Symplectic Extra-Gradient Methods for Fast Convergence in Non-Monotone and Min-Max Optimization Problems

Chris Junchi Li[◇]

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

September 28, 2024

Abstract

This paper presents a novel accelerated algorithmic framework, termed the Symplectic ExtraGradient (SEG) Method, for solving general inclusion problems, particularly focusing on non-monotone scenarios. Min-max problems, especially those characterized by nonconvexnonconcave structures, pose significant challenges due to their complex solution landscapes. By leveraging symplectic acceleration techniques, the SEG method achieves a faster convergence rate of $o\left(1/k^2\right)$, outperforming traditional extra-gradient methods. We also introduce an adaptive line search mechanism to handle unknown or rapidly varying Lipschitz constants and comonotone indices, further enhancing the robustness of our method. Theoretical analyses demonstrate the weak convergence properties of the SEG method, while numerical experiments validate its superior performance compared to existing extra-gradient techniques. Our results extend the applicability of extra-gradient algorithms to more general and difficult optimization problems, advancing the state-of-the-art in non-monotone inclusion problem-solving.

Keywords: Symplectic Extra-Gradient Method; Non-Monotone Inclusion Problem; Min-Max Optimization; Nonconvex Optimization; Accelerated Gradient Methods; Lipschitz Continuous Operator

1 Introduction

Min-max optimization problems lie at the heart of many fundamental areas in mathematics, economics, and computer science. From game theory to machine learning, the formulation and efficient solution of min-max problems are pivotal. The rise of adversarial machine learning and Generative Adversarial Networks (GANs) has further accelerated interest in solving nonconvex-nonconcave min-max problems, where the classical duality approaches often fail due to the inherent complexity of such problems.

Traditionally, optimization methods for convex-concave min-max problems rely on monotonicity assumptions, making them well-suited for problems with known structural properties. However, many real-world applications require solving inclusion problems where the monotonicity assumption does not hold, leading to non-monotone systems. Solving such problems demands new algorithmic advancements that can handle the increased complexity and lack of convexity.

The extra-gradient (EG) method has been a successful tool for solving monotone inclusion problems due to its simplicity and favorable theoretical properties. However, when dealing with non-monotone systems or systems with rapid local variations in Lipschitz continuity, traditional EG methods struggle. The challenge becomes even more pronounced in stochastic settings or when attempting to generalize beyond convex-concave assumptions.

In this paper, we address these challenges by proposing a new method, termed the Symplectic Extra-Gradient (SEG) method, which builds on the symplectic acceleration framework recently developed for proximal point algorithms. This method is designed to solve both monotone and non-monotone inclusion problems while achieving superior convergence rates compared to existing approaches. Our method also incorporates a line search mechanism to adapt to varying Lipschitz constants, making it more robust in practice.

Background Min-max problem and min-max duality theory lie at the foundations of game theory, designing algorithms [Val14, ZC08] and duality theory of mathematical programming [JFB00], and have found far-reaching applications across a range of disciplines, including decision theory [Mye97], economics [vNM47], structural design [TB84], control theory [VKA10] and robust optimization [BTGN09]. Recently, as burgeoning of Generative Adversarial Networks [GBV+18, GPAM+14] and adversarial attacks [MMS+18], solving min-max problem under nonconvex-nonconcave assumption has gained researchers' attention. However, due to the nonconvex-nonconcave assumption, solving min-max problem exactly is nearly impossible. Similar to nonlinear programming, solving first-order stationary point of nonconvex-nonconcave min-max problem is relatively easier. The first-order stationary condition of most min-max problems can be described by the following inclusion problem

$$0 \in T(z) := F(z) + G(z), \quad z \in \mathcal{H} \tag{1}$$

where $\mathcal{H}$ is a real Hilbert space equipped with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. Let $\text{zero}(T)$ be the set of all solutions of (1). Throughout this paper, we assume $F : \mathcal{H} \to \mathcal{H}$ is a single-value $L-$Lipschitz continuous operator, $G : \mathcal{H} \to 2^{\mathcal{H}}$ is a set-valued maximally monotone operator and $\text{zero}(T)$ is non-empty.

When $G = 0$, F is monotone and $L-$Lipschitz continuous, the extra-gradient (EG) method proposed in [Kor76] and [Pov80] is an efficient method for solving (1). Theoretically, the EG method has $O(1/k)$ ergodic convergence rate without additional requirement on F. Practically, the EG method can be described as applying gradient descent step twice per iteration. Owning to its fascinating theoretical properties and simple recursive rule, extensions of the EG method have been studied, such as the mirror-prox method in [KS11] and [Nem04] when G is the normal cone of convex set C, the Tseng's splitting method or forward-backward splitting method in [RFP13] and [Tse00] when $G \neq 0$ and the EG+ method [DDJ21] to solve (1) under non-monotone assumption. The adaptive EG+ method proposed in [FLC23] and [PLP+22] further generalizes the EG+ method by considering an adaptive step-size. Here, we refer to the survey [TD23] of the EG method for more existing theory about the EG method. As stochastic optimization continues to gain significant attention, research on stochastic extra-gradient method has seen a rise, such as [GBGL22], [IJOT17], [KS19] and [MKS+20].

Nesterov's accelerated gradient method [Nes83] has stimulated research on acceleration technique. The first acceleration form of EG method for solving monotone inclusion problem, called extra anchored gradient (EAG) algorithm, was presented by [YR21]. From the observation that the EG method is closed to proximal point algorithm [MOP20], [YR21] uses the Halpern's iteration [DDJ21, Hal67, QX21], which is a widely studied acceleration technique for the proximal point algorithm, to derive the EAG algorithm. Also, the complexity lower bound of EG type method was proven to be $O(1/k^2)$ in [YR21]. Nowadays, studies of acceleration form of the EG method have increased. [LK21] combined the EAG algorithm and the EG+ method and obtained the fast EG method, which can solve (1) under non-monotone assumption. [COZ22] obtained an accelerated forward-backward splitting method by further generalizing the fast EG method.

However, the EAG type method is prone to oscillatory phenomena, which eventually slows down numerical performance. That is because the Halpern's iteration can be characterized as calculating convex combination of the initial starting point and the current feasible point. The use of the initial starting point may cause the current feasible point far away from the solution set, leading to oscillation phenomena.

1.1 Our Contributions

In Section 3, we exploit a recently proposed acceleration method, called *symplectic acceleration* [YZ23], to devise a novel accelerated variant of the extra-gradient method, named *Symplectic Extra-Gradient (SEG) Method*. Capitalizing on the Lyapunov function framework proposed in [YZ23], we establish that both the SEG and its extended versions exhibit a convergence rate of $O(1/k^2)$. Moreover, in Section 4, we demonstrate that by imposing stricter conditions, we can prove the SEG type methods admits a faster $o(1/k^2)$ convergence rate and weak convergence property. To the best of our knowledge, the SEG type method is the first EG type method with proved $o(1/k^2)$ convergence rate. A concise summary of these theoretical results is presented in Table 1.

In Section 5, we address the challenge posed by the difficulty in precisely estimating the Lipschitz constant and the comonotone index, as well as handling situations involving rapid variations in both the local Lipschitz constant and the local comonotone index. To this end, we integrate the line search framework into the SEG framework. Under specified assumptions, we prove the convergence of the SEG method equipped with line search. Our numerical experiments demonstrate that our method performs better than several existing EG type methods.

2 Preliminaries

2.1 Basic Concepts

Let $\mathcal{H}$ be a real Hilbert space, and let $T : \mathcal{H} \to 2^{\mathcal{H}}$ is a set-valued operator. The operator T is said to be $\rho-$*comonotone* if

$$\langle u - v, x - y \rangle \geqslant \rho \left\| u - v \right\|^2, \quad \forall x,\ y \in \mathcal{H},\ u \in T(x),\ v \in T(y)$$

When $\rho = 0$, the concept of $0-$comonotone coincide with the concept of monotone. If $\rho > 0$, $\rho-$comonotone is the same as $\rho-$coercive. A monotone operator G is *maximally* if the graph of G is not a proper subset of the graph of another monotone operator. The Minty surjectivity theorem [Min62] shows that a monotone operator G is maximal if and only if the domain of resolvent $J_G = (I + G)^{-1}$ of G is $\mathcal{H}$. Also, a continuous and monotone single-value operator is a maximally monotone operator. For further theories on comonotone operator, we refer to [BMW21].

Methods	Convergence Rate	Weak Convergence Property
EG+ [DDJ21]	$O(1/k)$	-
FEG [LK21]	$O(1/k^2)$	-
SEG+ (This paper)	$o(1/k^2)$	✓

Table 1: Illustration of theoretical results in this paper.

2.2 Extra-gradient Method and Extra-gradient+ Method

The extra-gradient method for solving zero-point of $L-$Lipschitz continuous and monotone operator F, i. e. solving

$$0 = F(z)$$

is given as follows:

$$\begin{aligned} z_{k+\frac{1}{2}} &= z_k - sF(z_k) \\ z_{k+1} &= z_k - sF(z_{k+\frac{1}{2}}) \end{aligned} \tag{2}$$

In [Kor76] and [Tse95], the proof of convergence results of EG method is based on the following inequality:

$$\|z_{k+1} - z^*\|^2 \leqslant \|z_k - z^*\|^2 - (1 - s^2L^2)\|F(z_k)\|^2, \quad \forall 0 = F(z^*)$$

By summing the above inequality respect to k, one can easily show that the ergodic convergence rate of the EG method on $\|F(z_k)\|^2$ is $O(1/k)$ when $0 < s < \dfrac{1}{L}$. The result that the sequence $\{z_k\}$ converges weakly to a zero-point of F relies on the following propositions.

Proposition 1 (Lemma 2.47 in [BC17]). *Let $\{z_k\}$ be a sequence in $\mathcal{H}$ and let C be a nonempty subset of $\mathcal{H}$. Suppose that, for every $z \in C$, $\|z_k - z\|$ converges and that every weak sequential cluster point of $\{z_k\}$ belongs to C. Then $\{z_k\}$ converges weakly to a point in C.*

Proposition 2 (Proposition 20.38 (ii) in [BC17]). *Let $T : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, $\{z_k\}$ and $\{v_k\}$ be the sequences in $\mathcal{H}$ such that $v_k \in T(z_k)$. If $z_k \rightharpoonup z$ and $v_k \to v$. Then $v \in T(z)$.*

The details of the proof of weak convergence property of the EG method can be found in [NT06]. The EG+ method proposed in [DDJ21] for solving zero-point of $L-$Lipschitz continuous and $\rho-$comonotone operator F is given as follows:

$$\begin{aligned} z_{k+\frac{1}{2}} &= z_k - \frac{s}{\beta}F(z_k) \\ z_{k+1} &= z_k - sF(z_{k+\frac{1}{2}}) \end{aligned} \tag{3}$$

Similar to the EG method, [DDJ21] shows that if $\beta = \dfrac{1}{2}$, $s = \dfrac{1}{2L}$, $\rho > -\dfrac{1}{8L}$, then we have

$$\|z_{k+1} - z^*\|^2 \leqslant \|z_k - z^*\| - \frac{1}{4L}\left(\frac{1}{4L} + 2\rho\right)\left\|F(z_{k+\frac{1}{2}})\right\|^2, \quad \forall 0 = F(z^*)$$

By using the above inequality respect to k, one can prove the ergodic convergence rate of the EG+ method.

2.3 Anchor Acceleration

Since [MOP20] suggested that the sequence generated by the EG method is closed to the sequence generated by the proximal point algorithm, recent study of acceleration form of the EG method is based on existing acceleration technique for proximal point algorithm. Among most acceleration form of proximal point algorithm, the Halpern's iteration [Hal67] is one of the most wildly studied acceleration technique. The Halpern's iteration for solving zero-point of maximally monotone operator F is

$$z_{k+1} = \alpha_k z_0 + (1-\alpha_k) J_F(z_k) \tag{4}$$

Replacing the proximal point term by the EG step, we obtain the extra anchored gradient algorithm [YR21]:

$$\begin{aligned} z_{k+\frac{1}{2}} &= \alpha_k z_0 + (1-\alpha_k) z_k - \beta_k F(z_k) \\ z_{k+1} &= \alpha_k z_0 + (1-\alpha_k) z_k - \beta_k F(z_{k+\frac{1}{2}}) \end{aligned} \tag{5}$$

In [YR21], (5) was proven to be $O(1/k^2)$ convergence rate if F is an $L-$Lipschitz continuous and monotone operator and the parameters in (5) satisfy several assumptions. Combining the EG+ method and (5), [LK21] gave the fast EG (FEG) method as follows:

$$\begin{aligned} z_{k+\frac{1}{2}} &= \alpha_k z_0 + (1-\alpha_k) z_k - (1-\alpha_k)\left(\frac{1}{L} + 2\rho\right) F(z_k) \\ z_{k+1} &= \alpha_k z_0 + (1-\alpha_k) z_k - \frac{1}{L} F(z_{k+\frac{1}{2}}) - (1-\alpha_k) 2\rho F(z_k) \end{aligned} \tag{6}$$

If $\alpha_k = \dfrac{1}{k+1}$, (6) admits $O(1/k^2)$ convergence rate. The convergence results of accelerated forward-backward splitting method in [COZ22] is similar to the convergence results (6).

2.4 Symplectic Acceleration

The symplectic acceleration technique proposed in [YZ23] is a new way to accelerate proximal point algorithm. The symplectic proximal point algorithm for solving zero-point of maximally monotone operator F is given as follows:

$$\begin{aligned} \widetilde{z}_{k+1} &= \frac{k}{k+r} z_k + \frac{r}{k+r} u_k \\ z_{k+1} &= J_F(z_k) \\ u_{k+1} &= u_k + \frac{D}{r}(z_{k+1} - \widetilde{z}_{k+1}) \end{aligned} \tag{7}$$

The reason why (7) is called symplectic proximal point algorithm is that (7) is obtained by applying first-order implicit symplectic method to a first-order ODEs system. If $r > 1$, $0 < D < r-1$, then (7) admits $O(1/k^2)$ convergence rate. Moreover, $o(1/k^2)$ convergence rate and weak convergence property of (7) can be established by considering two auxiliary functions.

3 Symplectic Extra-gradient Type Method

In this section, we introduce the symplectic extra-gradient+ framework for solving general inclusion problem

$$0 \in T(z) := F(z) + G(z)$$

At first, we explain how to propose symplectic extra-gradient framework by considering the zero-point problem $0 = F(z)$. In Section 2.4, we introduce the symplectic proximal point algorithm (7).

(7) can be generalized as follows:

$$
\begin{aligned}
\widetilde{z}_{k+1} &= (1-\alpha_k)z_k + \alpha_k u_k \\
z_{k+1} &= J_F(\widetilde{z}_{k+1}) \\
u_{k+1} &= u_k + C_k(z_{k+1} - \widetilde{z}_{k+1})
\end{aligned} \tag{8}
$$

To formulate the desired algorithm, we need a two-step modification to equation (8). Due to the fact that $\widetilde{z}_{k+1} - z_{k+1} = F(z_{k+1})$, the update rule for u_{k+1} is transformed to

$$
u_{k+1} = u_k - C_k F(z_{k+1})
$$

Moreover, similar to (5), we replace the proximal point step by the EG step and obtain the *symplectic extra-gradient (SEG) method* described as follows:

$$
\widetilde{z}_{k+1} = (1-\alpha_k)z_k + \alpha_k u_k \tag{9a}
$$

$$
z_{k+\frac{1}{2}} = \widetilde{z}_{k+1} - \beta_k F(z_k) \tag{9b}
$$

$$
z_{k+1} = \widetilde{z}_{k+1} - \beta_k F(z_{k+\frac{1}{2}}) \tag{9c}
$$

$$
u_{k+1} = u_k - C_k F(z_{k+1}) \tag{9d}
$$

The coefficients $\{\alpha_k\}, \{\beta_k\}$ in (9) are positive for all $k \geqslant 1$. First, we notice that if we does not impose any constrains on $\{\alpha_k\}$ and $\{C_k\}$, the EG method and EAG algorithm can be included in (9). Specifically, let $\alpha_k \equiv 1, \beta_k \equiv \dfrac{1}{L}$, (9) can be simplified as (2), and let $C_k \equiv 0$, (9) is the same as (5). However, it is noteworthy that despite the apparent requirement of three gradient evaluations per iteration in (9), a strategic reordering of the iterative process can economize computations to merely two gradient evaluations per loop. This is achieved through the adoption of the following alternative iteration rule:

$$
\begin{aligned}
u_{k+1} &= u_k - C_k F(z_k) \\
\widetilde{z}_{k+1} &= (1-\alpha_k)z_k + \alpha_k u_{k+1} \\
z_{k+\frac{1}{2}} &= \widetilde{z}_{k+1} - \beta_k F(z_k) \\
z_{k+1} &= \widetilde{z}_{k+1} - \beta_k F(z_{k+\frac{1}{2}})
\end{aligned}
$$

The employment of such a reordering methodology can decrease the quantity of necessary gradient evaluations within subsequent algorithms. Consequently, further expounding on this technique will be omitted in the remaining text.

It should be noted that the SEG method is only suitable for solving zero-point of F under monotone assumption, coupled with complicate convergence results discussed in Section A. As such, we focus on the symplectic extra-gradient+ framework in the remaining text. Inspired by the FEG method (6) proposed in [LK21], the corresponding *symplectic extra-gradient+ (SEG+) method* for solving zero-point of $L-$Lipschitz continuous and $\rho-$comonotone operator F is given as follows:

$$
\widetilde{z}_{k+1} = (1-\alpha_k)z_k + \alpha_k u_k \tag{10a}
$$

$$
z_{k+\frac{1}{2}} = \widetilde{z}_{k+1} - (1-\alpha_k)\left(\frac{1}{L} + 2\rho\right) F(z_k) \tag{10b}
$$

Algorithm 1: Symplectic Forward-Backward Splitting Method, SFBS

Input: Operators F and G, $L \in (0, +\infty), \rho \in \left(-\dfrac{1}{2L}, +\infty\right)$;

Initialization: z_0, $u_0 = z_0$, $\widetilde{G}_0 = 0$;

for $k = 0, 1, \cdots$ **do**

$$\widetilde{z}_{k+1} = \frac{k}{k+r} z_k + \frac{r}{k+r} u_k;$$
$$z_{k+\frac{1}{2}} = \widetilde{z}_{k+1} - \frac{k}{k+r}\left(\frac{1}{L} + 2\rho\right)(F(z_k) + \widetilde{G}_k);$$
$$z_{k+1} = J_{L^{-1}G}\left(\widetilde{z}_{k+1} - \frac{1}{L}F(z_{k+\frac{1}{2}}) - \frac{2\rho k}{k+r}(F(z_k) + \widetilde{G}_k)\right);$$
$$\widetilde{G}_{k+1} = L\left[\widetilde{z}_{k+1} - z_{k+1} - \frac{1}{L}F(z_{k+\frac{1}{2}}) - \frac{2\rho k}{k+r}(F(z_k) + \widetilde{G}_k)\right];$$
$$u_{k+1} = u_k - \frac{D}{r}(F(z_{k+1}) + \widetilde{G}_{k+1}).$$

$$z_{k+1} = \widetilde{z}_{k+1} - \frac{1}{L}F(z_{k+\frac{1}{2}}) - (1-\alpha_k)2\rho F(z_k) \tag{10c}$$
$$u_{k+1} = u_k - C_k F(z_{k+1}) \tag{10d}$$

Motivated by Tseng's splitting algorithm [Tse00] and the accelerated forward-backward splitting method [COZ22], we can further generalize the SEG+ method and propose the *symplectic forward-backward splitting (SFBS) method* with $\alpha_k = \dfrac{r}{k+r}$ and $C_k = \dfrac{D}{r}$, which is described in Algorithm 1.

Next, we apply the Lyapunov function technique to demonstrate the convergence rates of (10). Here, inspired by the Lyapunov function in [LK21] for proving the convergence rates of the FEG method and the Lyapunov function in [YZ23] for proving the convergence rates of the symplectic proximal point algorithm, the Lyapunov function for Algorithm 1 is

$$\begin{aligned}\mathcal{E}(k) := &\left[\frac{Dk^2}{2L} + \rho Dk(k-r)\right]\left\|F(z_k) + \widetilde{G}_k\right\|^2 \\ &+ Drk\left\langle F(z_k) + \widetilde{G}_k, z_k - u_k\right\rangle + \frac{r^3 - r^2}{2}\left\|u_k - z^*\right\|^2,\end{aligned} \tag{11}$$

where $z^* \in \text{zero}(T)$. Given the fundamental role that the difference of the Lyapunov function, i. e. $\mathcal{E}(k+1) - \mathcal{E}(k)$, plays in establishing convergence rates for (10), it becomes imperative to estimate the upper bound of the difference. By estimating the upper bound of $\mathcal{E}(k+1) - \mathcal{E}(k)$, we can give the convergence results of Algorithm 1 in Theorem 1.

Theorem 1. *Let $\{z_k\}$ and $\{u_k\}$ be the sequences generated by Algorithm 1. If $r > 1$ and $0 < D \leqslant (r-1)\left(\dfrac{1}{L} + 2\rho\right)$, then we have*

$$\left(\frac{1}{2L} + \rho - \frac{D}{2(r-1)}\right)k^2 \text{dist}(0, T(z_k))^2 \leqslant \frac{r^3 - r^2}{2D} \cdot \text{dist}(z_0, \text{zero}(T))^2, \quad \forall k \geqslant 1$$

Thus if $D < (r-1)\left(\frac{1}{L}+2\rho\right)$, the last-iterate convergence rate of Algorithm 1 is

$$\operatorname{dist}(0, T(z_k))^2 \leqslant \frac{(r-1)^2 r^2}{\left[(r-1)\left(\frac{1}{L}+2\rho\right)D - D^2\right]k^2} \cdot \operatorname{dist}(z_0, \operatorname{zero}(T))^2, \quad \forall k \geqslant 1$$

In addition, we have

$$\begin{aligned}&\sum_{k=0}^{\infty} \frac{1}{2}\left[(r-1)\left(\frac{1}{L}+2\rho\right) - D\right](2k+r+1)\left\|F(z_{k+1}) + \widetilde{G}_{k+1}\right\|^2 \\ \leqslant\ & \frac{r^3 - r^2}{2D} \cdot \operatorname{dist}(z_0, \operatorname{zero}(T))^2\end{aligned}$$

Proof. Let $\widetilde{T}(z_k) = F(z_k) + \widetilde{G}_k$. By the definition of $\widetilde{G}_k$, we have

$$z_{k+1} = \widetilde{z}_{k+1} - \frac{1}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{G}_{k+1}] - \frac{2\rho k}{k+r}\widetilde{T}(z_k)$$

With above equation, we obtain the following useful equations

$$k(z_{k+1} - z_k) = r(u_k - z_{k+1}) - \frac{k+r}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{G}_{k+1}] - 2\rho k\widetilde{T}(z_k) \tag{12}$$

$$r(z_k - u_k) = -(k+r)(z_{k+1} - z_k) - \frac{k+r}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{G}_{k+1}] - 2\rho k\widetilde{T}(z_k) \tag{13}$$

Also, by $L-$Lipschitz continuity of F, we have

$$\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2 \leqslant L^2\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2$$

Since

$$z_{k+1} - z_{k+\frac{1}{2}} = \frac{1}{L}\left[\frac{k}{k+r}\widetilde{T}(z_k) - F(z_{k+\frac{1}{2}}) - \widetilde{G}_{k+1}\right]$$

we have

$$\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2 \leqslant \left\|\frac{k}{k+r}\widetilde{T}(z_k) - F(z_{k+\frac{1}{2}}) - \widetilde{G}_{k+1}\right\|^2 \tag{14}$$

Recall the Lyapunov function

$$\mathcal{E}(k) = \left[\frac{Dk^2}{2L} + \rho Dk(k-r)\right]\left\|\widetilde{T}(z_k)\right\|^2 + Drk\left\langle \widetilde{T}(z_k), z_k - u_k\right\rangle + \frac{r^3 - r^2}{2}\|u_k - z^*\|^2$$

First, we divide the difference $\mathcal{E}(k+1) - \mathcal{E}(k)$ into three parts.

$$\begin{aligned}&\mathcal{E}(k+1) - \mathcal{E}(k)\\ =\ & \underbrace{\left[\frac{D(k+1)^2}{2L} + \rho D(k+1)(k+1-r)\right]\left\|\widetilde{T}(z_{k+1})\right\|^2 - \left[\frac{Dk^2}{2L} + \rho Dk(k-r)\right]\left\|\widetilde{T}(z_k)\right\|^2}_{\mathrm{I}}\\ & + \underbrace{Dr(k+1)\left\langle \widetilde{T}(z_{k+1}), z_{k+1} - u_{k+1}\right\rangle - Drk\left\langle \widetilde{T}(z_k), z_k - u_k\right\rangle}_{\mathrm{II}}\end{aligned}$$

$$+ \underbrace{\frac{r^3 - r^2}{2} \left[\|u_{k+1} - z^*\|^2 - \|u_k - z^*\|^2 \right]}_{\text{III}}$$

Secondly, we reckon the upper bound of II and III. First we consider II.

$$\begin{aligned} \text{II} = & \, Dr \left\langle \widetilde{T}(z_{k+1}), z_{k+1} - u_{k+1} \right\rangle + Drk \left\langle \widetilde{T}(z_{k+1}), z_{k+1} - z_k - u_{k+1} + u_k \right\rangle \\ & + Drk \left\langle \widetilde{T}(z_{k+1}) - \widetilde{T}(z_k), z_k - u_k \right\rangle \end{aligned}$$

Due to (12) and the equality $u_{k+1} - u_k = -\frac{D}{r}[F(z_{k+1}) + \widetilde{G}_{k+1}]$, we have

$$\begin{aligned} & Drk \left\langle \widetilde{T}(z_{k+1}), z_{k+1} - z_k - u_{k+1} + u_k \right\rangle \\ = & \, Dr \left\langle \widetilde{T}(z_{k+1}), r(u_k - z_{k+1}) - \frac{k+r}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{G}_{k+1}] - 2\rho k \widetilde{T}(z_k) \right\rangle \\ & + D^2 k \left\| \widetilde{T}(z_{k+1}) \right\|^2 \end{aligned} \tag{15}$$

In addition, because of (13), we have

$$\begin{aligned} & Drk \left\langle \widetilde{T}(z_{k+1}) - \widetilde{T}(z_k), z_k - u_k \right\rangle \\ = & \, Dk \left\langle \widetilde{T}(z_{k+1}) - \widetilde{T}(z_k), -(k+r)(z_{k+1} - z_k) - \frac{k+r}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{G}_{k+1}] - 2\rho k \widetilde{T}(z_k) \right\rangle \end{aligned} \tag{16}$$

Also,

$$Dr \left\langle \widetilde{T}(z_{k+1}), z_{k+1} - u_{k+1} \right\rangle = D^2 \left\| \widetilde{T}(z_{k+1}) \right\|^2 + Dr \left\langle \widetilde{T}(z_{k+1}), z_{k+1} - u_k \right\rangle. \tag{17}$$

By summing (44), (45) and (46), we have

$$\begin{aligned} & \text{II} \\ = & \, -\frac{D(k+r)^2}{L} \left\langle F(z_{k+1}), F(z_{k+\frac{1}{2}}) \right\rangle + \frac{D}{L} \left\langle F(z_{k+\frac{1}{2}}), k(k+r)\widetilde{T}(z_k) - (k+r)^2 \widetilde{G}_{k+1} \right\rangle \\ & + \frac{D}{L} \left\langle \widetilde{G}_{k+1}, k(k+r)\widetilde{T}(z_k) - (k+r)^2 \widetilde{T}(z_{k+1}) \right\rangle \\ & + D^2(k+1) \left\| \widetilde{T}(z_{k+1}) \right\|^2 - Dk(k+r) \left\langle \widetilde{T}(z_{k+1}) - \widetilde{T}(z_k), z_{k+1} - z_k \right\rangle \\ & - 2D\rho k(k+r) \left\langle \widetilde{T}(z_{k+1}), \widetilde{T}(z_k) \right\rangle + 2D\rho k^2 \left\| \widetilde{T}(z_k) \right\|^2 \\ & + D(r - r^2) \left\langle \widetilde{T}(z_{k+1}), z_{k+1} - u_k \right\rangle \end{aligned}$$

Combining $\rho-$comonotonicity of F, (14) and the following equality

$$-\left\langle F(z_{k+1}), F(z_{k+\frac{1}{2}}) \right\rangle = \frac{1}{2} \left[\left\| F(z_{k+1}) - F(z_{k+\frac{1}{2}}) \right\|^2 - \|F(z_{k+1})\|^2 - \left\| F(z_{k+\frac{1}{2}}) \right\|^2 \right]$$

we obtain the following estimation

$$\text{II} \leqslant -\frac{D(k+r)^2}{2L} \left\| \widetilde{T}(z_{k+1}) \right\|^2 - \rho Dk(k+r) \left(\left\| \widetilde{T}(z_{k+1}) \right\|^2 + \left\| \widetilde{T}(z_k) \right\|^2 \right)$$

$$+\underbrace{D(r-r^2)\left\langle \widetilde{T}(z_{k+1}), z_{k+1}-u_k\right\rangle}_{\mathrm{II}_1} + D^2(k+1)\left\|\widetilde{T}(z_{k+1})\right\|^2 + \frac{Dk^2}{2L}\left\|\widetilde{T}(z_k)\right\|^2$$

Now estimate III. Owing to the equation

$$\frac{1}{2}\|a\|^2 - \frac{1}{2}\|b\|^2 = \langle a-b, b\rangle + \frac{1}{2}\|a-b\|^2 \tag{18}$$

we have

$$\begin{aligned}
\mathrm{III} &= (r^3-r^2)\langle u_{k+1}-u_k, u_k-z^*\rangle + \frac{r^3-r^2}{2}\|u_{k+1}-u_k\|^2 \\
&= -\underbrace{D(r^2-r)\left\langle \widetilde{T}(z_{k+1}), u_k-z^*\right\rangle}_{\mathrm{III}_1} + \frac{D^2(r-1)}{2}\left\|\widetilde{T}(z_{k+1})\right\|^2
\end{aligned}$$

By summing previous estimation, we can deduce the upper bound of $\mathcal{E}(k+1)-\mathcal{E}(k)$. Since

$$\mathrm{II}_1 - \mathrm{III}_1 = -D(r^2-r)\left\langle \widetilde{T}(z_{k+1}), z_{k+1}-z^*\right\rangle,$$

the upper bound of $\mathcal{E}(k+1)-\mathcal{E}(k)$ can be simplified as follows:

$$\begin{aligned}
&\mathcal{E}(k+1)-\mathcal{E}(k) \\
\leqslant &\left[-(r-1)\left(\frac{1}{L}+2\rho\right)Dk + D^2k - \frac{Dr^2-D}{2L} + \rho D(1-r) + \frac{D^2(r+1)}{2}\right]\left\|\widetilde{T}(z_{k+1})\right\|^2 \\
&- D(r^2-r)\left\langle F(z_{k+1}) + \widetilde{G}_{k+1}, z_{k+1}-z^*\right\rangle \\
\leqslant &-\frac{D}{2}\left[(r-1)\left(\frac{1}{L}+2\rho\right) - D\right](2k+r+1)\left\|\widetilde{T}(z_{k+1})\right\|^2
\end{aligned}$$

Since $r>1$, $D \leqslant (r-1)\left(\frac{1}{L}+2\rho\right)$, we have

$$\mathcal{E}(k+1)-\mathcal{E}(k) \leqslant 0$$

Finally, we can obtain the convergence rates of Algorithm 1. First, we translate $\mathcal{E}(k)$ into the following form.

$$\begin{aligned}
\mathcal{E}(k) = &\left[\frac{Dk^2}{2L} + \rho Dk(k-r) - \frac{D^2k^2}{2(r-1)}\right]\left\|\widetilde{T}(z_k)\right\|^2 + Drk\left\langle \widetilde{T}(z_k), z_k-z^*\right\rangle \\
&+ \frac{1}{2}\left\|\frac{Dk}{\sqrt{r-1}}\widetilde{T}(z_k) - \sqrt{r^3-r^2}(u_k-z^*)\right\|^2
\end{aligned}$$

By $\rho-$comonotonicity of T, we have

$$\begin{aligned}
\mathcal{E}(k) \geqslant &\left[\left(\frac{1}{2L}+\rho\right)Dk^2 - \frac{D^2k^2}{2(r-1)}\right]\left\|\widetilde{T}(z_k)\right\|^2 \\
&+ \frac{1}{2}\left\|\frac{Dk}{\sqrt{r-1}}\widetilde{T}(z_k) - \sqrt{r^3-r^2}(u_k-z^*)\right\|^2
\end{aligned}$$

Since $D \leqslant (r-1)\left(\frac{1}{L}+2\rho\right)$, $\left(\frac{1}{2L}+\rho\right)Dk^2 - \frac{D^2k^2}{2(r-1)} \geqslant 0$. Thus $\mathcal{E}(k) \geqslant 0$ and

$$\left[\left(\frac{1}{2L}+\rho\right)Dk^2 - \frac{D^2k^2}{2(r-1)}\right]\left\|\widetilde{T}(z_k)\right\|^2 \leqslant \mathcal{E}(k) \leqslant \mathcal{E}(0) = \frac{r^3-r^2}{2}\left\|z_0 - z^*\right\|^2$$

By taking infimum respect to all $z^* \in \mathrm{zero}(T)$, we obtain the desire result. Moreover, by summing $\mathcal{E}(k+1)-\mathcal{E}(k)$ from 0 to ∞ and taking infimum respect to all $z^* \in \mathrm{zero}(T)$, we have

$$\begin{aligned}&\sum_{k=0}^{\infty}\frac{1}{2}\left[(r-1)\left(\frac{1}{L}+2\rho\right)-D\right](2k+r+1)\left\|\widetilde{T}(z_{k+1})\right\|^2\\ \leqslant\ &\frac{r^3-r^2}{2D}\mathrm{dist}(z_0,\mathrm{zero}(T))^2\end{aligned}$$

□

Next, we consider a special monotone inclusion problem:

$$0 \in T(z) = F(z) + N_C(z) \tag{19}$$

Here $N_C(z) = \{z^* \in \mathcal{H} | \langle z^*, z'-z\rangle \leqslant 0, \forall z' \in C\}$ is the normal cone of convex subset C at z and F is $L-$Lipschitz continuous and monotone. The inclusion problem is equivalent to the Stampacchia Variational Inequality problem

$$\left\langle F(z), z'-z\right\rangle \geqslant 0, \quad \forall z' \in C. \tag{20}$$

Example 1. *Consider the matrix game problem*

$$\min_{x\in\Delta_m}\max_{y\in\Delta_n} x^{\mathrm{T}}Ay$$

where Δ_m is the $m-1$-dimensional unit simplex. Let $C = \Delta_m \times \Delta_n$, the first-order characterization of matrix game is

$$0 \in \begin{pmatrix} Ay \\ -A^{\mathrm{T}}x\end{pmatrix} + N_C(x,y)$$

Based on the projected EG method in [Kor76] and the accelerated projected EG method in [COZ22], we propose the *symplectc projected extra-gradient+ (SPEG+) method* , described in Algorithm 2.

The corresponding convergence results in given as follows.

Theorem 2. *Let $\{z_k\}$ and $\{u_k\}$ be the sequences generated by Algorithm 2. If F is $L-$Lipschitz and monotone, C is closed and convex, $r > 1$ and $0 < D \leqslant \frac{r-1}{L}$, then we have*

$$\left(\frac{1}{2L} - \frac{D}{2(r-1)}\right)k^2\mathrm{dist}(0,T(z_k))^2 \leqslant \frac{r^3-r^2}{2D}\cdot\mathrm{dist}(z_0,\mathrm{zero}(T))^2, \quad \forall k \geqslant 1$$

$$Drk\left\langle F(z_k)+\widetilde{C}_k, z_k - z^*\right\rangle \leqslant \frac{r^3-r^2}{2}\left\|z_0-z^*\right\|^2, \quad \forall z^* \in \mathrm{zero}(T)$$

Algorithm 2: Symplectic Projected Extra-gradient+ Method, SPEG+

Input: Operator F, Convex set C, $L \in (0, +\infty), \rho \in \left(-\frac{1}{2L}, +\infty\right)$;

Initialization: $z_0, u_0 = z_0$;

for $k = 0, 1, \cdots$ **do**

$$\widetilde{z}_{k+1} = \frac{k}{k+r} z_k + \frac{r}{k+r} u_k;$$
$$z_{k+\frac{1}{2}} = P_C\left(\widetilde{z}_{k+1} - \frac{k}{(k+r)L} F(z_k)\right);$$
$$z_{k+1} = P_C\left(\widetilde{z}_{k+1} - \frac{1}{L} F(z_{k+\frac{1}{2}})\right);$$
$$\widetilde{C}_{k+1} = L(\widetilde{z}_{k+1} - z_{k+1}) - F(z_{k+\frac{1}{2}});$$
$$u_{k+1} = u_k - \frac{D}{r}[F(z_{k+1}) + \widetilde{C}_{k+1}].$$

Thus, if $D < \frac{r-1}{L}$, the last-iterate convergence rates of Algorithm 2 are

$$\mathrm{dist}(0, T(z_k))^2 \leqslant \frac{r^2(r-1)^2}{\left[\frac{r-1}{L} D - D^2\right] k^2} \cdot \mathrm{dist}(z_0, \mathrm{zero}(T))^2, \quad \forall k \geqslant 1$$

$$\left\langle F(z_k) + \widetilde{C}_k, z_k - z^* \right\rangle \leqslant \frac{r^2 - r}{2Dk} \|z_0 - z^*\|^2, \quad \forall z^* \in \mathrm{zero}(T)$$

In addition, we have

$$\sum_{k=0}^{\infty} \frac{1}{2}\left(\frac{r-1}{L} - D\right)(2k + r + 1) \left\|F(z_{k+1}) + \widetilde{C}_{k+1}\right\|^2 \leqslant \frac{r^3 - r^2}{2} \mathrm{dist}(z_0, \mathrm{zero}(T))^2$$
$$\sum_{k=0}^{\infty} \frac{(k+r)^2}{2L} \left\|\widetilde{C}_{k+\frac{1}{2}} - \frac{k}{k+r}\widetilde{C}_k\right\|^2 \leqslant \frac{r^3 - r^2}{2D} \mathrm{dist}(z_0, \mathrm{zero}(T))^2$$

Proof. Similar to the proof for Theorem 1 , we need to obtain some useful equalities and inequalities similar to (12)-(14). Here we let $\widetilde{C}_{k+\frac{1}{2}} = L\left(\widetilde{z}_{k+1} - z_{k+\frac{1}{2}} - \frac{k}{(k+r)L} F(z_k)\right)$. By equivalent characterization of projection operator, we have

$$\widetilde{C}_{k+1} \in N_C(z_{k+1}), \quad \widetilde{C}_{k+\frac{1}{2}} \in N_C(z_{k+\frac{1}{2}})$$

Also by the definition of $\widetilde{C}_{k+1}$ and $\widetilde{C}_{k+\frac{1}{2}}$, we have

$$z_{k+1} = \widetilde{z}_{k+1} - \frac{1}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{C}_{k+1}] \tag{21}$$
$$z_{k+\frac{1}{2}} = \widetilde{z}_{k+1} - \frac{k}{(k+r)L} F(z_k) - \frac{1}{L}\widetilde{C}_{k+\frac{1}{2}} \tag{22}$$

Using (21) and (22), we obtain the following useful equations

$$k(z_{k+1} - z_k) = r(u_k - z_{k+1}) - \frac{k+r}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{C}_{k+1}] \tag{23}$$

$$r(z_k - u_k) = -(k+r)(z_{k+1} - z_k) - \frac{k+r}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{C}_{k+1}]. \tag{24}$$

Also. by $L-$Lipschitz continuity of F, we have

$$\left\| F(z_{k+1}) - F(z_{k+\frac{1}{2}}) \right\|^2 \leqslant L^2 \left\| z_{k+1} - z_{k+\frac{1}{2}} \right\|^2$$

Since $\widetilde{C}_{k+\frac{1}{2}} \in N_C(z_{k+\frac{1}{2}}), z_{k+1} \in C$, we have

$$\left\langle \widetilde{C}_{k+\frac{1}{2}}, z_{k+\frac{1}{2}} - z_{k+1} \right\rangle \geqslant 0$$

Thus

$$\begin{aligned} \left\| z_{k+1} - z_{k+\frac{1}{2}} \right\|^2 &\leqslant \left\| z_{k+1} - z_{k+\frac{1}{2}} \right\|^2 + \frac{2}{L} \left\langle \widetilde{C}_{k+\frac{1}{2}}, z_{k+\frac{1}{2}} - z_{k+1} \right\rangle \\ &= \left\| z_{k+1} - z_{k+\frac{1}{2}} - \frac{1}{L} \widetilde{C}_{k+\frac{1}{2}} \right\|^2 - \frac{1}{L^2} \left\| \widetilde{C}_{k+\frac{1}{2}} \right\|^2 \end{aligned}$$

By subtracting (21) by (22), we obtain

$$\left\| F(z_{k+1}) - F(z_{k+\frac{1}{2}}) \right\|^2 \leqslant \left\| \frac{k}{k+r} F(z_k) - F(z_{k+\frac{1}{2}}) - \widetilde{C}_{k+1} \right\|^2 - \left\| \widetilde{C}_{k+\frac{1}{2}} \right\| \tag{25}$$

Next, we consider the following Lyapunov function

$$\mathcal{E}(k) = \frac{Dk^2}{2L} \left\| F(z_k) + \widetilde{C}_k \right\|^2 + Drk \left\langle F(z_k) + \widetilde{C}_k, z_k - u_k \right\rangle + \frac{r^3 - r^2}{2} \left\| u_k - z^* \right\|^2, \tag{26}$$

where $z^* \in \text{zero}(T)$. First, we divide the difference $\mathcal{E}(k+1) - \mathcal{E}(k)$ into three parts.

$$\begin{aligned} &\mathcal{E}(k+1) - \mathcal{E}(k) \\ = &\underbrace{\frac{D(k+1)^2}{2L} \left\| F(z_{k+1}) + \widetilde{C}_{k+1} \right\|^2 - \frac{Dk^2}{2L} \left\| F(z_k) + \widetilde{C}_k \right\|^2}_{\text{I}} \\ &+ \underbrace{Dr(k+1) \left\langle F(z_{k+1}) + \widetilde{C}_{k+1}, z_{k+1} - u_{k+1} \right\rangle - Drk \left\langle F(z_k) + \widetilde{C}_k, z_k - u_k \right\rangle}_{\text{II}} \\ &+ \underbrace{\frac{r^3 - r^2}{2} \left[\left\| u_{k+1} - z^* \right\|^2 - \left\| u_k - z^* \right\|^2 \right]}_{\text{III}} \end{aligned}$$

Secondly, we reckon the upper bound of II and III. First we consider II. II can be split into the following three terms.

$$\begin{aligned} \text{II} = Dr &\left\langle F(z_{k+1}) + \widetilde{C}_{k+1}, z_{k+1} - u_{k+1} \right\rangle \\ &+ Drk \left\langle F(z_{k+1}) + \widetilde{C}_{k+1}, z_{k+1} - z_k - u_{k+1} + u_k \right\rangle \\ &+ Drk \left\langle F(z_{k+1}) + \widetilde{C}_{k+1} - F(z_k) - \widetilde{C}_k, z_k - u_k \right\rangle \end{aligned}$$

Due to (23) and the equality $u_{k+1} - u_k = -\frac{D}{r}[F(z_{k+1}) + \widetilde{C}_{k+1}]$, we have

$$
\begin{aligned}
&Drk \left\langle F(z_{k+1}), z_{k+1} - z_k - u_{k+1} + u_k \right\rangle \\
= &Dr \left\langle F(z_{k+1}) + \widetilde{C}_{k+1}, r(u_k - z_{k+1}) - \frac{k+r}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{C}_{k+1}] \right\rangle \\
&+ D^2 k \left\| F(z_{k+1}) + \widetilde{C}_{k+1} \right\|^2
\end{aligned} \tag{27}
$$

In addition, because of (24), we have

$$
\begin{aligned}
&Drk \left\langle F(z_{k+1}) - F(z_k) + \widetilde{C}_{k+1} - \widetilde{C}_k, z_k - u_k \right\rangle \\
= &Dk \left\langle F(z_{k+1}) + \widetilde{C}_{k+1} - F(z_k) - \widetilde{C}_k, -(k+r)(z_{k+1} - z_k) - \frac{k+r}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{C}_{k+1}] \right\rangle
\end{aligned} \tag{28}
$$

Also,

$$
\begin{aligned}
&Dr \left\langle F(z_{k+1}) + \widetilde{C}_{k+1}, z_{k+1} - u_{k+1} \right\rangle \\
= &D^2 \left\| F(z_{k+1}) + \widetilde{C}_{k+1} \right\|^2 + Dr \left\langle F(z_{k+1}) + \widetilde{C}_{k+1}, z_{k+1} - u_k \right\rangle
\end{aligned} \tag{29}
$$

By summing (27), (28) and (29), we have

$$
\begin{aligned}
&\mathrm{II} \\
= &-\frac{D(k+r)^2}{L} \left\langle F(z_{k+1}), F(z_{k+\frac{1}{2}}) \right\rangle + \frac{Dk(k+r)}{L} \left\langle F(z_k), F(z_{k+\frac{1}{2}}) \right\rangle - \frac{D(k+r)^2}{L} \left\| \widetilde{C}_{k+1} \right\|^2 \\
&- \frac{D}{L} \left\langle \widetilde{C}_{k+1}, (k+r)^2 [F(z_{k+\frac{1}{2}}) + F(z_{k+1})] - k(k+r)[F(z_k) + \widetilde{C}_k] \right\rangle + \frac{Dk(k+r)}{L} \left\langle \widetilde{C}_k, F(z_{k+\frac{1}{2}}) \right\rangle \\
&- Dk(k+r) \left\langle F(z_{k+1}) + \widetilde{C}_{k+1} - F(z_k) - \widetilde{C}_k, z_{k+1} - z_k \right\rangle \\
&+ D(r - r^2) \left\langle F(z_{k+1}) + \widetilde{C}_{k+1}, z_{k+1} - u_k \right\rangle + D^2(k+1) \left\| F(z_{k+1}) + \widetilde{C}_{k+1} \right\|^2
\end{aligned}
$$

Combining $\rho-$comonotonicity of F, (25) and the following equality

$$
-\left\langle F(z_{k+1}), F(z_{k+\frac{1}{2}}) \right\rangle = \frac{1}{2} \left[\left\| F(z_{k+1}) - F(z_{k+\frac{1}{2}}) \right\|^2 - \| F(z_{k+1}) \|^2 - \left\| F(z_{k+\frac{1}{2}}) \right\|^2 \right]
$$

we can obtain the upper bound of II as follows:

$$
\begin{aligned}
\mathrm{II} \leqslant &-\frac{D(k+r)^2}{2L} \left\| F(z_{k+1}) + \widetilde{C}_{k+1} \right\|^2 + Dk(k+r) \left\langle \widetilde{C}_k, z_{k+1} - z_k + \frac{1}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{C}_{k+1}] \right\rangle \\
&- Dk(k+r) \left\langle F(z_{k+1}) + \widetilde{C}_{k+1} - F(z_k), z_{k+1} - z_k \right\rangle - \frac{D(k+r)^2}{2L} \left\| \widetilde{C}_{k+\frac{1}{2}} \right\| \\
&+ D(r - r^2) \left\langle F(z_{k+1}) + \widetilde{C}_{k+1}, z_{k+1} - u_k \right\rangle + D^2(k+1) \left\| F(z_{k+1}) + \widetilde{C}_{k+1} \right\|^2 + \frac{Dk^2}{2L} \| F(z_k) \|^2
\end{aligned}
$$

Since

$$
z_{k+1} + \frac{1}{L}[F(z_{k+\frac{1}{2}}) + \widetilde{C}_{k+1}] = z_{k+\frac{1}{2}} + \frac{k}{(k+r)L} F(z_k) + \frac{1}{L} \widetilde{C}_{k+1},
$$

we have

$$
\begin{aligned}
&Dk(k+r)\left\langle \widetilde{C}_k, z_{k+1}-z_k+\frac{1}{L}[F(z_{k+\frac{1}{2}})+\widetilde{C}_{k+1}]\right\rangle \\
=\ &Dk(k+r)\left\langle \widetilde{C}_k, z_{k+\frac{1}{2}}-z_k+\frac{1}{L}\left(\frac{k}{k+r}F(z_k)+\widetilde{C}_{k+\frac{1}{2}}\right)\right\rangle
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mathrm{II} \leqslant\ & -\frac{D(k+r)^2}{2L}\left\|F(z_{k+1})+\widetilde{C}_{k+1}\right\|^2+\frac{Dk^2}{2L}\left\|F(z_k)+\widetilde{C}_k\right\|^2-\frac{D}{2L}\left\|(k+r)\widetilde{C}_{k+\frac{1}{2}}-k\widetilde{C}_k\right\|^2 \\
& -Dk(k+r)\left\langle F(z_{k+1})+\widetilde{C}_{k+1}-F(z_k), z_{k+1}-z_k\right\rangle+Dk(k+r)\left\langle \widetilde{C}_k, z_{k+\frac{1}{2}}-z_k\right\rangle \\
& +D(r-r^2)\left\langle F(z_{k+1})+\widetilde{C}_{k+1}, z_{k+1}-u_k\right\rangle+D^2(k+1)\left\|F(z_{k+1})+\widetilde{C}_{k+1}\right\|^2
\end{aligned}
$$

Now we estimate III. Owing to (18), we have

$$
\begin{aligned}
\mathrm{III} &= (r^3-r^2)\langle u_{k+1}-u_k, u_k-z^*\rangle+\frac{r^3-r^2}{2}\|u_{k+1}-u_k\|^2 \\
&= \underbrace{-D(r^2-r)\left\langle F(z_{k+1})+\widetilde{C}_{k+1}, u_k-z^*\right\rangle}_{\mathrm{III}_1}+\frac{D^2(r-1)}{2}\left\|F(z_{k+1})+\widetilde{C}_{k+1}\right\|^2
\end{aligned}
$$

Next, we can deduce the upper bound of $\mathcal{E}(k+1)-\mathcal{E}(k)$. Due to

$$
\mathrm{II}_1-\mathrm{III}_1=-D(r^2-r)\left\langle F(z_{k+1})+\widetilde{C}_{k+1}, z_{k+1}-z^*\right\rangle,
$$

the upper bound of $\mathcal{E}(k+1)-\mathcal{E}(k)$ can be simplified as follows:

$$
\begin{aligned}
& \mathcal{E}(k+1)-\mathcal{E}(k) \\
\leqslant\ & -\frac{D}{2}\left(\frac{r-1}{L}-D\right)(2k+r+1)\left\|F(z_{k+1})+\widetilde{C}_{k+1}\right\|^2 \\
& -D(r^2-r)\left\langle F(z_{k+1})+\widetilde{C}_{k+1}, z_{k+1}-z^*\right\rangle-Dk(k+r)\left\langle F(z_{k+1})+\widetilde{C}_{k+1}-F(z_k), z_{k+1}-z_k\right\rangle \\
& +Dk(k+r)\left\langle \widetilde{C}_k, z_{k+\frac{1}{2}}-z_k\right\rangle-\frac{D}{2L}\left\|(k+r)\widetilde{C}_{k+\frac{1}{2}}-k\widetilde{C}_k\right\|^2
\end{aligned}
$$

Because $\widetilde{C}_{k+1}\in N_C(z_{k+1})$, $\widetilde{C}_k\in N_C(z_k)$, we have

$$
\left\langle \widetilde{C}_{k+1}, z_k-z_{k+1}\right\rangle \leqslant 0, \quad \left\langle \widetilde{C}_k, z_{k+\frac{1}{2}}-z_k\right\rangle \leqslant 0
$$

In conclusion

$$
\begin{aligned}
& \mathcal{E}(k+1)-\mathcal{E}(k) \\
\leqslant\ & -\frac{D}{2}\left(\frac{r-1}{L}-D\right)(2k+r+1)\left\|F(z_{k+1})+\widetilde{C}_{k+1}\right\|^2-\frac{D}{2L}\left\|(k+r)\widetilde{C}_{k+\frac{1}{2}}-k\widetilde{C}_k\right\|^2 \\
\leqslant\ & 0
\end{aligned}
$$

Finally, we can obtain the convergence results of Algorithm 2. First, we translate $\mathcal{E}(k)$ into the following form.

$$\begin{aligned}\mathcal{E}(k) = & \left(\frac{D}{2L} - \frac{D^2}{2(r-1)}\right) k^2 \left\|F(z_k) + \widetilde{C}_k\right\|^2 + Drk \left\langle F(z_k) + \widetilde{C}_k, z_k - z^*\right\rangle \\ & + \frac{1}{2}\left\|\frac{Dk}{\sqrt{r-1}}[F(z_k) + \widetilde{C}_k] - \sqrt{r^3 - r^2}(u_k - z^*)\right\|^2\end{aligned}$$

Since F is monotone, $\mathcal{E}(k)$ can be represented by the sum of three non-negative terms. Then we have

$$\left(\frac{D}{2L} - \frac{D^2}{2(r-1)}\right) k^2 \mathrm{dist}(0, T(z_k))^2 \leqslant \mathcal{E}(k) \leqslant \mathcal{E}(0) = \frac{r^3 - r^2}{2}\|z_0 - z^*\|^2 .$$
$$Drk \left\langle F(z_k) + \widetilde{C}_k, z_k - z^*\right\rangle \leqslant \mathcal{E}(k) \leqslant \mathcal{E}(0) = \frac{r^3 - r^2}{2}\|z_0 - z^*\|^2$$

Taking infimum respect to all $z^* \in \mathrm{zero}(T)$ on both sides of the first above inequality and using the inequality $\mathrm{dist}(0, T(z_k)) \leqslant \left\|F(z_k) + \widetilde{C}_k\right\|$, we obtain the desire result. Moreover, by summing $\mathcal{E}(k+1) - \mathcal{E}(k)$ from 0 to ∞, we have

$$\sum_{k=0}^{\infty} \frac{D}{2}\left(\frac{r-1}{L} - D\right)(2k + r + 1)\left\|F(z_{k+1}) + \widetilde{C}_{k+1}\right\|^2 \leqslant \frac{r^3 - r^2}{2}\|z_0 - z^*\|^2$$
$$\sum_{k=0}^{\infty} \frac{(k+r)^2}{2L}\left\|\widetilde{C}_{k+\frac{1}{2}} - \frac{k}{k+r}\widetilde{C}_k\right\|^2 \leqslant \frac{r^3 - r^2}{2}\|z_0 - z^*\|^2$$

Also, taking infimum respect to all $z^* \in \mathrm{zero}(T)$, we obtain the desire results. □

4 Sharper Rate

In this section, we will build up the $o(1/k^2)$ convergence rate of the SEG type method. It is noteworthy that the symplectic proximal point algorithm admits an $o(1/k^2)$ convergence rate if the assumption $D < r - 1$ holds. Analogously, one may suggest that the symplectic EG type method admits $o(1/k^2)$ convergence rate when $D < (r-1)\left(\frac{1}{L} + 2\rho\right)$. However, the assumption $D < (r-1)\left(\frac{1}{L} + 2\rho\right)$ alone is not enough. The main reason is that the EG type method tries to use $F(z_{k+\frac{1}{2}})$ to approximate $F(z_{k+1})$, but the error term $\left\|F(z_{k+\frac{1}{2}}) - F(z_{k+1})\right\|^2$ can not be estimated directly. To overcome this problem, we need additional assumption on the step-size in the SEG type method. Here, we make use of Algorithm 1 to illustrate the desire result.

Lemma 3 (Estimation of error). *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_{k+1}\}$ and $\{u_k\}$ be the sequences generated by*

$$\widetilde{z}_{k+1} = \frac{k}{k+r}z_k + \frac{r}{k+r}u_k \tag{30}$$

$$z_{k+\frac{1}{2}} = \widetilde{z}_{k+1} - \frac{k}{k+r}(s + 2\rho)(F(z_k) + \widetilde{G}_k) \tag{31}$$

$$z_{k+1} = J_{\frac{1}{s}G}\left[\widetilde{z}_{k+1} - sF(z_{k+\frac{1}{2}}) - \frac{k}{k+r}2\rho(F(z_k) + \widetilde{G}_k)\right] \tag{32}$$

$$\widetilde{G}_{k+1} = s^{-1}\left[\widetilde{z}_{k+1} - z_{k+1} - sF(z_{k+\frac{1}{2}}) - \frac{k}{k+r}2\rho(F(z_k) + \widetilde{G}_k)\right] \tag{33}$$

$$u_{k+1} = u_k - \frac{D}{r}[F(z_{k+1}) + \widetilde{G}_{k+1}]. \tag{34}$$

If F is $L-$Lipschitz continuous and $s < \dfrac{1}{L}$, there exists a positive constant C such that

$$\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2 \leqslant C\left\|F(z_{k+1}) + \widetilde{G}_{k+1} - F(z_k) - \widetilde{G}_k\right\|^2$$

Proof. First, we show that for all $C_1 > 1$, there exists a constant $C_2 > 1$ such that

$$\|x+y\|^2 \leqslant C_2\|x\|^2 + C_1\|y\|^2, \quad \forall x, y \in \mathcal{H} \tag{35}$$

By using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\|x+y\|^2 &= \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle \\ &= \|x\|^2 + \|y\|^2 + 2\left\langle (C_1-1)^{-\frac{1}{2}}x, \sqrt{C_1-1}y\right\rangle \\ &\leqslant \frac{C_1}{C_1-1}\|x\|^2 + C_1\|y\|^2, \quad \forall C_1 > 1\end{aligned}$$

Let $C_2 = \dfrac{C_1}{C_1-1}$, we obtain the desire result. Let $\widetilde{T}(z_k) = F(z_k) + \widetilde{G}_k$. Due to $L-$Lipschitz continuity of F, we have

$$\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2 \leqslant L^2s^2\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}}) - \widetilde{T}(z_{k+1}) + \widetilde{T}(z_k)\right\|^2$$

Let $y = F(z_{k+1}) - F(z_{k+\frac{1}{2}})$ and $x = -\widetilde{T}(z_{k+1}) + \widetilde{T}(z_k)$. Since $s < \dfrac{1}{L}$, there exists a constant C_1 such that $C_1 > 1$ and $L^2s^2C_1 < 1$. By applying (35), we have

$$\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2 \leqslant L^2s^2C_2\left\|\widetilde{T}(z_{k+1}) - \widetilde{T}(z_k)\right\|^2 + L^2s^2C_1\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2$$

Let $C = \dfrac{L^2s^2C_2}{1 - L^2s^2C_1}$, we have

$$\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2 \leqslant C\left\|\widetilde{T}(z_{k+1}) - \widetilde{T}(z_k)\right\|^2$$

□

With Lemma 3, we can demonstrate the $o(1/k^2)$ convergence rate of the SFBS method.

Theorem 4 (Faster convergence rate)**.** *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_{k+1}\}$ and $\{u_k\}$ be the sequences generated by (30)-(34). If F is $L-$Lipschitz continuous, G is maximally monotone, $T = F+G$ is $\rho-$comonotone, $0 < s < \dfrac{1}{L}$, $\rho > -\dfrac{s}{2}$ and $0 < D < (r-1)(s+2\rho)$, we have*

$$\lim_{k\to\infty}\|z_k - u_k\|^2 = 0, \qquad \lim_{k\to\infty}\|u_k - z^*\|^2 \text{ exists for all } z^* \in \mathrm{zero}(T)$$

$$\lim_{k\to\infty} k^2\mathrm{dist}(0, T(z_k))^2 = 0, \qquad \lim_{k\to\infty} k^2\|z_{k+1} - z_k\|^2 = 0$$

Proof. Let $\widetilde{T}(z_k) = F(z_k) + \widetilde{G}_k$. Owing to $s < \frac{1}{L}$, the operator F can be regarded as $s^{-1}-$Lipschitz continuous. By Theorem 1, we have

$$\left\|\widetilde{T}(z_k)\right\|^2 \leqslant O\left(\frac{1}{k^2}\right), \quad \sum_{k=0}^{\infty} k\left\|\widetilde{T}(z_k)\right\|^2 < +\infty$$

Due to Lemma 3, we have

$$\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2 \leqslant O\left(\frac{1}{k^2}\right), \quad \sum_{k=0}^{\infty} k\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2 < +\infty$$

Next, we show that

$$\sum_{k=1}^{\infty} \frac{\|z_{k+1} - u_{k+1}\|^2}{2k} < +\infty$$

To obtain such result, we need to study the following auxiliary Lyapunov function:

$$\mathcal{G}(k) = \frac{1}{2}\|z_k - u_k\|^2 + \frac{r}{2D} \cdot \frac{(k+r)^2}{k^2}\left(s + 2\rho - \frac{D}{r}\right)\|u_k - z^*\|^2$$

First, consider the difference $\frac{1}{2}\|z_{k+1} - u_{k+1}\|^2 - \frac{1}{2}\|z_k - u_k\|^2$.

$$\begin{aligned}
&\frac{1}{2}\|z_{k+1} - u_{k+1}\|^2 - \frac{1}{2}\|z_k - u_k\|^2 \\
=\ &\frac{1}{2}\langle z_{k+1} + z_k - u_{k+1} - u_k, z_{k+1} - z_k - u_{k+1} + u_k\rangle \\
=\ &\frac{1}{2}\left\langle \left(2 + \frac{r}{k}\right) z_{k+1} - u_{k+1} - \left(1 + \frac{r}{k}\right) u_k + \frac{k+r}{k} s[F(z_{k+\frac{1}{2}}) + \widetilde{G}_{k+1}] + 2\rho\widetilde{T}(z_k), \right. \\
&\qquad \left. -\frac{r}{k} z_{k+1} + \left(1 + \frac{r}{k}\right) u_k - u_{k+1} - \frac{k+r}{k} s[F(z_{k+\frac{1}{2}}) + \widetilde{G}_{k+1}] - 2\rho\widetilde{T}(z_k)\right\rangle \\
=\ &-\frac{(2k+r)r}{2k^2}\|z_{k+1} - u_{k+1}\|^2 - \frac{(k+r)^2}{k^2} s\left\langle F(z_{k+1}) - F(z_{k+\frac{1}{2}}), z_{k+1} - u_{k+1}\right\rangle \\
&- \frac{k+r}{k} 2\rho s\left\langle F(z_{k+\frac{1}{2}}) - F(z_{k+1}), \widetilde{T}(z_k) - \frac{k+r}{k}\widetilde{T}(z_{k+1})\right\rangle \\
&- \frac{(k+r)^2}{k^2}\left(s + 2\rho - \frac{D}{r}\right)\left\langle \widetilde{T}(z_{k+1}), z_{k+1} - u_{k+1}\right\rangle \\
&- \frac{k+r}{k}\left(s + 2\rho - \frac{D}{r}\right) 2\rho s\left\langle \widetilde{T}(z_k) - \frac{k+r}{k}\widetilde{T}(z_{k+1}), \widetilde{T}(z_{k+1})\right\rangle \\
\leqslant\ &-\frac{(2k+r)(r-1)}{2k^2}\|z_{k+1} - u_{k+1}\|^2 + \frac{(k+r)^4 s^2}{2k^2(2k+r)}\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2 \\
&- \frac{(k+r)^2}{k^2}\left(s + 2\rho - \frac{D}{r}\right)\left\langle \widetilde{T}(z_{k+1}), z_{k+1} - u_{k+1}\right\rangle \\
&+ \frac{k+r}{k}\rho s\left(\left\|F(z_{k+\frac{1}{2}}) - F(z_{k+1})\right\|^2 + \left\|\widetilde{T}(z_k) - \frac{k+r}{k}\widetilde{T}(z_{k+1})\right\|^2\right)
\end{aligned}$$

$$+\frac{k+r}{k}\left(s+2\rho-\frac{D}{r}\right)\rho s\left(\left\|\widetilde{T}(z_k)-\frac{k+r}{k}\widetilde{T}(z_{k+1})\right\|^2+\left\|\widetilde{T}(z_{k+1})\right\|^2\right)$$

Also, since $\frac{(k+r)^2}{k^2}$ is non-increasing, we have

$$\begin{aligned}
&\frac{r}{2D}\cdot\frac{(k+1+r)^2}{(k+1)^2}\left(s+2\rho-\frac{D}{r}\right)\|u_{k+1}-z^*\|^2-\frac{r}{2D}\cdot\frac{(k+r)^2}{k^2}\left(s+2\rho-\frac{D}{r}\right)\|u_k-z^*\|^2\\
\leqslant&\frac{r}{2D}\cdot\frac{(k+r)^2}{k^2}\left(s+2\rho-\frac{D}{r}\right)\left(\|u_{k+1}-z^*\|^2-\|u_k-z^*\|^2\right)\\
=&\frac{r}{2D}\cdot\frac{(k+r)^2}{k^2}\left(s+2\rho-\frac{D}{r}\right)\left(2\left\langle u_{k+1}-u_k,u_{k+1}-z^*\right\rangle-\|u_{k+1}-u_k\|^2\right)\\
=&\frac{r}{2D}\cdot\frac{(k+r)^2}{k^2}\left(s+2\rho-\frac{D}{r}\right)\left(-\frac{2D}{r}\left\langle\widetilde{T}(z_{k+1}),u_{k+1}-z^*\right\rangle-\frac{D^2}{r^2}\left\|\widetilde{T}(z_{k+1})\right\|^2\right)
\end{aligned}$$

Let

$$\begin{aligned}
\varphi(k)=&\frac{(k+r)^4s^2}{2k^2(2k+r)}\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2\\
&-\frac{(k+r)^2}{k^2}\left(s+2\rho-\frac{D}{r}\right)\left(\rho+\frac{D}{2r}\right)\left\|\widetilde{T}(z_{k+1})\right\|^2\\
&+\frac{k+r}{k}\rho s\left(\left\|F(z_{k+\frac{1}{2}})-F(z_{k+1})\right\|^2+\left\|\widetilde{T}(z_k)-\frac{k+r}{k}\widetilde{T}(z_{k+1})\right\|^2\right)\\
&+\frac{k+r}{k}\left(s+2\rho-\frac{D}{r}\right)\rho s\left(\left\|\widetilde{T}(z_k)-\frac{k+r}{k}\widetilde{T}(z_{k+1})\right\|^2+\left\|\widetilde{T}(z_{k+1})\right\|^2\right)
\end{aligned}$$

By previous argument, we have

$$\mathcal{G}(k+1)-\mathcal{G}(k)\leqslant-\frac{(2k+r)(r-1)}{2k^2}\|z_{k+1}-u_{k+1}\|^2+\varphi(k)$$

Since $\sum\limits_{k=1}^{\infty}\varphi(k)<+\infty$, we have

$$\mathcal{G}(k+1)+\sum_{i=k+1}^{\infty}\varphi(i)\leqslant\mathcal{G}(k)+\sum_{i=k}^{\infty}\varphi(k)-\frac{(2k+r)(r-1)}{2k^2}\|z_{k+1}-u_{k+1}\|^2$$

Then we have

$$\lim_{k\to\infty}\mathcal{G}(k)\text{ exists},\qquad\sum_{k=1}^{\infty}\frac{\|z_k-u_k\|^2}{k}\leqslant+\infty$$

The next thing we need to do is to show that for all $z^*\in\text{zero}(T)$, $\lim\limits_{k\to\infty}\|u_k-z^*\|^2$ exists. If $\lim\limits_{k\to\infty}\|u_k-z^*\|^2$ exists, $\lim\limits_{k\to\infty}\mathcal{G}(k)$ exists and $\sum\limits_{k=1}^{\infty}\frac{\|z_{k+1}-u_{k+1}\|^2}{2k}<+\infty$, we can easily show that

$$\lim_{k\to\infty}\|z_k-u_k\|^2=0$$

Calculating the difference of $\frac{1}{2}\|u_k - z^*\|^2$, we have

$$\begin{aligned}
&\frac{1}{2}\|u_{k+1} - z^*\|^2 - \frac{1}{2}\|u_k - z^*\|^2 \\
\leqslant\ & -\frac{D}{r}\left\langle \widetilde{T}(z_{k+1}), u_{k+1} - z^* \right\rangle \\
\leqslant\ & \frac{D}{2r}\left(k\left\|\widetilde{T}(z_{k+1})\right\|^2 + k^{-1}\|u_{k+1} - z_{k+1}\|^2\right) - \frac{D\rho}{r}\left\|\widetilde{T}(z_{k+1})\right\|^2
\end{aligned}$$

Since

$$\sum_{k=1}^{\infty}\frac{D}{2r}\left(k\|T_{k+1}\|^2 + k^{-1}\|u_{k+1} - z_{k+1}\|^2\right) - \frac{D\rho}{r}\|T_{k+1}\|^2 < +\infty$$

we prove the existence of the limitation $\lim\limits_{k\to\infty}\|u_k - z^*\|^2$.

Finally, we can prove the results presented in Theorem 1. Since

$$\left\|\widetilde{T}(z_k)\right\|^2 \leqslant O\left(\frac{1}{k^2}\right), \quad \lim_{k\to\infty}\|z_k - u_k\|^2 = 0$$

we have

$$\lim_{k\to\infty} Drk\left\langle \widetilde{T}(z_k), z_k - u_k \right\rangle = 0$$

Consider the following Lyapunov function

$$\begin{aligned}
\mathcal{E}(k) = &\left[\frac{Dk^2 s}{2} + \rho Dk(k-r)\right]\left\|\widetilde{T}(z_k)\right\|^2 \\
&+ Drk\left\langle \widetilde{T}(z_k), z_k - u_k \right\rangle + \frac{r^3 - r^2}{2}\|u_k - z^*\|
\end{aligned}$$

By Theorem 4, we have $\{\mathcal{E}(k)\}$ is non-increasing and non-negetive. Thus, $\lim\limits_{k\to\infty}\mathcal{E}(k)$ exists. Also, by the existence of $\lim\limits_{k\to\infty}\|u_k - z^*\|^2$, we have

$$\lim_{k\to\infty}\left[\frac{Dk^2 s}{2} + \rho Dk(k-r)\right]\left\|\widetilde{T}(z_k)\right\|^2 \text{ exists}$$

Since $\sum\limits_{k=0}^{\infty} k\left\|\widetilde{T}(z_k)\right\|^2 < +\infty$, we have

$$\lim_{k\to\infty} k^2\left\|\widetilde{T}(z_k)\right\|^2 = 0$$

By the inequality $0 \leqslant \mathrm{dist}(0, T(z_k))^2 \leqslant \left\|\widetilde{T}(z_k)\right\|^2$, we can obtain

$$\lim_{k\to\infty} k^2 \mathrm{dist}(0, T(z_k))^2 = 0$$

Also, by the inequality

$$k(z_{k+1} - z_k) = r(u_{k+1} - z_{k+1}) - [(k+r)s - D]\widetilde{T}(z_{k+1}) - (k+r)s[F(z_{k+\frac{1}{2}}) - F(z_{k+1})] - 2\rho k\widetilde{T}(z_k)$$

we have

$$\lim_{k\to\infty} k^2\|z_{k+1} - z_k\|^2 = 0$$

□

Although Theorem 4 requires that the step-size $s \in \left(-2\rho, \frac{1}{L}\right)$, the existence of desired step-size s can be easily ensure if $\rho > -\frac{1}{2L}$.

Theorem 5 (Weak convergence property)**.** *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_{k+1}\}$ and $\{u_k\}$ be the sequences generated by* (30)-(34) *with $\rho = 0$. If F is $L-$Lipschitz continuous and monotone, G is maximally monotone, $0 < s < \frac{1}{L}$, and $0 < D < (r-1)s$, then both of the sequences $\{z_k\}$ and $\{u_k\}$ converge weakly to a point in* $\mathrm{zero}(T)$.

Proof. By Theorem 4, $\lim_{k\to\infty} \|u_k - z^*\|^2$ exists. Let u_∞ be any cluster point of $\{u_k\}$, and $\{u_{k_j}\}$ be the subsequence of $\{u_k\}$ such that $u_{k_j} \rightharpoonup u_\infty$. Since $\lim_{k\to\infty} \|z_k - u_k\|^2 = 0$, $z_{k_j} \rightharpoonup u_\infty$. As a result of Theorem 1, $F(z_k) + \widetilde{G}_k \to 0$. Because of Proposition 2, $0 \in T(u_\infty)$. By Proposition 1, we can show that $\{u_k\}$ converge weakly to a point in $\mathrm{zero}(T)$. Also by $\lim_{k\to\infty} \|u_k - z_k\|^2 = 0$, $\{z_k\}$ converge weakly to a point in $\mathrm{zero}(T)$ and the weak limitations of $\{z_k\}$ and $\{u_k\}$ are the same. □

However, the weak convergence property of the SFBS method requires that F is monotone because of Proposition 2. If one can prove a similar result to Proposition 2 under comonotonicity assumption, the weak convergence property of the SFBS method under comonotonicity assumption can be ensured.

The method to demonstrate the $o(1/k^2)$ convergence rate and the weak convergence property of the SPEG+ method is similar to the proof for Theorem 4 and Theorem 5.

Corollary 6. *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_{k+1}\}$ and $\{u_k\}$ be the sequences generated by*

$$
\begin{aligned}
\widetilde{z}_{k+1} &= \frac{k}{k+r} z_k + \frac{r}{k+r} u_k \\
z_{k+\frac{1}{2}} &= P_C\left(\widetilde{z}_{k+1} - \frac{ks}{(k+r)} F(z_k)\right) \\
z_{k+1} &= P_C\left(\widetilde{z}_{k+1} - sF(z_{k+\frac{1}{2}})\right) \\
\widetilde{C}_{k+1} &= s^{-1}\left(\widetilde{z}_{k+1} - sF(z_{k+\frac{1}{2}}) - z_{k+1}\right) \\
u_{k+1} &= u_k - \frac{D}{r}[F(z_{k+1}) + \widetilde{C}_{k+1}]
\end{aligned}
$$

If F is $L-$Lipschitz continuous and monotone, $0 < s < \frac{1}{L}$, $r > 1$ and $0 < D < (r-1)s$, then we have

$$
\begin{aligned}
&\lim_{k\to\infty} \|z_k - u_k\|^2 = 0 \\
&\lim_{k\to\infty} \|u_k - z^*\|^2 \textit{ exists for all } z^* \in \mathrm{zero}(T) \\
&\lim_{k\to\infty} k^2 \mathrm{dist}(0, T(z_k))^2 = 0 \\
&\lim_{k\to\infty} k \left\langle F(z_k) + \widetilde{C}_k, z_k - z^* \right\rangle = 0, \quad \forall z^* \in \mathrm{zero}(T) \\
&\lim_{k\to\infty} k^2 \|z_{k+1} - z_k\|^2 = 0
\end{aligned}
$$

Moreover, both of the sequences $\{z_k\}$ and $\{u_k\}$ converge weakly to a point in $\mathrm{zero}(T)$.

Proof. The main difficultly is to estimate the error term $\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|$. If

$$\sum_{k=0}^{\infty} k\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2<+\infty$$

is true, then by applying the proofing technique in the proof of Theorem 4, we can obtain the results in Corollary 6. Since F is $L-$Lipschitz continuous, we have

$$\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2 \leqslant L^2s^2\left\|F(z_{k+\frac{1}{2}})+\widetilde{C}_{k+1}-\frac{k}{k+r}F(z_k)-\widetilde{C}_{k+\frac{1}{2}}\right\|^2.$$

Since $s<\dfrac{1}{L}$, there exists a constant $C_1>1$ such that $L^2s^2C_1<1$. By (35), there exists a constant C_2 such that

$$\begin{aligned}
&\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2\\
\leqslant{}& L^2s^2C_1\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2\\
&+L^2s^2C_2\left\|F(z_{k+1})+\widetilde{C}_{k+1}-\frac{k}{k+r}F(z_k)-\widetilde{C}_{k+\frac{1}{2}}\right\|^2\\
\leqslant{}& L^2s^2C_1\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2+2L^2s^2C_2\left\|\frac{k}{k+r}\widetilde{C}_k-\widetilde{C}_{k+1}\right\|^2\\
&+2L^2s^2C_2\left\|F(z_{k+1})+\widetilde{C}_{k+1}-\frac{k}{k+r}F(z_k)-\frac{k}{k+r}\widetilde{C}_k\right\|^2
\end{aligned}$$

The proof of Theorem 2 shows that

$$\sum_{k=0}^{\infty}(k+r)^2\left\|\frac{k}{k+r}\widetilde{C}_k-\widetilde{C}_{k+\frac{1}{2}}\right\|^2<+\infty$$

In conclusion,

$$\sum_{k=0}^{\infty} k\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2<+\infty$$

holds. □

5 Line Search Framework

The theorems presented in Section 3 indicate that the SEG+ type method exhibits its maximum theoretical convergence speed when $D=(r-1)\left(\dfrac{1}{2L}+\rho\right)$. This prompts the question as to whether the SEG+ type method with $D=(r-1)\left(\dfrac{1}{2L}+\rho\right)$ indeed converges most rapidly in practical applications. We conduct a straightforward numerical experiment aimed at verifying this question. Here we consider the following min-max problem:

$$\min_x\max_y f(x,y)=-\frac{1}{6}x^2+\frac{2\sqrt{2}}{3}xy+\frac{1}{6}y^2$$

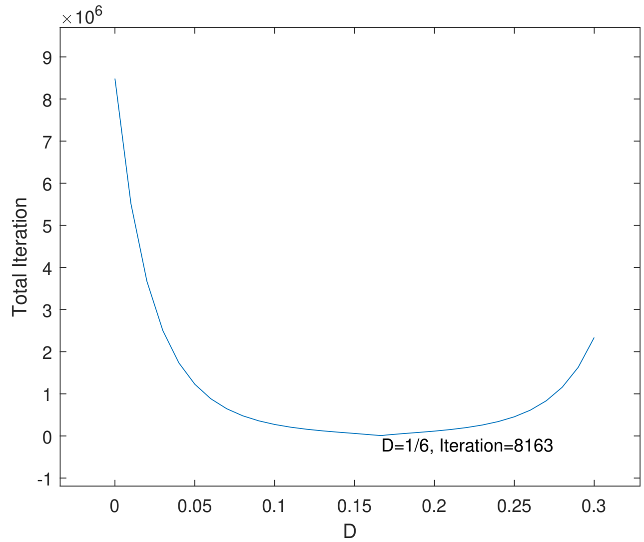

Figure 1. The required number of iterations Algorithm 1 with $G = 0$ respect to D. The termination condition is $\|F(x_k, y_k)\| \leqslant 10^{-6}$.

Such min-max problem is considered in section 4.4 in [LK21]. The first-order criteria of above min-max problem is

$$0 = F(x, y) = \begin{pmatrix} -\frac{1}{3}x + \frac{2\sqrt{2}}{3}y \\ -\frac{2\sqrt{2}}{3}x - \frac{1}{3}y \end{pmatrix} \tag{36}$$

It is easily to verify that

$$\begin{aligned} \left\|F(x, y) - F(x', y')\right\| &= \left\|(x, y) - (x', y')\right\|, \quad \forall (x, y),\ (x', y'), \\ \left\langle F(x, y) - F(x', y'), x - y \right\rangle &= -\frac{1}{3}\left\|F(x, y) - F(x', y')\right\|^2, \quad \forall (x, y),\ (x', y') \end{aligned}$$

We apply Algorithm 1 with parameter $r = 2$ and different parameters D to above zero-point problem until the termination condition is satisfied. The required number of iterations of Algorithm 1 respect to D is plotted in Figure 1.

As we can see from Figure 1, when $D = (r-1)\left(\frac{1}{2L} + \rho\right)$, i. e. $D = \frac{1}{6}$, the corresponding instance of Algorithm 1 requires least number of iterations. However, the Lipschitz constant L and the comonotone index ρ of the operator F described in (36) can be deduced in advance. In practice, the Lipschitz constant of most of the operators F and the comonotone index of most of the operators T can not be determined. In addition, even though we can determine the Lipschitz constant and the comonotone index in advance, the operator F may have smaller local Lipschitz constant or the operator T may have larger local comonotone index as z_k approaching zero(T). Algorithm 1 can not catch these local information. To overcome these problems, we apply the line search framework to the SFBS method to estimate these local information.

Since there is a term $\left\langle F(z_k) + \widetilde{G}_k, z_k - u_k \right\rangle$ in the Lyapunov function, when we consider the difference of $\left\langle F(z_k) + \widetilde{G}_k, z_k - u_k \right\rangle$, $\left\langle F(z_{k+1}) + \widetilde{G}_{k+1} - F(z_k) - \widetilde{G}_k, z_{k+1} - z_k \right\rangle$ appears. This is why we need to assume comonotonicity for the operator $T = F + G$. Thus we propose the termination

Algorithm 3: Line Search Framework for SEG+

Input: Operator F and Operator G;

Initialize: $z_0, u_0 = z_0$;

for $k = 0, 1, \cdots$ **do**

  repeat

    Choosing α_k, L_k, ρ_k with $\rho_k > -\frac{1}{2L_k}$;

    $\widetilde{z}_{k+1} = \alpha_k u_k + (1-\alpha_k) z_k$;

    $z_{k+\frac{1}{2}} = \widetilde{z}_{k+1} - (1-\alpha_k)\left(\frac{1}{L_k} + 2\rho_k\right)(F(z_k) + \widetilde{G}_k)$;

    $z_{k+1} = J_{L_k G}\left(\widetilde{z}_{k+1} - \frac{1}{L_k} F(z_{k+\frac{1}{2}}) - (1-\alpha_k) 2\rho_k (F(z_k) + \widetilde{G}_k)\right)$;

    $\widetilde{G}_{k+1} = L_k \left[\widetilde{z}_{k+1} - z_{k+1} - \frac{1}{L_k} F(z_{k+\frac{1}{2}}) - (1-\alpha_k) 2\rho_k (F(z_k) + \widetilde{G}_k)\right]$;

  until

  $\left\langle F(z_{k+1}) + \widetilde{G}_{k+1} - F(z_k) - \widetilde{G}_k, z_{k+1} - z_k \right\rangle \geqslant \rho_k \left\| F(z_{k+1}) + \widetilde{G}_{k+1} - F(z_k) - \widetilde{G}_k \right\|^2$,

  $\left\| F(z_{k+1}) - F(z_{k+\frac{1}{2}}) \right\| \leqslant L_k \left\| z_{k+1} - z_{k+\frac{1}{2}} \right\|$;

  $u_{k+1} = u_k - \frac{D}{r}\left(\frac{1}{2L_k} + \rho_k\right)(F(z_{k+1}) + \widetilde{G}_{k+1})$.

condition for ρ_k is

$$\rho_k \leqslant \frac{\left\langle F(z_{k+1}) + \widetilde{G}_{k+1} - F(z_k) + \widetilde{G}_k, z_{k+1} - z_k \right\rangle}{\left\| F(z_{k+1}) + \widetilde{G}_{k+1} - F(z_k) - \widetilde{G}_k \right\|^2}$$

Also, the idea that $F(z_{k+\frac{1}{2}})$ can be seen as an approximation of $F(z_{k+1})$ inspires us that the Lipschitz continuity on F helps us give an estimation of the error $\left\| F(z_{k+\frac{1}{2}}) - F(z_{k+1}) \right\|$. Because the parameter L_k serves as an estimation of local Lipschitz constant, the termination condition for L_k should be

$$L_k \geqslant \frac{\left\| F(z_{k+\frac{1}{2}}) - F(z_{k+1}) \right\|}{\left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|}$$

By summarizing previous argument, we obtain Algorithm 3. The convergence results of Algorithm 3 is given in the following theorems.

Theorem 7. *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_{k+1}\}$ and $\{u_k\}$ be the sequences generated by Algorithm 3, and let $\{\mathcal{E}(k)\}$ be the Lyapunov function such that*

$$\mathcal{E}(k) = D\left[\left(\sum_{i=0}^{k-1} \frac{1}{2L_i} + \rho_i\right)^2 - r\rho_{k-1}\left(\sum_{i=0}^{k-1} \frac{1}{2L_i} + \rho_i\right)\right]\left\| F(z_k) + \widetilde{G}_k \right\|^2$$
$$+ Dr\left(\sum_{i=0}^{k-1} \frac{1}{2L_i} + \rho_i\right)\left\langle F(z_k) + \widetilde{G}_k, z_k - u_k \right\rangle + \frac{r^3 - r^2}{2}\left\| u_k - z^* \right\|^2,$$

where $z^* \in \mathcal{H}$*. If* $r > 1$*,* $0 < D < 2(r-1)$*,*

$$\alpha_k = \frac{r\left(\dfrac{1}{2L_k}+\rho_k\right)}{\left(\sum_{i=0}^{k-1}\dfrac{1}{2L_i}+\rho_i\right)+r\left(\dfrac{1}{2L_k}+\rho_k\right)}$$

$$\left\langle F(z_{k+1})+\widetilde{G}_{k+1}-F(z_k)-\widetilde{G}_k, z_{k+1}-z_k\right\rangle \geqslant \rho_k \left\|F(z_{k+1})+\widetilde{G}_{k+1}-F(z_k)-\widetilde{G}_k\right\|^2$$

$$\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\| \leqslant L_k \left\|z_{k+1}-z_{k+\frac{1}{2}}\right\|,$$

$\rho_k > -\dfrac{1}{2L_k}$ *and* $\{\rho_k\}$ *is non-decreasing, then the upper bound of the difference* $\mathcal{E}(k+1)-\mathcal{E}(k)$ *is*

$$\begin{aligned}
&\mathcal{E}(k+1)-\mathcal{E}(k)\\
\leqslant & -\frac{D}{2}\left(\frac{1}{2L_k}+\rho_k\right)\left(r-1-\frac{D}{2}\right)\left[(r-1)\left(\frac{1}{2L_k}+\rho_k\right)+2\Sigma_{k+1}\right]\left\|F(z_{k+1})+\widetilde{G}_{k+1}\right\|^2\\
&+D(r^2-r)\left(\frac{1}{2L_k}+\rho_k\right)\left(\rho_k\left\|\widetilde{T}(z_{k+1})\right\|^2-\left\langle\widetilde{T}(z_{k+1}), z_{k+1}-z^*\right\rangle\right)
\end{aligned}$$

Proof. For the sake of simplicity, let $\Sigma_k = \sum\limits_{i=0}^{k-1}\dfrac{1}{2L_i}+\rho_i$ and $\widetilde{T}(z_k) = F(z_k)+\widetilde{G}_k$. First, we have

$$\begin{aligned}
\Sigma_k(z_{k+1}-z_k) = & \, r\left(\frac{1}{2L_k}+\rho_k\right)(u_k-z_{k+1})-2\rho_k\Sigma_k\widetilde{T}(z_k)\\
&-\frac{1}{L_k}\cdot\left[r\left(\frac{1}{2L_k}+\rho_k\right)+\Sigma_k\right](F(z_{k+\frac{1}{2}})+\widetilde{G}_{k+1}) \qquad (37)\\
r\left(\frac{1}{2L_k}+\rho_k\right)(z_k-u_k) = & -\left[r\left(\frac{1}{2L_k}+\rho_k\right)+\Sigma_k\right](z_{k+1}-z_k)-2\rho_k\Sigma_k(F(z_k)+\widetilde{G}_k)\\
&-\frac{1}{L_k}\left[r\left(\frac{1}{2L_k}+\rho_k\right)+\Sigma_k\right](F(z_{k+\frac{1}{2}})+\widetilde{G}_{k+1}). \qquad (38)
\end{aligned}$$

Next, we divide the difference $\mathcal{E}(k+1)-\mathcal{E}(k)$ into three parts.

$$\begin{aligned}
\mathcal{E}(k+1)-\mathcal{E}(k) = & \underbrace{D(\Sigma_{k+1}^2-r\rho_k\Sigma_{k+1})\left\|\widetilde{T}(z_{k+1})\right\|^2-D(\Sigma_k^2-r\rho_{k-1}\Sigma_k)\left\|\widetilde{T}(z_k)\right\|^2}_{\text{I}}\\
&+\underbrace{Dr\Sigma_{k+1}\left\langle\widetilde{T}(z_{k+1}), z_{k+1}-u_{k+1}\right\rangle-Dr\Sigma_k\left\langle\widetilde{T}(z_k), z_k-u_k\right\rangle}_{\text{II}}\\
&+\underbrace{\frac{r^3-r^2}{2}\left[\|u_{k+1}-z^*\|^2-\|u_k-z^*\|^2\right]}_{\text{III}}
\end{aligned}$$

The upper bound of II and III can be estimated as follows. First, we consider II. We split II into three parts as follows:

$$\text{II} = Dr\left(\frac{1}{2L_k}+\rho_k\right)\left\langle\widetilde{T}(z_{k+1}), z_{k+1}-u_{k+1}\right\rangle+Dr\Sigma_k\left\langle\widetilde{T}(z_{k+1}), z_{k+1}-z_k-u_{k+1}+u_k\right\rangle$$

$$+ Dr\Sigma_k \left\langle \widetilde{T}(z_{k+1}) - \widetilde{T}(z_k), z_k - u_k \right\rangle$$

By (37), (38) and $u_{k+1} = u_k - \frac{D}{r}\left(\frac{1}{2L_k} + \rho_k\right)\widetilde{T}(z_{k+1})$, we have

$$\begin{aligned}
\mathrm{II} = {} & Dr\left(\frac{1}{2L_k} + \rho_k\right)\left\langle \widetilde{T}(z_{k+1}), z_{k+1} - u_k \right\rangle + D^2\left(\frac{1}{2L_k} + \rho_k\right)\Sigma_{k+1}\left\|\widetilde{T}(z_{k+1})\right\|^2 \\
& + Dr\left\langle \widetilde{T}(z_{k+1}), r\left(\frac{1}{2L_k} + \rho_k\right)(u_k - z_{k+1}) - 2\rho_k\Sigma_k\widetilde{T}(z_k) \right\rangle \\
& - Dr\left\langle \widetilde{T}(z_{k+1}), \frac{1}{L_k}\left[r\left(\frac{1}{2L_k} + \rho_k\right) + \Sigma_k\right](F(z_{k+\frac{1}{2}}) + \widetilde{G}_{k+1}) \right\rangle \\
& - D\left(\frac{1}{2L_k} + \rho_k\right)^{-1}\Sigma_k\left[r\left(\frac{1}{2L_k} + \rho_k\right) + \Sigma_k\right]\left\langle \widetilde{T}(z_{k+1}) - \widetilde{T}(z_k), z_{k+1} - z_k \right\rangle \\
& - D\left(\frac{1}{2L_k} + \rho_k\right)^{-1}\Sigma_k\left\langle \widetilde{T}(z_{k+1}) - \widetilde{T}(z_k), \frac{1}{L_k}\left[r\left(\frac{1}{2L_k} + \rho_k\right) + \Sigma_k\right](F(z_{k+\frac{1}{2}}) + \widetilde{G}_{k+1}) \right\rangle \\
& - D\left(\frac{1}{2L_k} + \rho_k\right)^{-1}\Sigma_k\left\langle \widetilde{T}(z_{k+1}) - \widetilde{T}(z_k), 2\rho_k\Sigma_k\widetilde{T}(z_k) \right\rangle
\end{aligned}$$

By the assumption

$$\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\| \leqslant L_k\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|$$

we have

$$\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|^2 \leqslant \left\|F(z_{k+\frac{1}{2}}) + \widetilde{G}_{k+1} - \left(\Sigma_k + \frac{r}{2L_k} + r\rho_k\right)^{-1}\Sigma_k\widetilde{T}(z_k)\right\|^2.$$

Combining previous estimation and the assumption

$$\left\langle \widetilde{T}(z_{k+1}) - \widetilde{T}(z_k), z_{k+1} - z_k \right\rangle \geqslant \rho_k\left\|\widetilde{T}(z_{k+1}) - \widetilde{T}(z_k)\right\|^2$$

we can deduce the upper bound of II as follows:

$$\begin{aligned}
\mathrm{II} \leqslant {} & -D\frac{\left[r\left(\frac{1}{2L_k} + \rho_k\right) + \Sigma_k\right]^2}{1 + 2L_k\rho_k}\left\|\widetilde{T}(z_{k+1})\right\|^2 + D^2\left(\frac{1}{2L_k} + \rho_k\right)\Sigma_{k+1}\left\|\widetilde{T}(z_{k+1})\right\|^2 \\
& - D\rho_k\left(\frac{1}{2L_k} + \rho_k\right)^{-1}\Sigma_k\left[r\left(\frac{1}{2L_k} + \rho_k\right) + \Sigma_k\right]\left\|\widetilde{T}(z_{k+1})\right\|^2 \\
& + D(\Sigma_k^2 - r\rho_k\Sigma_k)\left\|\widetilde{T}(z_k)\right\|^2 + D(r^2 - r)\left(\frac{1}{2L_k} + \rho_k\right)\left\langle \widetilde{T}(z_{k+1}), u_k - z_{k+1} \right\rangle
\end{aligned}$$

Next, we consider III. By (18), we have

$$\mathrm{III} = -D(r^2 - r)\left(\frac{1}{2L_k} + \rho_k\right)\left\langle \widetilde{T}(z_{k+1}), u_k - z^* \right\rangle + \frac{D(r-1)}{2}\left(\frac{1}{2L_k} + \rho_k\right)^2\left\|\widetilde{T}(z_{k+1})\right\|^2.$$

Finally, we can estimate the upper bound of $\mathcal{E}(k+1)-\mathcal{E}(k)$. Combining previous estimations and non-decreasing of $\{\rho_k\}$, we have

$$
\begin{aligned}
&\mathcal{E}(k+1)-\mathcal{E}(k)\\
\leqslant\ & -D\left[\frac{(r-1)^2}{2L_k}\left(\frac{1}{2L_k}+\rho_k\right)+\Sigma_{k+1}^2+\frac{r-1}{L_k}\Sigma_{k+1}-(r-1)\left(\frac{1}{2L_k}+\rho_k\right)\rho_k\right]\left\|\widetilde{T}(z_{k+1})\right\|^2\\
&+D\left[\Sigma_{k+1}^2-2(r-1)\rho_k\Sigma_{k+1}+D\left(\frac{1}{2L_k}+\rho_k\right)\Sigma_{k+1}+\frac{D(r-1)}{2}\left(\frac{1}{2L_k}+\rho_k\right)^2\right]\left\|\widetilde{T}(z_{k+1})\right\|^2\\
&-D(r^2-r)\left(\frac{1}{2L_k}+\rho_k\right)\left\langle\widetilde{T}(z_{k+1}),z_{k+1}-z^*\right\rangle
\end{aligned}
$$

By adding and subtracting $D(r^2-r)\rho_k\left(\frac{1}{2L_k}+\rho_k\right)\left\|\widetilde{T}(z_{k+1})\right\|^2$, we have

$$
\begin{aligned}
\mathcal{E}(k+1)-\mathcal{E}(k)\leqslant\ & -\frac{D}{2}\left(\frac{1}{2L_k}+\rho_k\right)\left(r-1-\frac{D}{2}\right)\left[(r-1)\left(\frac{1}{2L_k}+\rho_k\right)+2\Sigma_{k+1}\right]\left\|\widetilde{T}(z_{k+1})\right\|^2\\
&+D(r^2-r)\left(\frac{1}{2L_k}+\rho_k\right)\left(\rho_k\left\|\widetilde{T}(z_{k+1})\right\|^2-\left\langle\widetilde{T}(z_{k+1}),z_{k+1}-z^*\right\rangle\right)
\end{aligned}
$$

At this point, we obtain the desired result. □

By using Theorem 7, one can easily propose a sufficient condition to guarantee the convergence of Algorithm 3.

Corollary 8. *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_{k+1}\}$ and $\{u_k\}$ be the sequences generated by Algorithm 3. If the parameters $\{L_k\}$, $\{\rho_k\}$ and D satisfy the assumptions in Theorem 7 and there exists a z^* such that the inequality*

$$\left\langle F(z_{k+1}+\widetilde{G}_{k+1},z_{k+1}-z^*\right\rangle\geqslant\rho_k\left\|F(z_{k+1})+\widetilde{G}_{k+1}\right\|^2$$

is hold for all $k=0,\,1,\,\cdots$, then we have

$$\operatorname{dist}(0,T(z_k))^2\leqslant\frac{r^2(r-1)^2}{[2(r-1)D-D^2]\left(\sum\limits_{i=0}^{k-1}\frac{1}{2L_i}+\rho_i\right)^2}\|z_0-z^*\|^2,\quad\forall k\geqslant1$$

Proof. Consider the Lyapunov function $\mathcal{E}(k)$ defined in Theorem 7. Let z^* in $\mathcal{E}(k)$ be the point defined in Corollary 8. By the upper bound of $\mathcal{E}(k+1)-\mathcal{E}(k)$ in Theorem 7 , the assumption $D<2(r-1)$ and $\left\langle F(z_{k+1})+\widetilde{G}_{k+1},z_{k+1}-z^*\right\rangle\geqslant\rho_k\|F(z_{k+1})\|^2$, we have

$$\mathcal{E}(k+1)-\mathcal{E}(k)\leqslant0$$

Next, we deduce the lower bound of $\mathcal{E}(k)$. Since $\left\langle F(z_{k+1})+\widetilde{G}_{k+1},z_{k+1}-z^*\right\rangle\geqslant\rho_k\left\|F(z_{k+1})+\widetilde{G}_{k+1}\right\|^2$, we have

$$\mathcal{E}(k)\geqslant\left(D-\frac{D^2}{2(r-1)}\right)\left(\sum_{i=0}^{k-1}\frac{1}{2L_i}+\rho_i\right)^2\left\|F(z_k)+\widetilde{G}_k\right\|^2$$

$$+\frac{1}{2}\left\|\frac{D}{\sqrt{r-1}}\left(\sum_{i=0}^{k-1}\frac{1}{2L_i}+\rho_i\right)F(z_k)-\sqrt{r^3-r^2}(u_k-z^*)\right\|^2$$

Thus

$$\begin{aligned}
&\left(D-\frac{D^2}{2(r-1)}\right)\left(\sum_{i=0}^{k-1}\frac{1}{2L_i}+\rho_i\right)^2 \operatorname{dist}(0,T(z_k))^2\\
\leqslant\ &\left(D-\frac{D^2}{2(r-1)}\right)\left(\sum_{i=0}^{k-1}\frac{1}{2L_i}+\rho_i\right)^2 \left\|F(z_k)+\widetilde{G}_k\right\|^2\\
\leqslant\ &\mathcal{E}(k)\leqslant\mathcal{E}(0)=\frac{r^3-r^2}{2}\left\|z_0-z^*\right\|^2
\end{aligned}$$

By dividing $\left(D-\frac{D^2}{2(r-1)}\right)\left(\sum\limits_{i=0}^{k-1}\frac{1}{2L_i}+\rho_i\right)^2$ on both sides of above inequality, we obtain the desired result. □

However, the assumption $\left\langle F(z_{k+1})+\widetilde{G}_{k+1}, z_{k+1}-z^*\right\rangle \geqslant \rho_k\left\|F(z_{k+1})\right\|^2$ can not be determined in practice. To overcome such problem, we introduce an upper bound constraint on ρ_k.

Corollary 9. *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_{k+1}\}$ and $\{u_k\}$ be the sequences generated by Algorithm 3. If the parameters $\{L_k\}$, $\{\rho_k\}$ and D satisfy the assumptions in Theorem 7 and there exists a z^* such that*

$$\begin{gathered}
\left\langle F(z_k)+\widetilde{G}_k, z-z^*\right\rangle \geqslant \widetilde{\rho}_k\left\|F(z_k)+\widetilde{G}_k\right\|^2, \quad \forall z,\\
\rho_k-\widetilde{\rho}_k\leqslant\frac{E_k}{r}\left(1-\frac{D}{2(r-1)}\right)\left(\sum_{i=0}^{k}\frac{1}{2L_i}+\rho_i\right), \quad E_k\in[0,1],
\end{gathered}$$

then we have

$$\begin{aligned}
&\sum_{k=0}^{\infty}\frac{D}{2}\left(\frac{1}{2L_k}+\rho_k\right)\left(r-1-\frac{D}{2}\right)\left[(r-1)\left(\frac{1}{2L_k}+\rho_k\right)+2(1-E_k)\Sigma_{k+1}\right]\operatorname{dist}(0,T(z_{k+1}))^2\\
\leqslant\ &\frac{r^3-r^2}{2}\cdot\left\|z_0-z^*\right\|^2\\
&(1-E_k)\left(1-\frac{D}{2r-2}\right)\Sigma_k^2\operatorname{dist}(0,T(z_k))^2\leqslant\frac{r^3-r^2}{2D}\cdot\left\|z_0-z^*\right\|^2,
\end{aligned}$$

where $\Sigma_k=\sum\limits_{i=0}^{k-1}\frac{1}{2L_i}+\rho_i$.

Proof. For the sake of simplicity, let $\widetilde{T}(z_k)=F(z_k)+\widetilde{G}_k$. Let z^* in Theorem 7 be the point such that $\left\langle\widetilde{T}(z_k), z_k-z^*\right\rangle\geqslant\widetilde{\rho}_k\left\|\widetilde{T}(z_k)\right\|^2$, $\forall k$. Then we have

$$\begin{aligned}
\mathcal{E}(k+1)-\mathcal{E}(k)\leqslant\ &-\frac{D}{2}\left(\frac{1}{2L_k}+\rho_k\right)\left(r-1-\frac{D}{2}\right)\left[(r+1)\left(\frac{1}{2L_k}+\rho_k\right)+2\Sigma_{k+1}\right]\left\|\widetilde{T}(z_{k+1})\right\|^2\\
&+D(r^2-r)\left(\frac{1}{2L_k}+\rho_k\right)(\rho_k-\widetilde{\rho}_k)\left\|\widetilde{T}(z_{k+1})\right\|^2
\end{aligned}$$

By the assumption that $\rho_k - \widetilde{\rho}_k \leqslant \frac{E}{r}\left(1 - \frac{D}{2(r-1)}\right)\Sigma_{k+1}$, we have

$$\begin{aligned}
&\mathcal{E}(k+1) - \mathcal{E}(k) \\
\leqslant & -\frac{D}{2}\left(\frac{1}{2L_k} + \rho_k\right)\left(r - 1 - \frac{D}{2}\right)\left[(r-1)\left(\frac{1}{2L_k} + \rho_k\right) + 2(1 - E_k)\Sigma_{k+1}\right]\left\|\widetilde{T}(z_{k+1})\right\|^2 .zs
\end{aligned}$$

In addition, the lower bound of $\mathcal{E}(k)$ can be estimated as follows:

$$\mathcal{E}(k) \geqslant (1 - E_k)D\left(1 - \frac{D}{2r-2}\right)\Sigma_k^2\left\|\widetilde{T}(z_k)\right\|^2 + \frac{1}{2}\left\|\frac{D}{\sqrt{r-1}}\Sigma_k\widetilde{T}(z_k) - \sqrt{r^3 - r^2}(u_k - z^*)\right\|^2 \geqslant 0$$

Thus, we have

$$\begin{aligned}
&(1 - E_k)\left(1 - \frac{D}{2r-2}\right)\Sigma_k^2 \mathrm{dist}(0, T(z_k))^2 \\
\leqslant\ & (1 - E_k)\left(1 - \frac{D}{2r-2}\right)\Sigma_k^2\left\|\widetilde{T}(z_k)\right\|^2 \\
\leqslant\ & \mathcal{E}(k) \leqslant \mathcal{E}(0) = \frac{r^3 - r^2}{2}\left\|z_0 - z^*\right\|^2
\end{aligned}$$

Also, by summing $\mathcal{E}(k+1) - \mathcal{E}(k)$ from 0 to ∞, we can obtain the ergodic convergence rate of Algorithm 3. □

The line search framework of the SPEG+ method is similar to Algorithm 3, which is described as follows along with its convergence results.

Corollary 10 (SPEG+ with line search). *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_{k+1}\}$ and $\{u_k\}$ be the sequences generated by the following recursive rule:*

$$\begin{aligned}
\widetilde{z}_{k+1} &= (1 - \alpha_k)z_k + \alpha_k u_k \\
z_{k+\frac{1}{2}} &= P_C\left(\widetilde{z}_{k+1} - (1 - \alpha_k)\frac{1}{L_k}F(z_k)\right) \\
z_{k+1} &= P_C\left(\widetilde{z}_{k+1} - \frac{1}{L_k}F(z_{k+\frac{1}{2}})\right) \\
\widetilde{C}_{k+1} &= L_k(\widetilde{z}_{k+1} - z_{k+1}) - F(z_{k+\frac{1}{2}}) \\
u_{k+1} &= u_k - \frac{D}{2rL_k}[F(z_{k+1}) + \widetilde{C}_{k+1}]
\end{aligned}$$

If $\alpha_k = \frac{rL_k^{-1}}{\sum_{i=0}^{k-1} L_i^{-1} + rL_k}$, $r > 1$, $0 < D < 2(r-1)$ and $L_k \geqslant \frac{\left\|F(z_{k+\frac{1}{2}}) - F(z_{k+1})\right\|}{\left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|}$, then we have

$$\mathrm{dist}(0, T(z_k))^2 \leqslant \frac{r^2(r-1)^2}{[2(r-1)D - D^2]\left(\sum_{i=0}^{k-1}\frac{1}{2L_i}\right)^2} \cdot \mathrm{dist}(z_0, \mathrm{zero}(T))^2, \quad \forall k \geqslant 1$$

$$\left\langle F(z_k) + \widetilde{C}_k, z_k - z^*\right\rangle \leqslant \frac{r^2 - r}{Dr\sum_{i=0}^{k-1} L_i^{-1}}\left\|z_0 - z^*\right\|^2, \quad \forall z^* \in \mathrm{zero}(T)$$

The proof for Corollary 10 is similar to the proof for Theorem 7 and Corollary 8. Thus, we omit the proof of Corollary 10.

Remark. 1) Although previous arguments suggest that when $D = r - 1$, the corresponding instance of Algorithm 3 may obtain the fastest numerical performance, this result is not generally observed in practical numerical experiment. The reason is that the inequalities

$$\begin{aligned}
\rho_k &\leqslant \frac{\left\langle F(z_{k+1}) + \widetilde{G}_{k+1} - F(z_k) + \widetilde{G}_k, z_{k+1} - z_k \right\rangle}{\left\| F(z_{k+1}) + \widetilde{G}_{k+1} - F(z_k) - \widetilde{G}_k \right\|^2} \\
L_k &\geqslant \frac{\left\| F(z_{k+\frac{1}{2}}) - F(z_{k+1}) \right\|}{\left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|}
\end{aligned}$$

are generally strict. Thus we encourage that $D > r - 1$ when using Algorithm 3.

2) In Theorem 7, Algorithm 3 imposes a condition that the parameter ρ_k is non-decreasing. Nonetheless, a problem may occur in practice, that is ρ_k may grow too large so that there does not exist a feasible point z_{k+1} such that the line search principle can be fulfilled. To circumvent this problem, a simple solution is restarting the algorithm. Specifically, run Algorithm 3 again with initial point z_k.

6 Conclusion and Further Discussion

In this paper, we introduce the Symplectic Extra-Gradient (SEG) Method, a novel accelerated extra-gradient approach based on symplectic acceleration techniques. By leveraging the Lyapunov function framework, we establish a convergence rate of $O\left(1/k^2\right)$, and under stronger assumptions, demonstrate a faster $o\left(1/k^2\right)$ convergence rate along with weak convergence properties. To enhance computational efficiency, we incorporate a line search technique, which adapts the step-size dynamically during optimization.

While our results are promising, several theoretical and practical questions remain. Exploring a stochastic variant of the SEG method could provide more efficient training for applications like Generative Adversarial Networks (GANs). Additionally, inspired by the broader applicability of Tseng's methods for general monotone inclusion problems, extending the SEG framework to more generalized cases is a promising direction. Further practical improvements could focus on optimizing the line search strategy, particularly in estimating parameters like L_k and ρ_k, as current implementations show only modest improvements in some test cases.

Our numerical experiments demonstrate that the SEG method exhibits faster convergence compared to existing extra-gradient techniques, but further fine-tuning of its parameters and expanding its applicability remains an intriguing area for future research.

References

[BC17] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces (Second Edition).* Springer Cham, Cham, 03 2017.

[BMW21] Heinz H Bauschke, Walaa M Moursi, and Xianfu Wang. Generalized monotone operators and their averaged resolvents. *Mathematical Programming*, 189:55–74, 2021.

[BTGN09] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*, volume 28. Princeton university press, Princeton, 2009.

[COZ22] Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. Accelerated algorithms for constrained nonconvex-nonconcave min-max optimization and comonotone inclusion, 2022.

[DDJ21] Jelena Diakonikolas, Constantinos Daskalakis, and Michael I. Jordan. Efficient methods for structured nonconvex-nonconcave min-max optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 2746–2754. PMLR, 2021.

[FLC23] Yifeng Fan, Yongqiang Li, and Bo Chen. Weaker mvi condition: Extragradient methods with multi-step exploration. In *The Twelfth International Conference on Learning Representations*, 2023.

[GBGL22] Eduard Gorbunov, Hugo Berard, Gauthier Gidel, and Nicolas Loizou. Stochastic extragradient: General analysis and improved rates. In *International Conference on Artificial Intelligence and Statistics*, pages 7865–7901. PMLR, 2022.

[GBV+18] Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial networks. *arXiv preprint arXiv:1802.10551*, 2018.

[GPAM+14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[Hal67] Benjamin Halpern. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73:957–961, 1967.

[IJOT17] Alfredo N. Iusem, Alekandro Jofré, Roberto I. Oliveira, and Philip Thompson. Extragradient method with variance reduction for stochastic variational inequalities. *SIAM Journal on Optimization*, 27(2):686–724, 2017.

[JFB00] Alexander Shapiro J. Frédéric Bonnans. *Perturbation Analysis of Optimization Problems*. Springer New York, New York, 05 2000.

[Kor76] Galina M Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

[KS11] Aswin Kannan and Uday V. Shanbhag. Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58, 2011.

[KS19] Aswin Kannan and Uday V Shanbhag. Optimal stochastic extragradient schemes for pseudomonotone stochastic variational inequality problems and their variants. *Computational Optimization and Applications*, 74(3):779–820, 2019.

[LK21] Sucheol Lee and Donghwan Kim. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.

[Min62] George J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29(3):341 – 346, 1962.

[MKS+20] Konstantin Mishchenko, Dmitry Kovalev, Egor Shulgin, Peter Richtárik, and Yura Malitsky. Revisiting stochastic extragradient. In *International Conference on Artificial Intelligence and Statistics*, pages 4573–4582. PMLR, 2020.

[MMS+18] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*, 2018.

[MOP20] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. In *International Conference on Artificial Intelligence and Statistics*, pages 1497–1507. PMLR, 2020.

[Mye97] Roger B. Myerson. *Game Theory: Analysis of Conflict.* Harvard University Press, Cambridge, 09 1997.

[Nem04] Arkadi Nemirovski. Prox-method with rate of convergence $o(1/t)$ for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

[Nes83] Yurii Nesterov. A method of solving a convex programming problem with convergence rate $o(1/k^2)$. *Doklady Akademii Nauk SSSR*, 269(3):543, 1983.

[NT06] N. Nadezhkina and W. Takahashi. Weak convergence theorem by an extragradient method for nonexpansive mappings and monotone mappings. *Journal of Optimization Theory and Applications*, 128:191–201, 2006.

[PLP+22] Thomas Pethick, Puya Latafat, Panos Patrinos, Olivier Fercoq, and Volkan Cevher. Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems. In *International Conference on Learning Representations*, 2022.

[Pov80] Leonid D. Povov. A modification of the arrow-hurwitz method of search for saddle points. *Mat. Zametki*, 28(5):777–784, 1980.

[QX21] Huiqiang Qi and Hong-Kun Xu. Convergence of halpern's iteration method with applications in optimization. *Numerical Functional Analysis and Optimization*, 42(15):1839–1854, 2021.

[RFP13] Hugo Raguet, Jalal Fadili, and Gabriel Peyré. A generalized forward-backward splitting. *SIAM Journal on Imaging Sciences*, 6(3):1199–1226, 2013.

[TB84] J. E. Taylor and Martin P. Bendsøe. An interpretation for min-max structural design problems including a method for relaxing constraints. *International Journal of Solids and Structures*, 20(4):301–314, 1984.

[TD23] Quoc Tran-Dinh. Sublinear convergence rates of extragradient-type methods: A survey on classical and recent developments, 2023.

[Tse95] Paul Tseng. On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, 60(1-2):237–252, 1995.

[Tse00] Paul Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38(2):431–446, 2000.

[Val14] Tuomo Valkonen. A primal–dual hybrid gradient method for nonlinear operators with applications to mri. *Inverse Problems*, 30(5):055012, 2014.

[VKA10] F. P. Vasilyev, E. V. Khoroshilova, and A. S. Antipin. An extragradient method for finding the saddle point in an optimal control problem. *Moscow University Computational Mathematics and Cybernetics*, 34:113–118, 2010.

[vNM47] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior (2nd rev. ed.).* Princeton University Press, Princeton, 1947.

[YR21] TaeHo Yoon and Ernest K. Ryu. Accelerated algorithms for smooth convex-concave minimax problems with $o(1/k^2)$ rate on squared gradient norm. In *International Conference on Machine Learning*, pages 12098–12109. PMLR, 2021.

[YZ23] Ya-Xiang Yuan and Yi Zhang. Symplectic discretization approach for developing new proximal point algorithms, 2023.

[ZC08] Mingqiang Zhu and Tony Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *Ucla Cam Report*, 34(2), 2008.

A Convergence of Symplectic Extra-gradient Method

To proof convergence of the symplectic extra-gradient method, we need to study the following Lyapunov function:

$$\mathcal{E}(k) = \frac{A_k}{2} \|F(z_k)\|^2 + B_k \langle F(z_k), z_k - u_k \rangle . \tag{39}$$

The first thing we should do is to find out the sufficient condition to ensure that $\mathcal{E}$ is a Lyapunov function, i. e. $\{\mathcal{E}(k)\}$ is non-increasing. The conditions (40)-(42) are based on the conditions in Lemma 2 in [YR21].

Theorem 11. *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_k\}$ and $\{u_k\}$ be the sequences generated by* (9)*. If*

$$A_k = \frac{B_k \beta_k}{\alpha_k} \tag{40}$$

$$B_{k+1} = \frac{B_k}{1-\alpha_k} \tag{41}$$

$$\beta_{k+1} + 2C_k\alpha_{k+1} = \frac{\beta_k \alpha_{k+1}(1 - L^2\beta_k^2 - \alpha_k^2)}{\alpha_k(1-\alpha_k)(1-L^2\beta_k^2)} \tag{42}$$

$$\beta_k \in (0, \frac{1}{L}), \quad \forall k \geqslant 0. \tag{43}$$

then the sequence $\{\mathcal{E}(k)\}$ is non-increasing.

Proof. Step 1: Dividing the difference $\mathcal{E}(k+1) - \mathcal{E}(k)$ into 2 parts.

$$\begin{aligned}\mathcal{E}(k+1) - \mathcal{E}(k) = & \underbrace{\frac{A_{k+1}}{2} \left\|F(z_{k+1})^2\right\| - \frac{A_k}{2} \|F(z_k)\|^2}_{\mathrm{I}} \\ & + \underbrace{B_{k+1} \langle F(z_{k+1}), z_{k+1} - u_{k+1}\rangle - B_k \langle F(z_k), z_k - u_k\rangle}_{\mathrm{II}}\end{aligned}$$

Step 2: Estimate the upper bound of II. II can be divided into the following three pieces.

$$\begin{aligned} & B_{k+1} \langle F(z_{k+1}), z_{k+1} - u_{k+1}\rangle - B_k \langle F(z_k), z_k - u_k\rangle \\ = & (B_{k+1} - B_k) \langle F(z_{k+1}), z_{k+1} - u_{k+1}\rangle + B_k \langle F(z_{k+1}), z_{k+1} - z_k - u_{k+1} + u_k\rangle \\ & + B_k \langle F(z_{k+1}) - F(z_k), z_k - u_k\rangle \end{aligned}$$

Since

$$z_{k+1} = \alpha_k u_k + (1-\alpha_k) z_k - \beta_k F(x_{k+\frac{1}{2}})$$

we have

$$B_k \langle F(z_{k+1}) - F(z_k), z_k - u_k\rangle = B_k \left\langle F(z_{k+1}) - F(z_k), -\frac{1}{\alpha_k}(z_{k+1} - z_k) - \frac{\beta_k}{\alpha_k} F(z_{k+\frac{1}{2}}) \right\rangle \tag{44}$$

Also, we have

$$\begin{aligned} & B_k \langle F(z_{k+1}), z_{k+1} - z_k - u_{k+1} + u_k\rangle \\ = & B_k \left\langle F(z_{k+1}), \frac{\alpha_k}{1-\alpha_k}(u_k - z_{k+1}) - \frac{\beta_k}{1-\alpha_k} F(z_{k+\frac{1}{2}}) \right\rangle + B_k C_k \|F(z_{k+1})\|^2 \end{aligned} \tag{45}$$

In addition, we have

$$(B_{k+1}-B_k)\langle F(z_{k+1}), z_{k+1}-u_{k+1}\rangle = (B_{k+1}-B_k)C_k\|F(z_{k+1})\|^2+(B_{k+1}-B_k)\langle F(z_{k+1}), z_{k+1}-u_k\rangle \tag{46}$$

By summing (44), (45) and (46) and using (41), we obtain

$$\begin{aligned} \mathrm{II} = {} & B_{k+1}C_k\|F(x_{k+1})\|^2 - \frac{B_k}{a_k}\langle F(z_{k+1})-F(z_k), z_{k+1}-z_k\rangle \\ & -\left(\frac{B_k\beta_k}{\alpha_k}+\frac{B_k\beta_k}{1-\alpha_k}\right)\left\langle F(z_{k+1}), F(z_{k+\frac{1}{2}})\right\rangle + \frac{B_k\beta_k}{\alpha_k}\left\langle F(z_k), F(z_{k+\frac{1}{2}})\right\rangle \end{aligned}$$

Since F is $L-$Lipschitz continuous, we have

$$\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2 \leqslant L^2\beta_k^2\left\|F(z_{k+\frac{1}{2}})-F(z_k)\right\|^2$$

Because of the equality

$$-\left\langle F(z_{k+1}), F(z_{k+\frac{1}{2}})\right\rangle = \frac{1}{2}\left[\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2 - \|F(z_{k+1})\|^2 - \left\|F(z_{k+\frac{1}{2}})\right\|^2\right]$$

we have

$$\begin{aligned} \mathrm{II} \leqslant {} & \left(B_{k+1}C_k - \frac{B_k\beta_k}{2\alpha_k L^2\beta_k^2}\right)\|F(z_{k+1})\|^2 + \frac{B_k\beta_k}{2\alpha_k}\|F(z_k)\|^2 + \frac{B_k\beta_k}{2\alpha_k}(1-\frac{1}{L^2\beta_k^2})\left\|F(z_{k+\frac{1}{2}})\right\|^2 \\ & -\left(\frac{B_k\beta_k}{\alpha_k}(1-\frac{1}{L^2\beta_k^2})+\frac{B_k\beta_k}{1-\alpha_k}\right)\left\langle F(z_{k+\frac{1}{2}}), F(z_{k+1})\right\rangle \end{aligned}$$

Step 3: Estimate the upper bound of $\mathcal{E}(k+1)-\mathcal{E}(k)$. By previous estimation and the equality (40), we have

$$\begin{aligned} \mathcal{E}(k+1)-\mathcal{E}(k) \leqslant {} & \left(\frac{A_{k+1}}{2}+B_{k+1}C_k - \frac{B_k\beta_k}{2\alpha_k L^2\beta_k^2}\right)\|F(z_{k+1})\|^2 + \frac{B_k\beta_k}{2\alpha_k}(1-\frac{1}{L^2\beta_k^2})\left\|F(z_{k+\frac{1}{2}})\right\|^2 \\ & -\left(\frac{B_k\beta_k}{\alpha_k}(1-\frac{1}{L^2\beta_k^2})+\frac{B_k\beta_k}{1-\alpha_k}\right)\left\langle F(z_{k+\frac{1}{2}}), F(z_{k+1})\right\rangle \end{aligned}$$

By (40), (41) and (42), we have

$$\begin{aligned} \frac{A_{k+1}}{2}+B_{k+1}C_k &= \frac{B_{k+1}\beta_{k+1}}{2\alpha_{k+1}}+B_{k+1}C_k \\ &= \frac{B_k\beta_{k+1}}{2\alpha_{k+1}(1-\alpha_k)}+\frac{B_kC_k}{1-\alpha_k} \\ &= \frac{B_k(\beta_{k+1}+2C_k\alpha_{k+1})}{2\alpha_{k+1}(1-\alpha_k)} \\ &= \frac{B_k\beta_k(1-L^2\beta_k^2-\alpha_k^2)}{2\alpha_k(1-\alpha_k)^2(1-L^2\beta_k^2)} \\ &= \frac{A_k(1-L^2\beta_k^2-\alpha_k^2)}{2(1-\alpha_k)^2(1-L^2\beta_k^2)} \end{aligned}$$

Algorithm 4: SEG with Varying Step-size

Input: Operator F, $L \in (0, +\infty)$, ;

Initialization: z_0, $u_0 = z_0$;

for $k = 0, 1, \cdots$ **do**

$$\widetilde{z}_{k+1} = \frac{1}{k+r} z_k + \frac{k+r-1}{k+r} u_k;$$
$$z_{k+\frac{1}{2}} = \widetilde{z}_{k+1} - \beta_k F(z_k);$$
$$z_{k+1} = \widetilde{z}_{k+1} - \beta_k F(z_{k+\frac{1}{2}});$$
$$u_{k+1} = u_k - D \cdot \frac{L^2\beta_k^3}{2(k+r-1)(1-L^2\beta_k^2)} F(z_{k+1});$$
$$\beta_{k+1} = \beta_k - (1+D) \cdot \frac{L^2\beta_k^3}{[(k+r)^2-1](1-L^2\beta_k^2)};$$

Also by (40), we have

$$\frac{B_k\beta_k}{2\alpha_k}(1 - \frac{1}{L^2\beta_k^2}) = \frac{A_k(L^2\beta_k^2 - 1)}{2L^2\beta_k^2}$$
$$-\frac{B_k\beta_k}{\alpha_k}(1 - \frac{1}{L^2\beta_k^2}) + \frac{B_k\beta_k}{1-\alpha_k} = \frac{A_k(1 - L^2\beta_k^2 - \alpha_k)}{2(1-\alpha_k)L^2\beta_k^2}$$

In conclusion, we have

$$\begin{aligned}
\mathcal{E}(k+1) - \mathcal{E}(k) \leqslant & -\frac{A_k(1-L^2\beta_k^2-\alpha_k)^2}{2(1-\alpha_k)^2(1-L^2\beta_k^2)L^2\beta_k^2} \|F(z_{k+1})\|^2 - \frac{A_k(1-L^2\beta_k^2)}{2L^2\beta_k^2} \left\|F(z_{k+\frac{1}{2}})\right\|^2 \\
& + \frac{A_k(1-L^2\beta_k^2-\alpha_k)}{2(1-\alpha_k)L^2\beta_k^2} \left\langle F(z_{k+\frac{1}{2}}), F(z_{k+1})\right\rangle \\
= & -\frac{A_k}{2L^2\beta_k^2} \left\| \frac{(1-L^2\beta_k^2-\alpha_k)}{(1-\alpha_k)\sqrt{1-L^2\beta_k^2}} F(z_{k+1}) - \sqrt{1-L^2\beta_k^2} F(z_{k+\frac{1}{2}}) \right\|^2 \\
\leqslant & \ 0
\end{aligned}$$

□

However, unlike Theorem 2 in [YR21], we can not directly estimate the upper bounds of $\|F(z_k)\|^2$, and $\langle F(z_k), z_k - z^*\rangle$. The main reason is that the lower bound of $E(k)$ is

$$\mathcal{E}(k) \geqslant \frac{A_k}{4} \|F(z_k)\|^2 + B_k \langle F(z_k), z_k - z^*\rangle - \frac{B_k^2}{A_k^2} \|u_k - z^*\|^2$$

which leads to that both $\|F(z_k)\|^2$, and $\langle F(z_k), z_k - z^*\rangle$ are bounded by both $\|u_k - z^*\|^2$ and $\|z_0 - z^*\|^2$. Because determining all classes of SEG that satisfies (43) and admits $O(1/k^2)$ convergence rate is complicated, we first focus on the following instance of SEG, described in Algorithm 4.

Algorithm 4 can be seen as SEG with $\alpha_k = \dfrac{1}{k+r}, C_k = D \cdot \dfrac{L^2\beta_k^3}{2(k+r-1)(1-L^2\beta_k^2)}$. The first thing we need to do is propose the sufficient condition for SEG such that (43) hold.

Lemma 12. *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_k\}$ and $\{u_k\}$ be the sequences generated by Algorithm 4. If*

$$\beta_0 \in (0, \frac{1}{L}) \tag{47}$$

$$-1 < D < \frac{r(r-1)(1-L^2\beta_0^2)}{(2r-1)L^2\beta_0^2} - 1 \tag{48}$$

hold, $\{\beta_k\}$ is a non-increasing sequence and $\beta_\infty = \lim_{k\to\infty} \beta_k > 0$.

Proof. Without loss of generalization, we presume $L = 1$ because the general case can be obtained by replacing β_k by $L\beta_k$. By the assumption, the difference $\beta_{k+1} - \beta_k$ is given as follows:

$$\beta_{k+1} - \beta_k = -(1+D) \cdot \frac{\beta_k^3}{[(k+r)^2 - 1](1-\beta_k^2)}$$

Next, we verify that $\beta_k > 0$ by using mathematical induction. We assume that $\beta_i > 0$ is hold for all $i = 0, \cdots, k$, then we have $\beta_{i+1} - \beta_i \leqslant 0, \forall i = 0, \cdots, k$. Since the function $g(x) = \dfrac{x^3}{1-x^2}$ is non-increasing on $(0,1)$, we have

$$\frac{\beta_i^3}{1-\beta_i^2} \leqslant \frac{\beta_0^3}{1-\beta_0^2}$$

Thus

$$\begin{aligned}
\beta_{k+1} - \beta_0 &= \sum_{i=0}^{k} \beta_{i+1} - \beta_i \\
&\geqslant -(1+D) \cdot \sum_{i=0}^{k} \frac{\beta_0^3}{(i+r-1)(i+r-1)(1-\beta_0^2)} \\
&\geqslant -\frac{(1+D)\beta_0^3}{(1-\beta_0^2)} \sum_{i=0}^{\infty} \frac{1}{(i+r-1)(i+r+1)} \\
&= -\frac{(2r-1)(1+D)\beta_0^3}{r(r-1)(1-\beta_0^2)}
\end{aligned}$$

In conclusion, we have

$$\beta_{k+1} \geqslant \left(1 - \frac{(2r-1)(1+D)\beta_0^2}{r(r-1)(1-\beta_0^2)}\right) \beta_0$$

□

The following lemma is useful to show that $\|u_k - z^*\|^2$ can be uniformly bounded by $\|z_0 - z^*\|^2$.

Lemma 13. *Let $\{a_k\}$ be a non-negative sequence. If there exists two positive constants E_1, E_2 such that*

$$\left(1 - \frac{E_1}{(k+r)^2 - 1}\right) a_{k+1} \leqslant \frac{E_2}{(k+r-1)(k+r)} a_0 + a_k, \quad E_1 < r^2 - 1$$

the sequence $\{a_k\}$ can be uniformly bounded by a_0, i. e. there exists a positive constant E_3 such that

$$a_k \leqslant E_3 a_0, \quad \forall k \geqslant 1$$

Proof. Deriving $1-\dfrac{E_1}{(k+r)^2-1}$ on both sides of the inequality in Lemma 13, we have

$$a_{k+1} \leqslant \frac{(k+r+1)E_2}{[(k+r-1)^2-E_1](k+r)}a_0 + \frac{(k+r)^2-1}{(k+r)^2-1-E_1}a_k$$

Recursively using upper bound of $a_k, a_{k-1}, \cdots, a_1$, we have

$$a_{k+1} \leqslant \sum_{i=0}^{k}\prod_{j=i+1}^{k}\frac{(j+r)^2-1}{(j+r)^2-1-E_1}\frac{(i+r+1)E_2}{[(i+r)^2-1-E_1](i+r)}a_0$$

where $\prod_{j=k+1}^{k}\dfrac{(j+r)^2-1}{(j+r)^2-1-E_1}=1$. Since

$$\prod_{j=i+1}^{k}\frac{(j+r)^2-1}{(j+r)^2-1-E_1} \leqslant \prod_{j=0}^{\infty}\frac{(j+r)^2-1}{(j+r)^2-1-E_1} < \infty$$

then we have

$$a_{k+1} \leqslant E_4\frac{(i+r+1)E_2}{[(i+r)^2-1-E_1](i+r)}a_0$$

where

$$E_4 = \prod_{j=0}^{\infty}\frac{(j+r)^2-1}{(j+r)^2-1-E_1}$$

Also, because

$$\sum_{i=0}^{k}\frac{(i+r+1)E_2}{[(i+r)^2-1-E_1](i+r)} \leqslant \sum_{i=0}^{\infty}\frac{(i+r+1)E_2}{[(i+r)^2-1-E_1](i+r)} < \infty$$

we have

$$a_{k+1} \leqslant E_3 a_0$$

where

$$E_3 = E_4E_2\sum_{i=0}^{\infty}\frac{i+r+1}{[(i+r)^2-1-E_1](i+r)}$$

$\square$

With Lemma 12 and Lemma 13, we can show that Algorithm 4 exhibits $O(1/k^2)$ convergence rates.

Theorem 14. *Let* $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_k\}$ *and* $\{u_k\}$ *be the sequences generated by Algorithm 4. If* (47) *and* (48) *hold, we have*

$$\|F(z_k)\|^2 \leqslant O\left(\frac{\mathrm{dist}(z_0,F)^2}{(k+r-1)(k+r)}\right),$$
$$\langle F(z_k), z_k-z^*\rangle \leqslant O\left(\frac{\|z_0-z^*\|^2}{k+r-1}\right)$$

Proof. Let $B_0 = 1$. By (40) and (41), we have $B_k = k + r - 1, A_k = \beta_k(k + r - 1)(k + r)$. By Theorem 11, we have

$$\mathcal{E}(k) \leqslant \mathcal{E}(0) = \frac{(r-1)r\beta_0}{2} \|F(z_0)\|^2 \leqslant \frac{(r-1)rL^2\beta_0}{2} \|z_0 - z^*\|^2$$

The lower bound of $\mathcal{E}(k)$ can be estimated as followed.

$$\begin{aligned}
\mathcal{E}(k) &= \frac{A_k}{2} \|F(z_k)\|^2 + B_k \langle F(z_k), z_k - z^* \rangle + B_k \langle F(z_k), z^* - u_k \rangle \\
&\geqslant \frac{A_k}{2} \|F(z_k)\|^2 + B_k \langle F(z_k), z_k - z^* \rangle \\
&\quad + \frac{1}{2} \left[\left\| \sqrt{\frac{A_k}{2}} F(z_k) - \sqrt{\frac{2B_k^2}{A_k}} (u_k - z^*) \right\|^2 - \frac{A_k}{2} \|F(z_k)\|^2 - \frac{2B_k^2}{A_k} \|u_k - z^*\|^2 \right] \\
&\geqslant \frac{A_k}{4} \|F(z_k)\|^2 + B_k \langle F(z_k), z_k - z^* \rangle - \frac{B_k^2}{A_k} \|u_k - z^*\|^2
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\|F(z_k)\|^2 &\leqslant \frac{2(r-1)rL^2\beta_0}{\beta_k(k+r-1)(k+r)} \|z_0 - z^*\|^2 + \frac{4}{\beta_k^2(k+r)^2} \|u_k - z^*\|^2 \\
\langle F(z_k), z_k - z^* \rangle &\leqslant \frac{(r-1)rL^2\beta_0}{2(k+r-1)} \|z_0 - z^*\|^2 + \frac{1}{\beta_k(k+r)} \|u_0 - z^*\|^2
\end{aligned}$$

Since $\beta_k \geqslant \beta_\infty > 0$, the remaining problem is showing $\|u_k - z\|$ can be uniformly bounded by $\|u_0 - z\| = \|z_0 - z^*\|$. By triangle inequality, we have

$$\|u_{k+1} - z^*\| \leqslant |C_k| \|F(z_{k+1})\| + \|u_k - z^*\|$$

By the definition of C_k and (48), we have

$$|C_k| \leqslant \frac{r(r-1)(1 - L^2\beta_0^2)}{(2r-1)L^2\beta_0^2} \cdot \frac{L^2\beta_k^3}{2(k+r-1)(1 - L^2\beta_k^3)} \leqslant \frac{r(r-1)}{(4r-2)(k+r-1)} \beta_k$$

Also by $\sqrt{a+b} \leqslant \sqrt{a} + \sqrt{b}, \forall a > 0, b > 0$, we have

$$\|u_{k+1} - z^*\| \leqslant \frac{r(r-1)}{(2r-1)[(k+r)^2 - 1]} \|u_{k+1} - z^*\| + \frac{r^{\frac{3}{2}}(r-1)^{\frac{3}{2}} L\beta_0}{\sqrt{2}(2r-1)(k+r-1)(k+r)} \|u_0 - z^*\| + \|u_k - z^*\|$$

Let $E_1 = \dfrac{r(r-1)}{2r-1}, E_2 = \dfrac{r^{\frac{3}{2}}(r-1)^{\frac{3}{2}} L\beta_0}{\sqrt{2}(2r-1)}, a_k = \|u_k - z^*\|$. By using Lemma 13, we have $\|u_k - z^*\| \leqslant E_3 \|z_0 - z^*\|$.

In conclusion, we have

$$\begin{aligned}
\|F(z_k)\|^2 &\leqslant O\left(\frac{\|z_0 - z^*\|^2}{(k+r-1)(k+r)} \right), \\
\langle F(z_k), z_k - z^* \rangle &\leqslant O\left(\frac{\|z_0 - z^*\|^2}{k+r-1} \right)
\end{aligned}$$

By taking infimum respect to all $z^* \in \text{zero}(F)$ in the first above inequality, we obtain the desired results.

□

Algorithm 5: SEG with Constant Step-size

Input: Operator F, $L \in (0,+\infty)$, $\rho \in (-\frac{1}{2\rho},+\infty)$;

Initialization: z_0, $u_0 = z_0$;

for $k = 0, 1, \cdots$ **do**

$\widetilde{z}_{k+1} = \frac{1}{k+r} z_k + \frac{k+r-1}{k+r} u_k$;

$z_{k+\frac{1}{2}} = \widetilde{z}_{k+1} - \beta F(z_k)$;

$z_{k+1} = \widetilde{z}_{k+1} - \beta F(z_{k+\frac{1}{2}})$;

$u_{k+1} = u_k + \frac{L^2\beta^3}{2(k+r-1)(1-L^2\beta^2)} F(z_{k+1})$;

Since the step-size of Nesterov's accelerated gradient method can be constant, we also discuss the SEG with constant step-size. It is easy to show that if $\alpha_k = \frac{1}{k+r}, C_k = -\frac{L^2\beta^3}{(k+r-1)(1-L^2\beta_k^2)}$, $\beta_k \equiv \beta$. In conclusion, we obtain the SEG with constant step-size, described in Algorithm 5.

The convergence results of Algorithm 5 is given in Corollary 15.

Corollary 15. *Let $\{z_k\}$, $\{z_{k+\frac{1}{2}}\}$, $\{\widetilde{z}_k\}$ and $\{u_k\}$ be the sequences generated by Algorithm 5 with $\beta \in (0, \frac{1}{L})$. Then we have*

$$\|F(z_k)\|^2 \leqslant O\left(\frac{\operatorname{dist}(z_0, \operatorname{zero}(F))^2}{(k+r-1)(k+r)}\right),$$

$$\langle F(z_k), z_k - z^*\rangle \leqslant O\left(\frac{\|z_0 - z^*\|^2}{k+r-1}\right), \quad \forall z^* \in \operatorname{zero}(F)$$

B Discussion of Stochastic SEG+

B.1 Symplectic Stochastic Extragradient+ Method

In this section, we consider the following stochastic inclusion problem:

$$0 = F(z) = \mathop{\mathbb{E}}_{\zeta}[F(z;\zeta)]$$

Such stochastic inclusion problem has been wildly used to train GANs and train Adversarial training deep neural network classifiers. To solve the above problem, we proposed the following stochastic SEG+ framework.

$$\widetilde{z}_{k+1} = (1-\alpha_k) z_k + \alpha_k u_k \tag{49a}$$

$$z_{k+\frac{1}{2}} = \widetilde{z}_{k+1} - (1-\alpha_k)\beta_k[F(z_k) + \zeta_k] \tag{49b}$$

$$z_{k+1} = \widetilde{z}_{k+1} - \beta_k[F(z_{k+\frac{1}{2}}) + \zeta_{k+\frac{1}{2}}] \tag{49c}$$

$$u_{k+1} = u_k - C_k[F(z_{k+1}) + \zeta_{k+1}]. \tag{49d}$$

Algorithm 6: Symplectic Stochastic Extragradient+ Method, SSEG+

Input: Operator F, $L \in (0, +\infty), \rho \in (-\frac{1}{2\rho}, +\infty)$;

Initialization: $z_0, u_0 = z_0$;

for $k = 0, 1, \cdots$ **do**

$\widetilde{z}_{k+1} = \frac{k}{k+r} z_k + \frac{r}{k+r} u_k$;

$z_{k+\frac{1}{2}} = \widetilde{z}_{k+1} - \frac{k}{k+r}\frac{1}{L}[F(z_k) + \zeta_k]$;

$z_{k+1} = \widetilde{z}_{k+1} - \frac{1}{L}[F(z_{k+\frac{1}{2}}) + \zeta_{k+\frac{1}{2}}]$;

$u_{k+1} = u_k - \frac{D}{r}[F(z_{k+1}) + \zeta_{k+1}]$.

Theorem 16. *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_{k+1}\}$ and $\{u_k\}$ be the sequences generated by (49a)-(49d). If F is $L-$Lipschitz continuous and monotone, the parameters satisfy the following requirements:*

$$\beta_k \in (0, \frac{1}{L})$$
$$\frac{B_k\beta_k}{\alpha_k} = \left(\frac{B_k\beta_k}{\alpha_k} + \frac{B_k\beta_k}{1-\alpha_k}\right) L^2\beta_k^2(1-\alpha_k),$$
$$\frac{B_k\beta_k}{2\alpha_k} + \frac{B_k\beta_k}{2-2\alpha_k} \geqslant B_{k+1}C_k + \frac{C_k(1-\alpha_k)B_{k+1} - C_kB_k}{2-2\alpha_k} + A_{k+1}$$
$$\frac{B_k\beta_k}{2\alpha_k} \geqslant A_k, \quad \frac{(1-\alpha_k)B_{k+1} - B_k}{2C_k(1-\alpha_k)} \text{ is non-increasing,}$$

then we have

$$\mathbb{E}\left[\mathcal{E}(k)\right] \leqslant \mathbb{E}\left[\mathcal{E}(0)\right] + \sum_{i=0}^{k-1}\left(\frac{B_i\beta_i}{1-\alpha_i} + \frac{B_i\beta_i}{\alpha_i}\right) L\beta_i\sigma_{i+\frac{1}{2}}^2 + \frac{C_iB_i - C_k(1-\alpha_i)B_{i+1}}{2-2\alpha_i}\sigma_{i+1}^2$$
$$+ \sum_{i=0}^{k-1}\left(\frac{B_i\beta_i}{2\alpha_i} + \frac{B_i\beta_i}{2-2\alpha_i}\right)[L^2\beta_i^2\sigma_{i+\frac{1}{2}}^2 + L^2\beta_i^2(1-\alpha_i)^2\sigma_i^2 + 2L^3\beta_i^3(1-\alpha_i)^2\sigma_i^2]$$

Corollary 17. *Let $\{z_{\frac{k}{2}}\}$, $\{\widetilde{z}_{k+1}\}$ and $\{u_k\}$ be the sequences generated by Algorithm 6. If F is $L-$Lipschitz continuous and monotone, $r \geqslant 2$, $0 < D \leqslant 1$, and there exists two positive constant E_1 and E_2 such that $\sigma_k^2 \leqslant E_1\frac{\varepsilon}{rk}$, $\sigma_{k+\frac{1}{2}}^2 \leqslant E_2\frac{\varepsilon}{r(k+r)}$, there exists a positive constant E_3 such that*

$$\mathbb{E}\left[\|F(z_k)\|^2\right] \leqslant \frac{L(r^3 - r^2)}{Dk^2}\text{dist}(z_0, \text{zero}(F))^2 + E_3\varepsilon$$

B.2 Proof of Theorem 16

Before proving Theorem 16, we need the following lemma.

Lemma 18. *Let* $\{z_k\}, \{u_k\}$ *be the sequences generated by* (49a)-(49d)*. If the random sequence* $\{\zeta_{\frac{k}{2}}\}$ *satisfies* $\mathbb{E}\left[\zeta_{\frac{k}{2}}\right] = 0$, $\mathbb{E}\left[\zeta_{\frac{k}{2}}^2\right] = \sigma_{\frac{k}{2}}^2$ *for all* $k \geqslant 0$*, then we have*

$$\left|\mathbb{E}\left[F(z_{k+1}), \zeta_{k+\frac{1}{2}}\right]\right| \leqslant L\beta_k\sigma_{k+\frac{1}{2}}^2,$$
$$\left|\mathbb{E}\left[F(z_{k+\frac{1}{2}}), \zeta_k\right]\right| \leqslant L(1-\alpha_k)\beta_k\sigma_k^2$$

The proof for Lemma 18 is analogous to the proof for Lemma D.2 in [LK21]. We refer the proof to [LK21].

The proof of Theorem 16 Consider the following Lyapunov function

$$\mathcal{E}(k) = A_k \left\|F(z_k)\right\|^2 + B_k \left\langle F(z_k), z_k - u_k\right\rangle + \frac{B_k - (1-\alpha_k)B_{k+1}}{2(1-\alpha_k)C_k}\left\|u_k - z^*\right\|^2, \tag{50}$$

where z^* is the zero-point of F, i. e. $F(z^*) = 0$. We show that $\{\mathbb{E}\left[\mathcal{E}(k)\right]\}$ is non-increasing.

Step 1: Divide the difference $\mathbb{E}\left[\mathcal{E}(k+1)\right] - \mathbb{E}\left[\mathcal{E}(k)\right]$ into three parts.

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{E}(k+1)\right] - \mathbb{E}\left[\mathcal{E}(k)\right] \\
=&\underbrace{A_{k+1}\mathbb{E}\left[\left\|F(z_{k+1})\right\|^2\right] - A_k\mathbb{E}\left[\left\|F(z_k)\right\|^2\right]}_{\mathrm{I}} \\
&+ \underbrace{B_{k+1}\mathbb{E}\left[\left\langle F(z_{k+1}), z_{k+1} - u_{k+1}\right\rangle\right] - B_k\mathbb{E}\left[\left\langle F(z_k), z_k - u_k\right\rangle\right]}_{\mathrm{II}} \\
&+ \underbrace{\frac{B_{k+1} - (1-\alpha_{k+1})B_{k+2}}{2(1-\alpha_{k+1})C_{k+1}}\mathbb{E}\left[\left\|u_{k+1} - z^*\right\|^2\right] - \frac{B_k - (1-\alpha_k)B_{k+1}}{2(1-\alpha_k)C_k}\mathbb{E}\left[\left\|u_k - z^*\right\|^2\right]}_{\mathrm{III}}
\end{aligned}$$

Step 2: Reckon the upper bound of all three parts. First we consider II.

$$\begin{aligned}
\mathrm{II} =&(B_{k+1} - B_k)\mathbb{E}\left[\left\langle F(z_{k+1}), z_{k+1} - u_{k+1}\right\rangle\right] + B_k\mathbb{E}\left[\left\langle F(z_{k+1}), z_{k+1} - z_k - u_{k+1} + u_k\right\rangle\right] \\
&+ B_k\mathbb{E}\left[\left\langle F(z_{k+1}) - F(z_k), z_k - u_k\right\rangle\right]
\end{aligned}$$

Since $\mathbb{E}\left[\zeta_{k+1}|z_{k+1}\right] = 0$, we have

$$\mathbb{E}\left[F(z_{k+1}), \zeta_{k+1}\right] = 0$$

Combining the above equation, (49c) and (49d), we have

$$\begin{aligned}
&B_k\mathbb{E}\left[\left\langle F(z_{k+1}), z_{k+1} - z_k - u_{k+1} + u_k\right\rangle\right] \\
=& B_k\mathbb{E}\left[\left\langle F(z_{k+1}), \frac{\alpha_k}{1-\alpha_k}(u_k - z_{k+1}) - \frac{\beta_k}{1-\alpha_k}[F(z_{k+\frac{1}{2}}) + \zeta_{k+\frac{1}{2}}]\right\rangle\right] + B_kC_k\mathbb{E}\left[\left\|F(z_{k+1})\right\|^2\right] \\
\leqslant& B_k\mathbb{E}\left[\left\langle F(z_{k+1}), \frac{\alpha_k}{1-\alpha_k}(u_k - z_{k+1}) - \frac{\beta_k}{1-\alpha_k}F(z_{k+\frac{1}{2}})\right\rangle\right] + \frac{B_kL\beta_k^2}{1-\alpha_k}\sigma_{k+\frac{1}{2}}^2 + B_kC_k\mathbb{E}\left[\left\|F(z_{k+1})\right\|^2\right]
\end{aligned} \tag{51}$$

In addition, because of (49b), we have

$$\begin{aligned}
&B_k\mathbb{E}\left[\left\langle F(z_{k+1}) - F(z_k), z_k - u_k\right\rangle\right] \\
=& B_k\mathbb{E}\left[\left\langle F(z_{k+1}) - F(z_k), -\frac{1}{\alpha_k}(z_{k+1} - z_k) - \frac{\beta_k}{\alpha_k}[F(z_{k+\frac{1}{2}}) + \zeta_{k+\frac{1}{2}}]\right\rangle\right] \\
\leqslant& B_k\mathbb{E}\left[\left\langle F(z_{k+1}) - F(z_k), -\frac{1}{\alpha_k}(z_{k+1} - z_k) - \frac{\beta_k}{\alpha_k}F(z_{k+\frac{1}{2}})\right\rangle\right] + \frac{B_kL\beta_k^2\sigma_{k+\frac{1}{2}}^2}{\alpha_k}
\end{aligned} \tag{52}$$

Also,

$$\begin{aligned}
&(B_{k+1}-B_k)\mathbb{E}\left[\langle F(z_{k+1}), z_{k+1}-u_{k+1}\rangle\right] \\
=\ &(B_{k+1}-B_k)C_k\mathbb{E}\left[\|F(z_{k+1})\|^2\right] + (B_{k+1}-B_k)\mathbb{E}\left[\langle F(z_{k+1}), z_{k+1}-u_k+C_k\zeta_{k+1}\rangle\right] \\
=\ &(B_{k+1}-B_k)C_k\mathbb{E}\left[\|F(z_{k+1})\|^2\right] + (B_{k+1}-B_k)\mathbb{E}\left[\langle F(z_{k+1}), z_{k+1}-u_k\rangle\right]
\end{aligned} \tag{53}$$

By summing (51), (52) and (53), we have

$$\begin{aligned}
\mathrm{II} \leqslant\ & B_k\mathbb{E}\left[\left\langle F(z_{k+1}), \frac{\alpha_k}{1-\alpha_k}(u_k-z_{k+1}) - \frac{\beta_k}{1-\alpha_k}F(z_{k+\frac{1}{2}})\right\rangle\right] \\
&+ B_k\mathbb{E}\left[\left\langle F(z_{k+1})-F(z_k), -\frac{1}{\alpha_k}(z_{k+1}-z_k) - \frac{\beta_k}{\alpha_k}F(z_{k+\frac{1}{2}})\right\rangle\right] + \left(\frac{B_k\beta_k}{1-\alpha_k}+\frac{B_k\beta_k}{\alpha_k}\right)L\beta_k\sigma_{k+\frac{1}{2}}^2 \\
&+ B_{k+1}C_k\mathbb{E}\left[\|F(z_{k+1})\|^2\right] + (B_{k+1}-B_k)\mathbb{E}\left[\langle F(z_{k+1}), z_{k+1}-u_k\rangle\right]
\end{aligned}$$

Since F is $L-$Lipschitz continuous, we have

$$\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2 \leqslant L^2\left\|z_{k+1}-z_{k+\frac{1}{2}}\right\|^2$$

Due to (49b) and (49c), we have

$$\left\|z_{k+1}-z_{k+\frac{1}{2}}\right\|^2 = \beta_k^2\left\|F(z_{k+\frac{1}{2}})+\zeta_{k+\frac{1}{2}}-(1-\alpha_k)[F(z_k)+\zeta_k]\right\|^2$$

By using Lemma 18, we have

$$\begin{aligned}
\mathbb{E}\left[\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2\right] \leqslant\ & L^2\beta_k^2\mathbb{E}\left[\left\|F(z_{k+\frac{1}{2}})\right\|^2\right] + 2L^2\beta_k^2(1-\alpha_k)\mathbb{E}\left[\left\langle F(z_{k+\frac{1}{2}}), F(z_k)\right\rangle\right] + L^2\beta_k^2\sigma_{k+\frac{1}{2}}^2 \\
&+ L^2\beta_k^2(1-\alpha_k)^2\sigma_k^2 + L^2\beta_k^2(1-\alpha_k)^2\mathbb{E}\left[\|F(z_k)\|^2\right] + 2L^3\beta_k^3(1-\alpha_k)^2\sigma_k^2
\end{aligned}$$

Because of the following equality

$$-\left\langle F(z_{k+1}), F(z_{k+\frac{1}{2}})\right\rangle = \frac{1}{2}\left[\left\|F(z_{k+1})-F(z_{k+\frac{1}{2}})\right\|^2 - \|F(z_{k+1})\|^2 - \left\|F(z_{k+\frac{1}{2}})\right\|^2\right]$$

and the assumption

$$\frac{B_k\beta_k}{\alpha_k} = \left(\frac{B_k\beta_k}{\alpha_k}+\frac{B_k\beta_k}{1-\alpha_k}\right)L^2\beta_k^2(1-\alpha_k)$$

we obtain the following estimation of II:

$$\begin{aligned}
\mathrm{II} \leqslant\ & -\left(\frac{B_k\beta_k}{2\alpha_k}+\frac{B_k\beta_k}{2-2\alpha_k}-B_{k+1}C_k\right)\mathbb{E}\left[\|F(z_{k+1})\|^2\right] - \left(\frac{B_k\beta_k}{2\alpha_k}+\frac{B_k\beta_k}{2-2\alpha_k}\right)(1-L^2\beta_k^2)\mathbb{E}\left[\left\|F(z_{k+\frac{1}{2}})\right\|^2\right] \\
&- \frac{B_k\beta_k}{2\alpha_k}\mathbb{E}\left[\|F(z_k)\|^2\right] - \frac{B_k}{1-\alpha_k}\mathbb{E}\left[\langle F(z_{k+1})-F(z_k), z_{k+1}-z_k\rangle\right] \\
&+ \underbrace{\left(B_{k+1}-\frac{B_k}{1-\alpha_k}\right)\mathbb{E}\left[\langle F(z_{k+1}), z_{k+1}-u_k\rangle\right]}_{\mathrm{II}_1} + \left(\frac{B_k\beta_k}{1-\alpha_k}+\frac{B_k\beta_k}{\alpha_k}\right)L\beta_k\sigma_{k+\frac{1}{2}}^2
\end{aligned}$$

$$+\left(\frac{B_k\beta_k}{2\alpha_k}+\frac{B_k\beta_k}{2-2\alpha_k}\right)[L^2\beta_k^2\sigma_{k+\frac{1}{2}}^2+L^2\beta_k^2(1-\alpha_k)^2\sigma_k^2+2L^3\beta_k^3(1-\alpha_k)^2\sigma_k^2]$$

Now estimate III. Utilizing (18), we have

$$\begin{aligned}
\mathrm{III} \leqslant\ & \frac{B_k-(1-\alpha_k)B_{k+1}}{2(1-\alpha_k)C_k}\mathbb{E}\left[\|u_{k+1}-z^*\|^2-\|u_k-z^*\|^2\right] \\
=\ & \underbrace{\frac{B_k-(1-\alpha_k)B_{k+1}}{1-\alpha_k}\mathbb{E}\left[\langle F(z_{k+1}),u_k-z^*\rangle\right]}_{\mathrm{III}_1}+\frac{C_kB_k-C_k(1-\alpha_k)B_{k+1}}{2-2\alpha_k}\left\{\mathbb{E}\left[\|F(z_{k+1})\|^2\right]+\sigma_{k+1}^2\right\}
\end{aligned}$$

Step 3: Deduce the upper bound of $\mathcal{E}(k+1)-\mathcal{E}(k)$. Noticing that $\mathrm{II}_1+\mathrm{III}_1=\frac{(1-\alpha_k)B_{k+1}-B_k}{1-\alpha_k}\langle F(z_{k+1}),z_{k+1}$ - we can simplify the upper bound of $\mathbb{E}\left[\mathcal{E}(k+1)\right]-\mathbb{E}\left[\mathcal{E}(k)\right]$ as follows:

$$\begin{aligned}
& \mathbb{E}\left[\mathcal{E}(k+1)\right]-\mathbb{E}\left[\mathcal{E}(k)\right] \\
\leqslant\ & -\left(\frac{B_k\beta_k}{2\alpha_k}+\frac{B_k\beta_k}{2-2\alpha_k}-B_{k+1}C_k-\frac{C_kB_k-C_k(1-\alpha_k)B_{k+1}}{2-2\alpha_k}-A_{k+1}\right)\mathbb{E}\left[\|F(z_{k+1})\|^2\right] \\
& -\left(\frac{B_k\beta_k}{2\alpha_k}+\frac{B_k\beta_k}{2-2\alpha_k}\right)(1-L^2\beta_k^2)\mathbb{E}\left[\left\|F(z_{k+\frac{1}{2}})\right\|^2\right]-\left(\frac{B_k\beta_k}{2\alpha_k}-A_k\right)\mathbb{E}\left[\|F(z_k)\|^2\right] \\
& -\frac{B_k}{1-\alpha_k}\mathbb{E}\left[\langle F(z_{k+1})-F(z_k),z_{k+1}-z_k\rangle\right]+\left(B_{k+1}-\frac{B_k}{1-\alpha_k}\right)\mathbb{E}\left[\langle F(z_{k+1}),z_{k+1}-z^*\rangle\right] \\
& +\left(\frac{B_k\beta_k}{1-\alpha_k}+\frac{B_k\beta_k}{\alpha_k}\right)L\beta_k\sigma_{k+\frac{1}{2}}^2+\frac{C_kB_k-C_k(1-\alpha_k)B_{k+1}}{2-2\alpha_k}\sigma_{k+1}^2 \\
& +\left(\frac{B_k\beta_k}{2\alpha_k}+\frac{B_k\beta_k}{2-2\alpha_k}\right)[L^2\beta_k^2\sigma_{k+\frac{1}{2}}^2+L^2\beta_k^2(1-\alpha_k)^2\sigma_k^2+2L^3\beta_k^3(1-\alpha_k)^2\sigma_k^2]
\end{aligned}$$

$\square$