FAST CONVERGENCE OF OPTIMISTIC GRADIENT ASCENT IN NETWORK ZERO-SUM EXTENSIVE FORM GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

The study of learning in games has thus far focused primarily on normal form games. In contrast, our understanding of learning in *extensive form games* (EFGs) and particularly in EFGs with many agents lags far behind, despite them being closer in nature to many real world applications. We consider the natural class of *Network Zero-Sum Extensive Form Games*, which combines the global zero-sum property of agent payoffs, the efficient representation of graphical games as well the expressive power of EFGs. We examine the convergence properties of *Optimistic Gradient Ascent* (OGA) in these games. We prove that the time-average behavior of such online learning dynamics exhibits $O(1/T)$ rate convergence to the set of Nash Equilibria. Moreover, we show that the day-to-day behavior also converges to Nash with rate $O(c^{-t})$ for some game-dependent constant $c > 0$.

1 INTRODUCTION

Extensive Form Games (EFGs) are an important class of games which have been extensively studied for more than 50 years (Kuhn, 1950b; Koller and Megiddo, 1992). EFGs capture various settings where several selfish agents sequentially perform actions which change the *state of nature*, with the action-sequence finally leading to a *terminal state*, at which each agent receives a payoff. The most ubiquitous examples of EFGs are real-life games such as Chess, Poker, Go etc. Recently the application of the *online learning framework* has proven to be very successful in the design of modern AI which can beat even the best human players in real-life games (Tammelin et al., 2015; Brown and Sandholm, 2017a;b). At the same time, online learning in EFGs has many interesting applications in Economics, AI, machine learning and sequential decision making that extend far beyond the design of game-solvers (Arieli and Babichenko, 2016; Pérolat et al., 2021).

Despite its numerous applications, online learning in games is far from well understood. From a practical point of view, testing and experimenting with various online learning algorithms in EFGs requires a huge amount of computational resources due to the large number of states in EFGs of interest (Zinkevich et al., 2008; Kroer et al., 2018; Brown and Sandholm, 2018a; Rowland et al., 2019). From a theoretical perspective, it is known that online learning dynamics may oscillate, cycle or even admit chaotic behavior even in very simple settings (Piliouras and Shamma, 2014; Palaiopanos et al., 2017; Mertikopoulos et al., 2018b; Vlatakis-Gkaragkounis et al., 2019; Leonardos and Piliouras, 2021). On the positive side, there exists a recent line of research into the special but fairly interesting class of *two-player zero-sum EFGs*, which provides the following solid claim: *In two-player zero-sum EFGs, the time-average strategy vector produced by online learning dynamics converges to the Nash Equilibrium (NE), while there exist online learning dynamics which exhibit day-to-day convergence* (Zinkevich et al., 2008; Lanctot et al., 2009; Kroer et al., 2018; Farina et al., 2019a; Wei et al., 2020; 2021b). Since in most settings of interest there are multiple interacting agents, all the above motivates the following question:

Question. *Are there natural and important classes of multi-agent extensive form games for which online learning dynamics converge to a Nash Equilibrium? Furthermore, what type of convergence is possible? Can we only guarantee time-average convergence or can we also prove day-to-day convergence (also known as last-iterate convergence) of the dynamics?*

In this paper we answer the above questions in the positive for an interesting class of multi-agent EFGs called *Network Zero-Sum Extensive Form Games*. A Network EFG consists of a graph $\mathcal{G} = (V, E)$ where each vertex $u \in V$ represents a selfish agent and each edge $(u, v) \in E$ corresponds to an extensive form game Γ^{uv} played between the agents $u, v \in V$. Each agent $u \in V$ selects her strategy so as to maximize the overall payoff from the games corresponding to

her incident edges. The game is additionally called *zero-sum* if the sum of the agents' payoffs is equal to zero no matter the selected strategies.

Network Zero-Sum EFGs are an interesting class of multi-agent EFGs for various reasons. First of all, their global constant-sum property[1] (the edge-games are not necessarily zero-sum) is very natural for closed systems in which self-interested agents competes over a fixed set of resources (Daskalakis and Papadimitriou, 2009a; Cai and Daskalakis, 2011a). For example, consider the users of an online poker platform playing Heads-up Poker (2-player poker). Each user can be thought of as a node in a graph and two users are connected by an edge (corresponding to a Head's up Poker game) if they play against each other. Note that here, each edge/game differs from another due to the differences in the dollar/blind equivalence. Each user selects a poker-strategy to utilize against the other players, with the goal of maximizing her overall payoff. In addition, Network Zero-Sum EFGs are also interesting due to the fact that descriptive complexity scales polynomially with the number of agents. Multi-agent EFGs that cannot be decomposed into pairwise interactions (i.e., do not have a network structure) admit an exponentially large description with respect to the number of the agents (Kearns et al., 2001; Babichenko, 2014). In this work, we analyze the convergence properties of the online learning dynamics produced when all agents of a Network Zero-Sum EFG update their strategies according to Optimistic Gradient Ascent

Informal Theorem. When the agents of a network zero-sum extensive form game update their strategies through optimistic gradient ascent, their time-average strategies converge with rate $O(1=T)$ to Nash Equilibrium, while the day-to-day mixed strategies converge with rate $O(c^{t})$ for some game-dependent constant $c < 1$.

Our Contributions. To the best of our knowledge, this is the first work establishing convergence to equilibrium of online learning dynamics in network extensive form games with more than two agents. As already mentioned, there has been a stream of recent works establishing the convergence of online learning dynamics in two-player zero-sum EFGs. However, there are several key differences between the two-player and the network cases. All the previous works concerning the two-player case follow a *bilinear saddle point approach*. Specifically, due to the fact that in the two-agent case any Nash Equilibrium coincides with a min-max equilibrium, the set of Nash Equilibria can be expressed as the solution to the following bilinear saddle-point problem:

$$\min_{x \in X} \max_{y \in Y} x^{\top} A y = \max_{y \in Y} \min_{x \in X} x^{\top} A y$$

In the two-player case, one can show that online learning dynamics converge to Nash Equilibrium by showing that they converge to the solution of the above saddle-point problem.

However, in the network case there is no min-max equilibrium, and thus there is no such connection between the Nash Equilibrium and saddle-point optimization. To overcome this difficulty, we establish that optimistic gradient ascent in Network Zero-Sum EFGs can be equivalently described as optimistic gradient descent in a two-player symmetric game $(R; R)$ over a treeplex polytope X. We remark that both the matrix R and the treeplex polytope X are constructed from the Network Zero-Sum EFG. Using the zero-sum property of Network EFGs, we show that the constructed R matrix satisfies the following 'restricted' zero-sum property:

$$x^{\top} R y + y^{\top} R x = 0 \text{ for all } x; y \in X \tag{1}$$

Indeed, Property (1) is a generalization of the classical zero-sum property $A = -A^{\top}$. In general, the constructed matrix R does not satisfy $R = -R^{\top}$ and Property 1 simply ensures that the sum of payoffs equal to zero only when $x; y \in X$. Our technical contribution consists of generalizing the analysis of Wei et al. (2020) (which holds for classical two-player zero-sum games) to symmetric games satisfying Property (1).

Related Work. Network Zero-Sum Normal Form Games (Daskalakis and Papadimitriou, 2009a; Cai and Daskalakis, 2011a; Cai et al., 2016) are a special case of our setting, where each edge/game is a normal-form game. Network zero-sum normal-form games present major complications compared to their two-player counterparts. The most important of these complications is that in the network case, there is no min-max equilibrium. In fact, different Nash Equilibria can assign different values to the agents. All the above works study linear programs for computing Nash Equilibria in network zero-sum normal-form games. Cai and Daskalakis (2011a) introduce the idea of connecting a network zero-sum normal form game with an equivalent symmetric game $(R; R)$ which satisfies Property (1). This generalizes the linear programming approach of two-player zero-sum normal-form games to the network case. They also show that in network normal-form zero-sum games, the time-average behavior of online learning dynamics converge with rate $(1 = \sqrt{T})$ to the Nash Equilibrium.

[1] equivalent to the global zero-sum property.

The properties of Online Learning in two-player zero-sum EFGs have been studied extensively in literature. Zinkevich et al. (2008) and Lanctot et al. (2009) propose no-regret algorithms for extensive form games with $O(1/\sqrt{T})$ average regret and polynomial running time in the size of the game. More recently, regret-based algorithms with $O(1/T)$ time-average convergence to the min-max equilibrium (Hoda et al., 2010a; Kroer et al., 2018; Farina et al., 2019a) for two-player zero-sum EFGs. Finally, Lee et al. (2021a) and Wei et al. (2021a) establish that Online Mirror Descent achieves $O(c^{-t})$ last-iterate convergence (for some game-dependent constant $c \in (0, 1)$) in two-player zero-sum EFGs.

2 PRELIMINARIES

2.1 EXTENSIVE FORM GAMES

Definition 1. A two-player extensive form game is a tuple $\Gamma := \langle H, A, Z, p, I \rangle$ where

- H denotes the states of the game that are decision points for the agents. The states in H form a tree rooted at an initial state $r \in H$.
- Each state $h \in H$ is associated with a set of available actions $A(h)$.
- Each state $h \in H$ admits a label $\text{Label}(h) \in \{1, 2, c\}$ denoting the acting player at state h. The letter c denotes a special agent called a chance agent. Each state $h \in H$ with $\text{Label}(h) = c$ is additionally associated with a function $\sigma_h : A(h) \to [0, 1]$ where $\sigma_h(a)$ denotes the probability that the chance player selects action $a \in A(h)$ at state h, $\sum_{a \in A(h)} \sigma_h(a) = 1$.
- $\text{Next}(a, h)$ denotes the state $h' := \text{Next}(a, h)$ which is reached when agent $i = \text{Label}(h)$ takes action $a \in A(h)$ at state h. $H_i \subseteq H$ denotes the states $h \in H$ with $\text{Label}(h) = i$.
- Z denotes the terminal states of the game corresponding to the leaves of the tree. At $z \in Z$ no further action can be chosen, so $A(z) = \emptyset$ for all $z \in Z$. Each terminal state $z \in Z$ is associated with values $(u_1(z), u_2(z))$ where $p_i(z)$ denotes the payoff of agent i at terminal state z.
- Each set of states H_i is further partitioned into information sets $(I_1, \ldots, I_k)$ where $I(h)$ denotes the information set of state $h \in H_i$. In the case that $I(h_1) = I(h_2)$ for some $h_1, h_2 \in H_1$, then $A(h_1) = A(h_2)$.

Information sets model situations where the acting agent cannot differentiate between different states of the game due to a lack of information. Since the agent cannot differentiate between states of the same information set, the available actions at states h_1, h_2 in the same information set ($I(h_1) = I(h_2)$) must coincide, in particular $A(h_1) = A(h_2)$.

Definition 2. A behavioral plan π_i for agent i is a function such that for each state $h \in H_i$, $\pi_i(h)$ is a probability distribution over $A(h)$ i.e. $\pi_i(h, a)$ denotes the probability that agent i takes action $a \in A(h)$ at state $h \in H_i$. Furthermore it is required that $\pi_i(h_1) = \pi_i(h_2)$ for each $h_1, h_2 \in H_i$ with $I(h_1) = I(h_2)$. The set of all behavioral plans for agent i is denoted by Π_i.

The constraint $\pi_i(h_1) = \pi_i(h_2)$ for all $h_1, h_2 \in H_i$ with $I(h_1) = I(h_2)$ models the fact that since agent i cannot differentiate between states h_1, h_2, agent i must act in the exact same way at states $h_1, h_2 \in H_i$.

Definition 3. For a collection of behavioral plans $\pi = (\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ the payoff of agent i, denoted by $U_i(\pi)$, is defined as:

$$U_i(\pi) := \sum_{z \in Z} p_i(z) \underbrace{\prod_{(h, h') \in P(z)} \pi_{\text{Label}(h)}(h, a_{h'})}_{\text{probability that state } z \text{ is reached}}$$

where $P(z)$ denotes the path from the root state to the terminal state z and $a_{h'}$ denotes the action $a \in H_i$ such that $h' = \text{Next}(h, a)$.

Definition 4. A collection of behavioral plans $\pi = (\pi_1, \pi_2)$ is called a Nash Equilibrium if for all agents $i = 1, 2$,

$$U_i(\pi_i, \pi_{-i}) \geq U_i(\pi'_i, \pi_{-i}) \quad \text{for all } \pi'_i \in \Pi_i$$

The classical result of Nash (1951) proves the existence of Nash Equilibrium in normal form games. This result also generalizes to a wide class of extensive form games which satisfy a property called perfect recall (Kuhn and Tucker (1953); Selten (1965)).

De nition 5. A two-player extensive form game $= hH; A; Z; p; Ii$ has perfect recall if and only if for all states $h_1; h_2 \in H_i$ with $I(h_1) = I(h_2)$ the following holds: De ne the sets $P(h_1) \setminus H_i := (p_1; \ldots; p_k; h_1)$ and $P(h_2) \setminus H_i := (q_1; \ldots; q_m; h_2)$. Then:

1. $k = m$.
2. $I(p_\ell) = I(q_\ell)$ for all $\ell \in \{1; \ldots; k\}$.
3. $p_{\ell+1} \in Next(p_\ell; ; i)$ and $q_{\ell+1} \in Next(q_\ell; ; i)$ for some action $\in A(p_\ell)$ (since $A(p_\ell) = A(q_\ell)$).

Before proceeding, let us further explain the perfect recall property. As already mentioned, agent cannot differentiate between states $h_1; h_2 \in H_i$ when $I(h_1) = I(h_2)$. In order for the state h_1 to be reached, agent i must take some speci c actions along the path $P(h_1) \setminus H_i := (p_1; \ldots; p_k; h_1)$. The same logic holds for $P(h_2) \setminus H_i := (p_1; \ldots; p_k; h_1)$. In case where agent i could distinguish $P(h_1) \setminus H_i$ from set $P(h_2) \setminus H_i$, then she could distinguish state h_1 from h_2 by recalling the previous states from h_i. This is the reason for the second constraint in De nition 5. Even if $I(p_\ell) = I(q_\ell)$ for all $\ell \in \{1; \ldots; k\}$, agent i could still be able to distinguish h_1 from h_2 if $p_{\ell+1} \in Next(p_\ell; ; i)$ and $q_{\ell+1} \in Next(q_\ell; ^0; i)$. In such a case, agent i can distinguish h_1 from h_2 by recalling the actions that she previously played and checking if the ℓ-th action was $$ or ^{0}. The latter case is encompassed by the third constraint.

2.2 Extensive Form Games in Sequence Form

A two-player extensive form game can be captured by a two-player bilinear game where the action spaces of the agents are a speci c kind of polytope, commonly known as a treeplex (Hoda et al., 2010a). In order to formally de ne the notion of a treeplex, we rst need to introduce some additional notation.

De nition 6. Given an two-player extensive form game, we de ne the following:

- $P(h)$ denotes the path from the root state $\in H$ to the state $h \in H$.
- $Level(h)$ denotes the distance from the root state $\in H$ to state $h \in H$.
- $Prev(h; i)$ denotes the lowest ancestor of h in the set H_i. In particular,
$$Prev(h; i) = \operatorname{argmax}_{h' \in P(h) \setminus H_i} Level(h').$$
- The set of states $Next(h; ; i) \subseteq H$ denotes the highest descendants $h' \in H_i$ once action $\in A(h)$ has been taken at state h. More formally, $h' \in Next(h; ; i)$ if and only if in the path $P(h; h') = (h; h_1; \ldots; h_k; h')$, all states $h_\ell \in H_i$ and $h_1 = Next(h;)$.

De nition 7. Given a two-player extensive form game, the set X_i is composed by all vectors $x_i \in [0; 1]^{|H_i| + |Z|}$ which satisfy the following constraints:

1. $x_i(h) = 1$ for all $h \in H_i$ with $Prev(h; i) = ?$.
2. $x_i(h_1) = x_i(h_2)$ if there exists $h'_1; h'_2 \in H_i$ such that $h_1 \in Next(h'_1; ; i)$, $h_2 \in Next(h'_2; ; i)$ and $I(h'_1) = I(h'_2)$.
3. $\sum_{\in A(h)} x_i(Next(h; ; i)) = x_i(h)$ for all $h \in H_i$.

A vector $x_i \in X_i$ is typically referred to as an agent i's strategy in sequence form. Strategies in sequence form come as an alternative to the behavioral plans of De nition 2. As established in Lemma 1, there exists an equivalence between a behavioral plan $_i \in _i$ and a strategy in sequence form $x_i \in X_i$ for games with perfect recall.

Lemma 1. Consider a two-player extensive form game with perfect recall and the $(|H_1| + |Z|) \times (|H_2| + |Z|)$ dimensional matrices $A_1; A_2$ with $[A_i]_{zz} = p_i(z)$ for all terminal nodes $z \in Z$ and 0 otherwise. There exists a polynomial-time algorithm transforming any behavioral plan $_i \in _i$ to a vector $x_i \in X_i$ such that

$$U_1(_1; _2) = x_1^\top A_1 x_2 \quad \text{and} \quad U_2(_1; _2) = x_2^\top A_2 x_1$$

Conversely, there exists a polynomial-time algorithm transforming any vector $x_i \in X_i$ to a vector $_{x_i} \in _i$ such that

$$x_1^\top A_1 x_2 = U_1(_{x_1}; _{x_2}) \quad \text{and} \quad x_2^\top A_2 x_1 = U_2(_{x_1}; _{x_2})$$

To this end, one can understand why strategies in sequence form are of great use. Assume that agent 2 selects a behavioral plan $\pi_2 \in \Pi_2$. Then, agent 1 wants to compute a behavioral plan $\pi_1 \in \Pi_1$ which is the best response to π_2, namely $\pi_1 := \text{argmax}_{\pi_1 \in \Pi_1} U_1(\pi_1, \pi_2)$. This computation can be done in polynomial-time in the following manner: Agent 1 initially converts (in polynomial time) the behavioral plan π_2 to $x_2 \in \mathcal{X}_2$, which is the respective strategy in sequence form. Then, she can obtain a vector $x_1 := \text{argmax}_{x_1 \in \mathcal{X}_1} x_1^\top A_1 x_2$. The latter step can be done in polynomial-time by computing the solution of an appropriate linear program. Finally, she can convert the vector x_1 to a behavioral plan $\pi_1 \in \Pi_1$ in polynomial-time. Lemma 1 ensures that $\pi_1 = \text{argmax}_{\pi_1 \in \Pi_1} U_1(\pi_1, \pi_2)$.

The above reasoning can be used to establish an equivalence between the Nash Equilibrium (π_1, π_2) of an EFG $\Gamma := \langle H, A, Z, p, I \rangle$ with the Nash Equilibrium in its sequence form.

Definition 8. A Nash Equilibrium of a two-player EFG in sequence form is a vector $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ such that

- $(x_1)^\top A_1 x_2 \geq (x_1')^\top A_1 x_2$ for all $x_1' \in \mathcal{X}_1$
- $(x_2)^\top A_2 x_1 \geq (x_2')^\top A_2 x_1$ for all $x_2' \in \mathcal{X}_2$

Lemma 1 directly implies that any Nash Equilibrium of an EFG $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ of Definition 4 can be converted in polynomial-time to a Nash Equilibrium in the sequence form $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ and vice versa.

2.3 Optimistic Mirror Descent

In this section we introduce and provide the necessary background of Optimistic Mirror Descent (Rakhlin and Sridharan, 2013a). For a convex function $\psi : \mathbb{R}^d \to \mathbb{R}$, the corresponding Bregman divergence is defined as

$$D_\psi(x, y) := \psi(x) - \psi(y) - \langle \nabla\psi(y), x - y \rangle$$

If ψ is σ-strongly convex, then $D_\psi(x, y) \geq \frac{\sigma}{2}\|x - y\|$. Here and in the rest of the paper, we note that $\|\cdot\|$ is shorthand for the L_2-norm.

Now consider a game played by n agents, where the action of each agent i is a vector x_i from a convex set $\mathcal{X}_i$. Each agent selects its action $x_i \in \mathcal{X}_i$ so as to minimize her individual cost (denoted by $C_i(x_i, x_{-i})$), which is continuous, differentiable and convex with respect to x_i. Specifically,

$$C_i(\alpha x_i + (1-\alpha) x_i', x_{-i}) \leq \alpha C_i(x_i, x_{-i}) + (1-\alpha) C_i(x_i', x_{-i}) \text{ for all } \alpha \in [0, 1]$$

Given a step size $\eta > 0$ and a convex function $\psi(\cdot)$ (called a regularizer), Optimistic Mirror Descent (OMD) sequentially performs the following update step for $t = 1, 2, \ldots$:

$$x_i^t = \text{argmin}_{x \in \mathcal{X}_i} \langle x, F_i^{t-1}(x) \rangle + D_\psi(x, \hat{x}_i^t) \tag{2}$$

$$\hat{x}_i^{t+1} = \text{argmin}_{x \in \mathcal{X}_i} \langle x, F_i^t(x) \rangle + D_\psi(x, \hat{x}_i^t) \tag{3}$$

where $F_i^t(x_i) = \nabla_{x_i} C_i(x_i, x_{-i}^t)$ and $D_\psi(x, y)$ is the Bregman Divergence with respect to $\psi(\cdot)$. If the step-size selected is sufficiently small, then Optimistic Mirror Descent ensures the no-regret property (Rakhlin and Sridharan, 2013a), making it a natural update algorithm for self-play agents (Hazan, 2019). To simplify notation we denote the projection operator of a convex set $\mathcal{X}$ as $\Pi_{\mathcal{X}}(x) := \text{argmax}_{x' \in \mathcal{X}} \|x - x'\|$ and the squared distance of vector x from a convex set $\mathcal{X}$ as $\text{dist}^2(x, \mathcal{X}) := \|x - \Pi_{\mathcal{X}}(x)\|^2$.

3 Our Setting

We introduce the concept of a Network Zero-Sum Extensive Form Game, which is a network extension of the two player EFGs we have introduced in Section 2.

3.1 Network Zero-Sum Extensive Form Games

A network extensive form game is defined with respect to an undirected graph $G = (V, E)$ where nodes $v \in V$ ($|V| = n$) correspond to the set of players and each edge $(u, v) \in E$ represents a two-player extensive form game Γ^{uv} played between agents u, v. Each node/agent $u \in V$ selects a behavioral plan $\pi_u \in \Pi_u$ which they use to play all the two-player EFGs on its outgoing edges.

De nition 9 (Network Extensive Form Games)A network extensive form game is a tuple= hG;H;A;Z;Ii where

- G = (V;E) is an an undirected graph where the nodes represents the agents.
- Each agentu 2 V admits a set of statesH u at which the agentu plays. Each stateh 2 H u is associated with a setA(h) of possible actions that agentu can take at stateh.
- I (h) denotes the information set of 2 H u. If I (h) = I (h0) for someh;h0 2 H u thenA(h) = A(h0).
- For each edge(u;v) 2 E, uv is a two-player extensive form game with perfect recall. The states of uv are denoted byH uv H u [H v.
- For each edge(u;v) 2 E, Z uv is the set of terminal states of the two-player extensive form game uv where p u uv (z) denotes the payoffs of u;v at the terminal statez 2 Z uv. The overall set of terminal states of the network extensive form game is the setZ := [(u;v)2E Z uv.

In a network extensive form game, each agentu 2 V selects a behavioral plan u 2 u (see De nition 2) that they use to play the two-player EFG's uv with (u;v) 2 E. Each agent selects her behavioral plan so as to maximize the sum of the payoffs of the two-player EFGs in her outgoing edges.

De nition 10. Given a collection of behavioral plans= (1;:::; n) 2 1 ::: n the payoff of agentu, denoted by U u( ), equals

$$U_u(\ ) := \sum_{v:(u,v)\in E} p_u^{uv}(\ _u;\ _v)$$

Moreover a collection = (1;:::; n) 2 1 ::: n is called a Nash Equilibrium if and only if

$$U_u(\ _u;\ _u) \quad U_u(\ _u;\ _u) \quad \text{for all } \ _u \in \ _u$$

As already mentioned, each agentu 2 V plays all the two-player games uv for (u;v) 2 E with the same behavioral plan u 2 u. This is due to the fact that the agent cannot distinguish between all states h1;h2 2 H u with I (h1) = I (h2) even if h1;h2 are states of different EFG's uv and uv0. As in the case of perfect recall the latter implies that u cannot differentiate statesh1;h2 even when recalling the statesH u visited in the past and her past actions. In De nition 11 we introduce the notion ofconsistency(this corresponds to the notion ofperfect recallfor two-player extensive form games (De nition 5)). From now on we assume that the network EFG is consistent without mentioning it explicitly.

De nition 11. A network extensive form game= hG;H;A;Z;Ii is calledconsistentif and only if for all players u 2 V and statesh1;h2 2 H u with I (h1) = I (h2) the following holds: for any(u;v);(u;v0) 2 E the sets P uv(h1) \ H u := (p1;:::;pk;h1) and P uv0(h2) \ H u := (q1;:::;qm;h2) satisfy:

1. k = m.
2. I (p) = I (q) for all ` 2 f1;:::;kg.
3. p +1 2 Next uv (p ; ;u) and q +1 2 Next uv0(q ; ;u) for some action 2 A(p).

where P uv(h) denotes the path from the root state to stateh in the two-player extensive form game uv.

In this work we study the special class of networkzero-sumextensive form games. This class of games is a generalization of the network zero-sum normal-form games studied in Cai and Daskalakis (2011b).

De nition 12. A behavioral plan u 2 u of De nition 2 is called pure if and only if u(h;) either equals 0 or 1 for all actions 2 A(h). A network extensive form game is calledzero-sumif and only if for any collection := (1;:::; n) of pure behavioral plans,U u( ) = 0 for all u 2 V.

3.2 Network Extensive Form Games in Sequence Form

As in the case of two-player EFGs, there exists an equivalence between behavioral plans u and strategies in sequence form x u. As we shall later see, this equivalence is of great importance since it permits the design of natural and computationally ef cient learning dynamics that converge to Nash Equilibria both in terms of behavioral plans and strategies in sequence form.

De nition 13. Given a network extensive form game $= hG; H; A; Z; Ii$, the treeplex polytope $X_u \quad [0;1]^{jH_u j + jZ_u j}$ is the set de ned as follows: $x_u \; 2 \; X_u$ if and only if

1. $x_u \; 2 \; X_u{}^{uv}$ for all $(u;v) \; 2 \; E$.
2. $x_u(h_1) = x_u(h_2)$ in case there exists $(u;v); (u;v^0) \; 2 \; E$ and $h_1^0; h_2^0 \; 2 \; H_u$ with $I(h_1^0) = I(h_2^0)$ such that $h_1 \; 2 \; \text{Next}^{uv}(h_1^0; ; u), h_2 \; 2 \; \text{Next}^{uv^0}(h_2^0; ; u)$ and $I(h_1^0) = I(h_2^0)$.

The second constraint of De nition 13 is the equivalent of the second constraint in De nition 7. To this end, we remark that the linear equations describing the treeplex polytope X_u can be derived in polynomial-time with respect to the description of the network extensive form game. In Lemma 2 we formally state and prove the equivalence between behavioral plans and strategies in sequence form.

Lemma 2. Consider the matrix A^{uv} of dimensions $(jH_u j + jZ^u j) \quad (jH_v j + jZ^v j)$ such that

$$[A^{uv}]_{h_1 h_2} = \begin{cases} p_u{}^{uv}(h) & \text{if } h_1 = h_2 = h \; 2 \; Z^{uv} \\ 0 & \text{otherwise} \end{cases}$$

There exists a polynomial time algorithm converting any collection of behavioral plans $(_1; : : ; _n) \; 2 \; _1 \quad : : : \quad _n$ into a collection of vectors $(x_1; : : : ; x_n) \; 2 \; X_1 \quad : : : \quad X_n$ such that for any $u \; 2 \; V$,

$$U_u(\ ) = x_u^> \sum_{v:(u;v)2E} A^{uv} \; x_v$$

In the opposite direction, there exists a polynomial time algorithm converting any collection of vectors $(x_1; : : ; x_n) \; 2 \; X_1 \quad : : : \quad X_n$ into a collection of behavioral plans $(_1; : : : ; _n) \; 2 \; _1 \quad : : : \quad _n$ such that for any $u \; 2 \; V$,

$$x_u^> \sum_{v:(u;v)2E} A^{uv} \; x_v = U_u(\ )$$

De nition 14. A Nash Equilibrium of a network extensive form game in sequence form is a vector $(x_1; : : : ; x_n) \; 2 \; X_1 \quad : : : \quad X_n$ such that for all $u \; 2 \; V$:

$$(x_u)^> \sum_{v:(u;v)2E} A^{uv} \; x_v \quad x_u^> \sum_{v:(u;v)2E} A^{uv} \; x_v \quad \text{for all } x_u \; 2 \; X_u$$

Corollary 1. Given a network extensive form game, any Nash Equilibrium $(_1; : : ; _n) \; 2 \; _1 \quad : : : \quad _n$ of De nition 4 can be converted in polynomial-time to a Nash Equilibrium $(x_1; : : : ; x_n) \; 2 \; X_1 \quad : : : \quad X_n$ of De nition 14 and vice versa.

4 Our Results

In this work, we study the convergence properties of Optimistic Gradient Ascent (OGA) when applied to network zero-sum EFGs. OGA is a special case of Optimistic Mirror Descent where the regularizer is $(a) = \frac{1}{2}kak^2$, which means that the Bregman divergence $D(x;y)$ equals $\frac{1}{2}kx \quad yk^2$. Since in network zero-sum EFGs each agent tries to maximize her payoff, OGA takes the following form:

$$x_u^t = \text{argmax}_{x \; 2 X_u} \left\{ x; \sum_{v:(u;v)2E} A^{uv} \; x_v^{t \; 1} \quad D \quad x; x_u^t \right\} \tag{4}$$

$$x_u^{t+1} = \text{argmax}_{x \; 2 X_u} \left\{ x; \sum_{v:(u;v)2E} A^{uv} \; x_v^t \quad D \quad x; x_u^t \right\} \tag{5}$$

In Theorem 1 we claim and describe the $(1 =T)$ convergence rate for the time-average strategy produced by OGA.

Theorem 1. Let $fx^1; x^2; : : : x^T g$ be the vectors produced by Equations (4),(5) for some initial strategies $x^0 := (x_1^0; : : : ; x_n^0)$. There exist game-dependent constants $c_1; c_2 > 0$ such that if $1=c_1$ then for any $u \; 2 \; V$:

$$\hat{x}_u^> \sum_{v:(u;v)2E} A^{uv} \; \hat{x}_v \quad x^> \sum_{v:(u;v)2E} A^{uv} \; \hat{x}_v \quad \frac{c_1 \quad c_2}{T} \quad \text{for all } x \; 2 \; X_u$$

where $\hat{x}_u = \sum_{s=1}^{T} x_u^s = T$.

Applying the polynomial-time transformation of Lemma 2 to the time-average strategy vector $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)$ produced by Optimistic Gradient Ascent, we immediately get that for any agent u,

$$U_u(\hat{x}_u, \hat{x}_{-u}) \geq U_u(x_u, \hat{x}_{-u}) - (c_1 + c_2/T) \quad \text{for all } x_u \in \mathcal{X}_u$$

In Theorem 2 we establish the fact that OGA admits last-iterate convergence to NE in network zero-sum EFGs.

Theorem 2. Let $\{x^1, x^2, \dots, x^T\}$ be the vectors produced by Equations (4),(5) for $\eta = 1/c_3$ when applied to a network zero-sum extensive form game. Then, the following inequality holds:

$$\text{dist}^2(x^t, \mathcal{X}^*) \leq 64 \text{dist}^2(x^1, \mathcal{X}^*)(1 + c_1)^{-t}$$

where $\mathcal{X}^*$ denotes the set of Nash Equilibria, $c_1 := \min\left\{\frac{16\eta^2 c^2}{81}, \frac{1}{2}\right\}$ and c_3, c are positive game-dependent constants.

We conclude the section by providing the key ideas towards proving Theorems 1 and 2. For the rest of the section, we assume that the network extensive form game is zero-sum. Before proceeding, we introduce a few more necessary definitions and notations. We denote $\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ the product of treeplexes of Definition 13 and define the $|\mathcal{X}| \times |\mathcal{X}|$ matrix R as follows:

$$R_{(u:h_1),(v:h_2)} = \begin{cases} [A^{uv}]_{h_1 h_2} & \text{if } (u,v) \in E \\ 0 & \text{otherwise} \end{cases}$$

The matrix R can be used to derive a more concrete form of the Equations (4),(5):

Lemma 3. Let $\{x^1, x^2, \dots, x^T\}$ be the collection of strategy vectors produced by Equations (4),(5) initialized with $x^0 := (x_1^0, \dots, x_n^0) \in \mathcal{X}$. The equations

$$x^t = \text{argmin}_{x \in \mathcal{X}} \langle x, R x^{t-1} \rangle + D(x, x^t) \tag{6}$$

$$x^{t+1} = \text{argmin}_{x \in \mathcal{X}} \langle x, R x^t \rangle + D(x, x^t) \tag{7}$$

produce the exact same collection of strategy vectors $\{x^1, \dots, x^T\}$ when initialized with $x^0 \in \mathcal{X}$.

To this end, we derive a two-player symmetric game (R, R) defined over the polytope $\mathcal{X}$. More precisely, the x-agent selects $x \in \mathcal{X}$ so as to minimize $x^\top R y$ while the y-agent selects $y \in \mathcal{X}$ so as to minimize $y^\top R x$. Now consider the Optimistic Mirror Descent algorithm (described in Equations (2),(3)) applied to the above symmetric game. Notice that if $x^0 = y^0$, then by the symmetry of the game, the produced strategy vectors (x^t, y^t) will be of the form (x^t, x^t) and indeed x^t, x^t will satisfy Equations (6), (7). We prove that the produced vector sequence $\{x^t\}_{t \geq 1}$ converges to symmetric Nash Equilibrium

Lemma 4. A strategy vector x^* is an ε-symmetric Nash Equilibrium for the symmetric game (R, R) if the following holds:

$$(x^*)^\top R x - x^\top R x \leq \varepsilon \quad \text{for all } x \in \mathcal{X}$$

Any ε-symmetric Nash Equilibrium $x^* \in \mathcal{X}$ is also an ε-Nash Equilibrium for the network zero-sum EFG.

A key property of the constructed matrix is the one stated and proven in Lemma 5. Its proof follows the steps of the proof of Lemma B.3 in Cai and Daskalakis (2011b) and is presented in Appendix B.5.

Lemma 5. $x^\top R y + y^\top R x = 0$ for all $x, y \in \mathcal{X}$.

Once Lemma 5 is established, we can use to it to prove that the time-average strategy vector converges to a symmetric Nash Equilibrium in a two-player symmetric game.

Lemma 6. Let $(x^1, x^2, \dots, x^T)$ be the sequence of strategy vectors produced by Equations (6),(7) for $\eta \leq \min\{1/8\|R\|^2, 1\}$. Then,

$$\min_{x \in \mathcal{X}} x^\top R \hat{x} \geq -\frac{D^2 \|R\|^2}{T}$$

where $\hat{x} = \sum_{s=1}^{T} x^s/T$ and D is the diameter of the treeplex polytope $\mathcal{X}$.

Combining Lemma 5 with Lemma 6, we get that that the time-average vector $\hat{x}$ is a $\frac{D^2\|R\|^2}{T}$-symmetric Nash Equilibrium. This follows directly from the fact that $\hat{x}^\top R \hat{x} = 0$. Then, Theorem 1 follows by direct application of Lemma 4. For completeness, we present the complete proof in Appendix B.7.

By Lemma 5, it directly follows that the set of symmetric Nash Equilibria can be written as:

$$X^* = \{x \in X : \min_{x \in X} x^\top R x = 0\}.$$

Using this, we establish that Optimistic Gradient Descent admits last-iterate convergence to the symmetric NE of the (R, R) game. This result is formally stated and proven in Theorem 3, the proof of which is deferred to Appendix B.8. Theorem 2 then follows directly by Theorem 3 and Lemma 4.

Theorem 3. Let $\{x^1, x^2, \ldots x^T\}$ be the vectors produced by Equations (6),(7) for $\min(1/8\|R\|^2, 1)$. Then:

$$\text{dist}^2(x^t, X^*) \leq 64\,\text{dist}^2(x^1, X^*)(1 + C_2)^{-t}$$

where $C_2 := \min\left\{\frac{16\gamma^2 C^2}{81}, \frac{1}{2}\right\}$ with C being a positive game-dependent constant.

5 Experimental Results

In order to better visualize our theoretical results, we experimentally evaluate OGA when applied to various network extensive form games. As a sort of sanity check for our results, we first experimented with small random network games, where each bilinear game between the players on the nodes is a randomly generated extensive form game.

As part of the experimental process, for each randomly generated game we ran a hyperparameter search to find the value of γ which gave the best convergence rate. As can be seen in Figure 1(a)-(b), we are able to obtain fast convergence in the day-to-day sense to the Nash Equilibrium. We measure the log of the distance between each player's strategy at time t and the set of Nash Equilibria, given by $\log(\text{dist}^2(x^t, X^*))$. Note here that in all our experiments, we use the last iterate of the strategy vectors as the Nash Equilibrium value.

In Figure 1(c), we show results of our experiments with an oft-studied simplification of poker, namely Kuhn poker. In our experiment, we model a situation whereby each player is playing against multiple other agents. We see that although the time needed for the players to converge is significantly greater, with a careful choice of γ we can guarantee convergence to the set of Nash Equilibria for all players. Further experiments and detailed game descriptions can be found in Appendix C.

Figure 1: Simulations using OGA in network extensive form games, where each player is involved in 2 or more different games and must select their strategy accordingly. The plots shown are: (Left) Day-to-day convergence to the Nash in 3-player random network extensive form games. (Center) Day-to-day convergence to the Nash in 4-player random network extensive form games. Note the significantly longer time needed to achieve convergence compared to the 3-player experiment. (Right) Convergence to the Nash in 5-player Kuhn poker

6 Conclusion

In this paper, we provide a formulation of Network Zero-Sum Extensive Form Games, which encode the setting where multiple agents compete in pairwise games over a set of resources. We analyze the convergence properties of Optimistic Gradient Ascent in this setting, proving that OGA results in both time-average and day-to-day convergence to the set of Nash Equilibria. In order to show this, we utilize a transformation from network zero-sum extensive form games to two-player symmetric games and subsequently show the convergence results in the symmetric game setting. This work represents an initial foray into the world of online learning dynamics in network extensive form games, and we hope that this will lead to more research into the practical and theoretical applications of multi-agent extensive form games.

Reproducibility Statement

In order to ensure reproducibility for our theoretical results, we have made efforts to clearly describe the proof of all of our results in Appendix B. Moreover, due to the complicated nature of the network extensive form games we study, we have also used a substantial amount of space in the main text (see Sections 2 and 3) to fully describe our setting and all necessary details of such. In addition, for our experimental results we present the descriptions of all simulations run in Appendix C. Finally, within the supplementary material we have also included a folder which contains the source code used to generate the figures seen throughout the paper.

References

I. Arieli and Y. Babichenko. Random extensive form games. J. Econ. Theory, 166:517–535, 2016.

W. Azizian, I. Mitliagkas, S. Lacoste-Julien, and G. Gidel. A tight and unified analysis of gradient-based methods for a whole spectrum of differentiable games. In International Conference on Artificial Intelligence and Statistics, pages 2863–2873. PMLR, 2020.

Y. Babichenko. Query complexity of approximate nash equilibria. In D. B. Shmoys, editor, Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014, pages 535–544. ACM, 2014.

J. P. Bailey and G. Piliouras. Multiplicative weights update in zero-sum games. In Proceedings of the 2018 ACM Conference on Economics and Computation, pages 321–338, 2018.

M. Bowling, N. Burch, M. Johanson, and O. Tammelin. Heads-up limit hold'em poker is solved. Science, 347(6218): 145–149, 2015.

M. Bowling, N. Burch, M. Johanson, and O. Tammelin. Heads-up limit hold'em poker is solved. Commun. ACM, 60(11): 81–88, Oct. 2017. ISSN 0001-0782. doi: 10.1145/3131284. URL https://doi.org/10.1145/3131284.

N. Brown and T. Sandholm. Safe and nested endgame solving for imperfect-information games. In The Workshops of the The Thirty-First AAAI Conference on Artificial Intelligence, Saturday, February 4-9, 2017, San Francisco, California, USA, volume WS-17 of AAAI Workshops. AAAI Press, 2017a.

N. Brown and T. Sandholm. Libratus: The superhuman AI for no-limit poker. In C. Sierra, editor, Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, pages 5226–5228. ijcai.org, 2017b.

N. Brown and T. Sandholm. Superhuman ai for heads-up no-limit poker: Libratus beats top professionals. Science, 359 (6374):418–424, 2018a.

N. Brown and T. Sandholm. Superhuman ai for heads-up no-limit poker: Libratus beats top professionals. Science, 359 (6374):418–424, 2018b.

Y. Cai and C. Daskalakis. On minmax theorems for multiplayer games. In D. Randall, editor, Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011, pages 217–234. SIAM, 2011a. doi: 10.1137/1.9781611973082.20. URL https://doi.org/10.1137/1.9781611973082.20.

Y. Cai and C. Daskalakis. On minmax theorems for multiplayer games. In Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete algorithms, pages 217–234. SIAM, 2011b.

Y. Cai, O. Candogan, C. Daskalakis, and C. H. Papadimitriou. Zero-sum polymatrix games: A generalization of minmax. Math. Oper. Res., 41(2):648–655, 2016.

C. Daskalakis and I. Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. arXiv preprint arXiv:1807.04252, 2018a.

C. Daskalakis and I. Panageas. The limit points of (optimistic) gradient descent in min-max optimization. arXiv preprint arXiv:1807.03907, 2018b.

C. Daskalakis and C. H. Papadimitriou. On a network generalization of the minmax theorem. In S. Albers, A. Marchetti-Spaccamela, Y. Matias, S. E. Nikoletseas, and W. Thomas, editors, Automata, Languages and Programming, 36th Internatilonal Colloquium, ICALP 2009, Rhodes, Greece, July 5-12, 2009, Proceedings, Part I, volume 5556 of Lecture Notes in Computer Science, pages 423–434. Springer, 2009a. doi: 10.1007/978-3-642-02930-1_35. URL https://doi.org/10.1007/978-3-642-02930-1_35.

C. Daskalakis and C. H. Papadimitriou. On a network generalization of the minmax theorem. International Colloquium on Automata, Languages, and Programming, pages 423–434. Springer, 2009b.

C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. The complexity of computing a nash equilibrium. In J. M. Kleinberg, editor, Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006, pages 71–78. ACM, 2006.

C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng. Training gans with optimism. arXiv preprint arXiv:1711.00141, 2017.

G. Farina, C. Kroer, and T. Sandholm. Optimistic regret minimization for extensive-form games via dilated distance-generating functions. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, Advances in Neural Information Processing Systems 32, volume 32. Curran Associates, Inc., 2019a. URL https://proceedings.neurips.cc/paper/2019/file/b030afbb3a8af8fb0759241c97466ee4-Paper.pdf.

G. Farina, C. K. Ling, F. Fang, and T. Sandholm. Ef cient regret minimization algorithm for extensive-form correlated equilibrium. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, editors, Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, pages 5187–5197, 2019b.

G. Farina, A. Celli, A. Marchesi, and N. Gatti. Simple uncoupled no-regret learning dynamics for extensive-form correlated equilibrium. CoRR, abs/2104.01520, 2021. URL https://arxiv.org/abs/2104.01520.

Y. Gao, C. Kroer, and D. Goldfarb. Increasing iterate averaging for solving saddle-point problems. arXiv preprint arXiv:1903.10646, 2019.

N. Golowich, S. Pattathil, and C. Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. arXiv preprint arXiv:2010.13724, 2020.

E. Hazan. Introduction to online convex optimization. CoRR, abs/1909.05207, 2019. URL http://arxiv.org/abs/1909.05207.

S. Hoda, A. Gilpin, J. Peña, and T. Sandholm. Smoothing techniques for computing nash equilibria of sequential games. Math. Oper. Res., 35(2):494–512, 2010a.

S. Hoda, A. Gilpin, J. Pena, and T. Sandholm. Smoothing techniques for computing nash equilibria of sequential games. Mathematics of Operations Research, 35(2):494–512, 2010b.

M. Jain, D. Korzhyk, O. Vaněk, V. Conitzer, M. Pechoucek, and M. Tambe. A double oracle algorithm for zero-sum security games on graphs. The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1, pages 327–334, 2011.

M. J. Kearns, M. L. Littman, and S. P. Singh. Graphical models for game theory. Proceedings of the 17th Conference in Uncertainty in Arti cial Intelligence, UAI '01, page 253–260, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc. ISBN 1558608001.

D. Koller and N. Megiddo. The complexity of two-person zero-sum games in extensive form. Games and economic behavior, 4(4):528–552, 1992.

C. Kroer, G. Farina, and T. Sandholm. Smoothing method for approximate extensive-form perfect equilibrium. arXiv preprint arXiv:1705.09326, 2017.

C. Kroer, G. Farina, and T. Sandholm. Solving large sequential games with the excessive gap technique. In S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 872–882, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/e836d813fd184325132fca8edcdfb40e-Abstract.html.

A. Y. Kruger. Error bounds and metric subregularity. *Optimization*, 64(1):49–79, 2015.

H. W. Kuhn. Simplified two-person poker. *Contributions to the Theory of Games*, 1:97–103, 1950a.

H. Kuhn. Extensive form games. *Proceedings of National Academy of Sciences*, pages 570–576, 1950b.

H. W. Kuhn and A. W. Tucker. *Contributions to the Theory of Games*, volume 2. Princeton University Press, 1953.

M. Lanctot, K. Waugh, M. Zinkevich, and M. H. Bowling. Monte carlo sampling for regret minimization in extensive games. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada*, pages 1078–1086. Curran Associates, Inc., 2009.

C. Lee, C. Kroer, and H. Luo. Last-iterate convergence in extensive-form games. *CoRR*, abs/2106.14326, 2021a. URL https://arxiv.org/abs/2106.14326.

C.-W. Lee, C. Kroer, and H. Luo. Last-iterate convergence in extensive-form games. *arXiv preprint arXiv:2106.14326*, 2021b.

S. Leonardos and G. Piliouras. Exploration-exploitation in multi-agent learning: Catastrophe theory meets game theory. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 11263–11271. AAAI Press, 2021.

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

T. Lin, Z. Zhou, P. Mertikopoulos, and M. Jordan. Finite-time last-iterate convergence for multi-agent learning in games. In *International Conference on Machine Learning*, pages 6161–6171. PMLR, 2020.

H. B. McMahan, G. J. Gordon, and A. Blum. Planning in the presence of cost functions controlled by an adversary. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 536–543, 2003.

P. Mertikopoulos, B. Lecouat, H. Zenati, C.-S. Foo, V. Chandrasekhar, and G. Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. *arXiv preprint arXiv:1807.02629*, 2018a.

P. Mertikopoulos, C. H. Papadimitriou, and G. Piliouras. Cycles in adversarial regularized learning. In A. Czumaj, editor, *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 2703–2717. SIAM, 2018b.

M. Moravcík, M. Schmid, N. Burch, V. Lisý, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, and M. Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.

D. Morrill, R. D'Orazio, R. Sarfati, M. Lanctot, J. R. Wright, A. R. Greenwald, and M. Bowling. Hindsight and sequential rationality of correlated play. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 5584–5594. AAAI Press, 2021. URL https://ojs.aaai.org/index.php/AAAI/article/view/16702.

J. Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

G. Palaiopanos, I. Panageas, and G. Piliouras. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. In I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5872–5882, 2017.

J. Pérolat, R. Munos, J. Lespiau, S. Omidsha ei, M. Rowland, P. A. Ortega, N. Burch, T. W. Anthony, D. Balduzzi, B. D. Vylder, G. Piliouras, M. Lanctot, and K. Tuyls. From poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 8525–8535. PMLR, 2021.

J. Perolat, R. Munos, J.-B. Lespiau, S. Omidsha ei, M. Rowland, P. Ortega, N. Burch, T. Anthony, D. Balduzzi, B. De Vylder, et al. From poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization. In *International Conference on Machine Learning*, pages 8525–8535. PMLR, 2021.

G. Piliouras and J. S. Shamma. Optimization despite chaos: Convex relaxations to complex limit sets via poincaré recurrence. In C. Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 861–873. SIAM, 2014. doi: 10.1137/1.9781611973402.64. URL https://doi.org/10.1137/1.9781611973402.64.

L. D. Popov. A modi cation of the arrow-hurwicz method for search of saddle points. *Mathematical notes of the Academy of Sciences of the USSR*, 28(5):845–848, 1980.

A. Rakhlin and K. Sridharan. Optimization, learning, and games with predictable sequences. In C. J. C. Burges, L. Bottou, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 3066–3074, 2013a.

A. Rakhlin and K. Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019. PMLR, 2013b.

M. Rowland, S. Omidsha ei, K. Tuyls, J. Pérolat, M. Valko, G. Piliouras, and R. Munos. Multiagent evaluation under incomplete information. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 12270–12282, 2019.

R. Selten. Spieltheoretische behandlung eines oligopolmodells mit nachfrageträgheit: Teil i: Bestimmung des dynamischen preisgleichgewichts. *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics*, pages 301–324, 1965.

Y. Shoham and K. Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

O. Tammelin, N. Burch, M. Johanson, and M. Bowling. Solving heads-up limit texas hold'em. In Q. Yang and M. J. Wooldridge, editors, *Proceedings of the Twenty-Fourth International Joint Conference on Arti cial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 645–652. AAAI Press, 2015.

E. Vlatakis-Gkaragkounis, L. Flokas, and G. Piliouras. Poincaré recurrence, cycles and spurious equilibria in gradient-descent-ascent for non-convex non-concave zero-sum games. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 10450–10461, 2019.

B. Von Stengel. Ef cient computation of behavior strategies. *Games and Economic Behavior*, 14(2):220–246, 1996.

B. Von Stengel and F. Forges. Extensive-form correlated equilibrium: De nition and computational complexity. *Mathematics of Operations Research*, 33(4):1002–1022, 2008.

C. Wei, C. Lee, M. Zhang, and H. Luo. Linear last-iterate convergence in constrained saddle-point optimization. In 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021 OpenReview.net, 2021a. URL https://openreview.net/forum?id=dx11_7vm5_r .

C.-Y. Wei, C.-W. Lee, M. Zhang, and H. Luo. Linear last-iterate convergence in constrained saddle-point optimization. arXiv preprint arXiv:2006.09517, 2020.

C.-Y. Wei, C.-W. Lee, M. Zhang, and H. Luo. Last-iterate convergence of decentralized optimistic gradient descent/ascent in in nite-horizon competitive markov games. arXiv preprint arXiv:2102.04540, 2021b.

M. Zinkevich, M. Johanson, M. Bowling, and C. Piccione. Regret minimization in games with incomplete information. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, Advances in Neural Information Processing Systems volume 20. Curran Associates, Inc., 2008. URL https://proceedings.neurips.cc/paper/2007/file/08d98638c6fcd194a4b1e6992063e944-Paper.pdf .

APPENDIX

A ADDITIONAL RELATED WORK

The related works presented in Section 1 are primarily focused on research which is directly related to our topic of study, namely network generalizations of zero-sum extensive form games. However, there is a large body of work which studies many adjacent areas of interest.

Extensive Form Games. As elucidated in the main text, extensive form games are widely studied due to their numerous applications. The problem of computing Nash Equilibria in extensive form games is of major interest, with several works utilizing techniques such as CFR methods (Zinkevich et al., 2008), LP methods (Shoham and Leyton-Brown, 2008) and double-oracle algorithms (McMahan et al., 2003; Jain et al., 2011). Of particular note is the success of works utilizing CFR-based algorithms to study poker variants (Brown and Sandholm, 2018b; Bowling et al., 2015; Moravčík et al., 2017). Two-player EFGs can be written in sequence form (as described in the main text), which allows for them to be written as bilinear saddle-point problems. This connection allows for the design of algorithms that utilize rst order methods to achieve approximate convergence to the Nash (Kroer et al., 2017; Gao et al., 2019).

Online Learning in Games. In this paper we study the properties of a particular online learning algorithm, Optimistic Gradient Ascent, for network zero-sum extensive form games. In normal form zero-sum games, recent results have shown that algorithms such as Gradient Descent Ascent and Multiplicative Weights Update do not converge in the last-iterate sense, even in the simplest of instances (Bailey and Piliouras, 2018; Vlatakis-Gkaragkounis et al., 2019). In contrast, optimistic variants of these algorithms have been shown to be effective in guaranteeing last-iterate convergence (Daskalakis et al., 2017; Daskalakis and Panageas, 2018a). As described in the main text, some of these results have been extended to two-player extensive form games. Speci cally, optimistic gradient descent and multiplicative weights update, as well as the versions thereof with dilated regularizers, have been studied by Lee et al. (2021b) and Wei et al. (2020) in the two-player setting. This line of research into extensive form games is not limited to discrete time algorithms. Perolat et al. (2021) show that a continuous learning dynamic known as Follow the Regularized Leader (FTRL) exhibits last-iterate convergence in monotone two-player zero-sum EFGs.

B OMITTED PROOFS

B.1 PROOF OF LEMMA 1

We rst describe how a behavioral plan can be transformed to a vector. For any x_i we let $x_i(h) := \prod_{(h,h') \in P(h) \setminus X_i} x_i(h, h')$ where h' is the action $\in A(h)$ such that $h' = \text{Next}(h,)$. We set $x_i(h) := 1$ for all $h \in H_i$ with $\text{Prev}(h, i) = \emptyset$. Notice that by de nition $U_1( ) = \sum_{z \in Z} x_1(z)\, p_1(z)\, x_2(z) = x_1^\top A_1 x_2$ and respectively $U_2( ) = \sum_{z \in Z} x_2(z)\, p_2(z)\, x_1(z) = x_2^\top A_2 x_1$.

Up next we show that all the constraints are satisfied. Consider the arbitrary state $h \in H_i$ and the states $h' \in \text{Next}(h, \sigma_i)$ for some $\sigma \in A(h)$. Notice that for each $h' \in \text{Next}(h, \sigma_i)$, $x_i(h') = x_i(h)\,\pi_i(h, \sigma)$. This implies that $\sum_{\sigma \in A(h)} x_i(\text{Next}(h, \sigma_i)) = x_i(h)$ since $\sum_{\sigma \in A(h)} \pi_i(h, \sigma) = 1$.

Now let $h_1, h_2 \in H_i$ where $h_1 \in \text{Next}(h_1', a_i)$, $h_2 \in \text{Next}(h_2', a_i)$ and $I(h_1) = I(h_2)$. Consider the set $P(h_1) \cap X_i := \{p_1, \ldots, p_k, h_1\}$ and $P(h_2) \cap X_i := \{q_1, \ldots, q_k, h_2\}$. Due to the perfect recall property, $k = k$ and $I(p_\ell) = I(q_\ell)$. Thus, $x_i(h_1) = x_i(h_2)$.

Up next we show how a vector $x_i \in X_i$ can be converted to a behavioral plan $\pi_i \in \Pi_i$. Let $\pi_i(h, \sigma) := \frac{x_i(h')}{x(h)}$ for some $h' \in \text{Next}(h, \sigma_i)$. Notice that due the third constraint $x_i(h') = x_i(h'')$ for all $h', h'' \in \text{Next}(h, \sigma_i)$ and thus $\pi_i(h, \sigma)$ is well-defined. For $h \in H_i$ let $h' \in \text{Next}(h, \sigma_i)$. By the third constraint we get that $\sum_{\sigma \in A(h)} \pi_i(h, \sigma) = 1$. Finally let $h_1, h_2 \in H_i$ with $I(h_1) = I(h_2)$ then $\pi_i(h_1, \sigma) = \frac{x_i(h_1')}{x_i(h_1)}$ for some $h_1' \in \text{Next}(h_1, \sigma_i)$ and $\pi_i(h_2, \sigma) = \frac{x_i(h_2')}{x_i(h_2)}$ for some $h_2' \in \text{Next}(h_2, \sigma_i)$. As a result, by the second constraint we get that $\pi_i(h_1, \sigma) = \pi_i(h_2, \sigma)$ for all $\sigma \in A(h)$.

B.2 Proof of Lemma 2

We first describe how a behavioral plan $\pi_u \in \Pi_u$ can be transformed to a vector $x_u(h) \in X_u$. If there exists a game Γ^{uv} with $(u, v) \in E$ such that $\text{Prev}^{uv}(h, u) = \emptyset$ we set $x_u(h) := 1$. Let us first verify that the above assignment is valid i.e. if $\text{Prev}^{uv}(h, u) = \emptyset$ for some $(u, v) \in E$ then $\text{Prev}^{uv'}(h, u) = \emptyset$ for all $(u, v') \in E$. Notice that $P^{uv}(h) \cap X_u = \{h\}$ and thus by the second constraint of Definition 13, $P^{uv'}(h) \cap X_u = \{h\}$ for all $(u, v') \in E$. Now for the remaining nodes $h \in H_u$ we select an arbitrary two-player EFG Γ^{uv} ($(u, v) \in E$) containing the state h and set $x_u(h) := \sum_{(h, h') \in P^{uv}(h) \cap X_u} \pi_u(h, \sigma_{h'})$ where $\sigma_{h'}$ is the action $\sigma \in A(h)$ such that $h' = \text{Next}^{uv}(h, \sigma, u)$. We again need to argue that $x_u(h)$ is independent of the arbitrary choice of the game Γ^{uv}. Let assume that state h also belongs in the two-player EFG $\Gamma^{uv'}$ for some $(u, v') \in E$. Again by the second constraint of Definition 11 we know that for the sets $P^{uv}(h) \cap X_u = \{p_1, \ldots, p_k, h\}$ and $P^{uv'}(h) \cap X_u = \{q_1, \ldots, q_m, h\}$ the following holds:

1. $k = m$.
2. $I(p_\ell) = I(q_\ell)$ for all $\ell \in \{1, \ldots, k\}$.
3. $p_{\ell+1} \in \text{Next}^{uv}(p_\ell, \sigma, u)$ and $q_{\ell+1} \in \text{Next}^{uv'}(q_\ell, \sigma, u)$ for some action $\sigma \in A(p_\ell)$.

Since $I(p_\ell) = I(q_\ell)$ means that $\pi_u(p_\ell, \sigma) = \pi_u(q_\ell, \sigma)$ for all $\sigma \in A(p_\ell) = A(q_\ell)$, we get that

$$\sum_{(h, h') \in P^{uv}(h) \cap X_u} \pi_u(h, \sigma_{h'}) = \sum_{(h, h') \in P^{uv'}(h) \cap X_u} \pi_u(h, \sigma_{h'})$$

Conversely, we show how a strategy in sequence form $x_u \in X_u$ can be converted to behavioral plan $\pi_u \in \Pi_u$. Given a state $h \in H_u$ we consider an edge $(u, v) \in E$ such that Γ^{uv} containing $h \in H_u$ and set

$$\pi_u(h, \sigma) := \frac{x_u(h')}{x_u(h)} \quad \text{for some } h' \in \text{Next}^{uv}(h, \sigma, u)$$

We first need to show that this is a valid probability distribution, $\sum_{\sigma \in A(h)} \pi_u(h, \sigma) = 1$. Since $x_u \in X_u^{uv}$, the second constraint of Definition 7 ensures that

$$\sum_{\sigma \in A(h)} x_u(\text{Next}(h, \sigma, u)) = x_u(h)$$

The latter implies that $\sum_{\sigma \in A(h)} \pi_u(h, \sigma) = 1$.

We now need to establish that $\pi_u(h, \sigma)$ is independent of the selection of the edge $(u, v) \in E$. Let h be a state of the game $\Gamma^{uv'}$ for some $(u, v') \in E$. By constraint 2 of Definition 13, for any $h' \in \text{Next}^{uv}(h, \sigma, u)$ and $h'' \in \text{Next}^{uv'}(h, \sigma, u)$ we have that $x_u(h') = x_u(h'')$ and thus $\pi_u(h, \sigma) = \frac{x_u(h'')}{x_u(h)}$.

Finally we need to argue that if $h_1, h_2 \in H_u$ with $I(h_1) = I(h_2)$, then $\pi_u(h_1, \sigma) = \pi_u(h_2, \sigma)$ for all $\sigma \in A(h_1) = A(h_2)$. Let $\pi_u(h_1, \sigma) = \frac{x_u(h_1')}{x_u(h)}$ for some $h_1' \in \text{Next}(h_1, \sigma, u)$ and $\pi_u(h_1, \sigma) = \frac{x_u(h_2')}{x_u(h)}$ for some $h_2' \in \text{Next}(h_2, \sigma, u)$. Then by Constraint 3 of Definition 11 we get that $x(h_1') = x(h_2')$ and thus $\pi_u(h_1, \sigma) = \pi_u(h_2, \sigma)$.

B.3 Proof of Lemma 3

First, since Equations (6), (7) are defined on the product of treeplexes X, let us decompose the equations from the perspective of an arbitrary agent u. Specifically, for some x_u^t, $u \in \{1, \ldots, n\}$ it holds that the inner product $\langle x, R\, x^{t-1} \rangle$, $x \in X$ can be decomposed into inner products of the form $\langle x_u, R_u\, x_u^{t-1} \rangle$, where x_u is now in the individual treeplex X_u. Moreover, by the definition of matrix R, we can substitute the following:

$$R_{(u:h_1),(v:h_2)} = [A^{uv}]_{h_1 h_2}$$

for all $(u, v) \in E$ and 0 otherwise. Effectively, from the perspective of player u, the product of R and x^t gives us $\sum_{(u,v)\in E} A^{u,v}\, x_v^t$. This gives us the following:

$$x_u^t = \operatorname{argmin}_{x \in X_u} \left\{ \left\langle x, \sum_{v:(u,v)\in E} A^{uv}\, x_v^{t-1} \right\rangle + D(x, \hat{x}_u^t) \right\} \tag{8}$$

$$\hat{x}_u^{t+1} = \operatorname{argmin}_{x \in X_u} \left\{ \left\langle x, \sum_{v:(u,v)\in E} A^{uv}\, x_v^{t} \right\rangle + D(x, \hat{x}_u^t) \right\} \tag{9}$$

Finally, we can just take the negative of the terms inside the braces to obtain Equations (4), (5). Hence, for every strategy vector x updated using Equations (6),(7), the constituent strategy vectors for each player are exactly the same as Equations (4), (5). Thus if the initial conditions x^0 are the same, for all time t the collection of strategy vectors $\{x^1, \ldots, x^T\}$ are the same between both formulations.

B.4 Proof of Lemma 4

Let $\hat{x} := (\hat{x}_1, \ldots, \hat{x}_n)$ be an ϵ-symmetric Nash Equilibrium. Now consider the vector $x^0 \in X$ defined as follows: $x_{u'} = \hat{x}_{u'}$ for all $u' \neq u$ and x_u^0 is an arbitrary vector in X_u. By the definition of the ϵ-symmetric Nash Equilibrium we get that

$$\hat{x}^\top R\, \hat{x} - (x^0)^\top R\, \hat{x} \leq \epsilon$$

Notice that $(x^0)^\top R\, \hat{x} = \sum_{v:(u,v)\in E} (x_u^0)^\top A^{uv}\, \hat{x}_v + \sum_{u' \neq u} \sum_{v:(u',v)\in E} \hat{x}_{u'}^\top A^{u'v}\, \hat{x}_v$. Thus we get

$$\sum_{v:(u,v)\in E} (\hat{x}_u)^\top A^{uv}\, \hat{x}_v - \sum_{v:(u,v)\in E} (x_u^0)^\top A^{uv}\, \hat{x}_v \leq \epsilon \quad \text{for all } x_u \in X_u$$

Theorem 1 follows by repeating the same argument for all agents $u \in V$.

B.5 Proof of Lemma 5

We first prove a simpler version of Lemma 5 where $x = y \in X$.

Lemma 7. $x^\top R\, x = 0$ for all $x \in X$.

Proof. Consider a vector $x \in X$. To simplify notation let $x := (x_1, \ldots, x_n)$ where each vector $x_u \in X_u$. Let $\sigma_u^x \in \Sigma_u$ denote the behavioral plan for agent u constructed from the vector $x_u \in X_u$ as described in Lemma 2. By the zero-sum property of Definition 12, we get that

$$\sum_{u \in V} \sum_{v:(u,v)\in E} U_u^{uv}(\sigma_u^x, \sigma_v^x) = 0$$

By Lemma 2 we get that $U^u(\sigma^x) = \sum_{v:(u,v)\in E} U_u^{uv}(\sigma_u^x, \sigma_v^x) = \sum_{v:(u,v)\in E} x_u^\top A^{uv}\, x_v$ meaning that

$$\sum_{u \in V} \sum_{v:(u,v)\in E} x_u^\top A^{uv}\, x_v = 0$$

As a result, we get that $x^\top R\, x = 0$. □

We will also utilize the following result:

Lemma 8. Consider a node $u \in V$ and its neighbors $N_u = \{v_1, v_2, \ldots, v_k\}$. Let $x_u \in X_u$ represent a mixed strategy for u and x_v a mixed strategy of the neighbor $v \in N_u$. For any fixed collection $\{x_v\}_{v \in N_u}$ the quantity

$$\sum_{v \in N_u} x_v^\top A^{vu} x_u + \sum_{v \in N_u} x_u^\top A^{uv} x_v$$

remains constant over the range of x_u.

Proof. For any vector $x := (x_1, \ldots, x_n) \in X$, consider the vector $x^0 \in X$ such that $x_v^0 = x_v$ for all $v \neq u$. By Lemma 7 we get that

$$x^\top R x - (x^0)^\top R x^0 = 0$$

The latter directly implies that

$$\sum_{v \in N_u} x_v^\top A^{vu} x_u + \sum_{v \in N_u} x_u^\top A^{uv} x_v = \sum_{v \in N_u} x_v^\top A^{vu} x_u^0 + \sum_{v \in N_u} (x_u^0)^\top A^{uv} x_v$$

for all $x_u, x_u^0 \in X_u$. □

Proof of Lemma 5. Consider vectors $x, y \in X$. Consider the vector $y^0 \in X$ such that $y_v = y_v^0$ for all $v \neq u$. We first show that

$$x^\top R y + y^\top R x = x^\top R y^0 + (y^0)^\top R x$$

Let N_u denote the neighbors of agent $u \in V$,

$$\begin{aligned}
x^\top R y + y^\top R x - x^\top R y^0 - (y^0)^\top R x &= \sum_{v \in N_u} x_v^\top A^{vu} y_u + \sum_{v \in N_u} x_u^\top A^{uv} y_v \\
&+ \sum_{v \in N_u} y_v^\top A^{vu} x_u + \sum_{v \in N_u} y_u^\top A^{uv} x_v \\
&- \sum_{v \in N_u} x_v^\top A^{vu} y_u^0 - \sum_{v \in N_u} x_u^\top A^{uv} y_v^0 \\
&- \sum_{v \in N_u} (y_v^0)^\top A^{vu} x_u - \sum_{v \in N_u} (y_u^0)^\top A^{uv} x_v \\
&= \sum_{v \in N_u} x_v^\top A^{vu} y_u - \sum_{v \in N_u} x_u^\top A^{uv} y_v \\
&- \sum_{v \in N_u} x_v^\top A^{vu} y_u^0 + \sum_{v \in N_u} (y_u^0)^\top A^{uv} x_v \\
&= 0
\end{aligned}$$

where the last equality follows by Lemma 8. By gradually transforming vector y to vector x we get that $x^\top R y + y^\top R x = 2\, x^\top R x = 0$. □

B.6 Proof of Lemma 6

Applying Lemma 1 of Rakhlin and Sridharan (2013b) to our setting, we obtain:

Lemma 9 (Rakhlin and Sridharan (2013b)). Let $\{x^t, \hat{x}^t\}$ be the sequences produced by Equations (6),(7). Then,

$$\sum_{t=1}^T (x^t)^\top R x^t - \min_{x \in X} \sum_{t=1}^T x^\top R x^t \leq \frac{D^2}{\eta} + \frac{1}{2} \sum_{t=1}^T \eta \| R x^t - R x^{t-1} \|^2 + \frac{1}{2} \sum_{t=1}^T \frac{1}{\eta} \| x^t - \hat{x}^t \|^2 - \frac{1}{2} \sum_{t=1}^T \frac{1}{\eta} \| \hat{x}^t - x^t \|^2 + \| \hat{x}^t - x^{t+1} \|^2$$

where D is the diameter of the treeplex polytope X.

Setting $\eta = \min\{1/(8\sqrt{2}kRk^2), 1\}$ in Lemma 9 we get that

$$\sum_{t=1}^{T}(x^t)^\top R\, x^t - \min_{x\in X}\sum_{t=1}^{T} x^\top R\, x^t \le \frac{D^2}{\eta} + \frac{1}{2}\sum_{t=1}^{T} kR\, x^t - R\, x^{t-1}k^2 - \frac{1}{4}\sum_{t=1}^{T} k\hat{x}^t - x^t k^2 + k\hat{x}^t - x^{t+1}k^2$$

$$\le \frac{D^2}{\eta} + \frac{1}{2}\sum_{t=1}^{T} kR\, x^t - R\, x^{t-1}k^2 - 2kRk^2\sum_{t=1}^{T} k\hat{x}^t - x^t k^2 + k\hat{x}^t - x^{t+1}k^2$$

$$\le \frac{D^2}{\eta} + \frac{kRk^2}{2}\sum_{t=1}^{T} kx^t - x^{t-1}k^2 - kRk^2\sum_{t=1}^{T} kx^t - x^{t-1}k^2$$

$$\le \frac{D^2}{\eta}$$

Setting $\hat{x} = \sum_{s=1}^{T} x^s/T$ and using the fact that $(x^t)^\top R\, x^t = 0$ we get $\min_{x\in X} x^\top R\, \hat{x} \le \frac{D^2 kRk^2}{T}$.

B.7 Proof of Theorem 1

Let $\hat{x}$ the time-average vector produced by Equations (6),(7). By Lemma 6, we have

$$\min_{x\in X} x^\top R\, \hat{x} \le \frac{D^2 kRk^2}{T}$$

Using the fact that $\hat{x}^\top R\, \hat{x} = 0$ we get that

$$\hat{x}^\top R\, \hat{x} - \min_{x\in X} x^\top R\, \hat{x} \le \frac{D^2 kRk^2}{T}$$

meaning that $(\hat{x}, \hat{x})$ is a $\frac{D^2 kRk^2}{T}$-approximate symmetric Nash Equilibrium of the symmetric game (R, R). By Lemma 4 we get that $\hat{x}$ is a $\frac{D^2 kRk^2}{T}$-approximate NE for the original network zero-sum EFG.

B.8 Proof of Theorem 3

First of all, in the proof of this theorem and in the lemmas presented within the proof, let $X^* = \{x \in X : \min_{x\in X} x^\top R\, x = 0\}$, which describes the set of symmetric Nash Equilibria.

In order to establish Theorem 3, we follow the approach and notation of Wei et al. (2020), with minor modifications along the way to apply the steps to our setting. Applying Lemma 1 of Wei et al. (2020)) to the Equations (6), (7) we get the following lemma:

Lemma 10 (Wei et al. (2020)). Let $\{x^t; \hat{x}^t\}_{t\ge 1}$ be the sequence of strategy vectors produced by Equations (6), (7) for $\eta \le 1/8kRk^2$. Then,

$$(R\, x^t)^\top (x^t - x) \le D_\psi(x, \hat{x}^t) - D_\psi(x, \hat{x}^{t+1}) - D_\psi(\hat{x}^{t+1}, x^t) \le \frac{15}{16}D_\psi(x^t, \hat{x}^t) + \frac{1}{16}D_\psi(\hat{x}^t, x^{t-1})$$

Since for OGD we have that $D_\psi(x) = \frac{1}{2}kxk^2$, we can write the above inequality as:

$$2\,(R\, x^t)^\top (x^t - x) \le k\hat{x}^t - xk^2 - k\hat{x}^{t+1} - xk^2 \le k\hat{x}^{t+1} - x^t k^2 - \frac{15}{16}kx^t - \hat{x}^t k^2 + \frac{1}{16}k\hat{x}^t - x^{t-1}k^2 \quad (10)$$

To simplify notation let $x^* := \pi_X(\hat{x}^t) \in X^*$ meaning that x^* is a symmetric Nash Equilibrium for the symmetric game (R, R) and let us apply Equation 10 with $x = x^*$. Now the LHS of Equation 10 takes the following form

$$\begin{aligned} 2\,(x^t)^\top R^T (x^t - x^*) &= 2\,(x^t)^\top R^T x^* \quad ((x^t)^\top R^T x^t = 0) \\ &= 2\,(x^*)^\top R\, x^t \\ &= 2\,(x^t)^\top R\, x^* \quad \text{(by Lemma 5)} \\ &\le 0 \end{aligned}$$

where the last inequality follows by the fact that $(\hat{x}^t; x)$ is a symmetric Nash Equilibrium of the game $(R; R)$. Since the LHS of Equation 10 is greater or equal to 0 we get that,

$$\|\hat{x}^{t+1} - X(\hat{x}^t)\|^2 + \|\hat{x}^t - X(\hat{x}^t)\|^2 + \|\hat{x}^t - \hat{x}^{t+1}\| \leq \|x^t - X(\hat{x}^t)\|^2 + \|\hat{x}^t - \hat{x}^{t+1}\| \leq \|x^t - X(\hat{x}^t)\|^2 + \frac{15}{16}\|x^t - \hat{x}^t\|^2 + \frac{1}{16}\|\hat{x}^t - x^{t-1}\|^2$$

By definition, the left hand side of the above is bounded below by $\text{dist}^2(\hat{x}^{t+1}; X)$. Thus, we have the following inequality,

$$\text{dist}^2(\hat{x}^{t+1}; X) + \text{dist}^2(\hat{x}^t; X) + \|\hat{x}^t - \hat{x}^{t+1}\| \leq \|x^t - X(\hat{x}^t)\|^2 + \frac{15}{16}\|x^t - \hat{x}^t\|^2 + \frac{1}{16}\|\hat{x}^t - x^{t-1}\|^2 \tag{11}$$

Now, we define $\Phi^t := \|x^t - X(\hat{x}^t)\|^2 + \frac{1}{16}\|\hat{x}^t - x^{t-1}\|^2$ and $\Psi^t := \|\hat{x}^{t+1} - x^t\|^2 + \|x^t - \hat{x}^t\|^2$ and rewrite Equation 11 as follows,

$$\Phi^{t+1} \leq \Phi^t - \frac{15}{16}\Psi^t \tag{12}$$

As in Wei et al. (2020), we now lower bound Ψ^t by a quantity related to $\text{dist}^2(\hat{x}^{t+1}; X)$ which will then give us a convergence rate for Φ^t. To do so we need to establish a property that is known as saddle-point metric subregularity (Wei et al. (2020)).

Lemma 11. (Saddle-Point Metric Subregularity (SP-MS)) For any $x \in X$ and $x^0 \in X \setminus X^*$,

$$\sup_{x^0 \in X} \frac{(R - x)^\top (x - x^0)}{\|x - x^0\|} \leq c\,\|x - X(x)\|$$

for some game-dependent constant $c > 0$.

We present the proof of Lemma 11 in Section B.9. To this end, we remark that once the proof of Lemma 11 is established, the proof of Theorem 3 follows by the analysis of Wei et al. (2020). For the sake of completeness, we conclude the section with this analysis.

Lemma 12 (Wei et al. (2020)). If the parameter η in Equations (6), (7) is selected less than $\frac{1}{4\sqrt{2}\|R\|}$ then for any $t \geq 0$ and $x^0 \in X$ with $x^0 \neq \hat{x}^{t+1}$,

$$\|\hat{x}^{t+1} - x^t\|^2 + \|x^t - \hat{x}^t\|^2 \geq \frac{32}{81\eta^2}\frac{\left[(R\hat{x}^{t+1})^\top (\hat{x}^{t+1} - x^0)\right]_+^2}{\|\hat{x}^{t+1} - x^0\|^2}$$

where $[a]_+ := \max\{a, 0\}$, and similarly, for $x^0 \neq x^{t+1}$,

$$\|\hat{x}^{t+1} - x^{t+1}\|^2 + \|x^t - \hat{x}^{t+1}\|^2 \geq \frac{32}{81\eta^2}\frac{\left[(R\,x^{t+1})^\top (x^{t+1} - x^0)\right]_+^2}{\|x^{t+1} - x^0\|^2}$$

Now taking the telescoping sum of Equation 12 over t, we get:

$$\Phi^1 - \Phi^{T+1} \geq \frac{15}{16}\sum_{t=1}^{T} \Phi^t \geq \frac{15}{16}\sum_{t=1}^{T} \|\hat{x}^{t+1} - x^t\|^2 + \|x^t - \hat{x}^t\|^2 \geq \frac{15}{32}\sum_{t=2}^{T} \|x^t - x^{t-1}\|^2$$

where the final inequality follows due to strong convexity of $\frac{1}{2}\|x\|^2$. Now, since the rightmost term is a summation of nonnegative terms and is upper bounded by a finite constant, we have that $\|x^t - x^{t-1}\| \to 0$ as $T \to \infty$. Thus, x^t converges to a point as $T \to \infty$. In addition, due to Theorem 1, we know that the time-average value of the iterates converge to a Nash Equilibrium. Combining these two observations, we can thus conclude that x^t indeed converges to a Nash Equilibrium in the last-iterate sense.

To show the explicit rate of convergence, we will require a few additional observations. First, note that the following inequality holds for Equation 12:

$$\|\hat{x}^{t+1} - x^t\|^2 \geq \Phi^t \geq \frac{16}{15}\Phi^{t+1} \geq \left(\frac{16}{15}\right)^{-1} \tag{13}$$

Then we have:

$$
\begin{aligned}
{}^{t} \quad & \frac{1}{2}\|\hat{x}^{t+1} - x^{t}\|^2 + \frac{1}{2}(\|\hat{x}^{t+1} - x^{t}\|^2 + \|x^{t} - \hat{x}^{t}\|^2) \\
& \frac{1}{2}\|\hat{x}^{t+1} - x^{t}\|^2 + \frac{16\ ^2}{81} \sup_{x' \in X} \frac{(R\ \hat{x}^{t+1})^{>} (\hat{x}^{t+1} - x')^{2}_{+}}{\|\hat{x}^{t+1} - x'\|^2} && \text{(Lemma 12)} \\
& \frac{1}{2}\|\hat{x}^{t+1} - x^{t}\|^2 + \frac{16\ ^2 C^2}{81}\|\hat{x}^{t+1} - {}_{X}(\hat{x}^{t+1})\|^2 && \text{(SP-MS condition)} \\
& \min\left\{\frac{16\ ^2 C^2}{81}, \frac{1}{2}\right\} \|\hat{x}^{t+1} - x^{t}\|^2 + \|x^{t} - \hat{x}^{t+1}\|^2 - {}_{X}(\hat{x}^{t+1})\|^2 && \text{(Equation 13)} \\
= \ & C_2\ {}^{t+1}
\end{aligned}
$$

Now, we can show the explicit convergence rate as follows. Combining the above inequality with Equation 12, we obtain

$$
{}^{t+1} \quad {}^{t} \quad C_2\ {}^{t+1} \tag{14}
$$

This immediately implies that ${}^{t+1} \quad (1 + C_2)^{\ 1}\ {}^{t}$. By iteratively expanding the right hand side of the inequality, we can equivalently write:

$$
{}^{t} \quad (1 + C_2)^{\ t+1\ \ 1} \quad 2^{\ 1}(1 + C_2)^{\ t} \tag{15}
$$

Next, notice that ^{1} is precisely $\mathrm{dist}^2(\hat{x}^1; X)$. Moreover, by using the triangle inequality, we can write:

$$
\begin{aligned}
\mathrm{dist}^2(x^t; X) \quad & \|x^t - {}_{X}(\hat{x}^{t+1})\|^2 \\
& 2\|\hat{x}^{t+1} - {}_{X}(\hat{x}^{t+1})\|^2 + 2\|\hat{x}^{t+1} - x^t\|^2 \\
& 32\ {}^{t+1} \quad 32\ {}^{t}
\end{aligned}
$$

Combining this observation with Equation 15 we get that

$$
\mathrm{dist}^2(x^t; X) \quad 64\mathrm{dist}^2(x^1; X)(1 + C_2)^{\ t}
$$

where $C_2 = \min\left\{\frac{16\ ^2 C^2}{81}, \frac{1}{2}\right\}$, which completes the proof of Theorem 3.

B.9 Proof of Lemma 11

Lemma 11 follows easily from Lemma 13, the proof of which is presented in Section B.10.

Lemma 13. For any $x \in X$ the following holds:

$$
\min_{x' \in X} x'^{>} R\ x \quad c\ \|x - {}_{X}(x)\|: \tag{16}
$$

for some game-dependent constant $c > 0$.

Proof of Lemma 11. Consider the LHS of the inequality in Lemma 13 and note that $\min_{x' \in X} x'^{>} R x = 0$ if and only if $x \in X$.

Let D denote the diameter of $\mathcal{X}$ which is assumed to be finite. Then,

$$\begin{aligned}
\max_{x' \in \mathcal{X}} \frac{(R - x)^\top (x - x')}{\|x - x'\|} &\leq \max_{x' \in \mathcal{X}} \frac{1}{D} (R - x)^\top (x - x') \\
&= \frac{1}{D} \max_{x' \in \mathcal{X}} x^\top R^\top (x - x') \\
&= \frac{1}{D} \max_{x' \in \mathcal{X}} [x^\top R^\top x - x^\top R^\top x'] \\
&= \frac{1}{D} \max_{x' \in \mathcal{X}} [-x^\top R^\top x'] && (x^\top R^\top x = 0) \\
&= -\frac{1}{D} \min_{x' \in \mathcal{X}} x^\top R^\top x' \\
&= -\frac{1}{D} \min_{x' \in \mathcal{X}} x'^\top R x \\
&\leq \frac{c}{D} \|x - \Pi_{\mathcal{X}}(x)\| && \text{(Lemma 13)}
\end{aligned}$$

□

B.10 Proof of Lemma 13

The proof of this lemma follows the basic steps in the proof of Theorem 5 in Wei et al. (2020), with some necessary modifications. We remind the reader that for the purposes of the proof, we defined the set of symmetric Nash Equilibria as $\mathcal{X} = \{x : \min_{x' \in \mathcal{X}} x^\top R x' = 0\}$. The proof is split into several auxiliary lemmas/claims, which can then be combined to show the required result.

Lemma 14. The set $\mathcal{X}$ is a polytope.

Proof. Let $x \in \mathcal{X}$ then $\min_{x' \in \mathcal{X}} x^\top R x' = 0$. Since $\mathcal{X}$ is a polytope the minimum value is attained in one of the vertices of polytope $\mathcal{X}$, the set of which is denoted by $\mathcal{V}(\mathcal{X})$. Thus

$$\min_{x' \in \mathcal{X}} x^\top R x' = \min_{v \in \mathcal{V}(\mathcal{X})} v^\top R x = 0$$

As a result, the set $\mathcal{X}$ can be equivalently described as the set of vectors $x \in \mathcal{X}$ that additionally satisfy $v_i^\top R x \geq 0$ for all vertices $v_i \in \mathcal{V}(\mathcal{X})$. □

Let us describe $\mathcal{X}$ in the following polytopal form:

$$\mathcal{X} := \{x : a_i^\top x \geq b_i \text{ for } i = 1, \ldots, L\}$$

where L is a positive integer. Consider also the following polytopal form of the set $\mathcal{X}$ as

$$\mathcal{X} := \{x \in \mathcal{X} : c_i^\top x \geq 0 \text{ for } i = 1, \ldots, K\}$$

where $c_i := v_i^\top R$ with v_i denoting the i-th vertex of polytope $\mathcal{X}$ and K denotes the number of different vertices.

Now x a specific $x \in \mathcal{X} \setminus \mathcal{X}$ and let $x := \Pi_{\mathcal{X}}(x)$. The vector x satisfies some of the polytopal constraints with equality. These constraints are called tight, and without loss of generality we can assume that

- $a_i^\top x = b_i$ for $i = 1, \ldots, \ell$
- $c_i^\top x = 0$ for $i = 1, \ldots, k$

Lemma 15. The vector $x \in \mathcal{X}$ violates at least one tight constraint of the form $c_i^\top x = 0$ for $i = 1, \ldots, k\}$.

Proof. Let assume that $\{c_i^\top x = 0$ for $i = 1, \ldots, k\}$. Since $x \in \mathcal{X}$ there exists at least one vertex $v \in \mathcal{V}(\mathcal{X})$ such that $v^\top R x < 0$. The latter implies that there exists $x' \in \mathcal{X}$ lying in line segment between x and x such that $v^\top R x' < 0$ for all vertices $v \in \mathcal{V}(\mathcal{X})$. The latter implies that $x' \in \mathcal{X}$ which contradicts with the fact that $x = \Pi_{\mathcal{X}}(x)$. □

Now, note that the normal cone of X at x is

$$N_x = \{x^0 : x \cdot x = \max_{x \in X} (x^0) \cdot x\}$$

From a standard result in linear programming literature (Wei et al., 2020), we know that the normal cone can be written in the following form:

$$N_x = \left(\sum_{i=1}^{\ell} p_i a_i + \sum_{i=1}^{k} q_i c_i \text{ for some } p_i, q_i \geq 0 \right)$$

Again, following the steps of Wei et al. (2020), we have the following claim:

Claim 1. For any $x \in X$ such that $x = \max_{x \in X} (x)$ the vector $x \cdot x$ belongs in the set

$$M_x = \left(\sum_{i=1}^{\ell} p_i a_i + \sum_{i=1}^{k} q_i c_i : p_i, q_i \geq 0, \sum_{i=1}^{\ell} p_i a_i + \sum_{i=1}^{k} q_i c_i \neq 0 \right)$$

Proof. As mentioned previously, we know that x belongs in the normal cone of X, N_x. Thus it can be expressed as $\sum_{i=1}^{\ell} p_i a_i + \sum_{i=1}^{k} q_i c_i$ with $p_i, q_i \geq 0$. As such, we need only additionally show that x satisfies the following:

$$a_i \cdot (x \cdot x) \leq 0, \quad \forall i \in \{1, \ldots, \ell\}$$

Notice that for all $i = 1, \ldots, \ell$, we have:

$$\begin{aligned} a_i \cdot (x \cdot x) &= (a_i \cdot x - b_i) + a_i \cdot (x \cdot x) && (i\text{-th constraint is tight at } x) \\ &= a_i \cdot (x + x \cdot x) - b_i \\ &= a_i \cdot x - b_i \leq 0 && (x \in X) \end{aligned}$$

□

Claim 2. $x \cdot x$ can be written as $\sum_{i=1}^{\ell} p_i a_i + \sum_{i=1}^{k} q_i c_i$ with $0 \leq p_i, q_i \leq C^0 \|x \cdot x\|$ for all i and some problem-dependent constant $C^0 < 1$.

Proof. Note that $\frac{x \cdot x}{\|x \cdot x\|} \in M_x$ because $0 \leq x \cdot x \in M_x$ and M_x is a cone. Furthermore, $\frac{x \cdot x}{\|x \cdot x\|} \in \{v \in \mathbb{R}^M : \|v\|_1 \leq 1\}$. Thus, $\frac{x \cdot x}{\|x \cdot x\|} \in M_x \cap \{v \in \mathbb{R}^M : \|v\|_1 \leq 1\}$, which is a bounded subset of the cone M_x.

We will argue that there exists large enough $C^0 > 0$ such that:

$$\left(\sum_{i=1}^{\ell} p_i a_i + \sum_{i=1}^{k} q_i c_i : 0 \leq p_i, q_i \leq C^0, \forall i \right) \supseteq M_x \cap \{v \in \mathbb{R}^M : \|v\|_1 \leq 1\} := P.$$

First note that P is a polytope. For every vertex v of P, the smallest C^0 such that v belongs to the left-hand side set above is the solution to the following linear program:

$$\begin{aligned} \min_{p_i, q_i, C^0_v} \quad & C^0_v \\ \text{s.t.} \quad & v = \sum_{i=1}^{\ell} p_i a_i + \sum_{i=1}^{k} q_i c_i, \quad 0 \leq p_i, q_i \leq C^0_v. \end{aligned}$$

Since $v \in M_x$, this LP is always feasible and admits a finite solution $C^0_v < 1$. Now, let $C^0 = \max_{v \in V(P)}$ where $V(P)$ is the set of all vertices of P. Then, since any $v \in P$ can be expressed as a convex combination of points in $V(P)$, v can thus be expressed as $\sum_{i=1}^{\ell} p_i a_i + \sum_{i=1}^{k} q_i c_i$ where $0 \leq p_i, q_i \leq C^0$. As a result, $\frac{x \cdot x}{\|x \cdot x\|}$ can be written as $\sum_{i=1}^{\ell} p_i a_i + \sum_{i=1}^{k} q_i c_i$ where $0 \leq p_i, q_i \leq C^0$, so it follows that $x \cdot x$ can be written as: $\sum_{i=1}^{\ell} p_i a_i + \sum_{i=1}^{k} q_i c_i$ where $0 \leq p_i, q_i \leq C^0 \|x \cdot x\|$.

□

Now, again following Wei et al. (2020), we can piece together all of the auxiliary results to show Lemma 13. De ne $A_i := {}_i^{>}(x \quad x\ )$ and $C_i := c_i^{>}(x \quad x\ )$. By Claim 2, we can write $x \quad x$ as $\sum_{i=1}^{\ell} p_i \quad {}_i + \sum_{i=1}^{k} q_i \quad c_i$ where $0 \quad p_i ; q_i \quad C^0 kx \quad x\ k$. Thus:

$$\sum_{i=1}^{\ell} p_i \quad A_i + \sum_{i=1}^{k} q_i \quad C_i = \sum_{i=1}^{\ell} p_i \quad {}_i + \sum_{i=1}^{k} q_i \quad c_i \quad (x \quad x\ ) = kx \quad x\ k^2$$

Moreover, since $x \quad x\ 2 M_x$ by Claim 1, we have

$$\sum_{i=1}^{\ell} p_i \quad A_i = \sum_{i=1}^{\ell} p_i \quad {}_i \quad 0$$

and

$$\sum_{i=1}^{k} q_i \quad C_i \quad \max_{i 2f\ 1;:::;k\ g} C_i \quad \sum_{i=1}^{k} q_i \quad \max_{i 2f\ 1;:::;k\ g} C_i \quad kC^0 kx \quad x\ k$$

The rst inequality follows because $p_i \quad 0$. The second inequality follows because $\max_{i 2f\ 1;:::;k\ g} C_i > 0$ (by Lemma 15) and $0 \quad q_i \quad C^0 kx \quad x\ k$.

Combining the above, we obtain:

$$\max_{i 2f\ 1;:::;k\ g} C_i \quad \frac{1}{kC^0 kx \quad x\ k}$$

Now, note that:

$$\max_{i 2f\ 1;:::;k\ g} C_i = \max_{i 2f\ 1;:::;k\ g} (c_i^{>} x \quad d_i) \quad \max_{i 2f\ 1;:::;\ jV(X)jg} (c_i^{>} x \quad d_i) = \max_{x^0 2X} (x^{0>} Rx)$$

where the last equality follows from the formulation of problem constraints in the proof of Lemma 14. Finally, by combining the last two statements, we can conclude that

$$\min_{x^0 2X} (x^{0>} Rx) \quad \frac{1}{kC^0 kx \quad x\ k}:$$

Here k and C^0 only depend on the set of tight constraints at x. There are only nitely many sets of tight constraints, so there exists a constant $C > 0$ such that $\min_{x^0 2X} (x^{0>} Rx) \quad \frac{1}{kC^0 kx \quad x\ k}$ holds for all x and x, completing the proof.

C Additional Experimental Details

In this section we provide more details about our experimental results from Section 5.

Random Network Extensive Form Games. In our simulations, we rst generated random zero-sum extensive form games on both a 3-node graph where every player plays against the other two players, as well as a 4-node graph (shown in Figure 2). Speci cally, each game is characterized by a symmetric matrix which represents the sequence form of an extensive form game written as a matrix. For each run of the simulation, we rst create the games which are to be played. Then, we optimize for the choice of stepsize selecting the value that gives the fastest convergence rate to the Nash Equilibrium. In the plots, in order to reduce visual clutter we present the squared distance from the Nash for only one of the players. In addition, in order to more clearly show the fast rate of convergence, we compute the logarithm of $\text{dist}^2(x^t; X\ )$ in the plots. It is worth noting that the 4-node graph takes signi cantly longer to arrive at the last iterate compared to the 3-node graph.

Kuhn Poker. Kuhn poker is a simpli ed version of poker proposed by Kuhn (1950a). The deck contains only three cards, namely Jack, Queen and King. Each player is dealt one card, and the third is left unseen. Player 1 can either check or bet, and subsequently Player 2 can also either check or bet. Finally, if Player 1 checks in round 1 and Player 2 bets in round 2, Player 1 gets another round to fold or call. Eventually, the player with the highest card wins the pot. In the sequence form representation of the game, Kuhn poker has dimensions $jX_1 j; jX_2 j = 13 \quad 13$ and the corresponding payoff matrix can be easily computed by hand. For the simulation we show in Figure 1, we run an experiment with 5 players on a graph where each player plays in exactly two Kuhn poker games with randomized initial conditions. This limitation was set in order to reduce the convergence time, since empirically we observe that increasing the number of players greatly increases the convergence times.

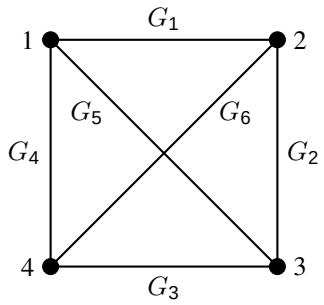

Figure 2: 4-node graph for randomized EFGs. Each node represents a player and each edge represents a game G_i between the corresponding players.

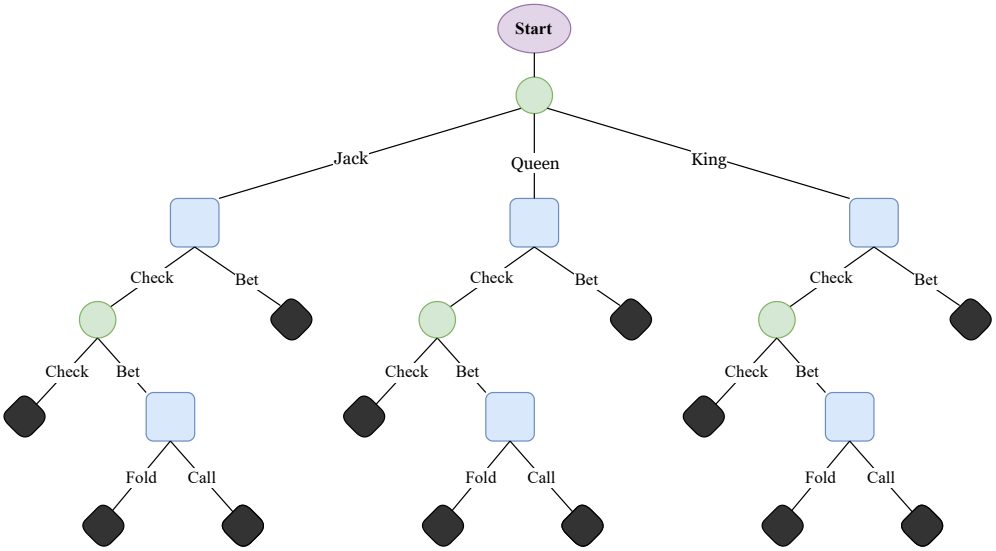

Figure 3: Extensive form representation of Kuhn poker from the perspective of one player. The blue nodes represent decision points for the player, green nodes represent observation points (either the player observes their card or the other player takes an action) and finally the black nodes denote the terminal states of the game.

Time Average Convergence. In the main text we show figures that exhibit last iterate convergence. Our theoretical results also guarantee time-average convergence to the Nash Equilibrium set (Theorem 1). In Figure 4 we show empirical evidence for time-average convergence in each of the simulations performed. In the plots, we take the difference between the cumulative averages of the strategy probabilities and the time-average value calculated from the data in order to centre the oscillations at 0. Note that in general, the time required to converge to the Nash is much faster compared to the last-iterate convergence times.

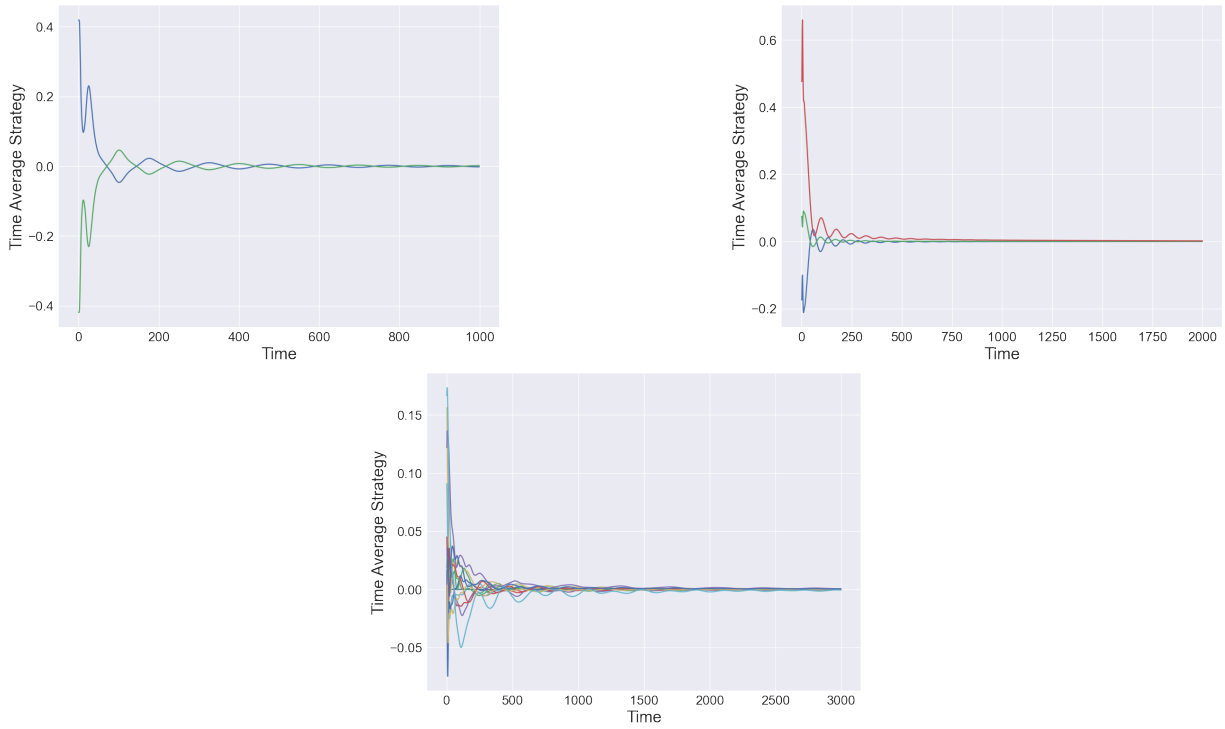

Figure 4: Time average convergence of all strategies in OGA simulations. (Left) 20-node Matching Pennies game; (Right) 4-node random extensive form game; (Bottom) 5-node Kuhn poker game.

A note on scaling. An empirical observation from our simulations is that the number of nodes in the network as well as the sparsity of the graph plays a major role in convergence times, particularly the *last-iterate* convergence times. This intuitive observation presents an interesting challenge when modeling truly large-scale problems. For instance, a setting such as Texas Hold'em poker admits a huge number of parameters (of order 10^{18}). Even in the two-player case this is prohibitively large, and this issue is compounded if we are in the multiplayer setting. As an illustrative example, consider a network game where every agent plays the ubiquitous zero-sum game, Matching Pennies, against two other players. Figure 5 shows that the convergence times drastically increase when we go from a 4-node graph to a 20-node graph. Similarly, in our experiments with extensive form games in sequence form, it becomes difficult to simulate larger games (such as Leduc poker, which has dimension $|\mathcal{X}| \times |\mathcal{X}| = 337$) once there are multiple players playing in several games. This is a practical limitation which represents an interesting divide between our theoretical results and the reality of many large-scale, real world games. It is certainly a fascinating research direction to find ways to bridge this gap in future research.