## 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 $\Sigma$ 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 [FP07].

**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

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

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

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

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 $\{(0,\dots,0)\}\subsetneq\widehat{N}_{\mathcal{Z}}(z)$,

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

which concludes the proof.

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

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

**(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 $x\in\mathbb{R}^n$,

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

Thus it is clear that we have

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

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

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

Thus

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

**(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 $a \neq (0, \dots, 0)$, $g(a) = \|F(z) + \langle F(z), \frac{a}{\|a\|}\rangle \frac{a}{\|a\|}\|$ and $g(0, \dots, 0) = \|F(z)\|$. 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 $g(0, \dots, 0) = \|F(z)\|$ 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

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

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

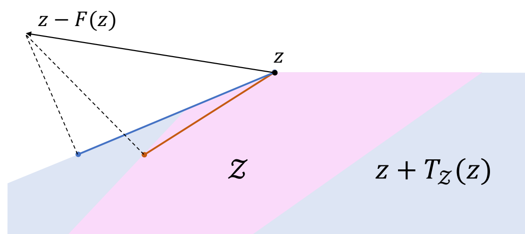


Figure 2: Illustration of the tangent residual and the natural residual. The blue line represents the tangent residual and the red line represents the natural residual. It is clear that the tangent residual upper bounds the natural residual.

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

If $\langle a, F(z)\rangle \geq 0$ 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

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

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

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

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

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

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

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

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

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

Thus,

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

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

Thus, when all the players update their strategies using the EG algorithm, by Theorem 6 and Lemma 9 we have that $r^{tan}_{\mathcal{G}}(z_T) = r^{tan}_{(F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}(z_T) \leq \frac{1}{\sqrt{T}} \frac{3D||z_0 - z^*||}{\eta\sqrt{1-(\eta L)^2}}$ and $r^{tan}_{\mathcal{G}}(z_{T+\frac{1}{2}}) = r^{tan}_{(F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}})}(z_{T+\frac{1}{2}}) \leq \frac{1}{\sqrt{T}} \frac{(1+\eta L)\cdot 3D||z_0 - z^*||}{\eta\sqrt{1-(\eta L)^2}}$. When all the players update their strategies using the OG algorithm, by Theorem 8 we have that $r^{tan}_{\mathcal{G}}(w_{T+1}) = r^{tan}_{F_{\mathcal{G}},\mathcal{Z}_{\mathcal{G}}}(w_{T+1}) \leq \frac{1}{\sqrt{T}} \cdot \frac{\sqrt{2}(2+\eta L)\cdot\sqrt{(4+6\eta^4L^4)\|z_0 - z^*\|^2 + (16\eta^2L^2 + 6\eta^4L^4)\|w_0 - z_0\|^2}}{\eta\cdot\sqrt{1-4\cdot(\eta L)^2}}$. 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 $k \geq 0$,*

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

*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 $\eta$ 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 $\eta$. 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 $\eta$, 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:

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

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, FP07]. 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, FP07])**.** *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:

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

The last inequality is because $\langle z_k - \eta F(z_k) - z_{k+\frac{1}{2}}, z_{k+\frac{1}{2}} - z_{k+1}\rangle \ge 0$, 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

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

So we can further simplify the inequality as follows:

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

Hence,

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

□

#### 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 $\{z_k, z_{k+\frac{1}{2}}\}_{k\ge 0}$ be the iterates of the EG algorithm with step size $\eta \in (0, \frac{1}{L})$. Then for all $k \ge 0$,*

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

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

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

By triangle inequality,

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

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 $\{z_i[\ell], \eta F(z_i)[\ell]\}_{i\in\{k,k+\frac{1}{2},k+1\},\ell\in[n]}$:

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

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

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

We always multiply $F$ with $\eta$ 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:

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

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:

$$\sum_{\ell\in[n]}\Big((\eta F(z_k)[\ell])^2-(\eta F(z_{k+1})[\ell])^2+\sum_{i>j\text{ and }i,j\in K}\lambda^*_{i,j}\Big(\big(\eta F(z_i)[\ell]-\eta F(z_j)[\ell]\big)^2-(\eta L)^2\big(z_i[\ell]-z_j[\ell]\big)^2\Big)$$
$$+\sum_{i>j\text{ and }i,j\in K}\mu^*_{i,j}\big(\eta F(z_i)[\ell]-\eta F(z_j)[\ell]\big)\big(z_j[\ell]-z_i[\ell]\big)+\gamma_1^{(\ell)}(\boldsymbol{w})(z_{k+\frac12}[\ell]-z_k[\ell]+\eta F(z_k)[\ell])$$
$$+\gamma_2^{(\ell)}(\boldsymbol{w})(z_{k+1}[\ell]-z_k[\ell]+\eta F(z_{k+\frac12})[\ell])\Big). \tag{10}$$

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 $\ell$-th summand and $\ell'$-th summand are identical subject to a change of variable;[13] (ii) the $\ell$-th summand only depends on the coordinate $\ell$, 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 $\ell$. $g^L_{i,j}(\boldsymbol{x},\boldsymbol{y})$ and $g^m_{i,j}(\boldsymbol{x},\boldsymbol{y})$ come from the $\ell$-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

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

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 $c(z) := \Pi_{N(z)}\left[ - F(z)\right]$ 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 $c(z) = \Pi_{N(z)}\left[ - F(z)\right]$, 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 $-a_{k+1} \in N(z_{k+1})$.

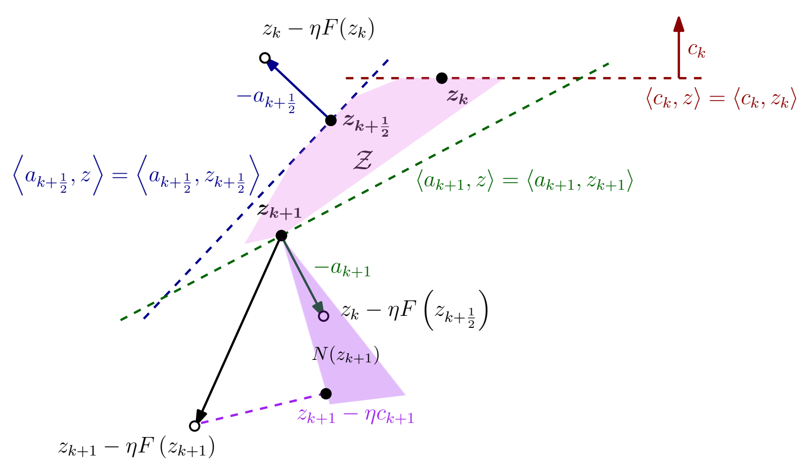


Figure 4: Reducing the number of constraints.

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 $\eta$ 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 $z_{k+\frac{1}{2}} = \Pi_{\mathcal{Z}}(z_k-\eta F(z_k))$ 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 $\ell$-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 $\ell$-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 $\ell$-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:**

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

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

Figure 5: Our SOS program in the constrained setting.

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

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 $T \geq 0$,*

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

**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 $k \geq 0$,*

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

*Moreover, for all $T \geq 0$,*

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

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

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

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

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

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). □

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

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

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 $\Phi(z_k, w_k)$

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)

$$\begin{aligned}
&= \left\|\frac{w_{k+1} - z_{k+1}}{2} + \eta F(w_k) - \eta F(z_k)\right\|^2 && (52)\\
&+ \left\|\eta F(z_k) + \eta c(z_k) - z_k + \frac{w_{k+1} + z_{k+1}}{2}\right\|^2 && (53)\\
&+ \|z_k - \eta F(w_{k+1}) - z_{k+1} - \eta c(z_{k+1})\|^2 && (54)\\
&\geq 0.
\end{aligned}$$

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

### 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 $D_0 := \sqrt{(4 + 6\eta^4 L^4)\|z_0 - z^*\|^2 + (16\eta^2 L^2 + 6\eta^4 L^4)\|w_0 - z_0\|^2}$. Then for all $T \geq 1$,*

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

# 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 $D_0 := \frac{\sqrt{(4+6\eta^4L^4)\|z_0 - z^*\|^2 + (16\eta^2L^2 + 6\eta^4L^4)\|w_0 - z_0\|^2}}{\sqrt{1-4(\eta L)^2}}$. 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 $\eta L \le 1$ and Inequality (55), we have

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

□

**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 $\{z_k, w_k\}_{k\ge 0}$ be the iterates of the OG algorithm with step size $\eta \in (0, \frac{1}{2L})$. Then for all $k \ge 1$,*

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

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

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

Thus, for all $k \ge 0$, 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 $\eta^2 L^2 \le \frac{1}{4}$ we have that for all $k \ge 0$,

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

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

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

□

# 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 $2\times 2$ 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

$$\begin{aligned} \left\| z_0 - z_{\frac{1}{2}} \right\|^2 =& 0.004527845815556556, \\ \left\| z_1 - z_{1+\frac{1}{2}} \right\|^2 =& 0.004552329544896258, \\ \left\| z_2 - z_{2+\frac{1}{2}} \right\|^2 =& 0.004552306444552208. \end{aligned}$$

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.

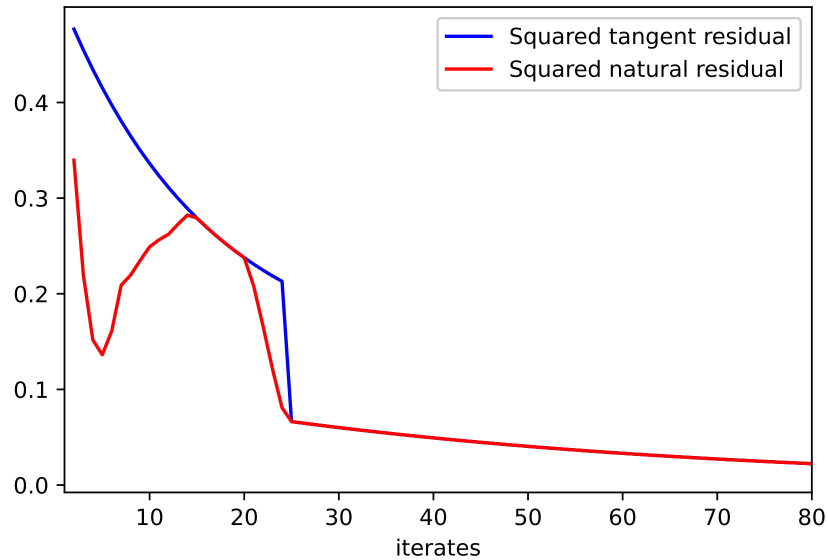


Figure 6: Non-monotonicity of the Natural Residual

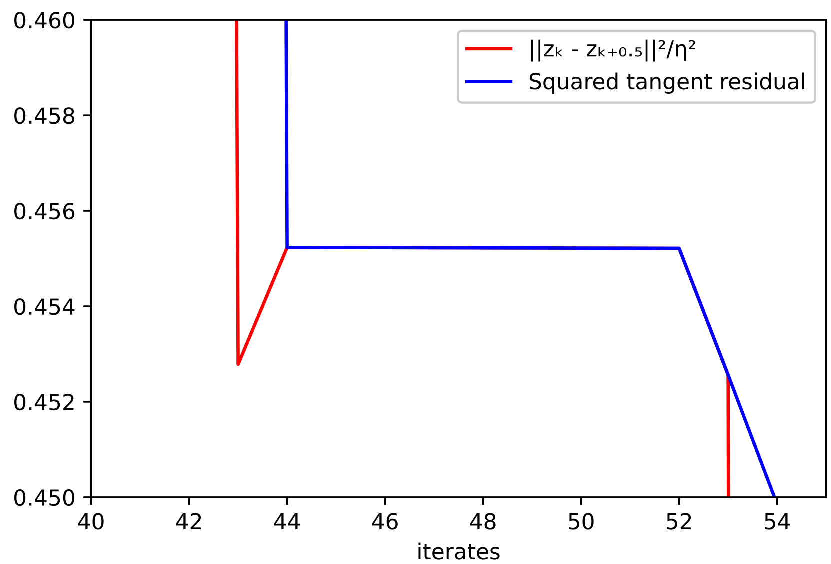


Figure 7: Non-monotonicity of $\|z_k - z_{k+1/2}\|^2$

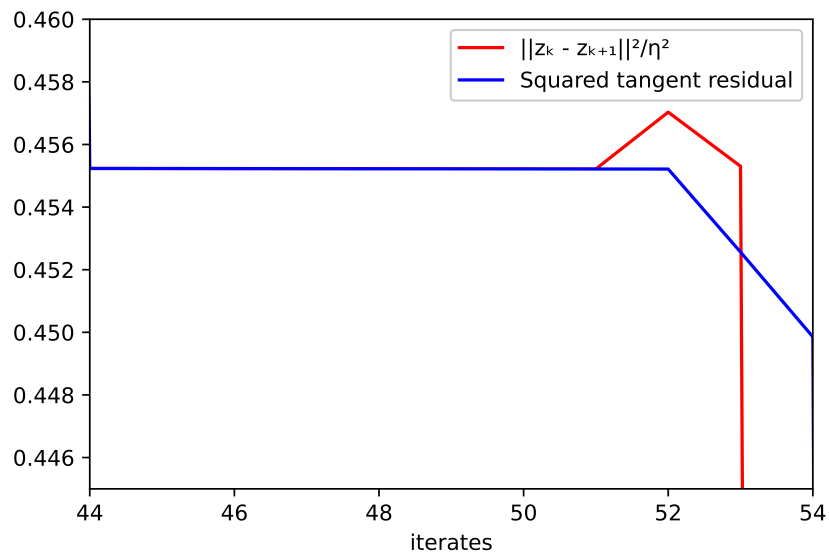


Figure 8: Non-monotonicity of $\|z_k - z_{k+1}\|^2$

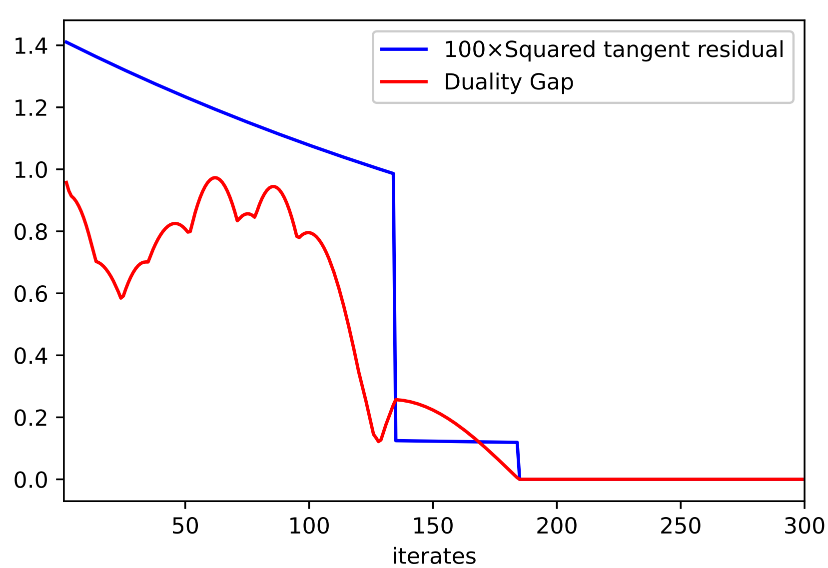


Figure 9: Non-monotonicity of Variants of Gap Functions. Here we have scaled the tangent residual $\times 100$ to make the plot clear.

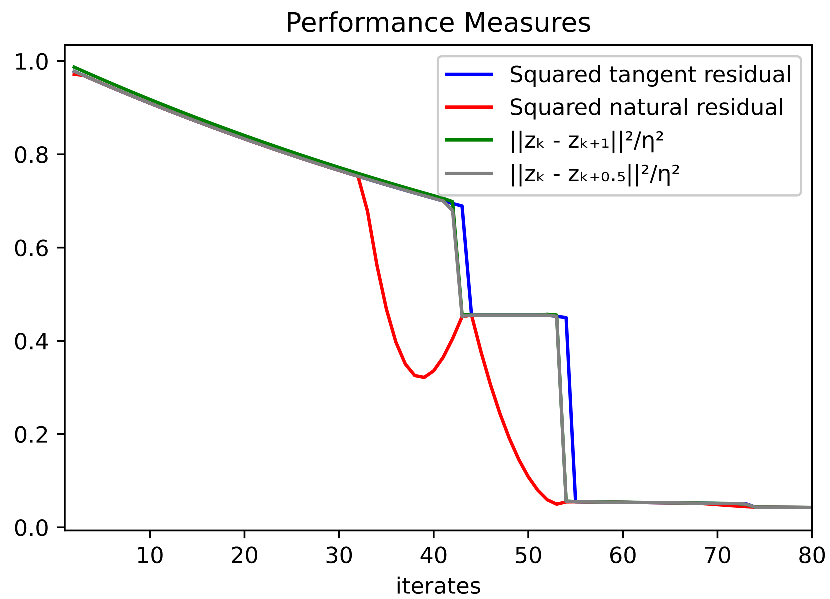


Figure 10: Numerical experiments on bilinear game (58) with $A = \begin{bmatrix} 0.50676631 & 0.15042569 \\ 0.46897595 & 0.96748026 \end{bmatrix}$, $b = c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, initial point $z_0 = (0.25, 0.25, 0.25, 0.25)^\top$ and step size $\eta = 0.1$. This is the same bilinear 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.