Accelerated Algorithms for Monotone Inclusions and Constrained Nonconvex-Nonconcave Min-Max Optimization

Yang Cai*† Yale University yang.cai@yale.edu

Argyris Oikonomou*† Yale University argyris.oikonomou@yale.edu

Weiqiang Zheng† Yale University weiqiang.zheng@yale.edu

June 13, 2022

Abstract

We study monotone inclusions and monotone variational inequalities, as well as their generalizations to non-monotone settings. We first show that the Extra Anchored Gradient (EAG) algorithm, originally proposed by Yoon and Ryu [2021] for unconstrained convex-concave min-max optimization, can be applied to solve the more general problem of Lipschitz monotone inclusion. More specifically, we prove that the EAG solves Lipschitz monotone inclusion problems with an *accelerated convergence rate* of $O(\frac{1}{T})$, which is *optimal among all first-order methods* [Diakonikolas, 2020, Yoon and Ryu, 2021]. Our second result is a new algorithm, called Extra Anchored Gradient Plus (EAG+), which not only achieves the accelerated $O(\frac{1}{T})$ convergence rate for all monotone inclusion problems, but also exhibits the same accelerated rate for a family of general (non-monotone) inclusion problems that concern negative comonotone operators. As a special case of our second result, EAG+ enjoys the $O(\frac{1}{T})$ convergence rate for solving a non-trivial class of nonconvex-nonconcave min-max optimization problems. Our analyses are based on simple potential function arguments, which might be useful for analysing other accelerated algorithms.

*Supported by a Sloan Foundation Research Fellowship and the NSF Award CCF-1942583 (CAREER).

†Part of this work was done while the author was visiting the Simons Institute for the Theory of Computing.

Contents

1 Introduction

We study the *monotone inclusion* problem and the *monotone variational inequality*, as well as their generalizations in non-monotone settings. Given a closed convex set $\mathcal{Z} \subseteq \mathbb{R}^n$ and a single-valued and *monotone* operator $F : \mathcal{Z} \to \mathbb{R}^n$, i.e.,

$$\langle F(z) - F(z'), z - z' \rangle \geq 0, \quad \forall z, z' \in \mathcal{Z},$$

the monotone inclusion problem (MI) consists in finding a $z^* \in \mathcal{Z}$ such that

$$\mathbf{0} \in F(z^*) + \partial \mathbb{I}_{\mathcal{Z}}(z^*),$$

where $\mathbb{I}_{\mathcal{Z}}(\cdot)$ is the indicator function for set $\mathcal{Z}$,[1] and $\partial \mathbb{I}_{\mathcal{Z}}(\cdot)$ is the subdifferential operator of $\mathbb{I}_{\mathcal{Z}}$. The corresponding monotone variational inequality shares the same input, and asks for a $z^* \in \mathcal{Z}$ such that

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

The monotone inclusion problem and the related monotone variational inequality play a crucial role in mathematical programming, providing unifying settings for the study of optimization and equilibrium problems. They also serve as computational frameworks for numerous important applications in fields such as economics, engineering, probability and statistics, and machine learning [Facchinei and Pang, 2003, Bauschke and Combettes, 2011, Ryu and Boyd, 2016]. It is not hard to observe that the exact solutions to the monotone inclusion problem coincide with the exact solutions to the corresponding variational inequality. Due to the different selected performance measures, the approximate solutions to these two problems differ. Take the unconstrained case for example, i.e., $\mathcal{Z} = \mathbb{R}^n$, an point z approximates the monotone inclusion problem implies that its operator norm $\|F(z)\|$ is small, while an approximate solution to the variational inequality only satisfies a weaker condition, i.e., its gap function is small.[2] Indeed, it is well-known that an approximate solution to the monotone inclusion problem is also an approximate solution to the monotone variational equality, but the reverse is not true in general. Additionally, the type of performance measure used in quantifying the sub-optimality of an approximate solution to monotone inclusion problems is readily extendable to non-monotone settings, e.g., nonconvex-nonconcave min-max optimization, while it is unclear how to provide a meaningful generalization of the gap function to non-monotone settings. We focus on algorithms for inclusion problems for the rest of the paper, but as explained earlier, these algorithms are also applicable to variational inequalities. For computational purposes, we make the standard assumption that the operator F is *L-Lipschitz*.

An important special case of the monotone inclusion problem is the convex-concave min-max optimization problem:

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y),$$

[1] $\mathbb{I}(z) = 0$ for all $z \in \mathcal{Z}$ and $+\infty$ otherwise.

[2] 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, or an approximate *strong* solution, i.e., approximately solve the Stampacchia Variational Inequality.

where $\mathcal{X}$ and $\mathcal{Y}$ are a closed convex sets in $\mathbb{R}^{n_x}$ and $\mathbb{R}^{n_y}$ respectively, and $f(\cdot,\cdot)$ is smooth, convex in x, and concave in y.[3] Besides its central importance in game theory, convex optimization, and online learning, the convex-concave min-max optimization problem has recently received a lot of attention from the machine learning community due to several novel applications such as the generative adversarial networks (GANS) (e.g., [Goodfellow et al., 2014, Arjovsky et al., 2017]), adversarial examples (e.g., [Madry et al., 2017]), robust optimization (e.g., [Ben-Tal et al., 2009]), and reinforcement learning (e.g., [Du et al., 2017, Dai et al., 2018]).

Given the importance of the monotone inclusion problem, it is crucial to understand the following open question.

What is the optimal convergence rate achievable by a first-order method for monotone inclusions? (*)

We provide the first algorithm that achieves the optimal convergence rate and further extend it to a nontrivial class of general inclusion problems that includes, for example, a family of nonconvex-nonconcave min-max optimization problems. Prior to our work, even for the special case of convex-concave min-max optimization, the optimal convergence rate is only known for the relatively weak notion of duality gap [Nemirovski, 2004, Nesterov, 2007], which is also difficult to generalize to nonconvex-nonconcave settings, see [Yoon and Ryu, 2021] for more discussion.

1.1 Our Contributions

A point $z \in \mathcal{Z}$ is an ϵ-approximate solution to a monotone inclusion problem if

$$\mathbf{0} \in F(z) + \partial \mathbb{I}_{\mathcal{Z}}(z) + \mathcal{B}(\mathbf{0}, \epsilon),$$

where $\mathcal{B}(\mathbf{0}, \epsilon)$ is the ball with radius ϵ centered at $\mathbf{0}$. As we argue in Section 2.3, this is equivalent to the tangent residual of z, a notion introduced in [Cai et al., 2022], being no more than ϵ. Our first contribution provides an answer to question (*).

Contribution 1: We extend the Extra Anchored Gradient algorithm (EAG) , originally proposed by Yoon and Ryu [2021] for unconstrained convex-concave min-max problems, to solve monotone inclusion problems, which include constrained convex-concave min-max optimization as a special case. We show in Theorem 4 that EAG finds an $O(\frac{L}{T})$-approximate solution in T iterations for monotone inclusions, where L is the Lipschitz constant of the operator F. The convergence rate we obtain for EAG matches the lower bound by Diakonikolas [2020], Yoon and Ryu [2021], and is therefore optimal for any first-order method.

For the second part of the paper, we go beyond the monotone case and study general inclusion problems (GI) with operators that are not necessarily monotone and only satisfy the weaker ρ-*comonotoncity* (Assumption 2) condition. Given a single-valued, L-Lipschitz, and possibly *non-monotone* operator F and a *set-valued maximally monotone* operator A, we denote $E = F + A$. The general inclusion problem (GI) consists in finding a point $z^* \in \mathbb{R}^n$ that satisfies

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

[3] If we set $F(x,y) = \begin{pmatrix} \nabla_x f(x,y) \\ -\nabla_y f(x,y) \end{pmatrix}$ and $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, then (i) $F(x,y)$ is a Lipschitz and monotone operator, and (ii) the set of saddle points coincide with the solutions to the monotone inclusion problem for operator F and domain $\mathcal{Z}$.

The general inclusion problem (GI) captures (MI) (when $\rho = 0$ and $A = \partial \mathbb{I}_{\mathcal{Z}}$) and a class of *non-smooth nonconvex-nonconcave* min-max optimization problems (when $\rho < 0$ and A chosen appropriately, see Example 1). Our second contribution is a new algorithm that achieves accelerated rate for solving GI.

Contribution 2: We design a new algorithm EAG+ that finds a $O(\frac{L}{T})$-approximate solution to the general inclusion problem (GI) in T iterations as long as $E = F + A$ is a ρ-comonotone operator with $\rho \geq -\frac{1}{c \cdot L}$ for some $c > 0$.[a] See Theorem 7 for the formal statement.

[a] We have not optimized the range of permissible ρ, and our constant c is slightly worse than some in the literature [Diakonikolas et al., 2021, Lee and Kim, 2021]

Our result is the first to obtain accelerated rate for constrained nonconvex-nonconcave min-max optimization and more generally non-monotone inclusion problems.

1.2 Related Work

There is a vast literature on general inclusion problems and variational inequalities, e.g., see [Facchinei and Pang, 2003, Bauschke and Combettes, 2011, Ryu and Boyd, 2016] and the references therein. We only provide a brief discussion of the most relevant and recent results.

1.2.1 Convex-Concave and Monotone Settings

Convergence in Gap Function. [Nemirovski, 2004, Nesterov, 2007] show that the average iterate of extragradient-type methods has $O(\frac{1}{T})$ convergence rate in terms of gap function defined as $\max_{z' \in \mathcal{Z}} \langle F(z'), z - z' \rangle$, which means that their result only provides an approximate solution to the weak solution. The $O(\frac{1}{T})$ rate is optimal for first-order methods due to the lower bound by [Ouyang and Xu, 2021].

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 [Korpelevich, 1976, Facchinei and Pang, 2003] show that the best-iterate of the extragradient method converges at a rate of $O(\frac{1}{\sqrt{T}})$. Recently, the same convergence rate is shown to hold even for the last-iterate of the extragradient method [Gorbunov et al., 2021, Cai et al., 2022]. Although $O(\frac{1}{\sqrt{T}})$ convergence on the residual is optimal for all p-SCIL algorithms [Golowich et al., 2020], 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.

Faster Convergence Rate in Operator Norm or Residual. We provide a brief overview of results that achieve faster convergence rate in terms of the operator norm or residual. Note that these results also imply essentially the same convergence rate in terms of the gap function. The literature here is rich and fast-growing, we only discuss the ones that are close related to our paper. Recent results show accelerated rates through Halpern iteration [Halpern, 1967] or a similar mechanism – anchoring. Implicit versions of Halpern iteration have $O(\frac{1}{T})$ convergence rate

[Kim, 2021, Lieder, 2021, Park and Ryu, 2022] for monotone operators and explicit variants of Halpern iteration achieve the same convergence rate when F is cocoercive [Diakonikolas, 2020, Kim, 2021]. Diakonikolas [2020] also provide a double-loop implementation of the algorithm for monotone operators at the expense of an additional logarithmic factor in the convergence rate. Yoon and Ryu [2021] propose the extra anchored gradient (EAG) method, which is the first explicit method with accelerated $O(\frac{1}{T})$ rate in the unconstrained setting for monotone operators. They also established a matching $\Omega(\frac{1}{T})$ lower bound that holds for all first-order methods. Convergence analysis of past extragradient method with anchoring in the unconstrained setting is provided in [Tran-Dinh and Luo, 2021]. Lee and Kim [2021] proposed a generalization of EAG called fast extraradient (FEG), which applies to comonotone operators and improves the constants in the convergence rate, but only for the unconstrained setting. Very recently, Tran-Dinh [2022] studies the connection between Halpern iteration and Nesterov accelerated method, and provides new algorithms for monotone operators and alternative analyses for EAG and FEG in the unconstrained setting. In Theorem 4, we show the projected version of EAG has $O(\frac{1}{T})$ convergence rate under arbitrary convex constraints, achieving the optimal convergence rate for all first-order methods in the constrained setting.

1.2.2 Nonconvex-Nonconcave and Non-Monotone Setting

Many practical applications of min-max optimization in modern machine learning, such as GANs and multi-agent reinforcement learning, are nonconvex-nonconcave. Without any additional structure, the problem is intractable [Daskalakis et al., 2021]. Hence, recent works study nonconvex-nonconcave min-max optimization problems under several structural assumptions. We only introduce the definitions in the unconstrained setting, as that is the setting considered by several of the results, and all convergence rates are in terms of the the operator norm. The *Minty variational inequality* (MVI) condition (also called coherence or variationally stable): there exits z^* such that

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

is studied in e.g., [Dang and 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 and Lan, 2015]. Diakonikolas et al. [2021] proposes a weaker condition called *weak MVI*: there exits z^* and $\rho \leq 0$ 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. Recently, Pethick et al. [2022] generalized EG+ to CEG+ algorithm which has $O(\frac{1}{\sqrt{T}})$ under weak MVI condition in general (constrained) setting. The result for accelerated algorithms in the nonconvex-nonconcave setting is more sparse. FEG achieves $O(\frac{1}{T})$ convergence rate for comonotone operators in the unconstrained setting [Lee and Kim, 2021]. In general (constrained) setting with comonotone operators, the proximal point algorithm is known to exhibit $O(\frac{1}{\sqrt{T}})$ convergence rate [Kohlenbach,

	Algorithm	Setting	Monotone	Non-Monotone		
				Comonotone	MVI	weak MVI
Normal	EG [Dang and Lan, 2015]	general	$O(\frac{1}{\sqrt{T}})$		$O(\frac{1}{\sqrt{T}})$	
	EG+ [Diakonikolas et al., 2021]	unconstrained	$O(\frac{1}{\sqrt{T}})$	$O(\frac{1}{\sqrt{T}})$	$O(\frac{1}{\sqrt{T}})$	$O(\frac{1}{\sqrt{T}})$
	CEG+ [Pethick et al., 2022]	general	$O(\frac{1}{\sqrt{T}})$	$O(\frac{1}{\sqrt{T}})$	$O(\frac{1}{\sqrt{T}})$	$O(\frac{1}{\sqrt{T}})$
Accelerated	Halpern [Diakonikolas, 2020]	general	$O(\frac{\log T}{T})$			
	EAG [Yoon and Ryu, 2021]	unconstrained	$O(\frac{1}{T})$			
	FEG [Lee and Kim, 2021]	unconstrained	$O(\frac{1}{T})$	$O(\frac{1}{T})$		
	EAG [**This paper**]	general	$O(\frac{1}{T})$			
	EAG+ [**This paper**]	general	$O(\frac{1}{T})$	$O(\frac{1}{T})$		

Table 1: Existing results for inclusion problem (min-max optimization problem) with monotone or non-monotone operators. The convergence rate is in terms of the operator norm (in the unconstrained setting) and the residual (in the constrained setting).

2022]. To the best of our knowledge, (EAG+) is the first explicit and efficient method that has an accelerated $O(\frac{1}{T})$ convergence rate in the constrained nonconvex-nonconcave setting (Theorem 7).

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 Monotone 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 $|A(z)| \le 1$ for all $z \in \mathbb{R}^n$. Tha *graph* of an operator A is defined as $\mathrm{Gra}_A = \{(z,u) : z \in \mathbb{R}^n, u \in A(z)\}$. The inverse operator of A is denoted as A^{-1} whose graph is $\mathrm{Gra}_{A^{-1}} = \{(u,z) : (z,u) \in \mathrm{Gra}_A\}$. For two operators A and B, we denote $A+B$ as the operator with graph $\mathrm{Gra}_{A+B} = \{(z, u_A + u_B) : (z,u_A) \in \mathrm{Gra}_A, (z,u_B) \in \mathrm{Gra}_B\}$. We denote the identity operator as $I : \mathbb{R}^n \to \mathbb{R}^n$.

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 $N_{\mathcal{Z}}$:

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

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 $\Pi_{\mathcal{Z}}[z] := \mathrm{argmin}_{z' \in \mathcal{Z}} \|z - z'\|^2$.

For $L \in (0, \infty)$, a single-valued operator $A : \mathbb{R}^n \to \mathbb{R}^n$ is L-Lipschitz if

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

Moreover, A is *non-expansive* if it is 1-Lipschitz. A set-valued operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *monotone* if

$$\langle u - u', z - z' \rangle \geq 0, \quad \forall (z, u), (z', u') \in \text{Gra}_A .$$

Maximally Monotone Operator. A is *maximally monotone* if A is monotone and there is no other monotone operator B such that $\text{Gra}_A \subset \text{Gra}_B$. When $f : \mathbb{R}^n \to \mathbb{R}$ is a convex closed proper function, then the subdifferential operator ∂f is maximally monotone. Therefore, $\partial \mathbb{I}_{\mathcal{Z}} = N_{\mathcal{Z}}$ is maximally monotone. We denote the **resolvent** of an operator A as $J_A := (I + A)^{-1}$. When A is maximally monotone, useful properties of J_A (See e.g., [Ryu and Boyd, 2016, Ryu and Yin, 2022]) include:

1. J_A is well-defined on $\mathbb{R}^n$;
2. J_A is non-expansive thus single-valued;
3. when $z = J_A(z')$, then $z' - z \in A(z)$;
4. when $A = \partial \mathbb{I}_{\mathcal{Z}}$ is the normal cone operator of a closed convex set $\mathcal{Z} \subseteq \mathbb{R}^n$, then $J_{\eta A} = \Pi_{\mathcal{Z}}$ is the projection operator for all $\eta > 0$.

ρ-comonotonicity. A generalized notion of monotonicity is the *ρ-comonotonicity* [Bauschke et al., 2021]: For $\rho \in \mathbb{R}$, an operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is ρ-comonotone if

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

Note that when 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.

2.1 Monotone Inclusion and Variational Inequality

Monotone Inclusion. Given a closed convex set $\mathcal{Z} \subseteq \mathbb{R}^n$ and a single-valued monotone operator F, the *monotone inclusion* problem is to find a point $z^* \in \mathbb{R}^n$ that satisfies

$$\mathbf{0} \in F(z^*) + \partial \mathbb{I}_{\mathcal{Z}}(z^*). \tag{MI}$$

We focus on monotone inclusion problem of a Lipschitz operator with a solution. The assumptions on (MI) is summarized in Assumption 1.

Assumption 1. *In* (MI) *problem,*

1. *F is monotone and L-Lipschitz on $\mathcal{Z}$, i.e.,*

$$\langle F(z) - F(z'), z - z' \rangle \geq 0 \text{ and } \|F(z) - F(z')\| \leq L \cdot \|z - z'\|, \quad \forall z, z' \in \mathcal{Z}.$$

2. *There exists a solution $z^* \in \mathcal{Z}$ such that $\mathbf{0} \in F(z^*) + \partial \mathbb{I}_{\mathcal{Z}}(z^*)$.*

Variational Inequality. A closely related problem to (MI) is the *monotone variational inequality* (VI) with operator F and feasible set $\mathcal{Z}$, which 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. Moreover, the solution set to (MI) is the same as the solution set to (SVI).

Approximate Solutions. We say $z \in \mathcal{Z}$ is an ϵ-approximate solution to (MI) if

$$\mathbf{0} \in F(z) + \partial \mathbb{I}_{\mathcal{Z}}(z) + \mathcal{B}(\mathbf{0}, \epsilon),$$

where we use $\mathcal{B}(u, r)$ to denote a ball in $\mathbb{R}^n$ centered at u with radius r. We say $z \in \mathcal{Z}$ is an ϵ-approximate solution to (SVI) or (MVI) if

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

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 (MI) 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 as

$$\begin{aligned} \textsc{Gap}^{SVI}_{F,D}(z) &:= \max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \langle F(z), z - z' \rangle, \\ \textsc{Gap}^{MVI}_{F,D}(z) &:= \max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \langle F(z'), z - z' \rangle. \end{aligned}$$

The $O(\frac{1}{T})$ convergence rate for extragradient-type algorithm [Nemirovski, 2004, Nesterov, 2007] is provided in terms of $\textsc{Gap}^{MVI}_{F,D}(z)$, which means convergence to an approximate *weak* solution. Prior to our work, the $O(\frac{1}{T})$ convergence rate on $\textsc{Gap}^{SVI}_{F,D}(z)$ was only known in the unconstrained setting [Yoon and Ryu, 2021]. When F is monotone, then the tangent residual $r^{tan}(z) \leq \frac{\epsilon}{D}$ (definition in section 2.3) implies $\textsc{Gap}^{SVI}_{F,D}(z) \leq \epsilon$ [Cai et al., 2022, Lemma 2]. Therefore, our result also implies an $O(\frac{1}{T})$ convergence rate on $\textsc{Gap}^{SVI}_{F,D}(z)$ when $\mathcal{Z}$ is arbitrary convex set (Theorem 4).

2.2 General Inclusion.

We study inclusion problem (GI) with (non)-monotone operators that satisfies *ρ-comonotoncity* (Assumption 2), which captures (MI) (when $\rho = 0$) and a class of nonconvex-nonconcave min-max optimization problems (when $\rho < 0$). Given a single-valued and possibly *non-monotone* operator F and a *set-valued maximally monotone* operator A, we denote $E = F + A$. The inclusion problem is to find a point $z^* \in \mathbb{R}^n$ that satisfies

$$\mathbf{0} \in E(z^*) = F(z^*) + A(z^*). \tag{GI}$$

Similar to (MI) , we say z is an ϵ-approximate solution to (GI) if

$$\mathbf{0} \in F(z) + A(z) + \mathcal{B}(\mathbf{0}, \epsilon).$$

We summarize the assumptions on (GI) below.

Assumption 2. *In* (GI) *problem,*

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 $\rho \le 0$ such that*

$$\langle u - u', z - z' \rangle \ge \rho \|u - u'\|^2, \quad \forall (z, u), (z', u') \in \mathrm{Gra}_E\,.$$

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

The formulation of (GI) provides a unified treatment for a range of problems, such as min-max optimization and multi-player games. We present one detailed example below and refer readers to [Facchinei and 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:*

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

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 $z = (x, y)$, if we define $F(z) = (\partial_x f(x, y), -\partial_y f(x, y))$ and $A(z) = (\partial g(x), \partial h(y))$, where A is maximally monotone, then the first-order optimality condition of (1) *has the form of* (GI) *. See [Lee and Kim, 2021, Example 1] for examples of nonconvex-nonconcave conditions that are implied by negative comonotonicity,.*

2.3 Convergence Criteria

An appropriate convergence criterion is the *tangent residual* $r^{tan}_{F,A}(z) := \min_{c\in A(z)} \|F(z)+c\|$ defined in [Cai et al., 2022]. It is not hard to see that $r^{tan}_{F,A}(z) \le \epsilon$ implies that z is an ϵ-approximate solution to (GI) . If $A = \partial \mathbb{I}_{\mathcal{Z}}$, and $\mathcal{Z}$ is bounded and has diameter no more than D, then z is an ϵ-approximate solution to (MI) and also an $(\epsilon \cdot D)$-approximate solution to (SVI).

Another commonly-used convergence criterion that captures the stationarity of a solution is the *natural residual* $r^{nat}_{F,A} := \|z - J_A\Big[z - F(z)\Big]\|$. Note that z^* is a solution to (GI) iff $z^* = J_A\Big[z^* - F(z^*)\Big]$. The definition of the natural residual for (MI) is similar: $r^{nat}_{F,\partial\mathbb{I}_{\mathcal{Z}}} := \|z - \Pi_{\mathcal{Z}}\Big[z - F(z)\Big]\|$.

Fact 1. *In* (GI) , $r^{nat}_{F,A}(z) \le r^{tan}_{F,A}(z)$.

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

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

Thus $r^{nat}_{F,A}(z) \le \min_{c\in A(z)} \|F(z)+c\| = r^{tan}_{F,A}(z)$. □

In this paper, we state our convergence rates in terms of the tangent residual $r^{tan}_{F,A}(z)$, which implies (i) convergence rates in terms of the natural residual $r^{nat}_{F,A}(z)$, and (ii) z is an approximate solution to (GI) , (MI) , (SVI), or (MVI).

3 Optimal Monotone Inclusion via EAG

In this section, we study constrained monotone inclusion problem (MI) with closed convex feasible set $\mathcal{Z} \subseteq \mathbb{R}^n$ and monotone and L-Lipschitz operator F, as summarized in Assumption 1. We analyse the (projected) Extra Anchored Gradient Method (EAG) , which is proposed by Yoon and Ryu [2021] in the unprojected form for $\mathcal{Z} = \mathbb{R}^n$. Let $z_0 \in \mathcal{Z}$ be an arbitrary starting point and $\{z_k, z_{k+\frac{1}{2}}\}_{k\ge 0}$ be the iterates of (EAG) with step size $\eta > 0$, whose update rule is as follows:

$$\begin{aligned} z_{k+\frac{1}{2}} &= \Pi_{\mathcal{Z}}\Big[z_k - \eta F(z_k) + \frac{1}{k+1}(z_0 - z_k)\Big], \\ z_{k+1} &= \Pi_{\mathcal{Z}}\Big[z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+1}(z_0 - z_k)\Big]. \end{aligned} \tag{EAG}$$

Our analysis is based on the following potential function

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

where

$$c_k := \frac{z_{k-1} - \eta F(z_{k-\frac{1}{2}}) + \frac{1}{k}(z_0 - z_{k-1}) - z_k}{\eta}, \quad k \geq 1.$$

From the update rule of (EAG) , we know $c_k \in N_{\mathcal{Z}}(z_k)$. Thus $\|F(z_k) + c_k\| \geq \min_{c \in N_{\mathcal{Z}}(z_k)} \|F(z_k) + c\| = r^{tan}(z_k)$. In Theorem 3, we show V_k is "approximately" non-increasing, i.e., $V_{k+1} \leq V_k + O(1) \cdot \|\eta F(z_{k+1}) + \eta c_{k+1}\|^2$ for all $k \geq 1$. From this property, we get $O(\frac{1}{T})$ last-iterate convergence rate in terms of $\|F(z_T) + c_T\|$ and thus the same convergence rate in $r^{tan}(z_T)$ (Theorem 4).

Potential functions in a similar form as V_k have been used to analyse (MI) by Diakonikolas [2020] for the Halpern iteration algorithm, and by Yoon and Ryu [2021] for (EAG) in the unconstrained setting ($\mathcal{Z} = \mathbb{R}^n$). We emphasize that we use a different potential function with different analysis.

Diakonikolas [2020] studied the Halpern iteration with operator $P := I - J_{F+\partial \mathbb{I}_{\mathcal{Z}}}$, which is $\frac{1}{2}$-cocoercive but can not be computed efficiently in general. She showed that the follwoing potential function is non-increasing.

$$P_k := \frac{k(k+1)}{2} \cdot \|P(z_k)\|^2 + k \cdot \langle P(z_k), z_k - z_0 \rangle,$$

which leads to $O(\frac{1}{T})$-approximate solution to (MI) after T iterations. However, since P can not be computed efficiently in general, the algorithm needs $O(\log(\frac{1}{\epsilon}))$ oracle queries for an $O(\epsilon)$-approximate value of P in each iteration, thus total oracle complexity $O(\frac{LD}{\epsilon} \cdot \log(\frac{1}{\epsilon}))$ for an ϵ-approximate solution to (MI) . In contrast, we use operator F in the potential function V_k, and we prove V_k is only "approximately" non-increasing (see Theorem 1 and 3). Moreover, (EAG) needs only 2 oracle query in each iteration and achieves optimal $O(\frac{LD}{\epsilon})$ oracle complexity for an ϵ-approximate solution to (MI) (Theorem 4).

Yoon and Ryu [2021] studied convergence of (EAG) for (MI) in the unconstrained setting ($\mathcal{Z} = \mathbb{R}^n$). The specific algorithm they analysed uses anchoring term $\frac{1}{k+2}(z_0 - z_k)$ while we use $\frac{1}{k+1}(z_0 - z_k)$ (see Remark 1 for more discussion on the choice of the constant in the anchoring term). They use the following potential function

$$P_k := A_k \cdot \|F(z_k)\|^2 + B_k \cdot \langle F(z_k), z_k - z_0 \rangle,$$

where $B_k = k + 1$, and $A_k = O(k^2)$ is updated adaptively in a sophisticated way for each k. Their potential function P_k is more complicated compared to V_k as we choose $B_k = k$ and $A_k = \frac{k(k+1)}{2}$. For the analysis, their proof of the monotonicity of P_k is relatively involved. In contrast, we use a simple proof to show that V_k is "approximately" non-increasing (Theorem 1) which suffices to establish the $O(\frac{1}{T})$ convergence rate. Moreover, our analysis can be naturally extended to the constrained setting where $\mathcal{Z} \subseteq \mathbb{R}^n$ is an arbitrary closed convex set (Theorem 3).

3.1 Warm Up: Unconstrained Case

We begin with the unconstrained setting $\mathcal{Z} = \mathbb{R}^n$, which illustrate our main idea and proof techniques. [Yoon and Ryu, 2021] also analyse the unconstrained setting but our proof is much simpler.

In the unconstrained setting, $c_k = 0$ by definition. Thus

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

It is not hard to see that $V_1 \leq (\eta^2 L^2 + 2\eta L)\|z_0 - z^*\|^2$: since the update rule for $z_{\frac{1}{2}}$ and z_1 of (EAG) coincides with the update rule of EG, by [Cai et al., 2022, Theorem 1], we have $\|\eta F(z_1)\|^2 \leq \|\eta F(z_0)\|^2 \leq \eta^2 L^2 \|z_0 - z^*\|^2$ and and by [Korpelevich, 1976], [Facchinei and Pang, 2003, Lemma 12.1.10] $\|z_1 - z^*\| \leq \|z_0 - z^*\|$

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

Theorem 1. *Suppose Assumption 1 holds with $\mathcal{Z} = \mathbb{R}^n$. Then for any $k \geq 1$, (EAG) with any step size $\eta \in (0, \frac{1}{L})$ satisfies $V_{k+1} \leq V_k + \frac{\eta^2 L^2}{1-\eta^2 L^2}\|\eta F(z_{k+1})\|^2$.*

Proof. Since F is monotone and L-Lipschitz, we have the following inequalities

$$\langle F(z_{k+1}) - F(z_k), z_k - z_{k+1}\rangle \leq 0$$

and

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

We simplify them using the update rule of (EAG) .

In particular, we replace $z_k - z_{k+1}$ with $\eta F(z_{k+\frac{1}{2}}) - \frac{1}{k+1}(z_0 - z_k)$ and $z_{k+\frac{1}{2}} - z_{k+1}$ with $\eta F(z_{k+\frac{1}{2}}) - \eta F(z_k)$.

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

$$\left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2 - \eta^2 L^2\left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_k)\right\|^2 \leq 0. \tag{3}$$

It is not hard to verify that the following identity holds.

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

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

Note that $\frac{k+1}{2k} \leq 1$ holds for all $k \geq 1$. Thus, $V_{k+1} \leq V_k + \frac{\eta^2 L^2}{1-\eta^2 L^2}\|\eta F(z_{k+1})\|^2$. □

Lemma 1. *For all $k \geq 2$,*

$$\left(\frac{k(k+1)}{4} - \frac{\eta^2 L^2}{1-\eta^2 L^2}\right) \|\eta F(z_k)\|^2 \leq (1+\eta L)^2 \|z_0 - z^*\|^2 + \frac{\eta^2 L^2}{1-\eta^2 L^2} \sum_{t=2}^{k-1} \|\eta F(z_t)\|^2.$$

Moreover, when $\eta \in (0, \frac{1}{\sqrt{3}L})$, we have

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

Proof. Fix any $k \geq 2$. By definition, we have

$$\begin{aligned}
V_k &= \frac{k(k+1)}{2} \|\eta F(z_k)\|^2 + k\langle \eta F(z_k), z_k - z_0 \rangle \\
&\geq \frac{k(k+1)}{2} \|\eta F(z_k)\|^2 + k\langle \eta F(z_k), z^* - z_0 \rangle && (\ \langle F(z_k), z^* - z_k \rangle \leq 0) \\
&\geq \frac{k(k+1)}{2} \|\eta F(z_k)\|^2 - \frac{k(k+1)}{4} \|\eta F(z_k)\|^2 - \frac{k^2}{k(k+1)} \|z_0 - z^*\|^2 && (\ \langle a, b \rangle \geq -\tfrac{c}{4}\|a\|^2 - \tfrac{1}{c}\|b\|^2) \\
&\geq \frac{k(k+1)}{4} \|\eta F(z_k)\|^2 - \|z_0 - z^*\|^2. && (\tfrac{k}{k+1} \leq 1)
\end{aligned}$$

Using Theorem 1, we have

$$V_k \leq V_1 + \frac{\eta^2 L^2}{1-\eta^2 L^2} \sum_{t=2}^{k} \|\eta F(z_t)\|^2.$$

Combining the two inequalities above and the fact that $V_1 \leq (2\eta L + \eta^2 L^2)\|z_0 - z^*\|^2$ yields the first inequality in the statement. When $\eta L \in (0, \frac{1}{\sqrt{3}})$, we have $\frac{\eta^2 L^2}{1-\eta^2 L^2} \leq \frac{1}{2} \leq \frac{k}{4}$ for $k \geq 2$. Hence the second inequality in the statement holds. □

Theorem 2. *Suppose Assumption 1 holds with $\mathcal{Z} = \mathbb{R}^n$. Let $z_0 \in \mathbb{R}^n$ be arbitrary starting point and $\{z_k, z_{k+\frac{1}{2}}\}_{k \geq 0}$ be the iterates of (EAG) with any step size $\eta \in (0, \frac{1}{\sqrt{3}L})$. Denote $D := \|z_0 - z^*\|$. Then for any $T \geq 1$,*

$$\begin{aligned}
\|F(z_T)\|^2 &\leq \frac{4(1+\eta L)^2}{\eta^2 L^2 (1 - 3\eta^2 L^2)} \cdot \frac{D^2 L^2}{T^2}, \\
\textsc{Gap}^{SVI}_{F,D}(z_T) &\leq \frac{2(1+\eta L)}{\eta L \sqrt{1 - 3\eta^2 L^2}} \cdot \frac{D^2 L}{T}.
\end{aligned}$$

If we set $\eta = \frac{1}{3L}$, then $\|F(z_T)\|^2 \leq \frac{96 \cdot D^2 L^2}{T^2}$.

Proof. Note that the second inequality is implied by the first inequality since $\text{GAP}^{SVI}_{F,D}(z) \le D \cdot \|F(z_T)\|$ [Cai et al., 2022, Lemma 2]. Denote $a_k := \frac{\|\eta F(z_k)^2\|}{\|z_0 - z^*\|^2}$. It suffices to prove for all $k \ge 1$,

$$a_k \le \frac{4(1+\eta L)^2}{(1-3\eta^2 L^2)k^2}. \tag{4}$$

Since the update rule for $z_{\frac{1}{2}}$ and z_1 of (EAG) coincides with the update rule of EG, by [Cai et al., 2022, Theorem 1], we have $\|\eta F(z_1)\|^2 \le \|\eta F(z_0)\|^2 \le \eta^2 L^2 \|z_0 - z^*\|^2$ and thus $a_1 \le \eta^2 L^2 < \frac{1}{3}$. Thus Equation (4) holds for $k = 1$.

From Lemma 1, we know for $k \ge 2$,

$$\frac{k^2}{4} \cdot a_k \le (1+\eta L)^2 + \frac{\eta^2 L^2}{1-\eta^2 L^2} \sum_{t=2}^{k-1} a_t.$$

Applying Proposition 4 with $C_1 = (1+\eta L)^2$ and $p = \eta^2 L^2 < \frac{1}{3}$ completes the proof. □

3.2 Convergence of EAG with Arbitrary Convex Constraints

Proposition 1. $V_1 \le \frac{(1+\eta L+\eta^2 L^2)(2+2\eta L+\eta^2 L^2)}{1-\eta^2 L^2} \|z_0 - z^*\|^2$.

Proof. We first upper bound $\|\eta F(z_1) + \eta c_1\|$ and $\|z_1 - z_0\|$. Note that $z_{\frac{1}{2}}, z_1$ are updated exactly as original EG. By definition, we have

$$\begin{aligned}
\|\eta F(z_1) + \eta c_1\| &= \left\|\eta F(z_1) + z_0 - \eta F(z_{\frac{1}{2}}) - z_1\right\| \\
&\le \left\|\eta F(z_1) - \eta F(z_{\frac{1}{2}})\right\| + \|z_0 - z_1\| \\
&\le \eta L \left\|z_1 - z_{\frac{1}{2}}\right\| + \|z_0 - z_1\| && (L\text{-Lipschitzness of } F) \\
&\le (1+\eta L)\left\|z_1 - z_{\frac{1}{2}}\right\| + \left\|z_{\frac{1}{2}} - z_0\right\| \\
&\le (1+\eta L+\eta^2 L^2)\left\|z_{\frac{1}{2}} - z_0\right\| \\
&\le \frac{1+\eta L+\eta^2 L^2}{\sqrt{1-\eta^2 L^2}} \|z_0 - z^*\|,
\end{aligned}$$

where in the last inequality we use a well-known result regarding EG: $\|z_{\frac{1}{2}} - z_0\|^2 \le \frac{\|z_0 - z^*\|^2 - \|z_1 - z^*\|^2}{1-\eta^2 L^2}$ [Facchinei and Pang, 2003, Lemma 12.1.10]. Note that in the above sequence of inequalities, we also prove that $\|z_1 - z_0\| \le \frac{1+\eta L}{\sqrt{1-\eta^2 L^2}} \|z_0 - z^*\|$.

By definition of V_1 and the above upper bound for $\|\eta F(z_1) + \eta c_1\|$ and $\|z_1 - z_0\|$, we have

$$\begin{aligned}
V_1 &= \|\eta F(z_1) + \eta c_1\|^2 + \langle \eta F(z_1) + \eta c_1, z_1 - z_0 \rangle \\
&\le \|\eta F(z_1) + \eta c_1\|^2 + \|\eta F(z_1) + \eta c_1\| \cdot \|z_1 - z_0\| \\
&\le \frac{(1+\eta L+\eta^2 L^2)(2+2\eta L+\eta^2 L^2)}{1-\eta^2 L^2} \|z_0 - z^*\|^2.
\end{aligned}$$

□

Theorem 3. *Suppose Assumption 1 holds. Let $z_0 \in \mathcal{Z}$ be any starting point and $\{z_k, z_{k+\frac{1}{2}}\}_{k\geq 0}$ be the iterates of (EAG) with step size $\eta \in (0, \frac{1}{L})$. Then for any $k \geq 1$, $V_{k+1} \leq V_k + \frac{\eta^2 L^2}{1-\eta^2 L^2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2$, where $c_{k+1} = \frac{z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+1}(z_0 - z_k) - z_{k+1}}{\eta}$.*

Proof. We first present several inequalities. From the monotonicity and L-Lipschitzness of F, we have

$$(-\frac{k(k+1)}{2\eta^2 L^2}) \cdot \left(\eta^2 L^2 \cdot \left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2 - \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2\right) \leq 0, \tag{5}$$

$$(-k(k+1)) \cdot \langle \eta F(z_{k+1}) - \eta F(z_k), z_{k+1} - z_k) \rangle \leq 0. \tag{6}$$

Since $z_{k+\frac{1}{2}} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(z_k) + \frac{1}{k+1}(z_0 - z_k)\right]$, 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

$$(-k(k+1)) \cdot \left\langle z_k - \eta F(z_k) - z_{k+\frac{1}{2}} + \frac{1}{k+1}(z_0 - z_k), z_{k+\frac{1}{2}} - z_{k+1} \right\rangle \leq 0, \tag{7}$$

$$(-k(k+1)) \cdot \langle \eta c_{k+1}, z_{k+1} - z_k \rangle \leq 0, \tag{8}$$

$$(-k(k+1)) \cdot \left\langle \eta c_k, z_k - z_{k+\frac{1}{2}} \right\rangle \leq 0. \tag{9}$$

The following identity holds when we substitute ηc_{k+1} on both sides using $\eta c_{k+1} = z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+1}(z_0 - z_k) - z_{k+1}$, which follows from the definition. The correctness of the identity follows from Identity (28) in Proposition 3: 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 $\eta^2 L^2$ and q as k.

$$\begin{aligned}
&V_k - V_{k+1} + \text{LHS of Inequality (5)} + \text{LHS of Inequality (6)} + \text{LHS of Inequality (7)} \\
&\quad + \text{LHS of Inequality (8)} + \text{LHS of Inequality (9)} \\
=&\frac{k(k+1)}{2} \cdot \left\| z_{k+\frac{1}{2}} - z_k + \eta F(z_k) + \eta c_k + \frac{1}{k+1}(z_k - z_0)\right\|^2 && (10) \\
&+ \frac{(1-\eta^2 L^2)k(k+1)}{2\eta^2 L^2} \cdot \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2 && (11) \\
&+ (k+1) \cdot \left\langle \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}), \eta F(z_{k+1}) + \eta c_{k+1} \right\rangle. && (12)
\end{aligned}$$

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

$$
\begin{aligned}
&\text{Expression}(11) + \text{Expression}(12) \\
=&\left\| \sqrt{\frac{(1-\eta^2L^2)k(k+1)}{2\eta^2L^2}} \cdot \left(\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right) + \sqrt{\frac{\eta^2L^2(k+1)}{2(1-\eta^2L^2)k}} \cdot (\eta F(z_{k+1}) + \eta c_{k+1}) \right\|^2 \\
&- \frac{k+1}{2k} \cdot \frac{\eta^2L^2}{1-\eta^2L^2} \|\eta F(z_{k+1}) + \eta c_{k+1}\|^2.
\end{aligned}
$$

Since $k \geq 1$, we have $\frac{k+1}{2k} \leq 1$. Hence, we have $V_{k+1} \leq V_k + \frac{\eta^2L^2}{1-\eta^2L^2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2$. □

Remark 1. *The proof of Theorem 3 naturally extends to the following algorithm and potential function: Fix any $z_0 \in \mathcal{Z}$ and $\eta \in (0, \frac{1}{L}), \delta \geq 0$. Update $z_{\frac{1}{2}}, z_1, c_1, V_1$ as (EAG) and for $k \geq 1$:*

$$
\begin{aligned}
z_{k+\frac{1}{2}} &= \Pi_{\mathcal{Z}}\left[z_k - \eta F(z_k) + \frac{1}{k+\delta+1}(z_0 - z_k)\right], \\
z_{k+1} &= \Pi_{\mathcal{Z}}\left[z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+\delta+1}(z_0 - z_k)\right], \\
c_{k+1} &= \frac{z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+\delta+1}(z_0 - z_k) - z_{k+1}}{\eta}, \\
V_{k+1} &= \frac{(k+\delta+1)(k+\delta+2)}{2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2 + (k+\delta+1) \cdot \langle \eta F(z_{k+1}) + \eta c_{k+1}, z_{k+1} - z_0\rangle.
\end{aligned}
$$

Since the identity in Proposition 3 holds for any $q \neq 0$, we only need to change every k to be $k + \delta$ in the proof of Theorem 3. It is possible that a choice of $\delta > 0$ leads to a better upper bound (better constant) than $\delta = 0$ which is chosen for (EAG) , but we do not optimize over δ here.

Lemma 2. *For $k \geq 2$,*

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

Moreover, when $\eta \in (0, \frac{1}{\sqrt{3}L})$, then

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

Proof. Fix any $k \geq 2$. By definition, we have

$$
\begin{aligned}
V_k &= \frac{k(k+1)}{2}\|\eta F(z_k) + \eta c_k\|^2 + k\langle \eta F(z_k) + \eta c_k, z_k - z_0\rangle \\
&\geq \frac{k(k+1)}{2}\|\eta F(z_k) + \eta c_k\|^2 + k\langle \eta F(z_k) + \eta c_k, z^* - z_0\rangle \\
&\qquad\qquad (F + \partial \mathbb{I}_{\mathcal{Z}} \text{ is monotone and } \mathbf{0} \in F(z^*) + \partial \mathbb{I}_{\mathcal{Z}}(z^*)) \\
&\geq \frac{k(k+1)}{2}\|\eta F(z_k) + \eta c_k\|^2 - \frac{k(k+1)}{4}\|\eta F(z_k) + \eta c_k\|^2 - \frac{k}{k+1}\|z_0 - z^*\|^2 \\
&\qquad\qquad (\langle a, b\rangle \geq -\tfrac{c}{4}\|a\|^2 - \tfrac{1}{c}\|b\|^2) \\
&\geq \frac{k(k+1)}{4}\|\eta F(z_k) + \eta c_k\|^2 - \|z_0 - z^*\|^2. \qquad (13)
\end{aligned}
$$

According to Theorem 3, $V_{t+1} - V_t \leq \frac{\eta^2 L^2}{1-\eta^2 L^2}\|\eta F(z_{t+1}) + \eta c_{t+1}\|^2$ for all $t \geq 1$. Through a telescoping sum, we obtain the following inequality:

$$
V_k \leq V_1 + \frac{\eta^2 L^2}{1 - \eta^2 L^2} \cdot \sum_{t=2}^{k} \|\eta F(z_t) + \eta c_t\|^2. \qquad (14)
$$

The first inequality in the statement follows from the combination of Inequality (13) and (14). The second inequality in the statement follows from the fact that $\frac{\eta^2 L^2}{1-\eta^2 L^2} \leq \frac{1}{2} \leq \frac{k}{4}$ when $\eta \in (0, \frac{1}{\sqrt{3}L})$. □

Theorem 4. *Suppose Assumption 1 holds. Let $z_0 \in \mathcal{Z}$ be any starting point and $\{z_k, z_{k+\frac{1}{2}}\}_{k\geq 0}$ be the iterates of (EAG) with step size $\eta \in (0, \frac{1}{\sqrt{3}L})$. Denote $D := \|z_0 - z^*\|^2$. Then for any $T \geq 1$,*

$$
r^{nat}_{F,\partial \mathbb{I}_{\mathcal{Z}}}(z_T)^2 = \left\| z_T - \Pi_{\mathcal{Z}}\left[z_T - F(z_T)\right] \right\|^2 \leq \|F(z_T) + c_T\|^2 \leq \frac{44}{\eta^2 L^2 (1 - 3\eta^2 L^2)} \cdot \frac{D^2 L^2}{T^2}
$$

$$
\textsc{Gap}^{SVI}_{F,D}(z_T) \leq \frac{\sqrt{44}}{\eta L \sqrt{1 - 3\eta^2 L^2}} \cdot \frac{D^2 L}{T}.
$$

Proof. Note that by non-expansiveness of $\Pi_{\mathcal{Z}}$, we have

$$
\left\| z_T - \Pi_{\mathcal{Z}}\left[z_T - F(z_T)\right] \right\|^2 = \left\| \Pi_{\mathcal{Z}}[z_T + c_T] - \Pi_{\mathcal{Z}}\left[z_T - F(z_T)\right] \right\|^2 \leq \|F(z_T) + c_T\|^2.
$$

The bound on $\textsc{Gap}^{SVI}_{F,D}(z_T)$ follows from the bound on $r^{tan}_{F,A}(z_T)$ since $\textsc{Gap}^{SVI}_{F,D}(z_T) \leq D \cdot r^{tan}_{F,A}(z_T)$ [Cai et al., 2022, Lemma 2].

Denote $a_k := \frac{\|\eta F(z_k) + \eta c_k\|^2}{\|z_0 - z^*\|^2}$. It suffices to prove that for all $k \geq 1$,

$$
a_k \leq \frac{44}{(1 - 3\eta^2 L^2) k^2}. \qquad (15)
$$

Note that from the proof of Proposition 1, we have

$$\|\eta F(z_1) + \eta c_1\|^2 \le \frac{(1+\eta L+\eta^2L^2)^2}{1-\eta^2L^2}\|z_0 - z^*\|^2 \quad \Rightarrow \quad a_1 \le \frac{(1+\eta L+\eta^2L^2)^2}{1-\eta^2L^2} \le 6.$$

Thus Equation (15) holds for $k=1$.

From Proposition 1, we also have

$$V_1 \le \frac{(1+\eta L+\eta^2L^2)(2+2\eta L+\eta^2L^2)}{1-\eta^2L^2}\|z_0 - z^*\|^2 \le 10 \cdot \|z_0 - z^*\|^2.$$

Thus by Lemma 2, we have

$$\begin{aligned}
&\frac{k^2}{4}\cdot \|\eta F(z_k) + \eta c_k\|^2 \le 11 \cdot \|z_0 - z^*\|^2 + \frac{\eta^2L^2}{1-\eta^2L^2}\sum_{t=2}^{k-1}\|\eta F(z_t) + \eta c_t\|^2 \\
\Rightarrow &\frac{k^2}{4}\cdot a_k \le 11 + \frac{\eta^2L^2}{1-\eta^2L^2}\sum_{t=2}^{k-1} a_t.
\end{aligned}$$

Applying Proposition 4 with $C_1 = 11$ and $p = \eta^2L^2 \in (0, \frac{1}{3})$ completes the proof. □

4 Optimal Algorithms for General Inclusions with Negatively Comonotone Operators

Algorithm. The following (EAG+) is a generalization of (EAG) to the general inclusion problem (GI) . Given any initial point $z_0 \in \mathbb{R}^n$ and step size $\eta > 0$, (EAG+) updates $\{z_{k+\frac{1}{2}}, z_{k+1}, c_{k+1}\}_{k\ge 0}$ as follows:

$$z_{\frac{1}{2}} = J_{\eta A}\Big[z_0 - \eta F(z_0)\Big], \quad z_1 = J_{\eta A}\Big[z_0 - \eta F(z_{\frac{1}{2}})\Big], \quad c_1 = \frac{z_0 - \eta F(z_{\frac{1}{2}}) - z_1}{\eta}, \tag{16}$$

and for $k \ge 1$:

$$\begin{aligned}
z_{k+\frac{1}{2}} &= z_k - \eta F(z_k) + \frac{1}{k+2}(z_0 - z_k) - \eta c_k \\
z_{k+1} &= J_{\eta A}\Big[z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+2}(z_0 - z_k)\Big] \\
c_{k+1} &= \frac{z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+2}(z_0 - z_k) - z_{k+1}}{\eta}
\end{aligned} \tag{EAG+}$$

Note that by definition we have $c_k \in A(z_k)$ for all $k \ge 1$.

Remark 2. *In the special case where $A = \partial \mathbb{I}_{\mathcal{Z}}$ is the normal cone operator, $J_{\eta A}$ is the projection operator $\Pi_{\mathcal{Z}}$. In this case,* (EAG+) *is still different from* (EAG) *. The major difference is that in each iteration,* (EAG) *performs two projections while* (EAG+) *only performs one projection. Consequently, the iterates*

$\{z_{k+\frac{1}{2}}, z_{k+1}\}_{k\geq 0}$ *produced by* (EAG) *are all in the feasible set* $\mathcal{Z}$*, while* $\{z_{k+\frac{1}{2}}\}_{k\geq 1}$ *produced by* (EAG+) *may not belong to* $\mathcal{Z}$*. As a result, the convergence guarantee for* (EAG+) *requires that monotonicity and Lipschitzness of* F *hold on* $\mathbb{R}^n$ *while the guarantee for* (EAG) *only requires those properties to hold on the feasible set* $\mathcal{Z}$*. Nevertheless, we believe in many natural settings, i.e., min-max optimization, the operator* F *is indeed monotone in the entire Euclidean space.*

We use the following potential function:

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

The following proposition provides a bound for V_1.

Proposition 2. *$V_1 \leq \eta^2(1 + \eta L + \eta^2 L^2)(5 + 5\eta L + 3\eta^2 L^2) \cdot r^{tan}(z_0)^2$. If $\eta \in (0, \frac{1}{2L})$, then $V_1 \leq \frac{15}{4L^2} \cdot r^{tan}_{F,A}(z_0)^2$.*

Proof. Let us bound $\|z_{\frac{1}{2}} - z_0\|$ first. For any $c \in A(z_0)$, we have

$$\begin{aligned}\left\|z_{\frac{1}{2}} - z_0\right\| &= \|J_{\eta A}(z_0 - \eta F(z_0)) - J_{\eta A}(z_0 + \eta c)\| \\ &\leq \|\eta F(z_0) + \eta c\|. && \text{(non-expansiveness of } J_{\eta A})\end{aligned}$$

Thus we have $\|z_{\frac{1}{2}} - z_0\| \leq \eta \cdot \min_{c\in A(z_0)} \|F(z_0) + c\| = \eta \cdot r^{tan}_{F,A}(z_0)$.

Now we bound $\|\eta F(z_1) + \eta c_1\|$ and $\|z_1 - z_0\|$ first. By definition of c_1, we have

$$\begin{aligned}\|\eta F(z_1) + \eta c_1\| &= \left\|\eta F(z_1) - \eta F(z_{\frac{1}{2}}) + z_0 - z_1\right\| \\ &\leq (\eta L) \cdot \left\|z_1 - z_{\frac{1}{2}}\right\| + \|z_0 - z_1\| && (F \text{ is } L\text{-Lipschitz}) \\ &\leq (1 + \eta L) \cdot \left\|z_1 - z_{\frac{1}{2}}\right\| + \left\|z_{\frac{1}{2}} - z_0\right\| \\ &\leq (1 + \eta L) \cdot \left\|\eta F(z_{\frac{1}{2}}) - \eta F(z_0)\right\| + \left\|z_{\frac{1}{2}} - z_0\right\| && (J_{\eta A} \text{ is non-expansive}) \\ &\leq (1 + \eta L + \eta^2 L^2) \cdot \left\|z_{\frac{1}{2}} - z_0\right\|. && (F \text{ is } L\text{-Lipschitz}) \\ &\leq \eta(1 + \eta L + \eta^2 L^2) \cdot r^{tan}_{F,A}(z_0).\end{aligned}$$

In the above chain of inequalies, we also prove $\|z_1 - z_0\| \leq \eta(1 + \eta L) \cdot r^{tan}(z_0)$.

Finally, by definition of V_1, we have

$$\begin{aligned}V_1 &= 3 \cdot \|\eta F(z_1) + \eta c_1\|^2 + 2 \cdot \langle \eta F(z_1) + \eta c_1, z_1 - z_0 \rangle \\ &\leq 3 \cdot \|\eta F(z_1) + \eta c_1\|^2 + 2 \cdot \|\eta F(z_1) + \eta c_1\| \cdot \|z_1 - z_0\| && \text{(Cauchy-Schwarz inequality)} \\ &\leq \eta^2(1 + \eta L + \eta^2 L^2)(5 + 5\eta L + 3\eta^2 L^2) \cdot r^{tan}_{F,A}(z_0)^2.\end{aligned}$$

If $\eta \in (0, \frac{1}{2L})$, then $V_1 \leq 15\eta^2 \cdot r^{tan}_{F,A}(z_0)^2 \leq \frac{15}{4L^2} \cdot r^{tan}(z_0)^2$. □

4.1 Monotone Case

Theorem 5. *Suppose Assumption 2 holds with $\rho = 0$. Let $z_0 \in \mathbb{R}^n$ be any starting point and $\{z_{k+\frac{1}{2}}, z_{k+1}\}_{k\geq 0}$ be the iterates of (EAG+) with step size $\eta \in (0, \frac{1}{L})$. Then for any $k \geq 1$, $V_{k+1} \leq V_k + \frac{\eta^2 L^2}{1-\eta^2 L^2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2$.*

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

$$\left(-\frac{(k+1)(k+2)}{2\eta^2 L^2}\right) \cdot \left(\eta^2 L^2 \cdot \left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2 - \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2\right) \leq 0. \tag{17}$$

Additionally, as F is monotone, A is maximally monotone,$c_k \in A(z_k)$, and $c_{k+1} \in A(z_{k+1})$, we have

$$(-(k+1)(k+2)) \cdot \langle \eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k, z_{k+1} - z_k) \rangle \leq 0. \tag{18}$$

The following identity holds due to Identity (29) in Proposition 3: 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 $\eta^2 L^2$, and q as $k+1$. Note that by the update rule of (EAG+) , we have $\eta c_k = z_k - \eta F(z_k) + \frac{1}{k+2}(z_0 - z_k) - z_{k+\frac{1}{2}}$, and by definition, we have $\eta c_{k+1} = z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+2}(z_0 - z_k) - z_{k+1}$.

$$\begin{aligned} & V_k - V_{k+1} + \text{LHS of Inequality (17)} + \text{LHS of Inequality (18)} \\ =& \frac{(1-\eta^2 L^2)(k+1)(k+2)}{2\eta^2 L^2} \cdot \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2 \qquad (19) \\ & + (k+2) \cdot \left\langle \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}), \eta F(z_{k+1}) + \eta c_{k+1} \right\rangle. \qquad (20) \end{aligned}$$

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

$$\begin{aligned} & \text{Expression(19)} + \text{Expression(20)} \\ =& \left\|\sqrt{\frac{(1-\eta^2 L^2)(k+1)(k+2)}{2\eta^2 L^2}} \cdot \left(\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right) + \sqrt{\frac{\eta^2 L^2 (k+2)}{2(1-\eta^2 L^2)(k+1)}} \cdot (\eta F(z_{k+1}) + \eta c_{k+1})\right\|^2 \\ & - \frac{k+2}{2(k+1)} \cdot \frac{\eta^2 L^2}{1-\eta^2 L^2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2. \end{aligned}$$

Since $k \geq 1$, we have $\frac{k+2}{2(k+1)} \leq 1$. Hence, $V_{k+1} \leq V_k + \frac{\eta^2 L^2}{1-\eta^2 L^2}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2$. □

4.2 Non-Monotone Case

Fact 2. *For any $\rho \geq -\frac{1}{24L}$, there exists $\eta \in (0, \frac{1}{2L})$ such that*

$$1 - \left(4 - \frac{4\rho}{\eta}\right)\eta^2 L^2 + \frac{4\rho}{\eta} \geq 0. \tag{21}$$

Moreover, any $\eta \in (0, \frac{1}{2L})$ that satisfies Inequality (21) also satisfies $\frac{\rho}{\eta} \geq -\frac{1}{4}$.

Proof. Rearranging Inequality (21), we get

$$\rho \geq \frac{\eta L(4\eta^2L^2-1)}{4(1+\eta^2L^2)} \cdot \frac{1}{L}$$

Denote $x = \eta L \in (0, \frac{1}{2})$ and consider function $f(x) = \frac{x(4x^2-1)}{4(1+x^2)}$. Note that

$$\min_{x\in(0,\frac{1}{2})} f(x) \leq f(\frac{1}{3}) = -\frac{1}{24}.$$

Thus there exists $\eta = \frac{1}{3L} \in (0, \frac{1}{2L})$ such that Inequality (21) holds.

Moreover, by rearranging Inequality (21), we get

$$\frac{\rho}{\eta} \geq -\frac{1-4\eta^2L^2}{4(1+\eta^2L^2)} \geq -\frac{1}{4}.$$

□

Theorem 6. *Suppose Assumption 2 holds with* $-\frac{1}{24L} \leq \rho \leq 0$. *Fix* $p = \frac{1}{4}$. *Let* $z_0 \in \mathbb{R}^n$ *be any starting point and* $\{z_{k+\frac{1}{2}}, z_{k+1}\}_{k\geq 0}$ *be the iterates of* (EAG+) *with step size* $\eta \in (0, \frac{1}{2L})$ *that satisfies Inequality* (21). *Then for any* $k \geq 1$, $V_{k+1} \leq V_k + \frac{p}{1-p}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2 = V_k + \frac{1}{3} \cdot \|\eta F(z_{k+1}) + \eta c_{k+1}\|^2$.

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

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

Denote $c = -\frac{4p\rho}{\eta} \geq 0$. Multiply both sides of the above inequality by $(1+c)$ and rearrange terms we get

$$\begin{aligned} &p \cdot \left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2 - \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2 \\ &+ ((1+c)\eta^2L^2 - p) \cdot \left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2 - c \cdot \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2 \geq 0. \end{aligned} \tag{22}$$

By definition, we have $c_k \in A(z_k)$ and $c_{k+1} \in A(z_{k+1})$. Since $E = F + A$ satisfies ρ-comonotonicity, we have

$$\langle \eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k, z_{k+1} - z_k\rangle - \frac{\rho}{\eta}\|\eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k\|^2 \geq 0. \tag{23}$$

The following identity holds due to Identity (29) in Proposition 3: 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 out choice of p in the statement, and q as $k+1$. Note that by update rule, we have $\eta c_k = z_k - \eta F(z_k) + \frac{1}{k+2}(z_0 - z_k) - z_{k+\frac{1}{2}}$, and by definition, we have $\eta c_{k+1} = z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+2}(z_0 - z_k) - z_{k+1}$. Note that

Term (24) and Term (25) come from the identity in Proposition 3, Term (26) directly comes from (22), and Term (27) directly comes from (23).

$$
\begin{aligned}
& V_k - V_{k+1} + \left(-\frac{(k+1)(k+2)}{2p}\right) \cdot \text{LHS of Inequality (22)} + (-(k+1)(k+2)) \cdot \text{LHS of Inequality (23)} \\
=& \frac{(1-p)(k+1)(k+2)}{2p} \cdot \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2 && (24) \\
& + (k+2) \cdot \left\langle \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}), \eta F(z_{k+1}) + \eta c_{k+1} \right\rangle && (25) \\
& + \frac{(k+1)(k+2)}{2p} \cdot \left((p - (1+c)\eta^2 L^2) \cdot \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 + c \cdot \left\| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}) \right\|^2 \right) && (26) \\
& + \frac{(k+1)(k+2)\rho}{\eta} \|\eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k\|^2. && (27)
\end{aligned}
$$

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

$$
\begin{aligned}
& \text{Expression(24)} + \text{Expression(25)} \\
=& \left\| \sqrt{\frac{(1-p)(k+1)(k+2)}{2p}} \cdot \left(\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right) + \sqrt{\frac{p(k+2)}{2(1-p)(k+1)}} \cdot (\eta F(z_{k+1}) + \eta c_{k+1}) \right\|^2 \\
& - \frac{k+2}{2(k+1)} \cdot \frac{p}{1-p} \|\eta F(z_{k+1}) + \eta c_{k+1}\|^2.
\end{aligned}
$$

Since $k \geq 1$, we have $\frac{k+2}{2(k+1)} \leq 1$.

Now it remains to prove Expression (26) + Expression (27) is non-negative. Recall that $c = -\frac{4p\rho}{\eta}$.

$$
\begin{aligned}
& \frac{2}{(k+1)(k+2)} \cdot (\text{Expression(26)} + \text{Expression(27)}) \\
=& \left(1 - \left(\frac{1}{p} - \frac{4\rho}{\eta}\right)\eta^2 L^2\right) \cdot \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 - \frac{4\rho}{\eta} \cdot \left\| \eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1}) \right\|^2 \\
& + \frac{2\rho}{\eta} \cdot \|\eta F(z_{k+1}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k\|^2 \\
\geq& \left(1 - \left(\frac{1}{p} - \frac{4\rho}{\eta}\right)\eta^2 L^2\right) \cdot \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 + \frac{4\rho}{\eta} \left\| \eta F(z_{k+\frac{1}{2}}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k \right\|^2 \\
& && (\|A\|^2 - \tfrac{1}{2}\|B\|^2 \geq -\|A+B\|^2) \\
=& \left(1 - \left(\frac{1}{p} - \frac{4\rho}{\eta}\right)\eta^2 L^2 + \frac{4\rho}{\eta}\right) \cdot \left\| z_{k+\frac{1}{2}} - z_{k+1} \right\|^2 \\
\geq& 0. && \text{(Fact 2: Inequality (21) and } p = 1/4)
\end{aligned}
$$

The last equality holds because, by the update rule of (EAG+)

$$
\begin{aligned}
z_{k+\frac{1}{2}} - z_{k+1} &= \left(z_k - \eta F(z_k) + \frac{1}{k+2}(z_0 - z_k) - \eta c_k \right) - \left(z_k - \eta F(z_{k+\frac{1}{2}}) + \frac{1}{k+2}(z_0 - z_k) - \eta c_{k+1} \right) \\
&= \eta F(z_{k+\frac{1}{2}}) + \eta c_{k+1} - \eta F(z_k) - \eta c_k.
\end{aligned}
$$

Hence, $V_{k+1} \le V_k + \frac{p}{1-p}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2 = V_k + \frac{1}{3}\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2$. □

Lemma 3. *Suppose Assumption 2 holds with $\rho \in [0, -\frac{1}{24L}]$. Let $z_0 \in \mathbb{R}^n$ be any starting point and $\{z_{k+\frac{1}{2}}, z_{k+1}\}_{k\ge 0}$ be the iterates of (EAG+) with step size $\eta \in (0, \frac{1}{2L})$ that satisfies Inequality (21). Then for any $k \ge 2$,*

$$
\frac{(k+1)^2}{4} \cdot \|\eta F(z_k) + \eta c_k\|^2 \le V_1 + \|z_0 - z^*\|^2 + \frac{1}{3} \cdot \sum_{t=2}^{k-1} \|\eta F(z_t) + \eta c_t\|^2.
$$

Proof. Fix any $k \ge 2$. By definition, we have

$$
\begin{aligned}
V_k &= \frac{(k+1)(k+2)}{2}\|\eta F(z_k) + \eta c_k\|^2 + (k+1)\langle \eta F(z_k) + \eta c_k, z_k - z_0\rangle \\
&= \frac{(k+1)(k+2)}{2}\|\eta F(z_k) + \eta c_k\|^2 + (k+1)\langle \eta F(z_k) + \eta c_k, z^* - z_0\rangle + (k+1)\langle \eta F(z_k) + \eta c_k, z_k - z^*\rangle \\
&\ge \frac{(k+1)(k+2)}{2}\|\eta F(z_k) + \eta c_k\|^2 + (k+1)\langle \eta F(z_k) + \eta c_k, z^* - z_0\rangle + \frac{(k+1)\rho}{\eta}\|\eta F(z_k) + \eta c_k\|^2 \\
&\qquad\qquad (\mathbf{0} \in F(z^*) + A(z^*) \text{ and } F + A \text{ is } \rho\text{-comonotone}) \\
&\ge \frac{(k+1)(k+\frac{3}{2})}{2}\|\eta F(z_k) + \eta c_k\|^2 + (k+1)\langle \eta F(z_k) + \eta c_k, z^* - z_0\rangle \\
&\qquad\qquad (\tfrac{\rho}{\eta} \ge -\tfrac{1}{4} \text{ according to Fact 2}) \\
&\ge \frac{(k+1)(k+\frac{3}{2})}{2}\|\eta F(z_k) - \eta c_k\|^2 - \frac{(k+1)(k+\frac{3}{2})}{4}\|\eta F(z_k) - \eta c_k\|^2 - \frac{k+1}{k+\frac{3}{2}}\|z_0 - z^*\|^2 \\
&\qquad\qquad (\ \langle a, b\rangle \ge -\tfrac{c}{4}\|a\|^2 - \tfrac{1}{c}\|b\|^2) \\
&\ge \frac{(k+1)(k+\frac{3}{2})}{4}\|\eta F(z_k) + \eta c_k\|^2 - \|z_0 - z^*\|^2. \qquad\qquad (k \ge 2)
\end{aligned}
$$

According to Theorem 6, we know that

$$
V_k \le V_1 + \frac{1}{3} \cdot \sum_{t=2}^{k} \|\eta F(z_t) + \eta c_t\|^2.
$$

Combing the above two inequalities and rearranging terms, we obtain the following inequality for any $k \ge 2$:

$$
\left(\frac{(k+1)(k+\frac{3}{2})}{4} - \frac{1}{3} \right) \|\eta F(z_k) + \eta c_k\|^2 \le V_1 + \|z_0 - z^*\|^2 + \frac{1}{3} \cdot \sum_{t=2}^{k-1} \|\eta F(z_t) + \eta c_t\|^2.
$$

Since $\frac{1}{3} \leq \frac{(k+1)}{8}$ for all $k \geq 2$, we can further simplify the inequality for any $k \geq 2$:

$$\frac{(k+1)^2}{4} \cdot \|\eta F(z_k) + \eta c_k\|^2 \leq V_1 + \|z_0 - z^*\|^2 + \frac{1}{3} \cdot \sum_{t=2}^{k-1} \|\eta F(z_t) + \eta c_t\|^2.$$

□

Theorem 7. *Suppose Assumption 2 holds with $\rho \in [0, -\frac{1}{24L}]$. Let $z_0 \in \mathbb{R}^n$ be any starting point and $\{z_{k+\frac{1}{2}}, z_{k+1}\}_{k\geq 0}$ be the iterates of (EAG+) with step size $\eta \in (0, \frac{1}{2L})$ that satisfies Inequality (21). Then for any $T \geq 1$,*

$$r^{nat}_{F,A}(z_T)^2 = \left\| z_T - J_A\left[z_T - F(z_T)\right] \right\|^2 \leq \min_{c \in A(z_T)} \|F(z_T) + c\|^2 = r^{tan}_{F,A}(z_T)^2 \leq \frac{16}{\eta^2 L^2} \cdot \frac{H_0^2 L^2}{T^2},$$

where $H_0^2 = \|z_0 - z^*\|^2 + \frac{15}{4L^2} \cdot r^{tan}_{F,A}(z_0)^2$.

Proof. The first inequality follows from Fact 1. To prove the second inequality, we denote $a_k := \frac{\|\eta F(z_k) + \eta c_k\|^2}{H_0^2}$ for $k \geq 1$. It suffices to prove for all $k \geq 1$,

$$a_k \leq \frac{16}{k^2}.$$

From the proof of Proposition 2, we have

$$\|\eta F(z_1) + \eta c_1\|^2 \leq \eta^2 (1 + \eta L + \eta^2 L^2)^2 \cdot r^{tan}_{F,A}(z_0)^2 \quad \Rightarrow \quad a_1 \leq 1.$$

From Proposition 2, we have $V_1 \leq \frac{15}{4L^2} \cdot r^{tan}_{F,A}(z_0)^2$ and thus $V_1 + \|z_0 - z^*\|^2 \leq H_0^2$. Then from Lemma 3, we have

$$\frac{k^2}{4} \cdot a_k \leq 1 + \frac{1}{3} \cdot \sum_{t=2}^{t-1} a_t.$$

Note that $\frac{1}{3} = \frac{\frac{1}{4}}{1 - \frac{1}{4}}$. Applying Proposition 4 with $C_1 = 1$ and $p = \frac{1}{4}$, we get for all $k \geq 2$,

$$a_k \leq \frac{4C_1}{(1 - 3p)k^2} = \frac{16}{k^2}.$$

This completes the proof. □

5 Auxiliary Propositions

Proposition 3. *Let $x_0, x_1, x_2, x_3, y_1, y_2, y_3, u_1, u_3$ be arbitrary vectors in $\mathbb{R}^n$ and $p, q \neq 0$ be real numbers. When $u_3 = x_1 - y_2 + \frac{1}{q+1}(x_0 - x_1) - x_3$, the following identity holds:*

$$
\begin{aligned}
&\frac{q(q+1)}{2} \cdot \|y_1 + u_1\|^2 + q \cdot \langle y_1 + u_1, x_1 - x_0 \rangle \\
&- \left(\frac{(q+1)(q+2)}{2} \cdot \|y_3 + u_3\|^2 + (q+1) \cdot \langle y_3 + u_3, x_3 - x_0 \rangle \right) \\
&- \frac{q(q+1)}{2p} \cdot \left(p \cdot \|x_2 - x_3\|^2 - \|y_2 - y_3\|^2 \right) \\
&- q(q+1) \cdot \langle y_3 - y_1, x_3 - x_1 \rangle \\
&- q(q+1) \cdot \left\langle x_1 - y_1 - x_2 + \frac{1}{q+1}(x_0 - x_1), x_2 - x_3 \right\rangle \\
&- q(q+1) \cdot \langle u_3, x_3 - x_1 \rangle \\
&- q(q+1) \cdot \langle u_1, x_1 - x_2 \rangle \\
=&\frac{q(q+1)}{2} \cdot \left\| x_2 - x_1 + y_1 + u_1 + \frac{1}{q+1}(x_1 - x_0) \right\|^2 \\
&+ \frac{(1-p)q(q+1)}{2p} \cdot \|y_2 - y_3\|^2 \\
&+ (q+1) \cdot \langle y_2 - y_3, y_3 + u_3 \rangle
\end{aligned}
\tag{28}
$$

Moreover, if $u_1 = x_1 - y_1 + \frac{1}{q+1}(x_0 - x_1) - x_2$, then the following identity holds:

$$
\begin{aligned}
&\frac{q(q+1)}{2} \cdot \|y_1 + u_1\|^2 + q \cdot \langle y_1 + u_1, x_1 - x_0 \rangle \\
&- \left(\frac{(q+1)(q+2)}{2} \cdot \|y_3 + u_3\|^2 + (q+1) \cdot \langle y_3 + u_3, x_3 - x_0 \rangle \right) \\
&- \frac{q(q+1)}{2p} \cdot \left(p \cdot \|x_2 - x_3\|^2 - \|y_2 - y_3\|^2 \right) \\
&- q(q+1) \cdot \langle y_3 + u_3 - y_1 - u_1, x_3 - x_1 \rangle \\
=&\frac{(1-p)q(q+1)}{2p} \cdot \|y_2 - y_3\|^2 \\
&+ (q+1) \cdot \langle y_2 - y_3, y_3 + u_3 \rangle
\end{aligned}
\tag{29}
$$

Proof. We verify the first identity using MATLAB. Here is the verification MATLAB code [4]. The second identity follows easily from the first identity. Notice that when $u_1 = x_1 - y_1 + \frac{1}{q+1}(x_0 - x_1) - x_2$, the first term on the RHS of Identity (28) is 0. Moreover, as we demonstrate below, the

[4] https://github.com/weiqiangzheng1999/Accelerated-Non-Monotone-Inclusion/blob/main/Identity_verification.m

sum of the fourth term to the seventh term on the LHS of Identity (28) is exactly identical to the fourth term on the LHS of Identity (29).

$$
\begin{aligned}
&\langle y_3 - y_1, x_3 - x_1 \rangle \\
&+ \left\langle x_1 - y_1 - x_2 + \frac{1}{q+1}(x_0 - x_1), x_2 - x_3 \right\rangle \\
&+ \langle u_3, x_3 - x_1 \rangle \\
&+ \langle u_1, x_1 - x_2 \rangle \\
=&\langle y_3 - y_1, x_3 - x_1 \rangle + \langle u_1, x_2 - x_3 \rangle + \langle u_3, x_3 - x_1 \rangle + \langle u_1, x_1 - x_2 \rangle \\
=&\langle y_3 + u_3 - y_1 - u_1, x_3 - x_1 \rangle
\end{aligned}
$$

Therefore, Identity (29) follows from Identity (28). □

Proposition 4. *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 $k \geq 2$,*

$$
\frac{k^2}{4} \cdot a_k \leq C_1 + \frac{p}{1-p} \cdot \sum_{t=2}^{k-1} a_t, \tag{30}
$$

then for each $k \geq 2$ we have

$$
a_k \leq \frac{4 \cdot C_1}{1-3p} \cdot \frac{1}{k^2}. \tag{31}
$$

Proof. We prove the statement by induction.
Base Case: $k = 2$**.** From Inequality (30), 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 (31) holds for $k = 2$.
Inductive Step: for any $k \geq 3$**.** Fix some $k \geq 3$ and assume that Inequality (31) 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 (30), we have

$$
\begin{aligned}
\frac{k^2}{4} \cdot a_k &\leq C_1 + \frac{p}{1-p} \cdot \sum_{t=2}^{k-1} a_t \\
&\leq \frac{C_1}{1-p} + \frac{p}{1-p} \cdot \sum_{t=3}^{k-1} a_t && (a_2 \leq C_1) \\
&\leq \frac{C_1}{1-p} + \frac{4p \cdot C_1}{(1-p)(1-3p)} \cdot \sum_{t=3}^{k-1} \frac{1}{t^2} && \text{(Induction assumption on Inequality (31))} \\
&\leq \frac{C_1}{1-p} + \frac{2p \cdot C_1}{(1-p)(1-3p)} && \left(\textstyle\sum_{t=3}^{\infty} \frac{1}{t^2} = \frac{\pi^2}{6} - \frac{5}{4} \leq \frac{1}{2}\right) \\
&= \frac{C_1}{1-3p}.
\end{aligned}
$$

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

References

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

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

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

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

Y. Cai, A. Oikonomou, and W. Zheng. Tight Last-Iterate Convergence of the Extragradient and the Optimistic Gradient Descent-Ascent Algorithm for Constrained Monotone Variational Inequalities. Technical Report arXiv:2204.09228, arXiv, May 2022. URL http://arxiv.org/abs/2204.09228. arXiv:2204.09228 [cs, math] type: article.

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

C. D. Dang and G. Lan. 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.

C. Daskalakis, S. Skoulakis, and M. Zampetakis. The complexity of constrained min-max optimization. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1466–1478, 2021.

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

J. Diakonikolas, C. Daskalakis, and M. I. Jordan. Efficient Methods for Structured Nonconvex-Nonconcave Min-Max Optimization. *arXiv:2011.00364 [cs, math, stat]*, Feb. 2021. URL http://arxiv.org/abs/2011.00364. arXiv: 2011.00364.

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

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

N. Golowich, S. Pattathil, and C. Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. *arXiv:2010.13724 [cs, math]*, Oct. 2020. URL http://arxiv.org/abs/2010.13724. arXiv: 2010.13724.

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

E. Gorbunov, N. Loizou, and G. Gidel. Extragradient Method: $O(1/K)$ Last-Iterate Convergence for Monotone Variational Inequalities and Connections With Cocoercivity. *arXiv:2110.04261 [cs, math]*, Oct. 2021. URL http://arxiv.org/abs/2110.04261. arXiv: 2110.04261.

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

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

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

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

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

F. Lieder. On the convergence rate of the Halpern-iteration. *Optimization Letters*, 15(2): 405–418, Mar. 2021. ISSN 1862-4472, 1862-4480. doi: 10.1007/s11590-020-01617-9. URL https://link.springer.com/10.1007/s11590-020-01617-9.

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

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

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

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

A. Nemirovski. Prox-method with rate of convergence O (1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004. Publisher: SIAM.

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

Y. Ouyang and Y. Xu. Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems. *Mathematical Programming*, 185(1):1–35, 2021.

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

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

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

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

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

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

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

T. Yoon and E. K. Ryu. Accelerated Algorithms for Smooth Convex-Concave Minimax Problems with $\mathcal{O}(1/k^2)$ Rate on Squared Gradient Norm. *arXiv:2102.07922 [math]*, June 2021. URL http://arxiv.org/abs/2102.07922. arXiv: 2102.07922.

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