+\omega-1)^2} \\ \|Gz_k\|^2 & \leq \frac{3C_0\|y_0-y^\star\|^2}{2(k+\omega-1)^2} \\ \psi \cdot \sum_{k=1}^{\infty}(k+\omega-1)^2\left[\|Gy_k - Gy_{k-1}\|^2 + L^2\|z_k - y_k\|^2\right] & \leq C_0\|y_0 - y^\star\|^2 \end{cases} \tag{62}$$

where $C_0 := \frac{2\omega\widehat{\eta}^2L^2\left[1+(\omega-1)^2\right]+4(\omega-1)}{\widehat{\eta}^2(\omega-1)} > 0$ and $\psi := \frac{1-4L^2\widehat{\eta}^2}{8L^2\widehat{\eta}} \geq 0$.

Proof. We consider a Lyapunov function $\mathcal{L}_k := p_k\|Gy_k\|^2 + q_k\langle Gy_k, y_k - y_0\rangle$. With a similar proof as in [50], by choosing $q_{k+1} = \frac{q_k}{1-\beta_k}$ and using (61), we have

$$\begin{aligned} \mathcal{L}_k - \mathcal{L}_{k+1} \geq & \left(p_k - \frac{q_k\eta_k}{2\beta_k}\right)\|Gy_k\|^2 + \left(\frac{q_k}{8L^2\eta_k\beta_k} - p_{k+1}\right)\|Gy_{k+1}\|^2 \\ & + \frac{q_k\left(1-4L^2\widehat{\eta}_k^2\right)}{8L^2\eta_k\beta_k}\|Gz_{k+1}\|^2 + \frac{q_k}{\beta_k}\left(\frac{\widehat{\eta}_k}{1-\beta_k} - \frac{1}{4L^2\eta_k}\right)\langle Gy_{k+1}, Gz_{k+1}\rangle \\ & + \frac{q_k}{8\eta_k\beta_k}\|z_{k+1} - y_{k+1}\|^2 - \frac{L^2q_k\eta_k}{2\beta_k}\|z_k - y_k\|^2 \end{aligned} \tag{63}$$

Let us choose $\beta_k := \frac{1}{k+\omega}$ for some $\omega > 1, \widehat{\eta}_k := \widehat{\eta} \in \left(0, \frac{1}{2L}\right], \eta_k := \widehat{\eta}_k(1-\beta_k) = \widehat{\eta}(1-\beta_k)$, and $p_k := \frac{\widehat{\eta}q_k}{2\beta_{k-1}}$. Then, we have $q_{k+1} = \frac{q_k}{1-\beta_k} = \frac{q_k(k+\omega)}{k+\omega-1} = \frac{q_0(k+\omega)}{\omega}$. Next, using these choices of parameters and $q_{k+1} = \frac{q_k}{1-\beta_k}$, (63) reduces to

$$\mathcal{E}_k \geq \mathcal{E}_{k+1} + \frac{q_{k+1}\left(1-4L^2\widehat{\eta}^2\right)(k+\omega)}{8L^2\widehat{\eta}}\left[\|Gy_{k+1} - Gy_k\|^2 + L^2\|z_{k+1} - y_{k+1}\|^2\right] \tag{64}$$

where $\mathcal{E}_k := \mathcal{L}_k + \frac{\widehat{\eta}L^2q_k(k+\omega-1)}{2}\|z_k - y_k\|^2$.

Now, since $\mathcal{L}_k \geq \frac{p_k}{2}\|Gy_k\|^2 - \frac{q_k^2}{2p_k}\|y_0 - y^\star\|^2$ (see [55]), we have

$$\begin{aligned} \mathcal{E}_k & \geq \frac{\widehat{\eta}q_k(k+\omega-1)}{4}\|Gy_k\|^2 + \frac{\widehat{\eta}L^2q_k(k+\omega-1)}{2}\|z_k - y_k\|^2 - \frac{q_k}{\widehat{\eta}(k+\omega-1)}\|y_0 - y^\star\|^2 \\ & = \frac{\widehat{\eta}q_0(k+\omega-1)^2}{4\omega}\left[\|Gy_k\|^2 + 2L^2\|y_k - z_k\|^2\right] - \frac{q_0}{\widehat{\eta}\omega}\|y_0 - y^\star\|^2 \end{aligned} \tag{65}$$

If $2L\widehat{\eta} \leq 1$, then from (64), we have $\mathcal{E}_{k+1} \leq \mathcal{E}_k$. By induction, it leads to $\mathcal{E}_k \leq \mathcal{E}_0$. Furthermore, by the Lipschitz continuity of G and $Gy^\star = 0$, we have $\|Gy_0\| \leq L\|y_0 - y^\star\|$. Using this estimate, we can easily show that

$$\mathcal{E}_0 = \frac{\widehat{\eta}q_0}{2(\omega-1)}\|Gy_0\|^2 + \frac{\widehat{\eta}L^2q_0(\omega-1)}{2}\|y_0 - y^\star\|^2 \leq \frac{\widehat{\eta}L^2q_0\left[1+(\omega-1)^2\right]}{2(\omega-1)}\|y_0 - y^\star\|^2$$

Hence, combining this bound and $\mathcal{E}_k \le \mathcal{E}_0$, we get $\mathcal{E}_k \le \frac{\widehat{\eta} L^2 q_0 \left[1+(\omega-1)^2\right]}{2(\omega-1)} \|y_0 - y^\star\|^2$. Utilizing (65), the last inequality leads to the first line of (62).

By Young's inequality and the Lipschitz continuity of G, we have $\|Gz_k\|^2 \le \frac{3}{2}\|Gy_k\|^2 + 3\|Gz_k - Gy_k\|^2 \le \frac{3}{2}\left[\|Gy_k\|^2 + 2L^2\|z_k - y_k\|^2\right]$. Combining this inequality and the first line of (62), we get the second line of (62).

Finally, summing up (65) from $k := 0$ to $k := K$, and then using (65) and the upper bound of $\mathcal{E}_0$, we have

$$
\begin{aligned}
\frac{q_0\left(1 - 4L^2\widehat{\eta}^2\right)}{8L^2\widehat{\eta}} \sum_{k=0}^{K} (k+\omega)^2 \left[\|Gy_{k+1} - Gy_k\|^2 + L^2\|z_{k+1} - y_{k+1}\|^2\right] &\le \mathcal{E}_0 - \mathcal{E}_{K+1} \\
\le \frac{\widehat{\eta} L^2 q_0\left[1 + (\omega-1)^2\right]}{2(\omega-1)} \|y_0 - y^\star\|^2 + \frac{q_0}{\widehat{\eta}\omega}\|y_0 - y^\star\|^2
\end{aligned}
$$

Simplifying this inequality and letting $K \to \infty$, we obtain the third line of (62). $\square$

Note that if we choose $\omega := 2$ in Theorem 6, then $\beta_k = \frac{1}{k+2}$ and $C_0 = \frac{4\left(2L^2\widehat{\eta}^2+1\right)}{\widehat{\eta}^2}$. This constant factor is larger than the one in (54) of Theorem 5. However, (61) only requires one evaluation of G per iteration compared to two evaluations as in (40). If $0 < \widehat{\eta} < \frac{1}{2L}$, then the last summable bound in the third line of (62) is not vanished.

By following the same arguments as (43), we can derive Nesterov's accelerated interpretation of (61). This result is stated in the following theorem.

Theorem 7. *Let $\beta_k \in (0,1)$ and η_k, $\widehat{\eta}_k > 0$ be given as in (61). Let $\{(\widehat{x}_k, z_k)\}$ be generated by the following scheme:*

$$
\begin{cases}
\widehat{x}_{k+1} := z_k - \widehat{\gamma}_k G z_k \\
z_{k+1} := \widehat{x}_{k+1} + \theta_k(\widehat{x}_{k+1} - \widehat{x}_k) + \nu_k(z_k - \widehat{x}_{k+1}) + \kappa_k(z_{k-1} - \widehat{x}_k) \\
-\zeta_k(z_{k-2} - \widehat{x}_{k-1})
\end{cases}
\tag{66}
$$

starting from $z_{-2} = z_{-1} = z_0 = \widehat{x}_{-1} = \widehat{x}_0 := y_0$, and using $\theta_k := \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}$,, $\nu_k := \frac{\beta_k}{\beta_{k-1}}$, $\widehat{\gamma}_k := \widehat{\eta}_{k-1} + \frac{\eta_k}{1-\beta_k}$, $\kappa_k := \frac{\eta_{k-1}(1-\beta_k)}{\widehat{\gamma}_{k-1}}$ and $\zeta_k := \frac{\theta_k\eta_{k-2}}{\widehat{\gamma}_{k-2}}$ provided that $\widehat{\eta}_{-2} = \widehat{\eta}_{-1} = \widehat{\eta}_0 = \eta_{-2} = \eta_{-1} = \eta_0$. Let $y_k := z_k - \widehat{\eta}_{k-1}Gz_k + \eta_{k-1}Gz_{k-1}$. Then, $\{(y_k, z_k)\}$ is identical to the one generated by (61) starting from y_0.

Proof. Similar to (43), the following scheme is equivalent to (61):

$$
\begin{cases}
x_{k+1} := y_k - \gamma_k G z_k \\
z_{k+1} := x_{k+1} + \theta_k(x_{k+1} - x_k) + \nu_k(z_k - x_{k+1}) \\
y_{k+1} := z_{k+1} - \widehat{\eta}_k G z_{k+1} + \eta_k G z_k
\end{cases}
\tag{67}
$$

where $\gamma_k := \frac{\eta_k}{1-\beta_k}$, $\theta_k := \frac{\beta_k(1-\beta_{k-1})}{\beta_{k-1}}$, and $\nu_k := \frac{\beta_k}{\beta_{k-1}}$.

Next, from the third line of (67), we have $y_k = z_k - \widehat{\eta}_{k-1}Gz_k + \eta_{k-1}Gz_{k-1}$. Hence, we can eliminate y_k in (67) to get $x_{k+1} = z_k - (\widehat{\eta}_{k-1} + \gamma_k)Gz_k + \eta_{k-1}Gz_{k-1}$. In this case, (67) can be written equivalently to

$$
\begin{cases}
x_{k+1} := z_k - (\widehat{\eta}_{k-1} + \gamma_k) G z_k + \eta_{k-1} G z_{k-1} \\
z_{k+1} := x_{k+1} + \theta_k(x_{k+1} - x_k) + \nu_k(z_k - x_{k+1})
\end{cases}
\tag{68}
$$

This scheme can be viewed as Nesterov's accelerated interpretation of (61). However, if we denote $\widehat{x}_{k+1} := z_k - (\widehat{\eta}_{k-1} + \gamma_k)Gz_k$, then $x_{k+1} = \widehat{x}_{k+1} + \eta_{k-1}Gz_{k-1}$. Substituting this expression into the second line of (68), we get

$$z_{k+1} = \widehat{x}_{k+1} + \theta_k(\widehat{x}_{k+1} - \widehat{x}_k) + v_k(z_k - \widehat{x}_{k+1}) + \eta_{k-1}(1 + \theta_k - v_k)Gz_{k-1} - \theta_k\eta_{k-2}Gz_{k-2}$$

Since $Gz_{k-1} = \frac{1}{\widehat{\eta}_{k-2}+\gamma_{k-1}}(z_{k-1} - \widehat{x}_k)$, we can further write (68) as

$$\begin{cases} \widehat{x}_{k+1} := z_k - \widehat{y}_k Gz_k \\ z_{k+1} := \widehat{x}_{k+1} + \theta_k(\widehat{x}_{k+1} - \widehat{x}_k) + v_k(z_k - \widehat{x}_{k+1}) \\ \qquad\quad + \kappa_k(z_{k-1} - \widehat{x}_k) - \zeta_k(z_{k-2} - \widehat{x}_{k-1}) \end{cases}$$

where $\widehat{\gamma}_k := \widehat{\eta}_{k-1} + \gamma_k$, $\kappa_k := \frac{\eta_{k-1}(1+\theta_k-\nu_k)}{\widehat{\eta}_{k-2}+\gamma_{k-1}} = \frac{\eta_{k-1}(1-\beta_k)}{\widehat{\gamma}_{k-1}}$ and $\zeta_k := \frac{\theta_k\eta_{k-2}}{\widehat{\eta}_{k-3}+\gamma_{k-2}} = \frac{\theta_k\eta_{k-2}}{\widehat{\gamma}_{k-2}}$. The last scheme is exactly (66). □

Clearly, the new Nesterov's accelerated scheme (66) for solving (1) has three correction terms instead of two as in (6). Since our transformation is equivalent, the convergence of (66) is still guaranteed by Theorem 6. More specifically, if we choose $\beta_k, \widehat{\eta}_k$, and η_k as in Theorem 6, then we obtain the following corollary.

Corollary 3. *Assume that G in (1) is monotone and L-Lipschitz continuous and* $\mathrm{zer}(G) \neq \emptyset$. *Let $\{(\widehat{x}_k, z_k)\}$ be generated by (66) to solve (1) using $\widehat{\gamma}_k := \widehat{\gamma} \in \left(0, \frac{1}{L}\right]$ for all $k \geq 0$, and*

$$\theta_k := \frac{k+\omega-2}{k+\omega}, \quad v_k := \frac{k+\omega-1}{k+\omega}, \quad \kappa_k := \frac{k+\omega-2}{2(k+\omega)}$$
$$\text{and } \zeta_k := \begin{cases} 0 & \text{if } k = 0 \\ \frac{k+\omega-3}{2(k+\omega)} & \text{if } k \geq 1 \end{cases}$$

where $\omega > 1$ is a given parameter. Then, we have

$$\begin{cases} \|G\widehat{x}_{k+1}\|^2 \leq \frac{3(1+L^2\widehat{\gamma}^2)\widetilde{C}_0\|y_0-y^\star\|^2}{(k+\omega-1)^2} \\ \|Gz_k\|^2 \leq \frac{3\widetilde{C}_0\|y_0-y^\star\|^2}{2(k+\omega-1)^2} \end{cases} \tag{69}$$

where $\widetilde{C}_0 := \frac{2\omega\widehat{\gamma}^2L^2\left[1+(\omega-1)^2\right]+16(\omega-1)}{\widehat{\gamma}^2(\omega-1)} > 0$ is rendered from C_0 of Theorem 6.

Proof. First, by the choice of $\beta_k := \frac{1}{k+\omega}$ for some $\omega > 1, \widehat{\eta}_k := \widehat{\eta} \in \left(0, \frac{1}{2L}\right]$, and $\eta_k := \widehat{\eta}(1-\beta_k)$ in Theorem 6, using the update rules of parameters in Theorem 7, we can easily show that $\theta_k = \frac{k+\omega-2}{k+\omega}$, $\nu_k = \frac{k+\omega-1}{k+\omega}$, $\widehat{\gamma}_k = \widehat{\gamma} := 2\widehat{\eta}$, $\kappa_k = \frac{k+\omega-2}{2(k+\omega)}$, and $\zeta_k = \frac{k+\omega-3}{2(k+\omega)}$ if $k \geq 1$, and $\zeta_k = 0$ if $k = 0$, as given in Corollary 3.

Next, since $\widehat{x}^{k+1} = z_k - \widehat{\gamma}_k Gz_k = z_k - \widehat{\gamma}Gz_k$ due to (66), we have $\|G\widehat{x}_{k+1}\|^2 \leq 2\left\|G\widehat{x}_{k+1} - Gz_k\right\|^2 + 2\|Gz_k\|^2 \leq 2L^2\left\|\widehat{x}_{k+1} - z_k\right\|^2 + 2\|Gz_k\|^2 = 2L^2\widehat{\gamma}^2\|Gz_k\|^2 + 2\|Gz_k\|^2 = 2(1+L^2\widehat{\gamma}^2)\|Gz_k\|^2$. Combining this inequality and the second line of (62), and noting that $\widehat{x}_0 = y_0$, we obtain the first line of (69). Finally, the second line of (69) directly comes from the second line of (62).

Unlike (6), the new scheme (66) has convergence without the co-coerciveness of G. It only requires G to be monotone and L-Lipschitz continuous, and one evaluation of G per iteration. □

6 Conclusion

In this work, we developed a comprehensive framework that unifies Halpern's and Nesterov's methods, highlighting their shared foundation in monotone inclusions and fixed-point theory. By translating between the two schemes, we offered new perspectives on their application to classical optimization methods, such as proximal-point and Douglas-Rachford splitting. This study lays the groundwork for continued research into accelerated algorithms in both deterministic and stochastic contexts, with potential to improve large-scale optimization techniques used in practical machine learning and engineering problems.

References

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

[Attouch & Fadili, 2022] Attouch, H. & Fadili, J. (2022). From the ravine method to the nesterov method and vice versa: A dynamical system perspective. *SIAM J. Optim.*, 32(3), 2074–2101.

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

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

[Bauschke, 1996] Bauschke, H. (1996). The approximation of fixed points of compositions of nonexpansive mappings in hilbert space. *J. Math. Anal. Appl.*, 202(1), 150–159.

[Bauschke & Combettes, 2017] Bauschke, H. & Combettes, P. (2017). *Convex analysis and monotone operators theory in Hilbert spaces*. Berlin: Springer, 2nd edition.

[Bauschke et al., 2021] Bauschke, H., Moursi, W., & Wang, X. (2021). Generalized monotone operators and their averaged resolvents. *Math. Program.*, 189, 55–74.

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

[Bot et al., 2022] Bot, R., Csetnek, E., & Nguyen, D. (2022). Fast ogda in continuous and discrete time. *arXiv preprint arXiv:2203.10947*.

[Bot & Nguyen, 2022] Bot, R. & Nguyen, D. (2022). Fast krasnoselśkii-mann algorithm with a convergence rate of the fixed point iteration of $o(1/k)$. *arXiv preprint arXiv:2206.09462*.

[Bubeck et al., 2015] Bubeck, S., Lee, Y., & Singh, M. (2015). A geometric alternative to nesterov's accelerated gradient descent. *arXiv preprint arXiv:1506.08187*.

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

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

[Combettes & Wajs, 2005] Combettes, P. & Wajs, V. (2005). Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4, 1168–1200.

[d'Aspremont et al., 2021] d'Aspremont, A., Scieur, D., & Taylor, A. (2021). Acceleration methods. *Found. Trends® Optim.*, 5(1-2), 1–245.

[Davis & Yin, 2017] Davis, D. & Yin, W. (2017). A three-operator splitting scheme and its optimization applications. *Set-Valued Var. Anal.*, 25(4), 829–858.

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

[Diakonikolas et al., 2021] Diakonikolas, J., Daskalakis, C., & Jordan, M. (2021). Efficient methods for structured nonconvex-nonconcave min-max optimization. In *International Conference on Artificial Intelligence and Statistics* (pp. 2746–2754).: PMLR.

[Facchinei & Pang, 2003] Facchinei, F. & Pang, J.-S. (2003). *Finite-Dimensional Variational Inequalities and Complementarity Problems*, volume 1-2. Berlin: Springer.

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

[He & Yuan, 2015] He, B. & Yuan, X. (2015). On the convergence rate of douglas-rachford operator splitting method. *Math. Program.*, 153(2), 715–722.

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

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

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

[Körnlein, 2015] Körnlein, D. (2015). Quantitative results for halpern iterations of nonexpansive mappings. *J. Math. Anal. Appl.*, 428(2), 1161–1172.

[Labarre & Maingé, 2022] Labarre, F. & Maingé, P.-E. (2022). First-order frameworks for continuous newton-like dynamics governed by maximally monotone operators. *Set-Valued Var. Anal.*, 30(2), 425–451.

[Lee & Kim, 2021] Lee, S. & Kim, D. (2021). Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. In *Thirty-fifth Conference on Neural Information Processing Systems (NeurIPs2021)*, volume 34 (pp. 22588–22600).

[Lieder, 2021] Lieder, F. (2021). On the convergence rate of the halpern-iteration. *Optim. Lett.*, 15(2), 405–418.

[Lions & Mercier, 1979] Lions, P. & Mercier, B. (1979). Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Num. Anal.*, 16, 964–979.

[Malitsky, 2015] Malitsky, Y. (2015). Projected reflected gradient methods for monotone variational inequalities. *SIAM J. Optim.*, 25(1), 502–520.

[Monteiro & Svaiter, 2010] Monteiro, R. & Svaiter, B. (2010). On the complexity of the hybrid proximal extragradient method for the interates and the ergodic mean. *SIAM J. Optim.*, 20(6), 2755–2787.

[Nemirovskii, 2004] Nemirovskii, A. (2004). Prox-method with rate of convergence $\mathcal{O}(1/t)$ for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.*, 15(1), 229–251.

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

[Nesterov, 2004] Nesterov, Y. (2004). *Introductory lectures on convex optimization: A basic course*, volume 87. Kluwer Academic Publishers.

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

[Ouyang & Xu, 2021] Ouyang, Y. & Xu, Y. (2021). Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems. *Math. Program.*, 185(1-2), 1–35.

[Park & Ryu, 2022] Park, J. & Ryu, E. (2022). Exact optimal accelerated complexity for fixed-point iterations. In *International Conference on Machine Learning (ICML)* (pp. 17420–17457).

[Phelps, 2009] Phelps, R. (2009). *Convex Functions, Monotone Operators and Differentiability*, volume 1364. Springer.

[Polyak, 1964] Polyak, B. (1964). Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.*, 4(5), 1–17.

[Popov, 1980] Popov, L. (1980). A modification of the arrow-hurwicz method for search of saddle points. *Math. Notes Acad. Sci. USSR*, 28(5), 845–848.

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

[Rockafellar & Wets, 2004] Rockafellar, R. & Wets, R. (2004). *Variational Analysis*, volume 317. Springer.

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

[Sabach & Shtern, 2017] Sabach, S. & Shtern, S. (2017). A first order method for solving convex bilevel optimization problems. *SIAM J. Optim.*, 27(2), 640–660.

[Shi et al., 2021] Shi, B., Du, S., Jordan, M., & Su, W. (2021). Understanding the acceleration phenomenon via high-resolution differential equations. *Math. Program.*, (pp. 1–70).

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

[Tran-Dinh & Luo, 2021] Tran-Dinh, Q. & Luo, Y. (2021). Halpern-type accelerated and splitting algorithms for monotone inclusions. *arXiv preprint*.

[Tran-Dinh & Luo, 2023] Tran-Dinh, Q. & Luo, Y. (2023). Randomized block-coordinate optimistic gradient algorithms for root-finding problems. *arXiv preprint*.

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

[Wibisono et al., 2016] Wibisono, A., Wilson, A., & Jordan, M. (2016). A variational perspective on accelerated methods in optimization. *Proc. Natl. Acad. Sci.*, 113(47), E7351–E7358.

[Wittmann, 1992] Wittmann, R. (1992). Approximation of fixed points of nonexpansive mappings. *Arch. Math.*, 58(5), 486–491.

[Xu, 2002] Xu, H.-K. (2002). Iterative algorithms for nonlinear operators. *J. Lond. Math. Soc.*, 66(1), 240–256.

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

[Yoon & Ryu, 2022] Yoon, T. & Ryu, E. (2022). Accelerated minimax algorithms flock together. *arXiv preprint*.