

SIMILARITY AND SEPARATION OF LAST-ITERATE CONVERGENCE BETWEEN OPTIMISM AND REFLECTED ALGORITHMS IN TIME-VARYING GAMES

Anonymous authors

Paper under double-blind review

ABSTRACT

In this paper, we investigate the behaviours of reflected gradient (RG), accelerated reflected gradient (ARG), and optimistic gradient (OG) algorithms in multi-player games modelled as variational inequalities with L -smooth continuous monotone limits in convergent time-varying cases and with L -smooth continuous and monotone games at each time in periodic cases, both in convex action sets. The RG, ARG, and OG algorithms require fewer complex calculations, i.e., on the gradients and projections per iteration. We prove that a convergence rate of $O(1/\sqrt{T})$ and $O(1/T)$ can be reached by the RG and ARG algorithms with bounded action sets for convergent perturbed monotone games, respectively, if the sequence of time-varying games converges to the limit fast enough, without additional assumptions like strong monotonicity, and such a result matches and improves the existing results on similar algorithms requiring calculations on two gradients with different actions. Besides, a surprising result is also shown that the standard OG algorithm in time-varying games behaves dramatically differently from its variant and other similar algorithms: the standard OG algorithm converges in any sequences of time-varying monotone L -smooth games with a common Nash equilibrium, including some periodic games, while at the same time, its variant with a slight difference diverges exponentially even in periodic games. We also show that the RG and ARG algorithms diverge exponentially in some periodic games.

1 INTRODUCTION

This paper discusses machine learning algorithms applied to a type of repeated multi-player game. In this type of game, each player selects an action from a continuous closed convex action set, and pays a certain cost depending on all players' actions. The relationship between these actions and costs is determined by any one game of this type evolving over time and is unknown to any player a priori. As the game evolves, each player changes their action according to their cost and other known information, followed by the game repeating.

We seek to address the following open questions: Do improved versions of the extra gradient (EG) algorithm with only one calculation on the gradient or the projection, such as the reflected gradient (RG) algorithm, the accelerated reflected gradient (ARG) algorithm and the optimistic gradient (OG) algorithm, converge in time-varying monotone games with constant step sizes? If so, how fast do these algorithms converge?

1.1 BACKGROUND

Machine learning and game theory problems have intrinsic connections (Zhu et al. (2023); Lv et al. (2023)). Monotone games, a class of multi-player games introduced by (Rosen (1965)), include extensively discussed games such as two-player zero-sum games, convex-concave games (both including two-player bilinear games (Feng et al. (2023; 2024)), λ -cocoercive games (Lin et al. (2020)), zero-sum polymatrix games (Daskalakis & Papadimitriou (2009); Cai & Daskalakis (2011)), and zero-sum socially-concave games (Even-dar et al. (2009)).

054 The problem of finding a Nash equilibrium in a smooth monotone game is equivalent to solving a
055 Lipschitz and monotone variational inequality (MVI). Many gradient-based algorithms have been
056 proposed to solve game problems modelled as monotone variational inequality problems. A well-
057 known result on the optimal convergence rate of such algorithms in general MVI problems is $O(1/T)$
058 achieved by the EG algorithm (Korpelevich (1976)) and its Bregman variant, the Mirror-Prox (MP)
059 algorithm (Nemirovski (2004)).

060 A major disadvantage of the EG algorithm is the requirement of two calculations on projections and
061 gradients per iteration, which means it requires more computation than the gradient descent (GD)
062 algorithm due to the significant computational cost on gradients and on projections in constrained
063 games. Hence, its improvements, avoiding calculating the gradient and projection a second time
064 like in the EG algorithm while retaining the algorithm’s “predictive” properties, have attracted much
065 interest from the machine learning community. There are two complementary approaches to reduce
066 the costly calculations on reducing the number of calculations of projections and gradients per
067 iteration. The OG algorithm (Tseng (2000)) applies the former approach, and the PEG algorithm
068 (Popov (1980)) focusing on gradient extrapolation mechanisms applies the latter. The RG (Malitsky
069 (2015)) and the newly proposed ARG (Cai & Zheng (2023)) algorithms combine both.

070 Unfortunately, in time-varying game problems, it is still inevitable to calculate each of the gradient
071 and the projection twice per iteration with the past extra gradient (PEG) algorithm (also known as
072 the optimistic gradient descent ascent (OGDA) algorithm in some papers Feng et al. (2023)), which
073 increases the computational cost and makes the PEG algorithm as inefficient as the EG one. In
074 contrast, both the RG algorithm and the ARG algorithm developed from it require calculating each of
075 the gradient and the projection only once per iteration, and the OG algorithm requires calculating
076 gradients twice, but the projection only once, per iteration. Inspired by the advantages of the RG,
077 ARG, and OG algorithms, we investigate their convergence behaviors to analyze whether the three
078 algorithms perform satisfactorily for time-varying games, a complex problem.

079 1.2 RELATED WORKS

081 1.2.1 PREVIOUS RESEARCH ON THE EXTRA GRADIENT ALGORITHM AND ITS VARIANTS

082 Most known machine learning algorithms applied to game problems are based on gradients. The
083 simplest one of them is the GD algorithm. However, the GD algorithm diverges even for time-
084 invariant two-player zero-sum bilinear games (Daskalakis et al. (2018)). To eliminate this limitation,
085 many improved learning algorithms have been proposed, including the EG algorithm (Korpelevich
086 (1976); Cai et al. (2022); Feng et al. (2023); Monteiro & Svaiter (2010)) and its variants like the
087 PEG (Popov (1980); Cai et al. (2022); Feng et al. (2023)), OG (Daskalakis et al. (2018)), and RG
088 (Chambolle & Pock (2011); d’Angelo et al. (2014); Hsieh et al. (2019)) algorithms. Other algorithms
089 like the negative momentum algorithm (Feng et al. (2023)) and weights update algorithms (Arora
090 et al. (2012); Cai et al. (2024a)) are also applied by some researchers.

091 For the two classical algorithms, the EG (Korpelevich (1976)) and PEG (Popov (1980)) algorithms,
092 their convergence rates in time-invariant games have been thoroughly investigated. The convergence
093 property of the EG algorithm has been investigated in both concave games (Monteiro & Svaiter (2010)
094 on last-iterate convergence and Nemirovski (2004)) and special non-concave games (Mertikopoulos
095 et al. (2019)). The EG, PEG and other related algorithms in unconstrained strongly monotone games
096 and unconstrained bilinear games have also been proven to have linear convergence rates (Daskalakis
097 et al. (2018)). Later papers have proved their asymptotic convergence (Daskalakis & Panageas (2019))
098 and last-iterate convergence for games with monotone gradients of the cost functions with the rate
099 $O(1/\sqrt{T})$ (Golowich et al. (2020); Cai et al. (2022)). A recent result (Wei et al. (2021)) shows that
100 there exists even always an exponential convergence rate for the PEG algorithm in time-invariant
101 games. However, a surprising result, the separation of last-iterate convergence behaviors between the
102 EG and PEG algorithms in time-varying games (Feng et al. (2023)), has also been shown recently.

103 Compared with the EG and PEG algorithms, investigations of the RG (Malitsky (2015)), OG (Tseng
104 (2000)) and ARG (Cai & Zheng (2023)) algorithms proposed more recently are less extensive. (Yang
105 & Liu (2018)) shows weak convergence of the RG algorithm with general closed convex action sets
106 in monotone L -smooth games and its linear convergence with general closed convex action sets in
107 strongly monotone L -smooth games. The RG algorithm has also been applied in some works and
its performances have been shown to be satisfactory (Hsieh et al. (2019)). Cai & Zheng (2023) has

108 showed convergence rate of the RG algorithm to be $O(1/\sqrt{T})$ in time-invariant monotone games.
 109 The OG algorithm has been shown to strongly converge (Yang & Liu (2019)) in time-invariant
 110 monotone games, and Cai & Zheng (2023) showed that its best-iterate convergence rate is $O(1/\sqrt{T})$
 111 with the same conditions. Dung et al. (2024) has proved the weak convergence of the OG algorithm
 112 with conditions weaker than monotone games. The ARG algorithm has been proposed very recently
 113 (Cai & Zheng (2023)) and its last-iterate convergence rate in time-invariant monotone games is
 114 proven to be $O(1/T)$ together. So far, investigations of that algorithm are rare. However, the thought
 115 applied to this algorithm, the Halpern iteration, is classical (Halpern (1967)). Such a technique is
 116 closely related to Nesterov’s accelerated method (Tran-Dinh (2022)) and related methods have been
 117 extensively researched. Other literature on similar algorithms has been summarized in (Cai et al.
 118 (2024b)).

1.2.2 PREVIOUS RESEARCH ON TIME-VARYING GAMES

121 In recent years, works on time-varying games have been emerging. Most of them focused on time-
 122 average convergence behaviors (Cardoso et al. (2019); Zhang et al. (2022); Yan et al. (2023)). To
 123 divert from the difficulties in calculating the Nash equilibrium (Daskalakis et al. (2009); Chen et al.
 124 (2009); Deligkas et al. (2022)), some researchers turned to a weakened concept, correlated equilibrium
 125 (Anagnostides et al. (2023)), with limited usage within the machine learning community. An important
 126 recent article by Duvocelle et al (Duvocelle et al. (2023)) has discussed strictly monotone games with
 127 decreasing step sizes and has shown that the probability of convergence is 1, despite not providing
 128 a convergence rate, a key measure of a learning algorithm. Decreasing step sizes is also unnatural
 129 since such a requirement considers new information decreasingly important instead of equally or
 130 increasingly as expected (Lin et al. (2020)). Hadji et al. (2024) investigated strongly monotone games
 131 with contractive algorithms thoroughly, showing that the total errors though time (called “tracking
 132 errors”) of contractive algorithms are sublinear if their optimal solution has sublinear quadratic path
 133 lengths and are logarithmic if they are periodic games with unique solutions with time-invariant step
 134 sizes and the online gradient descent algorithm.

135 However, the above results are based on the strong or strict monotonicity of variational inequalities,
 136 while simple games such as two-player zero-sum bilinear games are not even strictly monotone, with
 137 $\langle F(z_1) - F(z_2), z_1 - z_2 \rangle = 0$. For problems involving non-strictly monotone games, Feng et al.
 138 (2023) discussed unconstrained two-player zero-sum bilinear games without requiring decreasing
 139 step sizes and obtained the first known result on the last-iterate convergence in time-varying games.
 140 More recently, Feng et al. (2024) discussed constrained two-player zero-sum bilinear periodic games,
 141 and Chen & Yu (2025) has provided an important result that the EG and PEG algorithms are robust
 142 to convergent perturbation under the BAP assumption (defined in Assumption 1, Section 2.2, which
 143 means the time-varying game converges fast enough) for general convergent perturbed monotone
 144 games. Other known results involving games which are not necessarily strictly monotone include
 145 Zhang et al. (2022) focused on regret bounds (two-player) in time-varying bilinear saddle-point
 146 problems parameterized by the similarity of the payoff matrices and the equilibria of these games,
 147 Cardoso et al. (2019) providing an optimal solution based on the Nash equilibrium regret, and
 148 Duvocelle et al. (2023) discussing a general algorithm based on mirror descent for non-converging
 149 games and proved the convergence rate $O(T^{-(1-r)/3})$ where $r < 1$.

149 Besides, for the algorithms requiring less computation per iteration like the OG, RG and ARG
 150 algorithms, no known literature exists on whether they converge in the average sense or exhibit
 151 last-iterate convergence in time-varying games, leaving a big gap.

1.3 OUR CONTRIBUTIONS

154 Even though time-varying games are more realistic than time-invariant ones, existing results on
 155 convergence of machine learning algorithms in time-varying non-strong monotone games are limited.
 156 To fill such gaps, we investigate the last-iterate behaviors of the RG, ARG and OG algorithms in
 157 time-varying games. Our main results on their behaviors are summarized as follows.

- 158 • With methods inspired by (Yang & Liu (2018); Cai et al. (2022); Cai & Zheng (2023);
 159 Feng et al. (2023)), we prove that a time-invariant step size η is sufficient for last-iterate
 160 convergence at the rates of $O(1/\sqrt{T})$ for the RG algorithm and $O(1/T)$ for the ARG
 161 algorithm in fast-converging perturbed time-varying monotone games under well-established

measures of performance, like tangent residuals, which expands the results in (Cai & Zheng (2023)). Note that the convergence rate $O(1/T)$ for the ARG algorithm matches the optimal convergence rate for all first-order methods for monotone inclusion problems (Diakonikolas (2020); Yoon & Ryu (2021); Cai & Zheng (2023)), which include the problems in this paper.

- We show that both the RG and ARG algorithms may diverge in a periodic monotone game even with a common Nash equilibrium.
- With methods inspired by (Cai & Zheng (2023); Feng et al. (2023)), we show that the OG algorithm and one of its variants behave surprisingly differently. Weak convergence of the standard OG algorithm is robust to time-varying games being monotone at each time as long as they share the same Nash equilibrium and they are all L -smooth. This is different from another optimistic algorithm called the OGDA (i.e. PEG) algorithm, which may diverge in periodic games (Feng et al. (2023)). However, the variant of the OG (VOG) algorithm with the first step of each iteration in the OG algorithm applying the gradient function used in the second step of the previous iteration may diverge even if the game is periodic.

A comparison of the RG, ARG, OG and VOG algorithms with the PEG and EG algorithms is shown in Table 1, where the results on convergent perturbed games are for games with bounded action sets for PEG, EG, RG and ARG algorithms and games with common Nash equilibria for PEG, EG, RG and OG algorithms, the results on periodic games are for games with common Nash equilibria for OG algorithms, the convergence rates of PEG, EG, RG and ARG algorithms are last-iterate convergence rates and the convergence rates of OG and VOG algorithms are best-iterate convergence rates.

Remark 1. *We discuss the periodic and convergent games in our paper for the following reasons. Both kinds of games have been discussed in previous literature as testing grounds for learning algorithms in the machine learning community (Duvocelle et al. (2023); Fiez et al. (2021); Feng et al. (2023; 2024); Chen & Yu (2025)), and last-iterate behaviours are well-defined in those kinds of time-varying games. Besides, our paper focuses on convergence behaviors of algorithms with less computational cost than those discussed in existing research, making it their natural extension. Moreover, the periodic game and the convergent perturbed game are natural generalizations of the time-invariant games in theoretical problems and in reality: a periodic game with the period of 1 is a time-invariant game, and a convergent perturbed game with the perturbation of 0 is also a time-invariant game. In theoretical problems, for the last iterate behaviours, the generalization of convergent perturbed games has been demonstrated in (Feng et al. (2023); Chen & Yu (2025)), and the generalization of periodic games has been demonstrated in (Feng et al. (2023)). In reality, the convergent perturbed game is a good model for the game affected by feedback progress, while for the periodic games, since seasonal changes or daily changes affect many ecological or market competitions in reality, such examples can be modelled as periodic games, and frameworks of multi-agent contextual games in (Sessa et al. (2020)) can also be modelled as periodic games.*

Our remaining paper is organized as follows. Section 2 defines the game problems, describes the main definitions and lemmas involved in these problems and introduces the model of machine learning in our game problems. Sections 3 and 4 show last-iterate convergence of the RG and ARG algorithms defined with tangent residuals in both periodic and BAP cases. Section 5 show best-iterate convergence of the standard OG algorithm and its variant. Section 6 concludes the paper with proposals on future directions.

2 PRELIMINARIES

In this section, we introduce the machine learning dynamics of time-varying games, which involves the RG, ARG and OG algorithms.

2.1 NOTATIONS

In this paper, the Euclidean space $(\mathbb{R}^n, \|\cdot\|)$ is considered, where $\|\cdot\|$ is the 2-norm. $\langle \cdot, \cdot \rangle$ denotes a inner product on \mathbb{R}^n . For simplicity, $\{1, 2, \dots, N\}$ is denoted as $[[N]]$. The time-varying game is denoted as $\mathcal{G}_t := \{\mathcal{N}, \mathcal{Z}, f_t(z)\}$ where $\mathcal{N} := [[N]]$ denotes the set of players, \mathcal{Z} is the set of actions and also a closed convex set, and $f_t(z)$ is the vector of cost functions of the players at time t .

Table 1: Comparison of convergence rates of several algorithms in time-invariant and two types of time-varying games. Alg stands for Algorithm.

ALG	TIME-INVARIANT	FAST CONVERGENT PERTURBED	PERIODIC
PEG	$O(1/\sqrt{T})$ (CAI ET AL. (2022))	$O(1/\sqrt{T})$ (CHEN & YU (2025))	NONE (FENG ET AL. (2023))
EG	$O(1/\sqrt{T})$ (CAI ET AL. (2022))	$O(1/\sqrt{T})$ (CHEN & YU (2025))	OPEN
RG	$O(1/\sqrt{T})$ (CAI & ZHENG (2023))	$O(1/\sqrt{T})$ (THEOREM 2)	NONE (THEOREM 1)
ARG	$O(1/T)$ (CAI & ZHENG (2023))	$O(1/T)$ (THEOREM 4)	NONE (THEOREM 3)
OG	$O(1/\sqrt{T})$ (CAI & ZHENG (2023))	$O(1/\sqrt{T})$ (THEOREM 6)	$O(1/\sqrt{T})$ (THEOREM 6)
VOG	$O(1/\sqrt{T})$ (SAME AS OG)	OPEN	NONE (THEOREM 5)

2.2 MODEL OF TIME-VARYING GAMES

In this part, we establish the time-varying game model, in which each player faces different cost functions at different times. With such a model, we can deal with the exogenous disturbance and the endogenous varying property. To begin with, we construct the framework of cost functions. For regularity, our paper only discusses continuously differentiable cost functions. For any time-varying game involved, the cost function of such a game at time t is denoted as $f_t(z)$, and $F_t^{(i)}(z)$ denotes its gradient for Player i , i.e., $F_t^{(i)}(z) = \nabla f_t^{(i)}(z)$. By putting all players together, we define $F_t(z) = (F_t^{(1)}(z), \dots, F_t^{(N)}(z))$. Specifically, we study the following two kinds of time-varying games: periodic games and convergent perturbed games.

Our time-varying games satisfy the following conditions:

1. At each time $t = 1, 2, \dots$, each player selects their action $z^{(i)}$, $i \in \mathcal{N}$.
2. Each player pays a cost associated with \mathcal{G}_t and selects their strategy based on the gradient of the set of actions of all the players, i.e., with the RG algorithm introduced in this paper.
3. All players implement their strategies and the sequence of games continues.

The essential elements of the machine learning model for the time-varying games include:

1. the game \mathcal{G}_t played by all players at each time;
2. the gradients of time-varying cost functions of the players, $F_t^{(1)}(z), \dots, F_t^{(N)}(z)$, at each time;
3. the algorithm for players to select their actions based on the gradients at each time.

The following parts present the properties of the elements above in detail.

Definition 1 (Periodic games). *A periodic game with period \mathcal{T} is an infinite sequence of games $\{f_t\}_{t=0}^{\infty} \subset \mathbb{R}^n$, and $f_{t+\mathcal{T}} = f_t$ for all $t \geq 0$.*

Definition 2 (Convergent perturbed games). *A convergent perturbed game is an infinite sequence of games where $\lim_{t \rightarrow \infty} f_t$ exists. For simplicity, $\lim_{t \rightarrow \infty} f_t$ is denoted as f_{∞} and $g_t := f_t - f_{\infty}$. $\lim_{t \rightarrow \infty} g_t = 0$ by definition.*

Throughout the paper, sequences of time-varying smooth and monotone games with their limits being C^2 smooth cost functions are applied in our model. Besides, cost functions in time-invariant cases and their limits in time-varying cases are required to have monotone gradients. Following is the precise requirement of smoothness and monotonicity.

Definition 3. ((Rosen (1965); Cai et al. (2022); Chen & Yu (2025))) *A game at a certain time is C^2 smooth and monotone if its cost function is C^2 smooth and monotone at the time, i.e., $F(z)$ is L -Lipschitz and monotone, which means that both $\|F(z) - F(z')\| \leq L\|z - z'\|$ and $\langle F(z) - F(z'), z - z' \rangle \geq 0$, where $L \geq 0$, holds.*

A convergent perturbed game is C^2 smooth and monotone if the sequence of cost functions converges to a monotone function with Lipschitz gradient, i.e., $F_{\infty}(z) := \lim_{t \rightarrow \infty} F_t(z)$ is monotone and

270 L -Lipschitz, and a periodic game is C^2 smooth and monotone if the cost function $F_t(z)$ at each time
 271 t is a monotone function with Lipschitz gradient.

272
 273 In the convergent perturbed games, another important assumption is that the difference between
 274 the time-varying cost function and its limit $f_t - f_\infty =: g_t$ satisfies the well-known assumption of
 275 bounded accumulated perturbations (BAP) (Benzaid & Lutz (1987); Saber Elaydi & Kamiyama
 276 (1999); Elaydi & Györi (1995)). Following are our settings for the BAP assumption.

277 **Assumption 1** (BAP assumption). Denote $G_t := \nabla g_t$. \mathcal{Z} is bounded and $\sum_{t=0}^{\infty} \max \|G_t\| < \infty$.

279 2.3 NASH EQUILIBRIA IN TIME-VARYING GAMES

280
 281 The most prevalent concept of solutions in game theory is the Nash equilibrium. The general
 282 definition of Nash equilibrium of a game \mathcal{G} is an action $z^* \in \mathcal{Z}$ such that for each player i , it holds
 283 that $f^{(i)}(z^*) \leq f^{(i)}(z^{(i)}, z^{*(-i)})$ for any $z^{(i)} \in \mathcal{Z}^{(i)}$, where $\mathcal{Z}^{(i)}$ means the action set of Player i . By
 284 the general definition of Nash equilibrium, we obtain the following sufficient and necessary condition
 285 for an action to be a Nash equilibrium in our problems.

286 **Lemma 1.** (Facchinei & Pang (2007)) For a time-invariant game \mathcal{G} with monotone $F(z)$, an action
 287 z^* is a Nash equilibrium if and only if $\langle F(z^*), z^* - z \rangle \leq 0, \forall z \in \mathcal{Z}$.

288 Lemma 1 above means that solving the Nash equilibrium of a game in our paper requires solving
 289 the variational inequality in this lemma. Based on Lemma 1, for periodic games with common Nash
 290 equilibria and convergent perturbed games, we define the Nash equilibrium as follows.

291 **Definition 4.** For a periodic game \mathcal{G} with monotone $F_t(z)$ and an action z^* satisfying $\langle F_t(z^*), z^* - z \rangle \leq 0, \forall z \in \mathcal{Z}$, its Nash equilibrium is defined as such z^* .

292
 293 **Definition 5.** For a convergent perturbed game \mathcal{G} with monotone $F_\infty(z)$, its Nash equilibrium is
 294 defined as an action z^* satisfying $\langle F_\infty(z^*), z^* - z \rangle \leq 0, \forall z \in \mathcal{Z}$.

296 2.4 TANGENT RESIDUAL

297
 298 We apply the tangent residual proposed in (Cai et al. (2022)) as the measure of error in our paper. For
 299 time-varying games, we modify the definition of the tangent residual as follows.

300 **Definition 6.** Denote the tangent residual of $f(z)$ involved in the game \mathcal{G} as $r_{F, \mathcal{Z}}^{\tan}$ and denote $N_{\mathcal{Z}}(z)$
 301 as the normal cone of z for the set \mathcal{Z} (for simplicity \mathcal{Z} may be omitted). The tangent residual of
 302 a periodic game with a common Nash equilibrium is defined as $r_{\mathcal{G}_i, \mathcal{Z}}^{\tan} = \min_{c \in N(z)} \|F_i(z) + c\|$,
 303 where $i \in [[T]]$ is the iteration number and T is the period in Definition 1, while the tangent residual
 304 of a convergent perturbed game is defined as $r_{\mathcal{G}, \mathcal{Z}}^{\tan} = \min_{c \in N(z)} \|F_\infty(z) + c\|$.

306 2.5 MACHINE LEARNING ALGORITHMS IN GAMES

307
 308 We introduce the following algorithms for our game problems: reflected gradient (RG) algorithm,
 309 accelerated reflected gradient (ARG) algorithm and optimistic gradient (OG) algorithm. Their pseudo-
 310 code implementations are shown in Algorithms 1, 2 and 3. All those three algorithms are variants
 311 of inexact proximal point algorithms (He & Yuan (2012); Zhang et al. (2025)), like EG and PEG
 312 discussed in existing results.

313 **Remark 2.** In all three following learning algorithms, we do not limit the relationship between two
 314 initial points (Hsieh et al. (2019); Feng et al. (2023)) to avoid reducing generality of our results.

316 **Algorithm 1** Reflected gradient algorithm (Malitsky (2015))

317 **Input:** Step size: $\eta > 0$; gradient of cost function: $F_t(x)$;

318 1: initialize z_0, z_{-1}
 319 2: **for** $t = 0, 1, 2, 3, \dots$ **do**
 320 3: play $z_t \in \mathcal{Z}$
 321 4: set $z_{t+1/2} = 2z_t - z_{t-1}$
 322 5: set $z_{t+1} = \Pi_{\mathcal{Z}}(z_t - \eta F_t(z_{t+1/2}))$
 323 6: **end for**

Algorithm 2 Accelerated reflected gradient algorithm (Cai & Zheng (2023))

Input: Step size: $\eta > 0$; gradient of cost function: $F_t(x)$;

- 1: initialize $z_0, z_{1/2}$
- 2: **for** $t = 0, 1, 2, 3, \dots$ **do**
- 3: play $z_t \in \mathcal{Z}$
- 4: set $z_{t+1/2} = 2z_t - z_{t-1} + \frac{1}{t+1}(z_0 - z_t) - \frac{1}{t}(z_0 - z_{t-1})$ if $t \neq 0$
- 5: set $z_{t+1} = \Pi_{\mathcal{Z}}(z_t - \eta F_t(z_{t+1/2}) + \frac{1}{t+1}(z_0 - z_t))$
- 6: **end for**

Algorithm 3 Optimistic gradient algorithm (Hsieh et al. (2019); Mokhtari et al. (2020a;b); Daskalakis et al. (2018))

Input: Step size: $\eta > 0$; gradient of cost function: $F_t(x)$;

- 1: initialize $z_0, z_{-1/2}$
- 2: **for** $t = 0, 1, 2, 3, \dots$ **do**
- 3: play $z \in \mathcal{Z}$
- 4: set $z_{t+1/2} = \Pi_{\mathcal{Z}}(z_t - \eta F_t(z_{t-1/2}))$
- 5: set $z_{t+1} = z_{t+1/2} + \eta F_t(z_{t-1/2}) - \eta F_t(z_{t+1/2})$
- 6: **end for**

3 CONVERGENCE RESULTS OF THE RG ALGORITHM

In this section, we prove that the RG algorithm does not necessarily converge for a periodic game even if that game is a two-player zero-sum game, though it reaches a convergence result of $O(1/\sqrt{T})$ with a fast converging perturbation. To illustrate Theorem 1 and Theorem 2 in this section, numerical examples based on examples of bilinear games in (Feng et al. (2023)) are shown in Section I.

3.1 PERIODIC CASE

In this section, we show a negative result: the RG algorithm does not necessarily converge in monotone periodic games, even if the step size is small. To illustrate this negative result, we provide a simple counterexample with a bilinear game inspired from (Feng et al. (2023)).

Theorem 1. *In the following two-player game \mathcal{G}_t , no $\eta > 0$ guarantees that the RG algorithm converges, since the RG algorithm diverges at an exponential rate in this example.*

$$z^{(1)} = x \in \mathbb{R}, z^{(2)} = y \in \mathbb{R}^2, f^{(1)} = x^T A_t y, f^{(2)} = -x^T A_t y, \mathcal{Z} = \mathbb{R} \times \mathbb{R}^2$$

$$A_t = \begin{cases} \begin{bmatrix} 1 & -1 \end{bmatrix}, & t \text{ is odd} \\ \begin{bmatrix} -1 & 1 \end{bmatrix}, & t \text{ is even} \end{cases} \quad (1)$$

Proof sketch. For the two-player bilinear game with the period of 2 above, we prove that \mathcal{A}_t satisfying $[x_t \ y_t \ x_{t+1} \ y_{t+1}]^T = \mathcal{A}_t [x_{t-1} \ y_{t-1} \ x_t \ y_t]^T$ has eigenvalues greater than 1 so that $\|z_t\|$ increases exponentially with $\eta > 0$ in certain cases. Then, with $r^{tan}(z_t) = \|z_t\|$ in this game, we conclude that $r^{tan}(z_t)$ diverges with the RG algorithm in such cases, which results in the impossibility of the RG algorithm converging to the Nash equilibrium in such cases.

The full proof of Theorem 1 is deferred to Section C.

3.2 BAP CASE

In this part, we discuss the case of time-varying games for the RG algorithm under BAP assumption, i.e., Assumption 1. We show that L_{G_t} -smooth perturbation makes convergent perturbed games with a common Nash equilibrium satisfy Assumption 1 naturally.

Lemma 2. *If there exists a common Nash equilibrium z^* for each time of a sequence of convergent perturbed monotone games and $G_t(z)$ is L_{G_t} -Lipschitz with $\sum_{t=0}^{\infty} L_{G_t} < \infty$ and $G_t(z^*) = 0$, then there exists a bounded set \mathcal{Z}_L satisfying $z_t \in \mathcal{Z}_L$. Consequently, Assumption 1 is satisfied.*

In the following theorem, we show that the RG algorithm is robust to perturbation satisfying Assumption 1.

Theorem 2. *For the convergent perturbed game with a bounded action set and its limit having a L -Lipschitz gradient, if $\eta \in (0, \frac{1}{(\sqrt{2}+1)L})$, then the RG algorithm converges and the convergence rate of the tangent residual is $\max \left\{ O(1/\sqrt{T}), O \left(\sqrt{\sum_{t=T/2}^{\infty} \max \|G_t\|} \right) \right\}$ under Assumption 1.*

Proof sketch. We construct a potential function and show that it increases slowly or does not increase between two consecutive iterates. We prove that the best-iterate convergence rate of the RG algorithm is small, i.e., $\forall T \geq 1$, there exists one iterate $t^* \in [[T]]$ such that our potential function at t^* is small. We combine the above steps to show that the last iterate has the same convergence guarantee as the best iterate and show that the last-iterate convergence rate is $O(1/\sqrt{T})$ under Assumption 1.

The full proof of Theorem 2 is deferred to Section D and the proof of Lemma 2 is included there in Section D.3.

4 CONVERGENCE RESULTS OF THE ARG ALGORITHM

In this section, we prove the following results: the ARG algorithm does not necessarily converge for a periodic game even if that game is a two-player zero-sum game. For completeness, we also show the ARG algorithm reaches a convergence result of $O(1/T)$ with a fast converging perturbation and still converges with a perturbation converging slower. To illustrate Theorem 3 and Theorem 4 in this section, numerical examples based on bilinear games in (Feng et al. (2023)) are shown in Section I.

4.1 PERIODIC CASE

In this section, we show a negative result: the ARG algorithm does not necessarily converge in monotone periodic games, even if the step size is small. To illustrate this negative result, we provide a simple counterexample with a bilinear game inspired from (Feng et al. (2023)).

Theorem 3. *In the two-player game (1) shown in Theorem 1, no $\eta > 0$ guarantees the ARG algorithm converges, since the ARG algorithm diverges at an exponential rate in the example provided.*

Proof sketch. For the two-player bilinear game with the period of 2 above, we prove that for a t large enough, \mathcal{A}_t satisfying $[x_0 \ y_0 \ x_t \ y_t \ x_{t+1} \ y_{t+1}]^T = \mathcal{A}_t [x_0 \ y_0 \ x_{t-1} \ y_{t-1} \ x_t \ y_t]^T$ converges to a constant matrix with eigenvalues greater than 1, making $\|z_{t+T}\|$ is approximately equal to $C^T \|z_t\|$ with $\eta > 0$ and $C > 1$. Then, with $r^{tan}(z_t) = \|z_t\|$ in this game, we show that $r^{tan}(z_t)$ diverges with the ARG algorithm.

The full proof of Theorem 3 is deferred to Section E.

4.2 BAP CASE

In this part, we show the ARG algorithm is robust to perturbation with a fast convergence rate.

Theorem 4. *For the convergent perturbed game with a bounded action set and its limit having a L -Lipschitz gradient, if $\eta \in (0, \frac{1}{\sqrt{24}L})$ and $\sum_{t=0}^{\infty} t^2 \|G_t\| < \infty$, the ARG algorithm converges at the rate $O(1/T)$. Under Assumption 1, the tangent residual with the ARG algorithm is $\max \left\{ O(1/\sqrt{T}), O \left(\sqrt{\sum_{t=1}^T \max \|G_t\|} \right) \right\}$. This result can be extended to $\rho \in [-\frac{1}{60}, 0]$ and $\langle F_t(z) + N(z) - F_t(z') - N(z'), z - z' \rangle \geq \rho \|F_t(z) + N(z) - F_t(z') - N(z')\|^2, \forall z, z' \in \mathcal{Z}$ if $\frac{1}{2} - (12 - \frac{4\rho}{\eta})\eta^2 L^2 + \frac{2\rho}{\eta} \geq 0$.*

Proof sketch. We only need to prove the extended result. We first construct a potential function and show that it is approximately non-increasing. Then, we prove that it is upper bounded by a term independent of time and prove that the such function at step t is $O(t^2 r^{tan}(z_t)^2)$ if $\sum_{t=0}^{\infty} t(t+1) \|G_t\| < \infty$ and is $O(t r^{tan}(z_t)^2)$ under Assumption 1. Then, we conclude that the convergence rate of the ARG algorithm is $O(1/T)$ if $\sum_{t=0}^{\infty} t(t+1) \|G_t\| < \infty$ and is $O(1/\sqrt{T})$ under Assumption 1.

The full proof of Theorem 4 is deferred to Section F.

5 CONVERGENCE RESULTS OF THE OG ALGORITHM

In this section, we show that the OG algorithm converges for any L -smooth game, including any periodic game and convergent perturbed game, with a common Nash equilibrium, while surprisingly, a variant of the OG algorithm may diverge exponentially even if that game is a periodic two-player zero-sum game with a common Nash equilibrium. To illustrate Theorem 5 and Theorem 6 in this section, numerical examples based on bilinear games in (Feng et al. (2023)) are shown in Section I.

5.1 PERIODIC CASE OF A VARIANT OF THE OG ALGORITHM

In this part, we show a surprising fact that it is not necessarily possible for a natural variant of the OG algorithm with still only one calculation on the gradient and the projection to converge with periodic monotone games, since it is almost equivalent to the RG algorithm in two-player bilinear games.

Theorem 5. *If F_t in Line 5 of the OG algorithm (Algorithm 3) is modified to F_{t+1} , then the modified OG algorithm for $z_{t+1/2}$ is equivalent to the RG algorithm for z_t in two-player bilinear games, including (1) shown in Theorem 1. Alternatively, in periodic monotone games, no $\eta > 0$ guarantees that the tangent residual converges with the modified OG algorithm, since the modified OG algorithm diverges at an exponential rate in the example provided.*

Proof sketch. We construct a two-player bilinear game with the period of 2 and prove that for $t = 2k - 1$ there exists a initial point z_0 and $z_{-1/2}$ making $\|z_t\| \geq C\lambda^k$ with $\eta, C > 0$ and $\lambda > 1$. Then, with $r^{tan}(z_t) = \|z_t\|$ in this game, we show that $r^{tan}(z_t)$ diverges with the variant of the OG algorithm.

The full proof of Theorem 5 is deferred to Section G.

5.2 TIME-VARYING CASE OF THE STANDARD OG ALGORITHM APPLICABLE TO PERIODIC CASE

In this part, we show that the weak convergence of the standard OG algorithm is highly robust to the changing cost function as long as L_t is bounded.

Theorem 6. *If there exists a common Nash equilibrium z^* and the game at each time t is monotone and L_t -Lipschitz with the step size η satisfying $\eta \in (0, \frac{1}{2L_{\max}})$ for the OG algorithm, then $\forall T \geq 1$, under the Assumption*

$$\min_{t \in \llbracket T \rrbracket} r_{F, \mathcal{Z}}^{tan}(z_{t+\frac{1}{2}}) \leq \min_{t \in \llbracket T \rrbracket} \frac{\|z_{t+1} - z_t\|}{\eta} = O\left(\frac{1}{\sqrt{T}}\right) \quad (2)$$

where $L_{\max} = \max_{t \in \llbracket T \rrbracket} L_t$ and $\langle N(z), z - z^* \rangle \geq \rho \|N(z)\|^2$. This result can be extended to $\langle F_t(z) + N(z) - F_t(z') - N(z'), z - z' \rangle \geq \rho \|F_t(z) + N(z) - F_t(z') - N(z')\|^2, \forall z, z' \in \mathcal{Z}$ with $\rho \in (-\frac{1}{12\sqrt{3}L_{\max}}, 0]$ if $\frac{1}{2} + \frac{2\rho}{\eta} - 2\eta^2 L_{\max}^2 > 0$. Note that L_{\max} exists for periodic games since there are at most \mathcal{T} different games all the time so that L_t is in a limited set and there exists a greatest L_t .

The proof of Theorem 6 is deferred to Section H.

6 CONCLUSION

In this paper, we provide a surprising result that the RG, ARG and even modified OG algorithms may diverge at an exponential rate for periodic games even if only two players are involved where the standard OG algorithm converges. We also prove that the RG and ARG algorithms with constant step sizes converge to the Nash equilibrium in time-varying multi-player convergent perturbed monotone games with bounded action sets at satisfactory rates. This is comparable to those algorithms and related algorithms for both time-invariant and time-varying games.

There remain some interesting future research topics. One is whether a last-iterate convergence rate can be established for the OG algorithm. Another is whether and when the RG, OG, and ARG algorithms behave similarly for stochastic games no less general than those in this paper.

REFERENCES

- 486
487
488 Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, and Tuomas Sandholm. On the convergence
489 of no-regret learning dynamics in time-varying games. In *Thirty-seventh Conference on Neural*
490 *Information Processing Systems*, 2023. URL [https://openreview.net/forum?id=](https://openreview.net/forum?id=0x2Ou3xHbH)
491 [0x2Ou3xHbH](https://openreview.net/forum?id=0x2Ou3xHbH).
- 492 Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-
493 algorithm and applications. *Theory of Computing*, 8(6):121–164, 2012. doi: 10.4086/toc.2012.
494 v008a006. URL <https://theoryofcomputing.org/articles/v008a006>.
- 495 Z. Benzaid and D. A. Lutz. Asymptotic representation of solutions of perturbed systems of linear
496 difference equations. *Studies in Applied Mathematics*, 77(3):195–221, 1987. doi: 10.1002/
497 sapm1987773195. URL [https://onlinelibrary.wiley.com/doi/abs/10.1002/](https://onlinelibrary.wiley.com/doi/abs/10.1002/sapm1987773195)
498 [sapm1987773195](https://onlinelibrary.wiley.com/doi/abs/10.1002/sapm1987773195).
- 499 Yang Cai and Constantinos Daskalakis. On minmax theorems for multiplayer games. In *Proceedings*
500 *of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA ’11, pp.
501 217–234, USA, 2011. Society for Industrial and Applied Mathematics.
- 502 Yang Cai and Weiqiang Zheng. Accelerated single-call methods for constrained min-max optimization.
503 In *The Eleventh International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=HRwN7IQLUKA>.
- 504 Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. Finite-time last-iterate conver-
505 gence for learning in multi-player games. In S. Koyejo, S. Mohamed, A. Agarwal,
506 D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Process-*
507 *ing Systems*, volume 35, pp. 33904–33919. Curran Associates, Inc., 2022. URL
508 [https://proceedings.neurips.cc/paper_files/paper/2022/file/](https://proceedings.neurips.cc/paper_files/paper/2022/file/db2d2001f63e83214b08948b459f69f0-Paper-Conference.pdf)
509 [db2d2001f63e83214b08948b459f69f0-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/db2d2001f63e83214b08948b459f69f0-Paper-Conference.pdf).
- 510 Yang Cai, Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and
511 Weiqiang Zheng. Fast last-iterate convergence of learning in games requires forgetful algorithms,
512 2024a. URL <https://arxiv.org/abs/2406.10631>.
- 513 Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. Accelerated algorithms for constrained
514 nonconvex-nonconcave min-max optimization and comonotone inclusion. In Ruslan Salakhut-
515 dinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix
516 Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, vol-
517 *ume 235 of Proceedings of Machine Learning Research*, pp. 5312–5347. PMLR, 21–27 Jul 2024b.
518 URL <https://proceedings.mlr.press/v235/cai24f.html>.
- 519 Adrian Rivera Cardoso, Jacob Abernethy, He Wang, and Huan Xu. Competing against nash equilibria
520 in adversarially changing zero-sum games. In Kamalika Chaudhuri and Ruslan Salakhutdinov
521 (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of
522 *Proceedings of Machine Learning Research*, pp. 921–930. PMLR, 09–15 Jun 2019. URL [https://proceedings.mlr.press/v97/](https://proceedings.mlr.press/v97/cardoso19a.html)
523 [cardoso19a.html](https://proceedings.mlr.press/v97/cardoso19a.html).
- 524 Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with
525 applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, May 2011.
526 ISSN 1573-7683. doi: 10.1007/s10851-010-0251-1. URL [https://doi.org/10.1007/](https://doi.org/10.1007/s10851-010-0251-1)
527 [s10851-010-0251-1](https://doi.org/10.1007/s10851-010-0251-1).
- 528 Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player nash
529 equilibria. *J. ACM*, 56(3), May 2009. ISSN 0004-5411. doi: 10.1145/1516512.1516516. URL
530 <https://doi.org/10.1145/1516512.1516516>.
- 531 Yanzheng Chen and Jun Yu. Classic but everlasting: Traditional gradient-based algorithms converge
532 fast even in time-varying multi-player games. In *The Thirteenth International Conference on Learn-*
533 *ing Representations*, 2025. URL <https://openreview.net/forum?id=t8FG4cJuL3>.
- 534 Emmanuel d’Angelo, Laurent Jacques, Alexandre Alahi, and Pierre Vandergheynst. From bits to
535 images: Inversion of local binary descriptors. *IEEE Transactions on Pattern Analysis and Machine*
536 *Intelligence*, 36(5):874–887, 2014. doi: 10.1109/TPAMI.2013.228.
537
538
539

- 540 Constantinos Daskalakis and Ioannis Panageas. Last-iterate convergence: Zero-sum games and
541 constrained min-max optimization. In Avrim Blum (ed.), *10th Innovations in Theoretical
542 Computer Science Conference (ITCS 2019)*, volume 124 of *Leibniz International Proceedings
543 in Informatics (LIPIcs)*, pp. 27:1–27:18, Dagstuhl, Germany, 2019. Schloss Dagstuhl –
544 Leibniz-Zentrum für Informatik. ISBN 978-3-95977-095-8. doi: 10.4230/LIPIcs.ITCS.2019.
545 27. URL [https://drops-dev.dagstuhl.de/entities/document/10.4230/
546 LIPIcs.ITCS.2019.27](https://drops-dev.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2019.27).
- 547 Constantinos Daskalakis and Christos H. Papadimitriou. On a network generalization of the minmax
548 theorem. In Susanne Albers, Alberto Marchetti-Spaccamela, Yossi Matias, Sotiris Nikolettseas,
549 and Wolfgang Thomas (eds.), *Automata, Languages and Programming*, pp. 423–434, Berlin,
550 Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-02930-1.
- 552 Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of
553 computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009. doi: 10.1137/
554 070699652. URL <https://doi.org/10.1137/070699652>.
- 555 Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training GANs with
556 optimism. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver,
557 BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. URL
558 <https://openreview.net/forum?id=SJJySbbAZ>.
- 560 Argyrios Deligkas, John Fearnley, Alexandros Hollender, and Themistoklis Melissourgos. Pure-
561 circuit: Strong inapproximability for PPA. In *2022 IEEE 63rd Annual Symposium on Foundations
562 of Computer Science (FOCS)*, pp. 159–170, 2022. doi: 10.1109/FOCS54457.2022.00022.
- 563 Jelena Diakonikolas. Halpern iteration for near-optimal and parameter-free monotone inclu-
564 sion and strong solutions to variational inequalities. In Jacob Abernethy and Shivani Agar-
565 wal (eds.), *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Pro-
566 ceedings of Machine Learning Research*, pp. 1428–1451. PMLR, 09–12 Jul 2020. URL
567 <https://proceedings.mlr.press/v125/diakonikolas20a.html>.
- 568
569 Vu Tien Dung, Pham Ky Anh, and Duong Viet Thong. Convergence of two-step inertial tseng’s
570 extragradient methods for quasimonotone variational inequality problems. *Communications in
571 Nonlinear Science and Numerical Simulation*, 136:108110, 2024. ISSN 1007-5704. doi: 10.1016/
572 j.cnsns.2024.108110. URL [https://www.sciencedirect.com/science/article/
573 pii/S1007570424002958](https://www.sciencedirect.com/science/article/pii/S1007570424002958).
- 574 Benoit Duvocelle, Panayotis Mertikopoulos, Mathias Staudigl, and Dries Vermeulen. Multiagent
575 online learning in time-varying games. *Mathematics of Operations Research*, 48(2):914–941, 2023.
576 doi: 10.1287/moor.2022.1283. URL <https://doi.org/10.1287/moor.2022.1283>.
- 577
578 S. Elaydi and I. Györi. Asymptotic theory for delay differene equations. *Journal of Difference
579 Equations and Applications*, 1(2):99–116, 1995. doi: 10.1080/10236199508808012. URL [https://doi.org/
580 //doi.org/10.1080/10236199508808012](https://doi.org/10.1080/10236199508808012).
- 581 Eyal Even-dar, Yishay Mansour, and Uri Nadav. On the convergence of regret minimization dynamics
582 in concave games. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of
583 Computing, STOC ’09*, pp. 523–532, New York, NY, USA, 2009. Association for Computing
584 Machinery. ISBN 9781605585062. doi: 10.1145/1536414.1536486. URL [https://doi.org/
585 10.1145/1536414.1536486](https://doi.org/10.1145/1536414.1536486).
- 586
587 Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complemen-
588 tarity Problems*. Springer New York, Jun 2007. ISBN 978-0-387-95580-3. doi: 10.1007/b97543.
- 589 Yi Feng, Hu Fu, Qun Hu, Ping Li, Ioannis Panageas, bo peng, and Xiao Wang. On the last-
590 iterate convergence in time-varying zero-sum games: Extra gradient succeeds where optimism
591 fails. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Ad-
592 vances in Neural Information Processing Systems*, volume 36, pp. 21933–21944. Curran Asso-
593 ciates, Inc., 2023. URL [https://proceedings.neurips.cc/paper_files/paper/
2023/file/457ab261562014550e53351422f69834-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2023/file/457ab261562014550e53351422f69834-Paper-Conference.pdf).

- 594 Yi Feng, Ping Li, Ioannis Panageas, and Xiao Wang. Last-iterate convergence separation between
595 extra-gradient and optimism in constrained periodic games. In *The 40th Conference on Un-*
596 *certainty in Artificial Intelligence*, 2024. URL [https://openreview.net/forum?id=](https://openreview.net/forum?id=IaH9nprqiU)
597 [IaH9nprqiU](https://openreview.net/forum?id=IaH9nprqiU).
598
- 599 Tanner Fiez, Ryann Sim, Stratis Skoulakis, Georgios Piliouras, and Lillian Ratliff. Online learning in
600 periodic zero-sum games. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman
601 Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 10313–10325.
602 Curran Associates, Inc., 2021. URL [https://proceedings.neurips.cc/paper_](https://proceedings.neurips.cc/paper_files/paper/2021/file/55563844bcd4bba067fe86ac1f008c7e-Paper.pdf)
603 [files/paper/2021/file/55563844bcd4bba067fe86ac1f008c7e-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/55563844bcd4bba067fe86ac1f008c7e-Paper.pdf).
- 604 John E. Franke and James F. Selgrade. Attractors for discrete periodic dynamical systems. *Journal*
605 *of Mathematical Analysis and Applications*, 286(1):64–79, 2003. ISSN 0022-247X. doi: 10.
606 1016/S0022-247X(03)00417-7. URL [https://www.sciencedirect.com/science/](https://www.sciencedirect.com/science/article/pii/S0022247X03004177)
607 [article/pii/S0022247X03004177](https://www.sciencedirect.com/science/article/pii/S0022247X03004177).
608
- 609 Noah Golowich, Sarath Pattathil, and Constantinos Daskalakis. Tight last-iterate con-
610 vergence rates for no-regret learning in multi-player games. In H. Larochelle,
611 M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural In-*
612 *formation Processing Systems*, volume 33, pp. 20766–20778. Curran Associates, Inc.,
613 2020. URL [https://proceedings.neurips.cc/paper_files/paper/2020/](https://proceedings.neurips.cc/paper_files/paper/2020/file/eea5d933e9dce59c7dd0f6532f9ea81b-Paper.pdf)
614 [file/eea5d933e9dce59c7dd0f6532f9ea81b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/eea5d933e9dce59c7dd0f6532f9ea81b-Paper.pdf).
- 615 Hédi Hadiji, Sarah Sachs, and Cristóbal Guzmán. Tracking solutions of time-varying variational
616 inequalities, 2024. URL <https://arxiv.org/abs/2406.14059>.
617
- 618 Benjamin Halpern. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical*
619 *Society*, 73:957–961, 1967. ISSN 0273-0979. doi: 10.1090/S0002-9904-1967-11864-0.
620
- 621 Bingsheng He and Xiaoming Yuan. An accelerated inexact proximal point algorithm for convex
622 minimization. *Journal of Optimization Theory and Applications*, 154(2):536–548, August 2012.
- 623 Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. On
624 the convergence of single-call stochastic extra-gradient methods. In H. Wallach,
625 H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett (eds.), *Ad-*
626 *vances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.,
627 2019. URL [https://proceedings.neurips.cc/paper_files/paper/2019/](https://proceedings.neurips.cc/paper_files/paper/2019/file/4625d8e31dad7d1c4c83399a6eb62f0c-Paper.pdf)
628 [file/4625d8e31dad7d1c4c83399a6eb62f0c-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/4625d8e31dad7d1c4c83399a6eb62f0c-Paper.pdf).
629
- 630 G. M. Korpelevich. The extragradient method for finding saddle points and other problems.
631 *Matecon*, 12:747–756, 1976. URL [https://api.semanticscholar.org/CorpusID:](https://api.semanticscholar.org/CorpusID:118602977)
632 [118602977](https://api.semanticscholar.org/CorpusID:118602977).
- 633 Tianyi Lin, Zhengyuan Zhou, Panayotis Mertikopoulos, and Michael Jordan. Finite-time last-
634 iterate convergence for multi-agent learning in games. In Hal Daumé III and Aarti Singh
635 (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119
636 of *Proceedings of Machine Learning Research*, pp. 6161–6171. PMLR, 13–18 Jul 2020. URL
637 <https://proceedings.mlr.press/v119/lin20h.html>.
638
- 639 Charles H. C. Little, Kee L. Teo, and Bruce van Brunt. *Infinite Products*, pp. 39–129. Springer Inter-
640 national Publishing, Cham, 2022. ISBN 978-3-030-90646-7. doi: 10.1007/978-3-030-90646-7_2.
641 URL https://doi.org/10.1007/978-3-030-90646-7_2.
- 642 Fangrui Lv, Jian Liang, Shuang Li, Jinming Zhang, and Di Liu. Improving generalization with
643 domain convex game. In *Proceedings of the IEEE/CVF Conference on Computer Vision and*
644 *Pattern Recognition (CVPR)*, pp. 24315–24324, June 2023.
645
- 646 Yu. Malitsky. Projected reflected gradient methods for monotone variational inequalities. *SIAM*
647 *Journal on Optimization*, 25(1):502–520, 2015. doi: 10.1137/14097238X. URL [https://doi.](https://doi.org/10.1137/14097238X)
[org/10.1137/14097238X](https://doi.org/10.1137/14097238X).

- 648 Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar,
649 and Georgios Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra(-
650 gradient) mile. In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=Bkg8jjc9KQ>.
- 651
652 Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. A unified analysis of extra-gradient
653 and optimistic gradient methods for saddle point problems: Proximal point approach. In Silvia
654 Chiappa and Roberto Calandra (eds.), *Proceedings of the Twenty Third International Conference*
655 *on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*,
656 pp. 1497–1507. PMLR, 26–28 Aug 2020a. URL <https://proceedings.mlr.press/v108/mokhtari20a.html>.
- 657
658 Aryan Mokhtari, Asuman E. Ozdaglar, and Sarath Pattathil. Convergence rate of $\mathcal{O}(1/k)$ for
659 optimistic gradient and extragradient methods in smooth convex-concave saddle point problems.
660 *SIAM Journal on Optimization*, 30(4):3230–3251, 2020b. doi: 10.1137/19M127375X. URL
661 <https://doi.org/10.1137/19M127375X>.
- 662
663 Renato D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient
664 method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20(6):2755–2787,
665 2010. doi: 10.1137/090753127. URL <https://doi.org/10.1137/090753127>.
- 666
667 Arkadi Nemirovski. Prox-method with rate of convergence $\mathcal{O}(1/t)$ for variational inequalities with
668 lipschitz continuous monotone operators and smooth convex-concave saddle point problems.
669 *SIAM Journal on Optimization*, 15(1):229–251, 2004. doi: 10.1137/S1052623403425629. URL
670 <https://doi.org/10.1137/S1052623403425629>.
- 671
672 L. D. Popov. A modification of the arrow-hurwicz method for search of saddle points. *Mathematical*
673 *notes of the Academy of Sciences of the USSR*, 28(5):845–848, Nov 1980. ISSN 1573-8876. doi:
674 10.1007/BF01141092. URL <https://doi.org/10.1007/BF01141092>.
- 675
676 J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econo-*
677 *metrica*, 33(3):520–534, 1965. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/1911749>.
- 678
679 Satoru Murakami: Saber Elaydi and Etsuyo Kamiyama. Asymptotic equivalence for differ-
680 ence equations with infinite delay. *Journal of Difference Equations and Applications*, 5(1):
681 1–23, 1999. doi: 10.1080/10236199908808167. URL <https://doi.org/10.1080/10236199908808167>.
- 682
683 Pier Giuseppe Sessa, Ilija Bogunovic, Andreas Krause, and Maryam Kamgarpour. Con-
684 textual games: Multi-agent learning with side information. In H. Larochelle,
685 M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural In-*
686 *formation Processing Systems*, volume 33, pp. 21912–21922. Curran Associates, Inc.,
687 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/f9afa97535cf7c8789alc50a2cd83787-Paper.pdf.
- 688
689 Quoc Tran-Dinh. The connection between nesterov’s accelerated methods and halpern fixed-point
690 iterations, 2022. URL <https://arxiv.org/abs/2203.04869>.
- 691
692 Paul Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM*
693 *Journal on Control and Optimization*, 38(2):431–446, 2000. doi: 10.1137/S0363012998338806.
694 URL <https://doi.org/10.1137/S0363012998338806>.
- 695
696 Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Last-iterate convergence of
697 decentralized optimistic gradient descent/ascent in infinite-horizon competitive markov games. In
698 Mikhail Belkin and Samory Kpotufe (eds.), *Proceedings of Thirty Fourth Conference on Learning*
699 *Theory*, volume 134 of *Proceedings of Machine Learning Research*, pp. 4259–4299. PMLR, 15–19
700 Aug 2021. URL <https://proceedings.mlr.press/v134/wei21a.html>.
- 701
702 Yu-Hu Yan, Peng Zhao, and Zhi-Hua Zhou. Fast rates in time-varying strongly monotone games.
In *Proceedings of the 40th International Conference on Machine Learning*, ICML’23. JMLR.org,
2023.

702 Jun Yang and Hongwei Liu. A modified projected gradient method for monotone variational
703 inequalities. *Journal of Optimization Theory and Applications*, 179(1):197–211, Oct 2018.
704 ISSN 1573-2878. doi: 10.1007/s10957-018-1351-0. URL [https://doi.org/10.1007/
705 s10957-018-1351-0](https://doi.org/10.1007/s10957-018-1351-0).

706 Jun Yang and Hongwei Liu. Strong convergence result for solving monotone variational inequal-
707 ities in hilbert space. *Numerical Algorithms*, 80:741–752, 3 2019. ISSN 15729265. doi: 10.
708 1007/S11075-018-0504-4/METRICS. URL [https://link.springer.com/article/
709 10.1007/s11075-018-0504-4](https://link.springer.com/article/10.1007/s11075-018-0504-4).

710 Taeho Yoon and Ernest K Ryu. Accelerated algorithms for smooth convex-concave minimax problems
711 with $\mathcal{O}(1/k^2)$ rate on squared gradient norm. In Marina Meila and Tong Zhang (eds.), *Proceedings
712 of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine
713 Learning Research*, pp. 12098–12109. PMLR, 18–24 Jul 2021. URL [https://proceedings.
714 mlr.press/v139/yoon21d.html](https://proceedings.mlr.press/v139/yoon21d.html).

715 Mengxiao Zhang, Peng Zhao, Haipeng Luo, and Zhi-Hua Zhou. No-regret learning in time-varying
716 zero-sum games. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu,
717 and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*,
718 volume 162 of *Proceedings of Machine Learning Research*, pp. 26772–26808. PMLR, 17–23 Jul
719 2022. URL <https://proceedings.mlr.press/v162/zhang22an.html>.

720 Minxin Zhang, Fuqun Han, Yat Tin Chow, Stanley Osher, and Hayden Schaeffer. Inexact proximal
721 point algorithms for zeroth-order global optimization, 2025. URL [https://arxiv.org/
722 abs/2412.11485](https://arxiv.org/abs/2412.11485).

723 Jinjing Zhu, Haotian Bai, and Lin Wang. Patch-mix transformer for unsupervised domain adaptation:
724 A game perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
725 Recognition (CVPR)*, pp. 3561–3571, June 2023.

728 A USAGE OF LARGE LANGUAGE MODELS (LLMs) IN THE PAPER

729 LLMs are applied to revise the potential grammatical mistakes and improve wording of sentences.
730 No significant contents are generated by LLMs.

731 B ADDITIONAL LEMMAS

732 In our proof, the following lemmas are necessary.

733 **Lemma 3.** (Cai & Zheng (2023)) $\forall x_i, y_j, u_2, u_4 \in \mathbb{R}^n$ with $i = 0, 1, 2, 3, 4, j = 1, 2, 3, 4$ and
734 $k \geq 1, q \in (0, 1)$, if $x_3 = x_2 - y_1 - u_2$ and $x_4 = x_2 - y_3 - u_4$, then there exists

$$\begin{aligned} & \|y_2 + u_2\|^2 + \|y_2 - y_1\|^2 - \|y_4 + u_4\|^2 - \|y_4 - y_3\|^2 - 2\langle y_4 - y_2, x_4 - x_2 \rangle \\ & - 2 \left(\frac{1}{4} \|x_4 - x_3\|^2 - \|y_4 - y_3\|^2 \right) - 2\langle u_4 - u_2, x_4 - x_2 \rangle \\ & = \left\| \frac{x_3 - x_4}{2} + y_1 - y_2 \right\|^2 + \left\| \frac{x_3 + x_4}{2} - x_2 + y_2 + u_2 \right\|^2 \end{aligned} \quad (3)$$

735 **Lemma 4.** (Cai & Zheng (2023)) $\forall x_i, y_j, u_2, u_4 \in \mathbb{R}^n$ with $i = 0, 1, 2, 3, 4, j = 1, 2, 3, 4$ and
736 $k \geq 1, q \in (0, 1)$, if $x_3 = x_2 - y_1 - u_2 + \frac{1}{k+1}(x_0 - x_2)$ and $x_4 = x_2 - y_3 - u_4 + \frac{1}{k+1}(x_0 - x_2)$
737 then there exists

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

Lemma 5. (Little et al. (2022)) If $\{a_n\}$ is a sequence of non-negative numbers, then the series $\sum_{j=1}^{\infty} a_j$ and the product $\prod_{j=1}^{\infty} (1 + a_j)$ either both converge or both diverge.

C MISSING PROOFS OF THEOREM 1

In periodic games, each period as a whole is time-invariant (Franke & Selgrade (2003), Section 3). We take advantage of this property in our analysis.

$$z_{t+\frac{1}{2}} = 2z_t - z_{t-1} = \begin{bmatrix} 2x_t - x_{t-1} \\ 2y_t - y_{t-1} \end{bmatrix} \quad (5)$$

Hence, we have

$$\begin{aligned} z_{t+1} &= z_t - \eta F_t(z_{t+\frac{1}{2}}) = z_t - \eta \begin{bmatrix} A_t(2y_t - y_{t-1}) \\ -A_t^T(2x_t - x_{t-1}) \end{bmatrix} \\ &= \begin{bmatrix} x_t - 2\eta A_t y_t + \eta A_t y_{t-1} \\ y_t + 2\eta A_t^T x_t - \eta A_t^T x_{t-1} \end{bmatrix} \end{aligned} \quad (6)$$

Then, there exists

$$\begin{bmatrix} x_t \\ y_t \\ x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ 0 & \eta A_t & I & -2\eta A_t \\ -\eta A_t^T & 0 & 2\eta A_t^T & I \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ x_t \\ y_t \end{bmatrix} \quad (7)$$

Denote $\begin{bmatrix} 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ 0 & \eta A_t & I & -2\eta A_t \\ -\eta A_t^T & 0 & 2\eta A_t^T & I \end{bmatrix}$ as P_t . We have

$$P_t = \begin{cases} \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \eta & -\eta & 1 & -2\eta & 2\eta \\ -\eta & 0 & 0 & 2\eta & 1 & 0 \\ \eta & 0 & 0 & -2\eta & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & -\eta & \eta & 1 & 2\eta & -2\eta \\ \eta & 0 & 0 & -2\eta & 1 & 0 \\ -\eta & 0 & 0 & 2\eta & 0 & 1 \end{bmatrix}, & t \text{ is odd} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \eta & -\eta & 1 & -2\eta & 2\eta \\ -\eta & 0 & 0 & 2\eta & 1 & 0 \\ \eta & 0 & 0 & -2\eta & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & -\eta & \eta & 1 & 2\eta & -2\eta \\ \eta & 0 & 0 & -2\eta & 1 & 0 \\ -\eta & 0 & 0 & 2\eta & 0 & 1 \end{bmatrix}, & t \text{ is even} \end{cases} \quad (8)$$

Without loss of generality, assume t is odd. Let $\mathcal{A}_{t,t+1} = P_{t+1}P_t$. Then,

$$\begin{aligned} &\det(\lambda I - \mathcal{A}_{t,t+1}) \\ &= 64\eta^4 \lambda^4 - 96\eta^4 \lambda^3 + 36\eta^4 \lambda^2 - 4\eta^4 \lambda - 16\eta^2 \lambda^5 \\ &\quad + 36\eta^2 \lambda^4 - 24\eta^2 \lambda^3 + 4\eta^2 \lambda^2 + \lambda^6 - 3\lambda^5 + 3\lambda^4 - \lambda^3 \\ &= \lambda(\lambda - 1)(\lambda^4 - (16\eta^2 + 2)\lambda^3 + (64\eta^4 + 20\eta^2 + 1)\lambda^2 \\ &\quad - (32\eta^4 + 4\eta^2)\lambda + 4\eta^4) \\ &= \lambda(\lambda - 1)(\lambda^2 - (8\eta^2 + 1)\lambda + 2\eta^2)^2 \\ &= \left(\lambda - \left(\frac{8\eta^2 + 1 + \sqrt{64\eta^4 + 8\eta^2 + 1}}{2} \right) \right)^2 \\ &\quad \left(\lambda - \left(\frac{8\eta^2 + 1 - \sqrt{64\eta^4 + 8\eta^2 + 1}}{2} \right) \right)^2 \lambda(\lambda - 1) \end{aligned} \quad (9)$$

Then, $\mathcal{A}_{t,t+1}$ has an eigenvalue $\lambda_{1,2} = \frac{8\eta^2+1+\sqrt{64\eta^4+8\eta^2+1}}{2} > 1$. Since step size $\eta > 0$, the spectral radius of matrix $\mathcal{A}_{t,t+1}$ is no less than $\lambda_{1,2}$, and then greater than 1. Denote the corresponding eigenvector of $\lambda_{1,2}$ for $\mathcal{A}_{t,t+1}$ as v . Then, we obtain

$$\left\| \left(\prod_{i=0}^k \mathcal{A}_{2i,2i+1} \right) v \right\| = \|\lambda_{1,2}^{k+1} v\| \quad (10)$$

Since the action set is \mathbb{R}^3 , we have

$$r_t^{tan}(z) = \|F_t(z)\| = \left\| \begin{bmatrix} A_t y \\ -A_t^T x \end{bmatrix} \right\| \quad (11)$$

Hence, $\forall t \geq 1$,

$$\|r_{\mathcal{G}_t, \mathcal{Z}}^{tan}(z)\| = \left\| \begin{bmatrix} A_t y \\ -A_t^T x \end{bmatrix} \right\| = \|z\| \quad (12)$$

Since $\|\lambda_{1,2}^k v\| = \|\lambda_{1,2}\|^k \|v\|$ grows exponentially, $\|z_{2k}\|$ does not converge. This means that the tangent residuals of \mathcal{G} with the RG algorithm diverge exponentially. This completes the proof.

D MISSING PROOFS OF THEOREM 2

To prove Theorem 2, we apply a potential function argument. We first show that the potential function is approximately non-increasing and then prove that it is upper bounded by a term independent of T . As the potential function at step t is also $O(t)r^{tan}(z_t)^2$, we conclude that the RG algorithm converges at the rate of $\max \left\{ O\left(\frac{1}{\sqrt{T}}\right), O\left(\sqrt{\sum_{t=T/2}^{\infty} \max \|G_t\|}\right) \right\}$.

D.1 APPROXIMATE MONOTONICITY OF THE POTENTIAL FUNCTION

To construct a potential function, we denote

$$c_{t+1} = \frac{z_t - \eta F_t(z_{t+\frac{1}{2}}) - z_{t+1}}{\eta}, \forall t \geq 0 \quad (13)$$

Note that according to the update rules of RG, $z_{t+1} = \Pi_{\mathcal{Z}}[z_t - \eta F_t(z_{t+1/2})]$. This means $c_{t+1} \in N_{\mathcal{Z}}(z_{t+1})$. To our analysis on the RG algorithm, the potential function P_t is applied, which is defined as

$$P_t = \|F_{t-1}(z_t) + c_t\|^2 + \left\| F_{t-1}(z_t) - F_{t-1}(z_{t-\frac{1}{2}}) \right\|^2, \forall t \geq 1 \quad (14)$$

Lemma 6. *With the same settings as Theorem 2, $\forall t \geq 1$, $P_t - P_{t+1} \geq (-\frac{2}{\eta} - 8L)D_{\mathcal{Z}} \max \|G_t\| - \frac{2}{\eta}D_{\mathcal{Z}} \max \|G_{t-1}\| - 8 \max \|G_t\|^2$.*

Proof. We only need to prove that the sum of $P_t - P_{t+1} + (\frac{2}{\eta} + 8L)D_{\mathcal{Z}} \max \|G_t\| + \frac{2}{\eta}D_{\mathcal{Z}} \max \|G_{t-1}\| + 8 \max \|G_t\|^2$ and some non-positive terms is approximately non-negative. Since F_{∞} is monotone, we have

$$-2\langle \eta F_{\infty}(z_{t+1}) - \eta F_{\infty}(z_t), z_{t+1} - z_t \rangle \leq 0 \quad (15)$$

which means

$$-2\langle \eta F_t(z_{t+1}) - \eta F_{t-1}(z_t), z_{t+1} - z_t \rangle + 2\langle \eta G_t(z_{t+1}) - \eta G_{t-1}(z_t), z_{t+1} - z_t \rangle \leq 0 \quad (16)$$

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

$$\frac{1}{4} \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| \eta F_{\infty}(z_{t+1}) - \eta F_{\infty}(z_{t+\frac{1}{2}}) \right\|^2 \geq 0 \quad (17)$$

864 Since

$$\begin{aligned}
865 & \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 - \left\| \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\|^2 \\
866 & - 2 \left\langle \eta F_\infty(z_{t+1}) - \eta F_\infty(z_{t+\frac{1}{2}}), \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\rangle - \left\| \eta F_\infty(z_{t+1}) - \eta F_\infty(z_{t+\frac{1}{2}}) \right\|^2 \\
867 & = \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 - \left(\left\| \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\| + \left\| \eta F_\infty(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\| \right)^2 \\
868 & \leq 0 \\
869 & \\
870 & \\
871 & \\
872 & \tag{18}
\end{aligned}$$

873 we have

$$\begin{aligned}
874 & - 2 \left(\frac{1}{4} \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 \right) \\
875 & \leq - 2 \left(- \left\| \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\|^2 - 2 \left\langle \eta F_\infty(z_{t+1}) - \eta F_\infty(z_{t+\frac{1}{2}}), \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\rangle \right) \\
876 & \\
877 & \\
878 & \tag{19}
\end{aligned}$$

879 so that

$$\begin{aligned}
880 & - 2 \left(\frac{1}{4} \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 \right) - 2 \left\| \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\|^2 \\
881 & - 4 \left\langle \eta F_\infty(z_{t+1}) - \eta F_\infty(z_{t+\frac{1}{2}}), \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\rangle \\
882 & \leq 0 \\
883 & \\
884 & \tag{20}
\end{aligned}$$

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

$$886 -2 \langle \eta c_{t+1} - \eta c_t, z_{t+1} - z_t \rangle \leq 0 \tag{21}$$

887 \square

888 We use the following equivalent formations of $z_{t+1/2}$ and z_{t+1} .

$$\begin{aligned}
889 & z_{t+\frac{1}{2}} = 2z_t - z_{t-1} = z_t - (z_{t-1} - z_t) = z_t - \eta F_{t-1}(z_{t-\frac{1}{2}}) - \eta c_t \\
890 & z_{t+1} = \Pi_{\mathcal{Z}} \left[z_t - \eta F_t(z_{t+\frac{1}{2}}) \right] = z_t - \eta F_t(z_{t+\frac{1}{2}}) - \eta c_{t+1} \\
891 & \tag{22}
\end{aligned}$$

892 According to Lemma 3, by replacing x_k with $z_{t-1+\frac{k}{2}}$, y_k with $\eta F_{t-1}(z_{t-1+\frac{k}{2}})$, u_2 with ηc_t , and u_4 with ηc_{t+1} , since both $x_3 = x_2 - y_1 - u_2$ and $x_4 = x_2 - y_3 - u_4$ hold true, we have

$$\begin{aligned}
893 & \eta^2 (P_t - P_{t+1}) + \text{LHS of Inequality (16)} + \text{LHS of Inequality (20)} + \text{LHS of Inequality (21)} \\
894 & = \left\| \frac{z_{t+\frac{1}{2}} - z_{t+1}}{2} + \eta F_{t-1}(z_{t-\frac{1}{2}}) - \eta F_{t-1}(z_t) \right\|^2 + \left\| \frac{z_{t+\frac{1}{2}} + z_{t+1}}{2} - z_t + \eta F_{t-1}(z_t) + \eta c_t \right\|^2 \\
895 & + 2 \langle \eta G_t(z_{t+1}) - \eta G_{t-1}(z_t), z_{t+1} - z_t \rangle - 2 \left\| \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\|^2 \\
896 & - 4 \left\langle \eta F_\infty(z_{t+1}) - \eta F_\infty(z_{t+\frac{1}{2}}), \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\rangle \\
897 & \\
898 & \tag{23}
\end{aligned}$$

899 Since \mathcal{Z} is bounded and $z_t, z_{t+1} \in \mathcal{Z}$ there exists $D_{\mathcal{Z}} > 0$ so that $\|z_t - z_{t+1}\| \leq D_{\mathcal{Z}}$, we have

$$\begin{aligned}
900 & 2 \langle \eta G_t(z_{t+1}) - \eta G_{t-1}(z_t), z_{t+1} - z_t \rangle - 2 \left\| \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\|^2 \\
901 & - 4 \left\langle \eta F_\infty(z_{t+1}) - \eta F_\infty(z_{t+\frac{1}{2}}), \eta G_t(z_{t+1}) - \eta G_t(z_{t+\frac{1}{2}}) \right\rangle \\
902 & \geq - 2\eta D_{\mathcal{Z}} (\max \|G_t\| + \max \|G_{t-1}\|) - 8\eta^2 \max \|G_t\|^2 \\
903 & - 4\eta^2 \left\| F_\infty(z_{t+1}) - F_\infty(z_{t+\frac{1}{2}}) \right\| \left\| G_t(z_{t+1}) - G_t(z_{t+\frac{1}{2}}) \right\| \\
904 & \geq - 2\eta D_{\mathcal{Z}} (\max \|G_t\| + \max \|G_{t-1}\|) - 8\eta^2 \max \|G_t\|^2 - 8L\eta^2 \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\| \max \|G_t\| \\
905 & \geq - 2\eta D_{\mathcal{Z}} (\max \|G_t\| + \max \|G_{t-1}\|) - 8\eta^2 \max \|G_t\|^2 - 8L\eta^2 D_{\mathcal{Z}} \max \|G_t\| \\
906 & = (-2\eta - 8L\eta^2) D_{\mathcal{Z}} \max \|G_t\| - 2\eta D_{\mathcal{Z}} \max \|G_{t-1}\| - 8\eta^2 \max \|G_t\|^2 \\
907 & \\
908 & \tag{24}
\end{aligned}$$

Thus we conclude $P_t - P_{t+1} \geq (-\frac{2}{\eta} - 8L)D_{\mathcal{Z}} \max \|G_t\| - \frac{2}{\eta}D_{\mathcal{Z}} \max \|G_{t-1}\| - 8 \max \|G_t\|^2$.

D.2 BEST-ITERATE CONVERGENCE OF THE ERROR FROM THE OPTIMAL SOLUTION

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

Lemma 7. *In the settings of Theorem 2, for any $T \geq 1$, we have*

$$\sum_{t=1}^T \left\| z_{t+\frac{1}{2}} - z_t \right\|^2 = \sum_{t=1}^T \|z_t - z_{t-1}\|^2 \leq \frac{H^2}{1 - (1 + \sqrt{2})\eta L} \quad (25)$$

Proof. By the update rules of the RG algorithm, we have $z_{t+1/2} = 2z_t - z_{t-1}$ so that $z_{t+1/2} - z_t = z_t - z_{t-1}$. Therefore, it is only necessary to prove the inequality for $\sum_{t=1}^T \|z_{t+1/2} - z_t\|^2$. With the assistance of the proof of Lemma 2 in Hsieh et al. (2019), $\forall t \geq 1$ and $p \in \mathcal{Z}$, we have

$$\begin{aligned} & \|z_{t+1} - p\|^2 \\ & \leq \|z_t - p\|^2 + 2\langle \eta F_t(z_{t+\frac{1}{2}}) - (z_{t-1} - z_t), z_{t+\frac{1}{2}} - z_{t+1} \rangle - 2\eta \langle F_t(z_{t+\frac{1}{2}}), z_{t+\frac{1}{2}} - p \rangle \\ & \quad - \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| z_{t+\frac{1}{2}} - z_t \right\|^2 \end{aligned} \quad (26)$$

According to the update rule of the RG algorithm, we have

$$\langle z_t - (z_{t-1} - \eta F_{t-1}(z_{t-\frac{1}{2}})), z_t - z_{t-1} \rangle \leq 0 \quad (27)$$

$$\langle z_t - (z_{t-1} - \eta F_{t-1}(z_{t-\frac{1}{2}})), z_t - z_{t+1} \rangle \leq 0 \quad (28)$$

The sum of (27) and (28) is

$$\langle z_t - z_{t-1}, z_{t+\frac{1}{2}} - z_{t+1} \rangle \leq -\langle \eta F_{t-1}(z_{t-\frac{1}{2}}), z_{t+\frac{1}{2}} - z_{t+1} \rangle \quad (29)$$

Hence, we have

$$\begin{aligned} & 2\langle \eta F_t(z_{t+\frac{1}{2}}) - (z_{t-1} - z_t), z_{t+\frac{1}{2}} - z_{t+1} \rangle \\ & \leq 2\langle \eta F_t(z_{t+\frac{1}{2}}) - \eta F_t(z_{t-\frac{1}{2}}), z_{t+\frac{1}{2}} - z_{t+1} \rangle \\ & = 2\eta \left(\langle F_{\infty}(z_{t+\frac{1}{2}}) - F_{\infty}(z_{t-\frac{1}{2}}), z_{t+\frac{1}{2}} - z_{t+1} \rangle + \langle G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}}), z_{t+\frac{1}{2}} - z_{t+1} \rangle \right) \\ & \leq 2\eta \left(L \left\| z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}} \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\| + \langle G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}}), z_{t+\frac{1}{2}} - z_{t+1} \rangle \right) \\ & \leq 2\eta \left(L \left\| z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}} \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\| + \left\| G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}}) \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\| \right) \end{aligned} \quad (30)$$

By combining (26) and (30), we have

$$\begin{aligned} & \|z_{t+1} - p\|^2 \\ & \leq \|z_t - p\|^2 + 2\eta L \left\| z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}} \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\| + 2\eta \langle G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}}), z_{t+\frac{1}{2}} - z_{t+1} \rangle \\ & \quad - 2\eta \langle F_t(z_{t+\frac{1}{2}}), z_{t+\frac{1}{2}} - p \rangle - \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| z_{t+\frac{1}{2}} - z_t \right\|^2 \\ & \leq \|z_t - p\|^2 + 2\eta L \left\| z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}} \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\| + 2\eta \left\| G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}}) \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\| \\ & \quad - 2\eta \langle F_t(z_{t+\frac{1}{2}}), z_{t+\frac{1}{2}} - p \rangle - \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| z_{t+\frac{1}{2}} - z_t \right\|^2 \end{aligned} \quad (31)$$

Since

$$\begin{aligned} & 2 \left\| z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}} \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\| \\ & \leq \frac{1}{\sqrt{2}} \|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\|^2 + \sqrt{2} \|z_{t+\frac{1}{2}} - z_{t+1}\|^2 \\ & \leq (1 + \sqrt{2}) \|z_{t+\frac{1}{2}} - z_t\|^2 + \|z_t - z_{t-\frac{1}{2}}\|^2 + \sqrt{2} \|z_{t+\frac{1}{2}} - z_{t+1}\|^2 \end{aligned} \quad (32)$$

we obtain

$$\begin{aligned}
& \|z_{t+1} - p\|^2 \\
& \leq \|z_t - p\|^2 - 2\eta \langle F_t(z_{t+\frac{1}{2}}), z_{t+\frac{1}{2}} - p \rangle + ((1 + \sqrt{2})\eta L - 1) \|z_{t+\frac{1}{2}} - z_t\|^2 \\
& \quad + \eta L \|z_t - z_{t-\frac{1}{2}}\|^2 - (1 - \sqrt{2}\eta L) \|z_{t+1} - z_{t+\frac{1}{2}}\|^2 + 2\eta \left\| G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}}) \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\|
\end{aligned} \tag{33}$$

so that

$$\begin{aligned}
& \left(1 - (1 + \sqrt{2})\eta L\right) \left\| z_{t+\frac{1}{2}} - z_t \right\|^2 \\
& \leq \|z_t - p\|^2 - \|z_{t+1} - p\|^2 - 2\eta \left\langle F(z_{t+\frac{1}{2}}), z_{t+\frac{1}{2}} - p \right\rangle + \eta L \left(\left\| z_t - z_{t-\frac{1}{2}} \right\|^2 - \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 \right) \\
& \quad + 2\eta \left\| G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}}) \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\|
\end{aligned} \tag{34}$$

if $\eta < \frac{1}{(\sqrt{2}+1)L}$.

Let $p = z^*$ in the inequality above. Since F_∞ is monotone and $z_{t+1/2} = 2z_t - z_{t-1}$, we have

$$\begin{aligned}
& -2\eta \left\langle F_t(z_{t+\frac{1}{2}}), z_{t+\frac{1}{2}} - z^* \right\rangle \\
& = -2\eta \left\langle F_t(z_{t+\frac{1}{2}}) - F_t(z^*), z_{t+\frac{1}{2}} - z^* \right\rangle - 2\eta \left\langle F_t(z^*), z_{t+\frac{1}{2}} - z^* \right\rangle \\
& \leq -2\eta \left\langle 2G_t(z_{t+\frac{1}{2}}) - G(z^*), z_{t+\frac{1}{2}} - z^* \right\rangle - 2\eta \left\langle F_\infty(z^*), z_{t+\frac{1}{2}} - z^* \right\rangle \\
& = -2\eta \left\langle 2G_t(z_{t+\frac{1}{2}}) - G_t(z^*), z_{t+\frac{1}{2}} - z^* \right\rangle + 2\eta \left\langle F_\infty(z^*), z_{t-1} - z^* \right\rangle - 4\eta \left\langle F_\infty(z^*), z_t - z^* \right\rangle \\
& \leq -2\eta \left\langle 2G_t(z_{t+\frac{1}{2}}) - G_t(z^*), z_{t+\frac{1}{2}} - z^* \right\rangle + 2\eta \left\langle F_\infty(z^*), z_{t-1} - z^* \right\rangle - 2\eta \left\langle F_\infty(z^*), z_t - z^* \right\rangle
\end{aligned} \tag{35}$$

Also, since $z_t \in \mathcal{Z}$ and z^* is a Nash equilibrium, $\forall t \geq 0$, we have $\langle F_\infty(z^*), z_t - z^* \rangle \geq 0$. By combining (34) and (35), telescoping the terms for $t = 1, 2, \dots, T$ and dividing both sides by $1 - (1 + \sqrt{2})\eta L > 0$, we obtain

$$\begin{aligned}
& \sum_{t=1}^T \left\| z_{t+\frac{1}{2}} - z_t \right\|^2 \\
& \leq \frac{\|z_1 - z^*\|^2 + \left\| z_1 - z_{\frac{1}{2}} \right\|^2 + 2\eta \langle F_\infty(z^*), z_0 - z^* \rangle}{1 - (1 + \sqrt{2})\eta L} + \frac{1}{1 - (1 + \sqrt{2})\eta L} \\
& \quad \cdot \sum_{t=1}^T \left(2\eta \left\| G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}}) \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\| - 2\eta \left\langle 2G_t(z_{t+\frac{1}{2}}) - G_t(z^*), z_{t+\frac{1}{2}} - z^* \right\rangle \right) \\
& \leq \frac{1}{1 - (1 + \sqrt{2})\eta L} \left(\|z_1 - z^*\|^2 + \left\| z_1 - z_{\frac{1}{2}} \right\|^2 + 2\eta \langle F_\infty(z^*), z_0 - z^* \rangle + 8\eta D_{\mathcal{Z}} \sum_{t=1}^T \max \|G_t\| \right. \\
& \quad \left. + 2\eta D_{\mathcal{Z}} \sum_{t=1}^T \max \|G_{t-1}\| \right)
\end{aligned} \tag{36}$$

With the denotation $H^2 := \left(\|z_1 - z^*\|^2 + \left\| z_1 - z_{\frac{1}{2}} \right\|^2 + 2\eta \langle F_\infty(z^*), z_0 - z^* \rangle + 8\eta D_{\mathcal{Z}} \sum_{t=1}^T \max \|G_t\| + 2\eta D_{\mathcal{Z}} \sum_{t=1}^T \max \|G_{t-1}\| \right)$, the proof is completed. \square

Lemma 8. In the settings of Theorem 2, $\forall T \geq 1$, there exists $C_0 \geq 0$ satisfying

$$P_{t^*, t \in [[T]]} \leq \frac{C_0}{T} \tag{37}$$

where $P_{t^*, t \in [[T]]} := \min_{t \in [[T]]} P_t$.

1026 *Proof.* We first show an upper bound for P_t .
 1027

$$\begin{aligned}
 1028 \quad P_t &= \|F_{t-1}(z_t) + c_t\|^2 + \|F_{t-1}(z_t) - F_{t-1}(z_{t-\frac{1}{2}})\|^2 \\
 1029 &= \|F_{t-1}(z_t) - F_{t-1}(z_{t-\frac{1}{2}}) + \frac{z_{t-\frac{1}{2}} - z_{t-1}}{\eta}\|^2 + \|F(z_t) - F(z_{t-\frac{1}{2}})\|^2 \\
 1030 &\leq 3 \|F_{t-1}(z_t) - F_{t-1}(z_{t-\frac{1}{2}})\|^2 + \frac{2}{\eta^2} \|z_t - z_{t-1}\|^2 \\
 1031 &= 3 \|F_\infty(z_t) - F_\infty(z_{t-\frac{1}{2}})\|^2 + 6 \langle F_\infty(z_t) - F_\infty(z_{t-\frac{1}{2}}), G_{t-1}(z_t) - G_{t-1}(z_{t-\frac{1}{2}}) \rangle \\
 1032 &\quad + 3 \|G_{t-1}(z_t) - G_{t-1}(z_{t-\frac{1}{2}})\|^2 + \frac{2}{\eta^2} \|z_t - z_{t-1}\|^2 \\
 1033 &\leq 3L^2 \|z_t - z_{t-\frac{1}{2}}\|^2 + \frac{2}{\eta^2} \|z_{t-1} - z_{t-\frac{1}{2}}\|^2 \\
 1034 &\quad + 6 \|F_\infty(z_t) - F_\infty(z_{t-\frac{1}{2}})\| \|G_{t-1}(z_t) - G_{t-1}(z_{t-\frac{1}{2}})\| + 3 \|G_{t-1}(z_t) - G_{t-1}(z_{t-\frac{1}{2}})\|^2 \\
 1035 &\leq 3L^2 \|z_t - z_{t-1} + z_{t-1} - z_{t-\frac{1}{2}}\|^2 + \frac{2}{\eta^2} \|z_t - z_{t-1}\|^2 + 6L \|z_t - z_{t-\frac{1}{2}}\| \|G_{t-1}(z_t) - G_{t-1}(z_{t-\frac{1}{2}})\| \\
 1036 &\quad + 3 \|G_{t-1}(z_t) - G_{t-1}(z_{t-\frac{1}{2}})\|^2 \\
 1037 &\leq 6L^2 \|z_{t-\frac{1}{2}} - z_{t-1}\|^2 + \left(\frac{2}{\eta^2} + 6L^2\right) \|z_t - z_{t-1}\|^2 + 12LD_{\mathcal{Z}} \max \|G_{t-1}\| + 12 \max \|G_{t-1}\|^2 \\
 1038 &\leq \frac{2 + 6\eta^2 L^2}{\eta^2} \left(\|z_{t-\frac{1}{2}} - z_{t-1}\|^2 + \|z_t - z_{t-1}\|^2 \right) + 12LD_{\mathcal{Z}} \max \|G_{t-1}\| + 12 \max \|G_{t-1}\|^2 \\
 1039 &\hspace{15em} (38)
 \end{aligned}$$

1040 Summing the above inequality of $t = 1, 2, \dots, T$, we get
 1041

$$\begin{aligned}
 1042 \quad &\sum_{t=1}^T P_t \\
 1043 &\leq \frac{2 + 6\eta^2 L^2}{\eta^2} \sum_{t=1}^T \left(\|z_{t-\frac{1}{2}} - z_{t-1}\|^2 + \|z_t - z_{t-1}\|^2 \right) + 12LD_{\mathcal{Z}} \sum_{t=1}^T \max \|G_{t-1}\| + 12 \sum_{t=1}^T \max \|G_{t-1}\|^2 \\
 1044 &= \frac{2 + 6\eta^2 L^2}{\eta^2} \left(\|z_1 - z_0\|^2 + \sum_{t=1}^{T-1} \left(\|z_{t+\frac{1}{2}} - z_t\|^2 + \|z_{t+1} - z_t\|^2 \right) \right) + 12LD_{\mathcal{Z}} \sum_{t=1}^T \max \|G_{t-1}\| \\
 1045 &\quad + 12 \sum_{t=1}^T \max \|G_{t-1}\|^2 \\
 1046 &\leq \frac{2 + 6\eta^2 L^2}{\eta^2} \left(\|z_1 - z_0\|^2 + \frac{2H^2}{1 - (1 + \sqrt{2})\eta L} \right) + 12LD_{\mathcal{Z}} \sum_{t=1}^T \max \|G_{t-1}\| + 12 \sum_{t=1}^T \max \|G_{t-1}\|^2 \\
 1047 &\leq \frac{6(1 + 3\eta^2 L^2)H^2}{\eta^2(1 - (1 + \sqrt{2})\eta L)} + 12LD_{\mathcal{Z}} \sum_{t=1}^{\infty} \max \|G_{t-1}\| + 12 \sum_{t=1}^{\infty} \max \|G_{t-1}\|^2 \\
 1048 &\hspace{15em} (39)
 \end{aligned}$$

1049 The second last inequality holds by Lemma 7. The last inequality holds since $\|z_1 - z_0\|^2 \leq \frac{4}{L^2} \|F(z_0)\|^2 \leq H^2$. Denote RHS of the last inequality in (39) as C_0 , and we have
 1050

$$1051 \quad P_{t^*, t \in [[T]]} \leq \frac{C_0}{T} \quad (40)$$

1052 This completes the proof. \square
 1053

D.3 PROOF OF LEMMA 2

If G_t is L_{G_t} -Lipschitz and $G_t(z^*) = 0$, we have

$$\begin{aligned}
& -2\eta \langle F_t(z_{t+\frac{1}{2}}), z_{t+\frac{1}{2}} - z^* \rangle \\
& \leq -2\eta \langle 2G_t(z_{t+\frac{1}{2}}) - G_t(z^*), z_{t+\frac{1}{2}} - z^* \rangle + 2\eta \langle F_\infty(z^*), z_{t-1} - z^* \rangle - 2\eta \langle F_\infty(z^*), z_t - z^* \rangle \\
& = -2\eta \langle 2G_t(z_{t+\frac{1}{2}}) - 2G_t(z^*), z_{t+\frac{1}{2}} - z^* \rangle + 2\eta \langle F_\infty(z^*), z_{t-1} - z^* \rangle - 2\eta \langle F_\infty(z^*), z_t - z^* \rangle \\
& = -4\eta \langle G_t(z_{t+\frac{1}{2}}) - G_t(z_t), z_{t+\frac{1}{2}} - z_t \rangle - 4\eta \langle G_t(z_t) - G_t(z^*), z_t - z^* \rangle \\
& \quad + 2\eta \langle F_\infty(z^*), z_{t-1} - z^* \rangle - 2\eta \langle F_\infty(z^*), z_t - z^* \rangle \\
& \leq 4\eta L_{G_t} \|z_{t+\frac{1}{2}} - z_t\|^2 + 4\eta L_{G_t} \|z_t - z^*\|^2 + 2\eta \langle F_\infty(z^*), z_{t-1} - z^* \rangle - \frac{2\eta}{1+4\eta L_{G_t}} \langle F_\infty(z^*), z_t - z^* \rangle
\end{aligned} \tag{41}$$

so that

$$\begin{aligned}
& \left(1 - (1 + \sqrt{2})\eta L - 4\eta L_{G_t}\right) \|z_{t+\frac{1}{2}} - z_t\|^2 \\
& \leq (1 + 4\eta L_{G_t}) \|z_t - z^*\|^2 - \|z_{t+1} - z^*\|^2 + 2\eta \langle F_\infty(z^*), z_{t-1} - z^* \rangle - \frac{2\eta}{1+4\eta L_{G_t}} \langle F_\infty(z^*), z_t - z^* \rangle \\
& \quad + \eta L \left(\|z_t - z_{t-\frac{1}{2}}\|^2 - \|z_{t+1} - z_{t+\frac{1}{2}}\|^2 \right) + 2\eta \langle G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}}), z_{t+\frac{1}{2}} - z_{t+1} \rangle \\
& \leq (1 + 4\eta L_{G_t}) \|z_t - z^*\|^2 - \|z_{t+1} - z^*\|^2 + 2\eta \langle F_\infty(z^*), z_{t-1} - z^* \rangle - \frac{2\eta}{1+4\eta L_{G_t}} \langle F_\infty(z^*), z_t - z^* \rangle \\
& \quad + \eta L \left(\|z_t - z_{t-\frac{1}{2}}\|^2 - \|z_{t+1} - z_{t+\frac{1}{2}}\|^2 \right) + 2\eta \langle G_t(z_{t+\frac{1}{2}}) - G_t(z^*), z_{t+\frac{1}{2}} - z_{t+1} \rangle \\
& \quad + 2\eta \langle G_{t-1}(z^*) - G_{t-1}(z_{t-\frac{1}{2}}), z_{t+\frac{1}{2}} - z_{t+1} \rangle \\
& \leq (1 + 4\eta L_{G_t}) \|z_t - z^*\|^2 - \|z_{t+1} - z^*\|^2 + 2\eta \langle F_\infty(z^*), z_{t-1} - z^* \rangle - \frac{2\eta}{1+4\eta L_{G_t}} \langle F_\infty(z^*), z_t - z^* \rangle \\
& \quad + \eta L \left(\|z_t - z_{t-\frac{1}{2}}\|^2 - \frac{1}{1+4\eta L_{G_t}} \|z_{t+1} - z_{t+\frac{1}{2}}\|^2 \right) + 2\eta \|G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}})\| \|z_{t+\frac{1}{2}} - z_{t+1}\|
\end{aligned} \tag{42}$$

Since

$$\begin{aligned}
& 2\eta \|G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}})\| \|z_{t+\frac{1}{2}} - z_{t+1}\| \\
& \leq 2\eta \|G_t(z_{t+\frac{1}{2}}) - G_t(z^*)\| \|z_{t+\frac{1}{2}} - z_{t+1}\| + 2\eta \|G_{t-1}(z^*) - G_{t-1}(z_{t-\frac{1}{2}})\| \|z_{t+\frac{1}{2}} - z_{t+1}\| \\
& \leq 2\eta L_{G_t} \|z_{t+\frac{1}{2}} - z^*\| \|z_{t+\frac{1}{2}} - z_{t+1}\| + 2\eta L_{G_{t-1}} \|z^* - z_{t-\frac{1}{2}}\| \|z_{t+\frac{1}{2}} - z_{t+1}\| \\
& \leq 2\eta L_{G_t} \|z_{t+1} - z^*\| \|z_{t+\frac{1}{2}} - z_{t+1}\| + 2\eta L_{G_t} \|z_{t+\frac{1}{2}} - z_{t+1}\|^2 + 2\eta L_{G_{t-1}} \|z^* - z_{t-\frac{1}{2}}\| \|z_{t+\frac{1}{2}} - z_{t+1}\| \\
& \leq 2\eta L_{G_t} \|z_{t+1} - z^*\| \|z_{t+\frac{1}{2}} - z_{t+1}\| + 2\eta L_{G_t} \|z_{t+\frac{1}{2}} - z_{t+1}\|^2 + 2\eta L_{G_{t-1}} \|z^* - z_t\| \|z_{t+\frac{1}{2}} - z_{t+1}\| \\
& \quad + 2\eta L_{G_{t-1}} \|z_t - z_{t-\frac{1}{2}}\| \|z_{t+\frac{1}{2}} - z_{t+1}\| \\
& \leq \eta L_{G_t} \|z_{t+1} - z^*\|^2 + (3\eta L_{G_t} + 2\eta L_{G_{t-1}}) \|z_{t+\frac{1}{2}} - z_{t+1}\|^2 + \eta L_{G_{t-1}} \|z^* - z_t\|^2 + \eta L_{G_{t-1}} \|z_t - z_{t-\frac{1}{2}}\|^2
\end{aligned} \tag{43}$$

we have

$$\begin{aligned}
& \left(1 - (1 + \sqrt{2})\eta L - 4\eta L_{G_t}\right) \|z_{t+\frac{1}{2}} - z_t\|^2 \\
& \leq (1 + 4\eta L_{G_t}) \|z_t - z^*\|^2 - \|z_{t+1} - z^*\|^2 + 2\eta \langle F_\infty(z^*), z_{t-1} - z^* \rangle - \frac{2\eta}{1+4\eta L_{G_t}} \langle F_\infty(z^*), z_t - z^* \rangle
\end{aligned}$$

$$\begin{aligned}
& + \eta L \left(\left\| z_t - z_{t-\frac{1}{2}} \right\|^2 - \frac{1}{1+4\eta L_{G_t}} \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 \right) + 2\eta \left\| G_t(z_{t+\frac{1}{2}}) - G_{t-1}(z_{t-\frac{1}{2}}) \right\| \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\| \\
& \leq (1+4\eta L_{G_t}) \left\| z_t - z^* \right\|^2 - \left\| z_{t+1} - z^* \right\|^2 + 2\eta \langle F_\infty(z^*), z_{t-1} - z^* \rangle - \frac{2\eta}{1+4\eta L_{G_t}} \langle F_\infty(z^*), z_t - z^* \rangle \\
& + \eta L \left(\left\| z_t - z_{t-\frac{1}{2}} \right\|^2 - \frac{1}{1+4\eta L_{G_t}} \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 \right) + \eta L_{G_t} \left\| z_{t+1} - z^* \right\|^2 + (3\eta L_{G_t} + 2\eta L_{G_{t-1}}) \left\| z_{t+\frac{1}{2}} - z_{t+1} \right\|^2 \\
& + \eta L_{G_{t-1}} \left\| z^* - z_t \right\|^2 + \eta L_{G_{t-1}} \left\| z_t - z_{t-\frac{1}{2}} \right\|^2 \\
& = (1+4\eta L_{G_t} + \eta L_{G_{t-1}}) \left\| z_t - z^* \right\|^2 - (1-\eta L_{G_t}) \left\| z_{t+1} - z^* \right\|^2 + 2\eta \langle F_\infty(z^*), z_{t-1} - z^* \rangle \\
& - \frac{2\eta}{1+4\eta L_{G_t}} \langle F_\infty(z^*), z_t - z^* \rangle + \eta (L + L_{G_{t-1}}) \left\| z_t - z_{t-\frac{1}{2}} \right\|^2 \\
& - \left(\frac{\eta L}{1+4\eta L_{G_t}} - 3\eta L_{G_t} - 2\eta L_{G_{t-1}} \right) \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 \tag{44}
\end{aligned}$$

Since

$$\begin{aligned}
& \frac{1+4\eta L_{G_t} + \eta L_{G_{t-1}}}{1-\eta L_{G_t}} - \frac{\eta(L+L_{G_{t-1}})}{\frac{\eta L}{1+4\eta L_{G_t}} - 3\eta L_{G_t} - 2\eta L_{G_{t-1}}} \\
& = (\eta L - (1+4\eta L_{G_t})(3\eta L_{G_t} + 2\eta L_{G_{t-1}}))(4\eta L_{G_t} + \eta L_{G_{t-1}} + 1) - \eta(1-\eta L_{G_t})(1+4\eta L_{G_t})(L+L_{G_{t-1}}) \\
& \quad (1-\eta L_{G_t})^{-1} (\eta L(1+4\eta L_{G_t}) (-3\eta L_{G_t} - 2\eta L_{G_{t-1}}))^{-1} \\
& = (-48\eta^3 L_{G_t}^3 - 40\eta^3 L_{G_t}^2 L_{G_{t-1}} + (4L\eta^3 - 24\eta^2)L_{G_t}^2 - 8\eta^3 L_{G_t} L_{G_{t-1}}^2 - 22\eta^2 L_{G_t} L_{G_{t-1}} + (L\eta^2 - 3\eta)L_{G_t} \\
& \quad - 2\eta^2 L_{G_{t-1}}^2 + (L\eta^2 - 3\eta)L_{G_{t-1}})(1-\eta L_{G_t})^{-1} (\eta L(1+4\eta L_{G_t}) (-3\eta L_{G_t} - 2\eta L_{G_{t-1}}))^{-1} \tag{45}
\end{aligned}$$

and $\frac{1+4\eta L_{G_t} + \eta L_{G_{t-1}}}{1-\eta L_{G_t}} - (1+4\eta L_{G_t}) = \frac{4L_{G_t}^2 \eta^2 + L_{G_t} \eta + L_{G_{t-1}} \eta}{1-\eta L_{G_t}}$, there exists T_N and $\epsilon > 0$ so that

$\forall t > T_N$, it holds that $L_{G_t}, L_{G_{t-1}} < \epsilon$, $1 - (1 + \sqrt{2})\eta L - 4\eta L_{G_t} > 0$, $\frac{1+4\eta L_{G_t} + \eta L_{G_{t-1}}}{1-\eta L_{G_t}} <$

$\frac{\eta(L+L_{G_{t-1}})}{\frac{\eta L}{1+4\eta L_{G_t}} - 3\eta L_{G_t} - 2\eta L_{G_{t-1}}}$ with $L\eta^2 - 3\eta < 0$ and $\frac{1+4\eta L_{G_t} + \eta L_{G_{t-1}}}{1-\eta L_{G_t}} > 1 + 4\eta L_{G_t}$. Since

$\frac{\eta(L+L_{G_{t-1}})}{\frac{\eta L}{1+4\eta L_{G_t}} - 3\eta L_{G_t} - 2\eta L_{G_{t-1}}} = 1 + \frac{4\eta L + 6}{L}\epsilon + O(\epsilon^2)$, we have

$$\sum_{t=T_N+1}^{\infty} \left(\frac{4\eta L + 6}{L} L_{G_{t-1}} + O(L_{G_{t-1}}^2) \right) \leq \sum_{t=T_N+1}^{\infty} \left(\frac{4\eta L + 6}{L} L_{G_{t-1}} + O(L_{G_{t-1}}^2) \right) < \infty \tag{46}$$

Hence,

$$\prod_{t=T_N+1}^{\infty} \frac{\eta(L+L_{G_{t-1}})}{\frac{\eta L}{1+4\eta L_{G_t}} - 3\eta L_{G_t} - 2\eta L_{G_{t-1}}} < \infty \tag{47}$$

so that

$$E := \prod_{t=2}^{\infty} \frac{\eta(L+L_{G_{t-1}})}{\frac{\eta L}{1+4\eta L_{G_t}} - 3\eta L_{G_t} - 2\eta L_{G_{t-1}}} < \infty \tag{48}$$

Hence, $\forall t > T_N$,

$$\begin{aligned}
& \left\| z_{t+1} - z^* \right\|^2 \\
& \leq \left\| z_{t+1} - z^* \right\|^2 + \frac{2\eta}{(1+4\eta L_{G_t})(1-\eta L_{G_t})} \langle F_\infty(z^*), z_t - z^* \rangle \\
& \quad + \frac{\eta L + (3\eta L_{G_t} + 2\eta L_{G_{t-1}})(1+4\eta L_{G_t})}{(1+4\eta L_{G_t})(1-\eta L_{G_t})} \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 \\
& \leq E \left\| z_2 - z^* \right\|^2 + \frac{2E\eta}{(1+4\eta L_{G_1})(1-\eta L_{G_1})} \langle F_\infty(z^*), z_1 - z^* \rangle \\
& \quad + \frac{E\eta L + E(3\eta L_{G_1} + 2\eta L_{G_0})(1+4\eta L_{G_1})}{(1+4\eta L_{G_1})(1-\eta L_{G_1})} \left\| z_2 - z_{\frac{3}{2}} \right\|^2 \tag{49}
\end{aligned}$$

Define $C_1 = \max \left\{ E \|z_2 - z^*\|^2 + \frac{2E\eta}{(1+4\eta L_{G_1})(1-\eta L_{G_1})} \langle F_\infty(z^*), z_1 - z^* \rangle + \frac{E\eta L + E(3\eta L_{G_1} + 2\eta L_{G_0})(1+4\eta L_{G_1})}{(1+4\eta L_{G_1})(1-\eta L_{G_1})} \cdot \|z_2 - z_{\frac{3}{2}}\|^2, \max_{t=0,1,2,\dots,T_N} \|z_T - z^*\|^2 \right\}$. We have that $\|z_t - z^*\| < \sqrt{C_1}$, $t \in \mathbb{N}$. Hence, there exists \mathcal{Z}_L satisfying $z^*, z_t \in \mathcal{Z}_L$, $t \in \mathbb{N}$ and we can take $D_{\mathcal{Z}} = D_{\mathcal{Z}_L} = 2\sqrt{C_1}$.

D.4 FINAL PROOF

With the definition of P_T , we have

$$\begin{aligned} & r^{\tan}(z_T)^2 \\ & \leq \|F_{T-1}(z_T) - G_{T-1}(z_T) + c_T\|^2 \\ & \leq \|F_\infty(z_T) + c_T\|^2 \\ & \leq \|F_{T-1}(z_T) + c_T\|^2 + 2 \langle F_{T-1}(z_T) + c_T, -G_{T-1}(z_T) \rangle + \| -G_{T-1}(z_T) \|^2 \\ & \leq \|F_{T-1}(z_T) + c_T\|^2 + 2 \|F_{T-1}(z_T) + c_T\| \|G_{T-1}(z_T)\| + \|G_{T-1}(z_T)\|^2 + \left\| F_{T-1}(z_T) - F_{T-1}(z_{T-\frac{1}{2}}) \right\|^2 \\ & \leq P_T + 2 \|F_{T-1}(z_T)\| \max \|G_{T-1}\| + \max \|G_{T-1}\|^2 \end{aligned} \quad (50)$$

Under the BAP assumption, $\max \|G_{T-1}\| = O\left(\frac{1}{T}\right)$. Hence, according to Lemma 6, we have

$$\begin{aligned} P_{\lceil \frac{T}{2} \rceil} & \leq P_1 + \left(\frac{2}{\eta} + 8L\right) D_{\mathcal{Z}} \sum_{t=1}^{\lceil \frac{T}{2} \rceil - 1} \max \|G_t\| + \frac{2}{\eta} D_{\mathcal{Z}} \sum_{t=1}^{\lceil \frac{T}{2} \rceil - 1} \max \|G_{t-1}\| + 8 \sum_{t=1}^{\lceil \frac{T}{2} \rceil - 1} \max \|G_t\|^2 \\ & \leq P_1 + \left(\frac{2}{\eta} + 8L\right) D_{\mathcal{Z}} \sum_{t=1}^{\lceil \frac{T}{2} \rceil - 1} \max \|G_t\| + \frac{2}{\eta} D_{\mathcal{Z}} \sum_{t=1}^{\infty} \max \|G_{t-1}\| + 8 \sum_{t=1}^{\infty} \max \|G_t\|^2 =: P_{hs} \end{aligned} \quad (51)$$

$$\begin{aligned} P_T & \leq P_{t^*, t \in [\lceil \frac{T}{2} \rceil, T]} + \left(\frac{2}{\eta} + 8L\right) D_{\mathcal{Z}} \sum_{t=\lceil \frac{T}{2} \rceil - 1}^T \max \|G_t\| + \frac{2}{\eta} D_{\mathcal{Z}} \sum_{t=\lceil \frac{T}{2} \rceil - 1}^T \max \|G_{t-1}\| + 8 \sum_{t=\lceil \frac{T}{2} \rceil - 1}^T \max \|G_t\|^2 \\ & \leq P_{t^*, t \in [\lceil \frac{T}{2} \rceil, T]} + \left(\frac{2}{\eta} + 8L\right) D_{\mathcal{Z}} \sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_t\| + \frac{2}{\eta} D_{\mathcal{Z}} \sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_{t-1}\| + 8 \sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_t\|^2 \end{aligned} \quad (52)$$

Hence,

$$\begin{aligned} & r^{\tan}(z_T)^2 \\ & \leq P_{t^*, t \in [\lceil \frac{T}{2} \rceil, T]} + \left(\frac{2}{\eta} + 8L\right) D_{\mathcal{Z}} \sum_{t=t^*, t \in [\lceil \frac{T}{2} \rceil, T]} \max \|G_t\| + \frac{2}{\eta} D_{\mathcal{Z}} \sum_{t=t^*, t \in [\lceil \frac{T}{2} \rceil, T]} \max \|G_{t-1}\| \\ & \quad + 8 \sum_{t=t^*, t \in [\lceil \frac{T}{2} \rceil, T]} \max \|G_t\|^2 + 2 \|F_{T-1}(z_T)\| \max \|G_{T-1}\| + \max \|G_{T-1}\|^2 \\ & \leq P_{t^*, t \in [\lceil \frac{T}{2} \rceil, T]} + \left(\frac{2}{\eta} + 8L\right) D_{\mathcal{Z}} \sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_t\| + \frac{2}{\eta} D_{\mathcal{Z}} \sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_{t-1}\| + 8 \sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_t\|^2 \\ & \quad + 2 \|F_{T-1}(z_T)\| \max \|G_{T-1}\| + \max \|G_{T-1}\|^2 \end{aligned} \quad (53)$$

If $\sum_{t=T}^{\infty} \max \|G_t\| = O\left(\sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_t\|\right)$, $\sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_t\| = O\left(\frac{1}{T}\right)$. Since $\max \|G_{T-1}\|^2 = O\left(\sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_t\|^2\right)$, $\max \|G_{T-1}\| = O\left(\sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_t\|\right)$,

1242 $\sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_t\|^2 = O\left(\sum_{t=\lceil \frac{T}{2} \rceil - 1}^{\infty} \max \|G_t\|\right)$ and $P_{t^*, t \in [\lceil \frac{T}{2} \rceil, T]} = O\left(\frac{1}{T}\right)$, we have

1243
1244
1245
$$r^{\tan}(z_T)^2 = \max \left\{ O\left(\frac{1}{T}\right), O\left(\sum_{t=T/2}^{\infty} \max \|G_t\|\right) \right\} \quad (54)$$

1246
1247
1248 Hence,

1249
1250
$$r^{\tan}(z_T) = \max \left\{ O\left(\frac{1}{\sqrt{T}}\right), O\left(\sqrt{\sum_{t=T/2}^{\infty} \max \|G_t\|}\right) \right\} \quad (55)$$

1251 1252 1253 E MISSING PROOFS OF THEOREM 3

1254
1255 In periodic games, each period as a whole is time-invariant (Franke & Selgrade (2003), Section 3).
1256 We take advantage of this property in our analysis.

1257
1258 With the ARG algorithm, we have

1259
1260
$$\begin{bmatrix} x_0 \\ y_0 \\ x_t \\ y_t \\ x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} I_3 & O_{3 \times 3} & O_{3 \times 3} \\ O_{3 \times 3} & O_{3 \times 3} & I_3 \\ Q_t & R_t & S_t \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ x_{t-1} \\ y_{t-1} \\ x_t \\ y_t \end{bmatrix} \quad (56)$$

1261
1262
1263
1264
1265 where

1266
1267
$$Q_t = \begin{bmatrix} \frac{1}{t+1} I & \frac{\eta A_t}{t(t+1)} \\ -\frac{\eta A_t^T}{t(t+1)} & \frac{1}{t+1} I \end{bmatrix} \quad (57)$$

1268
1269
$$R_t = \begin{bmatrix} 0 & \eta \left(1 - \frac{1}{t}\right) A_t \\ -\eta \left(1 - \frac{1}{t}\right) A_t^T & 0 \end{bmatrix} \quad (58)$$

1270
1271
1272
$$S_t = \begin{bmatrix} \left(1 - \frac{1}{t+1}\right) I & -\eta \left(2 - \frac{1}{t+1}\right) A_t \\ \eta \left(2 - \frac{1}{t+1}\right) A_t^T & \left(1 - \frac{1}{t+1}\right) I \end{bmatrix} \quad (59)$$

1273
1274
1275 Denote $\begin{bmatrix} I_3 & O_{3 \times 3} & O_{3 \times 3} \\ O_{3 \times 3} & O_{3 \times 3} & I_3 \\ Q_t & R_t & S_t \end{bmatrix}$ as P_t . We have

1276
1277
$$\lambda I - \lim_{t \rightarrow \infty} P_{t+1} P_t = \lambda I - \begin{bmatrix} I_3 & O_{3 \times 3} & O_{3 \times 3} \\ O_{3 \times 3} & O_{3 \times 3} & I_3 \\ O_{3 \times 3} & R'_t & S'_t \end{bmatrix} \quad (60)$$

1278
1279
1280
1281
1282 where

1283
1284
$$R'_t = \begin{bmatrix} 0 & \eta A_t \\ -\eta A_t^T & 0 \end{bmatrix} = \begin{cases} \begin{bmatrix} 0 & \eta & -\eta \\ -\eta & 0 & 0 \\ \eta & 0 & 0 \end{bmatrix}, & t \text{ is odd} \\ \begin{bmatrix} 0 & -\eta & \eta \\ \eta & 0 & 0 \\ -\eta & 0 & 0 \end{bmatrix}, & t \text{ is even} \end{cases} \quad (61)$$

1285
1286
1287
1288
1289
1290
1291
1292
$$S'_t = \begin{bmatrix} I & -2\eta A_t \\ 2\eta A_t^T & I \end{bmatrix} = \begin{cases} \begin{bmatrix} 1 & -2\eta & 2\eta \\ 2\eta & 1 & 0 \\ -2\eta & 0 & 1 \end{bmatrix}, & t \text{ is odd} \\ \begin{bmatrix} 1 & 2\eta & -2\eta \\ -2\eta & 1 & 0 \\ 2\eta & 0 & 1 \end{bmatrix}, & t \text{ is even} \end{cases} \quad (62)$$

1293
1294
1295

Hence,

$$\det \left(\lambda I - \lim_{t \rightarrow \infty} P_{t+1} P_t \right) = \lambda(\lambda - 1)^4 (8\eta^2 \lambda - 2\eta^2 - \lambda^2 + \lambda)^2 \quad (63)$$

which means that $\frac{\sqrt{64\eta^4 + 8\eta^2 + 1} + 8\eta^2 + 1}{2}$ is an eigenvalue of $\lim_{t \rightarrow \infty} P_{t+1} P_t$ and it is greater than 1.

Since an eigenvalue of $\lim_{t \rightarrow \infty} P_{t+1} P_t$ is $\frac{\sqrt{64\eta^4 + 8\eta^2 + 1} + 8\eta^2 + 1}{2} > 1$, denote this eigenvalue as ρ_∞ and the corresponding eigenvector of $\lim_{t \rightarrow \infty} P_{t+1} P_t^2$ as v . Then, we obtain

$$\left(\lim_{t \rightarrow \infty} \prod_{i=0}^k P_{t+2i+1} P_{t+2i} \right) v = \left(\prod_{i=0}^k \lim_{t \rightarrow \infty} P_{t+1} P_t \right) v = \rho_\infty^{k+1} v \quad (64)$$

which means that $\forall \epsilon > 0$, there exists $M > 0$ so that $\forall t > M$, there exists

$$\left\| \left(\prod_{i=0}^k P_{t+2i+1} P_{t+2i} \right) v \right\| - \|\rho_\infty^{k+1} v\| < \epsilon \quad (65)$$

Since the action set is \mathbb{R}^3 , we have

$$r^{tan}(z) = F(z) = \begin{bmatrix} Ay \\ -A^T x \end{bmatrix} \quad (66)$$

Hence,

$$\|r^{tan}(z)\| = \left\| \begin{bmatrix} Ay \\ -A^T x \end{bmatrix} \right\| = \|z\| \quad (67)$$

Since $\|\rho_\infty^k v\| = \|\rho_\infty\|^k \|v\|$ grows exponentially, $\|z_{t+2k}\|$ does not converge. With the ARG algorithm, $\forall v$, corresponding z_0 and $z_{1/2}$ exist for $z_t = v$. This means that the tangent residual with the ARG algorithm diverges exponentially if such z_0 and $z_{1/2}$ are the initial actions. This completes the proof.

F MISSING PROOFS OF THEOREM 4

To prove Theorem 4, we apply a potential function argument. We first show the potential function is approximately non-increasing and then prove that it is upper bounded by a term independent of T . As the potential function at step t is also at least $\Omega(t^2) r^{tan}(z_t)^2$, we conclude that the ARG algorithm has a $O(\frac{1}{T})$ convergence rate.

F.1 POTENTIAL FUNCTION

With the update rules of the ARG algorithm, i.e., $z_0, z_{1/2} \in \mathbb{R}^n$ being initial points, for $t \geq 1$,

$$\begin{aligned} z_{t+\frac{1}{2}} &= 2z_t - z_{t-1} + \frac{1}{t+1}(z_0 - z_t) - \frac{1}{t}(z_0 - z_{t-1}) \\ z_{t+1} &= \Pi_{\mathcal{Z}} \left[z_t - \eta F_t(z_{t+\frac{1}{2}}) + \frac{1}{t+1}(z_0 - z_t) \right] \end{aligned} \quad (68)$$

Since $N_{\mathcal{Z}}$ is the normal cone of a closed convex set \mathcal{Z} and $\Pi_{\mathcal{Z}}$ is the projection to set \mathcal{Z} , if we apply the ARG algorithm to solve time-varying game problems, the algorithm calculates gradient function $F_t(z)$ once and a projection to \mathcal{Z} once per iteration. Next, we specify the potential function. Define

$$c_{t+1} := \frac{z_t - \eta F_t(z_{t+\frac{1}{2}}) + \frac{1}{t+1}(z_0 - z_t) - z_{t+1}}{\eta}, \quad \forall t \geq 0 \quad (69)$$

By update rule we have $c_t \in N_{\mathcal{Z}}$ for all $t \geq 1$. The potential function at $t \geq 1$ is defined as

$$V_t := \frac{t(t+1)}{2} \|\eta F_{t-1}(z_t) + \eta c_t\|^2 + t \langle \eta F_{t-1}(z_t) + \eta c_t, z_t - z_0 \rangle + \frac{t(t+1)}{2} \left\| \eta F_{t-1}(z_t) - \eta F_{t-1}(z_{t-\frac{1}{2}}) \right\|^2 \quad (70)$$

F.2 APPROXIMATE MONOTONICITY OF THE POTENTIAL FUNCTION

Lemma 9. $\forall L > 0$ and $\rho \geq -\frac{1}{60L}$, There exists $\eta > 0$ such that

$$\frac{1}{2} - \left(12 - \frac{4\rho}{\eta}\right)\eta^2 L^2 + \frac{2\rho}{\eta} \geq 0 \quad (71)$$

Moreover, every $\eta > 0$ satisfies (71) also satisfies $\frac{\rho}{\eta} \geq -\frac{1}{4}$.

Proof. Equation (71) means that

$$\rho > \frac{\eta L(24\eta^2 L^2 - 1)}{4 + 8\eta^2 L^2} \cdot \frac{1}{L} \quad (72)$$

Let $x = \eta L$ and $h(x) = \frac{x(24x^2 - 1)}{4 + 8x^2}$. Since $h(\frac{1}{12}) = -\frac{5}{292} < -\frac{1}{60}$, there exists $\eta = \frac{1}{12L}$ satisfying (71).

Besides, with $\eta L > 0$ and (71), we obtain

$$\frac{\rho}{\eta} \geq -\frac{1 - 72\eta^2 L^2}{4 + 8\eta^2 L^2} \geq -\frac{1}{4} \quad (73)$$

□

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

Lemma 10. With the settings in Theorem 4, $\forall t \geq 1$, we have

$$\begin{aligned} & V_t - V_{t+1} \\ & \leq -\frac{1}{8}\|\eta F_t(z_{t+1}) + \eta c_{t+1}\|^2 + t(t+1)\langle \eta G_t(z_{t+1}) - \eta G_{t-1}(z_t), z_{t+1} - z_t \rangle \\ & \quad + \frac{t(t+1)}{4p} \left(1 - \frac{\rho}{3\eta}\right) \left\| -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\|^2 \\ & \quad + \frac{t(t+1)}{2p} \left(1 - \frac{\rho}{3\eta}\right) \left\langle \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}), -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\rangle \end{aligned} \quad (74)$$

Proof. We show that $V_t - V_{t+1}$ plus a few non-positive terms is still $\geq -\frac{1}{8}\|\eta F(z_{t+1}) + \eta c_{t+1}\|^2 + t(t+1)\langle \eta G_t(z_{t+1})$

$$\begin{aligned} & - \eta G_{t-1}(z_t), z_{t+1} - z_t \rangle + \frac{t(t+1)}{4p} \left(1 - \frac{\rho}{3\eta}\right) \left\| -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\|^2 + \\ & \frac{t(t+1)}{2p} \left(1 - \frac{\rho}{3\eta}\right) \left\langle \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}), \right. \\ & \left. -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\rangle, \text{ where } p > 0. \end{aligned}$$

With the settings of F_∞ in Theorem 4, we have

$$\langle \eta F_\infty(z_{t+1}) + \eta c_{t+1} - \eta F_\infty(z_t) - \eta c_t, z_{t+1} - z_t \rangle - \frac{\rho}{\eta} \|\eta F_\infty(z_{t+1}) + \eta c_{t+1} - \eta F_\infty(z_t) - \eta c_t\|^2 \geq 0 \quad (75)$$

which means

$$\begin{aligned} & \langle \eta F_t(z_{t+1}) + \eta c_{t+1} - \eta F_{t-1}(z_t) - \eta c_t, z_{t+1} - z_t \rangle - \frac{\rho}{\eta} \|\eta F_\infty(z_{t+1}) + \eta c_{t+1} - \eta F_\infty(z_t) - \eta c_t\|^2 \\ & - \langle \eta G_t(z_{t+1}) - \eta G_{t-1}(z_t), z_{t+1} - z_t \rangle \\ & \geq 0 \end{aligned} \quad (76)$$

Since F_∞ is L -Lipschitz, we have

$$\eta^2 L^2 \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| \eta F_\infty(z_{t+1}) - \eta F_\infty(z_{t+\frac{1}{2}}) \right\|^2 \geq 0 \quad (77)$$

1404 which means

$$1405 \quad \eta^2 L^2 \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 - \left\| -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\|^2$$

$$1406 \quad - 2 \left\langle \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}), -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\rangle \quad (78)$$

$$1407 \quad \geq 0$$

1409 By multiplying the above inequality by $1 - \frac{\rho}{3\eta} > 0$, we get

$$1410 \quad p \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 - \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 + \left(\left(1 - \frac{\rho}{3\eta} \right) \eta^2 L^2 - p \right) \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2$$

$$1411 \quad + \frac{\rho}{3\eta} \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 - \left(1 - \frac{\rho}{3\eta} \right) \left\| -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\|^2$$

$$1412 \quad - 2 \left(1 - \frac{\rho}{3\eta} \right) \left\langle \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}), -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\rangle$$

$$1413 \quad \geq 0 \quad (79)$$

1414 We show the following two equations involving $z_{t+1/2}$ and z_{t+1} with the update rules of the ARG algorithm.

$$1415 \quad z_{t+\frac{1}{2}} = 2z_t - z_{t-1} + \frac{1}{t+1}(z_0 - z_t) - \frac{1}{t}(z_0 - z_{t-1})$$

$$1416 \quad = z_t + (z_t - z_{t-1}) + \frac{1}{t+1}(z_0 - z_t) - \frac{1}{t}(z_0 - z_{t-1}) \quad (80)$$

$$1417 \quad = z_t - \eta F_{t-1}(z_{t-\frac{1}{2}}) - \eta c_t + \frac{1}{t+1}(z_0 - z_t)$$

$$1418 \quad z_{t+1} = z_t - \eta F_t(z_{t+\frac{1}{2}}) - \eta c_{t+1} + \frac{1}{t+1}(z_0 - z_t) \quad (81)$$

1419 Hence, we have

$$1420 \quad z_{t+1} - z_{t+\frac{1}{2}} = \eta F_{t-1}(z_{t-\frac{1}{2}}) + \eta c_t - \eta F_t(z_{t+\frac{1}{2}}) - \eta c_{t+1} \quad (82)$$

1421 Next, we simplify

$$1422 \quad V_t - V_{t+1} - t(t+1) \cdot \text{LHS of Inequality (76)} - \frac{t(t+1)}{4p} \cdot \text{LHS of Inequality (79)} \quad (83)$$

1423 using the second identity in Lemma 4: replace x_0 with z_0 ; for $k \in [[4]]$, replace x_k with $z_{t-1+\frac{k}{2}}$ and

1424 replace y_k with $\eta F(z_{t-1+\frac{k}{2}})$; replace u_2 with ηc_t ; replace u_4 with ηc_{t+1} ; replace k with t ; replace p

1425 with q . Note that $x_3 = x_2 - y_1 - u_2 + \frac{1}{k+1}(x_0 - x_2)$ and $x_4 = x_2 - y_3 - u_4 + \frac{1}{k+1}(x_0 - x_2)$ hold

1426 due to the above equivalent formations of $z_{t+1/2}$ and z_{t+1} .

$$1427 \quad V_t - V_{t+1} - t(t+1) \cdot \text{LHS of Inequality (76)} - \frac{t(t+1)}{4p} \cdot \text{LHS of Inequality (79)}$$

$$1428 \quad = \frac{t(t+1)}{4} \left\| \eta c_{t+1} - \eta c_t + \eta F_{t-1}(z_{t-\frac{1}{2}}) - 2\eta F_{t-1}(z_t) + \eta F_t(z_{t+\frac{1}{2}}) \right\|^2$$

$$1429 \quad + \left(\frac{(1-4p)t - 4p}{4p}(t+1) \right) \left\| \eta F_t(z_{t+\frac{1}{2}}) - \eta F_t(z_{t+1}) \right\|^2 + (t+1) \left\langle \eta F_t(z_{t+\frac{1}{2}}) - \eta F_t(z_{t+1}), \eta F_t(z_{t+1}) + \eta c_{t+1} \right\rangle$$

$$1430 \quad + t(t+1) \frac{\rho}{\eta} \left\| \eta F_\infty(z_{t+1}) + \eta c_{t+1} - \eta F_\infty(z_t) - \eta c_t \right\|^2$$

$$1431 \quad - \frac{t(t+1)}{4p} \left(\left(\left(1 - \frac{\rho}{3\eta} \right) \eta^2 L^2 - p \right) \left\| z_{t+1} - z_{t+\frac{1}{2}} \right\|^2 + \frac{\rho}{3\eta} \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 \right)$$

$$1432 \quad + t(t+1) \left\langle \eta G_t(z_{t+1}) - \eta G_{t-1}(z_t), z_{t+1} - z_t \right\rangle + \frac{t(t+1)}{4p} \left(1 - \frac{\rho}{3\eta} \right) \left\| -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\|^2$$

$$1433 \quad + \frac{t(t+1)}{2p} \left(1 - \frac{\rho}{3\eta} \right) \left\langle \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}), -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\rangle \quad (84)$$

1458 Since $\|a\|^2 + \langle a, b \rangle = \|a + \frac{b}{2}\|^2 - \frac{\|b\|^2}{4}$, with $p = \frac{1}{24}$, we have

$$\begin{aligned}
1461 & \left(\frac{(1-4p)t-4p}{4p}(t+1) \right) \left\| \eta F_t(z_{t+\frac{1}{2}}) - \eta F_t(z_{t+1}) \right\|^2 + (t+1) \langle \eta F_t(z_{t+\frac{1}{2}}) - \eta F_t(z_{t+1}), \eta F_t(z_{t+1}) + \eta c_{t+1} \rangle \\
1462 & \\
1463 & = \left\| \sqrt{\frac{(1-4p)t-4p}{4p}(t+1)} (\eta F_t(z_{t+\frac{1}{2}}) - \eta F_t(z_{t+1})) + \sqrt{\frac{p(t+1)}{(1-4p)t-4p}} (\eta F_t(z_{t+1}) + \eta c_{t+1}) \right\|^2 \\
1464 & \\
1465 & \quad - \frac{p(t+1)}{(1-4p)t-4p} \|\eta F_t(z_{t+1}) + \eta c_{t+1}\|^2 \\
1466 & \\
1467 & \geq -\frac{p(t+1)}{(1-8p)t} \|\eta F_t(z_{t+1}) + \eta c_{t+1}\|^2 \quad (t \geq 1) \\
1468 & \\
1469 & \geq -\frac{2p}{1-8p} \|\eta F_t(z_{t+1}) + \eta c_{t+1}\|^2 \quad \left(\frac{t+1}{t} \leq 2\right) \\
1470 & \\
1471 & = -\frac{1}{8} \|\eta F_t(z_{t+1}) + \eta c_{t+1}\|^2 \\
1472 & \\
1473 & \\
1474 & \\
1475 & \tag{85}
\end{aligned}$$

1476 Then, we get

$$\begin{aligned}
1478 & \frac{4}{t(t+1)} \left(\frac{t(t+1)}{4} \left\| \eta c_{t+1} - \eta c_t + \eta F_{t-1}(z_{t-\frac{1}{2}}) - 2\eta F_{t-1}(z_t) + \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 \right. \\
1479 & \\
1480 & \quad \left. + t(t+1) \frac{\rho}{\eta} \|\eta F_\infty(z_{t+1}) + \eta c_{t+1} - \eta F_\infty(z_t) - \eta c_t\|^2 \right. \\
1481 & \\
1482 & \quad \left. - \frac{t(t+1)}{4p} \left(\left(\left(1 - \frac{\rho}{3\eta}\right) \eta^2 L^2 - p \right) \|z_{t+1} - z_{t+\frac{1}{2}}\|^2 + \frac{\rho}{3\eta} \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 \right) \right) \\
1483 & \\
1484 & = \left\| \eta c_{t+1} - \eta c_t + \eta F_{t-1}(z_{t-\frac{1}{2}}) - 2\eta F_{t-1}(z_t) + \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 + \left(1 - \left(24 - \frac{8\rho}{\eta}\right) \eta^2 L^2\right) \|z_{t+1} - z_{t+\frac{1}{2}}\|^2 \\
1485 & \\
1486 & \quad + \frac{4\rho}{\eta} \|\eta F_\infty(z_{t+1}) + \eta c_{t+1} - \eta F_\infty(z_t) - \eta c_t\|^2 - \frac{8\rho}{\eta} \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 \\
1487 & \\
1488 & \\
1489 & \tag{86}
\end{aligned}$$

1490 When there exists

$$\begin{aligned}
1492 & B_1 = \eta c_{t+1} - \eta c_t + \eta F_{t-1}(z_{t-\frac{1}{2}}) - 2\eta F_{t-1}(z_t) + \eta F_t(z_{t+\frac{1}{2}}) \\
1493 & \\
1494 & B_2 = z_{t+1} - z_{t+\frac{1}{2}} = \eta F_{t-1}(z_{t-\frac{1}{2}}) + \eta c_t - \eta F_t(z_{t+\frac{1}{2}}) - \eta c_{t+1} \\
1495 & \\
1496 & B_3 = \eta F_t(z_{t+1}) + \eta c_{t+1} - \eta F_{t-1}(z_t) - \eta c_t \\
1497 & \\
1498 & B_4 = \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \\
1499 & \\
1500 & \tag{87}
\end{aligned}$$

1499 we have

$$1501 \quad B_1 - B_2 = 2\eta c_{t+1} - 2\eta c_t - 2\eta F_{t-1}(z_t) + 2\eta F_t(z_{t+\frac{1}{2}}) = 2(B_3 - B_4) \tag{88}$$

1504 Note that ρ is non-positive and we have

$$\begin{aligned}
1506 & \frac{4}{t(t+1)} \left(\frac{t(t+1)}{4} \left\| \eta c_{t+1} - \eta c_t + \eta F_{t-1}(z_{t-\frac{1}{2}}) - 2\eta F_{t-1}(z_t) + \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 \right. \\
1507 & \\
1508 & \quad \left. + t(t+1) \frac{\rho}{\eta} \|\eta F_\infty(z_{t+1}) + \eta c_{t+1} - \eta F_\infty(z_t) - \eta c_t\|^2 \right. \\
1509 & \\
1510 & \quad \left. - \frac{t(t+1)}{4p} \left(\left(\left(1 - \frac{\rho}{3\eta}\right) \eta^2 L^2 - p \right) \|z_{t+1} - z_{t+\frac{1}{2}}\|^2 + \frac{\rho}{3\eta} \left\| \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}) \right\|^2 \right) \right) \\
1511 &
\end{aligned}$$

$$\begin{aligned}
&= \|B_1\|^2 + \left(1 - \left(24 - \frac{8\rho}{\eta}\right) \eta^2 L^2\right) \|B_2\|^2 + \frac{\rho}{\eta} \|2B_3 - 2G_t(z_{t+1}) + 2G_{t-1}(z_t)\|^2 - \frac{2\rho}{\eta} \|2B_4\|^2 \\
&\geq \left(\frac{1}{2} - \left(12 - \frac{4\rho}{\eta}\right) \eta^2 L^2\right) \|B_1 - B_2\|^2 + \frac{\rho}{\eta} \|2B_3 - 2G_t(z_{t+1}) + 2G_{t-1}(z_t)\|^2 - \frac{2\rho}{\eta} \|2B_4\|^2 \\
&\quad (\|a\|^2 + \|b\|^2 \geq \frac{1}{2} \|a - b\|^2 \text{ and } \left(24 - \frac{8\rho}{\eta}\right) \eta^2 L^2 \geq 0) \\
&\geq \left(\frac{1}{2} - \left(12 - \frac{4\rho}{\eta}\right) \eta^2 L^2\right) \|B_1 - B_2\|^2 + \frac{2\rho}{\eta} \|2B_3 - 2B_4 - 2G_t(z_{t+1}) + 2G_{t-1}(z_t)\|^2 \\
&\quad (-\|a\|^2 + 2\|b\|^2 \geq -2\|a - b\|^2 \text{ and } -\frac{\rho}{\eta} \geq 0) \\
&= \left(\frac{1}{2} - \left(12 - \frac{4\rho}{\eta}\right) \eta^2 L^2 + \frac{2\rho}{\eta}\right) \|B_1 - B_2\|^2 + \frac{4\rho}{\eta} \langle 2(B_3 - B_4), -2G_t(z_{t+1}) + 2G_{t-1}(z_t) \rangle \\
&\quad + \frac{2\rho}{\eta} \| -2G_t(z_{t+1}) + 2G_{t-1}(z_t) \|^2 \\
&\geq 0 \\
&\quad (\eta \text{ is chosen as shown in Lemma 9})
\end{aligned} \tag{89}$$

Hence, we have

$$\begin{aligned}
&V_t - V_{t+1} \\
&\geq -\frac{1}{8} \|\eta F_t(z_{t+1}) + \eta c_{t+1}\|^2 + t(t+1) \langle \eta G_t(z_{t+1}) - \eta G_{t-1}(z_t), z_{t+1} - z_t \rangle \\
&\quad + \frac{t(t+1)}{4p} \left(1 - \frac{\rho}{3\eta}\right) \left\| -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\|^2 \\
&\quad + \frac{t(t+1)}{2p} \left(1 - \frac{\rho}{3\eta}\right) \left\langle \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}), -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\rangle
\end{aligned} \tag{90}$$

□

F.3 FINAL PROOF

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

Lemma 11. *With the settings of Theorem 4, $\forall t \geq 1$, we have*

$$\begin{aligned}
&\frac{t(t+\frac{1}{2})}{4} \|\eta F_{t-1}(z_t) + \eta c_t\|^2 \\
&\leq V_t + \|z^* - z_0\|^2 + \frac{t}{4} \|\eta G_{t-1}(z_t)\|^2 + \frac{t}{2} \langle \eta F_{t-1}(z_t) + \eta c_t, -\eta G_{t-1}(z_t) \rangle - t \langle \eta G_{t-1}(z_t), z_t - z^* \rangle
\end{aligned} \tag{91}$$

Proof. Since $F_\infty(z^*) + N_{\mathcal{Z}}(z^*) = 0$, by the settings in Theorem 4 and Lemma 9, we have

$$\langle \eta F_\infty(z_t) + \eta c_t, z_t - z^* \rangle \geq \frac{\rho}{\eta} \|\eta F_\infty(z_t) + \eta c_t\|^2 \geq -\frac{1}{4} \|\eta F_\infty(z_t) + \eta c_t\|^2 \tag{92}$$

With definition of V_t in Equation (70), $\forall t \geq 1$, we obtain

$$\begin{aligned}
&V_t \\
&= \frac{t(t+1)}{2} \|\eta F_{t-1}(z_t) + \eta c_t\|^2 + t \langle \eta F_{t-1}(z_t) + \eta c_t, z_t - z_0 \rangle + \frac{t(t+1)}{2} \left\| \eta F_{t-1}(z_t) - \eta F_{t-1}(z_{t-\frac{1}{2}}) \right\|^2 \\
&\geq \frac{t(t+1)}{2} \|\eta F_{t-1}(z_t) + \eta c_t\|^2 + t \langle \eta F_{t-1}(z_t) + \eta c_t, z_t - z^* \rangle + t \langle \eta F_{t-1}(z_t) + \eta c_t, z^* - z_0 \rangle
\end{aligned}$$

$$\begin{aligned}
&\geq \frac{t(t+1)}{2} \|\eta F_{t-1}(z_t) + \eta c_t\|^2 - \frac{t}{4} \|\eta F_\infty(z_t) + \eta c_t\|^2 + t \langle \eta G_{t-1}(z_t), z_t - z^* \rangle + t \langle \eta F_{t-1}(z_t) + \eta c_t, z^* - z_0 \rangle \\
&\geq \frac{t(t+1)}{2} \|\eta F_{t-1}(z_t) + \eta c_t\|^2 - \frac{t}{4} \|\eta F_{t-1}(z_t) + \eta c_t\|^2 - \frac{t}{4} \|\eta G_{t-1}(z_t)\|^2 - \frac{t}{2} \langle \eta F_{t-1}(z_t) + \eta c_t, -\eta G_{t-1}(z_t) \rangle \\
&\quad - \frac{t(t+\frac{1}{2})}{4} \|\eta F_{t-1}(z_t) + \eta c_t\|^2 - \frac{t}{t+\frac{1}{2}} \|z^* - z_0\|^2 + t \langle \eta G_{t-1}(z_t), z_t - z^* \rangle \\
&\geq \frac{t(t+\frac{1}{2})}{4} \|\eta F_{t-1}(z_t) + \eta c_t\|^2 - \|z^* - z_0\|^2 - \frac{t}{4} \|\eta G_{t-1}(z_t)\|^2 - \frac{t}{2} \langle \eta F_{t-1}(z_t) + \eta c_t, -\eta G_{t-1}(z_t) \rangle \\
&\quad + t \langle \eta G_{t-1}(z_t), z_t - z^* \rangle
\end{aligned} \tag{93}$$

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

Proof of Theorem 4. There exists $H > 0$ satisfying

$$\|\eta F_0(z_1) + \eta c_1\|^2 \leq H^2 \tag{94}$$

So the theorem holds for $T = 1$. $\forall T \geq 2$, with Lemma 11, we have

$$\begin{aligned}
&\frac{T(T+\frac{1}{2})}{4} \|\eta F(z_T) + \eta c_T\|^2 \\
&\leq V_T + \|z_0 - z^*\|^2 - \left(-\frac{t}{4} \|\eta G_{t-1}(z_t)\|^2 - \frac{t}{2} \langle \eta F_{t-1}(z_t) + \eta c_t, -\eta G_{t-1}(z_t) \rangle + t \langle \eta G_{t-1}(z_t), z_t - z^* \rangle \right) \\
&\leq V_T + \|z_0 - z^*\|^2 + \frac{t}{4} \|\eta G_{t-1}(z_t)\|^2 + \frac{t}{2} \langle \eta F_{t-1}(z_t) + \eta c_t, -\eta G_{t-1}(z_t) \rangle - t \langle \eta G_{t-1}(z_t), z_t - z^* \rangle \\
&\leq V_1 + \|z_0 - z^*\|^2 + \frac{t}{4} \|\eta G_{t-1}(z_t)\|^2 + \frac{t}{2} \langle \eta F_{t-1}(z_t) + \eta c_t, -\eta G_{t-1}(z_t) \rangle - t \langle \eta G_{t-1}(z_t), z_t - z^* \rangle \\
&\quad + \frac{1}{8} \sum_{t=2}^T \|\eta F(z_t) + \eta c_t\|^2 - \sum_{t=1}^{T-1} t(t+1) \langle \eta G_t(z_{t+1}) - \eta G_{t-1}(z_t), z_{t+1} - z_t \rangle \\
&\quad - \sum_{t=1}^{T-1} \frac{t(t+1)}{4p} \left(1 - \frac{\rho}{3\eta} \right) \left\| -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\|^2 \\
&\quad - \sum_{t=1}^{T-1} \frac{t(t+1)}{2p} \left(1 - \frac{\rho}{3\eta} \right) \left\langle \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}), -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\rangle
\end{aligned} \tag{95}$$

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

$$\begin{aligned}
&\frac{T^2}{4} \|\eta F_{T-1}(z_T) + \eta c_T\|^2 \\
&\leq \frac{1}{8} \sum_{t=2}^{T-1} \|\eta F_{t-1}(z_t) + \eta c_t\|^2 + \|z_0 - z^*\|^2 + \frac{t}{4} \|\eta G_{t-1}(z_t)\|^2 + \frac{t}{2} \langle \eta F_{t-1}(z_t) + \eta c_t, -\eta G_{t-1}(z_t) \rangle \\
&\quad - t \langle \eta G_{t-1}(z_t), z_t - z^* \rangle - \sum_{t=1}^{T-1} t(t+1) \langle \eta G_t(z_{t+1}) - \eta G_{t-1}(z_t), z_{t+1} - z_t \rangle \\
&\quad - \sum_{t=1}^{T-1} \frac{t(t+1)}{4p} \left(1 - \frac{\rho}{3\eta} \right) \left\| -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\|^2 \\
&\quad - \sum_{t=1}^{T-1} \frac{t(t+1)}{2p} \left(1 - \frac{\rho}{3\eta} \right) \left\langle \eta F_t(z_{t+1}) - \eta F_t(z_{t+\frac{1}{2}}), -\eta G_t(z_{t+1}) + \eta G_t(z_{t+\frac{1}{2}}) \right\rangle
\end{aligned} \tag{96}$$

Since Assumption 1 holds, we have $\max \|G_t\| = O(\frac{1}{t})$.

We show that there exists $D_1, D_2 > 0$ satisfying

$$S_T = \|\eta F_{T-1}(z_T) + \eta c_T\| \leq \frac{D_1}{T} + D_2 \sum_{t=2}^T \max \|G_{t-1}\| \quad (97)$$

as follows. There exists $D_1, D_2 \geq 0$ satisfying

$$S_2 \leq \frac{D_1}{2} + D_2 \max \|G_1\| \quad (98)$$

$$S_T \leq \frac{1}{2T^2} \sum_{t=2}^{T-1} S_{t-1} + D_2 \sum_{t=2}^{T-1} \max \|G_{t-1}\| \quad (99)$$

If $D_1 > 0$ and $D_1 \geq D_2 \sum_{t=2}^{\infty} \max \|G_{t-1}\|$, with the assumption that Inequality (97) holds true when T is replaced by $2, 3, \dots, T-1$, we have

$$\begin{aligned} S_T &\leq \frac{D_1}{2T^2} \sum_{t=2}^{T-1} \frac{1}{t-1} + \frac{D_2}{2T^2} \sum_{t=2}^{T-1} \sum_{s=2}^t \max \|G_{s-1}\| + D_2 \sum_{t=2}^{T-1} \max \|G_{t-1}\| \\ &\leq \frac{D_1(T-2)}{2T^2} + \frac{D_2(T-2)}{2T^2} \sum_{s=2}^{T-1} \max \|G_{s-1}\| + D_2 \sum_{t=2}^{T-1} \max \|G_{t-1}\| \\ &\leq \frac{D_1(T-2)}{T^2} + D_2 \sum_{t=2}^{T-1} \max \|G_{t-1}\| \\ &\leq \frac{D_1}{T} + D_2 \sum_{t=2}^{T-1} \max \|G_{t-1}\| \end{aligned} \quad (100)$$

Since there exists $D_1 \geq D_2 \sum_{t=2}^{\infty} \max \|G_{t-1}\|$ satisfying (98), Inequality (97) holds true.

Fianlly, we show that there exists $D_3, D_4 > 0$ satisfying

$$S_T = \|\eta F_{T-1}(z_T) + \eta c_T\| \leq \frac{D_3}{T} \quad (101)$$

as follows. With the assumption that Inequality (101) holds true when T is replaced by $2, 3, \dots, T-1$, we have

$$\begin{aligned} S_T &\leq \frac{D_1}{2T^2} \sum_{t=2}^{T-1} \frac{1}{t-1} + \frac{D_2}{2T^2} \sum_{t=2}^{T-1} \sum_{s=2}^t \max \|G_{s-1}\| + D_2 \sum_{t=2}^{T-1} \max \|G_{t-1}\| \\ &\leq \frac{D_1(T-2)}{2T^2} + \frac{D_2(T-2)}{2T^2} \sum_{s=2}^{T-1} \max \|G_{s-1}\| + D_2 \sum_{t=2}^{T-1} \max \|G_{t-1}\| \\ &\leq \frac{D_1(T-2)}{T^2} + D_2 \sum_{t=2}^{T-1} \max \|G_{t-1}\| \\ &\leq \frac{D_1}{T} + D_2 \sum_{t=2}^{T-1} \max \|G_{t-1}\| \end{aligned} \quad (102)$$

If $\sum_{t=0}^{\infty} t^2 \|G_t\| < \infty$, there exists $D_4 > 0$ satisfying

$$S_T \leq \frac{1}{2T^2} \left(\sum_{t=2}^{T-1} S_{t-1} + D_4 \right) \quad (103)$$

If Inequality (101) holds when T is replaced by $2, 3, \dots, T-1$, then

$$S_T \leq \frac{D_3}{2T^2} \left(\sum_{t=2}^{T-1} \frac{1}{t-1} + D_4 \right) \leq \frac{D_3}{2T} + \frac{D_3}{2T^2 D_4} = \frac{D_3}{T} \left(\frac{1}{2} + \frac{1}{2D_4 T} \right) \leq \frac{D_3}{T} \quad (104)$$

When $D_3 = 2S_2$, Inequality (101) holds with T replaced by 2. Hence, Inequality (101) holds with $D_3 = 2S_2$.

This completes the proof. \square

1674 G MISSING PROOFS OF THEOREM 5
1675

1676 We apply the modified OG algorithm to the two-player bilinear game (1) shown in Theorem 1.
1677

1678 For the general two-player bilinear game

$$\begin{aligned}
 1679 \quad z^{(1)} &= x \in \mathbb{R}^m, \\
 1680 \quad z^{(2)} &= y \in \mathbb{R}^n, \\
 1681 \quad f^{(1)}(z) &= x^T A_t y, \\
 1682 \quad f^{(2)}(z) &= -x^T A_t y, \\
 1683 \quad \mathcal{Z} &= \mathbb{R}^m \times \mathbb{R}^n
 \end{aligned} \tag{105}$$

1684 with the RG algorithm, there exists

$$\begin{bmatrix} x_t \\ y_t \\ x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ 0 & \eta A_t & I & -2\eta A_t \\ -\eta A_t^T & 0 & 2\eta A_t^T & I \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ x_t \\ y_t \end{bmatrix} \tag{106}$$

1685 With the modified OG algorithm, there exists

$$\begin{bmatrix} x_{t+\frac{1}{2}} \\ y_{t+\frac{1}{2}} \\ x_{t+\frac{3}{2}} \\ y_{t+\frac{3}{2}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ 0 & \eta A_{t+1} & I & -2\eta A_{t+1} \\ -\eta A_{t+1}^T & 0 & 2\eta A_{t+1}^T & I \end{bmatrix} \begin{bmatrix} x_{t-\frac{1}{2}} \\ y_{t-\frac{1}{2}} \\ x_{t+\frac{1}{2}} \\ y_{t+\frac{1}{2}} \end{bmatrix} \tag{107}$$

1686 The equivalence of the RG and the variant of the OG algorithm algorithms is proven.
1687

1688 In the settings provided in Theorem 5, with the modified OG algorithm, as the relation shown in
1689 (107), we have

$$\begin{bmatrix} x_{t+\frac{1}{2}} \\ y_{t+\frac{1}{2}} \\ x_{t+\frac{3}{2}} \\ y_{t+\frac{3}{2}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ 0 & \eta A_{t+1} & I & -2\eta A_{t+1} \\ -\eta A_{t+1}^T & 0 & 2\eta A_{t+1}^T & I \end{bmatrix} \begin{bmatrix} x_{t-\frac{1}{2}} \\ y_{t-\frac{1}{2}} \\ x_{t+\frac{1}{2}} \\ y_{t+\frac{1}{2}} \end{bmatrix} \tag{108}$$

1690 Hence, with the notation in the proof of Theorem 1, we have

$$\begin{bmatrix} x_{t-\frac{1}{2}} \\ y_{t-\frac{1}{2}} \\ x_{t+\frac{1}{2}} \\ y_{t+\frac{1}{2}} \end{bmatrix} = P_t \begin{bmatrix} x_{t-\frac{3}{2}} \\ y_{t-\frac{3}{2}} \\ x_{t-\frac{1}{2}} \\ y_{t-\frac{1}{2}} \end{bmatrix} \tag{109}$$

1691 Under the modified OG algorithm, there exists $z_0, z_{-1/2}$ satisfying

$$\begin{bmatrix} z_{-\frac{1}{2}} \\ z_{\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} x_{-\frac{1}{2}} \\ y_{-\frac{1}{2}} \\ x_{\frac{1}{2}} \\ y_{\frac{1}{2}} \end{bmatrix} = c \tag{110}$$

1692 where c is defined in Theorem 1. Without loss of generality, assume t is odd. Then, we have

$$\begin{bmatrix} z_{t+\frac{1}{2}} \\ z_{t+\frac{3}{2}} \end{bmatrix} = P_{t+1} P_t \begin{bmatrix} z_{t-\frac{3}{2}} \\ z_{t-\frac{1}{2}} \end{bmatrix} \tag{111}$$

1693 Since $P_{t+1} P_t$ has an eigenvalue $\lambda_{1,2} = \frac{8\eta^2+1+\sqrt{64\eta^4+8\eta^2+1}}{2} > 1$, there exists a corresponding
1694 eigenvector

$$\begin{bmatrix} \frac{16\eta^2-2\sqrt{64\eta^4+8\eta^2+1}+4}{2(8\eta^2+1)} \\ \frac{\sqrt{64\eta^4+8\eta^2+1}-1}{4\eta(8\eta^2+1)} \\ -\frac{\sqrt{64\eta^4+8\eta^2+1}-1}{4\eta(8\eta^2+1)} \\ 1 \\ 0 \\ 0 \end{bmatrix} \tag{112}$$

1728 If c is equal to this eigenvector, we have $\|z_{1/2}\| = 1$ and $\|z_{-1/2}\| = \frac{\sqrt{8\eta^2 - \sqrt{64\eta^4 + 8\eta^2 + 1} + 1}}{2\eta}$.
 1729
 1730 Since $8\eta^2 - \sqrt{64\eta^4 + 8\eta^2 + 1} + 1 = 8\eta^2 + \frac{1}{2} - \sqrt{64\eta^4 + 8\eta^2 + 1} + \frac{1}{2} = \sqrt{64\eta^4 + 8\eta^2 + \frac{1}{4}} -$
 1731 $\sqrt{64\eta^4 + 8\eta^2 + 1} + \frac{1}{2} < \frac{1}{2}, \forall \eta > 0, \eta \|z_{-1/2}\| < \|z_{1/2}\|$. With such z_0 and $z_{-1/2}$, $\|z_{t-1/2}\|$ and
 1732 $\|z_{t+1/2}\|$ diverges exponentially. Since

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_{t+\frac{1}{2}} \\ y_{t+\frac{1}{2}} \end{bmatrix} + \eta \begin{bmatrix} A_t y_{t-\frac{1}{2}} \\ -A_t^T x_{t-\frac{1}{2}} \end{bmatrix} \quad (113)$$

1733 we have

$$\|z_t\| = \|z_{t+\frac{1}{2}} + F_t(z_{t-\frac{1}{2}})\| \geq \| \|z_{t+\frac{1}{2}}\| - \|F_t(z_{t-\frac{1}{2}})\| \| = \| \|z_{t+\frac{1}{2}}\| - \eta \|z_{t-\frac{1}{2}}\| \| \quad (114)$$

1734 Since $\|z_{1/2}\| - \eta \|z_{-1/2}\| \neq 0$ in the c equal to (112), for $t = 2k - 1$, there exists

$$\|r^{tan}(z_{2k-1})\| = \|F_{2k-1}(z_{2k-1})\| = \|z_{2k-1}\| \geq \| \|z_{2k+\frac{1}{2}}\| - \eta \|z_{2k-\frac{1}{2}}\| \| = \lambda_{1,2}^k \| \|z_{\frac{1}{2}}\| - \eta \|z_{-\frac{1}{2}}\| \| > 0 \quad (115)$$

1744 which means $\|r^{tan}(z_{2k-1})\|$ also diverges at an exponential rate. This completes the proof.

1746 H MISSING PROOFS OF THEOREM 6

1747 We only need to prove the extended result. With the update rules of the OG algorithm, we have the
 1748 following identity (Hsieh et al. (2019)): $\forall p \in \mathcal{Z}$,

$$\|z_{t+1} - p\|^2 = \|z_t - p\|^2 + \|z_{t+1} - z_{t+\frac{1}{2}}\|^2 - \|z_{t+\frac{1}{2}} - z_t\|^2 + 2 \left\langle z_t - \eta F_t(z_{t-\frac{1}{2}}) - z_{t+\frac{1}{2}} + \eta F_t(z_{t+\frac{1}{2}}), p - z_{t+\frac{1}{2}} \right\rangle \quad (116)$$

1754 Since $z_{t+1/2} = \Pi_{\mathcal{Z}}(z_t - \eta F_t(z_{t-1/2}))$, we have

$$\frac{z_t - \eta F_t(z_{t-\frac{1}{2}}) - z_{t+\frac{1}{2}}}{\eta} \in N_{\mathcal{Z}}(z_{t+\frac{1}{2}}) \quad (117)$$

1758 Then

$$\frac{z_t - z_{t+1}}{\eta} = \frac{z_t - \eta F_t(z_{t-\frac{1}{2}}) - z_{t+\frac{1}{2}}}{\eta} + F_t(z_{t+\frac{1}{2}}) \in F_t(z_{t+\frac{1}{2}}) + N_{\mathcal{Z}}(z_{t+\frac{1}{2}}) \quad (118)$$

1762 which means

$$r_{F, \mathcal{Z}}^{tan}(z_{t+\frac{1}{2}}) \leq \frac{\|z_{t+1} - z_t\|}{\eta} \quad (119)$$

1765 For $p = z^*$, we have

$$2 \left\langle z_t - \eta F_t(z_{t-\frac{1}{2}}) - z_{t+\frac{1}{2}} + \eta F_t(z_{t+\frac{1}{2}}), z^* - z_{t+\frac{1}{2}} \right\rangle = 2\eta \left\langle \frac{z_t - z_{t+1}}{\eta}, z^* - z_{t+\frac{1}{2}} \right\rangle \leq -\frac{2\rho}{\eta} \|z_t - z_{t+1}\|^2 \quad (120)$$

1769 Define $c = \frac{1 - 2L_t^2\eta^2 - 2L_{t-1}^2\eta^2}{2(2L_t^2\eta^2 - 2L_{t-1}^2\eta^2 + 1)} > 0$. We have identity

$$(1 - 2c)\eta^2 L_{t-1}^2 = \frac{1}{2} - c - (1 + 2c)\eta^2 L_t^2 \quad (121)$$

1774 Combining (116) and (120), since $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$, we have

$$\begin{aligned} & \|z_{t+1} - z^*\|^2 \\ & \leq \|z_t - z^*\|^2 + \|z_{t+1} - z_{t+\frac{1}{2}}\|^2 - \|z_{t+\frac{1}{2}} - z_t\|^2 + c\|z_t - z_{t+1}\|^2 - \left(c + \frac{2\rho}{\eta}\right) \|z_t - z_{t+1}\|^2 \\ & \leq \|z_t - z^*\|^2 + (1 + 2c) \|z_{t+1} - z_{t+\frac{1}{2}}\|^2 - (1 - 2c) \|z_{t+\frac{1}{2}} - z_t\|^2 - \left(c + \frac{2\rho}{\eta}\right) \|z_t - z_{t+1}\|^2 \end{aligned} \quad (122)$$

Using the update rules of OG and L_t -Lipschitzness of F_t , we have that for any $t \geq 0$,

$$\|z_{t+1} - z_{t+\frac{1}{2}}\|^2 = \|\eta F_t(z_{t-\frac{1}{2}}) - \eta F_t(z_{t+\frac{1}{2}})\|^2 \leq \eta^2 L_t^2 \|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\|^2 \quad (123)$$

Moreover, using $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$ and (123), we have that for any $t \geq 1$,

$$\|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\|^2 \leq 2\|z_{t+\frac{1}{2}} - z_t\|^2 + 2\|z_t - z_{t-\frac{1}{2}}\|^2 \leq 2\|z_{t+\frac{1}{2}} - z_t\|^2 + 2\eta^2 L_{t-1}^2 \|z_{t-\frac{1}{2}} - z_{t-\frac{3}{2}}\|^2 \quad (124)$$

which means

$$\|z_{t+\frac{1}{2}} - z_t\|^2 \geq \frac{1}{2} \|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\|^2 - \eta^2 L_{t-1}^2 \|z_{t-\frac{1}{2}} - z_{t-\frac{3}{2}}\|^2 \quad (125)$$

Hence,

$$\begin{aligned} & \|z_{t+1} - z^*\|^2 \\ & \leq \|z_t - z^*\|^2 + (1+2c)\|z_{t+1} - z_{t+\frac{1}{2}}\|^2 - (1-2c)\|z_{t+\frac{1}{2}} - z_t\|^2 - \left(c + \frac{2\rho}{\eta}\right) \|z_t - z_{t+1}\|^2 \\ & \leq \|z_t - z^*\|^2 + (1-2c)\eta^2 L_{t-1}^2 \|z_{t-\frac{1}{2}} - z_{t-\frac{3}{2}}\|^2 - \left(\frac{1}{2} - c - (1+2c)\eta^2 L_t^2\right) \|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\|^2 \\ & \quad - \left(c + \frac{2\rho}{\eta}\right) \|z_t - z_{t+1}\|^2 \\ & = \|z_t - z^*\|^2 + \frac{4L_t^2 L_{t-1}^2 \eta^4}{2L_t^2 \eta^2 - 2L_{t-1}^2 \eta^2 + 1} \left(\|z_{t-\frac{1}{2}} - z_{t-\frac{3}{2}}\|^2 - \|z_{t+\frac{1}{2}} - z_{t-\frac{1}{2}}\|^2\right) - \left(c + \frac{2\rho}{\eta}\right) \|z_t - z_{t+1}\|^2 \end{aligned} \quad (126)$$

By summing the above inequality for $t = 1, 2, \dots, T$, with $c = \frac{1-2L_t^2 \eta^2 - 2L_{t-1}^2 \eta^2}{2(2L_t^2 \eta^2 - 2L_{t-1}^2 \eta^2 + 1)}$, considering L_s -Lipschitz means L_g -Lipschitz if $L_s < L_g$, L_t with $t = 1, 2, \dots, T$ can be replaced by L_{\max} . Hence, we get

$$\left(\frac{1}{2} - 2\eta^2 L_{\max}^2 + \frac{2\rho}{\eta}\right) \sum_{t=1}^T \|z_t - z_{t+1}\|^2 \leq \|z_1 - z^*\|^2 + \frac{1}{4} \|z_{\frac{1}{2}} - z_{-\frac{1}{2}}\|^2 \quad (127)$$

which means

$$\min_{t \in \llbracket T \rrbracket} \left(\frac{\|z_{t+1} - z_t\|}{\eta}\right)^2 = O\left(\frac{1}{T}\right) \quad (128)$$

Hence, we have

$$\min_{t \in \llbracket T \rrbracket} r_{F, \mathcal{Z}}^{\tan}(z_{t+\frac{1}{2}}) \leq \min_{t \in \llbracket T \rrbracket} \frac{\|z_{t+1} - z_t\|}{\eta} = O\left(\frac{1}{\sqrt{T}}\right) \quad (129)$$

This completes the proof.

I NUMERICAL EXAMPLES

Several visual representations are provided to illustrate and verify the theoretical results, i.e., Theorem 1, Theorem 2, Theorem 3, Theorem 4, Theorem 5 and Theorem 6. These examples are based on examples of bilinear games shown in Feng et al. (2023). Running all the programs for the following experiments costs no more than dozens of minutes in total with the corresponding code and the computer resources of an ordinary laptop.

I.1 EXPERIMENTS ON THEOREM 1

We verify Theorem 1 by the example

$$\begin{aligned} z^{(1)} &= x \in \mathbb{R}, z^{(2)} = y \in \mathbb{R}^2 \\ f^{(1)} &= x^T A_t y, f^{(2)} = -x^T A_t y \\ \mathcal{Z} &= \mathbb{R} \times \mathbb{R}^2 \end{aligned} \quad (130)$$

1836 where

$$1837 \quad A_t = \begin{cases} [1 & -1], & t \text{ is odd} \\ [-1 & 1], & t \text{ is even} \end{cases} \quad (131)$$

1838 The step size η is chosen to be 0.005. The initial points are chosen to be $x_0 = 1, x_{-1} = 4$ and
 1839 $y_0 = [2, 3], y_{-1} = [5, 6]$. The experimental results are presented in Figure 1, where we can see
 1840 the RG algorithm makes $r^{tan}(z_t)$ diverge. This result supports the exponential divergence result in
 1841 Theorem 1.
 1842
 1843

1844 I.2 EXPERIMENTS ON THEOREM 2

1845 We verify Theorem 2 by the example

$$1846 \quad z^{(1)} = x \in [-100, 100]^2, z^{(2)} = y \in [-100, 100]^2$$

$$1847 \quad f^{(1)}(z) = (x^T - [2 \ 3]) \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} \left(y - \begin{bmatrix} 5 \\ 3 \end{bmatrix} \right) + 100t^i(-25 \sin(z_1) + 30 \sin(z_2) + 60 \sin(z_3) + 50 \sin(z_4))$$

$$1848 \quad f^{(2)}(z) = -(x^T - [2 \ 3]) \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} \left(y - \begin{bmatrix} 5 \\ 3 \end{bmatrix} \right) + 100t^i(-25 \sin(z_1) + 30 \sin(z_2) + 60 \sin(z_3) + 50 \sin(z_4))$$

$$1849 \quad \mathcal{Z} = [-100, 100]^4 \quad (132)$$

1850 where $i = -1.1, -2, -9$ in three cases. The step size is chosen to be 0.005. The initial points are
 1851 chosen to be $x_0 = [15, 40], x_{-1} = [44, 35]$ and $y_0 = [3, 51], y_{-1} = [3, 21]$. The experimental results
 1852 are presented in Figure 2, all of the three dynamics make $r^{tan}(z_t)$ converge to 0, thus support the
 1853 convergence result in Theorem 2.
 1854
 1855

1856 I.3 EXPERIMENTS ON THEOREM 3

1857 We verify Theorem 3 by the example

$$1858 \quad z^{(1)} = x \in \mathbb{R}, z^{(2)} = y \in \mathbb{R}^2$$

$$1859 \quad f^{(1)}(z) = x^T A_t y, f^{(2)}(z) = -x^T A_t y \quad (133)$$

$$1860 \quad \mathcal{Z} = \mathbb{R} \times \mathbb{R}^2$$

1861 where

$$1862 \quad A_t = \begin{cases} [1 & -1], & t \text{ is odd} \\ [-1 & 1], & t \text{ is even} \end{cases} \quad (134)$$

1863 The step size η is chosen to be 0.005. The initial points are chosen to be $x_0 = 1, x_{-\frac{1}{2}} = 4$ and
 1864 $y_0 = [2, 3], y_{-\frac{1}{2}} = [5, 6]$. The experimental results are presented in Figure 3, where we can see the
 1865 ARG algorithm makes $r^{tan}(z_t)$ diverge. This result supports the exponentially divergence result in
 1866 Theorem 3.
 1867
 1868

1869 I.4 EXPERIMENTS ON THEOREM 4

1870 We verify Theorem 4 by the example

$$1871 \quad z^{(1)} = x \in [-100, 100]^2, z^{(2)} = y \in [-100, 100]^2$$

$$1872 \quad f^{(1)}(z) = (x^T - [2 \ 3]) \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} \left(y - \begin{bmatrix} 5 \\ 3 \end{bmatrix} \right) + 100t^i(-25 \sin(z_1) + 30 \sin(z_2) + 60 \sin(z_3) + 50 \sin(z_4))$$

$$1873 \quad f^{(2)}(z) = -(x^T - [2 \ 3]) \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} \left(y - \begin{bmatrix} 5 \\ 3 \end{bmatrix} \right) + 100t^i(-25 \sin(z_1) + 30 \sin(z_2) + 60 \sin(z_3) + 50 \sin(z_4))$$

$$1874 \quad \mathcal{Z} = [-100, 100]^4 \quad (135)$$

1875 where $i = -1.1, -2, -9$ in three cases. The step size is chosen to be 0.005. The initial points are
 1876 chosen to be $x_0 = [15, 40], x_{-1} = [44, 35]$ and $y_0 = [3, 51], y_{-1} = [3, 21]$. The experimental results
 1877 are presented in Figure 4, all of the three dynamics make $r^{tan}(z_t)$ converge to 0, thus support the
 1878 convergence result in Theorem 4.
 1879
 1880

1890 I.5 EXPERIMENTS ON THEOREM 5

1891 We verify Theorem 5 by the example

$$\begin{aligned}
 1892 \quad & z^{(1)} = x \in \mathbb{R}, z^{(2)} = y \in \mathbb{R}^2 \\
 1893 \quad & f^{(1)}(z) = x^T A_t y, f^{(2)}(z) = -x^T A_t y \\
 1894 \quad & \mathcal{Z} = \mathbb{R} \times \mathbb{R}^2
 \end{aligned} \tag{136}$$

1895 where

$$A_t = \begin{cases} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, t \text{ is odd} \\ \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}, t \text{ is even} \end{cases} \tag{137}$$

1896 The step size η is chosen to be 0.005. The initial points are chosen to be $x_0 = 1, x_{\frac{1}{2}} = 4$ and
 1897 $y_0 = [2, 3], y_{\frac{1}{2}} = [5, 6]$. The experimental results are presented in Figure 5, where we can see the
 1898 variant of the OG algorithm makes $r_i^{tan}(z_t)$ diverge. This result supports the exponential divergence
 1899 result in Theorem 5.

1900 I.6 EXPERIMENTS ON THEOREM 6

1901 We verify Theorem 6 by examples with $k \in \mathbb{N}^*$:

$$\begin{aligned}
 1902 \quad & z^{(1)} = x \in \mathbb{R}^2, z^{(2)} = y \in \mathbb{R}^2 \\
 1903 \quad & f^{(1)}(z) = x^T A_t y, f^{(2)}(z) = -x^T A_t y \\
 1904 \quad & \mathcal{Z} = \mathbb{R}^2 \times \mathbb{R}^2
 \end{aligned} \tag{138}$$

1905 where

$$A_t = \begin{cases} \begin{bmatrix} 2 & -2 \\ -4 & 1 \end{bmatrix}, t = 3k \\ \begin{bmatrix} 1 & 4 \\ 2 & 2 \end{bmatrix}, t = 3k + 1 \\ \begin{bmatrix} 2 & 2 \\ 6 & 1 \end{bmatrix}, t = 3k + 2 \end{cases} \tag{139}$$

1906 The step size η is chosen to be 0.005. The initial points are chosen to be $z_0 = [4, 5, 2, 3], z_{\frac{1}{2}} =$
 1907 $[5, 6, 3, 1]$. The experimental results are presented in Figure 6, where we can see the OG algorithm
 1908 makes $r_i^{tan}(z_t)$ converge, $i = 0, 1, 2$. This result supports the convergence result in Theorem 6.

1944
 1945
 1946
 1947
 1948
 1949
 1950
 1951
 1952
 1953
 1954
 1955
 1956
 1957
 1958
 1959
 1960
 1961
 1962
 1963
 1964
 1965
 1966
 1967
 1968
 1969
 1970
 1971
 1972
 1973
 1974
 1975
 1976
 1977
 1978
 1979
 1980
 1981
 1982
 1983
 1984
 1985
 1986
 1987
 1988
 1989
 1990
 1991
 1992
 1993
 1994
 1995
 1996
 1997

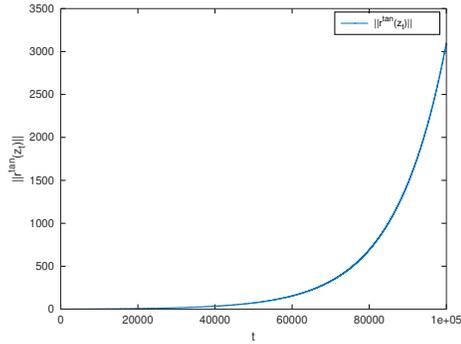


Figure 1: $\|r^{tan}(z_t)\|$ in a periodic game for the RG algorithm

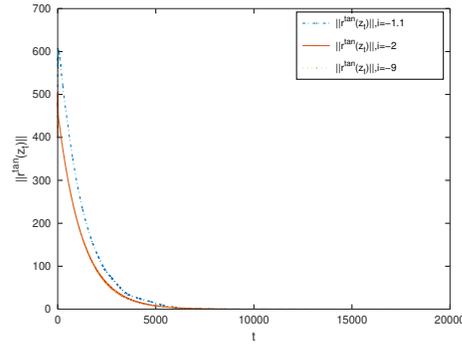


Figure 2: $\|r^{tan}(z_t)\|$ in a convergent perturbed game for the RG algorithm

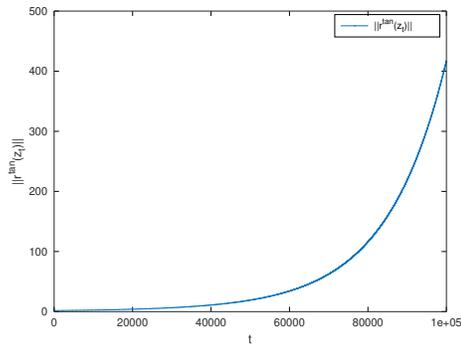


Figure 3: $\|r^{tan}(z_t)\|$ in a periodic game for the ARG algorithm

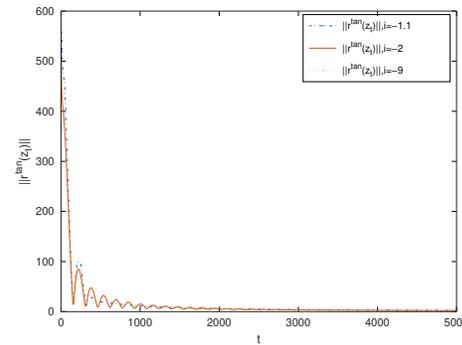


Figure 4: $\|r^{tan}(z_t)\|$ in a convergent perturbed game for the ARG algorithm

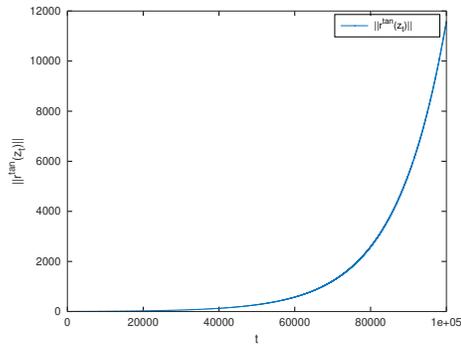


Figure 5: $\|r^{tan}(z_t)\|$ in a periodic game for the modified OG algorithm

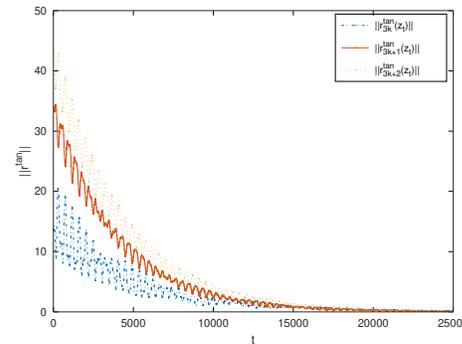


Figure 6: $\|r^{tan}(z_t)\|$ in a periodic game for the OG algorithm