Finite-Time Last-Iterate Convergence for Learning in Multi-Player Games

Yang Cai
Yale University
yang.cai@yale.edu

Argyris Oikonomou
Yale University
argyris.oikonomou@yale.edu

Weiqiang Zheng
Yale University
weiqiang.zheng@yale.edu

Abstract

We study the question of last-iterate convergence rate of the *extragradient algorithm* by [Kor76] and the *optimistic gradient algorithm* by [Pop80] in multi-player games. We show that both algorithms with *constant step-size* have last-iterate convergence rate of $O(\frac{1}{\sqrt{T}})$ to a Nash equilibrium in terms of the gap function in smooth monotone games, where each player's action set is an *arbitrary convex set*. Previous results only study the unconstrained setting, where each player's action set is the entire Euclidean space. Our results address an open question raised in several recent works [HIMM19, GPD20, GPDO20], which ask for last-iterate convergence rate of either the extragradient or the optimistic gradient algorithm in the constrained setting. Our convergence rates for both algorithms are tight and match the lower bounds by [GPD20, GPDO20]. At the core of our results lies a new notion – the *tangent residual*, which we use to measure the proximity to a Nash equilibrium. We use the tangent residual (or a modification of the tangent residual) as the the potential function in our analysis of the extragradient algorithm (or the optimistic gradient algorithm).

1 Introduction

We consider learning in *monotone games*, a class of multi-player games introduced by [Ros65] that include many well studied games, e.g., two-player zero-sum games, convex-concave games, λ-cocoercive games [LZMJ20], zero-sum polymatrix games [BF87, DP09, CD11], and zero-sum socially-concave games [EDMN09]. We focus on the following question: *Can we obtain **last-iterate convergence rate** to a **Nash equilibrium** in monotone games when all players act according to a simple learning algorithm?*

We adopt the multi-player online learning model as introduced in [CBL06], where players interact with each other repeatedly. At every time step t, every player $i \in \{1, \dots, N\}$ chooses an action $z_t^{(i)}$ from her action set $\mathcal{Z}^{(i)}$, which we assume to be a closed convex set in $\mathbb{R}^{n_i}$. We say the game is *unconstrained* if $\mathcal{Z}^{(i)} = \mathbb{R}^{n_i}$ for each player i. Player i's loss function $\ell_t^{(i)}(\cdot)$ is determined based on the underlying game and the actions of the other players in round t. Player i receives the loss $\ell_t^{(i)}(z_t^{(i)})$ as well as some additional feedback that informs her how to improve her decisions in the future. In this paper, we assume that each player receives the gradient feedback, i.e., player i receives the vector $\nabla \ell_t^{(i)}(z_t^{(i)})$. We make an additional mild assumption that the game is *smooth*, i.e., all players' gradients are Lipschitz. Smoothness is a natural assumption that is satisfied in most applications and is also made in the majority of works concerning monotone games.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Game class	Setting	Step size	Convergence rate
Strongly monotone	general	constant	$O(c^{-T})$ (see e.g., [Tse95] [LS19, MOP19, ZMM+21])
Cocoercive	unconstrained	constant	$O(\frac{1}{\sqrt{T}})$ [LZMJ20]
Monotone	general	constant	Asymptotic [Pop80, HIMM19]
	general	decreasing	Asymptotic * (see e.g., [ZMM+17] [ZMA+18, MZ19, HAM21])
	unconstrained	constant	$O(\frac{1}{\sqrt{T}})^{\dagger}$ [GPD20]
	general	constant	$O(\frac{1}{\sqrt{T}})$ **[This paper]**

Table 1: Last-iterate convergence for no-regret learning in smooth monotone games with perfect gradient feedback. (*) The results hold for variationally stable games. (†) The result holds under an additional second-order smoothness assumption.

The standard metric to quantify an online learning algorithm's performance is the regret. Formally, the regret of player i is defined as the difference between $\sum_{t=1}^{T} \ell_t^{(i)}(z_t^{(i)})$, player i's cumulative loss, and $\min_{z\in\mathcal{Z}^{(i)}} \sum_{t=1}^{T} \ell_t^{(i)}(z)$, the loss incurred by the best fixed action in hindsight. An online learning algorithm is no-regret if, even under an adversarially chosen loss sequence $\{\ell_t^{(i)}(\cdot)\}_{t\in[T]}$, its regret at the end of round T is sublinear in T.

A vast literature on learning in games discusses the convergence to a Nash equilibrium using no-regret learning algorithms. However, most of the results concern only the *time-average* convergence, i.e., the convergence of the time average of the joint action profile, rather than the last-iterate convergence, i.e., the convergence of the joint action profile. From a game-theoretic perspective, the last-iterate convergence is more appealing compared to the time-average convergence, as only the last-iterate convergence provides a description of the evolution of the overall behavior of the players. In contrast, the trajectory of the players' joint action may be cycling around in the space perpetually while still converges in the time-average sense as demonstrated by [MPP18]. Recently, a line of work is devoted to obtain last-iterate convergence in smooth monotone games [ZMB+17, ZMM+17, ZMA+18, DP18, MOP19, HIMM19, LNPW20, GPD20, LZMJ20, ZMM+21]. However, unless the game is strongly monotone or unconstrained, only asymptotic convergence is known. Moreover, many of these results crucially rely on decreasing step-size, which, as pointed out by [LZMJ20], is unnatural from an economic point of view, because it treats newly acquired information with decreasing importance. Hence, the following question is of particular interest and is raised as an open question in [HIMM19, LZMJ20, GPD20].

*Can we establish **last-iterate rates** if all players of a **constrained** smooth monotone game act according to a no-regret learning algorithm with **constant step size**?* (*)

[LZMJ20] first realizes the importance of question (*) and takes initial steps towards addressing it. They show that if all players follow the *gradient descent* algorithm with constant step size, then for all smooth λ-cocoercive games, the joint action $(z_t^{(1)}, \dots, z_t^{(N)})$ has last-iterate convergence rate of $O(\frac{1}{\sqrt{T}})$ to a Nash equilibrium in terms of the gap function. For smooth strongly monotone games, a subclass of λ-cocoercive games, linear last-iterate convergence rates are known [Tse95, Mal15, LS19, MOP19, ZMM+21]. Despite the generality of λ-cocoercive games, several fundamental classes of games such as two-player zero-sum games, zero-sum polymatrix games [BF87, DGP09, CD11] and its generalization zero-sum socially-concave games [EDMN09] are monotone but not λ-cocoercive. [GPD20] extends the result to smooth monotone games with an $O(\frac{1}{\sqrt{T}})$ last-iterate convergence rate using a different algorithm – the *optimistic gradient* by [Pop80] under an additional second-order smoothness assumption. Note that gradient descent has been observed to fail to converge in even two-player zero-sum games (see e.g., [DISZ17]), so a different algorithm is indeed needed. However, the results by [LZMJ20] and [GPD20] only consider *unconstrained* games, while in most game-theoretic settings, the players' actions are constrained. For example, in finite games, a player is restricted to choose a distribution over her finite set of actions. We summarize the results for last-iterate convergence in monotone games in Table 1.

Our Contributions. Our first contribution is to provide an affirmative answer to question (*).

> **Contribution 1:** In Theorem 3, we show that if all players of a constrained smooth monotone game act according to the optimistic gradient algorithm, which is no-regret, with a **constant step size**, then their joint action exhibits a last-iterate convergence rate of $O\left(\frac{1}{\sqrt{T}}\right)$ in terms of the gap function (Definition 2) to a Nash equilibrium.

Our result holds in the constrained setting and does not rely on the second-order smoothness assumption made in [GPD20]. Moreover, our upper bound is tight and matches the lower bound of [GPD20].

The problem of finding a Nash equilibrium in a smooth monotone game is essentially equivalent to solving a Lipschitz and monotone variational inequality (VI) (Definition 5 in Appendix D),[1] which has been studied since the 1960s [HS66, Bro65, LS67, BS68, Sib70]. There is a vast literature on solving VIs, and we refer the reader to [**?**] for further references. The extragradient (EG) algorithm by [Kor76] and the optimistic gradient (OG) algorithm by [Pop80] are arguably the two most classical and popular methods for solving monotone VIs. Despite the long history, the last-iterates of both algorithms are only known to asymptotically converge to a solution of the monotone and Lipschitz VI,[2] but no upper bounds on the rate of convergence had been provided for the general setting.

> **Contribution 2:** We provide the first and tight last-iterate convergence rate of $O(\frac{1}{\sqrt{T}})$ in terms of the gap function for both EG and OG with constant step size for solving Lipschitz and monotone VIs (Theorem 9 and Theorem 10 in Appendix I).

Our analysis of OG for monotone games directly applies to monotone VIs. To be more consistent with our discussion of OG, we state our last-iterate convergence rate of EG in the context of online learning in monotone games (Theorem 3).[3] Prior to our work, last-iterate convergence rate of EG only exists in the unconstrained setting. [GPDO20] shows a $O(\frac{1}{\sqrt{T}})$ upper bound in terms of the gap function under an additional second-order smoothness condition. [GLG21] improves the result and shows that the same upper bound holds without the second-order smoothness condition but still requires the setting to be unconstrained.

Our Analysis. As mentioned in [GPD20], "The lack of existence of a natural potential function in general monotone games is a significant challenge in establishing last-iterate convergence." Indeed, many of the natural quantities such as the gap function, the norm of the gradient, the difference of two consecutive iterates are provably non-monotone even in bilinear games. See Appendix J for more discussion and examples. We propose a notion that (i) measures the proximity to a Nash equilibrium, and (ii) can be used to construct natural potential functions to analyze both EG and OG in monotone games. We call the new notion the **tangent residual**, which can be viewed as the norm of the gradient projected to the tangent cone of the current iterate (Definition 4). The tangent residual plays a crucial role in our analyses for both algorithms. Unlike the other quantities mentioned above, we show that the *tangent residual is monotonically decreasing* and has *a last-iterate convergence rate of* $O(\frac{1}{\sqrt{T}})$ for EG. For OG, we prove that a small modification of the tangent residual is monotonically decreasing, which implies that the tangent residual has a last-iterate convergence rate of $O(\frac{1}{\sqrt{T}})$. Using the convergence rate of the tangent residual, we can easily derive the last-iterate convergence rate of other classical performance measures such as the gap function and the total gap function. However, we suspect these rates can be challenging to obtain directly without the help of a potential function. Finally, we establish the monotonicity of our potential functions using computer-aided proofs based on **sum-of-squares (SOS)**

[1] Technically, the definition of Lipschitz and monotone VI captures finding a Nash equilibrium in a smooth monotone game as a special case, but the difference has little impact on our analyses, and our results hold for Lipschitz and monotone VIs.

[2] The last-iterate asymptotic convergence of EG can be found in [Kor76] and [**?**], and the last-iterate asymptotic convergence of OG can be found in [Pop80] and [HIMM19].

[3] Although EG is not a no-regret learning algorithm (see Proposition 10 in [GPD20]), it is nevertheless a simple and natural learning algorithm.

programming [Nes00, Par00, Par03, Las01, Lau09]. Indeed, our potential function of OG is directly constructed using SOS programming. Additionally, our computer-aided proofs can be easily verified by humans. We think the tangent residual and the SOS-based analysis might be of independent interest. See Section 4.2, Appendix B, and Appendix C for more discussion.

1.1 Related Work

Last-Iterate Convergence Rate for EG/OG like Algorithms. [GPDO20, GPD20] show a lower bound of $\Omega(\frac{1}{\sqrt{T}})$ for solving bilinear games using any p-SCLI algorithms, which include EG and OG. In the unconstrained setting, if we further assume that either the game is strongly monotone or the payoff matrix A in a bilinear game has all singular values bounded away from 0, linear convergence rate is known for EG, OG, and several of their variants [DISZ17, GBV$^+$18, LS19, MOP19, PDZC20, ZY19]. The results for the constrained setting are sparser. Unless the game is strongly monotone, most results only guarantee asymptotic convergence, i.e., converge in the limit, [DP18, LNPW20]. Finally, a recent paper by [WLZL21b] provides a linear convergence rate of OG for bilinear games when the domain is a polytope. They show that there is a *problem dependent* constant $0 < c < 1$ that depends on the payoff matrix of the game as well as the constraint set, so that the error shrinks by a $1 - c$ factor per iteration. However, c may be arbitrarily close to 0, even if we assume the corresponding operator to be L-Lipschitz. Overall, their "problem-dependent" bound is incomparable and complements the worst-case view taken in this paper, where we want to derive the worst-case convergence rate for all smooth and monotone games. Our results are the first last-iterate convergence rates in this worst-case view and match the lower bounds by [GPDO20, GPD20].

Other Algorithms and Performance Measures. It is well-known that both EG [Nem04] and OG [HIMM19, MOP20] have a time-average convergence rate of $O(\frac{1}{T})$ in terms of the gap function for smooth monotone games. Other than the gap function, one can also measure the convergence using the norm of the operator if the setting is unconstrained, or the natural residual (Definition 7 in Appendix E) or similar notions if the setting is constrained. In the unconstrained setting, [Kim21], [YR21], and [LK21] provide algorithms that obtain $O(\frac{1}{T})$ convergence rate in terms of the norm of the operator, which is shown to be optimal by [YR21] for Lipschitz and monotone VIs. In the constrained setting, [Dia20] shows the same $O(\frac{1}{T})$ convergence rate under the extra assumption that the operator is cocoercive and loses an additional logarithmic factor when the operator is only monotone. Our result implies a $O(\frac{1}{\sqrt{T}})$ last-iterate convergence rate in terms of the natural residual or the gap function for both EG and OG. A main motivation of this paper is game-theoretic, that is, we would like to view *natural* learning algorithms as models of agents' behavior in online learning and understand the speed for the overall behavior to converge to a Nash equilibrium. Although some of the results above are natural for games, they either only hold in the unconstrained setting or do not converge in the last-iterate sense. From this game-theoretic view point, we believe understanding the last-iterate convergence rate of these natural learning algorithms is an important problem.

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 $\mathbb{R}^n$. A **continuous game** $\mathcal{G}$ is denoted as $(\mathcal{N}, \{\mathcal{Z}^{(i)}\}_{i\in[N]}, \{f^{(i)}\}_{i\in[N]})$ where there are N players $\mathcal{N} = \{1, \cdots, N\}$. Player $i \in \mathcal{N}$ chooses action from a closed convex set $\mathcal{Z}^{(i)} \subseteq \mathbb{R}^{n_i}$ such that $\mathcal{Z}_\mathcal{G} := \Pi_{i\in\mathcal{N}}\mathcal{Z}^{(i)} \subseteq \mathbb{R}^n$ and wants to minimize her cost function $f^{(i)} : \mathcal{Z}_\mathcal{G} \to \mathbb{R}$. For each player i, we denote by $z^{(-i)}$ the vector of actions of all the other players and by $z^{(i)}$ the action of player i. When players play according to action profile $z \in \mathcal{Z}_\mathcal{G}$, player i receives gradient feedback $\nabla_{z^{(i)}} f^{(i)}(z^{(1)}, \ldots, z^{(N)})$. A *Nash equilibrium* of game $\mathcal{G}$ is an action profile $z^* \in \mathcal{Z}_\mathcal{G}$ such that $f^{(i)}(z^*) \le f^{(i)}(z'^{(i)}, z^{*(-i)})$ for any $z'^{(i)} \in \mathcal{Z}^{(i)}$.

Let $F_{\mathcal{G}}(\cdot) = (\nabla_{z^{(1)}} f^{(1)}(\cdot), \cdots, \nabla_{z^{(N)}} f^{(N)}(\cdot))$ be an operator that maps any joint action in $\mathcal{Z}$ to the corresponding joint gradient feedback vector in $\mathbb{R}^n$. When $\mathcal{G}$ is clear from context, we omit the subscript and write $F_{\mathcal{G}}(z)$ ($\mathcal{Z}_{\mathcal{G}}$ resp.) as $F(z)$ ($\mathcal{Z}$ resp.).

Nash Equilibria of Monotone Games. Throughout this paper, we focus on smooth monotone games (Definition 1). It is well known that finding a Nash equilibrium of a monotone game is exactly the same as finding a solution to the variational inequality with monotone operator $F_{\mathcal{G}}$ (Lemma 1).

Definition 1 ([Ros65]). *We say game $\mathcal{G}$ is L-*smooth *and* monotone *if the operator $F_{\mathcal{G}}$ is L-Lipschitz (i.e., $\forall z, z' \in \mathcal{Z}$, $L \cdot \|z - z'\| \geq \|F_{\mathcal{G}}(z) - F_{\mathcal{G}}(z')\|$) and monotone (i.e., $\forall z, z' \in \mathcal{Z}$, $\langle F_{\mathcal{G}}(z) - F_{\mathcal{G}}(z'), z - z'\rangle \geq 0$).*

Remark 1 (Monotone and Concave Games). *For any monotone game $\mathcal{G} = (\mathcal{N}, \{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}, \{f^{(i)}\}_{i\in[N]})$, the monotonicity condition implies that for any fixed $z^{(-i)}$, player i's cost function is convex in $z^{(i)}$. Thus, all monotone games are concave games. However, the converse is not true as illustrated in the following example.*

Consider a two player game $\mathcal{G}$ where player 1 (or player 2) chooses action $x \in \mathbb{R}$ (or $y \in \mathbb{R}$), and their cost functions are $f^{(1)}(x, y) = f^{(2)}(x, y) = x \cdot y$. Clearly, $f^{(1)}(x, y)$ (or $f^{(2)}(x, y)$) is convex in x (or y) if we fix y (or x). It is not hard to see that $F_{\mathcal{G}}(x, y) = (y, x)$ for any $x, y \in \mathbb{R}$. Therefore, the game is not monotone as $\langle F_{\mathcal{G}}(x, y) - F_{\mathcal{G}}(y, x), (x, y) - (y, x)\rangle = -2(x - y)^2 < 0$ for any $x \neq y$.

Lemma 1 (1.4.2 Proposition [?]). *For a monotone game $\mathcal{G}$, an action profile z^* is a Nash equilibrium if and only if $\langle F_{\mathcal{G}}(z^*), z^* - z\rangle \leq 0, \forall z \in \mathcal{Z}$.*

Remark 2. *One sufficient condition for a Nash equilibrium z^* to exist is when the set $\mathcal{Z}$ is bounded, but there are also other sufficient conditions that apply to unbounded $\mathcal{Z}$. See [?] for more details. Throughout this paper, we only consider monotone games that have a Nash equilibrium.*

Definition 2 (Gap and Total Gap Function). *For a monotone game $\mathcal{G}$, two standard ways to measure the proximity of an action profile $z \in \mathcal{Z}$ to Nash equilibrium, are by its gap function and total gap function. Let D be a fixed parameter. The gap function is defined as $\text{GAP}_{\mathcal{G},D}(z) = \max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \langle F_{\mathcal{G}}(z), z - z'\rangle$, where $\mathcal{B}(z, D)$ is a ball with radius D centered at z.*[4] *The total gap function is defined as $\text{TGAP}_{\mathcal{G},D}(z) = \sum_{i\in\mathcal{N}} \left(f^{(i)}(z) - \min_{z'^{(i)} \in \mathcal{Z}^{(i)} \cap \mathcal{B}(z^{(i)},D)} f^{(i)}(z'^{(i)}, z^{(-i)})\right)$.*

When $\mathcal{G}$ and D are clear from context, we omit subscripts and write the gap function (total gap function resp.) at z as $\text{GAP}(z)$ ($\text{TGAP}(z)$ resp.).

Remark 3. *The total gap function is the sum of the suboptimality gaps, i.e., a player i's suboptimality gap is her cost under z minus her cost after best responding to $z^{(-i)}$ within $\mathcal{B}(z^{(i)}, D)$. As the suboptimality gap is nonnegative for every player, a small total gap implies that no player can deviate from their current action to significantly improve their cost, and the action profile is close to a Nash equilibrium. Thus the total gap is 0 if and only if players are at a Nash Equilibrium.*

In Lemma 2, we show that the gap function with radius D is an upper bound of the total gap with radius $\frac{D}{\sqrt{N}}$, where N is the number of players, so if an action profile has small gap function, then again no player can deviate to significantly reduce their cost. A corollary of Lemma 1, is that the gap function is 0 if and only if players are at a Nash Equilibrium.

The Optimistic Gradient Algorithm. Let w_k be the action profile played at day k and assume that player i updates her action according to the OG algorithm. For arbitrary $z_0^{(i)}$ and $w_0^{(i)}$, player's i action at day $k+1$ is $w_{k+1}^{(i)}$ such that $z_k^{(i)} = \Pi_{\mathcal{Z}^{(i)}}\left[z_{k-1}^{(i)} - \eta \nabla_{z^{(i)}} f^{(i)}(w_k^{(i)})\right]$ and $w_{k+1}^{(i)} = \Pi_{\mathcal{Z}^{(i)}}\left[z_k^{(i)} - \eta \nabla_{z^{(i)}} f^{(i)}(w_k^{(i)})\right]$.

[4] When $\mathcal{Z}$ is bounded, we choose D to be the diameter of $\mathcal{Z}$. When $\mathcal{Z}$ is unbounded, the standard choice is to choose D so that all the iterates of the algorithms are guaranteed to maintain in the ball $\mathcal{B}(z, D)$. See Appendix D for more details.

Note that action $z_k^{(i)}$ is not being played by player i and is only used to compute action $w_{k+1}^{(i)}$. When all the players update their actions according to OG with step-size η, let $w_0 = (w_0^{(1)}, \ldots, w_0^{(N)})$ and $z_0 = (z_0^{(1)}, \ldots, z_0^{(N)})$. Then at day $k+1$, players pick action profile w_{k+1}, where:

$$z_k = \Pi_{\mathcal{Z}}\left[z_{k-1} - \eta F(w_k)\right], \qquad w_{k+1} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(w_k)\right] \tag{1}$$

Clearly, the OG update rule is well defined for any operator F and any closed convex set $\mathcal{Z}$, and this is how the OG algorithm is defined for variational inequalities.

The Extragradient Algorithm. Let z_k be the action profile played at day $2k$ and assume that all players update their actions according to EG with step-size η. Then players play according to action profile $z_{k+\frac{1}{2}}$ at day $2k+1$ and action profile z_{k+1} at day $2k+2$, where:

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

Similarly, the EG update rule is well defined for any operator F and any closed convex set $\mathcal{Z}$. For the rest of the paper, we only consider the case where all players use the OG (or EG) algorithm with constant step-size η. We omit superscripts that denote players' identity and use Expression (1) (or Expression (2)) for the update rule when players use OG (or EG).

3 The Tangent Residual and Its Properties

We formally introduce our key performance measure, the tangent residual . We define the tangent residual over operators rather than games, as it will be easier to provide intuition behind its formulation.

Definition 3 (Unit Normal Cone). *Given a closed convex set $\mathcal{Z} \subseteq \mathbb{R}^n$ and a point $z \in \mathcal{Z}$, we denote by $N_{\mathcal{Z}}(z) = \langle v \in \mathbb{R}^n : \langle v, z' - z\rangle \leq 0, \forall z' \in \mathcal{Z}\rangle$ the normal cone of $\mathcal{Z}$ at point z and by $\widehat{N}_{\mathcal{Z}}(z) = \{v \in N_{\mathcal{Z}}(z) : \|v\| \leq 1\}$ the intersection of the unit ball with the the normal cone of $\mathcal{Z}$ at z. Note that $\widehat{N}_{\mathcal{Z}}(z)$ is nonempty and compact for any $z \in \mathcal{Z}$, as $(0, \ldots, 0) \in \widehat{N}_{\mathcal{Z}}(z)$.*

Definition 4 (Tangent Residual). *Given an operator $F : \mathcal{Z} \to \mathbb{R}^n$ and a closed convex set $\mathcal{Z}$, let $T_{\mathcal{Z}}(z) := \{z' \in \mathbb{R}^n : \langle z', a\rangle \leq 0, \forall a \in N_{\mathcal{Z}}(z)\}$ be the tangent cone of z,[5] and define $J_{\mathcal{Z}}(z) := \{z\} + T_{\mathcal{Z}}(z)$. The tangent residual of F at $z \in \mathcal{Z}$ is defined as $r^{tan}_{(F,\mathcal{Z})}(z) := \|\Pi_{J_{\mathcal{Z}}(z)}[z - F(z)] - z\|$. An equivalent definition is $r^{tan}_{(F,\mathcal{Z})}(z) := \sqrt{\|F(z)\|^2 - \max_{\substack{a \in \widehat{N}_{\mathcal{Z}}(z), \\ \langle F(z), a\rangle \leq 0}} \langle a, F(z)\rangle^2}$.*

Remark 4. *We show the equivalence of the two definitions of tangent residual in Lemma 5 in Appendix D. We may use either of them depending on which one is more convenient.*

When the convex set $\mathcal{Z}$ and the operator F are clear from context, we are going to omit the subscript and denote the unit normal cone as $\widehat{N}(z) = \widehat{N}_{\mathcal{Z}}(z)$ and the tangent residual as $r^{tan}(z) = r^{tan}_{(F,\mathcal{Z})}(z)$. Although the definition is slightly technical, one can think of the tangent residual as the norm of another operator $\widehat{F}$, which is F projected to all directions that are not "blocked" by the boundary of $\mathcal{Z}$ if one takes an infinitesimally small step $\epsilon \cdot F(z)$, which is the same as projecting F to $J_{\mathcal{Z}}(z)$. Intuitively, if the tangent residual is small, then the next iterate will not be far away from the current one.

Tangent Residual and Its Connection with Games. Given a monotone game $\mathcal{G}$, we denote by $r^{tan}_{\mathcal{G}}(z) = r^{tan}_{(F_{\mathcal{G}},\mathcal{Z})}(z)$. In the next lemma, we argue that a small tangent residual implies a

[5]Interested readers can find a thorough introduction of the tangent cone and its definition for general feasible sets in Chapter 6.A of [RW09]. When the feasible set is convex, Theorem 6.9 of [RW09] provides a more succinct definition of the tangent cone and normal cone. When the feasible set is a closed convex set, Corollary 6.30 of [RW09] further states that the tangent cone is the polar cone of the normal cone. As we consider closed convex sets in our paper, we choose to define the tangent cone for closed and convex sets directly as the polar cone of the normal cone, as it is the most convenient definition for us.

small gap and total gap, hence it suffices to show that the last-iterate has a small tangent residual. The proof is postponed to Appendix E.

Lemma 2. *[Adapted from Theorem 10 in [GPDO20] and Proposition 2 in [GPD20]] Let* $\mathcal{G} = (\mathcal{N}, \{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}, \{f^{(i)}\}_{i\in[N]})$ *be a monotone game where* $\{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}$ *are closed convex sets. For* $z \in \mathcal{Z}$*, we have* $\mathrm{GAP}_{\mathcal{G},D}(z) \le D \cdot r^{tan}_{\mathcal{G}}(z)$ *and* $\mathrm{TGAP}_{\mathcal{G},D}(z) \le \mathrm{GAP}_{\mathcal{G},\sqrt{N}D}(z) \le \sqrt{N}D \cdot r^{tan}_{\mathcal{G}}(z)$*.*

4 Last-Iterate Convergence Rate for EG and OG

We prove the last-iterate convergence rate for EG and OG. We first describe our proof plan.

Proof Plan. Our analyses for both algorithms follow the same three-step procedure: (i) define a potential function that measures the proximity to a Nash equilibrium of the current iterate; (ii) prove a best-iterate convergence rate, that is, show that in T steps, there exists one iterate whose potential function is small; (iii) show that the potential function is non-increasing, so the last-iterate is the best-iterate, and the best-iterate convergence rate becomes the last-iterate convergence rate.

The first major challenge we face is to choose the appropriate potential functions. In the unconstrained case, the central quantity is the norm of the operator $F_{\mathcal{G}}$. The key component of the analyses [LZMJ20, GPDO20, GPD20, GLG21] is to establish that the norm of the operator $F_{\mathcal{G}}$ at the last iterate (also the T-th iterate) is upper bounded by $O(\frac{1}{\sqrt{T}})$, which implies a $O(\frac{1}{\sqrt{T}})$ last-iterate convergence rate for the gap function. In the constrained setting, the norm of the operator is a poor choice to measure the proximity to a Nash equilibrium, as it can be far away from 0 even at a Nash equilibrium.

Potential function for EG. We use the **tangent residual** as the potential function for EG. Our starting point is to find a suitable generalization for the norm of the operator in the constrained setting. A standard generalization is the natural residual (Definition 7 in Appendix E), which takes the constraints into account and is guaranteed to converge to 0 at the Nash equilibrium. Unfortunately, we observe that the natural residual is *not monotonically decreasing* even in basic bilinear games, making it difficult to directly analyze. Similar non-monotonicity has been observed for several other natural performance measures such as $\|z_k - z_{k+1/2}\|$,[6] $\|z_k - z_{k+1}\|$, $\max_{z\in\mathcal{Z}} \langle F(z), z_k - z\rangle$ and $\max_{z\in\mathcal{Z}} \langle F(z_k), z_k - z\rangle$, leaving all these quantities unsuitable as a potential function. See more discussion in Appendix J. Indeed, tangent residual is the only natural generalization of the norm of the operator that is always monotone in our numerical experiments.

Potential function for OG. We choose the potential function $\Phi(z_k, w_k) = \|F(z_k) - F(w_k)\|^2 + r^{tan}(z_k)^2$ in our analysis for OG. The potential function can be interpreted as the squared tangent residual $r^{tan}(z_k)^2$ and an extra correction term $\|F(z_k) - F(w_k)\|^2$. The potential function is discovered via SOS programming, with more details in Section 4.2.

4.1 Best-Iterate Convergence

En route to establish the last-iterate convergence of EG and OG, we first show a weaker guarantee known as the best-iterate convergence. In Lemma 3 we show that when all the players use EG for $2 \cdot T$ steps (OG for T steps), then there exists $t^* \in [T]$ where $r^{tan}(z_{t^*+1}) \le O(\frac{1}{\sqrt{T}})$ ($\Phi(z_{t^*}, w_{t^*}) \le O(\frac{1}{T})$ resp.). The proof for EG builds upon the proof for the best-iterate w.r.t. $\|z_k - z_{k+\frac{1}{2}}\|$ by [Kor76**?**], while the proof for OG builds upon the best-iterate proof for $\|z_k - w_{k+1}\|$ by [WLZL21a, HIMM19]. The proof of Lemma 3 is in Appendix F.

Lemma 3. *Let* $\mathcal{G} = (\mathcal{N}, \{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}, \{f^{(i)}\}_{i\in[N]})$ *be an L-smooth and monotone game where* $\{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}$ *are closed convex sets and let* z^* *be a Nash Equilibrium of* $\mathcal{G}$*. Assume that all the players*

[6] $\|z_k - z_{k+1/2}\|$ is proportional to the norm of the operator mapping introduced in [Dia20].

update their actions using the EG algorithm with arbitrary starting action profile z_0 and step-size $\eta \in (0, \frac{1}{L})$. Then for and any $T > 0$, there exists $t^* \in [T]$ such that:

$$r^{tan}(z_{t^*+1}) \leq \frac{1 + \eta L + (\eta L)^2}{\eta} \frac{1}{\sqrt{T}} \frac{\|z_0 - z^*\|}{\sqrt{1 - (\eta L)^2}}.$$

Assume that all the players update their actions using the OG algorithm with arbitrary starting action profiles z_0,w_0 and step-size $\eta \in (0, \frac{1}{2L})$. Then for and any $T > 0$, there exists $t^* \in [T]$ such that:

$$\Phi(z_{t^*}, w_{t^*}) \leq \frac{1}{\eta^2 \cdot T} \left(\frac{4 + 6\eta^4 L^4}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2 + \frac{16\eta^2 L^2 + 6\eta^4 L^4}{1 - 4\eta^2 L^2} \|w_0 - z_0\|^2 \right).$$

4.2 Monotonicity of the Potentials

In this section, we prove that $r^{tan}(z_k)$ ($\Phi(z_k, w_k)$ resp.) is non-increasing across iterates of EG (OG resp.), which, in combination with Lemma 3, implies the last-iterate convergence rate of smooth monotone games when all players update their actions using EG (OG resp.).

SOS Programming. Suppose we want to prove that a polynomial $g(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}]$ is non-negative over a semialgebraic set $\mathcal{S} = \{\boldsymbol{x} : g_i(\boldsymbol{x}) \leq 0, \forall i \in [M], h_i(\boldsymbol{x}) = 0, \forall i \in [N]\}$, where each $g_i(\boldsymbol{x})$ ($h_i(\boldsymbol{x})$ resp.) is also a polynomial. One way is to construct a *certificate of non-negativity*, by providing a set of nonnegative coefficients $\{p_i \geq 0\}_{i \in [M]}$ and $\{q_i\}_{i \in [N]}$ such that $g(\boldsymbol{x}) + \sum_{i \in [M]} p_i \cdot g_i(\boldsymbol{x}) + \sum_{i \in [N]} q_i \cdot h_i(\boldsymbol{x})$ is a SOS polynomial. Surprisingly, if $g(\boldsymbol{x})$ is indeed non-negative over $\mathcal{S}$, a certificate of non-negativity always exists as guaranteed by a foundational result in real algebraic geometry – the Krivine-Stengle Positivestellensatz [Kri64, Ste74], a generalization of Artin's resolution of Hilbert's 17th problem [Art27]. The SOS programming [Nes00, Par00, Par03, Las01, Lau09] is a systematic way to search for such a certificate using semidefinite programming. See Appendix B for details.

Our approach is to apply SOS programming to search for a certificate of non-negativity for $r^{tan}(z_k)^2 - r^{tan}(z_{k+1})^2$ and $\Phi(z_k, w_k) - \Phi(z_{k+1}, w_{k+1})$ for every k, over a semialgebraic set defined by the update rule of the corresponding algorithm and set of constraints $\mathcal{Z}$.

Two important challenges with this approach is that the number of variables depends on the dimension of $\mathcal{Z}$ and there are infinitely many constraints associated with the problem (e.g. the set $\{\langle F(z) - F(z'), z - z' \rangle \geq 0, \forall z, z' \in \mathcal{Z}\}$). We provide a detailed exposition for the reduction of the number of constraints and the efficient formulation into a SOS program in Appendix G.3, where we prove the monotonicity of our potential function for EG in both the unconstrained and constrained settings.

Searching for the potential functions. The potential function of our analysis of OG is directly discovered using an SOS program. The program is formulated by searching over linear combinations of $||F(z_k)||^2 - ||F(z_{k+1})||^2$,$||F(w_k)||^2 - ||F(w_{k+1})||^2$,$\langle F(z_k), F(w_k) \rangle - \langle F(z_{k+1}), F(w_{k+1}) \rangle$ and $r^{tan}(z_k)^2 - r^{tan}(z_{k+1})^2$, under (i) the constraint that the linear combination is non-increasing,[7] and (ii) the constraints induced by properties of the operator $F(\cdot)$, the update rule of OG and the set $\mathcal{Z}$ (See Appendix G.3 for a demonstration of the induced semialgebraic set of EG algorithm).

Our basis functions are chosen in a way so that we can search over all candidate potential functions that are a linear combination of (i) $r^{tan}(z_k)^2$ and (ii) any squared norms of a linear combination of $F(z_k), F(w_k)$. Observe that the difference between two consecutive iterates of any of the above functions can be expressed as a linear combination of the basis functions we chose.

We then use the linear combination output by the SOS program as the potential function in our analysis. We believe our heuristic for finding a potential function could be useful in

[7] To avoid finding the trivial linear combination, i.e., all coefficients equal to 0, we also use the objective function in the SOS program to encourage a non-trivial solution if one exists by, for example, maximizing the sum of the coefficients of the linear combination.

other settings. In general, one can first choose a collection of basis functions that may be part of a potential function, then use SOS programming to search over all linear combinations of the basis functions subject to the constraint that the linear combination is non-negative to discover the potential function.

Initially, our candidate potential functions included the squared tangent residuals evaluated at z_k and w_k and all degree-2 monomials of all vectors of interest. In other words, we searched over all linear combinations of (i) $r^{tan}(z_k)^2, r^{tan}(w_k)^2$ and (ii) any linear combination of degree-2 monomials of $z_k, F(z_k), w_k, F(w_k)$, subject to the constraint that the candidate potential function is non-negative. We included the tangent residuals in the basis of our potential functions due to their central role in the analysis of the EG algorithm (we show that the tangent residual is non-increasing across the iterates of the EG algorithm in Theorem 2).

We also observed that any non-zero linear combination of only the tangent residuals evaluated at z_k and w_k is not monotone for the OG algorithm. Motivated by this, we enlarged the basis to include the degree-2 monomials of $z_k, F(z_k), w_k, F(w_k)$ so that we had the flexibility to introduce an extra correction term in the potential function. Starting from this very flexible class of potentials functions, we gradually removed elements in the basis, and the basis that we selected in the paper was a minimal basis such that: (i) it contains a non-increasing potential function and (ii) it enjoys best-iterate convergence (See Lemma 17 in Appendix H.2) and (iii) the discovered potential function bounds the gap functions.

We establish the monotonicity of the potential functions in Theorem 1 and Theorem 2, which are both stated for general monotone operators over convex sets rather than games for technical convenience. We turn our attention back to games in Section 4.3 where we provide formal last-iterate convergence guarantees for smooth monotone games.

4.2.1 Monotonicity of $\Phi(z_k, w_k)$ for OG in the Unconstrained Setting

To better illustrate our approach, we prove in Theorem 1 the monotonicity of $\Phi(z_k, w_k)$ for the OG algorithm in the unconstrained setting. Note that this is already a strengthening of [GPD20], where we provide a simple potential function that can be used to directly argue that $\|F(z_k)\|^2 = O(\frac{1}{k})$ without making any second-order smoothness assumption.

Theorem 1. *Let $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator. For any $z_k, w_k \in \mathbb{R}^n$, the unconstrained OG algorithm with step-size $\eta \in (0, \frac{1}{2L})$ satisfies $\Phi(z_k, w_k) \geq \Phi(z_{k+1}, w_{k+1})$.* [8]

Proof. Since F is monotone and L-Lipschitz, we have $\langle F(z_{k+1}) - F(z_k), z_k - z_{k+1}\rangle \leq 0$ and $\|F(w_{k+1}) - F(z_{k+1})\|^2 - L^2\|w_{k+1} - z_{k+1}\|^2 \leq 0$. We simplify them using the update rule of OG and $\eta^2 L^2 < \frac{1}{4}$. In particular, we replace $z_k - z_{k+1}$ by $\eta F(w_{k+1})$ and $w_{k+1} - z_{k+1}$ with $\eta F(w_{k+1}) - \eta F(w_k)$.

$$\langle F(z_{k+1}) - F(z_k), F(w_{k+1})\rangle \leq 0, \tag{3}$$

$$\|F(w_{k+1}) - F(z_{k+1})\|^2 - \frac{1}{4}\|F(w_{k+1}) - F(w_k)\|^2 \leq 0. \tag{4}$$

It is not hard to verify the following identity:

$$\begin{aligned}&\|F(z_k) - F(w_k)\|^2 + \|F(z_k)\|^2 - \|F(z_{k+1}) - F(w_{k+1})\|^2 - \|F(z_{k+1})\|^2 \\ &+ 2 \cdot \text{LHS of Inequality}(3) + 2 \cdot \text{LHS of Inequality}(4) = \frac{1}{2}\|F(w_k) + F(w_{k+1}) - 2F(z_k))\|^2.\end{aligned}$$

Thus, $\|F(z_k) - F(w_k)\|^2 + \|F(z_k)\|^2 \geq \|F(z_{k+1}) - F(w_{k+1})\|^2 + \|F(z_{k+1})\|^2$. □

4.2.2 Constrained EG and OG

In the constrained setting, the approach is more complicated due to the projections and we postpone any exposition to Appendix G.3. We state the results in the following theorem.

[8] In the unconstrained setting, $r^{tan}_{(F,\mathbb{R}^n)}(z) = \|F(z)\|^2$.

Theorem 2. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set and $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator. For any $z_k \in \mathcal{Z}$, the EG algorithm with step-size $\eta \in (0, \frac{1}{L})$ satisfies $r^{tan}(z_k) \geq r^{tan}(z_{k+1})$. Moereover, for any $z_k, w_k \in \mathcal{Z}$, the OG algorithm with step-size $\eta \in (0, \frac{1}{2L})$ satisfies $\Phi(z_k, w_k) \geq \Phi(z_{k+1}, w_{k+1})$, where $\Phi(z_k, w_k) = \|F(z_k) - F(w_k)\|^2 + r^{tan}(z_k)^2$ for all k.*

4.3 Last-Iterate Convergence of EG and OG

In this section, we formally combine Lemma 3 and Theorem 2 to show the last-iterate convergence of EG and OG with respect to the tangent residual, gap function, and the total gap function. Recall that when all the players update their actions using the EG algorithm, then z_k is the action profile played at day $2 \cdot k$ and $z_{k+\frac{1}{2}}$ is the action played at day $2 \cdot k + 1$, while when all the players update their action profile using the OG algorithm, then w_k is the action profile they play at day k. The formal proof of Theorem 3 is postponed at Appendix F.

Theorem 3. *Let $\mathcal{G} = (\mathcal{N}, \{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}, \{f^{(i)}\}_{i\in[N]})$ be an L-smooth and monotone game where $\{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}$ are closed convex sets and let z^* be a Nash Equilibrium of $\mathcal{G}$. Assume that all the players update their actions using the EG algorithm with arbitrary starting action profile z_0 and step-size $\eta \in (0, \frac{1}{L})$. Let $D_0 = \frac{3\|z_0 - z^*\|}{\sqrt{1-(\eta L)^2}}$, then for and any $T > 0$ and $D > 0$,*

$$\max\left(r^{tan}(z_T), \frac{\textsc{Gap}(z_T)}{D}, \frac{\textsc{TGap}(z_T)}{\sqrt{N}\cdot D}\right) \leq \frac{D_0}{\eta\sqrt{T}},$$
$$\max\left(r^{tan}(z_{T+\frac{1}{2}}), \frac{\textsc{Gap}(z_{T+\frac{1}{2}})}{D}, \frac{\textsc{TGap}(z_{T+\frac{1}{2}})}{\sqrt{N}\cdot D}\right) \leq \frac{(1+\eta L)\cdot D_0}{\eta\sqrt{T}}.$$

When all the players update their actions using the OG algorithm with arbitrary starting action profiles z_0, w_0 and step-size $\eta \in (0, \frac{1}{2L})$. Denote $D_0 := \frac{\sqrt{2}(2+\eta L)}{\sqrt{1-4\cdot(\eta L)^2}} \cdot \sqrt{(4+6\eta^4L^4)\|z_0 - z^*\|^2 + (16\eta^2L^2 + 6\eta^4L^4)\|w_0 - z_0\|^2}$. Then for any $T > 0$ and $D > 0$,

$$\max\left(r^{tan}(w_{T+1}), \frac{\textsc{Gap}(w_{T+1})}{D}, \frac{\textsc{TGap}(w_{T+1})}{\sqrt{N}\cdot D}\right) \leq \frac{D_0}{\eta\sqrt{T}}.$$

For EG, according to our bound in Theorem 3, the optimal value of η is $\frac{1}{\sqrt{2}L}$, which implies that for any $D \geq 0$, $\textsc{Gap}(z_T) \leq \frac{6\|z_0 - z^*\|DL}{\sqrt{T}}$, and when $D \geq \|z_0 - z^*\|$, then $\textsc{Gap}(z_T) \leq \frac{6D^2L}{\sqrt{T}}$; For OG initialized with $w_0 = z_0$, by numerically optimizing the bound on Theorem 3, we set η to be $\frac{0.34}{L}$, in which case for $D \geq \|z_0 - z^*\|$, $\textsc{Gap}(w_{T+1}) \leq \frac{26.82D^2L}{\sqrt{T}}$. Both upper bounds matches the $\Omega(\frac{D^2L}{\sqrt{T}})$ lower bound for EG, OG, and more generally all p-SCLI algorithms [GPDO20, GPD20] in terms of the dependence on D, L, and T.

In Appendix I we show that EG and OG algorithm also have last-iterate convergence for monotone variational inequalities with respect to several standard performance measures (See Appendix D for the formal definition of monotone variational inequality).

5 Conclusion

We provide the first and tight last-iterate convergence rate of the EG and OG algorithm for smooth monotone games and Lipschitzs and monotone VIs. Our proof is based on the tangent residual, which is a new proximity measure to a Nash equilibrium, and an accompanying SOS-based analysis. We believe our techniques may be useful in the study of last-iterate convergence for other algorithms.

Acknowledgement: This work was supported by a Sloan Foundation Research Fellowship and the NSF Award CCF-1942583 (CAREER). Part of this work was done while the authors were visiting the Simons Institute for the Theory of Computing.

References

[Art27] Emil Artin. Über die zerlegung definiter funktionen in quadrate. In *Abhandlungen aus dem mathematischen Seminar der Universität Hamburg*, volume 5, pages 100–115. Springer, 1927. Issue: 1.

[BF87] LM Bregman and IN Fokin. Methods of Determining Equilibrium Situations in Zero-Sum Polymatrix Games. *Optimizatsia*, 40(57):70–82, 1987.

[Bro65] Felix E. Browder. Nonlinear monotone operators and convex sets in Banach spaces. *Bulletin of the American Mathematical Society*, 71(5):780–785, 1965.

[BS68] Haïm Brezis and Moïse Sibony. Méthodes d'approximation et d'itération pour les opérateurs monotones. *Archive for Rational Mechanics and Analysis*, 28(1):59–82, January 1968.

[CBL06] Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[CD11] Yang Cai and Constantinos Daskalakis. On Minmax Theorems for Multiplayer Games. In *Proceedings of the 2011 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Proceedings, pages 217–234. Society for Industrial and Applied Mathematics, January 2011.

[COZ22] 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 (version 2), 2022.

[DGP09] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The Complexity of Computing a Nash Equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

[Dia20] Jelena Diakonikolas. Halpern Iteration for Near-Optimal and Parameter-Free Monotone Inclusion and Strong Solutions to Variational Inequalities. *arXiv:2002.08872 [cs, math]*, April 2020. arXiv: 2002.08872.

[DISZ17] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training gans with optimism. *arXiv preprint arXiv:1711.00141*, 2017.

[DP09] Constantinos Daskalakis and Christos H. Papadimitriou. On a Network Generalization of the Minmax Theorem. In *Proceedings of the 36th Internatilonal Collogquium on Automata, Languages and Programming: Part II*, ICALP '09, pages 423–434, Berlin, Heidelberg, July 2009. Springer-Verlag.

[DP18] Constantinos Daskalakis and Ioannis Panageas. Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization. page 18 pages, 2018. Artwork Size: 18 pages Medium: application/pdf Publisher: Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany Version Number: 1.0.

[DT14] Yoel Drori and Marc Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145(1):451–482, 2014. Publisher: Springer.

[DTdB21] Radu-Alexandru Dragomir, Adrien B. Taylor, Alexandre d'Aspremont, and Jérôme Bolte. Optimal complexity and certification of bregman first-order methods. *Mathematical Programming*, 2021.

[EDMN09] Eyal Even-Dar, Yishay Mansour, and Uri Nadav. On the convergence of regret minimization dynamics in concave games. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 523–532, 2009.

[FMP18] Mahyar Fazlyab, Manfred Morari, and Victor M Preciado. Design of first-order optimization algorithms via sum-of-squares programming. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 4445–4452. IEEE, 2018.

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

[GLG21] Eduard Gorbunov, Nicolas Loizou, and Gauthier Gidel. Extragradient Method: \$O(1/K)\$ Last-Iterate Convergence for Monotone Variational Inequalities and Connections With Cocoercivity. *arXiv:2110.04261 [cs, math]*, October 2021. arXiv: 2110.04261.

[GMG+22] Baptiste Goujaud, Céline Moucer, François Glineur, Julien M. Hendrickx, Adrien B. Taylor, and Aymeric Dieuleveut. Pepit: computer-assisted worst-case analyses of first-order optimization methods in python. *CoRR*, abs/2201.04040, 2022.

[GPD20] Noah Golowich, Sarath Pattathil, and Constantinos Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. *arXiv:2010.13724 [cs, math]*, October 2020. arXiv: 2010.13724.

[GPDO20] Noah Golowich, Sarath Pattathil, Constantinos Daskalakis, and Asuman Ozdaglar. Last Iterate is Slower than Averaged Iterate in Smooth Convex-Concave Saddle Point Problems. *arXiv:2002.00057 [cs, math, stat]*, July 2020. arXiv: 2002.00057.

[HAM21] Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. Adaptive Learning in Continuous Games: Optimal Regret Bounds and Convergence to Nash Equilibrium. *arXiv:2104.12761 [cs, math]*, April 2021. arXiv: 2104.12761.

[HIMM19] 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.

[HS66] Philip Hartman and Guido Stampacchia. On some non-linear elliptic differential-functional equations. *Acta Mathematica*, 115(none):271–310, January 1966. Publisher: Institut Mittag-Leffler.

[Kim21] Donghwan Kim. Accelerated proximal point method for maximally monotone operators. *Mathematical Programming*, 190(1):57–87, November 2021.

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

[Kri64] Jean-Louis Krivine. Anneaux préordonnés. *Journal d'analyse mathématique*, 12:p. 307–326, 1964. Publisher: Springer.

[Las01] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on optimization*, 11(3):796–817, 2001. Publisher: SIAM.

[Lau09] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2009.

[LK21] 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.

[LNPW20] Qi Lei, Sai Ganesh Nagarajan, Ioannis Panageas, and Xiao Wang. Last iterate convergence in no-regret learning: constrained min-max optimization for convex-concave landscapes. *arXiv:2002.06768 [cs, stat]*, February 2020. arXiv: 2002.06768.

[LRP16] Laurent Lessard, Benjamin Recht, and Andrew Packard. Analysis and Design of Optimization Algorithms via Integral Quadratic Constraints. *SIAM Journal on Optimization*, 26(1):57–95, January 2016. arXiv: 1408.3595.

[LS67] J. L. Lions and G. Stampacchia. Variational inequalities. *Communications on Pure and Applied Mathematics*, 20(3):493–519, 1967. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160200302.

[LS19] Tengyuan Liang and James Stokes. Interaction matters: A note on non-asymptotic local convergence of generative adversarial networks. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 907–915. PMLR, 2019.

[LZMJ20] Tianyi Lin, Zhengyuan Zhou, Panayotis Mertikopoulos, and Michael Jordan. Finite-Time Last-Iterate Convergence for Multi-Agent Learning in Games. In *Proceedings of the 37th International Conference on Machine Learning*, pages 6161–6171. PMLR, November 2020. ISSN: 2640-3498.

[Mal15] Yu Malitsky. Projected Reflected Gradient Methods for Monotone Variational Inequalities. *SIAM Journal on Optimization*, 25(1):502–520, January 2015. arXiv: 1502.04968.

[MOP19] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. A Unified Analysis of Extra-gradient and Optimistic Gradient Methods for Saddle Point Problems: Proximal Point Approach. *arXiv:1901.08511 [cs, math, stat]*, September 2019. arXiv: 1901.08511.

[MOP20] Aryan Mokhtari, Asuman E. Ozdaglar, and Sarath Pattathil. Convergence rate of (1/k) for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems. *SIAM J. Optim.*, 30(4):3230–3251, 2020.

[MPP18] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2703–2717. SIAM, 2018.

[MZ19] Panayotis Mertikopoulos and Zhengyuan Zhou. Learning in games with continuous action sets and unknown payoff functions. *Math. Program.*, 173(1-2):465–507, 2019.

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

[Nes00] Yurii Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, pages 405–440. Springer, 2000.

[Par00] Pablo A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. California Institute of Technology, 2000.

[Par03] Pablo A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320, May 2003.

[PAV+13] Antonis Papachristodoulou, James Anderson, Giorgio Valmorbida, Stephen Prajna, Pete Seiler, Pablo Parrilo, Matthew M. Peet, and Declan Jagt. Sostools version 4.00 sum of squares optimization toolbox for matlab, 2013.

[PDZC20] 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.

[Pop80] 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.

[Ros65] J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):520–534, 1965.

[RTBG20] Ernest K. Ryu, Adrien B. Taylor, Carolina Bergeling, and Pontus Giselsson. Operator splitting performance estimation: Tight contraction factors and optimal parameter selection. *SIAM Journal on Optimization*, 30(3):2251–2271, 2020.

[RW09] R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*. Springer Science & Business Media, 2009.

[Sib70] Moïse Sibony. Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone. *CALCOLO*, 7(1):65–183, March 1970.

[Ste74] Gilbert Stengle. A nullstellensatz and a positivstellensatz in semialgebraic geometry. *Mathematische Annalen*, 207(2):87–97, June 1974.

[THG17a] Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Exact Worst-case Performance of First-order Methods for Composite Convex Optimization. *SIAM Journal on Optimization*, 27(3):1283–1313, January 2017. arXiv: 1512.07516.

[THG17b] Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Performance estimation toolbox (PESTO): automated worst-case analysis of first-order optimization methods. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 1278–1283. IEEE, 2017.

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

[TVT21] Sandra SY Tan, Antonios Varvitsiotis, and Vincent YF Tan. Analysis of optimization algorithms via sum-of-squares. *Journal of Optimization Theory and Applications*, 190(1):56–81, 2021.

[WLZL21a] Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Last-iterate Convergence of Decentralized Optimistic Gradient Descent/Ascent in Infinite-horizon Competitive Markov Games. *arXiv:2102.04540 [cs, stat]*, July 2021. arXiv: 2102.04540.

[WLZL21b] Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Linear Last-iterate Convergence in Constrained Saddle-point Optimization. *arXiv:2006.09517 [cs, stat]*, March 2021. arXiv: 2006.09517.

[YR21] TaeHo Yoon and Ernest 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. arXiv: 2102.07922.

[ZBLG21] Guodong Zhang, Xuchan Bao, Laurent Lessard, and Roger B. Grosse. A unified analysis of first-order methods for smooth games via integral quadratic constraints. *Journal of Machine Learning Research*, 22:103:1–103:39, 2021.

[ZMA+18] Zhengyuan Zhou, Panayotis Mertikopoulos, Susan Athey, Nicholas Bambos, Peter W. Glynn, and Yinyu Ye. Learning in games with lossy feedback. In *Annual Conference on Neural Information Processing Systems*, pages 5140–5150, 2018.

[ZMB+17] Zhengyuan Zhou, Panayotis Mertikopoulos, Nicholas Bambos, Peter W Glynn, and Claire Tomlin. Countering Feedback Delays in Multi-Agent Learning. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[ZMM+17] Zhengyuan Zhou, Panayotis Mertikopoulos, Aris L. Moustakas, Nicholas Bambos, and Peter Glynn. Mirror descent learning in continuous games. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017.

[ZMM+21] Zhengyuan Zhou, Panayotis Mertikopoulos, Aris L. Moustakas, Nicholas Bambos, and Peter Glynn. Robust Power Management via Learning and Game Design. *Operations Research*, 69(1):331–345, January 2021.

[ZY19] Guojun Zhang and Yaoliang Yu. Convergence of gradient methods on bilinear zero-sum games. *arXiv preprint arXiv:1908.05699*, 2019.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] Our results are theoretical. We state the assumptions of our results in the statements of the lemmas.
   (c) Did you discuss any potential negative societal impacts of your work? [No] Societal Impacts are discussed in Appendix A.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] All assumptions are clearly stated in the statement of the theorems (e.g. see Theorem 3).
   (b) Did you include complete proofs of all theoretical results? [Yes] We include proof for all theoretical results either in the main body or the appendix.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

Contents

A Potential Societal Impact

This work provides theoretical results for the convergence rate of natural online learning algorithms in multi-player games. Online learning in multi-player games is a mathematical model that captures the strategic interaction between agents in multi-agent systems. From this perspective, our convergence results provide new understandings of the evolution of the overall behavior of agents in multi-agent systems. More specifically, our results imply that certain natural dynamics will lead the agents' joint action profile to a stable state, i.e., a Nash equilibrium, efficiently. As a direct application, a designer of a multi-agent system can prescribe the learning algorithms studied in this paper, i.e., optimistic gradient or extragradient, to agents, so that the system stabilizes quickly. Moreover, practical applications of min-max optimization (a special case of the games studied in this paper) include Generative Adversarial Networks (GANs) and adversarial examples. Therefore, our results might also provide useful insights on the training of GANs and adversarial training. To our best knowledge, we do not envision any immediate negative societal impacts of our results, such as security, privacy, and fairness issues.

B Sum-of-Squares Programming

We first formally define SOS polynomials and SOS programs. Then we discuss how to use SOS programs to construct certificate of non-negativity to prove the monotonicity of the potential functions of EG and OG.

Sum-of-Squares (SOS) Polynomials. Let $\boldsymbol{x}$ be a set of variables. We denote the set of real polynomials in $\boldsymbol{x}$ as $\mathbb{R}[\boldsymbol{x}]$. We say that polynomial $p(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}]$ is an SOS polynomial if there exist polynomials $\{q_i(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}]\}_{i\in[M]}$ such that $p(\boldsymbol{x}) = \sum_{i\in[M]} q_i(\boldsymbol{x})^2$. We denote the set of SOS polynomials in $\boldsymbol{x}$ as $\mathrm{SOS}[\boldsymbol{x}]$. Note that any SOS polynomial is non-negative.

SOS Programs. Suppose we want to prove that a polynomial $g(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}]$ is non-negative over a semialgebraic set $\mathcal{S} = \{\boldsymbol{x} : g_i(\boldsymbol{x}) \le 0, \forall i \in [M], h_i(\boldsymbol{x}) = 0, \forall i \in [N]\}$, where each $g_i(\boldsymbol{x})$ ($h_i(\boldsymbol{x})$ resp.) is also a polynomial. One way is to construct a *certificate of non-negativity*, for example, by providing a set of nonnegative coefficients $\{p_i\}_{i\in[M]} \in \mathbb{R}^M_{\ge 0}$ and $\{q_i\}_{i\in[\mathcal{N}]} \in \mathbb{R}^N$ such that $g(\boldsymbol{x}) + \sum_{i\in[M]} p_i \cdot g_i(\boldsymbol{x}) + \sum_{i\in[N]} q_i \cdot h_i(\boldsymbol{x})$ is a SOS polynomial. Surprisingly, if $g(\boldsymbol{x})$ is indeed non-negative over $\mathcal{S}$, a certificate of non-negativity always exists as guaranteed by a foundational result in real algebraic geometry – the Krivine-Stengle Positivestellensatz [Kri64, Ste74], a generalization of Artin's resolution of Hilbert's 17th problem [Art27]. Note that, it is sometimes necessary to allow more sophisticated forms of certificates than in the example above, e.g., replacing each coefficient p_i with a SOS polynomial $p_i(\boldsymbol{x})$, etc. The complexity of a certificate is parametrized by the highest degree of the polynomial involved. The SOS programming consists of a hierarchy of algorithms, where the d-th hierarchy is an algorithm that searches for a certificate of non-negativity up to degree $2d$ based on semidefinite programming.

In Figure 1 we present a generic formulation of a degree-$2d$ SOS program. The SOS program takes three kinds of input, a polynomial $g(\boldsymbol{x})$, sets of polynomials $\{g_i(\boldsymbol{x})\}_{i\in[M]}$ and $\{h_i(\boldsymbol{x})\}_{i\in[N]}$. Each polynomial in $\{g(\boldsymbol{x})\} \cup \{g_i(\boldsymbol{x})\}_{i\in[M]} \cup \{h_i(\boldsymbol{x})\}_{i\in[N]}$ has degree of at most $2d$. The SOS program searches for an SOS polynomial in the set of polynomials $\Sigma = \{g(\boldsymbol{x}) + \sum_{i\in[M]} p_i(\boldsymbol{x}) \cdot g_i(\boldsymbol{x}) + \sum_{i\in[N]} q_i(\boldsymbol{x}) \cdot h_i(\boldsymbol{x})\}$, where $\{p_i(\boldsymbol{x})\}_{i\in[M]}$ and $\{q_i(\boldsymbol{x})\}_{i\in[N]}$ are polynomials in $\boldsymbol{x}$. More precisely for each $i \in [M]$, $p_i(\boldsymbol{x})$ is an SOS polynomial with degree at most $2d - \deg(g_i(\boldsymbol{x}))$. For each $i \in [N]$, $q_i(\boldsymbol{x})$ is a (not necessarily SOS) polynomial with degree at most $2d - \deg(h_i(\boldsymbol{x}))$. Note that any polynomial in set Σ is at most degree $2d$. In our applications, we choose $\{g_i(\boldsymbol{x})\}_{i\in[M]}$ to be non-positive polynomials and $\{h_i(\boldsymbol{x})\}_{i\in[N]}$ to be polynomials that are equal to 0. Any feasible solution to the program certifies the non-negativity of $g(\boldsymbol{x})$. We used SOSTOOLS package in MATLAB to solve any SOS program encountered in this paper [PAV+13].

Input Fixed Polynomials.

- Polynomial $g(\boldsymbol{x})$
- Polynomial $g_i(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}]$ for all $i \in [M]$.
- Polynomial $h_i(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}]$ for all $i \in [N]$.

Decision Variables of the SOS Program:

- $p_i(\boldsymbol{x}) \in \mathrm{SOS}[\boldsymbol{x}]$ is an SOS polynomial with degree at most $2d - \deg(g_i)$, for all $i \in [M]$.
- $q_i(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}]$ is a polynomial with degree at most $2d - \deg(h_i)$, for all $i \in [N]$.

Constraints of the SOS Program:

$$g(\boldsymbol{x}) + \sum_{i\in[M]} p_i(\boldsymbol{x}) \cdot g_i(\boldsymbol{x}) + \sum_{i\in[N]} q_i(\boldsymbol{x}) \cdot h_i(\boldsymbol{x}) \in \mathrm{SOS}[\boldsymbol{x}]$$

Figure 1: Generic degree $2d$ SOS program.

SOS-based Analysis of EG and OG We mainly discuss the analysis of EG here, as the analysis of OG is similar and also based on SOS programming. At the core of our analysis of the EG algorithm lies the monotonicity of the squared tangent residual, which can be formulated as the non-negativity of a **degree-4 polynomial** in the iterates.[9] Our original proof directly applies SOS programming to certify the non-negativity of this degree-4 polynomial. The certificate is rather complex and involves a polynomial identity of a degree-8 polynomial in 27 variables, which we discover by solving a degree-8 SOS program. Interested readers can find the proof in version 2 of [COZ22] in arXiv. In this version, we include a simplified proof. By introducing *auxiliary vectors* that are not part of the update rule of EG, we provide an equivalent formulation of the squared tangent residual (Lemma 13) that is a degree-2 polynomial, which allows us to prove the monotonicity of the squared tangent residual using a degree-2 SOS program. Detailed proof can be found in Appendix G.3.

For OG, we are not able to show that the squared tangent residual is monotone. Inspired by the adaptive potential proof in [GPD20], we suspect that some extra correction term is needed to construct the potential function. Instead of trying to devise such a correction term manually, we manage to directly find one by searching over a family of performance measures using SOS programming. The search we perform is heuristic but might be helpful to discover potential functions in other problems. See Section 4.2 for a more detailed discussion.

[9]The tangent residual is not a polynomial, but the squared tangent residual is a degree-4 polynomial

C Additional Related Work – other Computer-Aided Proofs

A powerful computer-aided proof framework – the *performance estimation problem (PEP)* technique (e.g., [DT14, THG17b]) is widely applied to analyze first-order iterative methods. Indeed, the last-iterate convergence rate of EG in the unconstrained setting by [GLG21] is obtained via the PEP technique. Although the PEP framework can handle projections [THG17a, RTBG20, GMG+22, DTdB21], the main challenge for applying it to the constrained setting is that, the PEP framework requires the performance measures to be polynomials of degree 2 or less (see e.g., [THG17a]).[10] In fact, solving the PEP is equivalent to solving a degree-2 SOS program, which can be viewed as the dual of the PEP [TVT21]. In the unconstrained setting for EG, the performance measure is a degree-2 polynomial – the squared norm of the operator, and that is why one can either use the PEP (as in [GLG21]) or a degree-2 SOS to certify its monotonicity (Theorem 4). In the constrained setting for EG, we use the squared tangent residual to measure the algorithm's progress, which in our original formulation is a degree 4 polynomial, making the PEP framework not directly applicable.[11] As the SOS approach can accommodate polynomial objectives and constraints of any degree, we could directly apply it to certify the monotonicity of the tangent residual in the constrained setting, although the resulting proof is complex. With the new formulation of the squared tangent residual (Lemma 13), we manage to simplify our proof and derive it using a degree-2 SOS program. We believe an interesting future direction is to understand whether there are natural settings in optimization where degree-2 SOS programs are provably insufficient and higher degree SOS programs are necessary.

[LRP16] analyze first-order iterative algorithms for convex optimization using a technique inspired by the stability analysis from control theory. They model first-order iterative algorithms using discrete-time dynamical systems and search over quadratic potential functions that satisfy a set of Integral Quadratic Constraints (IQC). [ZBLG21] extend the IQC framework to study smooth and *strongly* monotone VIs in the unconstrained setting.

SOS programming has been employed in the design and analysis of algorithms in convex optimization. To the best of our knowledge, these results only concern minimization of smooth and strongly-convex functions in the unconstrained setting. [FMP18] propose a framework to search the optimal parameters of the algorithm, e.g., step size. They use SOS programming to search over quadratic potential functions and parameters of the algorithm with the goal of optimizing the exponential decay rate of the potential function. [TVT21] proposes to use SOS programming to study the convergence rates of first-order methods in unconstrained convex optimization.

D Additional Preliminaries

We use $z[i]$ to denote the i-th coordinate of $z \in \mathbb{R}^n$ and e_i to denote the unit vector such that $e_i[j] := \mathbb{1}[i = j]$, the dimension of e_i is going to be clear from context. For $z \in \mathbb{R}^n$ and $D > 0$, we use $\mathcal{B}(z, D) = \{z' \in \mathbb{R}^n : \|z' - z\| \leq D\}$ to denote the ball of radius D, centered at z.

Definition 5 (Variational Inequality)**.** *Given a closed convex set $\mathcal{Z} \subseteq \mathbb{R}^n$ and an operator $F : \mathcal{Z} \to \mathbb{R}^n$, a variational inequality problem is defined as follows: find $z^* \in \mathcal{Z}$ such that*

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

Min-Max Saddle Points. A special case of the variational inequality problem is the constrained min-max problem $\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y)$, where $\mathcal{X}$ and $\mathcal{Y}$ are closed convex sets

[10] More specifically, the PEP framework requires the performance measure as well as the constraints to be linear in (i) the function values at the iterates and (ii) the Gram matrix of a set of vectors consisting of the iterates and their gradients.

[11] The tangent residual is the square root of a rational function and can only be even harder to handle.

in $\mathbb{R}^n$, and $f(\cdot,\cdot)$ is smooth, convex in x, and concave in y. It is well known that if one set $F(x,y) = \begin{pmatrix} \nabla_x f(x,y) \\ -\nabla_y f(x,y) \end{pmatrix}$, then $F(x,y)$ is a monotone and Lipschitz operator [?].

Definition 6 (Gap Function for monotone VIs). *Similar to games, for a monotone VI with operator $F : \mathcal{Z} \to \mathbb{R}^n$ on closed convex set $\mathcal{Z}$, a standard way to measure the proximity of $z \in \mathcal{Z}$ to the solution of the monotone VI, is through the gap function for VIs: $\max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \langle F(z), z - z' \rangle$. We abuse notation and for a monotone operator F and closed convex set $\mathcal{Z}$, we denote by $\textsc{Gap}_{F,\mathcal{Z},D}(z) = \max_{z' \in \mathcal{Z} \cap \mathcal{B}(z,D)} \langle F(z), z - z' \rangle$.*

When F, $\mathcal{Z}$ and D are clear from context, we omit subscripts and write the gap function for VIs at z as $\textsc{Gap}(z)$. Moreover, we refer to the gap function for VIs, simply as the gap function, when there is no ambiguity if we are refer to the gap function for games or the gap function for VIs.

Lemma 4. *[Restatement of Lemma 2 for VIs] Let $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone operator on convex closed set $\mathcal{Z}$. For $z \in \mathcal{Z}$, we have $\textsc{Gap}_{F,\mathcal{Z},D}(z) \le D \cdot r^{tan}_{(F,\mathcal{Z})}(z)$.*

Proof. The proof follows in the exact same way as the proof of Lemma 2 for the gap function for monotone games (see Appendix E.2). □

D.1 Remark about choice of D in Definition 2

Remark 5. *Consider a smooth monotone game $\mathcal{G}$, and let $\{z_k^{EG}, z_{k+\frac{1}{2}}^{EG}\}_{k\ge 0}$ ($\{z_k^{OG}, w_k^{OG}\}_{k\ge 0}$ resp.) be the action profile when all players update their actions using the EG (OG resp.) algorithm and let z^* be a Nash equilibrium of $\mathcal{G}$. Sometimes the gap function is defined to allow z' to take value in $\mathcal{Z} \cap \mathcal{B}(z^*, \Theta(\|z^* - z_0^{EG}\|))$ for the EG algorithm and $\mathcal{Z} \cap \mathcal{B}(z^*, \Theta(\|z^* - z_0^{OG}\| + \|z_0^{OG} - w_0^{OG}\|))$ for the OG algorithm.*

In Lemma 8, by choosing the step size appropriately, we show that

$$\begin{aligned}
\max_{k\ge 0} \left(\left\| z_k^{EG} - z^* \right\|, \left\| z_{k+\frac{1}{2}}^{EG} - z^* \right\| \right) &= O\left(\left\| z_0^{EG} - z^* \right\| \right), \\
\max_{k\ge 0} \left(\left\| z_k^{OG} - z^* \right\|, \left\| w_k^{OG} - z^* \right\| \right) &= O\left(\max(\left\| z_0^{OG} - z^* \right\|, \left\| w_0^{OG} - z_0^{OG} \right\|) \right).
\end{aligned}$$

Thus, the set $\{z_k^{EG}, z_{k+\frac{1}{2}}^{EG}\}_{k\ge 0}$ is contained in $\mathcal{B}(z^*, \Theta(\|z^* - z_0^{EG}\|))$ and set $\{z_k^{OG}, w_k^{OG}\}_{k\ge 0}$ is contained in $\mathcal{B}(z^*, \Theta(\|z^* - z_0^{OG}\| + \|z_0^{OG} - w_0^{OG}\|))$.

D.2 Equivalent Definitions of the Tangent Residual

In Lemma 5 we present several equivalent formulations of the tangent residual.

Lemma 5. *Let $\mathcal{Z}$ be a closed convex set and $F : \mathcal{Z} \to \mathbb{R}^n$ be an operator. Denote $N_{\mathcal{Z}}(z)$ the normal cone of z and $J_{\mathcal{Z}}(z) := \{z\} + T_{\mathcal{Z}}(z)$, where $T_{\mathcal{Z}}(z) = \{z' \in \mathbb{R}^n : \langle z', a \rangle \le 0, \forall a \in N_{\mathcal{Z}}(z)\}$ is the tangent cone of z. Then all of the following quantities are equivalent:*

1. $\sqrt{\|F(z)\|^2 - \max_{\substack{a \in \widehat{N}_{\mathcal{Z}}(z), \\ \langle F(z), a \rangle \le 0}} \langle F(z), a \rangle^2}$
2. $\min_{\substack{a \in \widehat{N}_{\mathcal{Z}}(z), \\ \langle F(z), a \rangle \le 0}} \|F(z) - \langle F(z), a \rangle \cdot a\|$
3. $\left\| \Pi_{T_{\mathcal{Z}}(z)} \left[-F(z) \right] \right\|$
4. $\left\| \Pi_{J_{\mathcal{Z}}(z)} \left[z - F(z) \right] - z \right\|$
5. $\left\| -F(z) - \Pi_{N_{\mathcal{Z}}(z)} \left[-F(z) \right] \right\|$
6. $\min_{a \in N_{\mathcal{Z}}(z)} \|F(z) + a\|$

Proof. **(quantity 1 = quantity 2).** Observe that

$$\min_{\substack{a\in\widehat{N}_{\mathcal{Z}}(z),\\ \langle F(z),a\rangle\leq 0}} \|F(z)-\langle F(z),a\rangle\cdot a\|^2 = \|F(z)\|^2 - \max_{\substack{a\in\widehat{N}_{\mathcal{Z}}(z),\\ \langle F(z),a\rangle\leq 0}} \langle F(z),a\rangle^2\cdot\Big(2-\|a\|^2\Big).$$

Therefore, it is enough to show that $\max_{\substack{a\in\widehat{N}_{\mathcal{Z}}(z),\\ \langle F(z),a\rangle\leq 0}} \langle F(z),a\rangle^2\cdot(2-\|a\|^2) = \max_{\substack{a\in\widehat{N}_{\mathcal{Z}}(z),\\ \langle F(z),a\rangle\leq 0}} \langle F(z),a\rangle^2$. If $\widehat{N}_{\mathcal{Z}}(z) = \{(0,\ldots,0)\}$, then the equality holds trivially. Now we assume that $\{(0,\ldots,0)\}\subsetneq\widehat{N}_{\mathcal{Z}}(z)$ and consider any $a\in\widehat{N}_{\mathcal{Z}}(z)\backslash(0,\ldots,0)$. Let $c\in\Big[1,\frac{1}{\|a\|}\Big]$. By Definition 3, $\|a\|\leq 1$, which implies that $c\cdot a\in\widehat{N}_{\mathcal{Z}}(z)$. We try to maximize the following objective

$$\langle F(z),c\cdot a\rangle^2\cdot\Big(2-c^2\|a\|^2\Big) = \frac{\langle F(z),a\rangle^2}{\|a\|^2}\cdot c^2\|a\|^2\cdot\Big(2-c^2\|a\|^2\Big).$$

One can easily verify that function $c^2\|a\|^2\cdot(2-c^2\|a\|^2)$ is maximized when $c^2\|a\|^2=1\Leftrightarrow c=\frac{1}{\|a\|}$. Thus when $\{(0,\ldots,0)\}\subsetneq\widehat{N}_{\mathcal{Z}}(z)$,

$$\begin{aligned}
\max_{\substack{a\in\widehat{N}_{\mathcal{Z}}(z),\\ \langle F(z),a\rangle\leq 0}} \langle F(z),a\rangle^2\cdot\Big(2-\|a\|^2\Big) &= \max_{\substack{a\in\widehat{N}_{\mathcal{Z}}(z),\\ \langle F(z),a\rangle\leq 0,\\ \|a\|=1}} \langle F(z),a\rangle^2\cdot\Big(2-\|a\|^2\Big)\\
&= \max_{\substack{a\in\widehat{N}_{\mathcal{Z}}(z),\\ \langle F(z),a\rangle\leq 0,\\ \|a\|=1}} \langle F(z),a\rangle^2\\
&= \max_{\substack{a\in\widehat{N}_{\mathcal{Z}}(z),\\ \langle F(z),a\rangle\leq 0}} \langle F(z),a\rangle^2,
\end{aligned}$$

which concludes the proof.

(quantity 3 = quantity 4). By definition, $J_{\mathcal{Z}}(z)=\{z\}+T_{\mathcal{Z}}(z)$. Thus we have

$$\Big\|\Pi_{J_{\mathcal{Z}}(z)}\Big[z-F(z)\Big]-z\Big\| = \Big\|\Pi_{T_{\mathcal{Z}}(z)}\Big[-F(z)\Big]\Big\|.$$

(quantity 4 = quantity 5). By definition, the tangent cone $T_{\mathcal{Z}}(z)$ is the polar cone of the normal cone $N_{\mathcal{Z}}(z)$. Since $N_{\mathcal{Z}}(z)$ is a closed convex cone, by Moreau's decomposition theorem, we have for any vector $x\in\mathbb{R}^n$,

$$x=\Pi_{N_{\mathcal{Z}}(z)}(x)+\Pi_{T_{\mathcal{Z}}(z)}(x),\qquad \Big\langle\Pi_{N_{\mathcal{Z}}(z)}(x),\Pi_{T_{\mathcal{Z}}(z)}(x)\Big\rangle=0.$$

Thus it is clear that we have

$$\begin{aligned}
\Big\|\Pi_{J_{\mathcal{Z}}(z)}\Big[z-F(z)\Big]-z\Big\| &= \Big\|\Pi_{T_{\mathcal{Z}}(z)}\Big[-F(z)\Big]\Big\|\\
&= \Big\|-F(z)-\Pi_{N_{\mathcal{Z}}(z)}\Big[-F(z)\Big]\Big\|.
\end{aligned}$$

(quantity 5 = quantity 6). Denote $a^*:=\Pi_{N_{\mathcal{Z}}(z)}\Big[-F(z)\Big]$. By definition of projection, we have

$$a^*=\operatorname*{argmin}_{a\in N_{\mathcal{Z}}(z)}\|F(z)+a\|^2.$$

Thus

$$\Big\|-F(z)-\Pi_{N_{\mathcal{Z}}(z)}\Big[-F(z)\Big]\Big\|^2 = \|F(z)+a^*\|^2 = \min_{a\in N_{\mathcal{Z}}(z)}\|F(z)+a\|^2.$$

(quantity 6 = quantity 2). Let $a \in N_{\mathcal{Z}}(z)$ such that $\langle F(z), a\rangle \leq 0$. Observe that for any $c \geq 0$, $c \cdot a \in N_{\mathcal{Z}}(z)$ and $\langle F(z), c \cdot a\rangle \leq 0$. Consider the following minimization problem,

$$g(a) = \min_{c \geq 0} \|F(z) + c \cdot a\|$$

By taking first-order optimality conditions, one can easily verify that when $a \neq (0, \ldots, 0)$, $g(a) = \|F(z) + \langle F(z), \frac{a}{\|a\|}\rangle \frac{a}{\|a\|}\|$ and $g(0, \ldots, 0) = \|F(z)\|$. Since $N_{\mathcal{Z}}(z)$ is a cone and for any a and $c \in \arg\min_{c \geq 0} \|F(z) + c \cdot a\|$, we have that $\|c \cdot a\| \leq 1$, we infer that

$$\min_{a \in N_{\mathcal{Z}}(z)} \|F(z) + a\| = \min_{a \in N_{\mathcal{Z}}(z)} g(a) = \min_{a \in \widehat{N}_{\mathcal{Z}}(z)} \|F(z) + \langle F(z), a\rangle \cdot a\|$$

Observe that for any $a \in \widehat{N}_{\mathcal{Z}}(z)$ such that $\langle F(z), a\rangle \geq 0$, then $\|F(z) + \langle F(z), a\rangle \cdot a\| \geq \|F(z)\|$. Since $g(0, \ldots, 0) = \|F(z)\|$ we have that,

$$\min_{a \in \widehat{N}_{\mathcal{Z}}(z)} \|F(z) + \langle F(z), a\rangle \cdot a\| = \min_{\substack{a \in \widehat{N}_{\mathcal{Z}}(z), \\ \langle F(z), a\rangle \leq 0}} \|F(z) + \langle F(z), a\rangle \cdot a\|,$$

which concludes the proof.

□

In the following Lemma, we show a useful property of the tangent residual that we use repeatedly.

Lemma 6. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set and $F : \mathcal{Z} \to \mathbb{R}^n$ be an operator. Let $\eta > 0$ and $z_1, z_2, z_3 \in \mathcal{Z}$ be three points such that $z_1 = \Pi_{\mathcal{Z}}[z_2 - \eta F(z_3)]$, then we have*

$$r^{tan}(z_1) \leq \left\| \frac{z_2 - z_1}{\eta} + F(z_1) - F(z_3) \right\|.$$

Proof. Since $z_1 = \Pi_{\mathcal{Z}}[z_2 - \eta F(z_3)]$, we have $\frac{z_2 - \eta F(z_3) - z_1}{\eta} = \frac{z_2 - z_1}{\eta} - F(z_3) \in N_{\mathcal{Z}}(z_1)$. Then by item 6 in Lemma 5 we have

$$r^{tan}(z_1) = \min_{c \in N_{\mathcal{Z}}(z_1)} \|F(z_1) + c\| \leq \left\| \frac{z_2 - z_1}{\eta} + F(z_1) - F(z_3) \right\|.$$

□

E Missing Proofs and Details from Section 3

E.1 The Natural Residual and Its Relation to the Tangent Residual

We formally define the natural residual for monotone operators over closed convex sets in Definition 7, and show in Lemma 7 how it is related to the tangent residual.

Definition 7. *Let $\mathcal{Z}$ be a closed convex set in $\mathbb{R}^n$ and consider a monotone operator $F : \mathcal{Z} \to \mathbb{R}^n$. The natural residual at $z \in \mathcal{Z}$ is defined as follows*

$$r^{nat}_{(F,\mathcal{Z})}(z) = \|z - \Pi_{\mathcal{Z}}(z - F(z))\|.$$

Given a monotone game $\mathcal{G}$, an action profile z^* is a Nash equilibrium iff $r^{nat}_{(F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}(z^*) = 0$. In Lemma 7, we show that the tangent residual upper bounds the the natural residual. See Figure 2 for illustration of how the tangent residual relates to the natural residual.

Lemma 7. *Let $\mathcal{Z}$ be a closed convex set and consider a monotone operator $F : \mathcal{Z} \to \mathbb{R}^n$. For any $z \in \mathcal{Z}$, $r^{tan}_{(F,\mathcal{Z})}(z) \geq r^{nat}_{(F,\mathcal{Z})}(z)$.*

Proof. Note that for any $c \in N_{\mathcal{Z}}(z)$, $\Pi_{\mathcal{Z}}(z+c) = z$. Thus for any $c \in N_{\mathcal{Z}}(z)$, we have

$$\begin{aligned} r^{nat}_{(F,\mathcal{Z})}(z) &= \|z - \Pi_{\mathcal{Z}}(z - F(z))\| \\ &= \|\Pi_{\mathcal{Z}}(z+c) - \Pi_{\mathcal{Z}}(z - F(z))\| \\ &\le \|F(z) + c\|, \end{aligned}$$

where the last inequality holds because $\Pi_{\mathcal{Z}}(\cdot)$ is non-expansive. Thus $r^{nat}_{(F,\mathcal{Z})}(z) \le \min_{c \in N_{\mathcal{Z}}(z)} \|F(z) + c\| = r^{tan}_{(F,\mathcal{Z})}(z)$ by item 6 in Lemma 5. $\square$

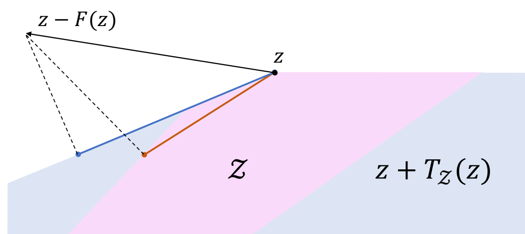

Figure 2: Illustration of the tangent residual and the natural residual. The blue line represents the tangent residual and the red line represents the natural residual. It is clear that the tangent residual upper bounds the natural residual.

Due to the above lemma, an upper bound of the tangent residual is also an upper bound of the natural residual.

E.2 Proof of Lemma 2

Proof of Lemma 2: We first show that $\text{GAP}_{\mathcal{G},D}(z) \le D \cdot r^{tan}_{\mathcal{G}}(z)$.

If $\langle a, F(z)\rangle \ge 0$ for all $a \in \widehat{N}(z)$, then by item 1 of Lemma 5 we have $r^{tan}(z) = \|F(z)\|$. Thus for any $z' \in \mathcal{Z}$, by Cauchy-Schwarz inequality, we have

$$\langle F(z), z - z'\rangle \le \|F(z)\|\|z - z'\| \le D \cdot r^{tan}(z).$$

Otherwise, by item 2 in Lemma 5 there exists $a \in \widehat{N}(z)$ such that $\|a\| = 1$, $\langle a, F(z)\rangle < 0$ and $r^{tan}(z) = \|F(z) - \langle a, F(z)\rangle a\|$. Then for any $z' \in \mathcal{Z}$, we have

$$\begin{aligned} \left\langle F(z), z - z'\right\rangle &= \left\langle F(z) - \langle a, F(z)\rangle a, z - z'\right\rangle + \langle a, F(z)\rangle \cdot \left\langle a, z - z'\right\rangle \\ &\le \left\langle F(z) - \langle a, F(z)\rangle a, z - z'\right\rangle \\ &\le \|F(z) - \langle a, F(z)\rangle a\|\|z - z'\| \\ &\le D \cdot r^{tan}(z), \end{aligned}$$

where we use $\langle a, F(z)\rangle < 0$ and $\langle a, z - z'\rangle \ge 0$ in the first inequality and Cauchy-Schwarz inequality in the second inequality. Thus for all $D > 0$,

$$\text{GAP}_{\mathcal{G},D}(z) \le D \cdot r^{tan}_{\mathcal{G}}(z). \tag{5}$$

Now we prove that $\text{TGAP}_{\mathcal{G},D}(z) \le \sqrt{N} \cdot D \cdot r^{tan}_{\mathcal{G}}(z)$. Let $z^{*(i)} = \min_{z'^{(i)} \in Z^{(i)} \cap \mathcal{B}(z^{(i)}, D)} f(z^{(i)}, z^{(-i)})$ and $z^* = (z^{*(1)}, \dots, z^{*(N)})$. By monotonicity of F, we have that for any $i \in \mathcal{N}$ and $z'^{(i)} \in \mathcal{Z}^{(i)}$

$$\left\langle F(z'^{(i)}, z^{(-i)}) - F(z), (z'^{(i)}, z^{(-i)}) - z\right\rangle = \left\langle \nabla_{z^{(i)}} f^{(i)}(z'^{(i)}, z^{(-i)}) - \nabla_{z^{(i)}} f^{(i)}(z), z'^{(i)} - z^{(i)}\right\rangle \ge 0,$$

Moreover, since $f^{(i)}$ is a continues differentiable function, then $g(z^{(i)}) = f(z'^{(i)}, z^{(-i)}) : \mathcal{Z}^{(i)} \to \mathbb{R}$ is a convex function, which further implies that

$$f^{(i)}(z) - f(z^{*(i)}, z^{(-i)}) \le \left\langle \nabla_{z^{(i)}} f^{(i)}(z), z^{(i)} - z^{*(i)}\right\rangle.$$

Thus,

$$\begin{aligned}\text{TGAP}_{\mathcal{G},D}(z) &= \sum_{i\in\mathcal{N}} f^{(i)}(z) - f(z^{*(i)}, z^{(-i)}) \\ &\le \langle F(z), z - z^* \rangle \\ &\le \max_{z'\in\mathcal{Z}\cap\mathcal{B}(z,\sqrt{N}\cdot D)} \langle F(z), z - z' \rangle = \text{GAP}_{\mathcal{G},\sqrt{N}\cdot D}(z).\end{aligned}$$

The second inequality follows because for each $i \in \mathcal{N}$, $\|z^{(i)} - z^{*(i)}\| \le D$, which implies that $\|z - z^*\| = \sqrt{\sum_{i\in\mathcal{N}} \|z^{(i)} - z^{*(i)}\|^2} \le \sqrt{N} \cdot D$. The proof follows by Inequality (5). ■

F Missing Proofs in Section 4

In this section, we present the missing proofs in Section 4. Finding a Nash Equilibrium for a smooth monotone game is a special instance of solving a monotone VI (Definition 5). Thus, for simplicity and technical convenience, we show the last-iterate convergence rate of EG (OG resp.) for monotone VIs in Appendix G (Appendix H resp.) with respect to the tangent residual (Definition 4), the gap function for VIs (Definition 6), and the natural residual (Definition 7). In this section, we show how to apply the last-iterate convergence rate of EG (Appendix G) and OG (Appendix H) for smooth monotone games and we also show last-iterate convergence rates for some additional performance measure.

Proof of Lemma 3: Consider an instance $(\mathcal{I})$ of the monotone VI on operator $F_{\mathcal{G}}$ on closed convex set $\mathcal{Z}_{\mathcal{G}}$. By Lemma 1, z^* is a solution to the monotone VI $(\mathcal{I})$.

Observe that the updates of EG (OG resp.) algorithm with step-size η on the monotone VI $(\mathcal{I})$ coincide with the action profile when all players update their actions using EG (OG resp.) algorithm with step-size η. Thus, the proof follows by Lemma 12 and Corollary 4. ■

Proof of Theorem 2: Consider an instance $(\mathcal{I})$ of the monotone VI on operator $F_{\mathcal{G}}$ on closed convex set $\mathcal{Z}_{\mathcal{G}}$. By Lemma 1, z^* is a solution to the monotone VI $(\mathcal{I})$.

Observe that the updates of EG (OG resp.) algorithm with step-size η on the monotone VI $(\mathcal{I})$ coincide with the action profile when all players update their actions using EG (OG resp.) algorithm with step-size η. Thus, the proof follows by Theorem 5 and Theorem 7. ■

Proof of Theorem 3: Consider an instance $(\mathcal{I})$ of the monotone VI on operator $F_{\mathcal{G}}$ on closed convex set $\mathcal{Z}_{\mathcal{G}}$. By Lemma 1, z^* is a solution to the monotone VI $(\mathcal{I})$.

Observe that the updates of EG (OG resp.) algorithm with step-size η on the monotone VI $(\mathcal{I})$ coincide with the action profile when all players update their actions using EG (OG resp.) algorithm with step-size η.

Thus, when all the players update their strategies using the EG algorithm, by Theorem 6 and Lemma 9 we have that $r^{tan}_{\mathcal{G}}(z_T) = r^{tan}_{(F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}(z_T) \le \frac{1}{\sqrt{T}} \frac{3D||z_0 - z^*||}{\eta\sqrt{1-(\eta L)^2}}$ and $r^{tan}_{\mathcal{G}}(z_{T+\frac{1}{2}}) = r^{tan}_{(F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}(z_{T+\frac{1}{2}}) \le \frac{1}{\sqrt{T}} \frac{(1+\eta L)\cdot 3D||z_0 - z^*||}{\eta\sqrt{1-(\eta L)^2}}$. When all the players update their strategies using the OG algorithm, by Theorem 8 we have that $r^{tan}_{\mathcal{G}}(w_{T+1}) = r^{tan}_{F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}}}(w_{T+1}) \le \frac{1}{\sqrt{T}} \cdot \frac{\sqrt{2}(2+\eta L)\cdot\sqrt{(4+6\eta^4 L^4)\|z_0 - z^*\|^2 + (16\eta^2 L^2 + 6\eta^4 L^4)\|w_0 - z_0\|^2}}{\eta\cdot\sqrt{1-4\cdot(\eta L)^2}}$. The proof concludes by Lemma 2. ■

F.1 Bounded Iterates of EG and OG

Lemma 8. *Let $\mathcal{G} = (\mathcal{N}, \{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}, \{f^{(i)}\}_{i\in[N]})$ be an L-smooth and monotone game where $\{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}$ are closed convex sets and let z^* be a Nash Equilibrium of $\mathcal{G}$. Assume that all the players update their actions using the EG algorithm with arbitrary starting action profile z_0 and step-size*

$\eta \in (0, \frac{1}{L})$. *Then for and any $k \geq 0$,*

$$\|z_k - z^*\| \leq \|z_0 - z^*\|,$$
$$\left\|z_{k+\frac{1}{2}} - z^*\right\| \leq \left(1 + \frac{1}{\sqrt{1-\eta^2 L^2}}\right)\|z_0 - z^*\|.$$

Assume that all the players update their actions using the OG algorithm with arbitrary starting action profiles z_0,w_0 and step-size $\eta \in (0, \frac{1}{2L})$. Then for any $k \geq 1$,

$$\|z_k - z^*\| \leq \sqrt{\frac{1-2\eta^2L^2}{1-4\eta^2L^2}\|z_0 - z^*\|^2 + \frac{2\eta^2L^2}{1-4\eta^2L^2}\|w_0 - z_0\|^2},$$
$$\|w_k - z^*\| \leq 2 \cdot \sqrt{\frac{1-2\eta^2L^2}{1-4\eta^2L^2}\|z_0 - z^*\|^2 + \frac{2\eta^2L^2}{1-4\eta^2L^2}\|w_0 - z_0\|^2}.$$

Proof. Consider an instance $(\mathcal{I})$ of the monotone VI on operator $F_{\mathcal{G}}$ on closed convex set $\mathcal{Z}_{\mathcal{G}}$. By Lemma 1, z^* is a solution to the monotone VI $(\mathcal{I})$.

Observe that the updates of EG (OG resp.) algorithm with step-size η on the monotone VI $(\mathcal{I})$ coincide with the strategy profiles when all players update their strategies using EG (OG resp.) algorithm with step-size η. Thus, the proof follows by Corollary 1 and Corollary 3. $\square$

F.2 Auxiliary Lemma

Lemma 9. *Let $\mathcal{G} = (\mathcal{N}, \{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}, \{f^{(i)}\}_{i\in[N]})$ be an L-smooth and monotone game where $\{\mathcal{Z}^{(i)}\}_{i\in\mathcal{N}}$ are closed convex sets. Assume that all the players update their actions using the EG algorithm with arbitrary starting action profile z_0 and step-size η, then for any $k \geq 0$, $r^{tan}_{(F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}(z_{k+\frac{1}{2}}) \leq (1+\eta L) r^{tan}_{(F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}(z_k)$.*

Proof. The proof follows from the following sequence of inequalities,

$$\begin{aligned}
\eta r^{tan}_{(F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}\left(z_{k+\frac{1}{2}}\right) \leq& \left\|z_k - z_{k+\frac{1}{2}} + \eta F(z_{k+\frac{1}{2}}) - \eta F(z_k)\right\| \\
\leq& \left\|z_k - z_{k+\frac{1}{2}}\right\| + \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_k)\right\| \\
\leq& (1+\eta L)\left\|z_k - z_{k+\frac{1}{2}}\right\| \\
=& (1+\eta L) r^{nat}_{(\eta F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}(z_k) \\
\leq& (1+\eta L) r^{tan}_{(\eta F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}(z_k) \\
=& (1+\eta L)\eta r^{tan}_{(F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}(z_k).
\end{aligned}$$

The first inequality follows by Lemma 6, the third inequalit follows by L-lipschitzness of F. The first equality follows from the fact that $z_{k+\frac{1}{2}} = \Pi_{\mathcal{Z}}(z_k - \eta F(z_k))$ and Definition 7. The fourth inequality follows by Lemma 7. The last equality follows by Definition 4. $\square$

G Missing Details for the Analysis of the Extragradient Algorithm

In this section, we provide the last-iterate convergence rate of the EG algorithm for monotone VIs (Definition 5). We establish last-iterate convergence rate w.r.t. the gap function for VIs (Definition 6), the natural residual (Definition 7) and the tangent residual (Definition 4). For the rest of this section, we abuse notation and refer to the gap function for VIs as the gap function. We show in Appendix F (Appendix I resp.) last-iterate convergence rates for additional performance measures for smooth monotone games (monotone VIs resp.).

We prove last-iterate convergence rate for EG w.r.t. the gap function, natural residual and tangent residual in Theorem 6 at Appendix G.3. The last-iterate convergence rate for the performance measures we mentioned follow from the last-iterate convergence rate of the tangent residual $r^{tan}(z_T)$.

Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set and $F : \mathcal{Z} \to \mathbb{R}$ be an operator. Let $z_0 \in \mathcal{Z}$ be an arbitrary starting point and $\{z_k, z_{k+\frac{1}{2}}\}_{k\geq 0}$ be the iterates of the Extragradient algorithm, according to Expression (2), as follows:

$$
\begin{aligned}
z_{k+\frac{1}{2}} =& \Pi_{\mathcal{Z}}\left[z_k - \eta F(z_k)\right] = \arg\min_{z\in\mathcal{Z}} \|z - (z_k - \eta F(z_k))\|, \\
z_{k+1} =& \Pi_{\mathcal{Z}}\left[z_k - \eta F(z_{k+\frac{1}{2}})\right] = \arg\min_{z\in\mathcal{Z}} \left\|z - \left(z_k - \eta F(z_{k+\frac{1}{2}})\right)\right\|.
\end{aligned}
$$

This appendix is organized as follows. The best-iterate convergence rate for the EG algorithm w.r.t. $\|z_k - z_{k+\frac{1}{2}}\|$ is known [Kor76?]. In Appendix G.1 we include the proof for completeness. A known corollary of the best-iterate iterate for the EG, is that the EG algorithm has bounded iterates (e.g. for z^* be a solution to the monotone VI, then for all $k \geq 0$, $\|z_k - z^*\|, \|z_{k+\frac{1}{2}} - z^*\| \leq O(\|z_0 - z^*\|)$. We include the proof in Appendix G.1.1 for completeness. In Appendix G.2 we show how to upper bound the tangent residual at z_k ($r^{tan}(z_k)$) at the best-iterate. In Appendix G.3 we show that the tangent residual in non-increasing across iterates of the EG algorithm, and we conclude by showing last-iterate convergence rates of the EG algorithm.

G.1 Best-Iterate Convergence of EG

Lemma 10 ([Kor76?]). *Let $\mathcal{Z}$ be a closed convex set in $\mathbb{R}^n$, $F(\cdot)$ be a monotone and L-Lipschitz operator mapping from $\mathcal{Z}$ to $\mathbb{R}^n$ and let $z^* \in \mathcal{Z}$ be a solution of the monotone VI (See Definition 5). For any $z_k \in \mathcal{Z}$, the EG algorithm with step size $\eta \in (0, \frac{1}{L})$. satisfies,*

$$
\|z_k - z^*\|^2 \geq \|z_{k+1} - z^*\|^2 + (1 - \eta^2 L^2)\|z_k - z_{k+\frac{1}{2}}\|^2. \tag{6}
$$

Proof. By Pythagorean inequality,

$$
\begin{aligned}
\|z_{k+1} - z^*\|^2 &\leq \|z_k - \eta F(z_{k+\frac{1}{2}}) - z^*\|^2 - \|z_k - \eta F(z_{k+\frac{1}{2}}) - z_{k+1}\|^2 \\
&= \|z_k - z^*\|^2 - \|z_k - z_{k+1}\|^2 + 2\eta\langle F(z_{k+\frac{1}{2}}), z^* - z_{k+1}\rangle \\
&= \|z_k - z^*\|^2 - \|z_k - z_{k+1}\|^2 + 2\eta\langle F(z_{k+\frac{1}{2}}), z^* - z_{k+\frac{1}{2}}\rangle + 2\eta\langle F(z_{k+\frac{1}{2}}), z_{k+\frac{1}{2}} - z_{k+1}\rangle.
\end{aligned} \tag{7}
$$

We first use monotonicity of $F(\cdot)$ to argue that $\langle F(z_{k+\frac{1}{2}}), z^* - z_{k+\frac{1}{2}}\rangle \leq 0$.

Fact 1. *For all $z \in \mathcal{Z}$, $\langle F(z), z^* - z\rangle \leq 0$.*

Proof.

$$
\begin{aligned}
0 &\leq \langle F(z^*) - F(z), z^* - z\rangle && \text{(monotonicity of } F(\cdot)) \\
&= \langle F(z^*), z^* - z\rangle - \langle F(z), z^* - z\rangle \\
&\leq -\langle F(z), z^* - z\rangle && (z^* \text{ is a solution of the monotone VI})
\end{aligned}
$$

□

We can simplify Equation (7) using Fact 1:

$$\begin{aligned}
\|z_{k+1}-z^*\|^2 &\leq \|z_k-z^*\|^2-\|z_k-z_{k+1}\|^2+2\eta\langle F(z_{k+\frac{1}{2}}), z_{k+\frac{1}{2}}-z_{k+1}\rangle \\
&= \|z_k-z^*\|^2-\|z_k-z_{k+\frac{1}{2}}\|^2-\|z_{k+\frac{1}{2}}-z_{k+1}\|^2-2\langle z_k-\eta F(z_{k+\frac{1}{2}})-z_{k+\frac{1}{2}}, z_{k+\frac{1}{2}}-z_{k+1}\rangle \\
&= \|z_k-z^*\|^2-\|z_k-z_{k+\frac{1}{2}}\|^2-\|z_{k+\frac{1}{2}}-z_{k+1}\|^2 \\
&\qquad -2\langle z_k-\eta F(z_k)-z_{k+\frac{1}{2}}, z_{k+\frac{1}{2}}-z_{k+1}\rangle-2\langle \eta F(z_k)-\eta F(z_{k+\frac{1}{2}}), z_{k+\frac{1}{2}}-z_{k+1}\rangle \\
&\leq \|z_k-z^*\|^2-\|z_k-z_{k+\frac{1}{2}}\|^2-\|z_{k+\frac{1}{2}}-z_{k+1}\|^2-2\eta\langle F(z_k)-F(z_{k+\frac{1}{2}}), z_{k+\frac{1}{2}}-z_{k+1}\rangle
\end{aligned}$$

The last inequality is because $\langle z_k-\eta F(z_k)-z_{k+\frac{1}{2}}, z_{k+\frac{1}{2}}-z_{k+1}\rangle \geq 0$, which follows from the that fact that $z_{k+\frac{1}{2}}=\Pi_{\mathcal{Z}}[z_k-\eta F(z_k)]$ and $z_{k+1}\in\mathcal{Z}$.

Finally, since $F(\cdot)$ is L-Lipschitz, we know that

$$-\langle F(z_k)-F(z_{k+\frac{1}{2}}), z_{k+\frac{1}{2}}-z_{k+1}\rangle \leq \|F(z_k)-F(z_{k+\frac{1}{2}})\|\cdot\|z_{k+\frac{1}{2}}-z_{k+1}\| \leq L\|z_k-z_{k+\frac{1}{2}}\|\cdot\|z_{k+\frac{1}{2}}-z_{k+1}\|.$$

So we can further simplify the inequality as follows:

$$\begin{aligned}
\|z_{k+1}-z^*\|^2 \leq& \|z_k-z^*\|^2-\|z_k-z_{k+\frac{1}{2}}\|^2-\|z_{k+\frac{1}{2}}-z_{k+1}\|^2-2\eta\langle F(z_k)-F(z_{k+\frac{1}{2}}), z_{k+\frac{1}{2}}-z_{k+1}\rangle \\
\leq& \|z_k-z^*\|^2-\|z_k-z_{k+\frac{1}{2}}\|^2-\|z_{k+\frac{1}{2}}-z_{k+1}\|^2+2\eta L\|z_k-z_{k+\frac{1}{2}}\|\cdot\|z_{k+\frac{1}{2}}-z_{k+1}\| \\
\leq& \|z_k-z^*\|^2-(1-\eta^2L^2)\|z_k-z_{k+\frac{1}{2}}\|^2
\end{aligned}$$

Hence,

$$\|z_k-z^*\|^2 \geq \|z_{k+1}-z^*\|^2+(1-\eta^2L^2)\|z_k-z_{k+\frac{1}{2}}\|^2.$$

□

G.1.1 Bounded Iterates of EG with Constant Step Size

Corollary 1. *Let $\mathcal{Z}$ be a closed convex set in $\mathbb{R}^n$, $F(\cdot)$ be a monotone and L-Lipschitz operator mapping from $\mathcal{Z}$ to $\mathbb{R}^n$ and let $z^*\in\mathcal{Z}$ be a solution of the VI (See Definition 5). Let $z_0\in\mathcal{Z}$ be an arbitrary starting point and $\{z_k, z_{k+\frac{1}{2}}\}_{k\geq 0}$ be the iterates of the EG algorithm with step size $\eta\in(0,\frac{1}{L})$. Then for all $k\geq 0$,*

$$\begin{aligned}
\|z_k-z^*\| \leq& \|z_0-z^*\|, \\
\left\|z_{k+\frac{1}{2}}-z^*\right\| \leq& \left(1+\frac{1}{\sqrt{1-\eta^2L^2}}\right)\|z_0-z^*\|.
\end{aligned}$$

Proof. By Lemma 10 we have that for any $k\geq 0$,

$$\begin{aligned}
\|z_{k+1}-z^*\| \leq& \|z_k-z^*\|, \\
\left\|z_{k+\frac{1}{2}}-z_k\right\| \leq& \frac{1}{\sqrt{1-\eta^2L^2}}\|z_k-z^*\|.
\end{aligned}$$

By triangle inequality,

$$\left\|z_{k+\frac{1}{2}}-z^*\right\| \leq \left\|z_{k+\frac{1}{2}}-z_k\right\|+\|z_k-z^*\| \leq \left(1+\frac{1}{\sqrt{1-\eta^2L^2}}\right)\|z_k-z^*\|,$$

which concludes the proof □

G.2 Best-Iterate of Tangent Residual

Lemma 11. *Let $\mathcal{Z}$ be a closed convex set in $\mathbb{R}^n$, $F(\cdot)$ be a monotone and L-Lipschitz operator mapping from $\mathcal{Z}$ to $\mathbb{R}^n$. For any $z_k \in \mathcal{Z}$, the EG algorithm update satisfies $r^{tan}(z_{k+1}) \leq \left(1+\eta L+(\eta L)^2\right)\frac{||z_k - z_{k+1/2}||}{\eta}$.*

Proof. We need the following fact for our proof.

Fact 2. $\|z_{k+\frac{1}{2}} - z_{k+1}\| \leq \eta L \|z_k - z_{k+\frac{1}{2}}\|$. *Moreover, when $\eta L < 1$, $\|z_{k+\frac{1}{2}} - z_{k+1}\| \leq \frac{\|z_k - z_{k+1}\|}{1-\eta L}$.*

Proof. Recall that $z_{k+\frac{1}{2}} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(z_k)\right]$ and $z_{k+1} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(z_{k+\frac{1}{2}})\right]$. By the non-expansiveness of the projection operator and the L-Lipschitzness of operator F, we have that $\|z_{k+\frac{1}{2}} - z_{k+1}\| \leq \|\eta(F(z_{k+\frac{1}{2}}) - F(z_k))\| \leq \eta L \|z_k - z_{k+\frac{1}{2}}\|$.

Finally, by the triangle inequality

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

□

Now we prove Lemma 11. By the L-Lipschitzness of operator F we have

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

Recall that $z_{k+1} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(z_{k+\frac{1}{2}})\right]$. Using Lemma 6, we have

$$\begin{aligned}
r^{tan}(z_{k+1}) &\leq \left\|\frac{z_k - z_{k+1}}{\eta} + F(z_{k+1}) - F(z_{k+\frac{1}{2}})\right\| \\
&\leq \frac{\|z_k - z_{k+1}\|}{\eta} + \|F(z_{k+1}) - F(z_{k+\frac{1}{2}})\| \\
&\leq \frac{\|z_k - z_{k+1}\| + (\eta L)^2 \|z_k - z_{k+\frac{1}{2}}\|}{\eta} \\
&\leq \frac{||z_k - z_{k+\frac{1}{2}}|| + ||z_{k+\frac{1}{2}} - z_{k+1}|| + (\eta L)^2 ||z_k - z_{k+\frac{1}{2}}||}{\eta} \\
&\leq \left(1+\eta L+(\eta L)^2\right)\frac{||z_k - z_{k+\frac{1}{2}}||}{\eta}.
\end{aligned}$$

The second and the fourth inequality follow from the triangle inequality. The third inequality follows from Inequality (8). In the final inequality we use $||z_{k+\frac{1}{2}} - z_{k+1}|| \leq \eta L ||z_k - z_{k+\frac{1}{2}}||$ by Fact 2. □

Lemma 12. *Let $\mathcal{Z}$ be a closed convex set in $\mathbb{R}^n$, $F(\cdot)$ be a monotone and L-Lipschitz operator mapping from $\mathcal{Z}$ to $\mathbb{R}^n$ and let $z^* \in \mathcal{Z}$ be a solution of the VI. For arbitrary starting point $z_0 \in \mathcal{Z}$, let $\{z_k, z_{k+\frac{1}{2}}\}_{k\geq 0}$ be the iterates of the EG algorithm with step size $\eta \in (0, \frac{1}{L})$. For any $T > 0$, there exists $t^* \in [T]$ such that:*

$$\left\|z_{t^*} - z_{t^*+\frac{1}{2}}\right\|^2 \leq \frac{1}{T}\frac{\|z_0 - z^*\|^2}{1-(\eta L)^2}, \qquad \text{AND} \qquad r^{tan}(z_{t^*+1}) \leq \frac{1+\eta L+(\eta L)^2}{\eta}\frac{1}{\sqrt{T}}\frac{\|z_0 - z^*\|}{\sqrt{1-(\eta L)^2}}.$$

Proof. By Lemma 10 we have

$$\|z_0 - z^*\|^2 \geq \|z_{T+1} - z^*\|^2 + (1-\eta^2 L^2)\sum_{k=0}^{T}\|z_k - z_{k+\frac{1}{2}}\|^2 \geq (1-\eta^2 L^2)\sum_{k=0}^{T}\|z_k - z_{k+\frac{1}{2}}\|^2$$

Thus there exists a $t^* \in [T]$ such that $\|z_{t^*} - z_{t^*+\frac{1}{2}}\|^2 \leq \frac{\|z_0 - z^*\|^2}{T(1-\eta^2 L^2)}$. We conclude the proof by applying Lemma 11. □

G.3 Last-Iterate Convergence of EG with Constant Step Size

In this section, we show that the last-iterate convergence rate is $O(\frac{1}{\sqrt{T}})$. In particular, we prove that the tangent residual is non-increasing, which, in combination with Lemma 12, implies the last-iterate convergence rate of the tangent residual for monotone VIs. To establish the monotonicity of the tangent residual, we combine SOS programming with the low-dimensionality of the EG update rule. To better illustrate our approach, we first prove the result in the unconstrained setting (Appendix G.3.1), then show how to generalize it to the constrained setting (Appendix G.3.2).

G.3.1 Warm Up: Unconstrained Case

As a warm-up, we consider the unconstrained setting where $\mathcal{Z} = \mathbb{R}^n$. Although the last-iterate convergence rate is known in the unconstrained setting due to [GPDO20, GLG21], we provide a simpler proof that also permits a larger step size. Our analysis holds for any step size $\eta \in (0, \frac{1}{L})$, while the previous analysis requires $\eta \leq \frac{1}{\sqrt{2}L}$ [GLG21].

In Theorem 4, we show that the tangent residual is monotone in the unconstrained setting.[12] Our approach is to apply SOS programming to search for a certificate of non-negativity for $\|F(z_k)\|^2 - \|F(z_{k+1})\|^2$ for every k, over the semialgebraic set defined by the following polynomial constraints in variables $\{z_i[\ell], \eta F(z_i)[\ell]\}_{i \in \{k, k+\frac{1}{2}, k+1\}, \ell \in [n]}$:

$$z_{k+\frac{1}{2}}[\ell] - z_k[\ell] + \eta F(z_k)[\ell] = 0, \quad z_{k+1}[\ell] - z_k[\ell] + \eta F(z_{k+\frac{1}{2}})[\ell] = 0, \ \forall \ell \in [n], \quad \text{(EG Update)}$$

$$\left\|\eta F(z_i) - \eta F(z_j)\right\|^2 - (\eta L)^2 \|z_i - z_j\|^2 \leq 0, \quad \forall i, j \in \{k, k+\frac{1}{2}, k+1\}, \qquad \text{(Lipschitzness)}$$

$$\left\langle \eta F(z_i) - \eta F(z_j), z_j - z_i \right\rangle \leq 0, \quad \forall i, j \in \{k, k+\frac{1}{2}, k+1\}. \qquad \text{(Monotonicity)}$$

We always multiply F with η in the constraints as it will be convenient later. We use K to denote the set $\{k, k+\frac{1}{2}, k+1\}$. To obtain a certificate of non-negativity, we apply SOS programming to search for a degree-2 SOS proof. More specifically, we want to find non-negative coefficients $\{\lambda^*_{i,j}, \mu^*_{i,j}\}_{i>j, i,j \in K}$ and degree-1 polynomials $\gamma_1^{(\ell)}(\boldsymbol{w})$ and $\gamma_2^{(\ell)}(\boldsymbol{w})$ in $\mathbb{R}[\boldsymbol{w}]$ for each $\ell \in [n]$, where $\boldsymbol{w} := \{z_i[\ell], \eta F(z_i)[\ell]\}_{i \in K, \ell \in [n]}$, such that the following is an SOS polynomial:

$$\begin{aligned}
&\|\eta F(z_k)\|^2 - \|\eta F(z_{k+1})\|^2 + \sum_{i>j \text{ and } i,j \in K} \lambda^*_{i,j} \cdot \left(\left\|\eta F(z_i) - \eta F(z_j)\right\|^2 - (\eta L)^2 \|z_i - z_j\|^2 \right) \\
&+ \sum_{i>j \text{ and } i,j \in K} \mu^*_{i,j} \cdot \langle \eta F(z_i) - \eta F(z_j)), z_j - z_i \rangle + \sum_{\ell \in [n]} \gamma_1^{(\ell)}(\boldsymbol{w}) (z_{k+\frac{1}{2}}[\ell] - z_k[\ell] + \eta F(z_k)[\ell]) \\
&+ \sum_{\ell \in [n]} \gamma_2^{(\ell)}(\boldsymbol{w}) (z_{k+1}[\ell] - z_k[\ell] + \eta F(z_{k+\frac{1}{2}})[\ell]).
\end{aligned} \tag{9}$$

Due to constraints satisfied by the EG iterates, the non-negativity of Expression (9) clearly implies that $\|F(z_k)\|^2 - \|F(z_{k+1})\|^2$ is non-negative. However, Expression (9) is in fact an infinite family of polynomials rather than a single one. Expression (9) corresponds to a different polynomial for every integer n. To directly search for the solution, we would need to solve an infinitely large SOS program, which is clearly infeasible. By exploring the symmetry in Expression (9), we show that it suffices to solve a constant size SOS program.

[12] In the unconstrained setting, the tangent residual is simply the norm of the operator $r^{tan}_{(F,\mathbb{R}^n)}(z) = \|F(z)\|$.

Let us first expand Expression (9) as follows:

$$\sum_{\ell\in[n]}\Big((\eta F(z_k)[\ell])^2 - (\eta F(z_{k+1})[\ell])^2 + \sum_{i>j \text{ and } i,j\in K} \lambda^*_{i,j}\Big(\big(\eta F(z_i)[\ell] - \eta F(z_j)[\ell]\big)^2 - (\eta L)^2\big(z_i[\ell]-z_j[\ell]\big)^2\Big)$$
$$+ \sum_{i>j \text{ and } i,j\in K} \mu^*_{i,j}\big(\eta F(z_i)[\ell] - \eta F(z_j)[\ell]\big)\big)\big(z_j[\ell]-z_i[\ell]\big) + \gamma_1^{(\ell)}(\boldsymbol{w})(z_{k+\frac{1}{2}}[\ell] - z_k[\ell] + \eta F(z_k)[\ell])$$
$$+ \gamma_2^{(\ell)}(\boldsymbol{w})(z_{k+1}[\ell] - z_k[\ell] + \eta F(z_{k+\frac{1}{2}})[\ell])\Big). \tag{10}$$

What we will argue next is that, due to the *symmetry across coordinates*, it suffices to directly search for a single SOS proof that shows that each of the n summands in Expression (10) is an SOS polynomial. More specifically, we make use of the following two key properties. (i) For any $\ell, \ell' \in [n]$, the ℓ-th summand and ℓ'-th summand are identical subject to a change of variable;[13] (ii) the ℓ-th summand only depends on the coordinate ℓ, i.e., variables in $\{z_i[\ell], \eta F(z_i)[\ell]\}_{i\in K}$ and does not involve any other coordinates.[14] We solve the following SOS program, whose solution can be used to construct $\{\lambda^*_{i,j}, \mu^*_{i,j}\}_{i>j,i,j\in K}$ and $\{\gamma_1^{(\ell)}(\boldsymbol{w}), \gamma_2^{(\ell)}(\boldsymbol{w})\}_{\ell\in[n]}$ so that each of the summands in Expression (10) is an SOS polynomial.

Input Fixed Polynomials. We use $\boldsymbol{x}$ to denote (x_0, x_1, x_2) and $\boldsymbol{y}$ to denote (y_0, y_1, y_2). Interpret x_i as $z_{k+\frac{i}{2}}[\ell]$ and y_i as $\eta F(z_{k+\frac{i}{2}})[\ell]$ for $0 \le i \le 2$. Observe that $h_1(\boldsymbol{x},\boldsymbol{y})$ and $h_2(\boldsymbol{x},\boldsymbol{y})$ come from the EG update rule on coordinate ℓ. $g^L_{i,j}(\boldsymbol{x},\boldsymbol{y})$ and $g^m_{i,j}(\boldsymbol{x},\boldsymbol{y})$ come from the ℓ-th coordinate's contribution in the Lipschitzness and monotonicity constraints.

- $h_1(\boldsymbol{x},\boldsymbol{y}) := x_1 - x_0 + y_0$ and $h_2(\boldsymbol{x},\boldsymbol{y}) := x_2 - x_0 + y_1$.
- $g^L_{i,j}(\boldsymbol{x},\boldsymbol{y}) := (y_i - y_j)^2 - C\cdot(x_i - x_j)^2$ for any $0 \le j < i \le 2$.[a]
- $g^m_{i,j}(\boldsymbol{x},\boldsymbol{y}) := (y_i - y_j)(x_j - x_i)$ for any $0 \le j < i \le 2$.

Decision Variables of the SOS Program:

- $p^L_{i,j} \ge 0$, and $p^m_{i,j} \ge 0$, for all $0 \le j < i \le 2$.
- $q_1(\boldsymbol{x},\boldsymbol{y})$ and $q_2(\boldsymbol{x},\boldsymbol{y})$ are two degree 1 polynomials in $\mathbb{R}[\boldsymbol{x},\boldsymbol{y}]$.

Constraints of the SOS Program:

$$\text{s.t.}\quad y_0^2 - y_2^2 + \sum_{2\ge i>j\ge 0} p^L_{i,j}\cdot g^L_{i,j}(\boldsymbol{x},\boldsymbol{y}) + \sum_{2\ge i>j\ge 0} p^m_{i,j}\cdot g^m_{i,j}(\boldsymbol{x},\boldsymbol{y})$$
$$+ q_1(\boldsymbol{x},\boldsymbol{y})\cdot h_1(\boldsymbol{x},\boldsymbol{y}) + q_2(\boldsymbol{x},\boldsymbol{y})\cdot h_2(\boldsymbol{x},\boldsymbol{y}) \in \mathrm{SOS}[\boldsymbol{x},\boldsymbol{y}]. \tag{11}$$

[a] C represents $(\eta L)^2$. Larger C corresponds to a larger step size and makes the SOS program harder to satisfy. Through binary search, we find that the largest possible value of C is 1 while maintaining the feasibility of the SOS program.

Figure 3: Our SOS program in the unconstrained setting.

The proof of the following theorem is based on a feasible solution to the SOS program in Figure 3.

Theorem 4. *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator. Then for any $k \in \mathbb{N}$, the EG algorithm with step size $\eta \in (0, \frac{1}{L})$ satisfies $\|F(z_k)\|^2 \ge \|F(z_{k+1})\|^2$.*

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

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

[13] Simply replace $\{z_i[\ell]\}_{i\in K}$ and $\{\eta F(z_i)[\ell]\}_{i\in K}$ with $\{z_i[\ell']\}_{i\in K}$ and $\{\eta F(z_i)[\ell']\}_{i\in K}$.

[14] We mainly care about the polynomials arise from the constraints. Although $\gamma_1^{(\ell)}(\boldsymbol{w})$ and $\gamma_2^{(\ell)}(\boldsymbol{w})$ could depend on other coordinates, we show that it suffices to consider polynomials in $\{z_i[\ell], \eta F(z_i)[\ell]\}_{i\in K}$.

and

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

We simplify them using the update rule of EG and $\eta L < 1$. In particular, we replace $z_k - z_{k+1}$ with $\eta F(z_{k+\frac{1}{2}})$ and $z_{k+\frac{1}{2}} - z_{k+1}$ with $\eta F(z_{k+\frac{1}{2}}) - \eta F(z_k)$.

$$\left\langle F(z_{k+1}) - F(z_k), F(z_{k+\frac{1}{2}})\right\rangle \le 0, \tag{12}$$

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

Note that

$$\begin{aligned}
&\|F(z_k)\|^2 - \|F(z_{k+1})\|^2 + 2\cdot \text{LHS of Inequality}(12) + \text{LHS of Inequality}(13)\\
=&\|F(z_k)\|^2 - \|F(z_{k+1})\|^2 + 2\cdot\left\langle F(z_{k+1}), F(z_{k+\frac{1}{2}})\right\rangle - 2\cdot\left\langle F(z_k), F(z_{k+\frac{1}{2}})\right\rangle\\
&\quad + \left\|F(z_{k+\frac{1}{2}})\right\|^2 - 2\cdot\left\langle F(z_{k+1}), F(z_{k+\frac{1}{2}})\right\rangle + \|F(z_{k+1})\|^2\\
&\quad - \left\|F(z_{k+\frac{1}{2}})\right\|^2 + 2\cdot\left\langle F(z_k), F(z_{k+\frac{1}{2}})\right\rangle - \|F(z_k)\|^2 = 0.
\end{aligned}$$

Thus, $\|F(z_k)\|^2 - \|F(z_{k+1})\|^2 \ge 0$.

□

Corollary 2 is implied by combing Lemma 4, Lemma 12, Theorem 4 and the fact that $\eta \in (0, \frac{1}{L})$.

Corollary 2. *Let $F(\cdot)$ be a monotone and L-Lipschitz operator mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ and let $z^* \in \mathbb{R}^n$ be a solution of the VI. For arbitrary starting point $z_0 \in \mathcal{Z}$, let $\{z_k, z_{k+\frac{1}{2}}\}_{k\ge 0}$ be the iterates of the EG algorithm with step size $\eta \in (0, \frac{1}{L})$. For any $T \ge 1$ and $D > 0$, $\textsc{Gap}(z_T) \le \frac{1}{\sqrt{T}}\frac{3D\|z_0 - z^*\|}{\eta\sqrt{1-(\eta L)^2}}$.*

G.3.2 Last-Iterate Convergence of EG with Arbitrary Convex Constraints

We establish the last-iterate convergence rate of the EG algorithm in the constrained setting for monotone VIs in this section. The plan is similar to the one in Appendix G.3.1. First, we use the assistance of SOS programming to prove the monotonicity of the tangent residual (Theorem 5), then combine it with the best-iterate convergence guarantee from Lemma 12 to derive the last-iterate convergence rate (Theorem 6).

Due to the constraints, proving the monotonicity of the tangent residual becomes much more challenging. The tangent residual in the constrained setting (Definition 4) is significantly more complex than its counterpart in the unconstrained setting. In Lemma 13, we introduce an auxiliary point $c(z)$ for every point z that can be used to simplified the tangent residual.

Lemma 13. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set and $F : \mathcal{Z} \to \mathbb{R}^n$ be an operator. For any $z \in \mathcal{Z}$, denote $c(z) := \Pi_{N(z)}\Big[-F(z)\Big]$ the projection of $-F(z)$ on the normal cone $N(z)$. Then we have*

- $r^{tan}(z) = \|F(z) + c(z)\|$,
- $\langle F(z) + c(z), c(z)\rangle = 0$,
- $\langle F(z) + c(z), a\rangle \ge 0, \forall a \in N(z)$.

Proof. According to the definition of $c(z)$, $r^{tan}(z) = \|F(z) + c(z)\|$ follows from Lemma 5. Since $c(z) = \Pi_{N(z)}\Big[-F(z)\Big]$, we know that for all $a \in N(z)$,

$$\langle -F(z) - c(z), a - c(z)\rangle \le 0. \tag{14}$$

Note that $c(z) \in N(z)$ and $N(z)$ is a cone. By substituting $a = 0$ and $a = 2 \cdot c(z)$ in (14), we get

$$\langle -F(z) - c(z), c(z) \rangle = 0.$$

Therefore, for all $a \in N(z)$, we have

$$\langle -F(z) - c(z), a \rangle = \langle -F(z) - c(z), a - c(z) \rangle \leq 0.$$

□

Next, we need to decide over which semialgebraic set that we want to certify the non-negativity of $r^{tan}(z_k)^2 - r^{tan}(z_{k+1})^2$. Naturally, we would like to use all constraints of $\mathcal{Z}$, but there might be arbitrarily many of them. In the next paragraph, we argue how to reduce the number of constraints.

Reducing the Number of Constraints. Suppose we are not given the description of $\mathcal{Z} \subseteq \mathbb{R}^n$, and we only observe one iteration of the EG algorithm. In other words, we know z_k, $z_{k+\frac{1}{2}}$, and z_{k+1}, as well as $F(z_k)$, $F(z_{k+\frac{1}{2}})$, and $F(z_{k+1})$. To express the squared tangent residual at z_k and the squared tangent residual at z_{k+1}, let us also assume that the vector $c_k = \Pi_{N(z_k)}\Big[-F(z_k)\Big]$ and $c_{k+1} = \Pi_{N(z_{k+1})}\Big[-F(z_{k+1})\Big]$, and according to Lemma 13, we have $r^{tan}(z_k)^2 = \|F(z_k) + c_k\|^2$, and $r^{tan}(z_{k+1})^2 = \|F(z_{k+1}) + c_{k+1}\|^2$. Our plan is to derive a set of inequalities that must be satisfied by these vectors. From this limited information, what can we learn about $\mathcal{Z}$? We can conclude that $\mathcal{Z}$ must lie in the intersection of the following halfspaces: (a) $\langle c_k, z \rangle \leq \langle c_k, z_k \rangle$. This is true because $c_k \in N(z_k)$. (b) $\langle a_{k+\frac{1}{2}}, z \rangle \geq \langle a_{k+\frac{1}{2}}, z_{k+\frac{1}{2}} \rangle$, where $a_{k+\frac{1}{2}} = -(z_k - \eta F(z_k) - z_{k+\frac{1}{2}})$. This is true because $z_{k+\frac{1}{2}} = \Pi_{\mathcal{Z}}(z_k - \eta F(z_k))$, so $-a_{k+\frac{1}{2}} \in N(z_{k+\frac{1}{2}})$. (c) $\langle a_{k+1}, z \rangle \geq \langle a_{k+1}, z_{k+1} \rangle$, where $a_{k+1} = -(z_k - \eta F(z_{k+\frac{1}{2}}) - z_{k+1})$. This is true because $z_{k+1} = \Pi_{\mathcal{Z}}(z_k - \eta F(z_{k+\frac{1}{2}}))$, so $-a_{k+1} \in N(z_{k+1})$. See Figure 4 for illustration. Additionally, due to our definition of c_k and c_{k+1} and Lemma 13, we know that (d) $\langle \eta F(z_i) + \eta c_i, \eta c_i \rangle = 0$ for $i \in \{k, k+1\}$, and (e) $\langle \eta F(z_{k+1}) + \eta c_{k+1}, a_{k+1} \rangle \leq 0$ as $-a_{k+1} \in N(z_{k+1})$.

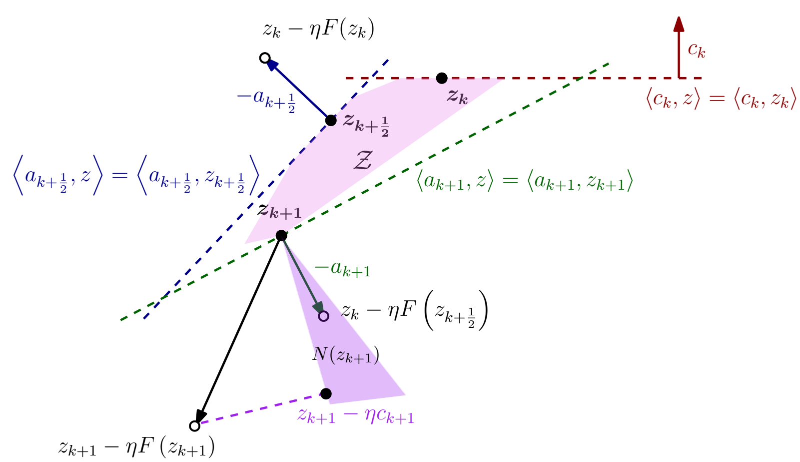

Figure 4: Reducing the number of constraints.

Clearly, for any $\mathcal{Z}$, the inequalities in (a) to (e) must hold, though there might be other inequalities that are also true. Our goal is to prove that the tangent residual is non-increasing even if only inequalities (a) to (e) hold. If we can do so, then we prove that tangent residual is non-increasing for an arbitrary $\mathcal{Z}$.

Formulation as SOS program. Similar to the unconstrained case, our plan is to search for a certificate of non-negativity of the following expression

$$\|F(z_k) + c_k\|^2 - \|F(z_{k+1}) + c_{k+1}\|^2 \tag{15}$$

over the semialgebraic set defined by the following polynomial constraints in variables $\left\{\{z_i[\ell], \eta F(z_i)[\ell]\}_{i\in\{k,k+\frac{1}{2},k+1\}} \cup \{c_i[\ell]\}_{i\in k,k+1}\right\}_{\ell\in[n]}$

$$\begin{aligned}
&\left\|\eta F(z_i) - \eta F(z_j)\right\|^2 - (\eta L)^2\left\|z_i - z_j\right\|^2 \le 0, && \forall i,j \in \{k, k+\frac{1}{2}, k+1\}, && \text{(Lipschitzness)}\\
&\left\langle \eta F(z_i) - \eta F(z_j), z_j - z_i\right\rangle \le 0, && \forall i,j \in \{k, k+\frac{1}{2}, k+1\}, && \text{(Monotonicity)}\\
&\left\langle a_i, z_i - z_j\right\rangle \le 0, && \forall i \in \{k+\frac{1}{2}, k+1\}, j \in \{k, k+\frac{1}{2}, k+1\}, && (-a_i \in N(z_i))\\
&\left\langle \eta c_i, z_j - z_i\right\rangle \le 0, && \forall i \in \{k, k+1\}, j \in \{k, k+\frac{1}{2}, k+1\}, && (c_i \in N(z_i))\\
&\langle \eta F(z_i) + \eta c_i, \eta c_i\rangle = 0, && \forall i \in \{k, k+1\}, && (\textit{Lemma } 13)\\
&\langle \eta F(z_{k+1}) + \eta c_{k+1}, a_{k+1}\rangle \le 0, && && (\textit{Lemma } 13).
\end{aligned}$$

Similar to Section G.3, we multiply the operators, c_k, and c_{k+1} with η for convenience. Fortunately, the dimensional-dependent Expression (15) and semialgebraic set are symmetric across coordinates, and more specifically, satisfy the two key properties in the unconstrained case – Property (i) and (ii). Hence, we can represent all of the coordinates $\ell \ge 1$ with one coordinate in the SOS program, and we can form a constant size SOS program to search for a certificate of non-negativity for Expression (15) as shown in Figure 5.

In Theorem 5, we establish the monotonicity of the tangent residual. Our proof is based on the solution to the degree-2 SOS program concerning polynomials in 8 variables (Figure 5).

Theorem 5. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set and $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator. For any step size $\eta \in (0, \frac{1}{L})$ and any $z_k \in \mathcal{Z}$, the EG algorithm update satisfies $r^{tan}_{(F,\mathcal{Z})}(z_k) \ge r^{tan}_{(F,\mathcal{Z})}(z_{k+1})$.*

Proof. Let $c_k = \Pi_{N_{\mathcal{Z}}(z_k)}(-F(z_k))$ and $c_{k+1} = \Pi_{N_{\mathcal{Z}}(z_{k+1})}(-F(z_{k+1}))$. By Lemma 13 we have

$$\eta^2 r^{tan}(z_k)^2 - \eta^2 r^{tan}(z_{k+1})^2 = \|\eta F(z_k) + \eta c_k\|^2 - \|\eta F(z_{k+1}) + \eta c_{k+1}\|^2 \tag{17}$$

Combining the monotonicity and L-Lipschitzness of F with the fact that $L \le \frac{1}{\eta}$, we have

$$(-1)\cdot\left(\left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|^2 - \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_{k+1})\right\|^2\right) \le 0, \tag{18}$$

$$(-2)\cdot\langle \eta F(z_{k+1}) - \eta F(z_k), z_{k+1} - z_k)\rangle \le 0,. \tag{19}$$

Since $z_{k+\frac{1}{2}} = \Pi_{\mathcal{Z}}\left(z_k - \eta F(z_k)\right)$ and $z_{k+1} = \Pi_{\mathcal{Z}}\left(z_k - \eta F(z_{k+\frac{1}{2}})\right)$, we can infer that $z_k - \eta F(z_k) - z_{k+\frac{1}{2}} \in N(z_{k+\frac{1}{2}})$ and $z_k - \eta F(z_{k+\frac{1}{2}}) - z_{k+1} \in N(z_{k+1})$, which further implies

$$(-2)\cdot\left\langle z_k - \eta F(z_k) - z_{k+\frac{1}{2}}, z_{k+\frac{1}{2}} - z_{k+1}\right\rangle \le 0, \tag{20}$$

$$(-2)\cdot\left\langle z_k - \eta F(z_{k+\frac{1}{2}}) - z_{k+1}, z_{k+1} - z_k\right\rangle \le 0, \tag{21}$$

$$(-2)\cdot\left\langle \eta c_k, z_k - z_{k+\frac{1}{2}}\right\rangle \le 0,. \tag{22}$$

Since $z_k - \eta F(z_{k+\frac{1}{2}}) - z_{k+1} \in N(z_{k+1})$ and $c_{k+1} = \Pi_{N_{\mathcal{Z}}(z_{k+1})}(-F(z_{k+1}))$, by Lemma 13 we have

$$(-2)\cdot\left\langle \eta c_{k+1} + \eta F(z_{k+1}), z_k - \eta F(z_{k+\frac{1}{2}}) - z_{k+1}\right\rangle \le 0, \tag{23}$$

$$(-2)\cdot\langle \eta c_{k+1} + \eta F(z_{k+1}), -\eta c_{k+1}\rangle = 0. \tag{24}$$

MATLAB code for the verification of the following identity is included in the supplementary material under the name "verify_identity_EG.m".

Input Fixed Polynomials. We use $\boldsymbol{x}$ to denote (x_0, x_1, x_2), $\boldsymbol{y}$ to denote (y_0, y_1, y_2) and $\boldsymbol{w}$ to denote (w_0, w_2). Interpret x_i as $z_{k+\frac{i}{2}}[\ell]$ and y_i as $\eta F(z_{k+\frac{i}{2}})[\ell]$ for $0 \le i \le 2$, w_0 as $\eta c_k[\ell]$ and w_2 as $\eta c_{k+1}[\ell]$. Let $b_1 = -(x_0 - y_0 - x_1)$ and $b_2 = -(x_0 - y_1 - x_2)$.

Origin of Constraints. $g^L_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ and $g^m_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ come from the ℓ-th coordinate's contribution in the Lipschitzness and monotonicity constraints. Similarly, $g^b_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ and $g^w_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ come from the ℓ-th coordinate contribution of fact that $-a_i$ and c_i are in the normal cone of z_i. Finally, $h^w_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ and $g^r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ comes from the ℓ-th coordinate contribution due to the inequalities of Lemma 13.

- $g^L_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) := (y_i - y_j)^2 - C \cdot (x_i - x_j)^2$ for any $0 \le j < i \le 2$.[a]
- $g^m_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) := (y_i - y_j)(x_j - x_i)$ for any $0 \le j < i \le 2$.
- $g^b_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) := b_i \cdot (x_i - x_j)$ for any $i \in \{1,2\}, 0 \le j \le 2$.
- $g^w_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) := w_i \cdot (x_j - x_i)$ for any $i \in \{0,2\}, 0 \le j \le 2$.
- $g^r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) := (y_2 + w_2) \cdot b_2$.
- $h^w_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) := (y_i + w_i) \cdot w_i$ for any $i \in \{0,2\}$.

Decision Variables of the SOS Program:

- $p^L_{i,j} \ge 0$, and $p^m_{i,j} \ge 0$, for all $0 \le j < i \le 2$.
- $p^b_{i,j} \ge 0$, for any $i \in \{1,2\}, 0 \le j \le 2$.
- $p^w_{i,j} \ge 0$, for any $i \in \{0,2\}, 0 \le j \le 2$.
- $p^r \ge 0$.
- $q^w_0, q^w_2 \in \mathbb{R}$.

Constraints of the SOS Program:

$$\begin{aligned}\text{s.t.}\quad & (y_0 + w_0)^2 - (y_2 + w_2)^2 + \sum_{2 \ge i > j \ge 0} p^L_{i,j} \cdot g^L_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) + \sum_{2 \ge i > j \ge 0} p^m_{i,j} \cdot g^m_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) \\ & \quad + \textstyle\sum_{i \in \{1,2\}, 2 \ge j \ge 0} p^b_{i,j} \cdot g^b_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) + \sum_{i \in \{0,2\}, 2 \ge j \ge 0} p^w_{i,j} \cdot g^w_{i,j}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) \qquad \in SOS[\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}] \\ & \quad + p^r \cdot g^r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) + \textstyle\sum_{i \in \{0,2\}} q^w_i \cdot h^w_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})\end{aligned} \tag{16}$$

[a] C represents $(\eta L)^2$.

Figure 5: Our SOS program in the constrained setting.

$$\begin{aligned}&\text{Expression (17)} + \text{LHS of Inequality (18)} + \text{LHS of Inequality (19)} \\ &+ \text{LHS of Inequality (20)} + \text{LHS of Inequality (21)} + \text{LHS of Inequality (22)} \\ &+ \text{LHS of Inequality (23)} + \text{LHS of Inequality (24)} \\ &= \|\eta F(z_k) + \eta c_k - z_k + z_{k+\frac{1}{2}}\|^2 && (25) \\ &+ \|\eta F(z_{k+\frac{1}{2}}) + \eta c_{k+1} - z_k + z_{k+1}\|^2 \ge 0, && (26)\end{aligned}$$

which concludes the proof. □

Theorem 6. *Let $\mathcal{Z}$ be a closed convex set in $\mathbb{R}^n$, $F(\cdot)$ be a monotone and L-Lipschitz operator mapping from $\mathcal{Z}$ to $\mathbb{R}^n$ and let $z^* \in \mathcal{Z}$ be a solution of the VI. For arbitrary starting point $z_0 \in \mathcal{Z}$, let $\{z_k, z_{k+\frac{1}{2}}\}_{k \ge 0}$ be the iterates of the EG algorithm with step size $\eta \in (0, \frac{1}{L})$. Then for any $T \ge 1$ and $D > 0$,*

- $\textsc{Gap}(z_T) \le \frac{1}{\sqrt{T}} \frac{3D\|z_0 - z^*\|}{\eta\sqrt{1-(\eta L)^2}}$,
- $r^{nat}(z_T) \le r^{tan}(z_T) \le \frac{1}{\sqrt{T}} \frac{3\|z_0 - z^*\|}{\eta\sqrt{1-(\eta L)^2}}$.

Theorem 6 is implied by combing Lemma 4, Lemma 7, Lemma 12, Theorem 5 and the fact that $\eta \in (0, \frac{1}{L})$.

H Optimistic Gradient Algorithm

In this section, we provide the last-iterate convergence rate of the OG algorithm. Similar to Appendix G, we only show the last-iterate convergence rate for monotone VIs w.r.t. the gap function for VIs (Definition 6), the natural residual (Definition 7) and the tangent residual (Definition 4) and the potential function $\Phi(z_k, w_k)$. For the rest of this section, we slightly abuse notation and refer to the gap function for VIs as the gap function. We show in Appendix F (Appendix I resp.) last-iterate convergence rates for additional performance measures for smooth monotone games (monotone VIs resp.).

Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set and $F : \mathcal{Z} \to \mathbb{R}$ be an operator. Let z_k and w_k be the k-th iterate of the Optimistic Gradient Descent Ascent algorithm (OG) algorithm. Let z_0, w_0 be arbitrary point in $\mathcal{Z}$ and $\{z_k, w_k\}_{k\geq 0}$ be the iterated of the OG algorithm. The update rule for any $k \geq 0$ is as follows:

$$\begin{aligned} w_{k+1} &= \Pi_{\mathcal{Z}}\left[z_k - \eta F(w_k)\right] = \arg\min_{z\in\mathcal{Z}} \|z - (z_k - \eta F(w_k))\| \\ z_{k+1} &= \Pi_{\mathcal{Z}}\left[z_k - \eta F(w_{k+1})\right] = \arg\min_{z\in\mathcal{Z}} \|z - (z_k - \eta F(w_{k+1}))\| \end{aligned} \tag{27}$$

We prove last-iterate convergence rate for OG w.r.t. the gap function, natural residual and tangent residual in Theorem 8 at Section H.4. The last-iterate convergence proof for OG is a simple extension of the proof for EG. The last-iterate convergence rate for the performance measures we mentioned follow from the last-iterate convergence rate of the following monotonically decreasing potential function:

$$\Phi(z_k, w_k) = \|F(z_k) - F(w_k)\|^2 + r^{tan}(z_k)^2 \tag{28}$$

We can interpret $\Phi(z_k, w_k)$ as an upper bound of $\|w_{k+1} - z_k\|$ (Lemma 14).

Lemma 14. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F : \mathcal{Z} \to \mathbb{R}^n$ be any operator and $z_1 = \Pi_{\mathcal{Z}}(z_2 - \eta F(z_3))$. Then $\|z_1 - z_2\|^2 \leq 2 \cdot (\eta^2 r^{tan}(z_2)^2 + \|\eta F(z_2) - \eta F(z_3)\|^2)$.*

Proof. Let $\widehat{z}_2 = \Pi_{\mathcal{Z}}(z_2 - \eta F(z_2))$. Then

$$\|z_1 - z_2\| \leq \|z_1 - \widehat{z}_2\| + \|z_2 - \widehat{z}_2\|. \tag{29}$$

By non-expansiveness of the projection mapping, we have that

$$\|z_1 - \widehat{z}_2\| \leq \|\eta F(z_2) - \eta F(z_3)\| \tag{30}$$

By Definition 7, Lemma 7 and Definition 4 we have that

$$\begin{aligned} \|z_2 - \widehat{z}_2\| &= r^{nat}_{(\eta F, \mathcal{Z})}(z_2) \\ &\leq r^{tan}_{(\eta F, \mathcal{Z})}(z_2) \\ &= \eta r^{tan}_{(F, \mathcal{Z})}(z_2). \end{aligned} \tag{31}$$

By combining Inequality (29), Inequality (30), Inequality (31) and the fact that $(a+b)^2 \leq 2a^2 + 2b^2$, we conclude

$$\|z_1 - z_2\|^2 \leq \left(\eta r^{tan}(z_2) + \|\eta F(z_2) - \eta F(z_3)\|\right)^2 \leq 2\Big(\eta^2 r^{tan}(z_2)^2 + \|\eta F(z_2) - \eta F(z_3)\|^2\Big).$$

∎

This appendix is organized as follows. In Section H.1, we derive best-iterate convergence rate of OG w.r.t. the quantity $\|z_k - w_{k+1}\|$. The rate of the best-iterate of OG is known [WLZL21a, HIMM19], but we include the proof for completeness. In Corollary 3, we show that the OG algorithm has bounded iterates (e.g. let z^* be a solution to the VI, then for all

$k \geq 0, \|z_k - z^*\|, \|w_k - z^*\| \leq O(\|z_0 - z^*\| + \|z_0 - w_0\|)$. In Section H.2, we show how to derive a best-iterate convergence rate w.r.t. the potential function $\Phi(z_k, w_k)$. In Section H.3 we show that the potential function $\Phi(z_k, w_k)$ is monotonically decreasing across iterates and finally in Section H.4 we show how to translate the last-iterate convergence rate w.r.t. the potential function $\Phi(z_k, w_k)$ to the last-iterate convergence rate of the performance measures of interest.

H.1 Best-Iterate Convergence of OG with Constant Step Size

The best-iterate convergence rate of OG is known [WLZL21a] and can easily be derived by [HIMM19]. We include the proof here for completeness.

Lemma 15 ([HIMM19, WLZL21a]). *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator, and z^* be a solution to the corresponding VI. Let $z_0, w_0 \in \mathcal{Z}$ be arbitrary starting points and $\{z_k, w_k\}_{k\geq 0}$ be the iterates of the OG algorithm with step size $\eta \in (0, \frac{1}{2L})$. Then for all $T \geq 0$,*

$$\|z_{T+1} - z^*\|^2 + \sum_{k=0}^{T} \|z_k - w_{k+1}\|^2 \leq \frac{1 - 2\eta^2 L^2}{1 - 4\eta^2 L^2}\|z_0 - z^*\|^2 + \frac{2\eta^2 L^2}{1 - 4\eta^2 L^2}\|w_0 - z_0\|^2. \tag{32}$$

Proof of Lemma 15: In order to upper bound $\sum_{k=0}^{T} \|w_k - w_{k+1}\|^2$, we first relate the quantity $\|w_k - w_{k+1}\|^2$ to the weighted sum of $\{\|z_t - w_{t+1}\|^2\}_{0\leq t\leq k}$.

Lemma 16. *For all $k \geq 0$,*

$$\|w_k - w_{k+1}\|^2 \leq 2(2\eta^2 L^2)^k \|w_0 - z_0\|^2 + \sum_{t=0}^{k} 2(2\eta^2 L^2)^t \|z_{k-t} - w_{k+1-t}\|^2. \tag{33}$$

Moreover, for all $T \geq 0$,

$$\sum_{k=0}^{T} \|w_k - w_{k+1}\|^2 \leq \frac{2}{1 - 2\eta^2 L^2}\left(\|w_0 - z_0\|^2 + \sum_{k=0}^{T} \|z_k - w_{k+1}\|^2\right). \tag{34}$$

Proof. We first prove Equation (33) by induction. Note that for all $k \geq 0$, we have

$$\begin{aligned}\|w_k - w_{k+1}\|^2 &= \|w_k - z_k + z_k - w_{k+1}\|^2 \\ &\leq 2\|w_k - z_k\|^2 + 2\|z_k - w_{k+1}\|^2.\end{aligned} \tag{35}$$

The inequality follows from the fact that $(a+b)^2 \leq 2a^2 + 2b^2$. Thus Equation (33) holds for the base case $k = 0$. For the sake of induction, we assume that Equation (33) holds for some $k - 1 \geq 0$. Using the update rule of OG, the non-expansiveness of the projection operator, and the L-Lipschitzness of F, for all $k \geq 1$ we have

$$\|w_k - z_k\|^2 \leq \eta^2 \|F(w_{k-1}) - F(w_k)\|^2 \leq \eta^2 L^2 \|w_{k-1} - w_k\|^2. \tag{36}$$

Combining Equation (35), Equation (36), and the induction assumption, we have

$$\begin{aligned}\|w_k - w_{k+1}\|^2 &\leq 2\|w_k - z_k\|^2 + 2\|z_k - w_{k+1}\|^2 \\ &\leq 2\eta^2 L^2 \|w_{k-1} - w_k\|^2 + 2\|z_k - w_{k+1}\|^2 \\ &\leq 2\eta^2 L^2 \left(2(2\eta^2 L^2)^{k-1}\|w_0 - z_0\|^2 + \sum_{t=0}^{k-1} 2(2\eta^2 L^2)^t \|z_{k-1-t} - w_{k-t}\|^2\right) + 2\|z_k - w_{k+1}\|^2 \\ &= 2(2\eta^2 L^2)^k \|w_0 - z_0\|^2 + \sum_{t=1}^{k} 2(2\eta^2 L^2)^t \|z_{k-t} - w_{k+1-t}\|^2 + 2\|z_k - w_{k+1}\|^2 \\ &= 2(2\eta^2 L^2)^k \|w_0 - z_0\|^2 + \sum_{t=0}^{k} 2(2\eta^2 L^2)^t \|z_{k-t} - w_{k+1-t}\|^2.\end{aligned}$$

This completes the proof of Equation (33).

Summing Equation (33) with $k = 0, 1, \cdots, T$, we have

$$\begin{aligned}
\sum_{k=0}^{T} \|w_k - w_{k+1}\|^2 &\le \sum_{k=0}^{T} 2(2\eta^2 L^2)^k \|w_0 - z_0\|^2 + \sum_{k=0}^{T}\sum_{t=0}^{k} 2(2\eta^2 L^2)^t \|z_{k-t} - w_{k+1-t}\|^2 \\
&= \sum_{k=0}^{T} 2(2\eta^2 L^2)^k \|w_0 - z_0\|^2 + \sum_{k=0}^{T} \left(\sum_{t=0}^{T-k} 2(2\eta^2 L^2)^t \right) \cdot \|z_k - w_{k+1}\|^2 \\
&\le \frac{2}{1 - 2\eta^2 L^2} \left(\|w_0 - z_0\|^2 + \sum_{k=0}^{T} \|z_k - w_{k+1}\|^2 \right).
\end{aligned}$$

This completes the proof of Equation (34). $\square$

Back to the proof of Lemma 15. For all $k \ge 0$, we have

$$\begin{aligned}
\|z_{k+1} - z^*\|^2 &= \|z_{k+1} - z_k + z_k - z^*\|^2 \\
&= \|z_k - z^*\|^2 + \|z_{k+1} - z_k\|^2 + 2\langle z_{k+1} - z_k, z_k - z^* \rangle \\
&= \|z_k - z^*\|^2 - \|z_{k+1} - z_k\|^2 + 2\langle z_{k+1} - z_k, z_{k+1} - z^* \rangle \\
&\le \|z_k - z^*\|^2 - \|z_{k+1} - z_k\|^2 - 2\eta \langle F(w_{k+1}), z_{k+1} - z^* \rangle. \qquad (37)
\end{aligned}$$

The last inequality follows from $\langle z_{k+1} - z_k + \eta F(w_{k+1}), z_{k+1} - z^* \rangle \le 0$ as $z_{k+1} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(w_{k+1})\right]$.

Similarly, for all $k \ge 0$, we have

$$\begin{aligned}
\|z_{k+1} - w_{k+1}\|^2 &= \|z_{k+1} - z_k + z_k - w_{k+1}\|^2 \\
&= \|z_{k+1} - z_k\|^2 + \|z_k - w_{k+1}\|^2 + 2\langle z_k - w_{k+1}, z_{k+1} - z_k \rangle \\
&= \|z_{k+1} - z_k\|^2 - \|z_k - w_{k+1}\|^2 + 2\langle z_k - w_{k+1}, z_{k+1} - w_{k+1} \rangle \\
&\le \|z_{k+1} - z_k\|^2 - \|z_k - w_{k+1}\|^2 + 2\eta \langle F(w_k), z_{k+1} - w_{k+1} \rangle. \qquad (38)
\end{aligned}$$

The last inequality follows from $\langle z_k - \eta F(w_k) - w_{k+1}, z_{k+1} - w_{k+1} \rangle \le 0$ as $w_{k+1} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(w_k)\right]$.

We can further simplify Equation (37) using Fact 1:

$$\begin{aligned}
\|z_{k+1} - z^*\|^2 &\le \|z_k - z^*\|^2 - \|z_{k+1} - z_k\|^2 - 2\eta \langle F(w_{k+1}), z_{k+1} - z^* \rangle \\
&= \|z_k - z^*\|^2 - \|z_{k+1} - z_k\|^2 - 2\eta \langle F(w_{k+1}), z_{k+1} - w_{k+1} \rangle + 2\eta \langle F(w_{k+1}), z^* - w_{k+1} \rangle \\
&\le \|z_k - z^*\|^2 - \|z_{k+1} - z_k\|^2 - 2\eta \langle F(w_{k+1}), z_{k+1} - w_{k+1} \rangle. \qquad (39)
\end{aligned}$$

Summing Equation (38) and Equation (39), we get

$$\begin{aligned}
\|z_{k+1} - z^*\|^2 &\le \|z_k - z^*\|^2 - \|z_k - w_{k+1}\|^2 - \|z_{k+1} - w_{k+1}\|^2 + 2\eta \langle F(w_k) - F(w_{k+1}), z_{k+1} - w_{k+1} \rangle \\
&\le \|z_k - z^*\|^2 - \|z_k - w_{k+1}\|^2 - \|z_{k+1} - w_{k+1}\|^2 + 2\eta \|F(w_k) - F(w_{k+1})\| \|z_{k+1} - w_{k+1}\| \\
&\le \|z_k - z^*\|^2 - \|z_k - w_{k+1}\|^2 - \|z_{k+1} - w_{k+1}\|^2 + 2\eta L \|w_k - w_{k+1}\| \|z_{k+1} - w_{k+1}\| \\
&\le \|z_k - z^*\|^2 - \|z_k - w_{k+1}\|^2 + \eta^2 L^2 \|w_k - w_{k+1}\|^2, \qquad (40)
\end{aligned}$$

where we use Cauchy-Schwarz inequality in the second inequality and L-Lipschitzness of $F(\cdot)$ in the third inequality. In the last inequality, we optimize the quadratic function in $\|z_{k+1} - w_{k+1}\|$.

Summing Equation (40) for $k = 0, 1, \cdots, T$ and using Lemma 16, we get

$$\begin{aligned}
\|z_{T+1} - z^*\|^2 &\le \|z_0 - z^*\|^2 - \sum_{k=0}^{T} \|z_k - w_{k+1}\|^2 + \eta^2 L^2 \sum_{k=0}^{T} \|w_k - w_{k+1}\|^2 \\
&\le \|z_0 - z^*\|^2 - \sum_{k=0}^{T} \|z_k - w_{k+1}\|^2 + \frac{2\eta^2 L^2}{1 - 2\eta^2 L^2} \left(\|w_0 - z_0\|^2 + \sum_{k=0}^{T} \|z_k - w_{k+1}\|^2 \right) \\
&\qquad \text{(Lemma 16)} \\
&= \|z_0 - z^*\|^2 - \frac{1 - 4\eta^2 L^2}{1 - 2\eta^2 L^2} \sum_{k=0}^{T} \|z_k - w_{k+1}\|^2 + \frac{2\eta^2 L^2}{1 - 2\eta^2 L^2} \|w_0 - z_0\|^2.
\end{aligned}$$

Since $\eta^2 L^2 < \frac{1}{4}$, we complete the proof by rearranging the above inequality. ■

H.1.1 Bounded Iterates of OG with Constant Step Size

Corollary 3. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator, and z^* be a solution to the corresponding VI. Let $z_0, w_0 \in \mathcal{Z}$ be arbitrary starting points and $\{z_k, w_k\}_{k \ge 0}$ be the iterates of the OG algorithm with step size $\eta \in (0, \frac{1}{2L})$. Then for all $k \ge 1$,*

$$\begin{aligned}
\|z_k - z^*\| &\le \sqrt{\frac{1 - 2\eta^2 L^2}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2 + \frac{2\eta^2 L^2}{1 - 4\eta^2 L^2} \|w_0 - z_0\|^2}, \\
\|w_k - z^*\| &\le 2 \cdot \sqrt{\frac{1 - 2\eta^2 L^2}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2 + \frac{2\eta^2 L^2}{1 - 4\eta^2 L^2} \|w_0 - z_0\|^2}.
\end{aligned}$$

Proof. By Lemma 15 for $k \ge 1$ we have that,

$$\|z_k - z^*\| \le \sqrt{\frac{1 - 2\eta^2 L^2}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2 + \frac{2\eta^2 L^2}{1 - 4\eta^2 L^2} \|w_0 - z_0\|^2}.$$

Since $\frac{1-2\eta^2 L^2}{1-4\eta^2 L^2} \ge 1$, $\|z_0 - z^*\| \le \sqrt{\frac{1-2\eta^2 L^2}{1-4\eta^2 L^2} \|z_0 - z^*\|^2 + \frac{2\eta^2 L^2}{1-4\eta^2 L^2} \|w_0 - z_0\|^2}$, which implies that for all $k \ge 0$,

$$\|z_k - z^*\| \le \sqrt{\frac{1 - 2\eta^2 L^2}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2 + \frac{2\eta^2 L^2}{1 - 4\eta^2 L^2} \|w_0 - z_0\|^2}. \tag{41}$$

For the second part of the proof, by Lemma 15 for all $k \ge 1$ we have that,

$$\|w_k - z_{k-1}\| \le \sqrt{\frac{1 - 2\eta^2 L^2}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2 + \frac{2\eta^2 L^2}{1 - 4\eta^2 L^2} \|w_0 - z_0\|^2}. \tag{42}$$

For all $k \ge 1$, by triangle inequality, Inequality (41) and Inequality (42) we have that,

$$\|w_k - z^*\| \le \|w_k - z_{k-1}\| + \|z_{k-1} - z^*\| \le 2 \cdot \sqrt{\frac{1 - 2\eta^2 L^2}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2 + \frac{2\eta^2 L^2}{1 - 4\eta^2 L^2} \|w_0 - z_0\|^2},$$

which concludes the proof. □

H.2 Best-Iterate of $\Phi(z_k, w_k)$

In this section, we use Lemma 15 to show that there exists $t^* \in [T]$ such that $\Phi(z_{t^*}, w_{t^*}) = O(\frac{1}{T})$.

Lemma 17. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator, and z^* be a solution to the corresponding monotone VI. Let $z_0, w_0 \in \mathcal{Z}$ be arbitrary starting point and $\{z_k, w_k\}_{k\geq 0}$ be the iterates of the OG algorithm with step size $\eta \in (0, \frac{1}{2L})$. Then for all $k \geq 1$,*

$$\sum_{k=1}^{T} \left(\|\eta F(z_k) - \eta F(w_k)\|^2 + \eta^2 r^{tan}(z_k)^2 \right) \leq \frac{4 + 6\eta^4 L^4}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2 + \frac{16\eta^2 L^2 + 6\eta^4 L^4}{1 - 4\eta^2 L^2} \|w_0 - z_0\|^2.$$

Moreover, when $w_0 = z_0$

$$\sum_{k=1}^{T} \left(\|\eta F(z_k) - \eta F(w_k)\|^2 + \eta^2 r^{tan}(z_k)^2 \right) \leq \frac{4 + 6\eta^4 L^4}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2.$$

Proof of Lemma 17: For all $k \geq 1$, we have

$$\begin{aligned}
\|\eta F(z_k) - \eta F(w_k)\|^2 &\leq \eta^2 L^2 \|z_k - w_k\|^2 && (L\text{-Lipschitzness of } F) \\
&\leq \eta^4 L^4 \|w_{k-1} - w_k\|^2. && (\text{Equation (36)})
\end{aligned}$$

Using Lemma 6 with the fact that $z_k = \Pi_{\mathcal{Z}}[z_{k-1} - \eta F(w_k)]$, we have for all $k \geq 1$,

$$\begin{aligned}
\eta^2 r^{tan}(z_k)^2 &\leq \|z_{k-1} - z_k + \eta F(z_k) - \eta F(w_k)\|^2 \\
&\leq 2\|z_{k-1} - z_k\|^2 + 2\eta^2 \|F(z_k) - F(w_k)\|^2 \\
&\leq 2\|z_{k-1} - w_k + w_k - z_k\|^2 + 2\eta^2 L^2 \|w_k - z_k\|^2 && (L\text{-Lipschitzness of } F) \\
&\leq 4\|z_{k-1} - w_k\|^2 + (4 + 2\eta^2 L^2)\|w_k - z_k\|^2 \\
&\leq 4\|z_{k-1} - w_k\|^2 + (4 + 2\eta^2 L^2)\eta^2 L^2 \|w_{k-1} - w_k\|^2. && (\text{Equation (36)})
\end{aligned}$$

Summing the above inequalities with $k = 1, \cdots, T$ and using Lemma 15 and Lemma 16, we have

$$\begin{aligned}
&\sum_{k=1}^{T} \left(\|\eta F(z_k) - \eta F(w_k)\|^2 + \eta^2 r^{tan}(z_k)^2 \right) \\
&\leq 4 \sum_{k=0}^{T-1} \|z_k - w_{k+1}\|^2 + (4 + 3\eta^2 L^2)\eta^2 L^2 \sum_{k=0}^{T-1} \|w_k - w_{k+1}\|^2 \\
&\leq \frac{2(4 + 3\eta^2 L^2)\eta^2 L^2}{1 - 2\eta^2 L^2} \|w_0 - z_0\|^2 + \left(4 + \frac{2(4 + 3\eta^2 L^2)\eta^2 L^2}{1 - 2\eta^2 L^2} \right) \sum_{k=0}^{T-1} \|z_k - w_{k+1}\|^2 \\
&\leq \frac{2(4 + 3\eta^2 L^2)\eta^2 L^2}{1 - 2\eta^2 L^2} \|w_0 - z_0\|^2 + \left(\frac{8\eta^2 L^2}{1 - 4\eta^2 L^2} + \frac{4(4 + 3\eta^2 L^2)\eta^4 L^4}{(1 - 2\eta^2 L^2) \cdot (1 - 4\eta^2 L^2)} \right) \|w_0 - z_0\|^2 \\
&\quad + \left(\frac{4 - 8\eta^2 L^2}{1 - 4\eta^2 L^2} + \frac{2(4 + 3\eta^2 L^2)\eta^2 L^2}{1 - 4\eta^2 L^2} \right) \|z_0 - z^*\|^2 \\
&= \frac{16\eta^2 L^2 + 6\eta^4 L^4}{1 - 4\eta^2 L^2} \|w_0 - z_0\|^2 + \frac{4 + 6\eta^4 L^4}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2,
\end{aligned}$$

which concludes the proof. ■

The following is a corollary of Lemma 17.

Corollary 4. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator, and z^* be a solution to the corresponding VI. Let $z_0, w_0 \in \mathcal{Z}$ be arbitrary starting point and $\{z_k, w_k\}_{k\geq 0}$ be the iterates of the OG algorithm with step size $\eta \in (0, \frac{1}{2L})$. Then for all $T \geq 1$, there exists $t^* \in [T]$ such that*

$$\|\eta F(z_{t^*}) - \eta F(w_{t^*})\|^2 + \eta^2 r^{tan}(z_{t^*})^2 \leq \frac{1}{T} \frac{4 + 6\eta^4 L^4}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2 + \frac{1}{T} \frac{16\eta^2 L^2 + 6\eta^4 L^4}{1 - 4\eta^2 L^2} \|w_0 - z_0\|^2.$$

Moreover, when $w_0 = z_0$

$$\|\eta F(z_{t^*}) - \eta F(w_{t^*})\|^2 + \eta^2 r^{tan}(z_{t^*})^2 \leq \frac{1}{T} \frac{4 + 6\eta^4 L^4}{1 - 4\eta^2 L^2} \|z_0 - z^*\|^2.$$

H.3 Monotonicity of the Potential

In this section, we show that the potential function $\Phi(z_k, w_k)$ is monotonically decreasing across iterates of OG. We only include the simplified proof discovered using a degree 2 SOS program.

Theorem 7. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set and $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator. Then for any $z_k, w_k \in \mathcal{Z}$, the OG algorithm with step size $\eta \in (0, \frac{1}{2L})$ produces $w_{k+1}, z_{k+1} \in \mathcal{Z}$ that satisfy $\|F(z_k) - F(w_k)\|^2 + r^{tan}(z_k)^2 \geq \|F(z_{k+1}) - F(w_{k+1})\|^2 + r^{tan}(z_{k+1})^2$.*

Proof. Let $c_k = \Pi_{N_{\mathcal{Z}}(z_k)}(-F(z_k))$ and $c_{k+1} = \Pi_{N_{\mathcal{Z}}(z_{k+1})}(-F(z_{k+1}))$. Lemma 13 implies that

$$\begin{aligned}
&\eta^2 r^{tan}(z_k)^2 + \eta^2\|F(z_k) - F(w_k)\|^2 - \left(\eta^2 r^{tan}(z_{k+1})^2 + \eta^2\|F(z_{k+1}) - F(w_{k+1})\|^2\right) \\
&= \|\eta F(z_k) + \eta c_k\|^2 + \eta^2\|F(z_k) - F(w_k)\|^2 \\
&\quad - \left(\|\eta F(z_{k+1}) + \eta c_{k+1}\|^2 + \eta^2\|F(z_{k+1}) - F(w_{k+1})\|^2\right) \qquad (43)
\end{aligned}$$

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

$$(-2) \cdot (\langle \eta F(z_{k+1}) - \eta F(z_k), z_{k+1} - z_k \rangle) \leq 0, \qquad (44)$$

$$(-2) \cdot \left(\frac{1}{4}\|z_{k+1} - w_{k+1}\|^2 - \|\eta F(z_{k+1}) - \eta F(w_{k+1})\|^2\right) \leq 0. \qquad (45)$$

Since $w_{k+1} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(w_k)\right]$ and $z_{k+1} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(w_{k+1})\right]$, we have that $z_k - \eta F(w_k) - w_{k+1} \in N(w_{k+1})$ and $z_k - \eta F(w_{k+1}) - z_{k+1} \in N(z_{k+1})$. Thus we have

$$(-1) \cdot \langle z_k - \eta F(w_k) - w_{k+1}, w_{k+1} - z_{k+1} \rangle \leq 0, \qquad (46)$$

$$(-2) \cdot \langle z_k - \eta F(w_{k+1}) - z_{k+1}, z_{k+1} - z_k \rangle \leq 0. \qquad (47)$$

Since $c(z_k) \in N(z_k)$, we have that

$$(-1) \cdot \langle \eta c(z_k), z_k - w_{k+1} \rangle \leq 0, \qquad (48)$$

$$(-1) \cdot \langle \eta c(z_k), z_k - z_{k+1} \rangle \leq 0. \qquad (49)$$

According to Lemma 13 and the fact that $z_k - \eta F(w_{k+1}) - z_{k+1} \in N(z_{k+1})$, $c_{k+1} \in \Pi_{N(z_{k+1})}(-F(z_{k+1}))$ we have

$$(-2) \cdot \langle \eta c(z_{k+1}) + \eta F(z_{k+1}), z_k - \eta F(w_{k+1}) - z_{k+1} \rangle \leq 0, \qquad (50)$$

$$(-2) \cdot \langle \eta c(z_{k+1}) + \eta F(z_{k+1}), -\eta c(z_{k+1}) \rangle = 0,. \qquad (51)$$

MATLAB code for the verification of the following identity is included in the supplementary material under the name "verify_identity_OG.m".

Expression (43) + LHS of Inequality (44) + LHS of Inequality (45) + LHS of Inequality (46)
+ LHS of Inequality (47) + LHS of Inequality (48) + LHS of Inequality (49)
+ LHS of Inequality (51) + LHS of Inequality (50)

$$= \left\|\frac{w_{k+1} - z_{k+1}}{2} + \eta F(w_k) - \eta F(z_k)\right\|^2 \qquad (52)$$

$$+ \left\|\eta F(z_k) + \eta c(z_k) - z_k + \frac{w_{k+1} + z_{k+1}}{2}\right\|^2 \qquad (53)$$

$$+ \|z_k - \eta F(w_{k+1}) - z_{k+1} - \eta c(z_{k+1})\|^2 \qquad (54)$$

$$\geq 0.$$

Thus, $\|F(z_k) - F(w_k)\|^2 + r^{tan}(z_k)^2 \geq \|F(z_{k+1}) - F(w_{k+1})\|^2 + r^{tan}(z_{k+1})^2$. □

H.4 Combining Everything

In this section, we combine the results of the previous sections and show that $\Phi(z_T, w_T) = O\left(\frac{1}{T}\right)$ and we show the last-iterate convergence rate for performance measures of iterest.

Lemma 18. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator. For any $z_k, w_k \in \mathcal{Z}$, the OG algorithm update satisfies,*

$$r^{tan}_{(F,\mathcal{Z})}(w_{k+1}) \leq \sqrt{2}(2+\eta L)\sqrt{r^{tan}_{(F,\mathcal{Z})}(z_k)^2 + \|F(w_k) - F(z_k)\|^2}.$$

Proof. Since $w_{k+1} = \Pi_{\mathcal{Z}}[z_k - F(w_k)]$, by using Lemma 6 we have

$$\begin{aligned} r^{tan}_{(F,\mathcal{Z})}(w_{k+1}) &\leq \left\| \frac{z_k - w_{k+1}}{\eta} + F(w_{k+1}) - F(w_k) \right\| \\ &\leq \left\| \frac{z_k - w_{k+1}}{\eta} \right\| + \|F(w_k) - F(z_k)\| + \|F(z_k) - F(w_{k+1})\| \\ &\leq \frac{1+\eta L}{\eta}\|z_k - w_{k+1}\| + \|F(w_k) - F(z_k)\|. \qquad (L\text{-Lipschitzness of } F) \end{aligned}$$

Using Lemma 7 and the non-expansiveness of the projection mapping, we have

$$\begin{aligned} \|z_k - w_{k+1}\| &\leq \|z_k - \Pi_{\mathcal{Z}}[z_k - \eta F(z_k)]\| + \|\Pi_{\mathcal{Z}}[z_k - \eta F(z_k)] - w_{k+1}\| \\ &= r^{nat}_{(\eta F,\mathcal{Z})}(z_k) + \|\Pi_{\mathcal{Z}}[z_k - \eta F(z_k)] - \Pi_{\mathcal{Z}}[z_k - \eta F(w_k)]\| \\ &\leq r^{tan}_{(\eta F,\mathcal{Z})}(z_k) + \eta\|F(z_k) - F(w_k)\| \\ &= \eta r^{tan}_{(F,\mathcal{Z})}(z_k) + \eta\|F(z_k) - F(w_k)\|. \end{aligned}$$

Combing the above two inequalities, we have

$$\begin{aligned} r^{tan}_{(F,\mathcal{Z})}(w_{k+1}) &\leq (1+\eta L) r^{tan}_{(F,\mathcal{Z})}(z_k) + (2+\eta L)\|F(w_k) - F(z_k)\| \\ &\leq \sqrt{2}(2+\eta L)\sqrt{r^{tan}_{(F,\mathcal{Z})}(z_k)^2 + \|F(w_k) - F(z_k)\|^2}. \qquad (a + b \leq \sqrt{2}\sqrt{a^2+b^2}) \end{aligned}$$

□

Combining Corollary 4, Theorem 7, Lemma 18, Lemma 7 and Lemma 4 we get $\mathcal{O}(\frac{1}{\sqrt{T}})$ last-iterate convergence rate in terms of the tangent residual, natural residual and gap function for both z_T and w_{T+1}. The result is formally stated in Theorem 8.

Theorem 8. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator, and z^* be a solution to the corresponding VI. Let $z_0, w_0 \in \mathcal{Z}$ be arbitrary starting point, $\{z_k, w_k\}_{k \geq 0}$ be the iterates of the OG algorithm with step size $\eta \in (0, \frac{1}{2L})$ and $D_0 := \sqrt{(4+6\eta^4L^4)\|z_0 - z^*\|^2 + (16\eta^2L^2 + 6\eta^4L^4)\|w_0 - z_0\|^2}$. Then for all $T \geq 1$,*

- $\sqrt{\Phi(z_T, w_T)} \leq \frac{1}{\sqrt{T}} \frac{D_0}{\eta\sqrt{1-4\eta^2L^2}}$
- $\mathrm{GAP}_{F,\mathcal{Z},D}(z_T) \leq \frac{1}{\sqrt{T}} \cdot \frac{D \cdot D_0}{\eta \cdot \sqrt{1-4\cdot(\eta L)^2}}$,
- $r^{nat}_{F,\mathcal{Z}}(z_T) \leq r^{tan}_{F,\mathcal{Z}}(z_T) \leq \frac{1}{\sqrt{T}} \cdot \frac{D_0}{\eta \cdot \sqrt{1-4\cdot(\eta L)^2}}$,
- $\mathrm{GAP}_{F,\mathcal{Z},D}(w_{T+1}) \leq \frac{1}{\sqrt{T}} \cdot \frac{\sqrt{2}(2+\eta L) \cdot D \cdot D_0}{\eta \cdot \sqrt{1-4\cdot(\eta L)^2}}$,
- $r^{nat}_{F,\mathcal{Z}}(w_{T+1}) \leq r^{tan}_{F,\mathcal{Z}}(w_{T+1}) \leq \frac{1}{\sqrt{T}} \cdot \frac{\sqrt{2}(2+\eta L) D_0}{\eta \cdot \sqrt{1-4\cdot(\eta L)^2}}$.

I Last-Iterate Convergence for Variational Inequalities

Theorem 9. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F(\cdot) : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator and $z^* \in \mathcal{Z}$ be a solution to the corresponding VI. Then for any $T \geq 1$, z_T produced by EG with any constant step size $\eta \in (0, \frac{1}{L})$ satisfies*

- $\textsc{Gap}(z_T) \leq \frac{1}{\sqrt{T}} \frac{3D\|z_0 - z^*\|}{\eta\sqrt{1-(\eta L)^2}}$,
- $r^{nat}(z_T) \leq r^{tan}(z_T) \leq \frac{1}{\sqrt{T}} \frac{3\|z_0 - z^*\|}{\eta\sqrt{1-(\eta L)^2}}$.
- $\max\{\|z_T - z_{T+\frac{1}{2}}\|, \frac{\|z_T - z_{T+1}\|}{2}\} \leq \frac{1}{\sqrt{T}} \frac{3\|z_0 - z^*\|}{\sqrt{1-(\eta L)^2}}$.

Proof. The proof follows by combining Theorem 6 and Lemma 19. □

Theorem 10. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator and $z^* \in \mathcal{Z}$ a solution to the corresponding VI. Let $z_0, w_0 \in \mathcal{Z}$ be arbitrary starting point and $\{z_k, w_k\}_{k\geq 0}$ be the iterates of the OG algorithm with any step size $\eta \in (0, \frac{1}{2L})$. Let $D_0 := \frac{\sqrt{(4+6\eta^4 L^4)\|z_0 - z^*\|^2 + (16\eta^2 L^2 + 6\eta^4 L^4)\|w_0 - z_0\|^2}}{\sqrt{1-4(\eta L)^2}}$. Then for any $T \geq 1$,*

- $\textsc{Gap}_{\mathcal{Z},F,D}(z_T) \leq \frac{DD_0}{\eta\sqrt{T}}$,
- $r^{nat}_{\mathcal{Z},F,D}(z_T) \leq r^{tan}_{\mathcal{Z},F,D}(z_T) \leq \frac{D_0}{\eta\sqrt{T}}$,
- $\|z_T - z_{T+1}\| \leq \frac{\sqrt{3}D_0}{\sqrt{T}}$,
- $\textsc{Gap}_{\mathcal{Z},F,D}(w_{T+1}) \leq \frac{\sqrt{2}(2+\eta L)\cdot D \cdot D_0}{\eta\sqrt{T}}$,
- $r^{nat}_{\mathcal{Z},F,D}(w_{T+1}) \leq r^{tan}_{\mathcal{Z},F,D}(w_{T+1}) \leq \frac{\sqrt{2}(2+\eta L)D_0}{\eta\sqrt{T}}$,
- $\|w_T - w_{T+1}\| \leq \frac{5D_0}{\sqrt{T-1}}$.

Proof. The proof follows by Theorem 8 and Lemma 20. □

I.1 Auxiliary Lemmas

Lemma 19. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set and $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator. For any $z_k \in \mathcal{Z}$, the EG algorithm update with step-size $\eta \in (0, \frac{1}{L})$ satisfies,*

$$\begin{aligned} \max\left\{\left\|z_k - z_{k+\frac{1}{2}}\right\|, \left\|z_{k+\frac{1}{2}} - z_{k+1}\right\|\right\} &\leq \eta r^{tan}_{(F,\mathcal{Z})}(z_k), \\ \|z_k - z_{k+1}\| &\leq 2 \cdot \eta r^{tan}_{(F,\mathcal{Z})}(z_k). \end{aligned}$$

Proof. We are only going to prove that $\max\left\{\|z_k - z_{k+\frac{1}{2}}\|, \|z_{k+\frac{1}{2}} - z_{k+1}\|\right\} \leq \eta r^{tan}_{(F,\mathcal{Z})}(z_k)$, since inequality $\|z_k - z_{k+1}\| \leq 2 \cdot \eta r^{tan}_{(F,\mathcal{Z})}(z_k)$ follows by triangle inequality.

Since $z_{k+\frac{1}{2}} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(z_k)\right]$, by Definition 7 we have that

$$\left\|z_k - z_{k+\frac{1}{2}}\right\| = r^{nat}_{(\eta F,\mathcal{Z})}(z_k) \leq r^{tan}_{(\eta F,\mathcal{Z})}(z_k) = \eta r^{tan}_{(F,\mathcal{Z})}(z_k), \tag{55}$$

where the first inequality follow by Lemma 7 and the second equality follows by Definition 4.

Moreover, since $z_{k+1} = \Pi_{\mathcal{Z}}\left[z_k - \eta F(z_{k+\frac{1}{2}})\right]$, by non-expansiveness of the projection mapping, the fact that F is L-Lipschitz, that $\eta L \leq 1$ and Inequality (55), we have

$$\left\|z_{k+\frac{1}{2}} - z_{k+1}\right\| \leq \left\|\eta F(z_{k+\frac{1}{2}}) - \eta F(z_k)\right\| \leq \eta L \left\|z_k - z_{k+\frac{1}{2}}\right\| \leq \eta r^{tan}_{(F,\mathcal{Z})}(z_k).$$

□

Lemma 20. *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set, $F : \mathcal{Z} \to \mathbb{R}^n$ be a monotone and L-Lipschitz operator, and z^* be a solution to the corresponding VI. Let $z_0, w_0 \in \mathcal{Z}$ be arbitrary starting point and $\{z_k, w_k\}_{k\geq 0}$ be the iterates of the OG algorithm with step size $\eta \in (0, \frac{1}{2L})$. Then for all $k \geq 1$,*

$$\begin{aligned}
\|z_k - z_{k+1}\| \leq& \sqrt{3}\eta\sqrt{\Phi(z_k, w_k)},\\
\|w_k - w_{k+1}\| \leq& 5\eta\sqrt{\Phi(z_{k-1}, w_{k-1})}.
\end{aligned}$$

Proof. By Lemma 14 and the fact that $w_{k+1} = \Pi_{\mathcal{Z}}(z_k - \eta F(w_k))$ for all $k \geq 0$, we have that,

$$\|w_{k+1} - z_k\|^2 \leq 2\left(\eta^2 r^{tan}(z_k)^2 + \eta^2\|F(w_k) - F(z_k)\|^2\right) = 2\eta^2\Phi(z_k, w_k). \tag{56}$$

Thus, for all $k \geq 0$, by combining Lemma 14, the fact that $z_{k+1} = \Pi_{\mathcal{Z}}(z_k - \eta F(w_{k+1})$, L-Lipschitzness of F, Inequality (56), the fact that $\eta^2 L^2 \leq \frac{1}{4}$ we have that for all $k \geq 0$,

$$\begin{aligned}
\|z_{k+1} - z_k\|^2 \leq& 2\left(\eta^2 r^{tan}(z_k)^2 + \eta^2\|F(w_{k+1}) - F(z_k)\|^2\right)\\
\leq& 2\left(\eta^2 r^{tan}(z_k)^2 + \eta^2 L^2\|w_{k+1} - z_k\|^2\right)\\
\leq& 2\eta^2\left(r^{tan}(z_k)^2 + 2\cdot\eta^2 L^2\Phi(z_k, w_k)\right)\\
\leq& 2\eta^2\left(r^{tan}(z_k)^2 + \frac{\Phi(z_k, w_k)}{2}\right)\\
\leq& 3\cdot\eta^2\Phi(z_k, w_k).
\end{aligned} \tag{57}$$

Moreover for all $k \geq 1$, by triangle inequality, Inequality (56), Inequality (57) and Theorem 7, we have that,

$$\begin{aligned}
\|w_{k+1} - w_k\| \leq& \|w_{k+1} - z_k\| + \|z_k - z_{k-1}\| + \|w_k - z_{k-1}\|\\
\leq& \eta\cdot\sqrt{2\Phi(z_k, w_k)} + \eta\cdot\sqrt{3\Phi(z_{k-1}, w_{k-1})} + \eta\sqrt{2\Phi(z_{k-1}, w_{k-1})}\\
\leq& 5\eta\sqrt{\Phi(z_{k-1}, w_{k-1})}.
\end{aligned}$$

□

J Non-Monotonicity of Several Standard Performance Measures

We conduct numerical experiments by trying to find saddle points in constrained bilinear games using EG, and verified that the following performance measures are not monotone: the (squared) natural residual, $\|z_k - z_{k+\frac{1}{2}}\|^2$, $\|z_k - z_{k+1}\|^2$, $\max_{z\in\mathcal{Z}}\langle F(z), z_k - z\rangle$, $\max_{z\in\mathcal{Z}}\langle F(z_k), z_k - z\rangle$.

All of our counterexamples are constructed by trying to find a saddle point in bilinear games of the following form:

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}} x^\top Ay - b^\top x - c^\top y \tag{58}$$

where $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^2$, A is a 2×2 matrix and b, c are 2-dimensional column vectors. All of the instances of the bilinear game considered in this section have $\mathcal{X}, \mathcal{Y} = [0, 10]^2$. We denote by $\mathcal{Z} = \mathcal{X}\times\mathcal{Y}$ and by $F(x, y) = \begin{pmatrix} Ay - b \\ -A^\top x + c\end{pmatrix} : \mathcal{Z} \to \mathbb{R}^n$. We remind readers that finding a saddle point of bilinear game (58), is equivalent to solving the corresponding monotone VI with operator $F(z)$ on set $\mathcal{Z}$.

J.1 Non-Monotonicity of the Natural Residual and its Variants

Performance Measure: Natural Residual. Let $A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$, $b = c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Running the EG algorithm on the corresponding VI problem with step-size $\eta = 0.1$ starting at $z_0 = (0.3108455, 0.4825575, 0.4621875, 0.5768655)^T$ has the following trajectory:

$$z_1 = (0.24923465, 0.47967569, 0.43497808, 0.57458145)^T,$$
$$z_2 = (0.19396855, 0.48164918, 0.40193211, 0.56061753)^T.$$

Thus we have

$$\begin{aligned} r^{nat}(z_0)^2 =& 0.15170013184049996, \\ r^{nat}(z_1)^2 =& 0.13617654362050116, \\ r^{nat}(z_2)^2 =& 0.16125792556139756. \end{aligned}$$

It is clear that the natural residual is not monotone.In Figure 6, the red line is the squared natural residual and the blue line is the squared tangent residual across many iterations.

Performance Measure: $\|z_k - z_{k+\frac{1}{2}}\|^2$. Note that the norm of the operator mapping defined in [Dia20] is exactly $\frac{1}{\eta} \cdot \|z_k - z_{k+\frac{1}{2}}\|$. Let $A = \begin{bmatrix} 0.50676631 & 0.15042569 \\ 0.46897595 & 0.96748026 \end{bmatrix}$, $b = c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Running the EG algorithm on the corresponding VI problem with step-size $\eta = 0.1$ starting at $z_0 = (2.35037432, 0.00333996, 1.70547279, 0.71065999)^T$ has the following trajectory:

$$\begin{aligned} z_{\frac{1}{2}} =& (2.35325656, 0, 1.72473848, 0.64633879)^T, \\ z_1 =& (2.35324779, 0, 1.72472791, 0.64605901)^T, \\ z_{1+\frac{1}{2}} =& (2.35612601, 0, 1.74398258, 0.58145791)^T \\ z_2 =& (2.35612201, 0, 1.74412844, 0.5815012)^T, \\ z_{2+\frac{1}{2}} =& (2.35898819, 0, 1.76352876, 0.51694333)^T. \end{aligned}$$

Thus we have

$$\begin{aligned} \left\| z_0 - z_{\frac{1}{2}} \right\|^2 =& 0.00452784581555656, \\ \left\| z_1 - z_{1+\frac{1}{2}} \right\|^2 =& 0.004552329544896258, \\ \left\| z_2 - z_{2+\frac{1}{2}} \right\|^2 =& 0.004552306444552208. \end{aligned}$$

It is clear that the $\|z_k - z_{k+\frac{1}{2}}\|^2$ is not monotone. In Figure 7, the red line is $\frac{\|z_k - z_{k+\frac{1}{2}}\|^2}{\eta^2}$ and the blue line is the squared tangent residual across many iterations.

Performance Measure: $\|z_k - z_{k+1}\|^2$. Let $A = \begin{bmatrix} 0.50676631 & 0.15042569 \\ 0.46897595 & 0.96748026 \end{bmatrix}$, $b = c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Running the EG algorithm on the corresponding VI problem with step-size $\eta = 0.1$ starting at $z_0 = (2.37003485, 0, 1.84327237, 0.25934775)^T$ has the following trajectory:

$$\begin{aligned} z_1 =& (2.37267186, 0, 1.86351397, 0.1950396)^T, \\ z_2 =& (2.37524308, 0, 1.88388624, 0.13077023)^T, \\ z_3 =& (2.37774149, 0.00426125, 1.90438549, 0.06653856)^T. \end{aligned}$$

Thus we have

$$
\begin{aligned}
\|z_0 - z_1\|^2 =& 0.004552214685275266, \\
\|z_1 - z_2\|^2 =& 0.004552191904998012, \\
\|z_2 - z_3\|^2 =& 0.004570327450598002.
\end{aligned}
$$

It is clear that the $\|z_k - z_{k+1}\|^2$ is not monotone. In Figure 8, the red line is $\frac{\|z_k - z_{k+1}\|^2}{\eta^2}$ and the blue line is the squared tangent residual across many iterations.

J.2 Non-Monotonicity of the Gap Functions and its Variant

Performance Measure: Gap Function and $\max_{z\in\mathcal{Z}} \langle F(z), z_k - z\rangle$. Let $A = \begin{bmatrix} -0.21025101 & 0.22360196 \\ 0.40667685 & -0.2922158 \end{bmatrix}$, $b = c = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. One can easily verify that $\langle F(z), z_k - z\rangle = \langle F(z_k), z_k - z\rangle$, which further implies that $\max_{z\in\mathcal{Z}} \langle F(z), z_k - z\rangle = \max_{z\in\mathcal{Z}} \langle F(z_k), z_k - z\rangle = \textsc{Gap}(z_k)$, which implies that non-monotonicity of the gap function implies non-monotonicity of $\max_{z\in\mathcal{Z}} \langle F(z), z_k - z\rangle$. Running the EG algorithm on the corresponding VI problem with step-size $\eta = 0.1$ starting at $z_0 = (0.53095379, 0.29084076, 0.62132986, 0.49440498)$ has the following trajectory:

$$
\begin{aligned}
z_1 &= (0.53290086, 0.28009156, 0.62151204, 0.4981395)^T, \\
z_2 &= (0.5347502, 0.26947398, 0.62122195, 0.50222691)^T.
\end{aligned}
$$

One can easily verify that

$$
\begin{aligned}
\textsc{Gap}(z_0) =& 0.6046398415472187, \\
\textsc{Gap}(z_1) =& 0.58462873354003214, \\
\textsc{Gap}(z_2) =& 0.5914026255469654.
\end{aligned}
$$

It is clear that the duality gap is not monotone. In Figure 9, the red line is the gap function and the blue line is the scaled squared tangent residual across many iterations.

Plots for the Numerical Experiments

In Figures 6-10, we plot the values of the non-monotone performance measures of interest as well as the tangent residual properly scaled so that it can fit in the figure for more iterations using the same instances as provided Appendix J.1 and J.2 with starting point $z_0 = (0.25, 0.25, 0.25, 0.25)^T$ and step size $\eta = 0.1$. Note that in Figures 6-9, the blue line always corresponds to (scaled) tangent residual – our potential function, and the red line corresponds to the performance measure stated at the top of the plot.

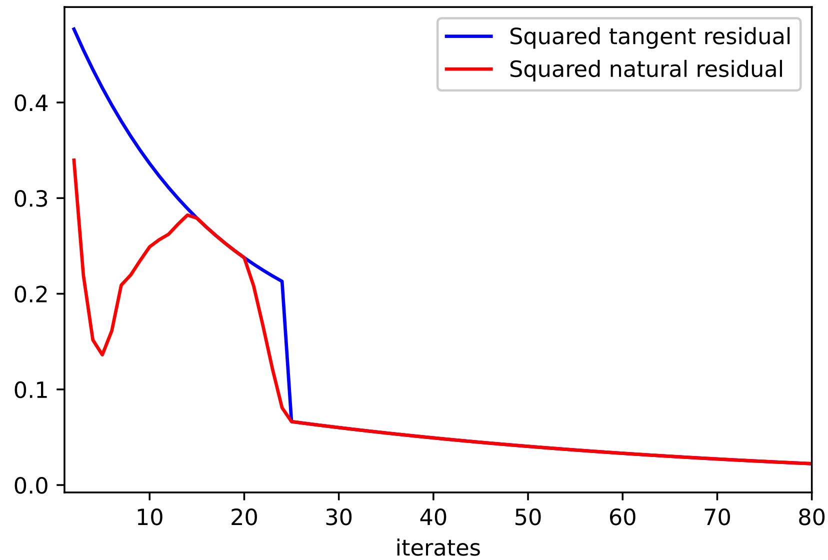

Figure 6: Non-monotonicity of the Natural Residual

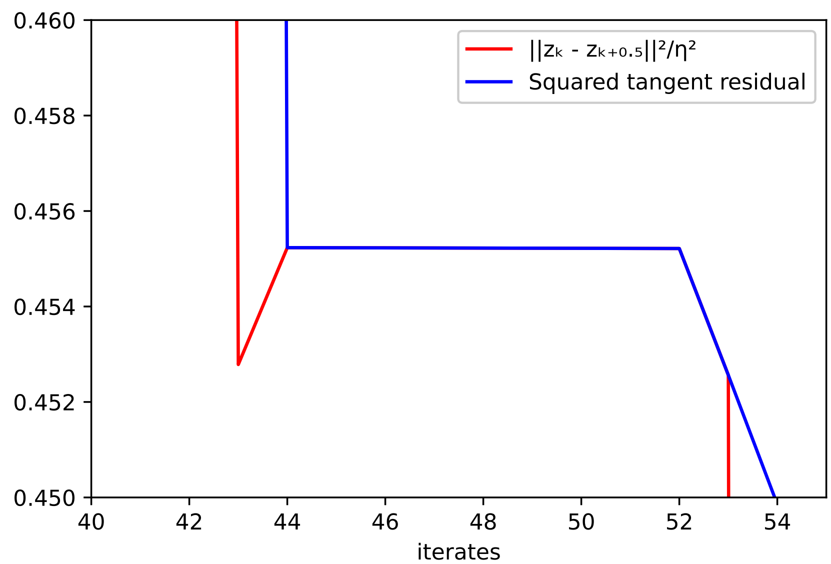

Figure 7: Non-monotonicity of $\|z_k - z_{k+1/2}\|^2$

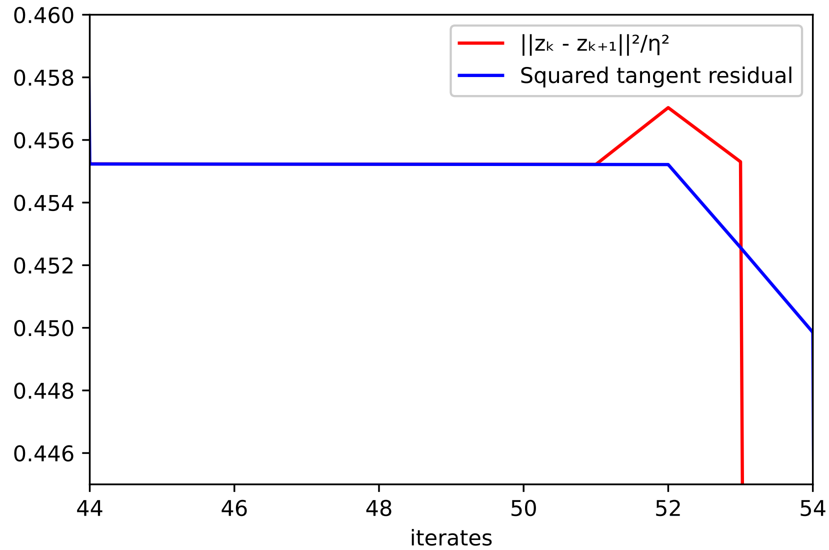

Figure 8: Non-monotonicity of $\|z_k - z_{k+1}\|^2$

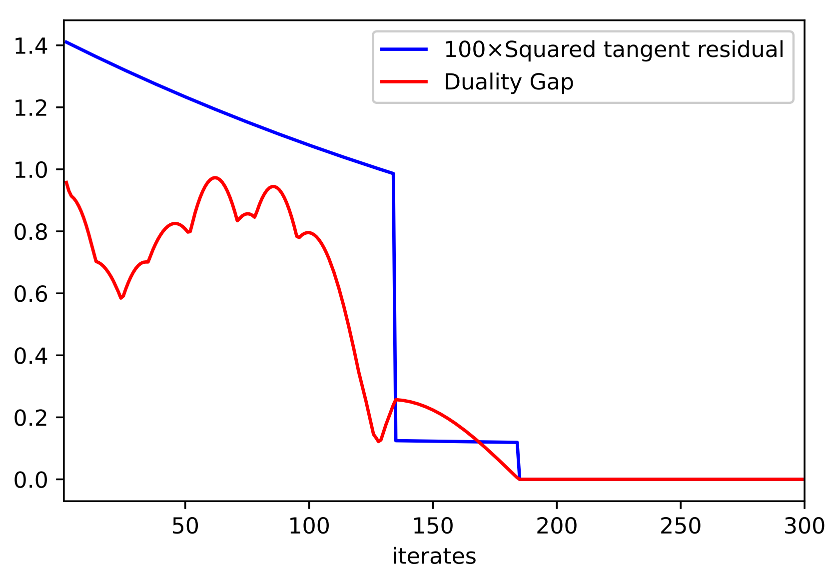

Figure 9: Non-monotonicity of Variants of Gap Functions. Here we have scaled the tangent residual $\times 100$ to make the plot clear.

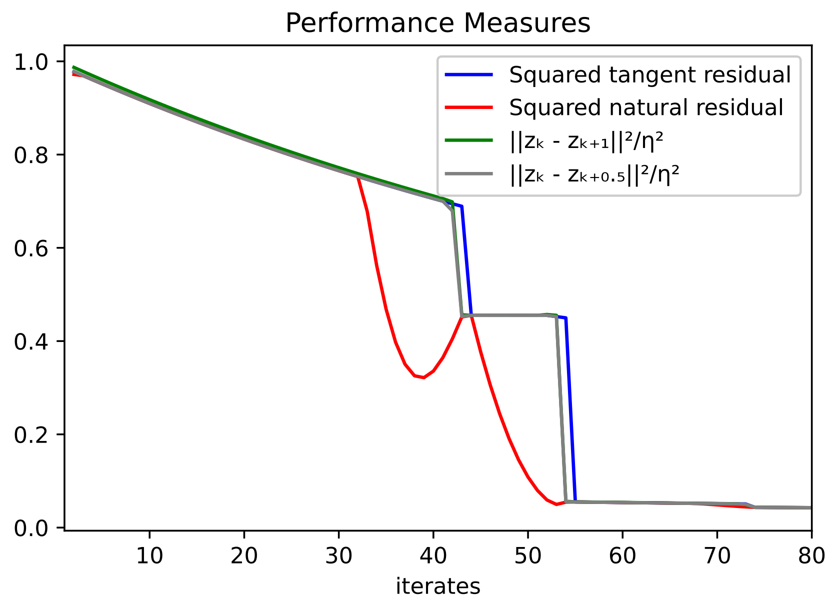

Figure 10: Numerical experiments on bilinear game (58) with $A = \begin{bmatrix} 0.50676631 & 0.15042569 \\ 0.46897595 & 0.96748026 \end{bmatrix}$, $b = c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, initial point $z_0 = (0.25, 0.25, 0.25, 0.25)^\top$ and step size $\eta = 0.1$. This is the same bilinear game we used in Figure 7 and 8.
Performance Measures: the blue line is tangent residual; the red line is natural residual; the gray line is $\|z_k - z_{k+1/2}\|^2/\eta^2$; the green line is $\|z_k - z_{k+1}\|^2/\eta^2$. Non-monotonicity of the natural residual is clear. Non-monotonicity of $\|z_k - z_{k+1/2}\|^2$ and $\|z_k - z_{k+1}\|^2$ are better illustrated in Figure 7 and 8.