Accelerated Single-Call Methods for Constrained Min-Max Optimization

Yang Cai YANG.CAI@YALE.EDU
Weiqiang Zheng WEIQIANG.ZHENG@YALE.EDU
Yale University

Abstract

We study first-order methods for constrained min-max optimization. Existing methods either requires two gradient calls or two projections in each iteration, which may be costly in applications. In this paper, we first show that the *Optimistic Gradient (OG)* method, a *single-call single-projection* algorithm, has $O(\frac{1}{\sqrt{T}})$ convergence rate for inclusion problems with operators that satisfy the weak Minty variation inequality (MVI). Our second result is the first single-call single-projection algorithm – the *Accelerated Reflected Gradient (ARG)* method that achieves the *optimal* $O(\frac{1}{T})$ convergence rate for inclusion problems that satisfy negative comonotonicity. Both the weak MVI and negative comonotonicity are well-studied assumptions and capture a rich set of non-convex non-concave min-max optimization problems. Finally, we show that the *Reflected Gradient (RG)* method, another *single-call single-projection* algorithm, has $O(\frac{1}{\sqrt{T}})$ last-iterate convergence rate for constrained convex-concave min-max optimization, answering an open problem of [26].

1. Introduction

Various Machine Learning applications, from the generative adversarial networks (GANs) (e.g., [1, 20]), adversarial examples (e.g., [33]), robust optimization (e.g., [4]), to reinforcement learning (e.g., [10, 16]), can be captured by constrained min-max optimization. Unlike the well-behaved convex-concave setting, these modern ML applications often require solving non-convex non-concave min-max optimization problems in high dimensional spaces.

Unfortunately, the general non-convex non-concave setting is intractable even for computing a *local* solution [13, 25, 44]. Motivated by the intractability, researchers turn their attention to non-convex non-concave settings *with structure*. Significant progress has been made for several interesting structured non-convex non-concave settings, such as the ones that satisfy the weak Minty variation inequality (MVI) (Definition 2) [15, 46] and the ones that satisfy the more strict negatively comonotone condition (Definition 3) [6, 30]. These algorithms are variations of the celebrated extragradient (EG) method [29], an iterative first-order method. Similar to the extragradient method, these algorithms all require *two oracle calls* per iteration, which may be costly in practice. We investigate the following important question in this paper:

Can we design efficient ***single-call*** *first-order methods for*
structured non-convex non-concave min-max optimization? (*)

We provide an affirmative answer to the question. We first show that a single-call method known as the *Optimistic Gradient* (OG) method [26] is applicable to all non-convex non-concave settings that satisfy the *weak MVI*. We then provide the *Accelerated Reflected Gradient* (ARG) method

Algorithm		Single-Call?	Constraints?	Non-Monotone	
				Comonotone	weak MVI
Normal	EG+ [15]	✗	✗	$O(\frac{1}{\sqrt{T}})$	$O(\frac{1}{\sqrt{T}})$
	CEG+ [46]	✗	✓	$O(\frac{1}{\sqrt{T}})$	$O(\frac{1}{\sqrt{T}})$
	OG[**This paper**]	✓	✓	$O(\frac{1}{\sqrt{T}})$	$O(\frac{1}{\sqrt{T}})$
Accelerated	FEG [30]	✗	✗	$O(\frac{1}{T})$	
	AS [6]	✗	✓	$O(\frac{1}{T})$	
	ARG [**This paper**]	✓	✓	$O(\frac{1}{T})$	

Table 1: Existing results for min-max optimization problem in the nonconvex-nonconcave setting. A ✓ in "Constraints?" means the algorithm works in the constrained setting. The convergence rate is in terms of the operator norm (in the unconstrained setting) and the residual (in the constrained setting).

that achieves the optimal convergence rate in all non-convex non-concave settings that satisfy the *negatively comonotone condition*. Single-call methods have been studied in the convex-concave settings [26] but not for the more general non-convex non-concave settings. See Table 1 for comparisons between our algorithms and other algorithms from the literature.

1.1. Our Contributions

Throughout the paper, we adopt the more general and abstract framework of inclusion problems, which includes constrained min-max optimization as a special case. More specifically, we consider the following problem.

Inclusion Problem. Given $E = F + A$ where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a single-valued (possibly non-monotone) operator and $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a set-valued maximally monotone operator, the *inclusion* problem is defined as follows

$$\text{find } z^* \in \mathcal{Z} \text{ such that } \mathbf{0} \in E(z^*) = F(z^*) + A(z^*). \tag{IP}$$

As shown in the following example, we can interpret a min-max optimization problem as an inclusion problem.

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 semi-continuous and convex functions g and h. Examples include the ℓ_1-norm, the ℓ_2-norm, and the indicator function of a closed 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 an inclusion problem.

[13] shows that without any assumption on the operator $E = F + A$, the problem is intractable.[1] The most well understood setting is when E is monotone, i.e., $\langle u - v, z - z' \rangle \geq 0$ for all z, z' and $u \in E(z)$, $v \in E(z')$, which captures convex-concave min-max optimization. Motivated by non-convex non-concave min-max optimization, we consider the two most widely studied families of non-monotone operators: (i) negatively comonotone operators and (ii) operators satisfy the less restrictive weak MVI. See Section 2 for more detailed discussion on their relationship. Here are the main contributions of this paper.

Contribution 1: We provide an extension of the *Optimistic Gradient (OG)* method for inclusion problems when the operator $E = F + A$ satisfies the weak MVI. More specifically, we prove that OG has a $O(\frac{1}{\sqrt{T}})$ convergence rate (Theorem 4) matching the state of the art algorithms [15, 46]. Importantly, our algorithm only requires a single oracle call to F and a single call to the resolvent of A.[a]

a. When A is the subdifferential of the indicator function of a closed convex set, the resolvent operator is exactly the Euclidean projection. Hence our algorithm performs a single projection in the constrained case.

Next, we provide an accelerated single-call method when the operator satisfies the stronger negatively comonotone condition.

Contribution 2: We design an accelerated version of the *Reflected Gradient (RG)* [8, 9, 26, 34] method that we call the *Accelerated Reflected Gradient* (ARG) method, which has the optimal $O(\frac{1}{T})$ convergence rate for inclusion problems whose operators $E = F + A$ are negatively comonotone (Theorem 5). Note that $O(\frac{1}{T})$ is the optimal convergence rate for any first-order methods even for monotone inclusion problems [14, 54]. Importantly, ARG only requires a single oracle call to F and a single call to the resolvent of A.

Finally, we resolve an open question from [26].

Contribution 3: We show that the *Reflected Gradient (RG)* method has a *last-iterate* convergence rate of $O(\frac{1}{\sqrt{T}})$ for constrained convex-concave min-max optimization (Theorem 6). [26] show that the RG algorithm asymptotically converges but fails to obtain a concrete rate. We strengthen their result to obtain a tight finite convergence rate for RG.

1.2. Related Works

We provide a brief discussion of the most relevant and recent results on nonconvex-nonconcave min-max optimization here and defer the discussion on related results in the convex-concave setting to Appendix A. We also refer readers to [2, 17, 49] and references therein for a comprehensive literature review on inclusion problems and related variational inequality problems.

Structured Nonconvex-Nonconcave Min-Max Optimization. Since in general nonconvex-nonconcave min-max optimization problems are intractable, recent works study problems under additional assumptions.

The *Minty variational inequality* (MVI) assumption (also called coherence or variationally stablility), which covers all quasiconvex-concave and starconvex-concave problems, is well-studied in e.g., [11, 31, 32, 35, 50, 56]. Extragradient-type algorithms has $O(\frac{1}{\sqrt{T}})$ convergence rate for problems that satisfies MVI [11].

1. Indeed, even if A is maximally monotone, [13] implies that the problem is still intractable without further assumptions on F.

[15] proposes a weaker assumption called *weak MVI*, which includes both MVI or negative comonotonicity [3] as special cases. Under the weak MVI, the EG+ [15] and OGDA+ [5] algorithms have $O(\frac{1}{\sqrt{T}})$ convergence rate in the unconstrained setting. Recently, [46] generalizes EG+ to CEG+ algorithm, achieving the same convergence rate in the general (constrained) setting. To the best of our knowledge, the OG algorithm is the only single-call single-resolvent algorithm with $O(\frac{1}{\sqrt{T}})$ convergence rate when we only assume weak MVI (Theorem 4).

The result for accelerated algorithms in the nonconvex-nonconcave setting is sparser. For negatively comonotone operators, optimal $O(\frac{1}{T})$ convergence rate is achieved by variants of the EG algorithm in the unconstrained setting [30] and in the constrained setting [6]. To the best of our knowledge, the ARG algorithm is the first efficient single-call single-resolvent method that achieves the accelerated and optimal $O(\frac{1}{T})$ convergence rate in the constrained nonconvex-nonconcave setting (Theorem 5). We summarize previous results and our results in Table 1.

2. Preliminaries

Basic Notations. Throughout the paper, we focus on the Euclidean space $\mathbb{R}^n$ equipped with inner product $\langle \cdot, \cdot \rangle$. We denote the standard ℓ_2-norm by $\| \cdot \|$. For any closed and convex set $\mathcal{Z} \subseteq \mathbb{R}^n$, $\Pi_Z[\cdot] : \mathbb{R}^n \to \mathcal{Z}$ denotes the Euclidean projection onto set $\mathcal{Z}$ such that for any $z \in \mathbb{R}^n, \Pi_{\mathcal{Z}}[z] = \operatorname{argmin}_{z' \in \mathcal{Z}} \|z - z'\|$. We denote $\mathcal{B}(z, r)$ the ℓ_2-ball centered at z with radius r.

Monotone Operator. We recall some standard definitions and results on monotone operators here and refer the readers to [2, 48, 49] for more detailed introduction. A *set-valued operator* $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ maps each point $z \in \mathbb{R}^n$ to a subset $A(z) \subseteq \mathbb{R}^n$. We denote the *graph* of A as $\mathrm{Gra}(A) := \{(z, u) : u \in A(z)\}$ and the *zeros* of A as $\mathrm{Zer}(A) = \{z : \mathbf{0} \in A(z)\}$. 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 $A + B$ to be 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 $I : \mathbb{R}^n \to \mathbb{R}^n$. We say operator A is *single valued* if $|A(z)| \leq 1$ for all $z \in \mathbb{R}^n$. Single-valued operator A is *L-Lipschitz* if

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

Moreover, we say A is *non-expansive* if it is 1-Lipschitz.

Definition 1 ((Maximally) monotonicity) *An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is* monotone *if*

$$\left\langle u - u', z - z' \right\rangle \geq 0, \quad \forall (z, u), (z', u') \in \mathrm{Gra}(A).$$

Moreover, A is maximally monotone *if A is monotone and $\mathrm{Gra}(A)$ is not properly contained in the graph of any other monotone operators.*

Non-Monotone Operator.

Definition 2 (Weak MVI [15, 46]) *An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ satisfies* weak MVI *if for some $z^* \in \mathrm{Zer}(A)$, there exists $\rho \leq 0$*

$$\langle u, z - z^* \rangle \geq \rho \|u\|^2, \quad \forall (z, u) \in \mathrm{Gra}(A).$$

Definition 3 (Comonotonicity [3]) *An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is* ρ-comonotone *if*

$$\left\langle u - u', z - z' \right\rangle \geq \rho \left\| u - u' \right\|^2, \quad \forall (z, u), (z', u') \in \mathrm{Gra}(A).$$

When A is ρ-comonotone for $\rho > 0$, then A is also known as ρ-cocoercive, which is a stronger condition than monotonicity. When A is ρ-comonotone for $\rho < 0$, then A is non-monotone. Weak MVI with $\rho = 0$ is also know as MVI, coherence, or variational stability. Note that the weak MVI is implied by negative comonotonicity. We refer the readers to [30, Example 1], [15, Section 2.2] and [46, Section 5] for examples of min-max optimization problems that satisfy the two conditions.

2.1. Problem Formulation

Inclusion Problem. We say z is an ϵ-approximate solution to an inclusion problem (IP) if

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

Throughout the paper, we study IP problems under the following assumption.

Assumption 1 *In the setup of IP,*

1. *there exists $z^* \in \mathrm{Zer}(E)$, i.e., $\mathbf{0} \in F(z^*) + A(z^*)$.*
2. *F is L-Lipschitz.*
3. *A is maximally monotone.*

When F is monotone, we refer to the corresponding IP problem as a *monotone inclusion* problem, which covers convex-concave min-max optimization. In the more general non-monotone setting, we would study problems that satisfy negative comonotonicity or weak MVI.

Assumption 2 *In the setup of IP, $E = F + A$ is ρ-comonotone, i.e.,*

$$\left\langle u - u', z - z' \right\rangle \geq \rho \left\| u - u' \right\|^2, \quad \forall (z, u), (z', u') \in \mathrm{Gra}(E).$$

Assumption 3 *In the setup of IP, $E = F + A$ satisfies weak MVI with $\rho \leq 0$, i.e., there exists $z^* \in \mathrm{Zer}(E)$,*

$$\langle u, z - z^* \rangle \geq \rho \|u\|^2, \quad \forall (z, u) \in \mathrm{Gra}(E).$$

Note that by definition, the parameter ρ in both assumption must be $O(\frac{1}{L})$. An important special case of inclusion problem is the variational inequality problem.

Variational Inequality. Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed and convex set and $F : \mathbb{R}^n \to \mathbb{R}^n$ be a single-valued operator. The *variation inequality* (VI) problem associated with $\mathcal{Z}$ and F is stated as

$$\text{find } z^* \in \mathcal{Z} \text{ such that } \langle F(z^*), z^* - z \rangle \leq 0, \ \forall z \in \mathcal{Z}. \tag{VI}$$

Note that VI is a special case of IP when $A = N_{\mathcal{Z}} = \partial \mathbb{I}_{\mathcal{Z}}$ is the normal cone operator:

$$\mathbf{0} \in F(z^*) + N_{\mathcal{Z}}(z^*) \Leftrightarrow -F(z^*) \in N_{\mathcal{Z}}(z^*) \Leftrightarrow \langle F(z^*), z^* - z \rangle \leq 0, \forall z \in \mathcal{Z}.$$

The general formulation of VI unifies many problems such as convex optimization, min-max optimization, computing Nash equilibria in multi-player concave games, and is extensively-studied since 1960s [17]. Definitions of the convergence measure for VI and the classical algorithms, EG and PEG, are presented in Appendix B.

2.2. Convergence Measure

We focus on a strong convergence measure called the *tangent residual*, defined as

$$r^{tan}_{F,A}(z) := \min_{c \in A(z)} \|F(z) + c\|$$

It is clear by definition that $r^{tan}_{F,A}(z) \le \epsilon$ implies z is an ϵ-approximate solution to the inclusion (IP) problem, and also an $(\epsilon \cdot D)$ approximate strong solution to the corresponding variational inequality (VI) problem when $\mathcal{Z}$ is bounded by D. Moreover, the tangent residual is an upper bound of other notion of residuals in the literature such as the natural residual $r^{nat}_{F,A}$ [14] or the forward-backward residual $r^{fb}_{F,A}$ [55] as shown in Fact 1. We defer the proof to Appendix B.4. Note that in the unconstrained setting where $A = 0$, these residuals are all equivalent to the *operator norm* $\|F(z)\|$.

Fact 1 *Let A be a maximally monotone operator and F be an single-valued operator. Then for any $z \in \mathbb{R}^n$ and $\alpha > 0$,*

$$\begin{aligned} r^{tan}_{F,A}(z) &\ge r^{nat}_{F,A}(z) := \|z - J_A(z - F(z))\| \\ r^{tan}_{F,A}(z) &\ge r^{fb}_{F,A,\alpha}(z) := \frac{1}{\alpha}\|z - J_{\alpha A}[z - \alpha F(z)]\| \end{aligned}$$

3. Optimistic Gradient Method for Weak MVI Problems

In this section, we consider an extension of the *Optimistic Gradient* (OG) algorithm [12, 26, 38, 39, 45] for inclusion problems: given arbitrary starting point $z_{-\frac{1}{2}} = z_0 \in \mathbb{R}^n$ and step size $\eta > 0$, the update rule is

$$\begin{aligned} z_{t+\frac{1}{2}} &= J_{\eta A}\Big[z_t - \eta F(z_{t-\frac{1}{2}})\Big] \\ z_{t+1} &= z_{t+\frac{1}{2}} + \eta F(z_{t-\frac{1}{2}}) - \eta F(z_{t+\frac{1}{2}}) \end{aligned} \tag{OG}$$

For $t \ge 1$, the update rule can also be written as $z_{t+\frac{3}{2}} = J_{\eta A}[z_{t+\frac{1}{2}} - 2\eta F(z_{t+\frac{1}{2}}) + \eta F(z_{t-\frac{1}{2}})]$, which coincides with the forward-reflected-backward algorithm [36]. We remark that the update rule of OG is different from the *Optimistic Gradient Descent/Ascent (OGDA)* algorithm (also known as *Past Extra Gradient* (PEG) *algorithm*) [47], which is single-call but requires two projections in each iteration.

Previous results for OG only hold in the convex-concave (monotone) setting. The main result in this section is that OG has $O(\frac{1}{\sqrt{T}})$ convergence rate even for nonconvex-nonconcave min-max optimization problems that satisfy weak MVI, matching the state of the art results achieved by two-call methods [15, 46]. Remarkably, OG only requires single call to F and single call to the resolvent $J_{\eta A}$ in each iteration. The main result is shown in Theorem 4. The proof relies on a simple yet important observation that $\frac{z_t - z_{t+1}}{\eta} \in F(z_{t+\frac{1}{2}}) + A(z_{t+\frac{1}{2}})$.

Theorem 4 *Assume Assumption 1 and 3 hold with $\rho \in (-\frac{1}{12\sqrt{3}L}, 0]$. Consider the iterates of* (OG) *with step size $\eta \in (0, \frac{1}{2L})$ satisfying $C = \frac{1}{2} + \frac{2\rho}{\eta} - 2\eta^2 L^2 > 0$ (existence of such η is guaranteed by Fact 2). Then for any $T \ge 1$,*

$$\min_{t \in [T]} r^{tan}_{F,A}(z_{t+\frac{1}{2}})^2 \le \min_{t \in [T]} \frac{\|z_{t+1} - z_t\|^2}{\eta^2} \le \frac{H^2}{C\eta^2} \cdot \frac{1}{T},$$

where $H^2 = \|z_1 - z^*\|^2 + \frac{1}{4}\|z_{\frac{1}{2}} - z_0\|^2$.

4. Accelerated Reflected Gradient For Negatively Comonotone Problems

In this section, we propose a new algorithm called the *Accelerated Reflected Gradient* (ARG) algorithm. We prove that ARG enjoys accelerated $O(\frac{1}{T})$ last-iterate convergence rate for inclusion problems with comonotone operators (Theorem 5). Note that the lower bound $\Omega(\frac{1}{T})$ holds even for the special case of convex-concave min-max optimization [14, 54].

Our algorithm is inspired by the *Reflected Gradient* (RG) algorithm [8, 9, 26, 34] for monotone variational inequalities. Starting at initial points $z_{-1} = z_0 \in \mathcal{Z}$, the update rule of RG with step size $\eta > 0$ is as follows: for $t = 0, 1, 2, \cdots$

$$\begin{aligned} z_{t+\frac{1}{2}} &= 2z_t - z_{t-1} \\ z_{t+1} &= \Pi_{\mathcal{Z}}\left[z_t - \eta F(z_{t+\frac{1}{2}})\right] \end{aligned} \tag{RG}$$

We propose the following Accelerated Reflected Gradient (ARG) algorithm, which is a single-call single-resolvent first-order method. Given arbitrary initial points $z_0 = z_{\frac{1}{2}} \in \mathbb{R}^n$ and step size $\eta > 0$, ARG sets $z_1 = J_{\eta A}[z_0 - \eta F(z_0)]$ and updates for $t = 1, 2, \cdots$

$$\begin{aligned} z_{t+\frac{1}{2}} &= 2z_t - z_{t-1} + \frac{1}{t+1}(z_0 - z_t) - \frac{1}{t}(z_0 - z_{t-1}) \\ z_{t+1} &= J_{\eta A}\left[z_t - \eta F(z_{t+\frac{1}{2}}) + \frac{1}{t+1}(z_0 - z_t)\right] \end{aligned} \tag{ARG}$$

We use the idea from *Halpern iteration* [23] to design the accelerated algorithm (ARG). This technique for deriving optimal first-order methods is also called *Anchoring* and receives intense attention recently [6, 14, 30, 51, 52, 54]. We defer detailed discussion on these works to Appendix A. We remark that the state of the art result from [6] is a two-call variant of the EG algorithm. Thus to the best of our knowledge, ARG is the first single-call single-resolvent algorithm with optimal convergence rate for general inclusion problems with comonotone operators.

Theorem 5 *Assume Assumption 1 and 2 hold with $\rho \in [-\frac{1}{60L}, 0]$, then the accelerated reflected gradient (ARG) algorithm with constant step size $\eta > 0$ satisfying Inequality (11) has the following last-iterate convergence rate: for any $T \geq 1$,*

$$r^{tan}_{F,A}(z_T) \leq \frac{\sqrt{6}H}{\eta} \cdot \frac{1}{T}$$

where $H^2 = \|z_0 - z^*\|^2 + 4\|z_1 - z_0\|^2 \leq \|z_0 - z^*\|^2 + 4r^{tan}_{F,A}(z_0)^2$.

We defer the proof to Appendix D.

5. Last-Iterate Convergence Rate of Reflected Gradient

In this section, we show that the *Reflected Gradient* (RG) algorithm [8, 9, 26, 34] has a last-iterate convergence rate of $O(\frac{1}{\sqrt{T}})$ with respect to tangent residual and gap function (see Definition 8) for solving monotone variational inequalities (Theorem 6). We defer the proof to Appendix E

Theorem 6 *For a variational inequality problem* (VI) *associated with a closed convex set $\mathcal{Z}$ and a monotone and L-Lipschitz operator F with a solution z^*, the* (RG) *algorithm with constant step size $\eta \in (0, \frac{1}{(1+\sqrt{2})L})$ has the following last-iterate convergence rate: for any $T \geq 1$,*

$$r^{tan}_{F,\mathcal{Z}}(z_T) \leq \frac{\lambda HL}{\sqrt{T}}, \quad \textsc{Gap}_{\mathcal{Z},F,D}(z_T) \leq \frac{\lambda DHL}{\sqrt{T}}$$

where $H^2 = 5\|z_0 - z^*\|^2 + \frac{13}{L^2}\|F(z_0)\|^2$ and $\lambda = \sqrt{\frac{6(1+3\eta^2L^2)}{\eta^2L^2(1-(1+\sqrt{2})\eta L)}}$.

We remark that the convergence rate of RG is slower than ARG and other optimal first-order algorithms even in the monotone setting. Nevertheless, understanding its last-iterate convergence rate is still interesting: (1) RG is simple and largely used in practice; (2) Last-iterate convergence rates of simple classic algorithms such as EG and RG are mentioned as open problems in [26]. The question is recently resolved for EG [7, 21] but remains open for RG; (3) Compared to EG, RG requires only a single call to F and a single projection in each iteration.

References

[1] Martin Arjovsky, Soumith Chintala, and Léon 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.

[2] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. {CMS} {Books} in {Mathematics}. Springer, 2011. URL https://link.springer.com/book/10.1007/978-1-4419-9467-7.

[3] Heinz H Bauschke, Walaa M Moursi, and Xianfu Wang. Generalized monotone operators and their averaged resolvents. *Mathematical Programming*, 189(1):55–74, 2021.

[4] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *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.

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

[6] Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. Accelerated Algorithms for Monotone Inclusions and Constrained Nonconvex-Nonconcave Min-Max Optimization. June 2022. doi: 10.48550/arXiv.2206.05248. URL http://arxiv.org/abs/2206.05248. arXiv:2206.05248 [cs, math].

[7] Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. Tight Last-Iterate Convergence of the Extragradient and the Optimistic Gradient Descent-Ascent Algorithm for Constrained Monotone Variational Inequalities. May 2022. doi: 10.48550/arXiv.2204.09228. URL http://arxiv.org/abs/2204.09228. arXiv:2204.09228 [cs, math].

[8] Antonin Chambolle and Thomas Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1): 120–145, May 2011. ISSN 1573-7683. doi: 10.1007/s10851-010-0251-1. URL https://doi.org/10.1007/s10851-010-0251-1.

[9] Shisheng Cui and Uday V Shanbhag. On the analysis of reflected gradient and splitting methods for monotone stochastic variational inequality problems. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 4510–4515. IEEE, 2016.

[10] Bo Dai, Albert Shaw, Lihong Li, Lin Xiao, Niao He, Zhen Liu, Jianshu Chen, and Le Song. Sbeed: Convergent reinforcement learning with nonlinear function approximation. In *International Conference on Machine Learning*, pages 1125–1134. PMLR, 2018.

[11] Cong D Dang and Guanghui 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.

[12] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training gans with optimism. *arXiv preprint arXiv:1711.00141*, 2017.

[13] Constantinos Daskalakis, Stratis Skoulakis, and Manolis 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.

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

[15] Jelena Diakonikolas, Constantinos Daskalakis, and Michael Jordan. Efficient methods for structured nonconvex-nonconcave min-max optimization. *International Conference on Artificial Intelligence and Statistics*, 2021.

[16] Simon S. Du, Jianshu Chen, Lihong Li, Lin Xiao, and Dengyong Zhou. Stochastic variance reduction methods for policy evaluation. In *International Conference on Machine Learning*, pages 1049–1058. PMLR, 2017.

[17] Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003. URL https://link.springer.com/book/10.1007/b97544.

[18] Tanner Fiez and Lillian J Ratliff. Local convergence analysis of gradient descent ascent with finite timescale separation. In *Proceedings of the International Conference on Learning Representation*, 2021.

[19] Noah Golowich, Sarath Pattathil, and Constantinos Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. *Advances in neural information processing systems*, 33:20766–20778, 2020.

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

[21] Eduard Gorbunov, Nicolas Loizou, and Gauthier Gidel. Extragradient method: O (1/k) last-iterate convergence for monotone variational inequalities and connections with cocoercivity. In *International Conference on Artificial Intelligence and Statistics*, pages 366–402. PMLR, 2022.

[22] Eduard Gorbunov, Adrien Taylor, and Gauthier Gidel. Last-iterate convergence of optimistic gradient method for monotone variational inequalities. *arXiv preprint arXiv:2205.08446*, 2022.

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

[24] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[25] Michael D Hirsch, Christos H Papadimitriou, and Stephen A Vavasis. Exponential lower bounds for finding Brouwer fix points. *Journal of Complexity*, 5(4):379–416, December 1989. ISSN 0885-064X. doi: 10.1016/0885-064X(89)90017-4. URL https://www.sciencedirect.com/science/article/pii/0885064X89900174.

[26] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. On the convergence of single-call stochastic extra-gradient methods. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/hash/4625d8e31dad7d1c4c83399a6eb62f0c-Abstract.html.

[27] Chi Jin, Praneeth Netrapalli, and Michael Jordan. What is local optimality in nonconvex-nonconcave minimax optimization? In *International conference on machine learning*, pages 4880–4889. PMLR, 2020.

[28] Donghwan Kim. 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.

[29] 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/.

[30] Sucheol Lee and Donghwan Kim. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. In *Annual Conference on Neural Information Processing Systems*, 2021.

[31] Mingrui Liu, Youssef Mroueh, Jerret Ross, Wei Zhang, Xiaodong Cui, Payel Das, and Tianbao Yang. Towards better understanding of adaptive gradient algorithms in generative adversarial nets. In *International Conference on Learning Representations*, 2019.

[32] Mingrui Liu, Hassan Rafique, Qihang Lin, and Tianbao Yang. First-order convergence theory for weakly-convex-weakly-concave min-max problems. *Journal of Machine Learning Research*, 22(169):1–34, 2021.

[33] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

[34] Yu Malitsky. Projected Reflected Gradient Methods for Monotone Variational Inequalities. *SIAM Journal on Optimization*, 25(1):502–520, January 2015. ISSN 1052-6234, 1095-7189. doi: 10.1137/14097238X. URL http://arxiv.org/abs/1502.04968. arXiv: 1502.04968.

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

[36] Yura Malitsky and Matthew K. Tam. A Forward-Backward Splitting Method for Monotone Inclusions Without Cocoercivity. *SIAM Journal on Optimization*, 30(2):1451–1472, January 2020. ISSN 1052-6234. doi: 10.1137/18M1207260. URL https://epubs.siam.org/doi/abs/10.1137/18M1207260. Publisher: Society for Industrial and Applied Mathematics.

[37] Eric V Mazumdar, Michael I Jordan, and S Shankar Sastry. On finding local nash equilibria (and only local nash equilibria) in zero-sum games. *arXiv preprint arXiv:1901.00838*, 2019.

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

[39] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. Convergence Rate of $\mathcal{O}(1/k)$ for Optimistic Gradient and Extra-gradient Methods in Smooth Convex-Concave Saddle Point Problems. *arXiv:1906.01115 [cs, math, stat]*, September 2020. URL http://arxiv.org/abs/1906.01115. arXiv: 1906.01115.

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

[41] Yurii Nesterov. 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.

[42] Maher Nouiehed, Maziar Sanjabi, Tianjian Huang, Jason D Lee, and Meisam Razaviyayn. Solving a class of non-convex min-max games using iterative first order methods. *Advances in Neural Information Processing Systems*, 32, 2019.

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

[44] Christos H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and System Sciences*, 48(3):498–532, June

1994. ISSN 0022-0000. doi: 10.1016/S0022-0000(05)80063-7. URL https://www.sciencedirect.com/science/article/pii/S0022000005800637.

[45] Wei Peng, Yu-Hong Dai, Hui Zhang, and Lizhi Cheng. Training GANs with centripetal acceleration. *Optimization Methods and Software*, 35(5):955–973, 2020. Publisher: Taylor & Francis.

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

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

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

[49] Ernest K. Ryu and Wotao Yin. *Large-Scale Convex Optimization via Monotone Operators*. To be published with Cambridge University Press, 2022.

[50] Chaobing Song, Zhengyuan Zhou, Yichao Zhou, Yong Jiang, and Yi Ma. Optimistic dual extrapolation for coherent non-monotone variational inequalities. *Advances in Neural Information Processing Systems*, 33:14303–14314, 2020.

[51] Quoc Tran-Dinh. The Connection Between Nesterov's Accelerated Methods and Halpern Fixed-Point Iterations, March 2022. URL http://arxiv.org/abs/2203.04869. arXiv:2203.04869 [math].

[52] Quoc Tran-Dinh and Yang Luo. Halpern-Type Accelerated and Splitting Algorithms For Monotone Inclusions. Technical Report arXiv:2110.08150, arXiv, December 2021. URL http://arxiv.org/abs/2110.08150. arXiv:2110.08150 [math, stat] type: article.

[53] Junchi Yang, Negar Kiyavash, and Niao He. Global convergence and variance reduction for a class of nonconvex-nonconcave minimax problems. *Advances in Neural Information Processing Systems*, 33:1153–1165, 2020.

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

[55] TaeHo Yoon and Ernest K Ryu. Accelerated minimax algorithms flock together. *arXiv preprint arXiv:2205.11093*, 2022.

[56] Zhengyuan Zhou, Panayotis Mertikopoulos, Nicholas Bambos, Stephen Boyd, and Peter W Glynn. Stochastic mirror descent in variationally coherent optimization problems. *Advances in Neural Information Processing Systems*, 30, 2017.

Contents

Appendix A. Additional Related Works

A.1. Convex-Concave and Monotone Setting

In the convex-concave setting, a weak convergence measure is the *gap* function (Definition 8). It is well-known that classic extragradient-type methods such as EG and PEG have $O(\frac{1}{T})$ average-iterate convergence rate in terms of gap function [26, 39–41] and the rate is optimal [43]. But the gap function or average-iterate convergence is not meaningful in the nonconvex-nonconcave setting. For convergence in terms of the residual in the constrained setting, EG and PEG has a slower rate of $O(\frac{1}{\sqrt{T}})$ for best-iterate convergence [17, 26, 29, 47] and the more desirable last-iterate convergence [7, 22]. We remark that the last-iterate convergence rate of the reflected gradient (RG) algorithm was unknown. The $O(\frac{1}{\sqrt{T}})$ rate is tight for p-SCIL algorithms [19], a subclass of first-order methods that includes EG, PEG, and many of its variations, but faster rate is possible for other first-order methods.

Accelerated Convergence Rate in Residual. Recent results with accelerated convergence rates in terms of the residual are based on Halpern iteration [23] (also called *Anchoring*). The vanilla Halpern iteration has $O(\frac{1}{T})$ convergence rate for cocoercive operators (stronger than monotonicity) [14, 28]. Recently, a line of works contributed to provide $O(\frac{1}{T})$ convergence rate for monotone operators in the constrained setting. [14, 55] provide double-loop algorithms with $O(\frac{\log T}{T})$ convergence rate for monotone operators in the constrained setting. In the unconstrained setting ($A = 0$), [54] propose the Extra Anchored Gradient (EAG) algorithm, the first efficient algorithm with $O(\frac{1}{T})$ convergence rate for monotone operators. They also establish matching lower bound for first-order methods. [30] generalize EAG to Fast Extragradient (FEG), which works even for negatively comonotone operators but still in the unconstrained setting. Analysis for variants of EAG and FEG in the unconstrained setting is provided in [51, 52]. Recently, [6] close the open problem by proving the projected version of EAG has $O(\frac{1}{T})$ convergence rate. They also propose the accelerated forward-backward splitting (AS) algorithm, a generalization of FEG, which has $O(\frac{1}{T})$ convergence rate for negatively comonotone operators in the constrained setting.

A.2. Nonconvex-Nonconcave Setting

This paper study structured nonconvex-nonconcave optimization problems from the general perspective of operator theory and focus on global convergence under weak MVI and negative comonotonicity. There is a line of works focusing on *local convergence*, e.g., [18, 24, 27, 37]. Another line of works focus on problems satisfying different structural assumptions, such as the Polyak Łojasiewicz condition [42, 53].

Appendix B. Additional Preliminary

B.1. Normal Cone and Resolvent

Normal Cone. We denote $N_{\mathcal{Z}} : \mathcal{Z} \to \mathbb{R}^n$ to be the normal cone operator such that for $z \in \mathcal{Z}$, $N_{\mathcal{Z}}(z) = \{a : \langle a, z' - z\rangle \le 0, \forall z' \in \mathcal{Z}\}$. Define the indicator function

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

It is not hard to see that the subdifferential operator $\partial \mathbb{I}_{\mathcal{Z}} = N_{\mathcal{Z}}$. A useful fact is that if $z = \Pi_{\mathcal{Z}}[z']$, then $\lambda(z' - z) \in N_{\mathcal{Z}}(z)$ for any $\lambda \geq 0$.

Some useful properties of the resolvent are summarized in the following proposition.

Proposition 7 *If A is maximally monotone, then J_A satisfies the following.*

1. *The domain of J_A is $\mathbb{R}^n$. J_A is non-expansive and single-valued on $\mathbb{R}^n$.*
2. *If $z = J_A(z')$, then $z' - z \in A(z)$. If $c \in A(z)$, then $z = J_A(z + c)$.*
3. *When $A = N_{\mathcal{Z}}$ is the normal cone operator for some closed convex set $\mathcal{Z}$, then $J_{\eta A} = \Pi_{\mathcal{Z}}$ is the Euclidean projection onto $\mathcal{Z}$ for all $\eta > 0$.*

B.2. Gap Function

A standard suboptimality measure for the variationaly inequalitt (VI) problem is the *gap function* defined as $\textsc{Gap}_{\mathcal{Z},F}(z) := \max_{z' \in \mathcal{Z}} \langle F(z), z - z' \rangle$. Note that when the feasible set $\mathcal{Z}$ is unbounded, approximating the gap function is impossible: consider the simple unconstrained saddle point problem $\min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} xy$, which has a unique saddle point $(0, 0)$ but any other point has an infinitely large gap. A refined notion is the following restricted gap function [41], which is meaningful for unbounded $\mathcal{Z}$.

Definition 8 (Restricted Gap Function) *Given a closed convex set $\mathcal{Z}$, a single-valued operator F, and a radius D, the restricted gap function at point $z \in \mathcal{Z}$ is*

$$\textsc{Gap}_{\mathcal{Z},F,D} := \max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \left\langle F(z), z - z' \right\rangle$$

where $\mathcal{B}(z, D)$ is a Euclidean ball centered at z with radius D.

In the rest of the paper, we call $\textsc{Gap}_{\mathcal{Z},F,D}$ the gap function (or gap) for convenience. The following Lemma relates $\|F(z) + c\|$ where $c \in N_{\mathcal{Z}}(z)$, and the gap function.

Lemma 9 *Let $\mathcal{Z}$ be a closed convex set $\mathcal{Z}$ and F be a monotone and L-Lipschitz operator. For any $z \in \mathcal{Z}$ and $c \in N_{\mathcal{Z}}(z)$, we have*

$$\textsc{Gap}_{\mathcal{Z},F,D}(z) := \max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \left\langle F(z), z - z' \right\rangle \leq D \cdot \|F(z) + c\|$$

Proof The proof is straightforward. Since $c \in N_{\mathcal{Z}}(z)$, we have $\langle c, z - z' \rangle \geq 0$ for any $z' \in \mathcal{Z}$. Therefor,

$$\begin{aligned}
\max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \left\langle F(z), z - z' \right\rangle &\leq \max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \left\langle F(z) + c, z - z' \right\rangle \\
&\leq \max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \left\| z - z' \right\| \cdot \|F(z) + c\| \quad \text{(Cauchy-Schwarz inequality)} \\
&\leq D \cdot \|F(z) + c\|.
\end{aligned}$$

■

B.3. Classical Algorithms for Variationaly Inequalities

The Extragradient Algorithm [29]. Starting at initial point $z_0 \in \mathcal{Z}$, the update rule of EG is: for $t = 0, 1, 2, \cdots$

$$\begin{aligned} z_{t+\frac{1}{2}} &= \Pi_{\mathcal{Z}}[z_t - \eta F(z_t)] \\ z_{t+1} &= \Pi_{\mathcal{Z}}\left[z_t - \eta F(z_{t+\frac{1}{2}})\right] \end{aligned} \tag{EG}$$

At each step $t \geq 0$, the EG algorithm makes an oracle call of $F(z_t)$ to produce an intermediate point $z_{t+\frac{1}{2}}$ (a gradient descent step if $F = \partial f$ is the gradient of some function f), then the algorithm makes another oracle call $F(z_{t+\frac{1}{2}})$ and updates z_t to z_{t+1}. In each step, EG needs two oracle calls to F and two projections $\Pi_{\mathcal{Z}}$.

The Past Extragradient Algorithm [47] Starting at initial point $z_0 = z_{-\frac{1}{2}} \in \mathcal{Z}$, the update rule of PEG with step size $\eta > 0$ is: for $t = 0, 1, 2, \cdots$

$$\begin{aligned} z_{t+\frac{1}{2}} &= \Pi_{\mathcal{Z}}\left[z_t - \eta F(z_{t-\frac{1}{2}})\right] \\ z_{t+1} &= \Pi_{\mathcal{Z}}\left[z_t - \eta F(z_{t+\frac{1}{2}})\right] \end{aligned} \tag{PEG}$$

Note that PEG is also known as the *Optimistic Gradient Descent/Ascent (OGDA)* algorithm in the literature. The update rule of PEG is similar to (EG) but only requires a single call to F in each iteration. Both of EG and PEG perform two projections in every iteration.

B.4. Proof of Fact 1

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

$$\begin{aligned} r^{nat}_{F,A}(z) &= \|z - J_A(z - F(z))\| \\ &= \|J_A(z + c) - J_A(z - F(z))\| \\ &\leq \|F(z) + c\| && (J_A \text{ is non-expansive}) \end{aligned}$$

and

$$\begin{aligned} r^{fb}_{F,A,\alpha}(z) &= \frac{1}{\alpha}\|z - J_{\alpha A}(z - \alpha F(z))\| \\ &= \frac{1}{\alpha}\|J_{\alpha A}(z + \alpha c) - J_{\alpha A}(z - \alpha F(z))\| \\ &\leq \|F(z) + c\| && (J_A \text{ is non-expansive}) \end{aligned}$$

Thus both $r^{tan}_{F,A}(z)$ and $r^{fb}_{F,A,\alpha}(z)$ are smaller than $r^{tan}_{F,A}(z) = \min_{c \in A(z)} \|F(z) + c\|$. ■

Appendix C. Missing Proofs in Section 3

We present the proof of Theorem 4 here.

Proof From the update rule of (OG), we have the following identity (see also [26, Appendix B]): for any $p \in \mathcal{Z}$,

$$\|z_{t+1} - p\|^2 = \|z_t - p\|^2 + \left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2 - \left\|z_{t+\frac{1}{2}} - z_t\right\|^2 + 2\left\langle z_t - \eta F(z_{t-\frac{1}{2}}) - z_{t+\frac{1}{2}} + \eta F(z_{t+\frac{1}{2}}), p - z_{t+\frac{1}{2}}\right\rangle \tag{2}$$

Since $z_{t+\frac{1}{2}} = J_{\eta A}[z_t - \eta F(z_{t-\frac{1}{2}})]$, we have $\frac{z_t - \eta F(z_{t-\frac{1}{2}}) - z_{t+\frac{1}{2}}}{\eta} \in A(z_{t+\frac{1}{2}})$ by Proposition 7. Then

$$\frac{z_t - z_{t+1}}{\eta} = \frac{z_t - \eta F(z_{t-\frac{1}{2}}) - z_{t+\frac{1}{2}}}{\eta} + F(z_{t+\frac{1}{2}}) \in F(z_{t+\frac{1}{2}}) + A(z_{t+\frac{1}{2}}).$$

Set $p = z^*$. By the weak MVI assumption, we have

$$\begin{aligned} 2\left\langle z_t - \eta F(z_{t-\frac{1}{2}}) - z_{t+\frac{1}{2}} + \eta F(z_{t+\frac{1}{2}}), z^* - z_{t+\frac{1}{2}}\right\rangle &= 2\eta\left\langle \frac{z_t - z_{t+1}}{\eta}, z^* - z_{t+\frac{1}{2}}\right\rangle \\ &\le -\frac{2\rho}{\eta}\|z_t - z_{t+1}\|^2 \end{aligned} \tag{3}$$

Define $c = \frac{1}{2} - 2\eta^2 L^2 > 0$. We have identity

$$(1-2c)\eta^2 L^2 = 4\eta^4 L^4 = \frac{1}{2} - c - (1+2c)\eta^2 L^2 \tag{4}$$

Combining Equation (2) and (3) and using $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$, we have

$$\begin{aligned} &\|z_{t+1} - z^*\|^2 \\ &\le \|z_t - z^*\|^2 + \left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2 - \left\|z_{t+\frac{1}{2}} - z_t\right\|^2 + c\|z_t - z_{t+1}\|^2 - (c + \frac{2\rho}{\eta})\|z_t - z_{t+1}\|^2 \\ &\le \|z_t - z^*\|^2 + (1+2c)\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2 - (1-2c)\left\|z_{t+\frac{1}{2}} - z_t\right\|^2 - (c + \frac{2\rho}{\eta})\|z_t - z_{t+1}\|^2 \end{aligned} \tag{5}$$

Using the update rule of OG and L-Lipschitzness of F, we have that for any $t \ge 0$,

$$\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2 = \left\|\eta F(z_{t-\frac{1}{2}}) - \eta F(z_{t+\frac{1}{2}})\right\|^2 \le \eta^2 L^2 \left\|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\right\|^2 \tag{6}$$

Moreover, using $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$ and Equation (6) , we have that for any $t \ge 1$,

$$\left\|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\right\|^2 \le 2\left\|z_{t+\frac{1}{2}} - z_t\right\|^2 + 2\left\|z_t - z_{t-\frac{1}{2}}\right\|^2 \le 2\left\|z_{t+\frac{1}{2}} - z_t\right\|^2 + 2\eta^2 L^2 \left\|z_{t-\frac{1}{2}} - z_{t-\frac{3}{2}}\right\|^2$$

which imples

$$\left\|z_{t+\frac{1}{2}} - z_t\right\|^2 \ge \frac{1}{2}\left\|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\right\|^2 - \eta^2 L^2 \left\|z_{t-\frac{1}{2}} - z_{t-\frac{3}{2}}\right\|^2. \tag{7}$$

Combining Equation (4), (5), (6), and (7), we have that for all $t \ge 1$.

$$
\begin{aligned}
&\|z_{t+1} - z^*\|^2 \\
&\le \|z_t - z^*\|^2 + (1+2c)\left\|z_{t+1} - z_{t+\frac{1}{2}}\right\|^2 - (1-2c)\left\|z_{t+\frac{1}{2}} - z_t\right\|^2 - (c + \frac{2\rho}{\eta})\|z_t - z_{t+1}\|^2 \\
&\le \|z_t - z^*\|^2 + (1-2c)\eta^2 L^2 \left\|z_{t-\frac{1}{2}} - z_{t-\frac{3}{2}}\right\|^2 - \left(\frac{1}{2} - c - (1+2c)\eta^2 L^2\right)\left\|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\right\|^2 \\
&\quad - (c + \frac{2\rho}{\eta})\|z_t - z_{t+1}\|^2 \\
&= \|z_t - z^*\|^2 + 4\eta^4 L^4 \left(\left\|z_{t-\frac{1}{2}} - z_{t-\frac{3}{2}}\right\|^2 - \left\|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\right\|^2\right) - (c + \frac{2\rho}{\eta})\|z_t - z_{t+1}\|^2
\end{aligned}
$$

Telescoping the above inequality and using $c = \frac{1}{2} - 2\eta^2 L^2$ and $\eta L < \frac{1}{2}$, we get

$$
(\frac{1}{2} + \frac{2\rho}{\beta} - 2\eta^2 L^2) \sum_{t=1}^{T} \|z_t - z_{t+1}\|^2 \le \|z_1 - z^*\|^2 + \frac{1}{4}\left\|z_{\frac{1}{2}} - z_{-\frac{1}{2}}\right\|^2.
$$

Note that z_0 is the same as $z_{-\frac{1}{2}}$. This completes the proof. ■

Fact 2 *For any $L > 0$ and $\rho > -\frac{1}{12\sqrt{3}L}$. There exists $\eta \in (0, \frac{1}{2L})$ such that*

$$
\frac{1}{2} + \frac{2\rho}{\eta} - 2\eta^2 L^2 > 0 \tag{8}
$$

Proof *Let $\eta = \frac{1}{2\sqrt{3}L}$, then Inequality (8) holds whenever*

$$
\rho > \frac{\eta L(1 - 4\eta^2 L^2)}{4} \cdot \frac{1}{L} = -\frac{1}{12\sqrt{3}L}.
$$

■

Appendix D. Missing Proofs in Section 4

To prove Theorem 5, we apply a potential function argument. We first show the potential function is approximately non-increasing and then prove that it is upper bounded by a term independent of T. As the potential function at step t is also at least $\Omega(t^2) \cdot r^{tan}(z_t)^2$, we conclude that ARG has an $O(\frac{1}{T})$ convergence rate .

D.1. Potential Function

Recall the update rule of ARG: $z_0 = z_{\frac{1}{2}} \in \mathbb{R}^n$ are initial points and $z_1 = J_{\eta A}[z_0 - \eta F(z_0)]$; for $t \ge 1$,

$$
\begin{aligned}
z_{t+\frac{1}{2}} &= 2z_t - z_{t-1} + \frac{1}{t+1}(z_0 - z_t) - \frac{1}{t}(z_0 - z_{t-1}) \\
z_{t+1} &= J_{\eta A}\left[z_t - \eta F(z_{t+\frac{1}{2}}) + \frac{1}{t+1}(z_0 - z_t)\right]
\end{aligned}
\tag{ARG}
$$

Recall that when A is the normal cone of a closed convex set $\mathcal{Z}$, the resolvent J_A is equivalent to Euclidean projection to set $\mathcal{Z}$. Hence, if we apply the ARG algorithm to solve monotone VI problems, the algorithm uses a single call to operator F and a single projection to $\mathcal{Z}$ per iteration. Here we allow A to be an arbitrary maximally monotone operator, and the ARG algorithm becomes a single-call single-resolvent algorithm in this more general setting.

Next, we specify the potential function. Define

$$c_{t+1} := \frac{z_t - \eta F(z_{t+\frac{1}{2}}) + \frac{1}{t+1}(z_0 - z_t) - z_{t+1}}{\eta}, \quad \forall t \geq 0. \tag{9}$$

By update rule we have $c_t \in A(z_t)$ for all $t \geq 1$. The potential function we track in iterate $t \geq 1$ is defined as

$$V_t := \frac{t(t+1)}{2} \|\eta F(z_t) + \eta c_t\|^2 + \frac{t(t+1)}{2} \left\|\eta F(z_t) - \eta F(z_{t-\frac{1}{2}})\right\|^2 + t\langle \eta F(z_t) + \eta c_t, z_t - z_0 \rangle \tag{10}$$

D.2. Approximately Non-Increasing Potential

Fact 3 *For any $L > 0$ and $\rho \geq -\frac{1}{60L}$. There exists $\eta > 0$ such that*

$$\frac{1}{2} - (12 - \frac{4\rho}{\eta})\eta^2 L^2 + \frac{2\rho}{\eta} \geq 0. \tag{11}$$

Moreover, every $\eta > 0$ satisfies (11) also satisfies $\frac{\rho}{\eta} \geq -\frac{1}{4}$.

Proof Rewriting (11), we get

$$\rho > \frac{\eta L(24\eta^2 L^2 - 1)}{4 + 8\eta^2 L^2} \cdot \frac{1}{L}.$$

Let $x = \eta L$ and $f(x) = \frac{x(24x^2-1)}{4+8x^2}$. Since $f(\frac{1}{12}) = -\frac{5}{292} < -\frac{1}{60}$. We know $\eta = \frac{1}{12L}$ satisfies (11).

Moreover, rewritng (11) and using $\eta L > 0$, we get

$$\frac{\rho}{\eta} \geq -\frac{1 - 72\eta^2 L^2}{4 + 8\eta^2 L^2} \geq -\frac{1}{4}.$$

■

We show in the following lemma that V_t is approximately non-increasing.

Lemma 10 *In the same setup as Theorem 5, for any $t \geq 1$, we have*

$$V_{t+1} \leq V_t + \frac{1}{8} \cdot \|\eta F(z_{t+1}) + \eta c_{t+1}\|^2.$$

Proof The plan is to show that $V_t - V_{t+1}$ plus a few non-positive terms is still $\geq -\frac{1}{8} \cdot \|\eta F(z_{t+1}) + \eta c_{t+1}\|^2$, which certifies the claim.

Two Positive Terms. Since $F + A$ is ρ-comonotone, we have

$$\langle \eta F(z_{t+1}) + \eta c_{t+1} - \eta F(z_t) - \eta c_t, z_{t+1} - z_t \rangle - \frac{\rho}{\eta} \|\eta F(z_{t+1}) + \eta c_{t+1} - \eta F(z_t) - \eta c_t\|^2 \geq 0. \tag{12}$$

Since F is L-Lipschitz, we have

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

Denote $p = \frac{1}{24}$. Multiplying the above inequality with $1 - \frac{\rho}{3\eta} > 0$ and rearranging terms, we get

$$\begin{aligned} & p \cdot \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| \eta F(z_{t+1}) - \eta F(z_{t+\frac{1}{2}}) \right\|^2 \\ & + \left((1 - \frac{\rho}{3\eta}) \eta^2 L^2 - p \right) \cdot \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 + \frac{\rho}{3\eta} \left\| \eta F(z_{t+1}) - \eta F(z_{t+\frac{1}{2}}) \right\|^2 \geq 0. \end{aligned} \tag{13}$$

Sum-of-Squares Identity. We show an equivalent formulation $z_{t+\frac{1}{2}}$ and z_{t+1} using definitions of $\eta c_t = z_{t-1} - z_t - \eta F(z_{t-\frac{1}{2}}) + \frac{1}{t}(z_0 - z_{t-1})$ and $\eta c_{t+1} = z_t - \eta F(z_{t+\frac{1}{2}}) + \frac{1}{t+1}(z_0 - z_t) - z_{t+1}$:

$$\begin{aligned} z_{t+\frac{1}{2}} &= 2z_t - z_{t-1} + \frac{1}{t+1}(z_0 - z_t) - \frac{1}{t}(z_0 - z_{t-1}) \\ &= z_t + (z_t - z_{t-1}) + \frac{1}{t+1}(z_0 - z_t) - \frac{1}{t}(z_0 - z_{t-1}) \\ &= z_t - \eta F(z_{t-\frac{1}{2}}) - \eta c_t + \frac{1}{t+1}(z_0 - z_t), \\ z_{t+1} &= z_t - \eta F(z_{t+\frac{1}{2}}) - \eta c_{t+1} + \frac{1}{t+1}(z_0 - z_t). \end{aligned}$$

We also have

$$z_{t+1} - z_{t+\frac{1}{2}} = \eta F(z_{t-\frac{1}{2}}) + \eta c_t - \eta F(z_{t+\frac{1}{2}}) - \eta c_{t+1}. \tag{14}$$

Next, we simplify

$$V_t - V_{t+1} - t(t+1) \times \text{LHS of Inequality (12)} - \frac{t(t+1)}{4p} \times \text{LHS of Inequality (13)}$$

using the second identity in Proposition 16: replace x_0 with z_0; for $k \in [4]$, replace x_k with $z_{t-1+\frac{k}{2}}$ and replace y_k with $\eta F(z_{t-1+\frac{k}{2}})$; replace u_2 with ηc_t; replace u_4 with ηc_{t+1}; replace k with t; replace p with q. Note that $x_3 = x_2 - y_1 - u_2 + \frac{1}{k+1}(x_0 - x_2)$ and $x_4 = x_2 - y_3 - u_4 + + \frac{1}{k+1}(x_0 - x_2)$ hold due to the above equivalent formations of $z_{t+\frac{1}{2}}$ and z_{t+1}. Expression (18) and (19) appear on both sides of the following equation.

$$
\begin{aligned}
& V_t - V_{t+1} - t(t+1) \times \text{LHS of Inequality (12)} - \frac{t(t+1)}{4p} \times \text{LHS of Inequality (13)} \\
&= \frac{t(t+1)}{4} \left\| \eta c_{t+1} - \eta c_t + \eta F(z_{t-\frac{1}{2}}) - 2\eta F(z_t) + \eta F(z_{t+\frac{1}{2}}) \right\|^2 && (15) \\
&\quad + \left(\frac{(1-4p)t - 4p}{4p}(t+1) \right) \cdot \left\| \eta F(z_{t+\frac{1}{2}}) - \eta F(z_{t+1}) \right\|^2 && (16) \\
&\quad + (t+1) \cdot \left\langle \eta F(z_{t+\frac{1}{2}}) - \eta F(z_{t+1}), \eta F(z_{t+1}) + \eta c_{t+1} \right\rangle && (17) \\
&\quad + t(t+1)\frac{\rho}{\eta} \cdot \| \eta F(z_{t+1}) + \eta c_{t+1} - \eta F(z_t) - \eta c_t \|^2 && (18) \\
&\quad - \frac{t(t+1)}{4p} \cdot \left(\left((1 - \frac{\rho}{3\eta})\eta^2 L^2 - p \right) \cdot \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 + \frac{\rho}{3\eta} \left\| \eta F(z_{t+1}) - \eta F(z_{t+\frac{1}{2}}) \right\|^2 \right). && (19)
\end{aligned}
$$

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

Expression (16) + Expression (17)

$$
\begin{aligned}
&= \left\| \sqrt{\frac{(1-4p)t-4p}{4p}(t+1)} \cdot \left(\eta F(z_{t+\frac{1}{2}}) - \eta F(z_{t+1}) \right) + \sqrt{\frac{p(t+1)}{(1-4p)t-4p}} \cdot (\eta F(z_{t+1}) + \eta c_{t+1}) \right\|^2 \\
&\quad - \frac{p(t+1)}{(1-4p)t-4p} \cdot \| \eta F(z_{t+1}) + \eta c_{t+1} \|^2 \\
&\geq -\frac{p(t+1)}{(1-8p)t} \cdot \| \eta F(z_{t+1}) + \eta c_{t+1} \|^2 && (t \geq 1) \\
&\geq -\frac{2p}{1-8p} \cdot \| \eta F(z_{t+1}) + \eta c_{t+1} \|^2 && (\tfrac{t+1}{t} \leq 2) \\
&= -\frac{1}{8} \| \eta F(z_{t+1}) + \eta c_{t+1} \|^2. && (p = \tfrac{1}{24})
\end{aligned}
$$

Now it remains to show that the sum of Expression (15), (18), and (19) is non-negative. Multiplying $\frac{4}{t(t+1)}$ and replacing $p = \frac{1}{24}$, we get

$$
\begin{aligned}
& \frac{4}{t(t+1)} \cdot (\text{Expression (15)} + \text{Expression (18)} + \text{Expression (19)}) \\
&= \left\| \eta c_{t+1} - \eta c_t + \eta F(z_{t-\frac{1}{2}}) - 2\eta F(z_t) + \eta F(z_{t+\frac{1}{2}}) \right\|^2 + \left(1 - (24 - \frac{8\rho}{\eta})\eta^2 L^2 \right) \cdot \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 \\
&\quad + \frac{4\rho}{\eta} \cdot \| \eta F(z_{t+1}) + \eta c_{t+1} - \eta F(z_t) - \eta c_t \|^2 - \frac{8\rho}{\eta} \left\| \eta F(z_{t+1}) - \eta F(z_{t+\frac{1}{2}}) \right\|^2.
\end{aligned}
$$

Denote

$$
\begin{aligned}
B_1 &= \eta c_{t+1} - \eta c_t + \eta F(z_{t-\frac{1}{2}}) - 2\eta F(z_t) + \eta F(z_{t+\frac{1}{2}}) \\
B_2 &= z_{t+1} - z_{t+\frac{1}{2}} = \eta F(z_{t-\frac{1}{2}}) + \eta c_t - \eta F(z_{t+\frac{1}{2}}) - \eta c_{t+1} && \text{(By (14))} \\
B_3 &= \eta F(z_{t+1}) + \eta c_{t+1} - \eta F(z_t) - \eta c_t \\
B_4 &= \eta F(z_{t+1}) - \eta F(z_{t+\frac{1}{2}}).
\end{aligned}
$$

It is not hard to check that $B_1 - B_2 = 2(B_3 - B_4)$:

$$B_1 - B_2 = 2\eta c_{t+1} - 2\eta c_t - 2\eta F(z_t) + 2\eta F(z_{t+\frac{1}{2}}) = 2(B_3 - B_4).$$

Note that ρ is non-positive and we have

$$\begin{aligned}
&\frac{4}{t(t+1)} \cdot (\text{Expression (15)} + \text{Expression (18)} + \text{Expression (19)}) \\
&= \|B_1\|^2 + \left(1 - (24 - \frac{8\rho}{\eta})\eta^2 L^2\right) \cdot \|B_2\|^2 + \frac{\rho}{\eta} \cdot \|2B_3\|^2 - \frac{2\rho}{\eta}\|2B_4\|^2 \\
&\geq \left(\frac{1}{2} - (12 - \frac{4\rho}{\eta})\eta^2 L^2\right) \cdot \|B_1 - B_2\|^2 + \frac{\rho}{\eta} \cdot \|2B_3\|^2 - \frac{2\rho}{\eta}\|2B_4\|^2 \\
&\qquad\qquad (\|a\|^2 + \|b\|^2 \geq \tfrac{1}{2}\|a-b\|^2 \text{ and } (24 - \tfrac{8\rho}{\eta})\eta^2 L^2 \geq 0) \\
&\geq \left(\frac{1}{2} - (12 - \frac{4\rho}{\eta})\eta^2 L^2\right) \cdot \|B_1 - B_2\|^2 + \frac{2\rho}{\eta} \cdot \|2B_3 - 2B_4\|^2 \\
&\qquad\qquad (-\|a\|^2 + 2\|b\|^2 \geq -2\|a-b\|^2 \text{ and } -\tfrac{\rho}{\eta} \geq 0) \\
&= \left(\frac{1}{2} - (12 - \frac{4\rho}{\eta})\eta^2 L^2 + \frac{2\rho}{\eta}\right) \cdot \|B_1 - B_2\|^2 \qquad (B_1 - B_2 = 2(B_3 - B_4)) \\
&\geq 0. \qquad\qquad (\text{Inequality (11)})
\end{aligned}$$

The last inequality holds by the choice of η as shown in Fact 3. ■

D.3. Bouding Potential at Iteration 1

Lemma 11 *Let F be a L-Lipschitz operator and A be a maximally monotone operator. For any $z_0 = z_{\frac{1}{2}} \in \mathbb{R}^n$, $\eta \in (0, \frac{1}{2L})$, and $z_1 = J_{\eta A}[z_0 - \eta F(z_0)]$, we have the following*

1. $\|z_1 - z_0\| \leq \eta \cdot r^{tan}_{F,A}(z_0)$.
2. $\|\eta F(z_1) + \eta c_1\| \leq (1 + \eta L)\|z_1 - z_0\|$
3. $V_1 \leq 4\|z_1 - z_0\|^2$ *where V_1 is defined in* (10)

Proof For any $c \in A(z_0)$, due to non-expansiveness of $J_{\eta A}$, we have

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

Thus $\|z_1 - z_0\| \leq \eta \cdot r^{tan}_{F,A}(z_0)$.

By definition of V_1 in (10), we have

$$V_1 = \|\eta F(z_1) + \eta c_1\|^2 + \|\eta F(z_1) - \eta F(z_0)\|^2 + \langle \eta F(z_1) + \eta c_1, z_1 - z_0\rangle.$$

We bound $\|\eta F(z_1) + \eta c_1\|$ first. Note that by definition, we have $\eta c_1 = z_0 - \eta F(z_0) - z_1$. Thus we have

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

Then we can apply the bound on $\|\eta F(z_1) + \eta c_1\|$ to bound V_1 as follows:

$$\begin{aligned} V_1 &= \|\eta F(z_1) + \eta c_1\|^2 + \|\eta F(z_1) - \eta F(z_0)\|^2 + \langle \eta F(z_1) + \eta c_1, z_1 - z_0 \rangle \\ &\le \|\eta F(z_1) + \eta c_1\|^2 + \eta^2 L^2 \|z_1 - z_0\|^2 + \|\eta F(z_1) + \eta c_1\| \|z_1 - z_0\| \\ &\le (1+\eta L)^2 \|z_1 - z_0\|^2 + \eta^2 L^2 \|z_1 - z_0\|^2 + (1+\eta L)\|z_1 - z_0\|^2 \\ &= (2 + 3\eta L + 2\eta^2 L^2)\|z_1 - z_0\|^2 \\ &\le 4\|z_1 - z_0\|^2 \end{aligned}$$

where we use L-Lipschitzness of F and Cauchy-Schwarz inequality in the first inequality; we use $\|\eta F(z_1) + \eta c_1\| \le (1+\eta L)\|z_1 - z_0\|$ in the second inequality; we use $\eta L \le \frac{1}{2}$ in the last inequality. ■

D.4. Proof of Theorem 5

We first show that the potential function $V_t = \Omega(t^2 \cdot r^{tan}(z_t)^2)$.

Lemma 12 *In the same setup as Theorem 5, for any $t \ge 1$, we have*

$$\frac{t(t+\frac{1}{2})}{4}\|\eta F(z_t) + \eta c_t\|^2 \le V_t + \|z^* - z_0\|^2$$

Proof Since $0 \in F(z^*) + A(z^*)$, by ρ-comonotonicity of $F + A$ and Fact 3, we have

$$\langle \eta F(z_t) + \eta c_t, z_t - z^* \rangle \ge \frac{\rho}{\eta}\|\eta F(z_t) + \eta c_t\|^2 \ge -\frac{1}{4}\|\eta F(z_t) + \eta c_t\|^2 \tag{20}$$

By definition of V_t in (10), for any $t \ge 1$, we have

$$\begin{aligned} V_t &= \frac{t(t+1)}{2}\|\eta F(z_t) + \eta c_t\|^2 + \frac{t(t+1)}{2}\left\|\eta F(z_t) - \eta F(z_{t-\frac{1}{2}})\right\|^2 + t\langle \eta F(z_t) + \eta c_t, z_t - z_0 \rangle \\ &\ge \frac{t(t+1)}{2}\|\eta F(z_t) + \eta c_t\|^2 + t\langle \eta F(z_t) + \eta c_t, z_t - z^* \rangle + t\langle \eta F(z_t) + \eta c_t, z^* - z_0 \rangle \\ &\ge \frac{t(t+1)}{2}\|\eta F(z_t) + \eta c_t\|^2 - \frac{1}{4}\|\eta F(z_t) + \eta c_t\|^2 + t\langle \eta F(z_t) + \eta c_t, z^* - z_0 \rangle \\ &\hspace{20em} \text{(By Inequality (20))} \\ &\ge \frac{t(t+\frac{1}{2})}{2}\|\eta F(z_t) + \eta c_t\|^2 - \frac{t(t+\frac{1}{2})}{4}\|\eta F(z_t) + \eta c_t\|^2 - \frac{t}{t+\frac{1}{2}}\|z^* - z_0\|^2 \\ &\ge \frac{t(t+\frac{1}{2})}{4}\|\eta F(z_t) + \eta c_t\|^2 - \|z^* - z_0\|^2 \qquad (\tfrac{t}{t+\frac{1}{2}} < 1) \end{aligned}$$

where in the second last inequality we we apply $\langle a, b \rangle \ge -\frac{\alpha}{4}\|a\|^2 - \frac{1}{\alpha}\|b\|^2$ with $a = \sqrt{t}(\eta F(z_t) + \eta c_t)$, $b = \sqrt{t}(z^* - z_0)$, and $\alpha = t + \frac{1}{2}$. ■

Proof [Proof of Theorem 5] It is equivalent to prove that for every $T \ge 1$, we have

$$\|\eta F(z_T) + \eta c_T\|^2 \le \frac{6H^2}{T^2}.$$

From Lemma 11, we have

$$\|\eta F(z_1) + \eta c_1\|^2 \le (1 + \eta L)^2 \|z_1 - z_0\|^2 \le H^2.$$

So the theorem holds for $T = 1$.

For any $T \ge 2$, by Lemma 12 we have

$$\begin{aligned}\frac{T(T+\frac{1}{2})}{4}\|\eta F(z_T) + \eta c_T\|^2 &\le V_T + \|z_0 - z^*\|^2 \\ &\le V_1 + \|z_0 - z^*\|^2 + \frac{1}{8}\sum_{t=2}^{T}\|\eta F(z_t) + \eta c_t\|^2 \\ &= H^2 + \frac{1}{8}\sum_{t=2}^{T}\|\eta F(z_t) + \eta c_t\|^2\end{aligned}$$

By subtracting $\frac{1}{8}\|\eta F(z_T) + \eta c_T\|^2$ from both sides of the above inequality, we get

$$\frac{T^2}{4}\|\eta F(z_T) + \eta c_T\|^2 \le H^2 + \frac{1}{8}\sum_{t=2}^{T-1}\|\eta F(z_t) + \eta c_t\|^2$$

which is in the form of Proposition 17 with $C_1 = H^2$ and $p = \frac{1}{9}$. Thus we have for any $T \ge 2$

$$\|\eta F(z_T) + \eta c_T\|^2 \le \frac{6H^2}{T^2}.$$

■

Appendix E. Missing Proofs in Section 5

To prove Theorem 6, our analysis is based on a potential function argument and can be summarized in the following three steps. (1) We construct a potential function and show that it is non-increasing between two consecutive iterates; (2) We prove that the RG algorithm has a best-iterate convergence rate, i.e., for any $T \ge 1$, there exists one iterate $t^* \in [T]$ such that our potential function at iterate t^* is small; (3) We combine the above steps to show that the the last iterate has the same convergence guarantee as the best iterate and derive the $O(\frac{1}{\sqrt{T}})$ last-iterate convergence rate.

E.1. Non-increasing Potential

Potential Function. We denote

$$c_{t+1} := \frac{z_t - \eta F(z_{t+\frac{1}{2}}) - z_{t+1}}{\eta}, \ \forall t \ge 0 \tag{21}$$

Note that according to the update rule of RG, $z_{t+1} = \Pi_{\mathcal{Z}}[z_t - \eta F(z_{t+\frac{1}{2}})]$, so $c_{t+1} \in N_{\mathcal{Z}}(z_{t+1})$.

The potential function we adopt is P_t defined as

$$P_t := \|F(z_t) + c_t\|^2 + \left\|F(z_t) - F(z_{t-\frac{1}{2}})\right\|^2, \ \forall t \ge 1. \tag{22}$$

Lemma 13 *In the same setup of Theorem 6, $P_t \ge P_{t+1}$ for any $t \ge 1$.*

Proof The plan is to show that $P_t - P_{t+1}$ plus a few non-positive terms is non-negative, which certifies that $P_t - P_{t+1} \ge 0$.

Three Non-Positive Terms. Since F is monotone, we have

$$(-2) \cdot \langle \eta F(z_{t+1}) - \eta F(z_t), z_{t+1} - z_t \rangle \le 0. \tag{23}$$

Since F is L-Lipschitz and $0 < \eta < \frac{1}{(1+\sqrt{2})L} < \frac{1}{2L}$, we have

$$(-2) \cdot \left(\frac{1}{4} \cdot \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| \eta F(z_{t+1}) - \eta F(z_{t+\frac{1}{2}}) \right\|^2 \right) \le 0. \tag{24}$$

By definition, we have $c_{t+1} \in N_{\mathcal{Z}}(z_{t+1})$ and $c_t \in N_{\mathcal{Z}}(z_t)$. Since the normal cone operator $N_{\mathcal{Z}}$ is maximally monotone, we have

$$(-2) \cdot \langle \eta c_{t+1} - \eta c_t, z_{t+1} - z_t \rangle \le 0. \tag{25}$$

Sum-of-Squares Identity. We use the following equivalent formations of $z_{t+\frac{1}{2}}$ and z_{t+1}.

$$\begin{aligned} z_{t+\frac{1}{2}} &= 2z_t - z_{t-1} = z_t - (z_{t-1} - z_t) = z_t - \eta F(z_{t-\frac{1}{2}}) - \eta c_t \\ z_{t+1} &= \Pi_{\mathcal{Z}}\left[z_t - \eta F(z_{t+\frac{1}{2}}) \right] = z_t - \eta F(z_{t+\frac{1}{2}}) - \eta c_{t+1} \end{aligned}$$

The following identity holds according to Proposition 16. To see this, we replace x_k with $z_{t-1+\frac{k}{2}}$; replace y_k with $\eta F(z_{t-1+\frac{k}{2}})$; replace u_2 with ηc_t; replace u_4 with ηc_{t+1}; also note that $x_3 = x_2 - y_1 - u_2$ and $x_4 = x_2 - y_3 - u_4$ hold due to the above equivalent formations of $z_{t+\frac{1}{2}}$ and z_{t+1}.

$$\begin{aligned} &\eta^2 \cdot (P_t - P_{t+1}) + \text{LHS of Inequality(23)} + \text{LHS of Inequality(24)} + \text{LHS of Inequality(25)} \\ &= \left\| \frac{z_{t+\frac{1}{2}} - z_{t+1}}{2} + \eta F(z_{t-\frac{1}{2}}) - \eta F(z_t) \right\|^2 + \left\| \frac{z_{t+\frac{1}{2}} + z_{t+1}}{2} - z_t + \eta F(z_t) + \eta c_t \right\|^2. \end{aligned}$$

The right-hand side of the above equality is clearly ≥ 0, thus we conclude $P_t - P_{t+1} \ge 0$. ■

E.2. Best-Iterate Convergence

In this section, we show that for any $T \ge 1$, there exists some iterate t^* such that $P_{t^*} = O(\frac{1}{T})$, which is implied by $\sum_{t=1}^T P_t = O(1)$. To prove this, we first show $\sum_{t=1}^T \|z_{t+\frac{1}{2}} - z_t\|^2 = \sum_{t=1}^T \|z_t - z_{t-1}\|^2 = O(1)$ and then relate $\sum_{t=1}^T P_t$ to these two quantities.

Lemma 14 *In the same setup of Theorem 6, for any $T \ge 1$, we have*

$$\sum_{t=1}^T \left\| z_{t+\frac{1}{2}} - z_t \right\|^2 = \sum_{t=1}^T \|z_t - z_{t-1}\|^2 \le \frac{H^2}{1 - (1+\sqrt{2})\eta L}$$

Proof First note that by the update rule of RG, we have $z_{t+\frac{1}{2}} = 2z_t - z_{t-1}$ thus $z_{t+\frac{1}{2}} - z_t = z_t - z_{t-1}$. Therefore, it suffices to only prove the inequality for $\sum_{t=1}^T \|z_{t+\frac{1}{2}} - z_t\|^2$.

From the proof of [26, Lemma 2], for any $t \geq 1$ and $p \in \mathcal{Z}$, we have

$$\Big(1 - (1+\sqrt{2})\eta L\Big) \cdot \Big\|z_{t+\frac{1}{2}} - z_t\Big\|^2 \leq \|z_t - p\|^2 - \|z_{t+1} - p\|^2 - 2\eta\Big\langle F(z_{t+\frac{1}{2}}), z_{t+\frac{1}{2}} - p\Big\rangle + \eta L\bigg(\Big\|z_t - z_{t-\frac{1}{2}}\Big\|^2 - \Big\|z_{t+1} - z_{t+\frac{1}{2}}\Big\|^2\bigg). \tag{26}$$

We set $p = z^*$ to be a solution of the variational inequality (VI) problem in the above inequality. Note that

$$\begin{aligned}
-2\eta\Big\langle F(z_{t+\frac{1}{2}}), z_{t+\frac{1}{2}} - z^*\Big\rangle &= -2\eta\Big\langle F(z_{t+\frac{1}{2}}) - F(z^*), z_{t+\frac{1}{2}} - z^*\Big\rangle - 2\eta\Big\langle F(z^*), z_{t+\frac{1}{2}} - z^*\Big\rangle \\
&\leq -2\eta\Big\langle F(z^*), z_{t+\frac{1}{2}} - z^*\Big\rangle \qquad (F \text{ is monotone}) \\
&= 2\eta\langle F(z^*), z_{t-1} - z^*\rangle - 4\eta\langle F(z^*), z_t - z^*\rangle \qquad (27)
\end{aligned}$$

where the last equality holds since $z_{t+\frac{1}{2}} = 2z_t - z_{t-1}$. Also note that $\langle F(z^*), z_t - z^*\rangle \geq 0$ for all $t \geq 0$ since $z_t \in \mathcal{Z}$ and z^* is a solution to (VI).

Combing Inequality (26) and Inequality (27), telescoping the terms for $t = 1, 2, \cdots, T$, and dividing both sides by $1 - (1+\sqrt{2})\eta L > 0$, we get

$$\sum_{t=1}^{T} \Big\|z_{t+\frac{1}{2}} - z_t\Big\|^2 \leq \frac{\|z_1 - z^*\|^2 + \|z_1 - z_{\frac{1}{2}}\|^2 + 2\eta\langle F(z^*), z_0 - z^*\rangle}{1 - (1+\sqrt{2})\eta L}.$$

To get a cleaner constant that only relies on the starting point $z_0 = z_{\frac{1}{2}}$, we further simplify the three terms on the right-hand side. Note that since $\eta < \frac{1}{2L}$ and $z_1 = \Pi_{\mathcal{Z}}[z_0 - \eta F(z_0)]$.

$$\Big\|z_1 - z_{\frac{1}{2}}\Big\|^2 = \|z_1 - z_0\|^2 \leq \eta^2\|F(z_0)\|^2 \leq \frac{4}{L^2}\|F(z_0)\|^2$$

Thus we have

$$\|z_1 - z^*\|^2 \leq 2\|z_1 - z_0\|^2 + 2\|z_0 - z^*\|^2 \leq \frac{8}{L^2}\|F(z_0)\|^2 + 2\|z_0 - z^*\|^2.$$

Moreover,

$$\begin{aligned}
2\eta\langle F(z^*), z_0 - z^*\rangle &\leq 2\eta\|F(z^*)\|\|z_0 - z^*\| \\
&\leq 4\eta(\|F(z^*) - F(z_0)\| + \|F(z_0)\|)\|z_0 - z^*\| \\
&\leq 4\eta L\|z_0 - z^*\|^2 + 4\eta\|F(z_0)\|\|z_0 - z^*\| \\
&\leq 2\|z_0 - z^*\|^2 + \frac{2}{L}\|F(z_0)\|\|z_0 - z^*\| \qquad (\eta < \tfrac{1}{2L}) \\
&\leq 3\|z_0 - z^*\|^2 + \frac{1}{L^2}\|F(z_0)\|^2 \qquad (2ab \leq a^2 + b^2)
\end{aligned}$$

Thus

$$\|z_1 - z^*\|^2 + \Big\|z_1 - z_{\frac{1}{2}}\Big\|^2 + 2\eta\langle F(z^*), z_0 - z^*\rangle \leq \frac{13}{L^2}\|F(z_0)\|^2 + 5\|z_0 - z^*\|^2 = H^2.$$

This completes the proof. ■

Lemma 15 *In the same setup of Theorem 6, for any $T \geq 1$, we have*

$$\sum_{t=1}^{T} P_t \leq \lambda^2 H^2 L^2.$$

Proof We first show an upper bound for P_t

$$\begin{aligned}
P_t &= \|F(z_t) + c_t\|^2 + \left\|F(z_t) - F(z_{t-\frac{1}{2}})\right\|^2 \\
&= \left\|F(z_t) - F(z_{t-\frac{1}{2}}) + \frac{z_t - z_{t-1}}{\eta}\right\|^2 + \left\|F(z_t) - F(z_{t-\frac{1}{2}})\right\|^2 && \text{(definition of } c_t\text{)} \\
&\leq 3\left\|F(z_t) - F(z_{t-\frac{1}{2}})\right\|^2 + \frac{2}{\eta^2}\|z_t - z_{t-1}\|^2 && (\|A+B\|^2 \leq 2\|A\|^2 + 2\|B\|^2) \\
&\leq 3L^2\left\|z_t - z_{t-\frac{1}{2}}\right\|^2 + \frac{2}{\eta^2}\|z_t - z_{t-1}\|^2 && (F \text{ is } L\text{-Lipschitz}) \\
&= 3L^2\left\|z_t - z_{t-1} + z_{t-1} - z_{t-\frac{1}{2}}\right\|^2 + \frac{2}{\eta^2}\|z_t - z_{t-1}\|^2 \\
&\leq 6L^2\left\|z_{t-\frac{1}{2}} - z_{t-1}\right\|^2 + \left(\frac{2}{\eta^2} + 6L^2\right)\|z_t - z_{t-1}\|^2 && (\|A+B\|^2 \leq 2\|A\|^2 + 2\|B\|^2) \\
&\leq \frac{2 + 6\eta^2 L^2}{\eta^2}\left(\left\|z_{t-\frac{1}{2}} - z_{t-1}\right\|^2 + \|z_t - z_{t-1}\|^2\right).
\end{aligned}$$

Summing the above inequality of $t = 1, 2, \cdots T$, we get

$$\begin{aligned}
\sum_{t=1}^{T} P_t &\leq \frac{2 + 6\eta^2 L^2}{\eta^2} \sum_{t=1}^{T}\left(\left\|z_{t-\frac{1}{2}} - z_{t-1}\right\|^2 + \|z_t - z_{t-1}\|^2\right) \\
&= \frac{2 + 6\eta^2 L^2}{\eta^2}\left(\|z_1 - z_0\|^2 + \sum_{t=1}^{T-1}\left(\left\|z_{t+\frac{1}{2}} - z_t\right\|^2 + \|z_{t+1} - z_t\|^2\right)\right) \\
&\leq \frac{2 + 6\eta^2 L^2}{\eta^2}\left(\|z_1 - z_0\|^2 + \frac{2H^2}{1 - (1+\sqrt{2})\eta L}\right) \\
&\leq \frac{6(1 + 3\eta^2 L^2)H^2}{\eta^2(1 - (1+\sqrt{2})\eta L)}.
\end{aligned}$$

The second last inequality holds by Lemma 14. The last inequality holds since $\|z_1 - z_0\|^2 \leq \frac{4}{L^2}\|F(z_0)\|^2 \leq H^2$. Recall that $\lambda = \sqrt{\frac{6(1+3\eta^2L^2)}{\eta^2L^2(1-(1+\sqrt{2})\eta L)}}$. This completes the proof. ■

E.3. Proof of Theorem 6

Fix any $T \geq 1$. From Lemma 13, we know that the potential function P_t is non-increasing for all $t \geq 1$. Lemma 15 guarantees that the sum of potential functions $\sum_{t=1}^{T} P_t$ is upper bounded by $\lambda^2 H^2 L^2$, where $\lambda^2 = \frac{6(1+3\eta^2L^2)}{\eta^2L^2(1-(1+\sqrt{2})\eta L)}$. Combining the above, we can conclude that the potential function at the last iterate P_T is upper bounded by $\frac{\lambda^2 H^2 L^2}{T}$. Since $P_T = \|F(z_T) + c_T\|^2 +$

$\|F(z_T) - F(z_{T-\frac{1}{2}})\|^2$, we obtain the last-iterate convergence rate $r^{tan}_{F,Z}(z_T)^2 \le \|F(z_T) + c_T\|^2 \le \frac{\lambda^2 H^2 L^2}{T}$.

The convergence rate on $\|F(z_T) + c_T\|^2$ implies a convergence rate on the gap function $\textsc{Gap}_{Z,F,D}(z_T)$ by Lemma 9:

$$\textsc{Gap}_{\mathcal{Z},F,D}(z_T) \le D \cdot \|F(z_T) + c_T\| \le \frac{\lambda DHL}{\sqrt{T}}.$$

Appendix F. Auxiliary Propositions

Proposition 16 (Two Identities) *Let $(x_k)_{k\in[4]}$, $(y_k)_{k\in[4]}$, x_0, u_2 and u_4 be arbitrary vectors in $\mathbb{R}^n$. Let $k \ge 1$ and $q \in (0,1)$ be two real numbers. If the following two equations holds:*

$$\begin{aligned} x_3 &= x_2 - y_1 - u_2 \\ x_4 &= x_2 - y_3 - u_4 \end{aligned}$$

then the following identity holds:

$$\begin{aligned} &\|y_2 + u_2\|^2 + \|y_2 - y_1\|^2 - \|y_4 + u_4\|^2 - \|y_4 - y_3\|^2 \\ &- 2 \cdot \langle y_4 - y_2, x_4 - x_2 \rangle \\ &- 2 \cdot \left(\frac{1}{4} \cdot \|x_4 - x_3\|^2 - \|y_4 - y_3\|^2 \right) \\ &- 2 \cdot \langle u_4 - u_2, x_4 - x_2 \rangle \\ =& \left\| \frac{x_3 - x_4}{2} + y_1 - y_2 \right\|^2 + \left\| \frac{x_3 + x_4}{2} - x_2 + y_2 + u_2 \right\|^2 \end{aligned}$$

If the following two equations holds:

$$\begin{aligned} x_3 &= x_2 - y_1 - u_2 + \frac{1}{k+1}(x_0 - x_2) \\ x_4 &= x_2 - y_3 - u_4 + \frac{1}{k+1}(x_0 - x_2) \end{aligned}$$

then the following identity holds:

$$\begin{aligned} &\frac{k(k+1)}{2}\left(\|y_2 + u_2\|^2 + \|y_2 - y_1\|^2 \right) + k\langle y_2 + u_2, x_2 - x_0 \rangle \\ &- \frac{(k+1)(k+2)}{2}\left(\|y_4 + u_4\|^2 + \|y_4 - y_3\|^2 \right) - (k+1)\langle y_4 + u_4, x_4 - x_0 \rangle \\ &- k(k+1) \cdot \langle y_4 + u_4 - y_2 - u_2, x_4 - x_2 \rangle \\ &- \frac{k(k+1)}{4q} \cdot \left\langle q \cdot \|x_4 - x_3\|^2 - \|y_4 - y_3\|^2 \right\rangle \\ =& \frac{k(k+1)}{4} \cdot \|u_4 - u_2 + y_1 - 2y_2 + y_3\|^2 \\ &+ \left(\frac{(1-4q)k - 4q}{4q}(k+1) \right) \cdot \|y_3 - y_4\|^2 \\ &+ (k+1) \cdot \langle y_3 - y_4, y_4 + u_4 \rangle \end{aligned}$$

Proof We verify the two identities by MATLAB. The code is available at
https://github.com/weiqiangzheng1999/Single-Call. ■

Proposition 17 ([6]) *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 $k \geq 2$,*

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

then for each $k \geq 2$ we have

$$a_k \leq \frac{4 \cdot C_1}{1-3p} \cdot \frac{1}{k^2}. \tag{29}$$

Proof We prove the statement by induction.
Base Case: $k = 2$**.** From Inequality (28), 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 (29) holds for $k = 2$.
Inductive Step: for any $k \geq 3$**.** Fix some $k \geq 3$ 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 $\sum_{t=3}^{2}$ as 0. By Inequality (28), we have

$$\begin{aligned}
\frac{k^2}{4} \cdot a_k &\leq C_1 + \frac{p}{1-p} \cdot \sum_{t=2}^{k-1} a_t \\
&\leq \frac{C_1}{1-p} + \frac{p}{1-p} \cdot \sum_{t=3}^{k-1} a_t && (a_2 \leq C_1) \\
&\leq \frac{C_1}{1-p} + \frac{4p \cdot C_1}{(1-p)(1-3p)} \cdot \sum_{t=3}^{k-1} \frac{1}{t^2} && \text{(Induction assumption on Inequality (29))} \\
&\leq \frac{C_1}{1-p} + \frac{2p \cdot C_1}{(1-p)(1-3p)} && \left(\textstyle\sum_{t=3}^{\infty} \frac{1}{t^2} = \frac{\pi^2}{6} - \frac{5}{4} \leq \frac{1}{2}\right) \\
&= \frac{C_1}{1-3p}.
\end{aligned}$$

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