Accelerated Algorithms for Constrained Nonconvex-Nonconcave Min-Max Optimization and Comonotone Inclusion

Yang Cai [* 1] **Argyris Oikonomou** [* 1] **Weiqiang Zheng** [* 1]

Abstract

We study constrained comonotone min-max optimization, a structured class of nonconvex-nonconcave min-max optimization problems, and their generalization to comonotone inclusion. In our first contribution, we extend the *Extra Anchored Gradient* (EAG) algorithm, originally proposed by Yoon & Ryu (2021) for unconstrained min-max optimization, to constrained comonotone min-max optimization and comonotone inclusion, achieving an optimal convergence rate of $O\left(\frac{1}{T}\right)$ among all first-order methods. Additionally, we prove that the algorithm's iterations converge to a point in the solution set. In our second contribution, we extend the *Fast Extra Gradient* (FEG) algorithm, as developed by Lee & Kim (2021), to constrained comonotone min-max optimization and comonotone inclusion, achieving the same $O\left(\frac{1}{T}\right)$ convergence rate. This rate is applicable to the broadest set of comonotone inclusion problems yet studied in the literature. Our analyses are based on simple potential function arguments, which might be useful for analyzing other accelerated algorithms.

1. Introduction

Min-max optimization is a cornerstone in game theory, optimization, and online learning. While classical theoretical studies have primarily focused on the convex-concave case,[1]

there has been growing interest in nonconvex-nonconcave min-max optimization problems within the machine learning and optimization community. This surge in attention is attributed to a range of innovative applications of nonconvex-nonconcave min-max such as generative adversarial networks (GANs) (see (Goodfellow et al., 2014; Arjovsky et al., 2017)), adversarial examples (see (Madry et al., 2017)), robust optimization (see (Ben-Tal et al., 2009)), and reinforcement learning (see (Du et al., 2017; Dai et al., 2018)).

Unfortunately, even finding a first-order stationary point is generally intractable for nonconvex-nonconcave min-max optimization problems (Daskalakis et al., 2021). In light of this obstacle, recent research has shifted focus towards classes of nonconvex-nonconcave min-max optimization problems that exhibit more structure. One class that has received extensive attention is the comonotone min-max optimization problem proposed by Bauschke et al. (2021) (see Example 1 for the definition), which captures the convex-concave setting as a special case.

During the past few years, we have witnessed a series of exciting advancements in comonotone min-max optimization (Diakonikolas et al., 2021; Pethick et al., 2022; Lee & Kim, 2021; Yoon & Ryu, 2022; Sedlmayer et al., 2023; Cai & Zheng, 2023a;b; Tran-Dinh, 2023; Gorbunov et al., 2023). However, the algorithms proposed in these studies often exhibit one or more of the following limitations: (a) they suffer a *sub-optimal* convergence rate, (b) they only apply to the *unconstrained setting* or (c) they *lack point convergence guarantees*.[2] Addressing these challenges, the primary goal of our paper is to design algorithms that simultaneously surmount all three of these limitations. More specifically, we seek to answer the following question:

> **Question:** Can we design algorithms for comonotone min-max optimization that have (i) the **optimal convergence rate** in the **constrained setting** and (ii) **point convergence** to the solution set?

We provide an affirmative answer to this question. In particular, we design algorithms that exhibit the optimal convergence rate of $O(1/T)$ in the constrained setting for comonotone min-max optimization, and our algorithms are guar-

---

*Authors are listed in alphabetical order [1]Department of Computer Science, Yale University, New Haven, USA. Correspondence to: Yang Cai <yang,cai@yale.edu>, Argyris Oikonomou <argyris.oikonomou@yale.edu>, Weiqiang Zheng <weiqiang.zheng@yale.edu>.

[1]For closed convex sets $\mathcal{X}$ and $\mathcal{Y}$, and a smooth function $f(\cdot, \cdot)$, the corresponding min-max optimization problem is formulated as: $\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y)$,. A min-max optimization problem is convex-concave when additionally, function $f(\cdot, \cdot)$ is convex in x and concave y.

[2]An algorithm has point convergence if its iterates has a limit.

anteed to converge to a point within the solution set. See Table 1 for comparison between our algorithms and the ones proposed in the literature. Indeed, our findings extend to the more general problem of *Comonotone Inclusion*. More specifically, we consider composite inclusion problems that involve a single-valued and Lipschitz continuous operator $F : \mathbb{R}^n \to \mathbb{R}^n$, and a maximally monotone set-valued operator $A : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$.[3] The goal is to find a point $z^* \in \mathbb{R}^n$ such that $\mathbf{0} \in F(z^*) + A(z^*)$. If the composite operator satisfies the ρ-comonotonicity condition:

$$\langle u - u', z - z' \rangle \geq \rho \|u - u'\|^2,$$
$$\forall z, z' \in \mathbb{R}^n \text{ and } \forall u \in F(z) + A(z), u' \in F(z') + A(z'),$$

we refer to these problems as Comonotone Inclusion Problems (CMI). It is not hard to see that both the unconstrained comonotone min-max optimization and the monotone variational inequality fall under the umbrella of CMI. The general formulation of CMI enables us to capture more complex settings, such as constrained or non-smooth min-max optimization. For a more detailed discussion, see Section 2.

1.1. Our Contributions

In this paper, we evaluate the quality of a solution by measuring the norm of the composite operator. Specifically, a point $z \in \mathbb{R}^n$ is an ϵ-approximate solution if $\mathbf{0} \in F(z) + A(z) + \mathcal{B}(\mathbf{0}, \epsilon)$, where $\mathcal{B}(\mathbf{0}, \epsilon)$ is the ball with radius ϵ centered at $\mathbf{0}$. As discussed in Section 2.2, this criterion is equivalent to the tangent residual of z, a notion introduced in (Cai et al., 2022), being no more than ϵ.

First, we extend the Extra Anchored Gradient algorithm (composite-EAG), originally proposed by Yoon & Ryu (2021) for unconstrained convex-concave min-max problems, to solve ρ-comonotone composite inclusion problems.

> **Contribution 1:** We show that the T-th iterate of (composite-EAG) is an $O(\frac{1}{T})$-approximate solution to CMI if (i) the single-valued operator F is L-Lipschitz, (ii) the set-valued operator A is maximally monotone, and (iii) the composite operator $F + A$ is $\rho > -\frac{1}{20L}$-comonotone. Additionally, our algorithm ensures point convergence to a point within the solution set when condition (i), (ii) and (iii) hold.

Next, we extend the Fast Extra-Gradient (FEG) algorithm Lee & Kim (2021) originally designed for unconstrained comonotone convex-concave min-max optimization to comonotone inclusion problems.

> **Contribution 2:** We show that the T-th iterate of (composite-FEG) is an $O(\frac{1}{T})$-approximate solution to CMI if (i) the single-valued operator F is L-Lipschitz, (ii) the set-valued operator A is maximally monotone, and (iii) the composite operator $F + A$ is $\rho > -\frac{1}{2L}$-comonotone. Additionally, our algorithm is guaranteed point convergence if (i) and (ii) hold, and the composite operator $F + A$ is monotone, i.e., 0-comonotone.

Note that the convergence rate of our algorithms composite-EAG and composite-FEG matches the lower bound by (Diakonikolas, 2020; Yoon & Ryu, 2021), and is therefore optimal for any first-order method. Some further remarks are in order.

- Our algorithms composite-EAG and composite-FEG are the *first* first-order algorithms with optimal $O\left(\frac{1}{T}\right)$ convergence rates for the constrained and negatively comonotone setting. We remark that Kovalev & Gasnikov (2022) concurrently propose composite-EAG and prove its convergence in the constrained and monotone setting.

- Our algorithm composite-FEG achieves the optimal convergence rate in the constrained and negatively monotone setting with $\rho > -\frac{1}{2L}$, which represents the widest range of ρ values among all single-loop algorithms.[4]

- We are the *first* to provide first-order methods that achieve both optimal convergence rates and point convergence in the constrained and negatively comonotone setting.

1.2. Related Work

There is a vast literature on inclusion problems and variational inequalities. See (Facchinei & Pang, 2003; Bauschke & Combettes, 2011; Ryu & Boyd, 2016) and the references therein. We provide a brief discussion of the most relevant results here and postpone the rest to Appendix A.

Accelerated Convergence Rate for Convex-Concave Min-Max Optimization and Monotone Variational Inequalities. We overview results that achieve the $O(\frac{1}{T})$ accelerated convergence rate in terms of the operator norm or residual. Note that these results imply $O(\frac{1}{T})$ convergence rate in terms of the gap function. We describe these results in the language of inclusion problems.

Recent results show accelerated rates through Halpern iteration (Halpern, 1967) or a similar mechanism – anchoring.

---

[3]The powerset of $\mathbb{R}^n$ is denoted as $\mathcal{P}(\mathbb{R}^n)$.

[4]Very recently, some works propose algorithms that work for a wider range of $\rho > -\frac{1}{L}$ in the unconstrained (Cai et al., 2024; Fan et al., 2024) and constrained setting (Alacaoglu et al., 2024). However, these algorithms are *double-loop* and are sub-optimal by $\log T$ in terms of the convergence rates. It is an open question to design a single-loop algorithm that works for $\rho > -\frac{1}{L}$ with no spurious log factors.

Table 1. Existing results for comonotone inclusion problem. The convergence rate is in terms of the operator norm (in the unconstrained setting) and the residual (in the constrained setting). (†): The $O(\frac{1}{\sqrt{T}})$ convergence rates hold under the weak Minty assumption, which is implied by the comonotonicity assumption.

Algorithm	Rate	Range of ρ	Constraints	Point Convergence
EG+ (Diakonikolas et al., 2021) (Gorbunov et al., 2023)†	$O(\frac{1}{\sqrt{T}})$	$\left(-\frac{1}{2L}, +\infty\right)$	✗	✓
CEG+ (Pethick et al., 2022; 2023)†	$O(\frac{1}{\sqrt{T}})$	$\left(-\frac{1}{2L}, +\infty\right)$	✓	✓
Double-Loop Halpern (Diakonikolas, 2020) (Kohlenbach, 2022)	$O(\frac{\log(T)}{T})$	$\left(-\frac{1}{2L}, +\infty\right]$	✓	Unknown
APG* (Yoon & Ryu, 2022)	$O(\frac{\log(T)}{T})$	$[0, +\infty]$	✓	✓
EAG (Yoon & Ryu, 2021)	$O(\frac{1}{T})$	$[0, +\infty]$	✗	✓
FEG (Lee & Kim, 2021)	$O(\frac{1}{T})$	$\left(-\frac{1}{2L}, +\infty\right)$	✗	✓(for $\rho \geq 0$)
ARG (Cai & Zheng, 2023b)	$O(\frac{1}{T})$	$\left[-\frac{1}{60L}, +\infty\right)$	✓	Unknown
AOG (Cai & Zheng, 2023a)	$O(\frac{1}{T})$	$[0, +\infty]$	✓	Unknown
fOGDA (Sedlmayer et al., 2023)	$O(\frac{1}{T})$	$[0, +\infty]$	✓	✓
composite-EAG [**This paper**]	$O(\frac{1}{T})$	$\left[-\frac{1}{20L}, +\infty\right)$	✓	✓
composite-FEG [**This paper**]	$O(\frac{1}{T})$	$\left(-\frac{1}{2L}, +\infty\right)$	✓	✓(for $\rho \geq 0$)

Implicit versions of Halpern iteration have $O(\frac{1}{T})$ convergence rate (Kim, 2021; Lieder, 2021; Park & Ryu, 2022) for monotone operators and explicit variants of Halpern iteration achieve the same convergence rate when F is co-coercive and A is the subdifferential of the indicator function of the feasible set (Diakonikolas, 2020; Kim, 2021). Diakonikolas (2020) also provides a double-loop implementation of the algorithm for monotone operators at the expense of an additional logarithmic factor in the convergence rate. Yoon & Ryu (2021) proposes the extra anchored gradient (EAG) method, which is the first explicit method with an accelerated $O(\frac{1}{T})$ rate in the unconstrained setting for monotone operators, i.e., F is monotone and $A = 0$.[5]. They also establish a matching $\Omega(\frac{1}{T})$ lower bound that holds for all first-order methods. Convergence analysis of the past extragradient method with anchoring in the unconstrained setting is provided by Tran-Dinh & Luo (2021). More recently, Tran-Dinh (2022) studies the connection between Halpern iteration and Nesterov's accelerated method and provides new algorithms for monotone operators and alternative analyses for EAG in the unconstrained setting. Sedlmayer et al. (2023) analyzes the fast optimistic gradient descent ascent (fOGDA) and shows that the algorithms achieve $O(\frac{1}{T})$ convergence rate for constrained monotone variational inequalities, i.e., F is monotone and A is the subdifferential of the indicator function of the feasible set. Cai & Zheng (2023a) proposes an anchored variant of optimistic

gradient descent ascent (AOG) for monotone games, and their analysis extends to constrained monotone variational inequalities.

Accelerated Convergence Rate for Comonotone Min-Max Optimization and Comonotone Inclusion. For unconstrained comonotone min-max optimization problems, i.e., F is comonotone and $A = 0$, Lee & Kim (2021) proposes a generalization of EAG called fast extra-gradient (FEG), which achieves $O(\frac{1}{T})$ convergence rate. For constrained comonotone min-max optimization problems, convergence of *implicit methods* such as the proximal point algorithm and the Halpern iteration has been analyzed (Bauschke et al., 2021; Kohlenbach, 2022). Moreover, the explicit double-loop variants of the Halpern iteration achieves $O(\frac{\log T}{T})$ convergence rates (Diakonikolas, 2020; Kohlenbach, 2022). Cai & Zheng (2023b) analyzes the anchored variant of Reflected Gradient (ARG) and shows a $O(\frac{1}{T})$ fast convergence rate for a subset of comonotone inclusion problems covered by our algorithms.[6] It is unknown whether their algorithm has point convergence.

Point Convergence in the Convex-Concave and Monotone Setting. For unconstrained convex-concave min-max optimization problems, Yoon & Ryu (2022) shows that a range of fast algorithms, namely EAG (Yoon & Ryu, 2021), FEG (Lee & Kim, 2021) and anchored Popov's scheme

---

[5]We use $A = 0$ to denote the set operator that maps any point to $\emptyset$

[6]See the paragraph of "Preliminary Version of this Paper" for discussion between their results and ours.

(Tran-Dinh & Luo, 2021), all converge to the Optimized Halpern Iteration (Lieder, 2021; Kim, 2021) – an implicit method. Consequentially, this implies that all aforementioned methods have point convergence, as the Optimized Halpern Iteration has point convergence. Additionally, Sedlmayer et al. (2023) shows that the fOGDA has point convergence to the solution set for constrained monotone variational inequalities.

Preliminary Version of this Paper. We would like to mention that the initial version of this paper was first made available online in June 2022. This early version included an algorithm similar to composite-EAG that is capable of achieving an $O(\frac{1}{T})$ accelerated convergence rate for CMI. In concurrent work, Kovalev & Gasnikov (2022) propose composite-EAG and prove its convergence for CMI with $\rho \geq 0$. The current version not only includes this result but also introduces a novel point convergence analysis. Note that Sedlmayer et al. (2023); Cai & Zheng (2023a;b) appeared after our manuscript was available online. In particular, the analyses used in Cai & Zheng (2023a;b) are similar to the one first proposed in this paper, building upon certain key lemmas and inequalities established here. Tran-Dinh (2023) provides an alternative proof for the convergence rates of our algorithm composite-FEG but does not provide any point convergence guarantees.[7]

We summarize all results discussed in this section and our results in Table 1.

2. Preliminaries

We consider the Euclidean Space $(\mathbb{R}^n, \|\cdot\|)$, where $\|\cdot\|$ is the ℓ_2 norm and $\langle\cdot,\cdot\rangle$ denotes inner product on $\mathbb{R}^n$.

Basic Notions about Operators. A set-valued operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ maps $z \in \mathbb{R}^n$ to a subset $A(z) \subseteq \mathbb{R}^n$. We say A is *single-valued* if $|A(z)| = 1$ for all $z \in \mathbb{R}^n$. The *graph* of an operator A is defined as $\text{Gra}_A = \{(z, u) : z \in \mathbb{R}^n, u \in A(z)\}$. The *inverse operator* of A is denoted as A^{-1} whose graph is $\text{Gra}_{A^{-1}} = \{(u, z) : (z, u) \in \text{Gra}_A\}$. We also denote $\text{Zer}(A) = \{z \in \mathbb{R}^n : 0 \in A(z)\}$. For two operators A and B, we denote $A + B$ as the operator with graph $\text{Gra}_{A+B} = \{(z, u_A + u_B) : (z, u_A) \in \text{Gra}_A, (z, u_B) \in \text{Gra}_B\}$. We denote the identity operator as $I : \mathbb{R}^n \to \mathbb{R}^n$. For $L \in (0, \infty)$, a single-valued operator $A : \mathbb{R}^n \to \mathbb{R}^n$ is *L-Lipschitz* if

$$\|A(z) - A(z')\| \leq L \cdot \|z - z'\|, \quad \forall z, z' \in \mathbb{R}^n.$$

Moreover, A is *non-expansive* if it is 1-Lipschitz. For a non-expansive operator A, we denote its fix points $\text{Fix}(A) = \{z \in \mathbb{R}^n : z = A(z)\}$. For a closed convex set $\mathcal{Z} \subseteq \mathbb{R}^n$

and point $z \in \mathbb{R}^n$, we denote the *normal cone operator* as $N_{\mathcal{Z}}$:

$$N_{\mathcal{Z}}(z) = \begin{cases} \emptyset, & z \notin \mathcal{Z}, \\ \{v \in \mathbb{R}^n : \langle v, z' - z \rangle \leq 0, \forall z' \in \mathcal{Z}\}, & z \in \mathcal{Z}. \end{cases}$$

Define the indicator function

$$\mathbb{I}_{\mathcal{Z}}(z) = \begin{cases} 0 & \text{if } z \in \mathcal{Z}, \\ +\infty & \text{otherwise}. \end{cases}$$

Then it is not hard to see that the subdifferential operator $\partial \mathbb{I}_{\mathcal{Z}} = N_{\mathcal{Z}}$. The projection operator $\Pi_{\mathcal{Z}} : \mathbb{R}^n \to \mathbb{R}^n$ is defined as $\Pi_{\mathcal{Z}}[z] := \text{argmin}_{z' \in \mathcal{Z}} \|z - z'\|^2$.

(Co)monotone Operators For $\rho \in \mathbb{R}$, an operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *ρ-comonotone* (Bauschke et al., 2021),[8] if

$$\langle u - u', z - z' \rangle \geq \rho \|u - u'\|^2, \quad \forall (z, u), (z', u') \in \text{Gra}_A.$$

If A is 0-comonotone, then A is *monotone*. If A is ρ-comonotone for $\rho > 0$, we also say A is *ρ-cocoercive* (a stronger assumption than monotonicity). When A satisfies negative comonotonicity, i.e., ρ-comonotonicity with $\rho < 0$, then A is possibly non-monotone. Negative comonotonicity is the focus of this paper in the non-monotone setting.

A is *maximally ρ-comonotone* if A is ρ-comonotone and there is no other ρ-comonotone operator B such that $\text{Gra}_A \subset \text{Gra}_B$ strictly. A is maximally monotone if it is maximally 0-comonotone. When $f : \mathbb{R}^n \to \mathbb{R}$ is a convex closed proper function, then its subdifferential operator ∂f is maximally monotone. As an example, $\partial \mathbb{I}_{\mathcal{Z}} = N_{\mathcal{Z}}$ is maximally monotone.

Resolvent Operators We denote the **resolvent** of an operator A as $J_A := (I + A)^{-1}$.

Proposition 2.1. *(Rockafellar, 1976; Ryu & Boyd, 2016; Bauschke & Combettes, 2011; Ryu & Yin, 2022) When A is maximally monotone, its resolvent $J_{\eta A}$ has the following properties: for any $\eta > 0$,*

1. *$J_{\eta A}$ is well-defined on $\mathbb{R}^n$ and non-expansive;*

2. *if $z = J_{\eta A}(z')$, $\frac{z'-z}{\eta} \in A(z)$;*

3. *when $A = N_{\mathcal{Z}}$ is the normal cone operator of a closed convex set $\mathcal{Z} \subseteq \mathbb{R}^n$, $J_{\eta A} = \Pi_{\mathcal{Z}}$ is the projection operator.*

2.1. Inclusion Problems with Negatively Comonotone Operators.

Given a single-valued and possibly *non-monotone* operator F and a *set-valued maximally monotone* operator A, we

---

[7]composite-FEG was proposed in the second version of this paper, which appeared online in August 2022.

[8]when $\rho < 0$, ρ-comonotonicity is also known as $|\rho|$-cohypomonotonicity (Combettes & Pennanen, 2004).

denote $E = F + A$. The comonotone inclusion problem is to find a point $z^* \in \mathbb{R}^n$ that satisfies

$$\mathbf{0} \in E(z^*) = F(z^*) + A(z^*). \qquad \text{(CMI)}$$

We say z is an ϵ-approximate solution to CMI if $\mathbf{0} \in F(z) + A(z) + \mathcal{B}(\mathbf{0}, \epsilon)$. We summarize the assumptions on CMI below.

Assumption 2.2. In CMI,

1. $F : \mathbb{R}^n \to \mathbb{R}^n$ is L-Lipschitz.

2. $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximally monotone.

3. $E = F + A$ is ρ-comonotone, i.e., there exists $\rho \leq 0$ such that
$$\langle u - u', z - z' \rangle \geq \rho \|u - u'\|^2, \forall (z, u), (z', u') \in \text{Gra}_E \,.$$

4. There exists a solution $z^* \in \mathbb{R}^n$ such that $\mathbf{0} \in E(z^*)$.

The formulation of CMI provides a unified treatment for a range of problems including variational inequalities (see Appendix B.1), min-max optimization, and multi-player games. We present one detailed example below and refer readers to (Facchinei & Pang, 2003) for more examples.

Example 1 (Min-Max Optimization). The following structured min-max optimization problem captures a wide range of applications in machine learning such as GANs, adversarial examples, robust optimization, and reinforcement learning:

$$\min_{x \in \mathbb{R}^{n_x}} \max_{y \in \mathbb{R}^{n_y}} f(x, y) + g(x) - h(y), \qquad (1)$$

where $f(\cdot, \cdot)$ is possibly non-convex in x and non-concave in y. Regularized and constrained min-max problems are covered by appropriate choices of lower semicontinuous and convex functions g and h. Examples include ℓ_1-norm, ℓ_2-norm, and indicator function of a convex feasible set. Let $z = (x, y)$, we define $F(z) = (\partial_x f(x, y), -\partial_y f(x, y))$ and $A(z) = (\partial g(x), \partial h(y))$. If F and A satisfies Assumption 2.2, then we call it a *comonotone min-max optimization* problem. See (Lee & Kim, 2021, Example 1) for examples of nonconvex-nonconcave conditions that imply negative comonotonicity. We also note that the resolvent of ∂g is also known as the *proximal operator* $\text{prox}_{\eta g} = (I + \lambda \eta \partial g)^{-1}$. The resolvent $J_{\eta \partial g}$ is efficiently computatable for g being the ℓ_1-norm, ℓ_2-norm, matrix norms, an indicator function for a convex feasible set, among many others (see (Parikh et al., 2014, Ch 6, 7) for a comprehensive overview of proximal operators and their efficient computation).

2.2. Convergence Criteria

An appropriate convergence criterion is the *tangent residual* $r_{F,A}^{tan}(z) := \min_{c \in A(z)} \|F(z) + c\|$, which is an extension of the operator norm $\|F(z)\|$ in the unconstrained setting ($A = 0$) to the composite setting. This is a natural quantity for inclusion problems, as $r_{F,A}^{tan}(z) \leq \epsilon$ is equivalent to z being an ϵ-approximate solution to CMI. In this paper, following (Cai et al., 2022), we refer to this quantity as the "tangent residual". Another commonly-used convergence criterion that captures the stationarity of a solution is the *natural residual* $r_{F,A}^{nat} := \|z - J_A[z - F(z)]\|$. Note that z^* is a solution to CMI iff $z^* = J_A[z^* - F(z^*)]$. The following fact shows that natural residual is upper bounded by the tangent residual (see proof in Appendix B.2).

Fact 1. In CMI, $r_{F,A}^{nat}(z) \leq r_{F,A}^{tan}(z)$.

In this paper, we state our convergence rates in terms of the tangent residual $r_{F,A}^{tan}(z)$, which implies (i) convergence rates in terms of the natural residual $r_{F,A}^{nat}(z)$, and (ii) z is an approximate solution to CMI and the variational inequality problems including SVI and MVI (see Definitions in Appendix B.1).

3. composite-EAG for Comonotone Inclusion with Point Convergence

We study CMI under Assumption 2.2 with ρ-comonotone operators. The extra anchored gradient (EAG) algorithm (Yoon & Ryu, 2021) has been shown to have optimal last-iterate convergence rate and point convergence (Yoon & Ryu, 2022). However, these results only hold when (i) $A = 0$, i.e., the *unconstrained* setting and (ii) F is *monotone* ($\rho \geq 0$). The state-of-the-art work of Sedlmayer et al. (2023) achieves similar results in the constrained setting, i.e., $A = \partial \mathbb{I}_S$ for the feasible set S, but still requires the monotonicity assumption on F. Thus the problem of achieving $O(1/T)$ convergence rate and point convergence without the monotonicity assumption is open even in the unconstrained case. In this section, we improve both results from (Yoon & Ryu, 2022; Sedlmayer et al., 2023) by proposing the composite extra anchored gradient algorithm composite-EAG that achieves both the optimal $O(1/T)$ last-iterate convergence rate (Theorem 3.2) and point convergence (Theorem 3.8) for CMI with $\rho \geq -\frac{1}{20L}$.

composite-EAG We propose *composite extra anchored gradient (composite-EAG)* algorithm (Algorithm 1). Specifically, given an arbitrary z_0 and step size $\eta > 0$, composite-EAG first set $z_1 = J_{\eta A}[z_0 - \eta F(z_0)]$ and $c_1 = \frac{z_0 - \eta F(z_0) - z_1}{\eta}$; then for any $k \geq 1$, it updates as shown in Algorithm 1. Recall that by definition of the resolvent $J_{\eta A}$, we have $c_k \in A(z_k)$. All missing proofs in this section can be found in Appendix C.

Remark 3.1. A keen reader may notice that in the important special case of constrained variational inequality problem where $A = N_\mathcal{Z}$ is the normal cone of a convex set $\mathcal{Z} \subseteq \mathbb{R}^n$, the $z_{k+\frac{1}{2}}$ produced by composite-EAG may not lie in the

---

Algorithm 1 composite-EAG for CMI with $\rho > -\frac{1}{2L}$

Initialization: $z^0 \in \mathbb{R}^n$. Step size $\eta > 0$. $\beta_k = \frac{1}{k+1}$ for $k \geq 0$. Set

$$z_1 = J_{\eta A}[z_0 - \eta F(z_0)], \quad c_1 = \frac{z_0 - \eta F(z_0) - z_1}{\eta}$$

for $k \geq 1$ **do**

$$z_{k+\frac{1}{2}} = \beta_k z_0 + (1 - \beta_k)z_k - \eta F(z_k) - \eta c_k$$
$$z_{k+1} = J_{\eta A}[\beta_k z_0 + (1 - \beta_k)z_k - \eta F(z_{k+\frac{1}{2}})] \qquad \text{(composite-EAG)}$$
$$c_{k+1} = \frac{\beta_k z_0 + (1 - \beta_k)z_k - \eta F(z_{k+\frac{1}{2}}) - z_{k+1}}{\eta}$$

end for

---

set $\mathcal{Z}$. In Appendix E, we present proj-EAG, a variant of composite-EAG, designed to consistently produce iterates within the set $\mathcal{Z}$. Furthermore, proj-EAG preserves the $O(1/T)$ convergence rate alongside point convergence.

3.1. Technical Novelty

We illustrate the main technical challenges and techniques developed in extending the algorithms and their convergence guarantees to the composite and comonotone settings, which may also be applicable to analyzing other algorithms.

Constraints. One key challenge is to design the correct update rule of *composite-EAG* in the composite setting. Standard approaches, such as adding two backward steps (i.e., projections) to EAG, do not yield algorithms that admit simple analysis. In composite-EAG, we introduce a new element $c_k \in A(z_k)$ to ensure that $F(z_k) + c_k \in (F + A)(z_k)$. If we consider $F + A$ as a single operator, the update rule of composite-EAG simplifies to that of EAG in the unconstrained setting. This property enables us to smoothly extend the analysis from the unconstrained setting to the composite setting. The design of composite-EAG also appears in the concurrent work of Kovalev & Gasnikov (2022). Our approach to handling constraints has inspired subsequent work by Sedlmayer et al. (2023), which extends fast OGDA from the unconstrained to the constrained setting.

Negative Comonotonicity. Negative comonotonicity introduces an additional quadratic term in the analysis of composite-EAG, and it is unclear how to cancel this term. We address it through a weighted sum of terms, employing Young's inequality and a carefully balanced set of coefficients. Although this approach enables our analysis, it results in a more restrictive range of $\rho \geq -\frac{1}{20L}$ compared to composite-FEG.

3.2. $O(1/T)$ **Last-Iterate Convergence Rate**

In this section, we show that composite-EAG has $O(1/T)$ last-iterate convergence rate with respect to the tangent residual, which matches the lower bound for any first-order methods (Diakonikolas, 2020; Yoon & Ryu, 2021).

Theorem 3.2 (Last-iterate convergence rate of composite-EAG). *Consider a comonotone inclusion problem with Assumption 2.2 holds with $\rho \geq -\frac{1}{20L}$. Let $\eta = \frac{0.31}{L}$ and $\{z_k, z_{k+\frac{1}{2}}\}_{k\geq 1}$ the iterates generated by composite-EAG. Then the following holds for any $T \geq 1$,*

$$r_{F,A}^{tan}(z_T)^2 \leq \|F(z_T) + c_T\|^2 \leq \frac{20H^2}{\eta^2 T^2},$$

where $H^2 := 6\|z_1 - z_0\|^2 + \|z_0 - z^*\|^2 \leq 6\eta^2 r_{F,A}^{tan}(z_0)^2 + \|z_0 - z^*\|^2$.

Proof Sketch We apply a potential function argument. Recall the definition of the tangent residual $r_{F,A}^{tan}(z) := \min_{c \in A(z)} \|F(z) + c\|$, which involves an optimization problem and can be hard to analyze directly. Since $c_k \in A(z_k)$, we have $\|F(z_k) + c_k\| \geq r_{F,A}^{tan}(z_k)$ and $\|F(z_k) + c_k\|$ can be used as a proxy for the tangent residual. We construct a potential function V_k that is of the order $\Omega(k^2 \cdot r_{F,A}^{tan}(z_k)^2)$ (Lemma 3.5). Although the potential function may not always decrease, we manage to prove that the increment in each iteration k is sufficiently small: $V_{k+1} - V_k \leq O(1) \cdot \|F(z_k) + c_k\|^2$ in Theorem 3.4. We then apply this *approximately non-increasing* property of the potential function to conclude $O(1/T)$ last-iterate convergence rate of $\|F(z_k) + c_k\|$ and thus $r_{F,A}^{tan}(z_k)$.

3.2.1. PROOF OF THEOREM 3.2

Potential function We formally define the potential function V_k for $k \geq 1$ as $V_k = \frac{k(k+1)}{2}\|\eta F(z_k) + \eta c_k\|^2 + k\langle \eta F(z_k) + \eta c_k, z_k - z_0 \rangle$. A bound on V_1 is immediate.

Lemma 3.3 (Upper bound of V_1). *In the same setup as Theorem 3.2, we have* $V_1 \le 6\|z_1 - z_0\|^2 \le 6\eta^2 r_{F,A}^{tan}(z_0)^2$

The following result that proves that V_k is approximately non-increasing is the core of the analysis.

Theorem 3.4 (Approximately non-increasing potential). *In the same setup as Theorem 3.2, we have for all $k \ge 1$,*

$$V_{k+1} - V_k$$
$$\le \frac{1}{2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2 - \frac{9k(k+1)}{4000}\left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2.$$

The following lemma shows that V_k is of order $\Omega(k^2 \cdot r^{tan}(z_k)^2)$.

Lemma 3.5. *In the same setup as Theorem 3.2, for any $k \ge 1$, we have*

$$\frac{k(k+1)}{4}\|\eta F(z_k) + \eta c_k\|^2 \le V_k + \|z_0 - z^*\|^2$$

In particular, $V_k \ge -\|z_0 - z^*\|^2$.

Proof of Theorem 3.2 Recall that $H^2 \ge V_1 + \|z_0 - z^*\|^2$. For $k = 1$, from Lemma 3.5, we directly get that $\|\eta F(z_1) + \eta c_1\|^2 \le 4(V_1 + \|z_0 - z^*\|^2)$.

Now fix any $k \ge 2$. Combining Theorem 3.4 and Lemma 3.5, we have

$$\frac{k(k+1)}{4}\|\eta F(z_k) + \eta c_k\|^2$$
$$\le V_k + \|z_0 - z^*\|^2$$
$$\le V_1 + \|z_0 - z^*\|^2 + \frac{1}{2} \cdot \sum_{t=2}^{k} \|\eta F(z_t) + \eta c_t\|^2.$$

Subtracting $\frac{1}{2}\|F(z_k) + \eta c_k\|^2$ from both sides and noting that $\frac{1}{4}k \ge \frac{1}{2}$ for $k \ge 2$ gives

$$\frac{k^2}{4}\|\eta F(z_k) + \eta c_k\|^2$$
$$\le V_1 + \|z_0 - z^*\|^2 + \frac{1}{2} \cdot \sum_{t=2}^{k-1} \|\eta F(z_t) + \eta c_t\|^2.$$

Now we can apply Proposition F.4 with $C_1 = V_1 + \|z_0 - z^*\|^2$ to conclude that $\|\eta F(z_k) + \eta c_k\|^2 \le \frac{20(V_1 + \|z_0 - z^*\|^2)}{k^2} \le \frac{20H^2}{k^2}$.

Corollary 3.6. *In the same setup as Theorem 3.4, we have*

$$\sum_{k=1}^{\infty} k(k+1)\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2 \le 4000H^2.$$

It also implies $\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2 \le \frac{8000H^2}{(k+1)^2}$ for all $k \ge 1$.

3.3. Point Convergence of composite-EAG

To further show point convergence of composite-EAG, we identify a *merging path (MP)* property, a notion proposed in (Yoon & Ryu, 2022), between composite-EAG and the *optimized Halpern's method* (OHM) (Lieder, 2021; Kim, 2021). We first present the definition of the merging path property and results for OHM.

The Merging Path Property (Yoon & Ryu, 2022) Let $\mathfrak{R}$ be a deterministic algorithm. We use $\mathfrak{R}(x_0; \mathcal{P}) = \{x_0, x_1, \ldots\}$ to denote that $\mathfrak{R}$ applied to problem $\mathcal{P}$ with initial point x_0 produces iterates $x_0, x_1, \ldots$. We say two algorithms $\mathfrak{R}_1$ and $\mathfrak{R}_1$ have $O(r(k))$-*merging path* if for any problem $\mathcal{P}$ and point $x_0 \in \mathbb{R}^n$, the iterates $\mathfrak{R}_\ell(x_0; \mathcal{P}) = (x_0, x_1^\ell, x_2^\ell, \ldots)$ for $\ell = 1, 2$ satisfy $\|x_k^1 - x_k^2\|^2 = O(r(k))$. We say $\mathfrak{R}_1$ and $\mathfrak{R}_1$ are $O(r(k))$-MP if they have $O(r(k))$-merging paths. The merging path property is a strong notion that quantifies the near-equivalence of algorithms. Yoon & Ryu (2022) establishes $O(1/k^2)$-MP for EAG, FEG, and the optimized Halpern's method (OHM) in the unconstrained ($A = 0$) and monotone ($\rho \ge 0$) setting. Since it is well-known that OHM has point convergence, the merging path property implies point convergence of EAG and FEG in the unconstrained and monotone setting.

Optimized Halpern's Method For the inclusion problem of $F + A$, OHM is an implicit method: let $w_0 = z_0$, it updates in each iteration $k \ge 0$:

$$w_{k+\frac{1}{2}} = \beta_k w_0 + (1 - \beta_k)w_k$$
$$w_{k+1} = J_{\eta(F+A)}[w_{k+\frac{1}{2}}] \tag{OHM}$$

where we denote $\beta_k = \frac{1}{k+1}$. For the ease of analysis, we define $d_{k+1} = \frac{w_{k+\frac{1}{2}} - \eta F(w_{k+1}) - w_{k+1}}{\eta}$. By definition of the resolvent $J_{\eta(F+A)}$, we have $d_{k+1} \in A(w_{k+1})$. We note that when $F + A$ is maximally ρ-comonotone with $\rho > -\frac{1}{2L}$, its resolvent $J_{\eta(F+A)}$ is nonexpansive (Bauschke et al., 2021). Thus, classical results on Halpern's iteration show that OHM converges (Bauschke et al., 2017, Theorem 30.1) and has optimal last-iterate convergence rate (Lieder, 2021). We summarize these results below.

Theorem 3.7 (Theorem 30.1 of (Bauschke et al., 2017), Theorem 2.1 of (Lieder, 2021)). *Let $w_0 = z_0 \in \mathcal{Z}$ be any point. If $J_{\eta(F+A)}$ has a fixed point,[9] then the iterates $\{w_k\}_{k\ge0}$ of OHM converge to $\Pi_{\text{Zer}(F+A)}[z_0] = \arg\min_{z \in \text{Zer}(F+A)} \|z - z_0\|$. Moreover, for any $k \ge 1$, it*

---

[9] $J_{\eta(F+A)}$ has a fixed point is equivalent to $\text{Zer}(F + A)$ is nonempty.

holds that

$$\|\eta F(w_{k+1}) + \eta d_{k+1}\|^2 = \left\|w_{k+\frac{1}{2}} - J_{\eta(F+N_{\mathcal{Z}})}[w_{k+\frac{1}{2}}]\right\|^2$$
$$\leq \frac{4\|w_0 - w^*\|^2}{(k+1)^2}.$$

Merging Path between composite-EAG and OHM The main result of this section is to establish that composite-EAG and OHM are $O(\frac{1}{\sqrt{k}})$-MP, i.e., $\|z_{k+1} - w_{k+1}\|^2 = O(\frac{1}{\sqrt{k}})$ (Theorem 3.8). Combining this result with the convergence guarantee of OHM (Theorem 3.7) directly implies that composite-EAG converges to the solution closest to the initial point z_0.

Theorem 3.8 (Merging path between composite-EAG and OHM). *Consider a comonotone inclusion problem with Assumption 2.2 holds with $\rho \geq -\frac{1}{20L}$. Let $\eta = \frac{0.31}{L}$. Denote $\{z_k\}_{k \geq 0}$ the iterates of composite-EAG and $\{w_k\}_{k \geq 0}$ the iterates of OHM. Let $w_0 = z_0$, then for any $k \geq 1$,*

$$\|z_{k+1} - w_{k+1}\|^2 \leq \frac{40000H^2}{(k+1)^{\frac{1}{2}}} + \frac{30000H^2}{(k+1)^2}.$$

The rate of $O(\frac{1}{\sqrt{k}})$-MP is worse than $O(1/k^2)$-MP obtained by Yoon & Ryu (2022) in the unconstrained and monotone setting. This is due to the need to tackle additional technical challenges introduced by the constraints and negative comonotonicity. Understanding the precise dependence on k for the merging path property is an interesting open problem. To our knowledge, composite-EAG is the first explicit algorithm with both optimal convergence rate and point convergence guarantee for CMI. Indeed, such guarantees were not known even in the unconstrained setting.

Proof Sketch Fix any k, we conduct a single-step analysis and prove the following inequality

$$(k+1)^2\|z_{k+1} - w_{k+1}\|^2 \leq k^2\|z_k - w_k\|^2 + (k+1)^2 E_k$$

where E_k represents certain terms. By telescoping the above inequality, it remains to show that $\sum_{t=1}^{k}(k+1)^2 E_k = O(k^{3/2})$, which follows from combining the $O(1/k)$ last-iterate convergence rate (Theorem 3.2) and Corollary 3.6.

4. composite-FEG: Optimal Convergence Rate for a Wider Range of ρ

The composite-EAG algorithm has optimal last-iterate convergence rate as well as guaranteed point convergence, but only for CMI with $\rho \geq -\frac{1}{20L}$. In this section, we propose the composite Fast Extra Gradient (composite-FEG) method, an explicit method that is applicable to CMI with $\rho > -\frac{1}{2L}$. We show that composite-FEG achieves optimal $O(1/T)$ last-iterate convergence rate. We remark that

CMI with $\rho > -\frac{1}{2L}$ is currently the largest class of non-monotone problems efficiently solvable by any single-loop algorithms. Our result generalizes previous results (Yoon & Ryu, 2021; Lee & Kim, 2021), which are limited to the unconstrained setting. Additionally, we show that composite-FEG has point convergence in the monotone setting ($\rho = 0$), which matches the state-of-the-art result by (Sedlmayer et al., 2023).

Extension of FEG for Comonotone Inclusion Given any initial point $z_0 \in \mathbb{R}^n$ and step size $\eta > 0$, composite-FEG sets $c_0 = 0$ and updates $\{z_{k+\frac{1}{2}}, z_{k+1}, c_{k+1}\}_{k \geq 0}$ as shown in Algorithm 2. Note that by definition, we have $c_k \in A(z_k)$ for all $k \geq 1$. Our algorithm is inspired by FEG (Lee & Kim, 2021). In particular, when $A(z) = 0$, i.e., the unconstrained setting, c_k is always $\mathbf{0}$, and our algorithm is identical to FEG.

Our main results of the section shows that composite-FEG has optimal $O(\frac{1}{T})$ for the range of comonotone inclusion problems with $\rho > -\frac{1}{2L}$. Moreover, composite-FEG enjoys point convergence in the monotone setting. All missing proofs in this section can be found in Appendix D.

4.1. Last-Iterate Convergence Rate of composite-FEG

Theorem 4.1. *Suppose Assumption 2.2 holds for some $\rho \in (-\frac{1}{2L}, 0]$. Let $z_0 \in \mathbb{R}^n$ be any starting point and $\{z_k, z_{k+\frac{1}{2}}\}_{k \geq 1}$ be the iterates of composite-FEG with step size $\eta \in (max(0, -2\rho), \frac{1}{L})$. Then for any $T \geq 1$,*

$$\min_{c \in A(z_T)} \|F(z_T) + c\|^2 = r_{F,A}^{tan}(z_T)^2 \leq \frac{4}{(\eta + 2\rho)^2 L^2} \frac{H_0^2 L^2}{T^2},$$

where $H_0^2 = 4\|z_1 - z_0\|^2 + \|z_0 - z^*\|^2 \leq \frac{4 \cdot r_{F,A}^{tan}(z_0)^2}{L^2} + \|z_0 - z^*\|^2$.

Remark 4.2. To interpret the convergence rate, one can think of a properly selected η such that $(\eta + 2\rho)L$ is an absolute constant, and $r_{F,A}^{tan}(z_T)^2$ will be $O(\frac{H_0^2 L^2}{T^2})$.

We show that composite-FEG enjoys the optimal convergence rate of $O(\frac{1}{T})$ via a potential function argument.

Potential function To analyze our algorithm, we adopt the following potential function: for $k \geq 1$,

$$U_k := \left(\frac{k^2}{2}\left(1 + \frac{2\rho}{\eta}\right) - \frac{\rho}{\eta}k\right) \cdot \|\eta F(z_k) + \eta c_k\|^2$$
$$+ k \cdot \langle \eta F(z_k) + \eta c_k, z_k - z_0 \rangle.$$

The potential function is the same as the one used in the analysis for FEG (Lee & Kim, 2021) when c_k is always 0, and we adapt it for non-zero c_k's. The analysis builds on the

---

Algorithm 2 composite-FEG for CMI with $\rho > -\frac{1}{2L}$

Initialization: $z^0 \in \mathbb{R}^n$, $c_0 = 0$, $\beta_k = \frac{1}{k+1}$ for $k \geq 0$
for $k \geq 0$ **do**

$$z_{k+\frac{1}{2}} = \beta_k z_0 + (1 - \beta_k)z_k - (1 - \beta_k)(\eta + 2\rho)(F(z_k) + c_k)$$

$$z_{k+1} = J_{\eta A}\Big[\beta_k z_0 + (1 - \beta_k)z_k - \eta F(z_{k+\frac{1}{2}}) - 2(1 - \beta_k)\rho(F(z_k) + c_k)\Big]$$ (composite-FEG)

$$c_{k+1} = \frac{\beta_k z_0 + (1 - \beta_k)z_k - \eta F(z_{k+\frac{1}{2}}) - 2(1 - \beta_k)\rho(F(z_k) + c_k) - z_{k+1}}{\eta}$$

end for

---

following two properties of the potential function: Proposition 4.3 establishes an upper bound of U_1; Lemma 4.4 shows $U_{k+1} \leq U_k$ for all $k \geq 1$.

Proposition 4.3. $U_1 \leq \frac{(1+\eta L)(3+\eta L)}{2} \cdot \|z_1 - z_0\|^2 \leq 4\|z_1 - z_0\|^2$ and $\|z_1 - z_0\|^2 \leq \eta^2 \cdot r_{F,A}^{tan}(z_0)^2$.

Lemma 4.4. *Assume Assumption 2.2 holds for some ρ. Let $z_0 \in \mathbb{R}^n$ be any initial point and $\{z_k, z_{k+\frac{1}{2}}\}_{k\geq 1}$ be the iterates of composite-FEG with step size $\eta > -2\rho$.[10] Then for all $k \geq 1$, we have*

$$U_{k+1} \leq U_k - \frac{(1 - \eta^2 L^2)(k+1)^2}{2}\Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\|^2.$$

The following lemma shows that U_t is of order $\Omega(t^2 \cdot r_{F,A}^{tan}(z_t)^2)$.

Lemma 4.5. *In the same setup as Theorem 4.1, for any $t \geq 1$, we have*

$$\frac{\eta(\eta + 2\rho)t^2}{4} \cdot \|F(z_t) + c_t\|^2 \leq U_t + \frac{\eta}{\eta + 2\rho}\|z_0 - z^*\|^2$$

In particular, $U_t \geq -\|z_0 - z^*\|^2$ if $\rho = 0$.

Now we are ready to prove $O(1/T)$ last-iterate convergence rate of composite-FEG.

Proof of Theorem 4.1 By Proposition 4.3 and the assumption that $\eta L < 1$, $U_1 \leq \frac{(1+\eta L)(3+\eta L)}{2} \cdot \|z_1 - z_0\|^2 \leq 4\|z_1 - z_0\|^2 \leq \frac{4}{L^2} \cdot r^{tan}(z_0)^2$. Now combining Lemma 4.4 and Lemma 4.5 gives

$$\|F(z_T) + c_T\|^2$$
$$\leq \frac{4}{\eta(\eta + 2\rho)T^2}\left(U_T + \frac{\eta}{\eta + 2\rho}\|z_0 - z^*\|^2\right)$$
$$\leq \frac{4}{(\eta + 2\rho)^2 T^2}\left(\frac{\eta + 2\rho}{\eta}U_1 + \|z_0 - z^*\|^2\right)$$
$$\leq \frac{4}{(\eta + 2\rho)^2 T^2}\left(4\|z_1 - z_0\|^2 + \|z_0 - z^*\|^2\right). \quad (\rho \leq 0)$$

This completes the proof.

---

[10]Lemma 4.4 holds for all step size η, but our potential function is no longer useful when $\eta \leq -2\rho$.

4.2. Point Convergence of composite-FEG

In this section, we show that the sequence of iterates of composite-FEG converges when F is monotone, i.e., $\rho = 0$. We adopt the same idea as in the previous section. Specifically, we prove the merging path (MP) property between composite-FEG and the *optimized Halpern's method* (OHM). We remark that MP property between FEG and OHM was established only in the unconstrained setting (Yoon & Ryu, 2022).

Merging Path between composite-FEG and OHM The main result of this section is to establish that composite-FEG and OHM are $O(\frac{1}{\sqrt{k}})$-MP, i.e., $\|z_{k+1} - w_{k+1}\| = O(\frac{1}{\sqrt{k}})$. Combining this result with the convergence guarantee of OHM (Theorem 3.7) directly proves that composite-EAG converges to the solution closest to the initial point z_0.

Theorem 4.6 (Merging path between composite-FEG and OHM)**.** *Suppose Assumption 2.2 holds for $\rho = 0$. Let $z_0 = w_0$ and $\eta \in (0, \frac{1}{L})$. Denote $\{z_k\}_{k\geq 0}$ the iterates of composite-FEG and $\{w_k\}_{k\geq 0}$ the iterates of OHM. Denote $H_0^2 = 4\|z_1 - z_0\|^2 + \|z_0 - z^*\|^2$. Then for any $k \geq 1$,*

$$\|z_{k+1} - w_{k+1}\|^2 \leq \frac{H_0^2}{1 - \eta^2 L^2} \cdot \left(\frac{24}{(k+1)^{\frac{1}{2}}} + \frac{36}{(k+1)^2}\right).$$

5. Conclusion and Discussion

In this work, we propose two novel algorithms for solving comonotone inclusion problems with provable convergence guarantees. The composite-EAG algorithm recovers EAG from (Yoon & Ryu, 2021) in the unconstrained case and has $O(1/T)$ last-iterate convergence rate and the favorable point convergence for ρ-comonotone with $\rho \geq -\frac{1}{20L}$. The composite-FEG algorithm recovers FEG from (Lee & Kim, 2021) in the unconstrained case, and has $O(1/T)$ last-iterate convergence rate for ρ-comonotone problems with $\rho > -\frac{1}{2L}$ while maintains point convergence in the monotone case. An interesting open problem is to design an algorithm that obtains both optimal convergence rate and point convergence for a wider range of ρ-comonotone problems.

Impact Statement

This paper presents work that aims to advance the field of Machine Learning. Our work has many potential societal consequences, none of which we feel must be specifically highlighted here.

Acknowledgment

References

Alacaoglu, A., Kim, D., and Wright, S. J. Extending the reach of first-order algorithms for nonconvex min-max problems with cohypomonotonicity. *arXiv preprint arXiv:2402.05071*, 2024.

Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 214–223. PMLR, July 2017. URL https://proceedings.mlr.press/v70/arjovsky17a.html. ISSN: 2640-3498.

Bauschke, H. H. and Combettes, P. L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011. URL https://link.springer.com/book/10.1007/978-1-4419-9467-7.

Bauschke, H. H., Combettes, P. L., Bauschke, H. H., and Combettes, P. L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2017.

Bauschke, H. H., Moursi, W. M., and Wang, X. Generalized monotone operators and their averaged resolvents. *Mathematical Programming*, 189(1):55–74, 2021.

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. *Robust Optimization*. Princeton University Press, December 2009. ISBN 978-1-4008-3105-0. doi: 10.1515/9781400831050. URL https://www.degruyter.com/document/doi/10.1515/9781400831050/html.

Böhm, A. Solving nonconvex-nonconcave min-max problems exhibiting weak minty solutions. *arXiv preprint arXiv:2201.12247*, 2022.

Cai, X., Alacaoglu, A., and Diakonikolas, J. Variance reduced halpern iteration for finite-sum monotone inclusions. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=0i6Z9N5MLY.

Cai, Y. and Zheng, W. Doubly optimal no-regret learning in monotone games. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 3507–3524. PMLR, 2023a. URL https://proceedings.mlr.press/v202/cai23g.html.

Cai, Y. and Zheng, W. Accelerated single-call methods for constrained min-max optimization. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023b. URL https://openreview.net/pdf?id=HRwN7IQLUKA.

Cai, Y., Oikonomou, A., and Zheng, W. Finite-time last-iterate convergence for learning in multi-player games. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Combettes, P. L. and Pennanen, T. Proximal methods for cohypomonotone operators. *SIAM journal on control and optimization*, 43(2):731–742, 2004.

Dai, B., Shaw, A., Li, L., Xiao, L., He, N., Liu, Z., Chen, J., and Song, L. Sbeed: Convergent reinforcement learning with nonlinear function approximation. In *International Conference on Machine Learning*, pp. 1125–1134. PMLR, 2018.

Dang, C. D. and Lan, G. On the convergence properties of non-euclidean extragradient methods for variational inequalities with generalized monotone operators. *Computational Optimization and applications*, 60(2):277–310, 2015.

Daskalakis, C., Skoulakis, S., and Zampetakis, M. The complexity of constrained min-max optimization. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2021.

Diakonikolas, J. Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. In *Conference on Learning Theory (COLT)*, 2020.

Diakonikolas, J., Daskalakis, C., and Jordan, M. Efficient methods for structured nonconvex-nonconcave min-max optimization. *International Conference on Artificial Intelligence and Statistics*, 2021.

Du, S. S., Chen, J., Li, L., Xiao, L., and Zhou, D. Stochastic variance reduction methods for policy evaluation. In *International Conference on Machine Learning*, pp. 1049–1058. PMLR, 2017.

Facchinei, F. and Pang, J.-S. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003. URL https://link.springer.com/book/10.1007/b97544.

Fan, Y., Li, Y., and Chen, B. Weaker MVI condition: Extragradient methods with multi-step exploration. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=RNGUbTYSjk.

Golowich, N., Pattathil, S., and Daskalakis, C. Tight last-iterate convergence rates for no-regret learning in multiplayer games. *Advances in neural information processing systems (NeurIPS)*, 2020a.

Golowich, N., Pattathil, S., Daskalakis, C., and Ozdaglar, A. Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems. In *Conference on Learning Theory (COLT)*, 2020b.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Gorbunov, E., Loizou, N., and Gidel, G. Extragradient method: $\mathcal{O}(1/k)$ last-iterate convergence for monotone variational inequalities and connections with cocoercivity. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 366–402. PMLR, 2022.

Gorbunov, E., Taylor, A. B., Horváth, S., and Gidel, G. Convergence of proximal point and extragradient-based methods beyond monotonicity: the case of negative comonotonicity. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 11614–11641. PMLR, 2023. URL https://proceedings.mlr.press/v202/gorbunov23a.html.

Halpern, B. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73(6):957–961, 1967.

Kim, D. Accelerated proximal point method for maximally monotone operators. *Mathematical Programming*, 190(1):57–87, November 2021. ISSN 1436-4646. doi: 10.1007/s10107-021-01643-0. URL https://doi.org/10.1007/s10107-021-01643-0.

Kohlenbach, U. On the proximal point algorithm and its halpern-type variant for generalized monotone operators in hilbert space. *Optimization Letters*, 16(2):611–621, 2022.

Korpelevich, G. M. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976. URL https://ci.nii.ac.jp/naid/10017556617/.

Kovalev, D. and Gasnikov, A. The first optimal acceleration of high-order methods in smooth convex optimization. *Advances in Neural Information Processing Systems*, 35:35339–35351, 2022.

Lee, S. and Kim, D. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2021.

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

Liu, M., Mroueh, Y., Ross, J., Zhang, W., Cui, X., Das, P., and Yang, T. Towards better understanding of adaptive gradient algorithms in generative adversarial nets. In *International Conference on Learning Representations (ICLR)*, 2019.

Liu, M., Rafique, H., Lin, Q., and Yang, T. First-order convergence theory for weakly-convex-weakly-concave min-max problems. *Journal of Machine Learning Research*, 22(169):1–34, 2021.

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., and Vladu, A. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

Malitsky, Y. Golden ratio algorithms for variational inequalities. *Mathematical Programming*, 184(1):383–410, 2020.

Nemirovski, A. 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. Publisher: SIAM.

Nesterov, Y. Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming*, 109(2):319–344, March 2007. ISSN 1436-4646. doi: 10.1007/s10107-006-0034-z. URL https://doi.org/10.1007/s10107-006-0034-z.

Ouyang, Y. and Xu, Y. Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point

problems. *Mathematical Programming*, 185(1):1–35, 2021.

Parikh, N., Boyd, S., et al. Proximal algorithms. *Foundations and trends® in Optimization*, 1(3):127–239, 2014.

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

Pethick, T., Latafat, P., Patrinos, P., Fercoq, O., and Cevherå, V. Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems. In *International Conference on Learning Representations (ICLR)*, 2022.

Pethick, T., Xie, W., and Cevher, V. Stable nonconvex-nonconcave training via linear interpolation. *CoRR*, abs/2310.13459, 2023. doi: 10.48550/ARXIV.2310.13459. URL https://doi.org/10.48550/arXiv.2310.13459.

Rockafellar, R. T. Monotone Operators and the Proximal Point Algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, August 1976. ISSN 0363-0129. doi: 10.1137/0314056. URL https://epubs.siam.org/doi/abs/10.1137/0314056. Publisher: Society for Industrial and Applied Mathematics.

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

Ryu, E. K. and Yin, W. *Large-Scale Convex Optimization via Monotone Operators*. Cambridge University Press, 2022.

Sedlmayer, M., Nguyen, D., and Bot, R. I. A fast optimistic method for monotone variational inequalities. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 30406–30438. PMLR, 2023. URL https://proceedings.mlr.press/v202/sedlmayer23a.html.

Song, C., Zhou, Z., Zhou, Y., Jiang, Y., and Ma, Y. Optimistic dual extrapolation for coherent non-monotone variational inequalities. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:14303–14314, 2020.

Tran-Dinh, Q. The connection between nesterov's accelerated methods and halpern fixed-point iterations. *arXiv preprint arXiv:2203.04869*, 2022.

Tran-Dinh, Q. Extragradient-type methods with o(1/k) last-iterate convergence rates for co-hypomonotone inclusions. *Journal of Global Optimization*, pp. 1–25, 2023.

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

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

Yoon, T. and Ryu, E. K. Accelerated minimax algorithms flock together, 2022.

Zhou, Z., Mertikopoulos, P., Bambos, N., Boyd, S., and Glynn, P. W. Stochastic mirror descent in variationally coherent optimization problems. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 2017.

A. Additional Related Work

A.1. Additional Classical Results for Convex-Concave Min-Max Optimization

Convergence in Gap Function. Nemirovski and Nesterov show that the average iterate of extragradient-type methods has $O(\frac{1}{T})$ convergence rate in terms of gap function defined as $\max_{z' \in \mathcal{Z}} \langle F(z'), z - z' \rangle$ (Nemirovski, 2004; Nesterov, 2007), matching the lower bound for first-order methods (Ouyang & Xu, 2021). Their results extend to the more general monotone variational inequalities. Convergence to an approximate solution of CMI implies convergences with the same rate for the gap function, while the converse is not true (see Example 2).

Convergence of the Extragradient Method in Stronger Performance Measures. For stronger performance measures such as the norm of the operator (when $\mathcal{Z} = \mathbb{R}^n$) or the residual (in constrained setting), classical results show that the best-iterate of the extragradient method converges at a rate of $O(\frac{1}{\sqrt{T}})$ (Korpelevich, 1976; Facchinei & Pang, 2003). Recently, the same convergence rate is shown to hold even for the last-iterate of the extragradient method (Gorbunov et al., 2022; Cai et al., 2022). Although $O(\frac{1}{\sqrt{T}})$ convergence on the residual is optimal for all p-SCIL algorithms (Golowich et al., 2020b;a), a subclass of first-order methods that includes the extragradient method and many of its variations, faster rate is possible for other first-order methods.

A.2. Additional Results on Nonconvex-Nonconcave Min-Max Optimization

(Weak) Minty Variational Inequality We only introduce the definitions in the unconstrained setting, as that is the setting considered by several recent results, and all convergence rates are with respect to the operator norm. The *Minty variational inequality* (MVI) condition (also called coherence or variationally stable) assumes the existence of point $z^* \in \mathbb{R}^n$ such that

$$\langle F(z), z - z^* \rangle \geq 0, \quad \forall z \in \mathbb{R}^n$$

is studied in e.g., (Dang & Lan, 2015; Zhou et al., 2017; Liu et al., 2019; Malitsky, 2020; Song et al., 2020; Liu et al., 2021). Extragradient-type algorithms has $O(\frac{1}{\sqrt{T}})$ convergence rate for Lipschitz operators that satisfy the MVI condition (Dang & Lan, 2015). Diakonikolas et al. (2021) proposes a weaker condition called *weak MVI*: there exits z^* and $\rho < 0$ such that

$$\langle F(z), z - z^* \rangle \geq \rho \cdot \|F(z)\|^2, \quad \forall z \in \mathbb{R}^n.$$

The weak MVI condition includes both MVI and negative comonotonicity (Bauschke et al., 2021) as special cases. Diakonikolas et al. (2021) proposes the EG+ algorithm, which has $O(\frac{1}{\sqrt{T}})$ convergence rate under the weak MVI condition in the unconstrained setting. Böhm (2022) propose OGDA+ in the same setting while Pethick et al. (2022) generalizes EG+ to CEG+ algorithm which has $O(\frac{1}{\sqrt{T}})$ convergence rate under weak MVI condition in general (constrained) setting.

B. Additional Preliminaries

B.1. Monotone Inclusion and Variational Inequality

Variational Inequality. An important special case of CMI is the *variational inequality* (VI) problem where the maximally monotone operator A is chosen to be a normal cone operator $N_{\mathcal{Z}}$ for a closed and convex feasible set $\mathcal{Z}$. The VI problem has two variants. The *Stampacchia Variational Inequality* (SVI) problem is to find $z^* \in \mathcal{Z}$ such that

$$\langle F(z^*), z^* - z \rangle \leq 0, \quad \forall z \in \mathcal{Z}. \tag{SVI}$$

Such z^* is called a *strong solution* to VI. The *Minty Variational Inequality* (MVI) problem is to find $z^* \in \mathcal{Z}$ such that

$$\langle F(z), z^* - z \rangle \leq 0, \quad \forall z \in \mathcal{Z}. \tag{MVI}$$

Such z^* is called a *weak solution* to VI.

When F is continuous, then every solution to MVI is also a solution to SVI. When F is monotone, every solution to SVI is also a solution to (MVI), and thus, the two solution sets are equivalent. We remark that the solution set to CMI is the same as the solution set to SVI. For a convex set $\mathcal{Z}$, $\partial \mathbb{I}_{\mathcal{Z}}$ is a maximally monotone set-valued operator, and as such, CMI generilizes MVI, and convex-concave min-max optimization problems for ($\rho = 0$ and $A = \partial \mathbb{I}_{\mathcal{Z}}$) and serves as

computational frameworks for numerous important applications in fields such as economics, engineering, probability and statistics, and machine learning (Facchinei & Pang, 2003; Bauschke & Combettes, 2011; Ryu & Boyd, 2016).

Although the set of exact solutions to CMI for $A = \partial \mathbb{I}_{\mathcal{Z}}$ coincides with the set of exact solutions to the corresponding variational inequality, the approximate solutions to these two problems differ due to different performance measures. An approximate solution to the comonotone inclusion problem CMI must have a small natural residual,[11] while an approximate solution to the variational inequality only satisfies a weaker condition, i.e., its gap function is small.[12] Indeed, it is well-known that an approximate solution to CMI is also an approximate solution to the monotone variational equality, but the reverse is not true in general.

Approximate Solutions. We say $z \in \mathcal{Z}$ is an ϵ-approximate solution to SVI or MVI if

$$\langle F(z), z - z' \rangle \leq \epsilon, \forall z' \in \mathcal{Z}, \text{ or}$$
$$\langle F(z'), z - z' \rangle \leq \epsilon, \forall z' \in \mathcal{Z}, \text{ respectively.}$$

When F is monotone, it is clear that every ϵ-approximate solution to SVI is also an ϵ-approximate solution to MVI; but the reverse does not hold in general. When F is monotone and $\mathcal{Z}$ is bounded by D, then any $\frac{\epsilon}{D}$-approximate solution to CMI is an ϵ-approximate solution to SVI (Diakonikolas, 2020, Fact 1). Note that when $\mathcal{Z}$ is unbounded, neither SVI nor MVI can be approximated. If we restrict the domain to be a bounded subset of (possibly unbounded) $\mathcal{Z}$, then we can define the (restricted) gap functions (Nesterov, 2007) as

$$\text{GAP}_{F,D}^{SVI}(z) := \max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \langle F(z), z - z' \rangle,$$
$$\text{GAP}_{F,D}^{MVI}(z) := \max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \langle F(z'), z - z' \rangle.$$

The $O(\frac{1}{T})$ convergence rate for extragradient-type algorithm (Nemirovski, 2004; Nesterov, 2007) is provided in terms of $\text{GAP}_{F,D}^{MVI}(z)$, which means convergence to an approximate *weak* solution. Prior to our work, the $O(\frac{1}{T})$ convergence rate on $\text{GAP}_{F,D}^{SVI}(z)$ was only known in the unconstrained setting (Yoon & Ryu, 2021). When F is monotone, then the tangent residual $r_{F,D}^{tan}(z) \leq \frac{\epsilon}{D}$ (definition in section 2.2) implies $\text{GAP}_{F,D}^{SVI}(z) \leq \epsilon$ (Cai et al., 2022, Lemma 2). Therefore, our results (Theorem 3.2, 4.1, and E.4) for the tangent residual and the natural residual also imply an $O(\frac{1}{T})$ convergence rate on $\text{GAP}_{F,D}^{SVI}(z)$ for SVI. In the following, we show that the converse is not true in general.

Example 2 (Gap function is weaker than natural residual). Consider an instance of the Monotone VI problem on the identity operator $F(x) = x$ in $\mathcal{Z} = [0, 1]$.

- Observe that the natural residual on $x \in \mathcal{Z}$ is $\|x - \Pi_{\mathcal{Z}}[x - F(x)]\| = x$.

- Moreover, since $\mathcal{Z} = [0, 1]$, observe that for any $x \in \mathcal{Z}$ and $D \geq 0$,

$$\text{GAP}_{F,D}^{SVI}(x) \leq \text{GAP}_{F,1}^{SVI}(x) = \max_{x' \in [0,1]} x \cdot (x - x') = x^2, \text{ and}$$
$$\text{GAP}_{F,D}^{MVI}(x) \leq \text{GAP}_{F,1}^{MVI}(x) = \max_{x' \in [0,1]} x' \cdot (x - x') = \frac{x^2}{4}.$$

As a result, any algorithm with $O(\frac{1}{T})$ convergence rate with respect to the gap function only implies a $O(\frac{1}{\sqrt{T}})$ convergence rate for the corresponding CMI problem or the natural residual.

---

[11]The natural residual of a point z is simply the operator norm $\|F(z)\|$ in the unconstrained case, i.e., $\mathcal{Z} = \mathbb{R}^n$, and equals to the norm of its natural map $z - \Pi_{\mathcal{Z}}[z - F(z)]$ (Facchinei & Pang, 2003).

[12]There are several variations of the gap function. Depending on the exact definition, a small gap function value could mean an approximate *weak* solution, i.e., approximately solve the Minty Variational Inequality (MVI), or an approximate *strong* solution, i.e., approximately solve the Stampacchia Variational Inequality (SVI). Formal definitions and discussions are in Section B.1.

B.2. Proof of Fact 1

Proof. For any $c \in A(z)$, we have

$$
\begin{aligned}
r_{F,A}^{nat}(z) &= \|z - J_A[z - F(z)]\| \\
&= \|J_A[z + c] - J_A[z - F(z)]\| && (z = J_A[z + c]) \\
&\leq \|F(z) + c\|. && \text{(non-expansiveness of } J_A)
\end{aligned}
$$

Thus $r_{F,A}^{nat}(z) \leq \min_{c \in A(z)} \|F(z) + c\| = r_{F,A}^{tan}(z)$. $\qquad\square$

C. Missing Proofs in Section 3

C.1. Proof of Lemma 3.3

Proof. We recall that $z_1 = J_{\eta A}[z_0 - \eta F(z_0)]$. Thus by non-expansiveness of the resolvent $J_{\eta A}$, we have for any $c_0 \in A(z_0)$,

$$
\|z_1 - z_0\| = \|J_{\eta A}[z_0 - \eta F(z_0)] - J_{\eta A}[z_0 - \eta c_0]\| \leq \eta \|F(z_0) + c_0\|.
$$

Since the above holds for all $c_0 \in A(z_0)$, we get $\|z_1 - z_0\| \leq \eta r_{F,A}^{tan}(z_0)$. By definition of c_1, we also have

$$
\begin{aligned}
\|\eta F(z_1) + \eta c_1\|^2 &= \|z_0 - z_1 + \eta F(z_1) - \eta F(z_0)\|^2 \\
&\leq 2\|z_0 - z_1\|^2 + 2\|\eta F(z_1) - \eta F(z_0)\|^2 \\
&\leq 4\|z_0 - z_1\|^2.
\end{aligned}
$$

Now by definition of V_1, we have

$$
\begin{aligned}
V_1 &= \|\eta(F(z_1) + c_1)\|^2 + \eta\langle F(z_1) + c_1, z_1 - z_0\rangle \\
&\leq \|\eta(F(z_1) + c_1)\|^2 + \|\eta(F(z_1) + c_1)\|\|z_1 - z_0\| \\
&\leq 6\|z_1 - z_0\|^2.
\end{aligned}
$$

This completes the proof. $\qquad\square$

C.2. Proof of Theorem 3.4

Proof. We require the following fact.

Fact 2. For any $L > 0$ and $\rho \geq -\frac{1}{20L}$, $\eta = \frac{0.31}{L}$ satisfies

$$
1 + \frac{4\rho}{\eta} - \left(3 - \frac{4\rho}{\eta}\right) \cdot \eta^2 L^2 \geq \frac{9}{2000}, \tag{2}
$$

and $\frac{\rho}{\eta} > -\frac{1}{4}$

Our plan is to show $V_k - V_{k+1}$ minus two non-negative terms is greater than $-\frac{1}{2}\|\eta F(z_{k+1}) + c_{k+1}\|^2 + \frac{9k(k+1)}{4000}\|z_{k+\frac{1}{2}} - z_{k+1}\|^2$.

Non-negative Terms Let $p = \frac{1}{3}$ and $c = -\frac{4p\rho}{\eta} \geq 0$. Since F is L-Lipschitz, we have

$$
\eta^2 L^2 \cdot \left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2 - \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2 \geq 0.
$$

Multiplying both sides of the above inequality by $(1 + c)$ and rearranging terms, we get

$$
\begin{aligned}
&p \cdot \left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2 - \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2 \\
&+ \left((1 + c)\eta^2 L^2 - p\right) \cdot \left\|z_{k+\frac{1}{2}}\right\|^2 - c \cdot \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2 \geq 0.
\end{aligned} \tag{3}
$$

Moreover, since $c_k \in A(z_k)$ and $c_{k+1} \in A(z_{k+1})$ and the fact that $F + A$ is ρ-comonotone, we have

$$
\langle \eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k, z_{k+1} - z_k\rangle - \frac{\rho}{\eta}\|\eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k\|^2 \geq 0. \tag{4}
$$

Descent Identity We have the following identity holds by Proposition F.2 (Equation (24)) since $\eta c_k = z_k - \eta F(z_k) + \frac{1}{k+1}(z_0 - z_k) - z_{k+\frac{1}{2}}$ and $\eta c_{k+1} = z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+1}(z_0 - z_k) - z_{k+1}$.

$$V_k - V_{k+1} - \frac{k(k+1)}{2p} \cdot \text{LHS of Inequality (3)} - k(k+1) \cdot \text{LHS of Inequality (4)}$$

$$= \frac{(1-p)k(k+1)}{2p}\left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2 \tag{5}$$

$$+ (k+1) \cdot \left\langle \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}), \eta F(z_{k+1}) + \eta c_{k+1} \right\rangle \tag{6}$$

$$+ \frac{k(k+1)}{2p}\left((p - (1+c)\eta^2 L^2) \cdot \left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2 + c\left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2\right) \tag{7}$$

$$+ \frac{k(k+1)\rho}{\eta}\|\eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k\|^2. \tag{8}$$

Since $\|a\|^2 + \langle a, b \rangle = \|a + \frac{b}{2}\|^2 - \frac{\|b\|^2}{4}$, we have

$$\text{Expression}(5) + \text{Expression}(6)$$

$$= \left\|\sqrt{\frac{(1-p)k(k+1)}{2p}} \cdot \left(\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right) + \sqrt{\frac{p(k+1)}{2(1-p)k}} \cdot (\eta F(z_{k+1}) + \eta c_{k+1})\right\|^2$$

$$- \frac{k+1}{2k} \cdot \frac{p}{1-p}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2$$

$$\geq -\frac{p}{1-p}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2. \tag{$k \geq 1$}$$

$$= -\frac{1}{2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2.$$

Now it remains to give a non-negative lower bound of Expression (7) + Expression (8). Recall that $p = \frac{1}{3}$ and $c = -\frac{4\rho p}{\eta}$, thus

$$(\frac{2}{k(k+1)}) \cdot (\text{Expression (20)} + \text{Expression (21)})$$

$$= \left(1 - \left(3 - \frac{4\rho}{\eta}\right) \cdot \eta^2 L^2\right) \cdot \left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2 - \frac{4\rho}{\eta} \cdot \cdot \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2$$

$$+ \frac{2\rho}{\eta} \cdot \|\eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k\|^2$$

$$\geq \left(1 - \left(3 - \frac{4\rho}{\eta}\right) \cdot \eta^2 L^2\right) \cdot \left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2 + \frac{4\rho}{\eta} \cdot \left\|\eta F(z_{k+\frac{1}{2}}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k\right\|^2$$

$$(\|A\|^2 - \tfrac{1}{2}\|B\|^2 \geq -\|A + B\|^2)$$

$$= \left(1 + \frac{4\rho}{\eta} - \left(3 - \frac{4\rho}{\eta}\right) \cdot \eta^2 L^2\right)\left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2$$

$$\geq \frac{9}{2000}\left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2,$$

where in the last equality we use the fact that $z_{k+\frac{1}{2}} - z_{k+1} = \eta F(z_{k+\frac{1}{2}}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k$ which holds by the update rule of composite-EAG, and in the last inequality we use Fact 2. $\qquad\square$

C.3. Proof of Lemma 3.5

Proof. Fix any $k \geq 1$. Since $0 \in F(z^*) + A(z^*)$, by ρ-comonotonicity and Fact 2, we have

$$\langle \eta F(z_k) + \eta c_k, z_k - z^* \rangle \geq \frac{\rho}{\eta}\|\eta F(z_k) + \eta c_k\|^2 > -\frac{1}{4}\|\eta F(z_k) + \eta c_k\|^2. \tag{9}$$

By definition of V_k, we have

$$
\begin{aligned}
V_k &= \frac{k(k+1)}{2} \cdot \|\eta F(z_k) + \eta c_k\|^2 + k \cdot \langle \eta F(z_k) + \eta c_k, z_k - z_0 \rangle \\
&= \frac{k(k+1)}{2} \cdot \|\eta F(z_k) + \eta c_k\|^2 + k \cdot \langle \eta F(z_k) + \eta c_k, z_k - z^* \rangle + k \langle \eta F(z_k) + \eta c_k, z^* - z_0 \rangle \\
&\geq \frac{k(k+1)}{2} \cdot \|\eta F(z_k) + \eta c_k\|^2 - \frac{1}{4}\|\eta F(z_k) + \eta c_k\|^2 + k \langle \eta F(z_k) + \eta c_k, z^* - z_0 \rangle && \text{(By (9))} \\
&\geq \frac{k(k+\frac{1}{2})}{2} \cdot \|\eta F(z_k) + \eta c_k\|^2 - \frac{k^2}{4}\|\eta F(z_k) + \eta c_k\|^2 - \|z_0 - z^*\|^2 \\
&= \frac{k(k+1)}{4} \cdot \|\eta F(z_k) + \eta c_k\|^2 - \|z_0 - z^*\|^2,
\end{aligned}
$$

where in the second last inequality, we apply $\langle a, b \rangle \geq -\frac{\alpha}{4}\|a^2\| - \frac{1}{\alpha}\|b^2\|$ with $a = \eta F(z_k) + \eta c_k$, $b = z^* - z_0$, and $\alpha = k$. This completes the proof. $\square$

C.4. Proof of Corollary 3.6

Proof. By Theorem 3.4, we have

$$
\frac{9k(k+1)}{4000}\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2 \leq V_k - V_{k+1} + \frac{1}{2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2
$$

Telescoping the above inequality for $k = 1, 2, \ldots, T$ gives

$$
\begin{aligned}
\frac{9}{4000}\sum_{k=1}^{T} k(k+1)\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2 &\leq V_1 - V_{T+1} + \frac{1}{2}\sum_{k=1}^{T}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2 \\
&\leq V_1 + \|z_0 - z^*\|^2 + 10H^2 \cdot \left(\sum_{k=1}^{T}\frac{1}{(t+1)^2}\right) \\
&\leq 9H^2.
\end{aligned}
$$

where in the second inequality we use (1) $V_{T+1} \geq -\|z_0 - z^*\|^2$ by Lemma 3.5 and (2) $\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2 \leq \frac{20H^2)}{(t+1)^2}$ by Theorem E.2, in the last inequality we use $\sum_{k=1}^{\infty}\frac{1}{(t+1)^2} = \frac{\pi^2}{6} - 1 \leq \frac{3}{4}$. This concludes $\sum_{k=1}^{\infty} k(k+1)\|z_{k+1} - z_{k+\frac{1}{2}}\|^2 \leq 4000H^2$. Since $\frac{1}{2}(k+1)^2 \leq k(k+1)$ for all $k \geq 1$, it further implies $\|z_{k+1} - z_{k+\frac{1}{2}}\|^2 \leq \frac{8000H^2}{(k+1)^2}$ for all $k \geq 1$. $\square$

C.5. Proof of Theorem 3.8

Proof. Using the update rule of composite-EAG and OHM and recall: $\eta c_{k+1} := z_k - \eta F(z_{k+\frac{1}{2}}) + \beta_k(z_0 - z_k) - z_{k+1}$ and $\eta d_{k+1} = w_{k+\frac{1}{2}} - \eta F(w_{k+1}) - w_{k+1}$ we have

$$
\begin{aligned}
z_{k+1} - w_{k+1} &= \left(\beta_k z_0 + (1 - \beta_k)z_k - \eta F(z_{k+\frac{1}{2}}) - \eta c_{k+1}\right) - \left(\beta_k w_0 + (1 - \beta_k)w_k - \eta F(w_{k+1}) - \eta d_{k+1}\right) \\
&= (1 - \beta_k)(z_k - w_k) + \eta\left(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right), && (w_0 = z_0)
\end{aligned}
$$

which implies

$$
\begin{aligned}
\|z_{k+1} - w_{k+1}\|^2 &= (1 - \beta_k)^2\|z_k - w_k\|^2 + \eta^2\left\|F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\|^2 \\
&+ \underbrace{2\left\langle (1 - \beta_k)(z_k - w_k), \eta\left(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right)\right\rangle}_{\mathbf{I}}.
\end{aligned}
$$

We focus on term **I**. We can verify

$$
z_{k+\frac{1}{2}} - w_{k+1} = (1 - \beta_k)(z_k - w_k) - \eta(F(z_k) + c_k - F(w_{k+1}) - d_{k+1}).
$$

17

Thus term $\mathbf{I}$ can be rewritten as

$$
\begin{aligned}
\mathbf{I} &= 2\Big\langle (1-\beta_k)(z_k - w_k), \eta\Big(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big)\Big\rangle \\
&= 2\Big\langle z_{k+\frac{1}{2}} - w_{k+1} + \eta(F(z_k) + c_k - F(w_{k+1}) - d_{k+1}), \eta\Big(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big)\Big\rangle \\
&= 2\eta^2\Big\langle F(z_k) + c_k - F(w_{k+1}) - d_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle + \\
&\quad + \underbrace{2\eta\Big\langle z_{k+\frac{1}{2}} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle}_{A_k}.
\end{aligned}
$$

Combining the above (we keep term A_k for now), we get

$$
\begin{aligned}
&\|z_{k+1} - w_{k+1}\|^2 \\
&\leq (1-\beta_k)^2\|z_k - w_k\|^2 + \eta^2\Big\|F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\|^2 \\
&\quad + 2\eta^2\Big\langle F(z_k) + c_k - F(w_{k+1}) - d_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle + A_k \\
&= (1-\beta_k)^2\|z_k - w_k\|^2 - \eta^2\|F(w_{k+1}) + d_{k+1}\|^2 + 2\eta^2\langle F(z_k) + c_k, F(w_{k+1}) + d_{k+1}\rangle \\
&\quad - 2\eta^2\Big\langle F(z_k) + c_k, F(z_{k+\frac{1}{2}}) + c_{k+1}\Big\rangle + \eta^2\Big\|F(z_{k+\frac{1}{2}}) + c_{k+1}\Big\|^2 + A_k. \\
&\leq (1-\beta_k)^2\|z_k - w_k\|^2 + \eta^2\|F(z_k) + c_k\|^2 \qquad\qquad\qquad\qquad (\text{We use } -a^2 + 2ab \leq b^2) \\
&\quad - 2\eta^2\Big\langle F(z_k) + c_k, F(z_{k+\frac{1}{2}}) + c_{k+1}\Big\rangle + \eta^2\Big\|F(z_{k+\frac{1}{2}}) + c_{k+1}\Big\|^2 + A_k \\
&= (1-\beta_k)^2\|z_k - w_k\|^2 + \eta^2\Big\|F(z_k) + c_k - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\|^2 + A_k \\
&= (1-\beta_k)^2\|z_k - w_k\|^2 + \Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\|^2 + A_k,
\end{aligned}
$$

where in the last equality we use the fact that $z_{k+1} - z_{k+\frac{1}{2}} = \eta(F(z_k) + c_k - F(z_{k+\frac{1}{2}}) - c_{k+1})$ by update rule of composite-EAG. Plugging $\beta_k = \frac{1}{k+1}$ and multiplying both sides with $(k+1)^2$ gives

$$
(k+1)^2\|z_{k+1} - w_{k+1}\|^2 \leq k^2\|z_k - w_k\|^2 + (k+1)^2\Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\|^2 + (k+1)^2 A_k.
$$

Telescoping the above gives

$$
(k+1)^2\|z_{k+1} - w_{k+1}\|^2 \leq \|z_1 - w_1\|^2 + \sum_{t=1}^{k}(t+1)^2\Big\|z_{t+1} - z_{t+\frac{1}{2}}\Big\|^2 + \sum_{t=1}^{k}(t+1)^2 A_t.
$$

$$
\Rightarrow \|z_{k+1} - w_{k+1}\|^2 \leq \frac{\|z_1 - w_1\|^2}{(k+1)^2} + \frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2\Big\|z_{t+1} - z_{t+\frac{1}{2}}\Big\|^2 + \frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2 A_t
$$

It remains to bound $\frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2\|z_{t+1} - z_{t+\frac{1}{2}}\|^2$ and $\frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2 A_t$.

For the first term, using Corollary 3.6, we have

$$
\sum_{t=1}^{k}(t+1)^2\Big\|z_{t+1} - z_{t+\frac{1}{2}}\Big\|^2 \leq 2\sum_{t=1}^{k}t(t+1)\Big\|z_{t+1} - z_{t+\frac{1}{2}}\Big\|^2 \leq 8000H^2.
$$

For the term with A_t, we need a more careful analysis. We decompose $A_k = B_k + C_k + D_k$ as follows.

$$
\begin{aligned}
A_k &= 2\eta\left\langle z_{k+\frac{1}{2}} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle \\
&= 2\eta\left\langle z_{k+\frac{1}{2}} - z_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle \\
&\quad + 2\eta\left\langle z_{k+1} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle \\
&= \underbrace{2\eta\left\langle z_{k+\frac{1}{2}} - z_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle}_{B_k} + \underbrace{2\eta\left\langle z_{k+1} - w_{k+1}, F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\rangle}_{C_k} \\
&\quad + \underbrace{2\eta\left\langle z_{k+1} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+1}) - c_{k+1}\right\rangle}_{D_k}.
\end{aligned}
$$

For B_k, we have

$$
\begin{aligned}
B_k &= 2\eta\left\langle z_{k+\frac{1}{2}} - z_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle \\
&\leq 2\eta\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|\left(\left\|F(w_{k+1}) + d_{k+1}\right\| + \left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\| + \left\|F(z_{k+1}) + c_{k+1}\right\|\right) \\
&\leq 2\eta L\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2 + 2\eta\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|\left(\left\|F(w_{k+1}) + d_{k+1}\right\| + \left\|F(z_{k+1}) + c_{k+1}\right\|\right) \\
&\leq \frac{16000H^2}{(k+1)^2} + 2\eta\sqrt{\frac{8000H^2}{(k+1)^2}}\cdot\left(\frac{2\|z_0 - z^*\|}{\eta(k+1)} + \frac{\sqrt{20H^2}}{\eta(k+1)}\right) \\
&\leq \frac{17400H^2}{(k+1)^2},
\end{aligned}
$$

where in the second last inequality we use the last-iterate convergence rate of OHM (Theorem 3.7) and composite-EAG (Theorem 3.2) as well as the bound on $\|z_{k+1} - z_{k+\frac{1}{2}}\|^2$ (Corollary 3.6).

For C_k, we have

$$
\begin{aligned}
C_k &= 2\eta\left\langle z_{k+1} - w_{k+1}, F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\rangle \\
&\leq 2\eta\|z_{k+1} - w_{k+1}\|\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\| \\
&\leq 2\|z_{k+1} - w_{k+1}\|\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\| \qquad\qquad (F \text{ is } L\text{-Lipschitz and } \eta L \leq 1) \\
&\leq 400H\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|,
\end{aligned}
$$

where in the last inequality we use $\|z_{k+1} - w_{k+1}\| \leq 200H$ is bounded by Corollary C.3.

For D_k, since $F + N_{\mathcal{Z}}$ is ρ-comonotone, we have

$$
\begin{aligned}
D_k &= 2\eta\langle z_{k+1} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+1}) - c_{k+1}\rangle \\
&\leq -2\rho\eta\|F(w_{k+1}) + d_{k+1} - F(z_{k+1}) - c_{k+1}\|^2 \\
&\leq -4\rho\eta\|F(w_{k+1}) + d_{k+1}\|^2 - 4\rho\eta\|F(z_{k+1}) + c_{k+1}\|^2 \\
&\leq -\frac{4\rho}{\eta}\cdot\frac{4\|z_0 - z^*\|^2}{(k+1)^2} - \frac{4\rho}{\eta}\cdot\frac{20H^2}{(k+1)^2} \qquad\qquad (\text{convergence rate of OHM and Theorem 3.2}) \\
&\leq \frac{24H^2}{(k+1)^2}. \qquad\qquad (\tfrac{\rho}{\eta} \geq -\tfrac{1}{4})
\end{aligned}
$$

Combining the above bounds for B_k, C_k, D_k, we get

$$
(k+1)^2 A_k \leq 17424H^2 + 400H(k+1)^2\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|.
$$

Since $\sum_{t=1}^{\infty}(t+1)^2\|z_{t+1} - z_{t+\frac{1}{2}}\|^2 \leq 3200H^2$, using Cauchy-Schwartz, we get

$$\sum_{t=1}^{k}(t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\| \leq \sqrt{\left(\sum_{t=1}^{k}(t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2\right)\cdot\left(\sum_{t=1}^{k}(t+1)^2\right)}$$
$$\leq 100H(k+1)^{3/2}.$$

Thus $\frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2\|z_{t+1} - z_{t+\frac{1}{2}}\| \leq 100H(k+1)^{-\frac{1}{2}}$.

Hence we get

$$\frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2 A_t = \frac{17424H^2}{(k+1)^2} + \frac{40000H^2}{(k+1)^{\frac{1}{2}}}$$

Combining all the above, we conclude that for all k,

$$\|z_{k+1} - w_{k+1}\|^2 \leq \frac{\|z_1 - w_1\|^2}{(k+1)^2} + \frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2 + \frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2 A_t$$
$$\leq \frac{40000H^2}{(k+1)^{\frac{1}{2}}} + \frac{30000H^2}{(k+1)^2}.$$

This completes the proof. $\qquad\square$

C.5.1. BOUNDED ITERATES OF OHM AND COMPOSITE-EAG

In this section, we prove auxiliary results needed in the proof of Theorem 3.8. We show that the iterates of OHM and composite-EAG all have bounded distance away from the initial point z_0.

Lemma C.1 (Bounded Iterates of OHM). *The iterates $\{w_k\}_{k\geq 0}$ of OHM satisfies for all $k \geq 0$*

$$\|w_k - w_0\|^2 \leq 4\|w_0 - w^*\|^2.$$

Proof. When $k = 1$, by nonexpansiveness of the resolvent $J_{\eta(F+A)}$, we know

$$\|w_1 - w^*\| = \left\|J_{\eta(F+A)}[w_0] - J_{\eta(F+A)}[w^*]\right\| \leq \|w_0 - w^*\|.$$

This implies

$$\|w_1 - w_0\|^2 \leq 2\|w_1 - w^*\|^2 + 2\|w_0 - w^*\|^2 \leq 4\|w_0 - w^*\|^2.$$

For $k \geq 1$, by definitions of OHM and d_{k+1}, we have

$$\|w_{k+1} - w_0\|^2 = \left\|\frac{k}{k+1}(w_k - w_0) - \eta(F(w_{k+1}) + d_{k+1})\right\|^2$$
$$\leq \left(1 + \frac{1}{k}\right)\frac{k^2}{(k+1)^2}\|w_k - w_0\|^2 + (1+k)\eta^2\|(F(w_{k+1}) + d_{k+1})\|^2 \qquad \text{(Young's inequality)}$$
$$\leq \frac{k}{k+1}\|w_k - w_0\|^2 + \frac{4\|w_0 - w^*\|^2}{k+1}. \qquad \text{(Theorem 3.7)}$$

This implies $(k+1)\|w_{k+1} - w_0\|^2 \leq k\|w_k - w_0\|^2 + 4\|w_0 - w^*\|^2$. By induction, it is easy to see that $\|w_k - w_0\|^2 \leq 4\|w_0 - w^*\|^2$ for all $k \geq 0$. $\qquad\square$

Lemma C.2 (Bounded Iterates of composite-EAG). *The iterates $\{z_k\}_{k\geq 0}$ of composite-EAG satisfies for all $k \geq 0$*

$$\|z_k - z_0\|^2 \leq 16040H^2.$$

Proof. For $k = 1$, $\|z_1 - z_0\|^2 \leq H^2$ is clear. By definitions of composite-EAG and c_{k+1}, we have

$$
\begin{aligned}
\|z_{k+1} - z_0\|^2 &= \left\| \frac{k}{k+1}(z_k - z_0) - \eta(F(z_{k+\frac{1}{2}}) + d_{k+1}) \right\|^2 \\
&\leq \left(1 + \frac{1}{k}\right) \frac{k^2}{(k+1)^2} \|z_k - z_0\|^2 + (1+k)\eta^2 \left\| (F(z_{k+\frac{1}{2}}) + d_{k+1}) \right\|^2 && \text{(Young's inequality)} \\
&\leq \frac{k}{k+1} \|z_k - z_0\|^2 + 2(1+k)\eta^2 \left(\left\| F(z_{k+\frac{1}{2}}) - F(z_{k+1}) \right\|^2 + \|F(z_{k+1}) + d_{k+1}\|^2 \right) \\
&\leq \frac{k}{k+1} \|z_k - z_0\|^2 + \frac{1}{k+1} 16040 H^2. && \text{(Corollary 3.6 and Theorem 3.2)}
\end{aligned}
$$

By induction, we get $\|z_{k+1} - z_0\|^2 \leq 16040 H^2$. $\qquad\qquad\square$

Combining Lemma C.1 and Lemma C.2 directly implies that the iterates of OHM and composite-EAG are at most constant away from each other.

Corollary C.3. *Let $w_0 = z_0$. The iterates $\{z_k\}_{k \geq 0}$ of composite-EAG and the iterates $\{w_k\}_{k \geq 0}$ of OHM satisfies for all $k \geq 0$,*

$$
\|w_k - z_k\| \leq \sqrt{(32080 + 8)H^2} \leq 200H.
$$

D. Missing Proofs in Section 4

D.1. Proof of Proposition 4.3

Proof. Note that $z_{\frac{1}{2}} = z_0$ and $z_1 = J_{\eta A}[z_0 - \eta F(z_0)]$. Thus we have $\eta c_1 = z_0 - \eta F(z_0) - z_1$. We first bound $\|\eta F(z_1) + \eta c_1\|$ as follows:

$$
\begin{aligned}
\|\eta F(z_1) + \eta c_1\| &= \|z_0 - z_1 + \eta F(z_1) - \eta F(z_0)\| \\
&\leq \|z_0 - z_1\| + \|\eta F(z_1) - \eta F(z_0)\| && \text{(Triangle inequality)} \\
&\leq (1 + \eta L) \cdot \|z_0 - z_1\|. && \text{(F is L-Lipschitz)}
\end{aligned}
$$

Then we can bound U_1 as follows:

$$
\begin{aligned}
U_1 &= \frac{1}{2} \cdot \|\eta F(z_1) + \eta c_1\|^2 + \langle \eta F(z_1) + \eta c_1, z_1 - z_0 \rangle \\
&\leq \frac{1}{2} \cdot \|\eta F(z_1) + \eta c_1\|^2 + \|\eta F(z_1) + \eta c_1\| \cdot \|z_1 - z_0\| && \text{(Cauchy-Schwarz Inequality)} \\
&\leq \left(\frac{(1 + \eta L)^2}{2} + (1 + \eta L) \right) \cdot \|z_1 - z_0\|^2 \\
&\leq \frac{(1 + \eta L)(3 + \eta L)}{2} \cdot \|z_1 - z_0\|^2.
\end{aligned}
$$

Moreover, for any $c \in A(z_0)$, we have

$$
\begin{aligned}
\|z_1 - z_0\| &= \|J_{\eta A}[z_0 - \eta F(z_0)] - J_{\eta A}[z_0 + \eta c]\| \\
&\leq \eta \cdot \|F(z_0) + c\|. && \text{($J_{\eta A}$ is non-expansive)}
\end{aligned}
$$

Hence $\|z_1 - z_0\| \leq \eta \cdot \min_{c \in A(z_0)} \|F(z_0) + c\| = \eta \cdot r_{F,A}^{tan}(z_0)$. $\qquad\qquad\square$

D.2. Proof of Lemma 4.4

Proof. Fix any $k \geq 1$. We first present several inequalities. Since F is L-Lipschitz, we have

$$
\left(-\frac{(k+1)^2}{2} \right) \cdot \left(\eta^2 L^2 \cdot \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 - \left\| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}) \right\|^2 \right) \leq 0. \tag{10}
$$

Additionally, since $F + A$ is ρ-comonotone, $c_k \in A(z_k)$, and $c_{k+1} \in A(z_{k+1})$, we have

$$(-k(k+1)) \cdot \Big(\langle \eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k, z_{k+1} - z_k \rangle$$

$$- \frac{\rho}{\eta} \| \eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k \|^2 \Big) \leq 0. \tag{11}$$

The following identity holds due to Identity (25) in Proposition F.2: we treat x_0 as z_0; x_t as $z_{k+\frac{t-1}{2}}$ for $t \in \{1, 2, 3\}$; y_t as $\eta F(z_{k+\frac{t-1}{2}})$ for $t \in \{1, 2, 3\}$; u_1 as ηc_k, and u_3 as ηc_{k+1}; p as $\eta^2 L^2$, q as k, and r as $\frac{\rho}{\eta}$. Note that by the update rule of composite-FEG, we have $\eta c_k = \frac{z_k + \frac{1}{k+1}(z_0 - z_k) - \frac{k}{k+1}(1 + 2\frac{\rho}{\eta}) \cdot \eta F(z_k) - z_{k+\frac{1}{2}}}{\frac{k}{k+1}(1 + 2\frac{\rho}{\eta})}$, and by definition, we have $\eta c_{k+1} = z_k + \frac{1}{k+1}(z_0 - z_k) - \eta F(z_{k+\frac{1}{2}}) - \frac{2k\frac{\rho}{\eta}}{k+1} \cdot (\eta F(z_k) + \eta c_k) - z_{k+1}$.

$$U_k - U_{k+1} + \text{LHS of Inequality (10)} + \text{LHS of Inequality (11)}$$

$$= \frac{(1 - \eta^2 L^2)(k+1)^2}{2} \cdot \left\| z_{k+1} - z_{k+\frac{1}{2}} \right\|^2.$$

This completes the proof. $\qquad\square$

The following corollary will be used in the proof of Theorem 4.6.

Corollary D.1. *In the same setup as Theorem 4.1 but we assume $\rho = 0$, then we have*

$$\sum_{k=1}^{\infty} (k+1)^2 \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 \leq \frac{2H_0^2}{1 - \eta^2 L^2}.$$

Proof. By Lemma 4.4, we get $\frac{1 - \eta^2 L^2}{2}(k+1)^2 \| z_{k+\frac{1}{2}} - z_{k+1} \|^2 \leq U_k - U_{k+1}$ holds for any $k \geq 1$. Then the claim follows by telescoping the above inequality and notice that $U_1 - U_{k+1} \leq H_0^2$. $\qquad\square$

D.3. Proof of Lemma 4.5

Proof. Fix any $T \geq 1$. According to Lemma 4.4, we have $U_T \leq U_1$. Then by definition of U_T, we have

$$U_T = \left(\frac{T^2}{2}\left(1 + \frac{2\rho}{\eta}\right) - \frac{\rho}{\eta}T \right) \cdot \| \eta F(z_T) + \eta c_T \|^2 + T \cdot \langle \eta F(z_T) + \eta c_T, z_T - z^* \rangle$$

$$+ T \cdot \langle \eta F(z_T) + \eta c_T, z^* - z_0 \rangle$$

$$\geq \left(\frac{T^2}{2}\left(1 + \frac{2\rho}{\eta}\right) - \frac{\rho}{\eta}T \right) \cdot \| \eta F(z_T) + \eta c_T \|^2 + \frac{\rho}{\eta}T \cdot \| \eta F(z_T) + \eta c_T \|^2$$

$$+ T \cdot \langle \eta F(z_T) + \eta c_T, z^* - z_0 \rangle$$

$$= \frac{\eta(\eta + 2\rho)T^2}{2} \cdot \| F(z_T) + c_T \|^2 + T \cdot \langle \eta F(z_T) + \eta c_T, z^* - z_0 \rangle$$

$$\geq \frac{\eta(\eta + 2\rho)T^2}{2} \cdot \| F(z_T) + c_T \|^2 - \frac{\eta(\eta + 2\rho)T^2}{4} \cdot \| F(z_T) + c_T \|^2$$

$$- \frac{\eta}{\eta + 2\rho} \| z_0 - z^* \|^2$$

$$= \frac{\eta(\eta + 2\rho)T^2}{4} \cdot \| F(z_T) + c_T \|^2 - \frac{\eta}{\eta + 2\rho} \| z_0 - z^* \|^2.$$

In the first inequality, we use the fact that z^* is a solution of the CMI with the ρ-comonotone operator $F + A$. In the second inequality, we use $\langle a, b \rangle \geq -\frac{\delta}{4} \| a \|^2 - \frac{1}{\delta} \| b \|^2$ for $\delta > 0$. $\qquad\square$

D.4. Proof of Theorem 4.6

Proof. Using the update rule of composite-FEG and OHM and recall: $\eta c_{k+1} := z_k - \eta F(z_{k+\frac{1}{2}}) + \beta_k(z_0 - z_k) - z_{k+1}$ and $\eta d_{k+1} = w_{k+\frac{1}{2}} - \eta F(w_{k+1}) - w_{k+1}$ we have

$$
\begin{aligned}
z_{k+1} - w_{k+1} &= \left(\beta_k z_0 + (1 - \beta_k)z_k - \eta F(z_{k+\frac{1}{2}}) - \eta c_{k+1}\right) - (\beta_k w_0 + (1 - \beta_k)w_k - \eta F(w_{k+1}) - \eta d_{k+1}) \\
&= (1 - \beta_k)(z_k - w_k) + \eta\left(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right), \qquad\qquad (w_0 = z_0)
\end{aligned}
$$

which implies

$$
\begin{aligned}
\|z_{k+1} - w_{k+1}\|^2 = {}& (1 - \beta_k)^2 \|z_k - w_k\|^2 + \eta^2 \left\|F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\|^2 \\
& + \underbrace{2\left\langle (1 - \beta_k)(z_k - w_k), \eta\left(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right)\right\rangle}_{\mathbf{I}}.
\end{aligned}
$$

We focus on term **I**. We can verify

$$
z_{k+\frac{1}{2}} - w_{k+1} = (1 - \beta_k)(z_k - w_k) - \eta(1 - \beta_k)(F(z_k) + c_k) - \eta(F(w_{k+1}) + d_{k+1}).
$$

Thus term **I** can be rewritten as

$$
\begin{aligned}
\mathbf{I} = {}& 2\left\langle (1 - \beta_k)(z_k - w_k), \eta\left(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right)\right\rangle \\
= {}& 2\left\langle z_{k+\frac{1}{2}} - w_{k+1} + \eta(1 - \beta_k)(F(z_k) + c_k) - \eta(F(w_{k+1}) + d_{k+1}), \eta\left(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right)\right\rangle \\
= {}& 2\eta^2\left\langle (1 - \beta_k)(F(z_k) + c_k) - (F(w_{k+1}) + d_{k+1}), F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle + \\
& + \underbrace{2\eta\left\langle z_{k+\frac{1}{2}} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle}_{A_k}.
\end{aligned}
$$

Combining the above (we keep term A_k for now), we get

$$
\begin{aligned}
& \|z_{k+1} - w_{k+1}\|^2 \\
&= (1 - \beta_k)^2 \|z_k - w_k\|^2 + \eta^2 \left\|F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\|^2 \\
&\quad + 2\eta^2\left\langle (1 - \beta_k)(F(z_k) + c_k) - (F(w_{k+1}) + d_{k+1}), F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle + A_k \\
&= (1 - \beta_k)^2 \|z_k - w_k\|^2 - \eta^2 \|F(w_{k+1}) + d_{k+1}\|^2 + 2\eta^2(1 - \beta_k)\langle F(z_k) + c_k, F(w_{k+1}) + d_{k+1}\rangle \\
&\quad - 2\eta^2(1 - \beta_k)\left\langle F(z_k) + c_k, F(z_{k+\frac{1}{2}}) + c_{k+1}\right\rangle + \eta^2\left\|F(z_{k+\frac{1}{2}}) + c_{k+1}\right\|^2 + A_k. \\
&\leq (1 - \beta_k)^2 \|z_k - w_k\|^2 + \eta^2 \|(1 - \beta_k)(F(z_k) + c_k)\|^2 \qquad\qquad \text{(We use } -a^2 + 2ab \leq b^2\text{)} \\
&\quad - 2\eta^2(1 - \beta_k)\left\langle F(z_k) + c_k, F(z_{k+\frac{1}{2}}) + c_{k+1}\right\rangle + \eta^2\left\|F(z_{k+\frac{1}{2}}) + c_{k+1}\right\|^2 + A_k \\
&= (1 - \beta_k)^2 \|z_k - w_k\|^2 + \eta^2 \left\|(1 - \beta_k)(F(z_k) + c_k) - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\|^2 + A_k \\
&= (1 - \beta_k)^2 \|z_k - w_k\|^2 + \left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2 + A_k,
\end{aligned}
$$

where in the last equality we use the fact that $z_{k+1} - z_{k+\frac{1}{2}} = \eta(1 - \beta_k)(F(z_k) + c_k) - \eta F(z_{k+\frac{1}{2}}) - \eta c_{k+1}$ by update rule of composite-FEG. Plugging $\beta_k = \frac{1}{k+1}$ and multiplying both sides with $(k + 1)^2$ gives

$$
(k + 1)^2 \|z_{k+1} - w_{k+1}\|^2 \leq k^2 \|z_k - w_k\|^2 + (k + 1)^2 \left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2 + (k + 1)^2 A_k.
$$

Telescoping the above gives

$$(k+1)^2\|z_{k+1} - w_{k+1}\|^2 \leq \|z_1 - w_1\|^2 + \sum_{t=1}^{k}(t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2 + \sum_{t=1}^{k}(t+1)^2 A_t.$$

$$\Rightarrow \|z_{k+1} - w_{k+1}\|^2 \leq \frac{\|z_1 - w_1\|^2}{(k+1)^2} + \frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2 + \frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2 A_t$$

It remains to bound $\frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2\|z_{t+1} - z_{t+\frac{1}{2}}\|^2$ and $\frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2 A_t$. By Corollary D.1, we have $\sum_{t=1}^{\infty}(t+1)^2\|z_{t+1} - z_{t+\frac{1}{2}}\|^2 \leq \frac{2H_0^2}{1-\eta^2 L^2}$.

For the term with A_t, we need a more careful analysis. We decompose $A_k = B_k + C_k + D_k$ as follows.

$$A_k = 2\eta\left\langle z_{k+\frac{1}{2}} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle$$

$$= 2\eta\left\langle z_{k+\frac{1}{2}} - z_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle$$

$$+ 2\eta\left\langle z_{k+1} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle$$

$$= \underbrace{2\eta\left\langle z_{k+\frac{1}{2}} - z_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle}_{B_k} + \underbrace{2\eta\left\langle z_{k+1} - w_{k+1}, F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\rangle}_{C_k}$$

$$+ \underbrace{2\eta\langle z_{k+1} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+1}) - c_{k+1}\rangle}_{D_k}.$$

For B_k, we have

$$B_k = 2\eta\left\langle z_{k+\frac{1}{2}} - z_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\rangle$$

$$\leq 2\eta\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|\left(\|F(w_{k+1}) + d_{k+1}\| + \left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\| + \|F(z_{k+1}) + c_{k+1}\|\right)$$

$$\leq 2\eta L\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2 + 2\eta\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|\left(\|F(w_{k+1}) + d_{k+1}\| + \|F(z_{k+1}) + c_{k+1}\|\right)$$

$$\leq \frac{4H_0^2}{(1-\eta^2 L^2)(k+1)^2} + 2\eta\sqrt{\frac{2H_0^2}{(1-\eta^2 L^2)(k+1)^2}}\cdot\left(\frac{2\|z_0 - z^*\|}{\eta(k+1)} + \frac{2H_0}{\eta(k+1)}\right)$$

$$\leq \frac{20H_0^2}{(1-\eta^2 L^2)(k+1)^2},$$

where in the second last inequality we use $\eta L \leq 1$, the last-iterate convergence rate of OHM (Theorem 3.7) and composite-FEG (Theorem 4.1) as well as the bound on $\|z_{k+1} - z_{k+\frac{1}{2}}\|^2$ (Corollary D.1); in the last inequality, we use $H_0 \geq \|z_0 - z^*\|$.

For C_k, we have

$$C_k = 2\eta\left\langle z_{k+1} - w_{k+1}, F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\rangle$$

$$\leq 2\eta\|z_{k+1} - w_{k+1}\|\left\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\|$$

$$\leq 2\|z_{k+1} - w_{k+1}\|\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\| \qquad (F \text{ is } L\text{-Lipschitz and } \eta L \leq 1)$$

$$\leq \frac{12H_0}{\sqrt{1-\eta^2 L^2}}\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|,$$

where in the last inequality we use $\|z_{k+1} - w_{k+1}\| \leq \frac{6H_0}{\sqrt{1-\eta^2 L^2}}$ is bounded by Corollary D.3.

24

For D_k, since $F + N_{\mathcal{Z}}$ is ρ-comonotone, we have

$$
\begin{aligned}
D_k &= 2\eta\langle z_{k+1} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+1}) - c_{k+1}\rangle \\
&\leq -2\rho\eta\|F(w_{k+1}) + d_{k+1} - F(z_{k+1}) - c_{k+1}\|^2 \\
&\leq -4\rho\eta\|F(w_{k+1}) + d_{k+1}\|^2 - 4\rho\eta\|F(z_{k+1}) + c_{k+1}\|^2 \\
&\leq -\frac{4\rho}{\eta} \cdot \frac{4\|z_0 - z^*\|^2}{(k+1)^2} - \frac{4\rho}{\eta} \cdot \frac{4H_0^2}{(1 - \eta^2 L^2)(k+1)^2} &&\text{(convergence rate of OHM and Theorem 4.1)} \\
&\leq \frac{8H_0^2}{(1 - \eta^2 L^2)(k+1)^2}. &&(\frac{\rho}{\eta} \geq -\frac{1}{4})
\end{aligned}
$$

Combining the above bounds for B_k, C_k, D_k, we get

$$
(k+1)^2 A_k \leq \frac{28H_0^2}{1 - \eta^2 L^2} + \frac{12H_0}{\sqrt{1 - \eta^2 L^2}}(k+1)^2\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|.
$$

Since $\sum_{t=1}^\infty (t+1)^2\|z_{t+1} - z_{t+\frac{1}{2}}\|^2 \leq \frac{2H_0^2}{1 - \eta^2 L^2}$, using Cauchy-Schwartz, we get

$$
\begin{aligned}
\sum_{t=1}^k (t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\| &\leq \sqrt{\left(\sum_{t=1}^k (t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2\right) \cdot \left(\sum_{t=1}^k (t+1)^2\right)} \\
&\leq \frac{2H_0}{\sqrt{1 - \eta^2 L^2}}(k+1)^{3/2}.
\end{aligned}
$$

Thus $\frac{12H_0}{\sqrt{1 - \eta^2 L^2}}\sum_{t=1}^k (t+1)^2\|z_{t+1} - z_{t+\frac{1}{2}}\| \leq \frac{24H_0^2}{(1 - \eta^2 L^2)}(k+1)^{3/2}$.

Hence we get

$$
\frac{1}{(k+1)^2}\sum_{t=1}^k (t+1)^2 A_t = \frac{28H_0^2}{(1 - \eta^2 L^2)(k+1)^2} + \frac{24H_0^2}{(1 - \eta^2 L^2)(k+1)^{\frac{1}{2}}}.
$$

Combining all the above, we concludes that for all k,

$$
\begin{aligned}
\|z_{k+1} - w_{k+1}\|^2 &\leq \frac{\|z_1 - w_1\|^2}{(k+1)^2} + \frac{1}{(k+1)^2}\sum_{t=1}^k (t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2 + \frac{1}{(k+1)^2}\sum_{t=1}^k (t+1)^2 A_t \\
&\leq \frac{H_0^2}{1 - \eta^2 L^2} \cdot \left(\frac{24}{(k+1)^{\frac{1}{2}}} + \frac{36}{(k+1)^2}\right).
\end{aligned}
$$

This completes the proof. $\qquad\square$

D.4.1. BOUNDED ITERATES OF COMPOSITE-FEG

In this section, we prove auxiliary results needed in the proof of Theorem 4.6. We show that the iterates of OHM and composite-FEG all have bounded distance away from the initial point z_0.

Lemma D.2 (Bounded Iterates of composite-FEG). *Let $\rho = 0$. The iterates $\{z_k\}_{k\geq 0}$ of composite-FEG satisfies for all $k \geq 0$*

$$
\|z_k - z_0\|^2 \leq \frac{12H_0^2}{1 - \eta^2 L^2}.
$$

Proof. For $k = 1$, $\|z_1 - z_0\|^2 \leq H_0^2$ is clear. By definitions of composite-FEG and c_{k+1}, we have

$$
\begin{aligned}
\|z_{k+1} - z_0\|^2 &= \left\| \frac{k}{k+1}(z_k - z_0) - \eta(F(z_{k+\frac{1}{2}}) + d_{k+1}) \right\|^2 \\
&\leq \left(1 + \frac{1}{k}\right) \frac{k^2}{(k+1)^2} \|z_k - z_0\|^2 + (1+k)\eta^2 \left\| (F(z_{k+\frac{1}{2}}) + d_{k+1}) \right\|^2 && \text{(Young's inequality)} \\
&\leq \frac{k}{k+1} \|z_k - z_0\|^2 + 2(1+k)\eta^2 \left(\left\| F(z_{k+\frac{1}{2}}) - F(z_{k+1}) \right\|^2 + \|F(z_{k+1}) + d_{k+1}\|^2 \right) \\
&\leq \frac{k}{k+1} \|z_k - z_0\|^2 + \frac{1}{k+1} \cdot \frac{12 H_0^2}{1 - \eta^2 L^2}. && \text{(Corollary D.1 and Theorem 4.1)}
\end{aligned}
$$

By induction, we get $\|z_{k+1} - z_0\|^2 \leq \frac{12 H_0^2}{1 - \eta^2 L^2}$. $\qquad\square$

Combining Lemma D.2 and Lemma C.1 gives the following cororllary.

Corollary D.3. *Let $w_0 = z_0$. The iterates $\{z_k\}_{k \geq 0}$ of composite-FEG and the iterates $\{w_k\}_{k \geq 0}$ of OHM satisfies for all $k \geq 0$,*

$$
\|w_k - z_k\| \leq \frac{6 H_0}{\sqrt{1 - \eta^2 L^2}}.
$$

E. proj-EAG for comonotone inclusion with point convergence

In this section, we study constrained variational inequality problem, a special case of CMI where $A = \partial \mathbb{I}_{\mathcal{Z}} = N_{\mathcal{Z}}$ is the normal cone operator of a closed convex set $\mathcal{Z} \subseteq \mathbb{R}^n$. This problem captures constrained non-convex-non-concave min-max optimization. We extend the extra anchored gradient (EAG) algorithm in the unconstrained case to this setting and propose the following projected extra anchored gradient (proj-EAG) algorithm.

$$
\begin{aligned}
z_{k+\frac{1}{2}} &= \Pi_{\mathcal{Z}}[\beta_k z_0 + (1 - \beta_k)z_k - \eta F(z_k)] \\
z_{k+1} &= \Pi_{\mathcal{Z}}[\beta_k z_0 + (1 - \beta_k)z_k - \eta F(z_{k+\frac{1}{2}})]
\end{aligned} \qquad \text{(proj-EAG)}
$$

where $\beta_k = \frac{1}{k+1}$. Note that all the iterates $\{z_k, z_{k+\frac{1}{2}}\}_{k \geq 0}$ lies in the feasible set $\mathcal{Z} \subseteq \mathbb{R}^n$. As a consequence, we only require L-Lipschitzness and ρ-comonotonicity hold on the feasible set $\mathcal{Z}$, but not the whole $\mathbb{R}^n$ space. Specifically, we assume $F : \mathcal{Z} \to \mathcal{Z}$ is L-Lipschitz and $F + N_{\mathcal{Z}}$ is ρ-comonotone with $\rho \geq -\frac{1}{20L}$.

The two main results of the section are:

1. proj-EAG has $O(\frac{1}{T})$ last-iterate convergence rate with respect to the tangent residual.

2. proj-EAG has point convergence

E.1. Last-Iterate Convergence Rate

We use the following potential function for

$$
V_k := \frac{k(k+1)}{2} \cdot \|\eta F(z_k) + \eta c_k\|^2 + k \cdot \langle \eta F(z_k) + \eta c_k, z_k - z_0 \rangle, \qquad k \geq 1,
$$

where $c_k := \frac{\beta_{k-1} z_0 + (1 - \beta_{k-1})z_{k-1} - \eta F(z_{k-\frac{1}{2}}) - z_k}{\eta}$ for $k \geq 1$. Introducing c_k gives the identity $z_{k+1} = \beta_k z_0 + (1 - \beta_k)z_k - \eta F(z_{k+\frac{1}{2}}) - \eta c_{k+1}$ for all $k \geq 0$. Recall that by definition of projection on $\mathcal{Z}$, we have $c_k \in N_{\mathcal{Z}}(z_k)$.

Lemma E.1 (Upper bound of V_1). *In the same setup as Theorem E.2, we have $V_1 \leq 16\|z_0 - z_{\frac{1}{2}}\|^2 \leq 16\eta^2 r_{F,N_{\mathcal{Z}}}^{tan}(z_0)^2$*

Proof. By nonexpansiveness of $\Pi_{\mathcal{Z}}$, L-Lipschitzness of F, and $\eta L \le 1$, we have $\|z_1 - z_{\frac{1}{2}}\|^2 \le \|z_{\frac{1}{2}} - z_0\|^2$. This also implies $\|z_1 - z_0\|^2 \le 2\|z_1 - z_{\frac{1}{2}}\|^2 + 2\|z_{\frac{1}{2}} - z_0\|^2 \le 4\|z_{\frac{1}{2}} - z_0\|^2$. Then we get

$$
\begin{aligned}
\|\eta F(z_1) + \eta c_1\|^2 &= \left\| z_0 - z_1 + \eta F(z_1) - \eta F(z_{\frac{1}{2}}) \right\|^2 \\
&\le 2\|z_0 - z_1\|^2 + 2\left\| z_{\frac{1}{2}} - z_1 \right\|^2 \\
&\le 10 \left\| z_0 - z_{\frac{1}{2}} \right\|^2.
\end{aligned}
$$

Now by definition of V_1,

$$
\begin{aligned}
V_1 &= \|\eta(F(z_1) + c_1)\|^2 + \eta\langle F(z_1) + c_1, z_1 - z_0 \rangle \\
&\le (10 + 2\sqrt{10}) \left\| z_0 - z_{\frac{1}{2}} \right\|^2 \le 16 \left\| z_0 - z_{\frac{1}{2}} \right\|^2.
\end{aligned}
$$

Let $c_0 \in N_{\mathcal{Z}}(z_0)$. Using nonexpansiveness of $J_{\eta A}$, we have $\|z_{\frac{1}{2}} - z_0\| \le \|\eta F(z_0) + \eta c_0\|$. Thus $\|z_{\frac{1}{2}} - z_0\| \le \eta r_{F, N_{\mathcal{Z}}}^{tan}(z_0)$. $\square$

Theorem E.2 (Approximate monotonicity of the potential). *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ is closed convex set and suppose $F : \mathcal{Z} \to \mathbb{R}^n$ is L-Lispchitz and $F + N_{\mathcal{Z}}$ is a ρ-comonotone operator with $0 \ge \rho \ge -\frac{1}{20L}$. Let $z_0 \in \mathcal{Z}$ be any starting point and $\{z_k, z_{k+\frac{1}{2}}\}_{k \ge 0}$ be the iterates of proj-EAG with step size $\eta = \frac{0.31}{L}$ that satisfies Fact 2. Then for any $k \ge 1$,*

$$
V_{k+1} \le V_k + \frac{1}{2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2 - \frac{9k(k+1)}{4000} \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2.
$$

Proof. We first present several inequalities. Let $p = \frac{1}{3}$ and $c = -\frac{4p\rho}{\eta} \ge 0$. Since F is L-Lipschitz, we have

$$
\eta^2 L^2 \cdot \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 - \left\| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}) \right\|^2 \ge 0.
$$

Multiplying the both sides of the above inequality by $(1 + c)$ and rearranging terms, we get

$$
\begin{aligned}
&p \cdot \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 - \left\| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}) \right\|^2 \\
&+ ((1+c)\eta^2 L^2 - p) \cdot \left\| z_{k+\frac{1}{2}} \right\|^2 - c \cdot \left\| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}) \right\|^2 \ge 0..
\end{aligned}
\tag{12}
$$

Since $F + N_{\mathcal{Z}}$ satisfies negative comonotonicity and $0 \in N_{\mathcal{Z}}(z)$ for any $z \in \mathcal{Z}$, we have

$$
(-k(k+1)) \cdot \left(\langle \eta F(z_{k+1}) - \eta F(z_k), z_{k+1} - z_k \rangle - \frac{\rho}{\eta} \cdot \|\eta F(z_{k+1}) - \eta F(z_k)\|^2 \right) \le 0.
\tag{13}
$$

Since $z_{k+\frac{1}{2}} = \Pi_{\mathcal{Z}}[z_k - \eta F(z_k) + \frac{1}{k+1}(z_0 - z_k)]$, we can infer that $z_k - \eta F(z_k) + \frac{1}{k+1}(z_0 - z_k) - z_{k+\frac{1}{2}} \in N_{\mathcal{Z}}(z_{k+\frac{1}{2}})$. Moreover, by definition of c_k and c_{k+1}, we know $c_k \in N_{\mathcal{Z}}(z_k)$ and $c_{k+1} \in N_{\mathcal{Z}}(z_{k+1})$. Therefore, we have

$$
(-k(k+1)) \cdot \left\langle z_k - \eta F(z_k) - z_{k+\frac{1}{2}} + \frac{1}{k+1}(z_0 - z_k), z_{k+\frac{1}{2}} - z_{k+1} \right\rangle \le 0,
\tag{14}
$$

$$
(-k(k+1)) \cdot \langle \eta c_{k+1}, z_{k+1} - z_k \rangle \le 0,
\tag{15}
$$

$$
(-k(k+1)) \cdot \left\langle \eta c_k, z_k - z_{k+\frac{1}{2}} \right\rangle \le 0.
\tag{16}
$$

Moreover, by definition, we have $\eta c_{k+1} = z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+1}(z_0 - z_k) - z_{k+1}$. Then have the following identity holds. The correctness of the identity follows by Equation (23) in Proposition F.2: we treat x_0 as z_0; x_t as $z_{k+\frac{t-1}{2}}$ for $t \in \{1, 2, 3\}$; y_t as $\eta F(z_{k+\frac{t-1}{2}})$ for $t \in \{1, 2, 3\}$; u_1 as ηc_k and u_3 as ηc_{k+1}; q as k. Note that Term (17), (18), and (19) comes from the

27

right hand side of Equation (23), while Term (20) directly comes from Inequality (12) and Term (21) directly comes from Inequality (13).

$$V_k - V_{k+1} - \frac{k(k+1)}{2p} \cdot \text{LHS of Inequality (12)} + \text{LHS of Inequality (13)} + \text{LHS of Inequality (14)}$$

$$+ \text{LHS of Inequality (15)} + \text{LHS of Inequality (16)}$$

$$= \frac{k(k+1)}{2} \cdot \left\| z_{k+\frac{1}{2}} - z_k + \eta F(z_k) + \eta c_k + \frac{1}{k+1}(z_k - z_0) \right\|^2 \tag{17}$$

$$+ \frac{(1-p)k(k+1)}{2p} \cdot \left\| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}) \right\|^2 \tag{18}$$

$$+ (k+1) \cdot \left\langle \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}), \eta F(z_{k+1}) + \eta c_{k+1} \right\rangle \tag{19}$$

$$+ \frac{k(k+1)}{2p} \cdot \left((p - (1+c)\eta^2 L^2) \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 + c \cdot \left\| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}) \right\|^2 \right) \tag{20}$$

$$+ k(k+1) \frac{\rho}{\eta} \| \eta F(z_{k+1}) - \eta F(z_k) \|^2. \tag{21}$$

Since $\|a\|^2 + \langle a, b \rangle = \|a + \frac{b}{2}\|^2 - \frac{\|b\|^2}{4}$, we have

$$\text{Expression}(18) + \text{Expression}(19)$$

$$= \left\| \sqrt{\frac{(1-p)k(k+1)}{2p}} \cdot \left(\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}) \right) + \sqrt{\frac{p(k+1)}{2(1-p)k}} \cdot (\eta F(z_{k+1}) + \eta c_{k+1}) \right\|^2$$

$$- \frac{k+1}{2k} \cdot \frac{p}{1-p} \| \eta F(z_{k+1}) + \eta c_{k+1} \|^2$$

$$\geq -\frac{p}{1-p} \| \eta F(z_{k+1}) + \eta c_{k+1} \|^2. \tag{$k \geq 1$}$$

$$= -\frac{1}{2} \| \eta F(z_{k+1}) + \eta c_{k+1} \|^2.$$

Now it remains to give a non-negative lower bound of Expression (20) + Expression (21). Recall that $p = \frac{1}{3}$ and $c = -\frac{4\rho p}{\eta}$, thus

$$\left(\frac{2}{k(k+1)} \right) \cdot (\text{Expression (20)} + \text{Expression (21)})$$

$$= \left(1 - \left(3 - \frac{4\rho}{\eta} \right) \cdot \eta^2 L^2 \right) \cdot \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 - \frac{4\rho}{\eta} \cdot \left(\cdot \left\| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}) \right\|^2 - \frac{1}{2} \cdot \| \eta F(z_{k+1}) - \eta F(z_k) \|^2 \right)$$

$$\geq \left(1 - \left(3 - \frac{4\rho}{\eta} \right) \cdot \eta^2 L^2 \right) \cdot \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 + \frac{4\rho}{\eta} \cdot \left\| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_k) \right\|^2 \qquad (\|A\|^2 - \frac{1}{2}\|B\|^2 \geq -\|A - B\|^2)$$

$$\geq \left(1 + \frac{4\rho}{\eta} - \left(3 - \frac{4\rho}{\eta} \right) \cdot \eta^2 L^2 \right) \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2,$$

where the last inequality holds because $\| z_{k+\frac{1}{2}} - z_{k+1} \|^2 \leq \| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_k) \|^2$, which is due to update rule of proj-EAG and the fact that Π_Z is non-expansive. Hence, we have $V_{k+1} \leq V_k + \frac{1}{2} \| \eta F(z_{k+1}) + \eta c_{k+1} \|^2 - (1 + \frac{4\rho}{\eta} - (3 - \frac{4\rho}{\eta}) \cdot \eta^2 L^2) \frac{k(k+1)}{2} \| z_{k+\frac{1}{2}} - z_{k+1} \|^2$. Plugging $\eta = \frac{0.31}{L}$, we have $(1 + \frac{4\rho}{\eta} - (3 - \frac{4\rho}{\eta}) \cdot \eta^2 L^2) \geq \frac{9}{2000}$ (by Fact 2). This completes the proof. $\qquad \square$

We also show that V_k is of order $\Omega(k^2 \cdot r^{tan}(z_k))$.

Lemma E.3. *In the same setup as Theorem E.2, for any $k \geq 1$, we have*

$$\frac{k(k+1)}{4} \| \eta F(z_k) + \eta c_k \|^2 \leq V_k + \| z_0 - z^* \|^2$$

In particular, $V_k \geq -\| z_0 - z^* \|^2$.

Proof. Fix any $k \geq 1$. Since $0 \in F(z^*) + N_{\mathcal{Z}}(z^*)$, by ρ-comonotonicity and Fact 2, we have

$$\langle \eta F(z_k) + \eta c_k, z_k - z^* \rangle \geq \frac{\rho}{\eta} \|\eta F(z_k) + \eta c_k\|^2 > -\frac{1}{4} \|\eta F(z_k) + \eta c_k\|^2. \tag{22}$$

By definition of V_k, we have

$$
\begin{aligned}
V_k &= \frac{k(k+1)}{2} \cdot \|\eta F(z_k) + \eta c_k\|^2 + k \cdot \langle \eta F(z_k) + \eta c_k, z_k - z_0 \rangle \\
&= \frac{k(k+1)}{2} \cdot \|\eta F(z_k) + \eta c_k\|^2 + k \cdot \langle \eta F(z_k) + \eta c_k, z_k - z^* \rangle + k \langle \eta F(z_k) + \eta c_k, z^* - z_0 \rangle \\
&\geq \frac{k(k+1)}{2} \cdot \|\eta F(z_k) + \eta c_k\|^2 - \frac{1}{4} \|\eta F(z_k) + \eta c_k\|^2 + k \langle \eta F(z_k) + \eta c_k, z^* - z_0 \rangle && \text{(By Equation (22))} \\
&\geq \frac{k(k+\frac{1}{2})}{2} \cdot \|\eta F(z_k) + \eta c_k\|^2 - \frac{k^2}{4} \|\eta F(z_k) + \eta c_k\|^2 - \|z_0 - z^*\|^2 \\
&= \frac{k(k+1)}{4} \cdot \|\eta F(z_k) + \eta c_k\|^2 - \|z_0 - z^*\|^2,
\end{aligned}
$$

where in the second last inequality, we apply $\langle a, b \rangle \geq -\frac{\alpha}{4}\|a^2\| - \frac{1}{\alpha}\|b^2\|$ with $a = \eta F(z_k) + \eta c_k$, $b = z^* - z_0$, and $\alpha = k$. This completes the proof. $\qquad \square$

Combining the approximate monotonicity of the potential function V_k in Theorem E.2 and the above lower bound of V_k, we can prove the $O(\frac{1}{T})$ last-iterate convergence rate easily.

Theorem E.4 (Last-iterate convergence rate of (proj-EAG) for comonotone inclusion problem). *In the same setup as Theorem E.2, we have for any $k \geq 1$,*

$$\|F(z_k) + c_k\|^2 \leq \frac{20(V_1 + \|z_0 - z^*\|^2)}{\eta^2 k^2} \leq \frac{20H^2}{\eta^2 k^2}.$$

where $H^2 = 16\|z_{\frac{1}{2}} - z_0\|^2 + \|z_0 - z^*\|^2 \leq 16\eta^2 r_{F,N_{\mathcal{Z}}}^{tan}(z_0)^2 + \|z_0 - z^*\|^2$.

Proof. For $k = 1$, from Lemma E.3, we directly get that $\|\eta F(z_1) + \eta c_1\|^2 \leq 4(V_1 + \|z_0 - z^*\|^2)$.

Now fix any $k \geq 2$. Combining Theorem E.2 and Lemma E.3, we have

$$
\begin{aligned}
\frac{k(k+1)}{4} \|\eta F(z_k) + \eta c_k\|^2 &\leq V_k + \|z_0 - z^*\|^2 \\
&\leq V_1 + \|z_0 - z^*\|^2 + \frac{1}{2} \cdot \sum_{t=2}^{k} \|\eta F(z_t) + \eta c_t\|^2.
\end{aligned}
$$

Subtracting $\frac{1}{2}\|F(z_k) + \eta c_k\|^2$ from both sides and noting that $\frac{1}{4}k \geq \frac{1}{2}$ for $k \geq 2$ gives

$$\frac{k^2}{4} \|\eta F(z_k) + \eta c_k\|^2 \leq V_1 + \|z_0 - z^*\|^2 + \frac{1}{2} \cdot \sum_{t=2}^{k-1} \|\eta F(z_t) + \eta c_t\|^2$$

Now we can apply Proposition F.3 with $C_1 = V_1 + \|z_0 - z^*\|^2$ to conclude that $\|\eta F(z_k) + \eta c_k\|^2 \leq \frac{20(V_1 + \|z_0 - z^*\|^2)}{k^2}$. Applying Lemma E.1, we know $V_1 \leq 16\|z_{\frac{1}{2}} - z_0\|^2 \leq 16\eta^2 r_{F,N_{\mathcal{Z}}}^{tan}(z_0)^2$ gives the desired result. $\qquad \square$

Corollary E.5. *In the same setup as Theorem E.2, we have*

$$\sum_{k=1}^{\infty} k(k+1) \left\| z_{k+1} - z_{k+\frac{1}{2}} \right\|^2 \leq 4000 H^2.$$

It also implies $\|z_{k+1} - z_{k+\frac{1}{2}}\|^2 \leq \frac{8000H^2}{(k+1)^2}$ for all $k \geq 1$.

Proof. By Theorem E.2, we have

$$\frac{9k(k+1)}{4000}\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2 \le V_k - V_{k+1} + \frac{1}{2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2$$

Telescoping the above inequality for $k = 1, 2, \ldots, T$ gives

$$\frac{9}{4000}\sum_{k=1}^{T} k(k+1)\left\|z_{k+1} - z_{k+\frac{1}{2}}\right\|^2 \le V_1 - V_{T+1} + \frac{1}{2}\sum_{k=1}^{T}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2$$

$$\le V_1 + \|z_0 - z^*\|^2 + 10H^2 \cdot \left(\sum_{k=1}^{T}\frac{1}{(t+1)^2}\right)$$

$$\le 9H^2.$$

where in the second inequality we use (1) $V_{T+1} \ge -\|z_0 - z^*\|^2$ by Lemma E.3 and (2) $\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2 \le \frac{20H^2}{(t+1)^2}$ by Theorem E.2, in the last inequality we use $\sum_{k=1}^{\infty}\frac{1}{(t+1)^2} = \frac{\pi^2}{6} - 1 \le \frac{3}{4}$. This concludes $\sum_{k=1}^{\infty} k(k+1)\|z_{k+1} - z_{k+\frac{1}{2}}\|^2 \le 4000H^2$. Since $\frac{1}{2}(k+1)^2 \le k(k+1)$ for all $k \ge 1$, it further implies $\|z_{k+1} - z_{k+\frac{1}{2}}\|^2 \le \frac{8000H^2}{(k+1)^2}$ for all $k \ge 1$. $\qquad\square$

E.2. Point Convergence of proj-EAG in the Comonotone Case

In the following, we show that $\|z_{k+1} - w_{k+1}\| \to 0$, i.e., the trajectory of proj-EAG and OHM merges. This would directly prove the point convergence of OHM.

Theorem E.6. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F : \mathcal{Z} \to \mathcal{Z}$ be L-Lipschitz, and $F + N_{\mathcal{Z}}$ be ρ-comonotone with $\rho \ge -\frac{1}{20L}$. Let $\beta_k = \frac{1}{k+1}$, $\eta = \frac{0.31}{L}$, and $w_0 = z_0$. Then the iterates $\{z_k\}_{k \ge 0}$ of proj-EAG and the iterates $\{w_k\}_{k \ge 0}$ of OHM merges, i.e.,*

$$\|z_{k+1} - w_{k+1}\|^2 \le \frac{40000H^2}{(k+1)^{\frac{1}{2}}} + \frac{30000H^2}{(k+1)^2} = O(\frac{H^2}{(k+1)^{\frac{1}{2}}})$$

In particular, $\lim_{k\to\infty}\|z_{k+1} - w_{k+1}\| = 0$.

Proof. Using the update rule of proj-EAG and OHM and recall: $\eta c_{k+1} := z_k - \eta F(z_{k+\frac{1}{2}}) + \beta_k(z_0 - z_k) - z_{k+1}$ and $\eta d_{k+1} = w_{k+\frac{1}{2}} - \eta F(w_{k+1}) - w_{k+1}$ we have

$$z_{k+1} - w_{k+1} = \left(\beta_k z_0 + (1-\beta_k)z_k - \eta F(z_{k+\frac{1}{2}}) - \eta c_{k+1}\right) - \left(\beta_k w_0 + (1-\beta_k)w_k - \eta F(w_{k+1}) - \eta d_{k+1}\right)$$

$$= (1-\beta_k)(z_k - w_k) + \eta\left(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right), \qquad (w_0 = z_0)$$

which implies

$$\|z_{k+1} - w_{k+1}\|^2 = (1-\beta_k)^2\|z_k - w_k\|^2 + \eta^2\left\|F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right\|^2$$

$$+ \underbrace{2\left\langle(1-\beta_k)(z_k - w_k), \eta\left(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\right)\right\rangle}_{\mathbf{I}}.$$

We focus on term **I**. Denote $c_{k+\frac{1}{2}} := \frac{\beta_k z_0 + (1-\beta_k)z_k - \eta F(z_k) - z_{k+\frac{1}{2}}}{\eta} \in N_{\mathcal{Z}}(z_{k+\frac{1}{2}})$. We can verify

$$z_{k+\frac{1}{2}} - w_{k+1} = (1-\beta_k)(z_k - w_k) - \eta\left(F(z_k) + c_{k+\frac{1}{2}} - F(w_{k+1}) - d_{k+1}\right).$$

30

Thus term **I** can be rewritten as

$$
\begin{aligned}
\mathbf{I} &= 2\Big\langle (1-\beta_k)(z_k - w_k), \eta\Big(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big)\Big\rangle \\
&= 2\Big\langle z_{k+\frac{1}{2}} - w_{k+1} + \eta\Big(F(z_k) + c_{k+\frac{1}{2}} - F(w_{k+1}) - d_{k+1}\Big), \eta\Big(F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big)\Big\rangle \\
&= 2\eta^2\Big\langle F(z_k) + c_{k+\frac{1}{2}} - F(w_{k+1}) - d_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle + \\
&\quad + \underbrace{2\eta\Big\langle z_{k+\frac{1}{2}} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle}_{A_k}.
\end{aligned}
$$

Combining the above (we keep term A_k for now), we get

$$
\begin{aligned}
\|z_{k+1} &- w_{k+1}\|^2 \\
&\leq (1-\beta_k)^2 \|z_k - w_k\|^2 + \eta^2\Big\|F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\|^2 \\
&\quad + 2\eta^2\Big\langle F(z_k) + c_{k+\frac{1}{2}} - F(w_{k+1}) - d_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle + A_k \\
&= (1-\beta_k)^2 \|z_k - w_k\|^2 - \eta^2\|F(w_{k+1}) + d_{k+1}\|^2 + 2\eta^2\Big\langle F(z_k) + c_{k+\frac{1}{2}}, F(w_{k+1}) + d_{k+1}\Big\rangle \\
&\quad - 2\eta^2\Big\langle F(z_k) + c_{k+\frac{1}{2}}, F(z_{k+\frac{1}{2}}) + c_{k+1}\Big\rangle + \eta^2\Big\|F(z_{k+\frac{1}{2}}) + c_{k+1}\Big\|^2 + A_k. \\
&\leq (1-\beta_k)^2 \|z_k - w_k\|^2 + \eta^2\Big\|F(z_k) + c_{k+\frac{1}{2}}\Big\|^2 \qquad\qquad (\text{We use } -a^2 + 2ab \leq b^2) \\
&\quad - 2\eta^2\Big\langle F(z_k) + c_{k+\frac{1}{2}}, F(z_{k+\frac{1}{2}}) + c_{k+1}\Big\rangle + \eta^2\Big\|F(z_{k+\frac{1}{2}}) + c_{k+1}\Big\|^2 + A_k \\
&= (1-\beta_k)^2 \|z_k - w_k\|^2 + \eta^2\Big\|F(z_k) + c_{k+\frac{1}{2}} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\|^2 + A_k \\
&= (1-\beta_k)^2 \|z_k - w_k\|^2 + \Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\|^2 + A_k
\end{aligned}
$$

Plugging $\beta_k = \frac{1}{k+1}$ and multiplying both sides with $(k+1)^2$ gives

$$
(k+1)^2 \|z_{k+1} - w_{k+1}\|^2 \leq k^2 \|z_k - w_k\|^2 + (k+1)^2\Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\|^2 + (k+1)^2 A_k.
$$

Telescoping the above gives

$$
(k+1)^2 \|z_{k+1} - w_{k+1}\|^2 \leq \|z_1 - w_1\|^2 + \sum_{t=1}^{k}(t+1)^2\Big\|z_{t+1} - z_{t+\frac{1}{2}}\Big\|^2 + \sum_{t=1}^{k}(t+1)^2 A_t.
$$

$$
\Rightarrow \|z_{k+1} - w_{k+1}\|^2 \leq \frac{\|z_1 - w_1\|^2}{(k+1)^2} + \frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2\Big\|z_{t+1} - z_{t+\frac{1}{2}}\Big\|^2 + \frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2 A_t
$$

It remains to show that $\frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2\|z_{t+1} - z_{t+\frac{1}{2}}\|^2 = o(1)$ and $\frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2 A_t = o(1)$.

For the first term, using Corollary E.5, we have

$$
\begin{aligned}
\sum_{t=1}^{k}(t+1)^2\Big\|z_{t+1} - z_{t+\frac{1}{2}}\Big\|^2 &\leq 2\sum_{t=1}^{k}t(t+1)\Big\|z_{t+1} - z_{t+\frac{1}{2}}\Big\|^2 \\
&\leq 8000 H^2.
\end{aligned}
$$

We need a more careful analysis of A_k for the second term. We decompose $A_k = B_k + C_k + D_k$ as follows.

$$
\begin{aligned}
A_k &= 2\eta\Big\langle z_{k+\frac{1}{2}} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle \\
&= 2\eta\Big\langle z_{k+\frac{1}{2}} - z_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle \\
&\quad + 2\eta\Big\langle z_{k+1} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle \\
&= \underbrace{2\eta\Big\langle z_{k+\frac{1}{2}} - z_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle}_{B_k} + \underbrace{2\eta\Big\langle z_{k+1} - w_{k+1}, F(z_{k+1}) - F(z_{k+\frac{1}{2}})\Big\rangle}_{C_k} \\
&\quad + \underbrace{2\eta\langle z_{k+1} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+1}) - c_{k+1}\rangle}_{D_k}.
\end{aligned}
$$

For B_k, we have

$$
\begin{aligned}
B_k &= 2\eta\Big\langle z_{k+\frac{1}{2}} - z_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+\frac{1}{2}}) - c_{k+1}\Big\rangle \\
&\leq 2\eta\Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\|\Big(\|F(w_{k+1}) + d_{k+1}\| + \Big\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\Big\| + \|F(z_{k+1}) + c_{k+1}\|\Big) \\
&\leq 2\eta L\Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\|^2 + 2\eta\Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\|\big(\|F(w_{k+1}) + d_{k+1}\| + \|F(z_{k+1}) + c_{k+1}\|\big) \\
&\leq \frac{16000H^2}{(k+1)^2} + 2\eta\sqrt{\frac{8000H^2}{(k+1)^2}}\cdot\left(\frac{2\|z_0 - z^*\|}{\eta(k+1)} + \frac{\sqrt{20H^2}}{\eta(k+1)}\right) \\
&\leq \frac{17140H^2}{(k+1)^2}.
\end{aligned}
$$

where in the second last inequality we use the last-iterate convergence rate of OHM and proj-EAG (Theorem E.4) as well as the bound on $\|z_{k+1} - z_{k+\frac{1}{2}}\|^2$ (Corollary E.5).

For C_k, we have

$$
\begin{aligned}
C_k &= 2\eta\Big\langle z_{k+1} - w_{k+1}, F(z_{k+1}) - F(z_{k+\frac{1}{2}})\Big\rangle \\
&\leq 2\eta\|z_{k+1} - w_{k+1}\|\Big\|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\Big\| \\
&\leq 2\|z_{k+1} - w_{k+1}\|\Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\| \qquad\qquad (F \text{ is } L\text{-Lipschitz and } \eta L \leq 1) \\
&\leq 400H\Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\|,
\end{aligned}
$$

where in the last inequality we use $\|z_{k+1} - w_{k+1}\| \leq 90H$ is bounded by Corollary E.8.

For D_k, since $F + N_{\mathcal{Z}}$ is ρ-comonotone, we have

$$
\begin{aligned}
D_k &= 2\eta\langle z_{k+1} - w_{k+1}, F(w_{k+1}) + d_{k+1} - F(z_{k+1}) - c_{k+1}\rangle \\
&\leq -2\rho\eta\|F(w_{k+1}) + d_{k+1} - F(z_{k+1}) - c_{k+1}\|^2 \\
&\leq -4\rho\eta\|F(w_{k+1}) + d_{k+1}\|^2 - 4\rho\eta\|F(z_{k+1}) + c_{k+1}\|^2 \\
&\leq -\frac{4\rho}{\eta}\cdot\frac{4\|z_0 - z^*\|^2}{(k+1)^2} - \frac{4\rho}{\eta}\cdot\frac{20H^2}{(k+1)^2} \qquad\qquad (\text{convergence rate of OHM and Theorem E.4}) \\
&\leq \frac{24H^2}{(k+1)^2}. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\tfrac{\rho}{\eta} \geq -\tfrac{1}{4})
\end{aligned}
$$

Combining the above bounds for B_k, C_k, D_k, we get

$$
(k+1)^2 A_k \leq 17424H^2 + 400H(k+1)^2\Big\|z_{k+1} - z_{k+\frac{1}{2}}\Big\|.
$$

32

Since $\sum_{t=1}^{\infty}(t+1)^2\|z_{t+1} - z_{t+\frac{1}{2}}\|^2 \leq 3200H^2$, using Cauchy-Schwartz, we get

$$\sum_{t=1}^{k}(t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\| \leq \sqrt{\left(\sum_{t=1}^{k}(t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2\right) \cdot \left(\sum_{t=1}^{k}(t+1)^2\right)}$$
$$\leq 100H(k+1)^{3/2}.$$

Thus $\frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2\|z_{t+1} - z_{t+\frac{1}{2}}\| \leq 100H(k+1)^{-\frac{1}{2}}$.

Hence we get

$$\frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2 A_t = \frac{17424H^2}{(k+1)^2} + \frac{40000H^2}{(k+1)^{\frac{1}{2}}}$$

Combining all the above, we conclude that for all k,

$$\|z_{k+1} - w_{k+1}\|^2 \leq \frac{\|z_1 - w_1\|^2}{(k+1)^2} + \frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2 + \frac{1}{(k+1)^2}\sum_{t=1}^{k}(t+1)^2 A_t$$
$$\leq \frac{40000H^2}{(k+1)^{\frac{1}{2}}} + \frac{30000H^2}{(k+1)^2}.$$

This completes the proof. $\qquad\square$

Lemma E.7 (Bounded Iterates of proj-EAG). *The iterates $\{z_k\}_{k\geq0}$ of proj-EAG satisfies for all $k \geq 0$*

$$\|z_k - z_0\|^2 \leq 16040H^2.$$

Proof. For $k = 1$, the proof of Lemma E.1 gives $\|z_1 - z_0\|^2 \leq H^2$. By definitions of proj-EAG and c_{k+1}, we have

$$\|z_{k+1} - z_0\|^2 = \left\|\frac{k}{k+1}(z_k - z_0) - \eta(F(z_{k+\frac{1}{2}}) + d_{k+1})\right\|^2$$
$$\leq \left(1 + \frac{1}{k}\right)\frac{k^2}{(k+1)^2}\|z_k - z_0\|^2 + (1+k)\eta^2\left\|(F(z_{k+\frac{1}{2}}) + d_{k+1})\right\|^2 \qquad \text{(Young's inequality)}$$
$$\leq \frac{k}{k+1}\|z_k - z_0\|^2 + 2(1+k)\eta^2\left(\left\|F(z_{k+\frac{1}{2}}) - F(z_{k+1})\right\|^2 + \|F(z_{k+1}) + d_{k+1}\|^2\right)$$
$$\leq \frac{k}{k+1}\|z_k - z_0\|^2 + \frac{1}{k+1}16040H^2. \qquad \text{(Corollary E.5 and Theorem E.4)}$$

By induction, we get $\|z_{k+1} - z_0\|^2 \leq 16040H^2$. $\qquad\square$

Corollary E.8. *Let $w_0 = z_0$. The iterates $\{z_k\}_{k\geq0}$ of proj-EAG and the iterates $\{w_k\}_{k\geq0}$ of OHM satisfies for all $k \geq 0$,*

$$\|w_k - z_k\| \leq \sqrt{(32080 + 8)H^2} \leq 200H.$$

F. Auxiliary Propositions

Remark F.1. The proofs for our algorithms' last-iterate convergence rates are all based on potential function arguments. The core of these arguments is to prove the potential function's (approximate) monotonicity. The three identities below simplify tedious algebraic calculations and are useful to show the potential functions' (approximate) monotonicity.

Proposition F.2. *Let $x_0, x_1, x_2, x_3, y_1, y_2, y_3, u_1, u_3$ be arbitrary vectors in $\mathbb{R}^n$. Let $q > 0$, $p > 0$, and $r \neq -\frac{1}{2}$ be real numbers.*

If $u_3 = x_1 - y_2 + \frac{1}{q+1}(x_0 - x_1) - x_3$, then the following identity holds:

$$
\begin{aligned}
&\frac{q(q+1)}{2} \cdot \|y_1 + u_1\|^2 + q \cdot \langle y_1 + u_1, x_1 - x_0 \rangle \\
&- \left(\frac{(q+1)(q+2)}{2} \cdot \|y_3 + u_3\|^2 + (q+1) \cdot \langle y_3 + u_3, x_3 - x_0 \rangle \right) \\
&- \frac{q(q+1)}{2p} \cdot \left(p \cdot \|x_2 - x_3\|^2 - \|y_2 - y_3\|^2 \right) \\
&- q(q+1) \cdot \langle y_3 - y_1, x_3 - x_1 \rangle \\
&- q(q+1) \cdot \left\langle x_1 - y_1 - x_2 + \frac{1}{q+1}(x_0 - x_1), x_2 - x_3 \right\rangle \\
&- q(q+1) \cdot \langle u_3, x_3 - x_1 \rangle \\
&- q(q+1) \cdot \langle u_1, x_1 - x_2 \rangle \\
=&\frac{q(q+1)}{2} \cdot \left\| x_2 - x_1 + y_1 + u_1 + \frac{1}{q+1}(x_1 - x_0) \right\|^2 \\
&+ \frac{(1-p)q(q+1)}{2p} \cdot \|y_2 - y_3\|^2 \\
&+ (q+1) \cdot \langle y_2 - y_3, y_3 + u_3 \rangle
\end{aligned}
\tag{23}
$$

If $u_1 = x_1 - y_1 + \frac{1}{q+1}(x_0 - x_1) - x_2$ and $u_3 = x_1 - y_2 + \frac{1}{q+1}(x_0 - x_1) - x_3$, then the following identity holds:

$$
\begin{aligned}
&\frac{q(q+1)}{2} \cdot \|y_1 + u_1\|^2 + q \cdot \langle y_1 + u_1, x_1 - x_0 \rangle \\
&- \left(\frac{(q+1)(q+2)}{2} \cdot \|y_3 + u_3\|^2 + (q+1) \cdot \langle y_3 + u_3, x_3 - x_0 \rangle \right) \\
&- \frac{q(q+1)}{2p} \cdot \left(p \cdot \|x_2 - x_3\|^2 - \|y_2 - y_3\|^2 \right) \\
&- q(q+1) \cdot \langle y_3 + u_3 - y_1 - u_1, x_3 - x_1 \rangle \\
=&\frac{(1-p)q(q+1)}{2p} \cdot \|y_2 - y_3\|^2 \\
&+ (q+1) \cdot \langle y_2 - y_3, y_3 + u_3 \rangle
\end{aligned}
\tag{24}
$$

If $u_1 = \frac{x_1 + \frac{1}{q+1}(x_0 - x_1) - \frac{q}{q+1}(1+2r)y_1 - x_2}{\frac{q}{q+1}(1+2r)}$ and $u_3 = x_1 + \frac{1}{q+1}(x_0 - x_1) - y_2 - \frac{2rq}{q+1}(y_1 + u_1) - x_3$, then the following identity holds:

$$
\begin{aligned}
&\left(\frac{q^2}{2}(1+2r) - rq \right) \cdot \|y_1 + u_1\|^2 + q \cdot \langle y_1 + u_1, x_1 - x_0 \rangle \\
&- \left(\frac{(q+1)^2}{2}(1+2r) - r(q+1) \right) \cdot \|y_3 + u_3\|^2 - (q+1) \cdot \langle y_3 + u_3, x_3 - x_0 \rangle \\
&- \frac{(q+1)^2}{2} \cdot \left(p \cdot \|x_2 - x_3\|^2 - \|y_2 - y_3\|^2 \right) \\
&- q(q+1) \cdot \left(\langle y_3 + u_3 - y_1 - u_1, x_3 - x_1 \rangle - r\|y_3 + u_3 - y_1 - u_1\|^2 \right) \\
=&\frac{(1-p)(q+1)^2}{2} \cdot \|x_2 - x_3\|^2.
\end{aligned}
\tag{25}
$$

Proof. We verify the three identities using MATLAB. Readers can find the verification code at https://github.com/weiqiangzheng1999/Accelerated-Comonotone-Inclusion. □

Proposition F.3. *Let $\{a_k \in \mathbb{R}^+\}_{k \geq 2}$ be a sequence of real numbers. Let $C_1 \geq 0$ and $p \in (0, \frac{1}{3})$ be two real numbers. If the following condition holds for every $k \geq 2$,*

$$\frac{k^2}{4} \cdot a_k \leq C_1 + \frac{p}{1-p} \cdot \sum_{t=2}^{k-1} a_t, \tag{26}$$

then for each $k \geq 2$ we have

$$a_k \leq \frac{4 \cdot C_1}{1 - 3p} \cdot \frac{1}{k^2}. \tag{27}$$

Proof. We prove the statement by induction.

Base Case: $k = 2$. From Inequality (26), we have

$$\frac{2^2}{4} \cdot a_2 \leq C_1 \quad \Rightarrow \quad a_2 \leq C_1 \leq \frac{4 \cdot C_1}{1 - 3p} \cdot \frac{1}{2^2}.$$

Thus, Inequality (27) holds for $k = 2$.

Inductive Step: for any $k \geq 3$. Fix some $k \geq 3$ and assume that Inequality (27) holds for all $2 \leq t \leq k - 1$. We slightly abuse notation and treat the summation in the form $\sum_{t=3}^{2}$ as 0. By Inequality (26), we have

$$
\begin{aligned}
\frac{k^2}{4} \cdot a_k &\leq C_1 + \frac{p}{1-p} \cdot \sum_{t=2}^{k-1} a_t \\
&\leq \frac{C_1}{1-p} + \frac{p}{1-p} \cdot \sum_{t=3}^{k-1} a_t && (a_2 \leq C_1) \\
&\leq \frac{C_1}{1-p} + \frac{4p \cdot C_1}{(1-p)(1-3p)} \cdot \sum_{t=3}^{k-1} \frac{1}{t^2} && \text{(Induction assumption on Inequality (27))} \\
&\leq \frac{C_1}{1-p} + \frac{2p \cdot C_1}{(1-p)(1-3p)} && \left(\sum_{t=3}^{\infty} \frac{1}{t^2} = \frac{\pi^2}{6} - \frac{5}{4} \leq \frac{1}{2}\right) \\
&= \frac{C_1}{1-3p}.
\end{aligned}
$$

This complete the inductive step. Therefore, for all $k \geq 2$, we have $a_k \leq \frac{4 \cdot C_1}{1-3p} \cdot \frac{1}{k^2}$. $\qquad \square$

Proposition F.4. *Let $\{a_k \in \mathbb{R}^+\}_{k \geq 2}$ be a sequence of real numbers and $C_1 \geq 0$. If the following condition holds for every $k \geq 2$,*

$$\frac{k^2}{4} \cdot a_k \leq C_1 + \frac{1}{2} \cdot \sum_{t=2}^{k-1} a_t, \tag{28}$$

then for each $k \geq 2$ we have

$$a_k \leq \frac{20C_1}{k^2} \tag{29}$$

Proof. We prove the statement by induction.

Base Case: $k = 2, 3, 4$. From Inequality (28), we can directly calculate that $a_2, a_3, a_4 \leq C_1$. Thus, Inequality (29) holds for $k = 2, 3, 4$.

Inductive Step: for any $k \geq 5$. Fix some $k \geq 5$ and assume that Inequality (29) holds for all $2 \leq t \leq k-1$. We slightly abuse notation and treat the summation in the form $\sum_{t=5}^{4}$ as 0. By Inequality (28), we have

$$
\begin{aligned}
\frac{k^2}{4} \cdot a_k &\leq C_1 + \frac{1}{2} \cdot \sum_{t=2}^{k-1} a_t \\
&\leq \frac{5}{2}C_1 + \frac{1}{2} \cdot \sum_{t=5}^{k-1} a_t & (a_2, a_3, a_4 \leq C_1) \\
&\leq \frac{5}{2}C_1 + 10C_1 \cdot \sum_{t=5}^{k-1} \frac{1}{t^2} & \text{(Induction assumption on Inequality (27))} \\
&\leq 5C_1. & (\textstyle\sum_{t=5}^{\infty} \frac{1}{t^2} = \leq \frac{1}{4})
\end{aligned}
$$

This complete the inductive step. Therefore, for all $k \geq 2$, we have $a_k \leq \frac{20C_1}{k^2}$. $\qquad\square$