

---

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

---

**Georgios Piliouras**

Singapore Univ. of Technology and Design  
Singapore  
georgios@sutd.edu.sg

**Lillian Ratliff**

University of Washington  
Seattle, Washington  
ratliff@uw.edu

**Ryann Sim**

Singapore Univ. of Technology and Design  
Singapore  
ryann\_sim@mymail.sutd.edu.sg

**Stratis Skoulakis**

EPFL  
Laussane, Switzerland  
efstratios.skoulakis@epfl.ch

## 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 studied for more than 50 years [19]. 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 [33, 4]. 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 [1, 28].

Despite its numerous applications, online learning in EFGs 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 [38, 30]. From a theoretical perspective, it is known that online learning dynamics may oscillate, cycle or even admit chaotic behavior even in very simple settings [27, 24, 23]. 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* [10, 35, 36]. Since in most settings of interest there are typically multiple interacting agents, the above results motivate 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  $\Gamma^{uv}$  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.

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*, and show the following result:

**Informal Theorem.** *When the agents of a network zero-sum extensive form game update their strategies using Optimistic Gradient Ascent, their time-average strategies converge with rate  $O(1/T)$  to a Nash Equilibrium, while the last-iterate mixed strategies converge to a Nash Equilibrium with rate  $O(c^{-t})$  for some game-dependent constant  $c > 0$ .*

Network Zero-Sum EFGs are an interesting class of multi-agent EFGs for much the same reasons that network zero-sum normal form games are interesting, with several additional challenges. Indeed, due to the prevalence of networks in computing systems, there has been increased interest in network formulations of normal form games [14], which have been applied to multi-agent reinforcement learning [37] and social networks [15].

Network Zero-Sum EFGs can be seen as a natural model of closed systems in which selfish agents compete over a fixed set of resources [8, 5], thanks to their global constant-sum property<sup>1</sup> (the edge-games are not necessarily zero-sum). For example, consider the users of an online poker platform playing *Heads-up Poker*, a two-player extensive form game. Each user can be thought of as a node in a graph and two users are connected by an edge (corresponding to a 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  $u$  selects a poker-strategy to utilize against the other players, with the goal of maximizing her overall payoff. This is an indicative example which can clearly be modeled as a Network Zero-Sum EFG.

In addition, Network Zero-Sum EFGs are also attractive to study due to the fact that their 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 [14]. Hence, by considering this class of games, we are able to exploit the decomposition to extend results that are known for network normal form games to the extensive form setting.

**Our Contributions.** To the best of our knowledge, this is the first work establishing convergence to Nash Equilibria 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 to Nash Equilibria 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:

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

Any online learning dynamic or algorithm that converges to the solution of the above saddle-point problem thus also converges to the Nash equilibrium in the two-player case.

However, in the network setting, there is no min-max equilibrium and hence no such connection between the Nash Equilibrium and saddle-point optimization. To overcome this difficulty, we establish that Optimistic Gradient Ascent in a class of EFGs known as *consistent* Network Zero-Sum EFGs (see Section 3) can be equivalently described as optimistic gradient descent in a *two-player symmetric game*  $(R, R)$  over a *treplex polytope*  $\mathcal{X}$ . We remark that both the matrix  $R$  and the treplex polytope  $\mathcal{X}$  are constructed from the Network Zero-Sum EFG. Using the zero-sum property of Network EFGs, we show that the constructed matrix  $R$  satisfies the following 'restricted' zero-sum property:

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

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

<sup>1</sup>equivalent to the global zero-sum property.

$x, y \in \mathcal{X}$ . Our technical contribution consists of generalizing the analysis of [35] (which holds for classical two-player zero-sum games) to symmetric games satisfying Property (1).

**Related Work.** *Network Zero-Sum Normal Form Games* [8, 5, 6] 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. [5] 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  $\Theta(1/\sqrt{T})$  to the Nash Equilibrium.

The properties of *online learning in two-player zero-sum EFGs* have been studied extensively in literature. [38] and [21] 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 achieve  $O(1/T)$  time-average convergence to the min-max equilibrium [13, 17, 10] for *two-player zero-sum EFGs*. Finally, [22] and [35] 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 Two-Player Extensive Form Games

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

- $\mathcal{H}$  denotes the states of the game that are decision points for the agents. The states  $h \in \mathcal{H}$  form a tree rooted at an initial state  $r \in \mathcal{H}$ .
- Each state  $h \in \mathcal{H}$  is associated with a set of available actions  $\mathcal{A}(h)$ .
- Each state  $h \in \mathcal{H}$  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  $h \in \mathcal{H}$  with  $\text{Label}(h) = c$  is additionally associated with a function  $\sigma_h : \mathcal{A}(h) \mapsto [0, 1]$  where  $\sigma_h(\alpha)$  denotes the probability that the chance player selects action  $\alpha \in \mathcal{A}(h)$  at state  $h$ ,  $\sum_{\alpha \in \mathcal{A}(h)} \sigma_h(\alpha) = 1$ .
- $\text{Next}(\alpha, h)$  denotes the state  $h' := \text{Next}(\alpha, h)$  which is reached when agent  $i := \text{Label}(h)$  takes action  $\alpha \in \mathcal{A}(h)$  at state  $h$ .  $\mathcal{H}_i \subseteq \mathcal{H}$  denotes the states  $h \in \mathcal{H}$  with  $\text{Label}(h) = i$ .
- $\mathcal{Z}$  denotes the terminal states of the game corresponding to the leaves of the tree. At each  $z \in \mathcal{Z}$  no further action can be chosen, so  $\mathcal{A}(z) = \emptyset$  for all  $z \in \mathcal{Z}$ . Each terminal state  $z \in \mathcal{Z}$  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  $\mathcal{H}_i$  is further partitioned into information sets  $(\mathcal{I}_1, \dots, \mathcal{I}_k)$  where  $\mathcal{I}(h)$  denotes the information set of state  $h \in \mathcal{H}_i$ . In the case that  $\mathcal{I}(h_1) = \mathcal{I}(h_2)$  for some  $h_1, h_2 \in \mathcal{H}_i$ , then  $\mathcal{A}(h_1) = \mathcal{A}(h_2)$ .

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 ( $\mathcal{I}(h_1) = \mathcal{I}(h_2)$ ) must coincide, in particular  $\mathcal{A}(h_1) = \mathcal{A}(h_2)$ .

**Definition 2.** A behavioral plan  $\sigma_i$  for agent  $i$  is a function such that for each state  $h \in \mathcal{H}_i$ ,  $\sigma_i(h)$  is a probability distribution over  $\mathcal{A}(h)$  i.e.  $\sigma_i(h, \alpha)$  denotes the probability that agent  $i$  takes action  $\alpha \in \mathcal{A}(h)$  at state  $h \in \mathcal{H}_i$ . Furthermore it is required that  $\sigma_i(h_1) = \sigma_i(h_2)$  for each  $h_1, h_2 \in \mathcal{H}_i$  with  $\mathcal{I}(h_1) = \mathcal{I}(h_2)$ . The set of all behavioral plans for agent  $i$  is denoted by  $\Sigma_i$ .

The constraint  $\sigma_i(h_1) = \sigma_i(h_2)$  for all  $h_1, h_2 \in \mathcal{H}_i$  with  $\mathcal{I}(h_1) = \mathcal{I}(h_2)$  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  $h_1, h_2 \in \mathcal{H}_i$ .

**Definition 3.** For a collection of behavioral plans  $\sigma = (\sigma_1, \sigma_2) \in \Sigma_1 \times \Sigma_2$  the payoff of agent  $i$ , denoted by  $U_i(\sigma)$ , is defined as:

$$U_i(\sigma) := \sum_{z \in \mathcal{Z}} p_i(z) \cdot \underbrace{\prod_{(h, h') \in \mathcal{P}(z)} \sigma_{\text{Label}(h)}(h, \alpha_{h'})}_{\text{probability that state } z \text{ is reached}}$$

where  $\mathcal{P}(z)$  denotes the path from the root state  $r$  to the terminal state  $z$  and  $\alpha_{h'}$  denotes the action  $\alpha \in \mathcal{H}_i$  such that  $h' = \text{Next}(h, \alpha)$ .

**Definition 4.** A collection of behavioral plans  $\sigma^* = (\sigma_1^*, \sigma_2^*)$  is called a Nash Equilibrium if for all agents  $i = \{1, 2\}$ ,

$$U_i(\sigma_i^*, \sigma_{-i}^*) \geq U_i(\sigma_i, \sigma_{-i}^*) \quad \text{for all } \sigma_i \in \Sigma_i$$

The classical result of [26] 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* ([20, 31]).

**Definition 5.** A two-player extensive form game  $\Gamma := \langle \mathcal{H}, \mathcal{A}, \mathcal{Z}, p, \mathcal{I} \rangle$  has **perfect recall** if and only if for all states  $h_1, h_2 \in \mathcal{H}_i$  with  $\mathcal{I}(h_1) = \mathcal{I}(h_2)$  the following holds: Define the sets  $\mathcal{P}(h_1) \cap \mathcal{H}_i := (p_1, \dots, p_k, h_1)$  and  $\mathcal{P}(h_2) \cap \mathcal{H}_i := (q_1, \dots, q_m, h_2)$ . Then:

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

Before proceeding, let us further explain the perfect recall property. As already mentioned, agent  $i$  cannot differentiate between states  $h_1, h_2 \in \mathcal{H}_i$  when  $\mathcal{I}(h_1) = \mathcal{I}(h_2)$ . In order for the state  $h_1$  to be reached, agent  $i$  must take some specific actions along the path  $\mathcal{P}(h_1) \cap \mathcal{H}_i := (p_1, \dots, p_k, h_1)$ . The same logic holds for  $\mathcal{P}(h_2) \cap \mathcal{H}_i := (p_1, \dots, p_k, h_1)$ . In case where agent  $i$  could distinguish  $\mathcal{P}(h_1) \cap \mathcal{H}_i$  from set  $\mathcal{P}(h_2) \cap \mathcal{H}_i$ , then she could distinguish state  $h_1$  from  $h_2$  by recalling the previous states in  $\mathcal{H}_i$ . This is the reason for the second constraint in Definition 5. Even if  $\mathcal{I}(p_\ell) = \mathcal{I}(q_\ell)$  for all  $\ell \in \{1, \dots, k\}$ , agent  $i$  could still distinguish  $h_1$  from  $h_2$  if  $p_{\ell+1} \in \text{Next}(p_\ell, \alpha, i)$  and  $q_{\ell+1} \in \text{Next}(q_\ell, \alpha', i)$ . 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  $\ell$ -th action was  $\alpha$  or  $\alpha'$ . This case is encompassed by the third constraint.

## 2.2 Two-Player Extensive Form Games in Sequence Form

A two-player extensive form game  $\Gamma$  can be captured by a two-player bilinear game where the action spaces of the agents are a specific kind of polytope, commonly known as a *treeplex* [13]. In order to formally define the notion of a treeplex, we first need to introduce some additional notation.

**Definition 6.** Given an two-player extensive form game  $\Gamma$ , we define the following:

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

**Definition 7.** Given a two-player extensive form game  $\Gamma$ , the set  $\mathcal{X}_i^\Gamma$  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  $h \in \mathcal{H}_i$  with  $\text{Prev}(h, i) = \emptyset$ .
2.  $x_i(h_1) = x_i(h_2)$  if there exists  $h'_1, h'_2 \in \mathcal{H}_i$  such that  $h_1 \in \text{Next}(h'_1, \alpha, i)$ ,  $h_2 \in \text{Next}(h'_2, \alpha, i)$  and  $\mathcal{I}(h'_1) = \mathcal{I}(h'_2)$ .
3.  $\sum_{\alpha \in \mathcal{A}(h)} x_i(\text{Next}(h, \alpha, i)) = x_i(h)$  for all  $h \in \mathcal{H}_i$ .

A vector  $x_i \in \mathcal{X}_i^\Gamma$  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 Definition 2. As established in Lemma 1, there exists an equivalence between a behavioral plan  $\sigma_i \in \Sigma_i$  and a strategy in sequence form  $x_i \in \mathcal{X}_i^\Gamma$  for games with perfect recall.

**Lemma 1.** Consider a two-player extensive form game  $\Gamma$  with perfect recall and the  $(|\mathcal{H}_1| + |\mathcal{Z}|) \times (|\mathcal{H}_2| + |\mathcal{Z}|)$  dimensional matrices  $A_1^\Gamma, A_2^\Gamma$  with  $[A_i^\Gamma]_{zz} = p_i(z)$  for all terminal nodes  $z \in \mathcal{Z}$  and 0 otherwise. There exists a polynomial-time algorithm transforming any behavioral plan  $\sigma_i \in \Sigma_i$  to a vector  $x_{\sigma_i} \in \mathcal{X}_i^\Gamma$  such that

$$U_1(\sigma_1, \sigma_2) = x_{\sigma_1}^\top \cdot A_1^\Gamma \cdot x_{\sigma_2} \quad \text{and} \quad U_2(\sigma_1, \sigma_2) = x_{\sigma_2}^\top \cdot A_2^\Gamma \cdot x_{\sigma_1}$$

Conversely, there exists a polynomial-time algorithm transforming any vector  $x_i \in \mathcal{X}_i^\Gamma$  to a vector  $\sigma_{x_i} \in \Sigma_i$  such that

$$x_1^\top \cdot A_1^\Gamma \cdot x_2 = U_1(\sigma_{x_1}, \sigma_{x_2}) \quad \text{and} \quad x_2^\top \cdot A_2^\Gamma \cdot x_1 = U_2(\sigma_{x_1}, \sigma_{x_2})$$

To this end, one can understand why *strategies in sequence form* are of great use. Assume that agent 2 selects a behavioral plan  $\sigma_2 \in \Sigma_2$ . Then, agent 1 wants to compute a behavioral plan  $\sigma_1^* \in \Sigma_1$  which is the *best response* to  $\sigma_2$ , namely  $\sigma_1^* := \operatorname{argmax}_{\sigma_1 \in \Sigma_1} U_1(\sigma_1, \sigma_2)$ . This computation can be done in polynomial-time in the following manner: Agent 1 initially converts (in polynomial time) the behavioral plan  $\sigma_2$  to  $x_{\sigma_2} \in \mathcal{X}_2^\Gamma$ , which is the respective strategy in sequence form. Then, she can obtain a vector  $x_1^* = \operatorname{argmax}_{x_1 \in \mathcal{X}_1^\Gamma} x_1^\top \cdot A_1^\Gamma \cdot x_2$ . 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  $\sigma_{x_1^*} \in \Sigma_1$  in polynomial-time. Lemma 1 ensures that  $\sigma_{x_1^*} = \operatorname{argmax}_{\sigma_1 \in \Sigma_1} U_1(\sigma_1, \sigma_2)$ .

The above reasoning can be used to establish an equivalence between the Nash Equilibrium  $(\sigma_1^*, \sigma_2^*)$  of an EFG  $\Gamma := \langle \mathcal{H}, \mathcal{A}, \mathcal{Z}, p, \mathcal{I} \rangle$  with the Nash Equilibrium in its sequence form.

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

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

Lemma 1 directly implies that any Nash Equilibrium of an EFG  $(\sigma_1^*, \sigma_2^*) \in \Sigma_1 \times \Sigma_2$  as per Definition 4 can be converted in polynomial-time to a Nash Equilibrium in the sequence form  $(x_1^*, x_2^*) \in \mathcal{X}_1^\Gamma \times \mathcal{X}_2^\Gamma$  and vice versa.

### 2.3 Optimistic Mirror Descent

In this section we introduce and provide the necessary background for *Optimistic Mirror Descent* [29]. For a convex function  $\psi : \mathbb{R}^d \mapsto \mathbb{R}$ , the corresponding *Bregman divergence* is defined as

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

If  $\psi$  is  $\gamma$ -strongly convex, then  $D_\psi(x, y) \geq \frac{\gamma}{2} \|x - y\|^2$ . 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,

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

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  $t = 1, 2, \dots$ :

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

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

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  $\eta$  selected is sufficiently small, then *Optimistic Mirror Descent* ensures the *no-regret property* [29], making it a natural update algorithm for selfish agents [12]. To simplify notation we denote the projection operator of a convex set  $\mathcal{X}^*$  as  $\Pi_{\mathcal{X}^*}(x) := \operatorname{argmin}_{x^* \in \mathcal{X}^*} \|x - x^*\|$  and the squared distance of vector  $x$  from a convex set  $\mathcal{X}^*$  as  $\operatorname{dist}^2(x, \mathcal{X}^*) := \|x - \Pi_{\mathcal{X}^*}(x)\|^2$ .

## 3 Our Setting

In this section of the paper, we introduce the concept of Network Zero-Sum Extensive Form Games, which are a network extension of the two player EFGs 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  $\mathcal{G} = (V, E)$  where nodes  $V$  ( $|V| = n$ ) correspond to the set of players and each edge  $(u, v) \in E$  represents a *two-player extensive form game*  $\Gamma^{uv}$  played between agents  $u, v$ . Each node/agent  $u \in V$  selects a behavioral plan  $\sigma_u \in \Sigma_u$  which they use to play all the *two-player EFGs* on its outgoing edges.

**Definition 9** (Network Extensive Form Games). A *network extensive form game* is a tuple  $\Gamma := \langle \mathcal{G}, \mathcal{H}, \mathcal{A}, \mathcal{Z}, \mathcal{I} \rangle$  where

- $\mathcal{G} = (V, E)$  is an undirected graph where the nodes  $V$  represents the agents.
- Each agent  $u \in V$  admits a set of states  $\mathcal{H}_u$  at which the agent  $u$  plays. Each state  $h \in \mathcal{H}_u$  is associated with a set  $\mathcal{A}(h)$  of possible actions that agent  $u$  can take at state  $h$ .
- $\mathcal{I}(h)$  denotes the information set of  $h \in \mathcal{H}_u$ . If  $\mathcal{I}(h) = \mathcal{I}(h')$  for some  $h, h' \in \mathcal{H}_u$  then  $\mathcal{A}(h) = \mathcal{A}(h')$ .
- For each edge  $(u, v) \in E$ ,  $\Gamma^{uv}$  is a two-player extensive form game with perfect recall. The states of  $\Gamma^{uv}$  are denoted by  $\mathcal{H}^{uv} \subseteq \mathcal{H}_u \cup \mathcal{H}_v$ .
- For each edge  $(u, v) \in E$ ,  $\mathcal{Z}^{uv}$  is the set of terminal states of the two-player extensive form game  $\Gamma^{uv}$  where  $p_u^{\Gamma^{uv}}(z)$  denotes the payoffs of  $u, v$  at the terminal state  $z \in \mathcal{Z}^{uv}$ . The overall set of terminal states of the network extensive form game is the set  $\mathcal{Z} := \cup_{(u,v) \in E} \mathcal{Z}^{uv}$ .

In a network extensive form game, each agent  $u \in V$  selects a behavioral plan  $\sigma_u \in \Sigma_u$  (see Definition 2) that they use to play the two-player EFG's  $\Gamma^{uv}$  with  $(u, v) \in E$ . Each agent selects her behavioral plan so as to maximize the sum of the payoffs of the two-player EFGs in her outgoing edges.

**Definition 10.** Given a collection of behavioral plans  $\sigma = (\sigma_1, \dots, \sigma_n) \in \Sigma_1 \times \dots \times \Sigma_n$  the payoff of agent  $u$ , denoted by  $U_u(\sigma)$ , equals

$$U_u(\sigma) := \sum_{v:(u,v) \in E} p_u^{\Gamma^{uv}}(\sigma_u, \sigma_v)$$

Moreover a collection  $\sigma^* = (\sigma_1^*, \dots, \sigma_n^*) \in \Sigma_1 \times \dots \times \Sigma_n$  is called a Nash Equilibrium if and only if

$$U_u(\sigma_u^*, \sigma_{-u}^*) \geq U_u(\sigma_u, \sigma_{-u}^*) \quad \text{for all } \sigma_u \in \Sigma_u$$

As already mentioned, each agent  $u \in V$  plays all the two-player games  $\Gamma^{uv}$  for  $(u, v) \in E$  with the same behavioral plan  $\sigma_u \in \Sigma_u$ . This is due to the fact that the agent cannot distinguish between a state  $h_1, h_2 \in \mathcal{H}_u$  with  $\mathcal{I}(h_1) = \mathcal{I}(h_2)$  even if  $h_1, h_2$  are states of different EFG's  $\Gamma^{uv}$  and  $\Gamma^{uv'}$ . As in the case of *perfect recall*, the latter implies that  $u$  cannot differentiate states  $h_1, h_2$  even when recalling the states  $\mathcal{H}_u$  visited in the past and her past actions. In Definition 11 we introduce the notion of *consistency* (this corresponds to the notion of *perfect recall* for two-player extensive form games (Definition 5)). From now on we assume that the network EFG is consistent without mentioning it explicitly.

**Definition 11.** A network extensive form game  $\Gamma := \langle \mathcal{G}, \mathcal{H}, \mathcal{A}, \mathcal{Z}, \mathcal{I} \rangle$  is called **consistent** if and only if for all players  $u \in V$  and states  $h_1, h_2 \in \mathcal{H}_u$  with  $\mathcal{I}(h_1) = \mathcal{I}(h_2)$  the following holds: for any  $(u, v), (u, v') \in E$  the sets  $\mathcal{P}^{uv}(h_1) \cap \mathcal{H}_u := (p_1, \dots, p_k, h_1)$  and  $\mathcal{P}^{uv'}(h_2) \cap \mathcal{H}_u := (q_1, \dots, q_m, h_2)$  satisfy:

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

where  $\mathcal{P}^{uv}(h)$  denotes the path from the root state to state  $h$  in the two-player extensive form game  $\Gamma^{uv}$ .

In this work we study the special class of network *zero-sum* extensive form games. This class of games is a generalization of the network zero-sum normal form games studied in [5].

**Definition 12.** A behavioral plan  $\sigma_u \in \Sigma_u$  of Definition 2 is called *pure* if and only if  $\sigma_u(h, \alpha)$  either equals 0 or 1 for all actions  $\alpha \in \mathcal{A}(h)$ . A network extensive form game is called **zero-sum** if and only if for any collection  $\sigma := (\sigma_1, \dots, \sigma_n)$  of pure behavioral plans,  $U_u(\sigma) = 0$  for all  $u \in V$ .

### 3.2 Network Extensive Form Games in Sequence Form

As in the case of two-player EFGs, there exists an equivalence between behavioral plans  $\sigma_u \in \Sigma_u$  and strategies in sequence form  $x_u$ . As we shall later see, this equivalence is of great importance since it allows for the design of natural and computationally efficient learning dynamics that converge to Nash Equilibria both in terms of behavioral plans and strategies in sequence form.

**Definition 13.** Given a network extensive form game  $\Gamma := \langle \mathcal{G}, \mathcal{H}, \mathcal{A}, \mathcal{Z}, \mathcal{I} \rangle$ , the treeplex polytope  $\mathcal{X}_u \subseteq [0, 1]^{|\mathcal{H}_u| + |\mathcal{Z}_u|}$  is the set defined as follows:  $x_u \in \mathcal{X}_u$  if and only if

1.  $x_u \in \mathcal{X}_u^{\Gamma^{uv}}$  for all  $(u, v) \in E$ .

2.  $x_u(h_1) = x_u(h_2)$  in case there exists  $(u, v), (u, v') \in E$  and  $h'_1, h'_2 \in \mathcal{H}_u$  with  $\mathcal{I}(h'_1) = I(h'_2)$  such that  $h_1 \in \text{Next}^{\Gamma_{uv}}(h'_1, \alpha, u)$ ,  $h_2 \in \text{Next}^{\Gamma_{uv'}}(h'_2, \alpha, u)$  and  $\mathcal{I}(h'_1) = I(h'_2)$ .

The second constraint in Definition 13 is the equivalent of the second constraint in Definition 7. We remark that the linear equations describing the treplex polytope  $\mathcal{X}_u$  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  $(|\mathcal{H}_u| + |\mathcal{Z}^u|) \times (|\mathcal{H}_v| + |\mathcal{Z}^v|)$  such that

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

There exists a polynomial time algorithm converting any collection of behavioral plans  $(\sigma_1, \dots, \sigma_n) \in \Sigma_1 \times \dots \times \Sigma_n$  into a collection of vectors  $(x_1, \dots, x_n) \in \mathcal{X}_1 \times \dots \times \mathcal{X}_n$  such that for any  $u \in V$ ,

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

In the opposite direction, there exists a polynomial time algorithm converting any collection of vectors  $(x_1, \dots, x_n) \in \mathcal{X}_1 \times \dots \times \mathcal{X}_n$  into a collection of behavioral plans  $(\sigma_1, \dots, \sigma_n) \in \Sigma_1 \times \dots \times \Sigma_n$  such that for any  $u \in V$ ,

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

**Definition 14.** A Nash Equilibrium of a network extensive form game  $\mathcal{G}$  in sequence form is a vector  $(x_1^*, \dots, x_n^*) \in \mathcal{X}_1 \times \dots \times \mathcal{X}_n$  such that for all  $u \in V$ :

$$(x_u^*)^\top \cdot \sum_{v:(u,v) \in E} A^{uv} \cdot x_v^* \geq x_u^\top \cdot \sum_{v:(u,v) \in E} A^{uv} \cdot x_v^* \text{ for all } x_u \in \mathcal{X}_u$$

**Corollary 1.** Given a network extensive form game, any Nash Equilibrium  $(\sigma_1^*, \dots, \sigma_n^*) \in \Sigma_1 \times \dots \times \Sigma_n$  (as per Definition 4) can be converted in polynomial-time to a Nash Equilibrium  $(x_1^*, \dots, x_n^*) \in \mathcal{X}_1 \times \dots \times \mathcal{X}_n$  (as per Definition 14) and vice versa.

The sequence form representation gives us a perspective with which we can analyze the theoretical properties of learning algorithms when applied to network zero-sum EFGs. In the following section, we utilize the sequence form representation to study a special case of Optimistic Mirror Descent known as Optimistic Gradient Ascent (OGA).

## 4 Our Convergence Results

In this work, we additionally 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  $\psi(a) = \frac{1}{2} \|a\|^2$ , which means that the Bregman divergence  $D_\psi(x, y)$  equals  $\frac{1}{2} \|x - y\|^2$ . Since in network zero-sum EFGs each agent tries to maximize her payoff, OGA takes the following form:

$$x_u^t = \operatorname{argmax}_{x \in \mathcal{X}_u} \left\{ \eta \left\langle x, \sum_{v:(u,v) \in E} A^{uv} \cdot x_v^{t-1} \right\rangle - D_\psi(x, \hat{x}_u^t) \right\} \quad (4)$$

$$\hat{x}_u^{t+1} = \operatorname{argmax}_{x \in \mathcal{X}_u} \left\{ \eta \left\langle x, \sum_{v:(u,v) \in E} A^{uv} \cdot x_v^t \right\rangle - D_\psi(x, \hat{x}_u^t) \right\} \quad (5)$$

In Theorem 1 we describe the  $\Theta(1/T)$  convergence rate to NE for the time-average strategies for any agent using OGA.

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

$$\hat{x}_u^\top \cdot \sum_{v:(u,v) \in E} A^{uv} \cdot \hat{x}_v \geq x^\top \cdot \sum_{v:(u,v) \in E} A^{uv} \cdot \hat{x}_v - \Theta\left(\frac{c_1 \cdot c_2}{T}\right) \text{ for all } x \in \mathcal{X}_u$$

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  $u \in V$ ,

$$U_u(\hat{\sigma}_u, \hat{\sigma}_{-u}) \geq U_u(\sigma_u, \hat{\sigma}_{-u}) - \Theta(c_1 \cdot c_2/T) \quad \text{for all } \sigma_u \in \Sigma_u$$

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  $\eta \leq 1/c_3$  when applied to a network zero-sum extensive form game. Then, the following inequality holds:*

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

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 consistent and zero-sum. Before proceeding, we introduce a few more necessary definitions and notations. We denote as  $\mathcal{X} := \mathcal{X}_1 \times \dots \times \mathcal{X}_n$  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*

$$x^t = \underset{x \in \mathcal{X}}{\text{argmin}} \left\{ \eta \langle x, R \cdot x^{t-1} \rangle + D_\psi(x, \hat{x}^t) \right\} \quad (6)$$

$$\hat{x}^{t+1} = \underset{x \in \mathcal{X}}{\text{argmin}} \left\{ \eta \langle x, R \cdot x^t \rangle + D_\psi(x, \hat{x}^t) \right\} \quad (7)$$

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 vector  $(x^t, y^t)$  will be of the form  $(x^t, x^t)$  and indeed,  $(x^t, \hat{x}^t)$  will satisfy Equations (6), (7). We prove that the produced vector sequence  $\{x^t\}_{t \geq 1}$  converges to a *symmetric Nash Equilibrium*.

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

$$(x^*)^\top \cdot R \cdot x^* \leq x^\top \cdot R \cdot x^* + \epsilon \quad \text{for all } x \in \mathcal{X}$$

Any  $\epsilon$ -symmetric Nash Equilibrium  $x^* \in \mathcal{X}$  is also an  $\epsilon$ -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 [5] and is presented in Appendix B.5.

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

Once Lemma 5 is established, we can use it to prove that the time-average strategy vector converges to an  $\epsilon$ -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,*

$$\min_{x \in \mathcal{X}} x^\top \cdot R \cdot \hat{x} \geq -\Theta \left( \frac{\mathcal{D}^2 \|R\|^2}{T} \right)$$

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

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



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

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

Using this, we further 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 adapted from the analysis of [36], with modifications to apply the steps to our setting. The proof of Theorem 3 is deferred to Appendix B.8.

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

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

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

The statement of Theorem 2 then follows directly by combining Theorem 3 and Lemma 4. This result generalizes previous last-iterate convergence results for the setting of two-player zero-sum EFGs, even for games without a unique Nash Equilibrium.

## 5 Experimental Results

In order to better visualize our theoretical results, we experimentally evaluate OGA when applied to various network extensive form games. As part of the experimental process, for each simulation we ran a hyperparameter search to find the value of  $\eta$  which gave the best convergence rate.

**Time-average Convergence.** Our theoretical results guarantee time-average convergence to the Nash Equilibrium set (Theorem 1). We experimentally confirm this by running OGA on a network version of the ubiquitous Matching Pennies game with 20 nodes (Figure 1 (a)), followed by a 4-node network zero-sum EFG (Figure 1 (b)). In particular, for the latter experiment each bilinear game between the players on the nodes is a randomly generated extensive form game with payoff values in  $[0, 1]$ . Next, we experimented with a well-studied simplification of poker known as Kuhn poker [18]. Emulating the illustrative example of a competitive online Poker lobby as described in Section 1, we modelled a situation whereby each agent is playing against multiple other agents, and ran simulations for such a game with 5 agents (Figure 1 (c)).

In the plots, we show on the  $y$ -axis the difference between the cumulative averages of the strategy probabilities and the Nash Equilibrium value calculated from the game. In each of the plots, we see that these time-average values go to 0, implying convergence to the NE set.

**Last-iterate Convergence.** Theorem 2 guarantees  $O(c^{-t})$  convergence in the last-iterate sense to a Nash Equilibrium for OGA. Similar to the time-average case, we ran simulations for randomly generated 3 and 4-node network extensive form games, where each bilinear game between two agents is a randomly generated matrix with values in  $[0, 1]$  (Figure 2 (a-b)). Moreover, we also simulated a 5-node game of Kuhn poker in order to generate Figure 2 (c). In order to generate the plots, we measured the log of the distance between each agent’s strategy at time  $t$  and the set of Nash Equilibria (computed *a priori*), given by  $\log(\text{dist}^2(x^t, \mathcal{X}^*))$ . As can be seen in Figure 2, OGA indeed obtains fast convergence in the last-iterate sense to a Nash Equilibrium in each of our experiments.

A point worth noting is that when the number of nodes increases, the empirical last-iterate convergence time also increases drastically. For example, in the 5-player Kuhn poker game we see that each agents’ convergence time is significantly greater compared to the smaller scale experiments. However, with a careful choice of  $\eta$ , we can still guarantee convergence to the set of Nash Equilibria for all players. Further discussion of these observations and detailed game descriptions can be found in Appendix C.

## 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, defined on a graph. 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 this class of games.

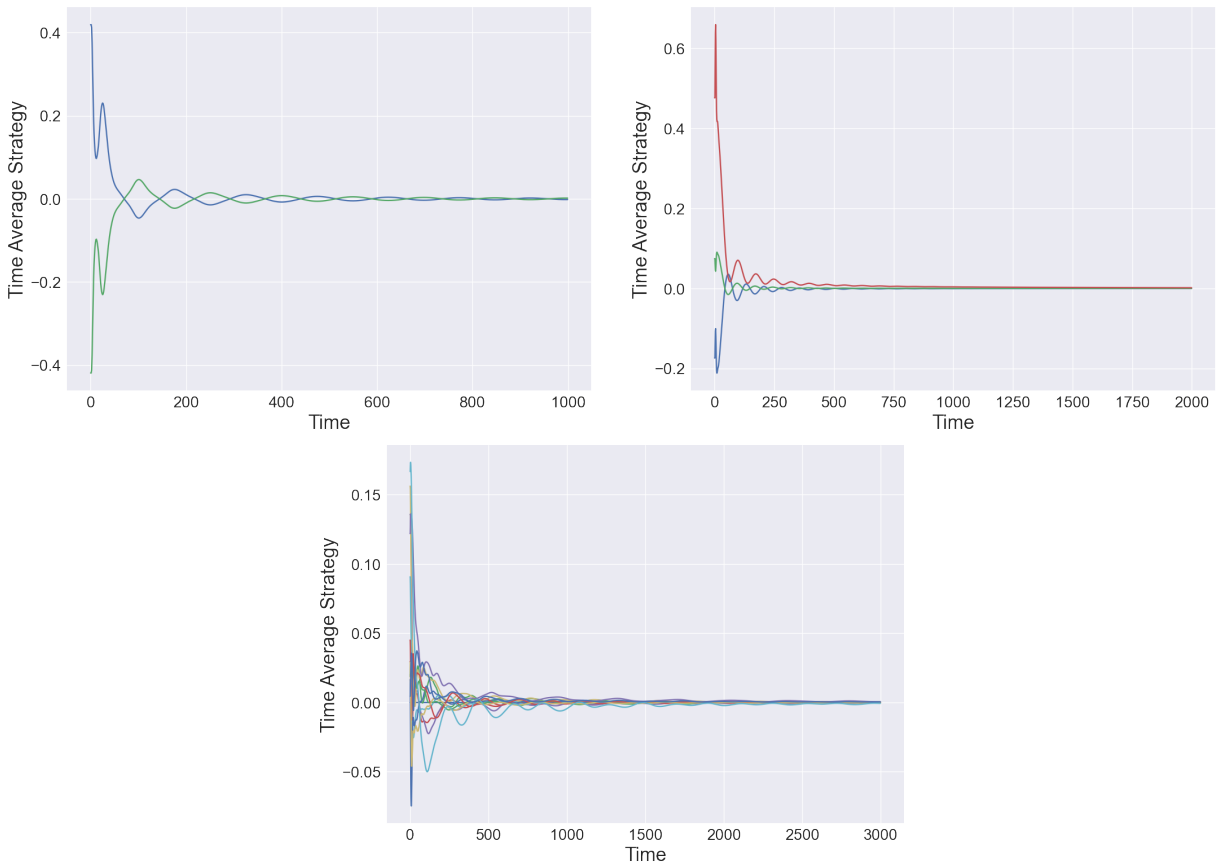


Figure 1: Time-average convergence of OGA in network zero-sum extensive form games, where each player is involved in 2 or more different games and must select their strategy accordingly. (a) 20-node Matching Pennies game. (b) 4-node random extensive form game. (c) 5-node *Kuhn poker* game.

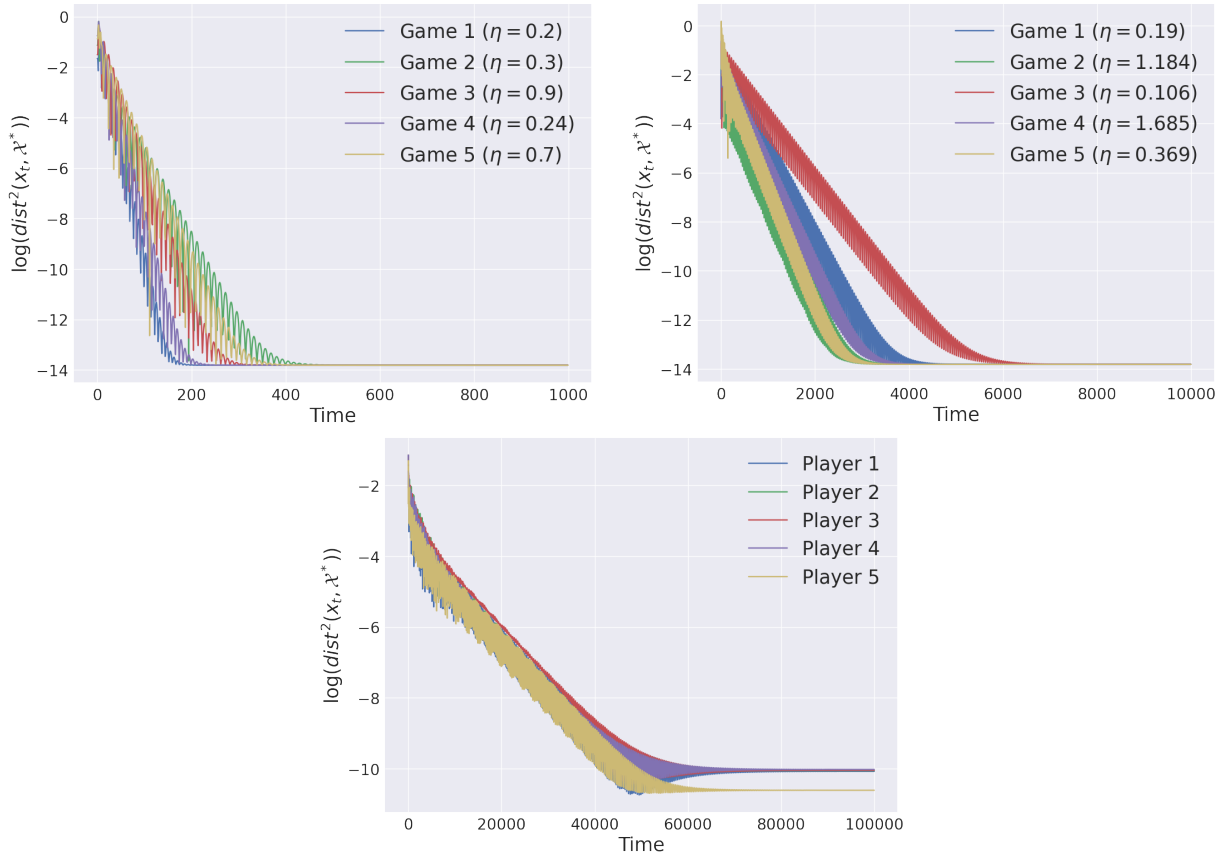


Figure 2: Last-iterate convergence of OGA to the NE in network zero-sum extensive form games. The plots shown are: (a) 3-node randomly generated network zero-sum extensive form game. (b) 4-node random network zero-sum extensive form game. Note the significantly longer time needed to achieve convergence compared to the 3-node experiment. (c) 5-node *Kuhn poker* game.

## Acknowledgements

This research/project is supported by the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Program (AISG Award No: AISG2-RP-2020-016), NRF2019-NRFANR095 ALIAS grant, grant PIE-SGP-AI-2020-01, NRF 2018 Fellowship NRF-NRFF2018-07 and AME Programmatic Fund (Grant No. A20H6b0151) from the Agency for Science, Technology and Research (A\*STAR). Ryann Sim gratefully acknowledges support from the SUTD President’s Graduate Fellowship (SUTD-PGF).

## References

- [1] I. Arieli and Y. Babichenko. Random extensive form games. *J. Econ. Theory*, 166:517–535, 2016.
- [2] 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.
- [3] M. Bowling, N. Burch, M. Johanson, and O. Tammelin. Heads-up limit hold’em poker is solved. *Science*, 347(6218):145–149, 2015.
- [4] N. Brown and T. Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.
- [5] 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, 2011.
- [6] 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.
- [7] C. Daskalakis and I. Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. *arXiv preprint arXiv:1807.04252*, 2018.
- [8] C. Daskalakis and C. H. Papadimitriou. On a network generalization of the minmax theorem. In *International Colloquium on Automata, Languages, and Programming*, pages 423–434. Springer, 2009.
- [9] C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng. Training gans with optimism. In *International Conference on Learning Representations (ICLR 2018)*, 2018.
- [10] G. Farina, C. Kroer, and T. Sandholm. Optimistic regret minimization for extensive-form games via dilated distance-generating functions. *Advances in neural information processing systems*, 32, 2019.
- [11] Y. Gao, C. Kroer, and D. Goldfarb. Increasing iterate averaging for solving saddle-point problems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7537–7544, 2021.
- [12] E. Hazan. Introduction to online convex optimization. *CoRR*, abs/1909.05207, 2019. URL <http://arxiv.org/abs/1909.05207>.
- [13] 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, 2010.
- [14] M. Kearns, M. L. Littman, and S. Singh. Graphical models for game theory. *arXiv preprint arXiv:1301.2281*, 2013.
- [15] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.
- [16] C. Kroer, G. Farina, and T. Sandholm. Smoothing method for approximate extensive-form perfect equilibrium. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 295–301, 2017.
- [17] C. Kroer, G. Farina, and T. Sandholm. Solving large sequential games with the excessive gap technique. *Advances in neural information processing systems*, 31, 2018.
- [18] Kuhn. Simplified two-person poker. *Contributions to the Theory of Games*, I:97–103, 1950.
- [19] H. Kuhn. Extensive form games. *Proceedings of National Academy of Science*, pages 570–576, 1950.
- [20] H. W. Kuhn and A. W. Tucker. *Contributions to the Theory of Games*, volume 2. Princeton University Press, 1953.
- [21] M. Lanctot, K. Waugh, M. Zinkevich, and M. Bowling. Monte Carlo sampling for regret minimization in extensive games. *Advances in neural information processing systems*, 22, 2009.
- [22] C.-W. Lee, C. Kroer, and H. Luo. Last-iterate convergence in extensive-form games. *Advances in Neural Information Processing Systems*, 34:14293–14305, 2021.

- [23] S. Leonardos and G. Piliouras. Exploration-exploitation in multi-agent learning: Catastrophe theory meets game theory. *Artificial Intelligence*, 304:103653, 2022.
- [24] P. Mertikopoulos, C. Papadimitriou, and G. Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2703–2717. SIAM, 2018.
- [25] M. Moravčík, 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.
- [26] J. Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.
- [27] G. Palaiopanos, I. Panageas, and G. Piliouras. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. *Advances in Neural Information Processing Systems*, 30, 2017.
- [28] J. Perolat, R. Munos, J.-B. Lespiau, S. Omidshafiei, 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.
- [29] A. Rakhlin and K. Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019. PMLR, 2013.
- [30] M. Rowland, S. Omidshafiei, K. Tuyls, J. Perolat, M. Valko, G. Piliouras, and R. Munos. Multiagent evaluation under incomplete information. *Advances in Neural Information Processing Systems*, 32, 2019.
- [31] 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.
- [32] Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-theoretic, and Logical Foundations*. Cambridge University Press, 2008.
- [33] O. Tammelin, N. Burch, M. Johanson, and M. Bowling. Solving heads-up limit Texas hold'em. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.
- [34] E.-V. 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. *Advances in Neural Information Processing Systems*, 32, 2019.
- [35] C.-Y. Wei, C.-W. Lee, M. Zhang, and H. Luo. Linear last-iterate convergence in constrained saddle-point optimization. In *International Conference on Learning Representations*, 2020.
- [36] C.-Y. Wei, C.-W. Lee, M. Zhang, and H. Luo. Last-iterate convergence of decentralized optimistic gradient descent/ascent in infinite-horizon competitive markov games. In *Conference on Learning Theory*, pages 4259–4299. PMLR, 2021.
- [37] Y. Yang and J. Wang. An overview of multi-agent reinforcement learning from game theoretical perspective. *arXiv preprint arXiv:2011.00583*, 2020.
- [38] M. Zinkevich, M. Johanson, M. Bowling, and C. Piccione. Regret minimization in games with incomplete information. *Advances in neural information processing systems*, 20, 2007.

## 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 [38] and LP methods [32]. Of particular note is the success of works utilizing CFR-based algorithms to study poker variants [4, 3, 25]. 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 first order methods to achieve approximate convergence to the Nash [16, 11].

**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 [2, 34]. In contrast, optimistic variants of these algorithms have been shown to be effective in guaranteeing last-iterate convergence [9, 7]. As described in the main text, some of these results have been extended to two-player extensive form games. Specifically, optimistic gradient descent and multiplicative weights update, as well as the versions thereof with *dilated* regularizers, have been studied by [22] and [35] in the two-player setting. This line of research into extensive form games is not limited to discrete time algorithms. [28] 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 first describe how a behavioral plan  $\sigma_i$  can be transformed to a vector  $x_i(h) \in \mathcal{X}_i$ . For any  $h \in \mathcal{X}_i$  we let  $x_i(h) := \prod_{(h, h') \in \mathcal{P}(h) \cap \mathcal{X}_i} \sigma_i(h, \alpha_{h'})$  where  $\alpha_{h'}$  is the action  $\alpha \in \mathcal{A}(h)$  such that  $h' = \text{Next}(h, \alpha)$ . We set  $x_i(h) := 1$  for all  $h \in \mathcal{H}_i$  with  $\text{Prev}(h, i) = \emptyset$ . Notice that by definition  $U_1(\sigma) = \sum_{z \in \mathcal{Z}} x_1(z) \cdot p_1(z) \cdot x_2(z) = x_1^\top \cdot A_1^\Gamma \cdot x_2$  and respectively  $U_2(\sigma) = \sum_{z \in \mathcal{Z}} x_2(z) \cdot p_2(z) \cdot x_1(z) = x_2^\top \cdot A_2^\Gamma \cdot x_1$ .

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

Now let  $h_1, h_2 \in \mathcal{H}_i$  where  $h_1 \in \text{Next}(h'_1, \alpha, i)$ ,  $h_2 \in \text{Next}(h'_2, \alpha, i)$  and  $\mathcal{I}(h_1) = \mathcal{I}(h_2)$ . Consider the set  $\mathcal{P}(h_1) \cap \mathcal{X}_i := \{p_1, \dots, p_k, h_1\}$  and  $\mathcal{P}(h_2) \cap \mathcal{X}_i := \{q_1, \dots, q_k, h_2\}$ . Due to the perfect recall property,  $m = k$  and  $\mathcal{I}(p_\ell) = \mathcal{I}(q_\ell)$ . Thus,  $x_i(h_1) = x_i(h_2)$ .

Up next we show how a vector  $x_i \in \mathcal{X}_i$  can be converted to a behavioral plan  $\sigma_i \in \Sigma_i$ . Let  $\sigma_i(h, \alpha) := \frac{x_i(h')}{x_i(h)}$  for some  $h' \in \text{Next}(h, \alpha, i)$ . Notice that due the third constraint,  $x_i(h') = x_i(h'')$  for all  $h', h'' \in \text{Next}(h, \alpha, i)$  and thus  $\sigma(h, \alpha)$  is well-defined. For  $h \in \mathcal{H}_i$  let  $h_\alpha \in \text{Next}(h, \alpha, i)$ . By the third constraint we get that  $\sum_{\alpha \in \mathcal{A}(h)} \sigma(h, \alpha) = 1$ . Finally let  $h_1, h_2 \in \mathcal{H}_i$  with  $\mathcal{I}(h_1) = \mathcal{I}(h_2)$  then  $\sigma(h_1, \alpha) = \frac{x_i(h'_1)}{x_i(h_1)}$  for some  $h'_1 \in \text{Next}(h_1, \alpha, i)$  and  $\sigma(h_2, \alpha) = \frac{x_i(h'_2)}{x_i(h_2)}$  for some  $h'_2 \in \text{Next}(h_2, \alpha, i)$ . As a result, by the second constraint we get that  $\sigma(h_1, \alpha) = \sigma(h_2, \alpha)$  for all  $\alpha \in \mathcal{A}(h)$ .

#### B.2 Proof of Lemma 2

We first describe how a behavioral plan  $\sigma_u \in \Sigma_u$  can be transformed to a vector  $x_u(h) \in \mathcal{X}_u$ . If there exists a game  $\Gamma^{uv}$  with  $(u, v) \in E$  such that  $\text{Prev}^{\Gamma^{uv}}(h, u) = \emptyset$  we set  $x_u(h) := 1$ . Let us first verify that the above assignment is valid i.e. if  $\text{Prev}^{\Gamma^{uv}}(h, u) = \emptyset$  for some  $(u, v) \in E$  then  $\text{Prev}^{\Gamma^{uv'}}(h, u) = \emptyset$  for all  $(u, v') \in E$ . Notice that  $\mathcal{P}^{uv}(h) \cap \mathcal{X}_u = \{h\}$  and thus by the second constraint of Definition 13,  $\mathcal{P}^{uv'}(h) \cap \mathcal{X}_u = \{h\}$  for all  $(u, v') \in E$ . Now for the remaining nodes  $h \in \mathcal{H}_u$  we select an arbitrary two-player EFG  $\Gamma^{uv}$  ( $(u, v) \in E$ ) containing the state  $h$  and set

$x_u(h) := \prod_{(h,h') \in \mathcal{P}^{uv}(h) \cap \mathcal{X}_u} \sigma_u(h, \alpha_{h'})$  where  $\alpha_{h'}$  is the action  $\alpha \in \mathcal{A}(h)$  such that  $h' = \text{Next}^{\Gamma^{uv}}(h, \alpha, u)$ . We again need to argue that  $x_u(h)$  is independent of the arbitrary choice of the game  $\Gamma^{uv}$ . 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  $\mathcal{P}^{uv}(h) \cap \mathcal{X}_u = \{p_1, \dots, p_k, h\}$  and  $\mathcal{P}^{uv'}(h) \cap \mathcal{X}_u = \{q_1, \dots, q_m, h\}$  the following holds:

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

Since  $\mathcal{I}(p_\ell) = \mathcal{I}(q_\ell)$  means that  $\sigma_u(p_\ell, \alpha) = \sigma_u(q_\ell, \alpha)$  for all  $\alpha \in \mathcal{A}(p_\ell) = \mathcal{A}(q_\ell)$ , we get that

$$\prod_{(h,h') \in \mathcal{P}^{uv}(h) \cap \mathcal{X}_u} \sigma_u(h, \alpha_{h'}) = \prod_{(h,h') \in \mathcal{P}^{uv'}(h) \cap \mathcal{X}_u} \sigma_u(h, \alpha_{h'})$$

Conversely, we show how a strategy in sequence form  $x_u \in \mathcal{X}_u$  can be converted to behavioral plan  $\sigma_u \in \Sigma_u$ . Given a state  $h \in \mathcal{H}_u$  we consider an edge  $(u, v) \in E$  such that  $\Gamma^{uv}$  containing  $h \in \mathcal{H}_u$  and set

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

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

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

The latter implies that  $\sum_{h \in \mathcal{A}(h)} \sigma_u(h, \alpha) = 1$ .

We now need to establish that  $\sigma(h, \cdot)$  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  $h' \in \text{Next}^{\Gamma^{uv}}(h, \alpha, u)$  and  $h'' \in \text{Next}^{\Gamma^{uv'}}(h, \alpha, u)$  we have that  $x_u(h') = x_u(h'')$  and thus  $\sigma(h, \alpha) = \frac{x_u(h'')}{x_u(h)}$ .

Finally we need to argue that if  $h_1, h_2 \in \mathcal{H}_u$  with  $\mathcal{I}(h_1) = \mathcal{I}(h_2)$ , then  $\sigma(h_1, \alpha) = \sigma(h_2, \alpha)$  for all  $\alpha \in \mathcal{A}(h_1) = \mathcal{A}(h_2)$ . Let  $\sigma(h_1, \alpha) = \frac{x_u(h'_1)}{x_u(h)}$  for some  $h'_1 \in \text{Next}(h_1, \alpha, u)$  and  $\sigma(h_2, \alpha) = \frac{x_u(h'_2)}{x_u(h)}$  for some  $h'_2 \in \text{Next}(h_2, \alpha, u)$ . Then by Constraint 3 of Definition 11 we get that  $x(h'_1) = x(h'_2)$  and thus  $\sigma(h_1, \alpha) = \sigma(h_2, \alpha)$ .

### B.3 Proof of Lemma 3

First, since Equations (6), (7) are defined on the product of treplexes  $\mathcal{X}$ , let us decompose the equations from the perspective of an arbitrary agent  $u$ . Specifically, for some  $x_u^t, u \in \{1, \dots, n\}$  it holds that the inner product  $\langle x, R \cdot x^{t-1} \rangle$ ,  $x \in \mathcal{X}$  can be decomposed into inner products of the form  $\langle x, R \cdot x^{t-1} \rangle$ , where  $x$  is now in the individual treplex  $\mathcal{X}_u$ . 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  $\sum_{(u,v) \in E} A^{u,v} \cdot x_v^t$ . This gives us the following:

$$x_u^t = \operatorname{argmin}_{x \in \mathcal{X}_u} \left\{ \eta \left\langle x, - \sum_{v:(u,v) \in E} A^{uv} \cdot x_v^{t-1} \right\rangle + D_\psi(x, \hat{x}_u^t) \right\} \quad (8)$$

$$\hat{x}_u^{t+1} = \operatorname{argmin}_{x \in \mathcal{X}_u} \left\{ \eta \left\langle x, - \sum_{v:(u,v) \in E} A^{uv} \cdot x_v^t \right\rangle + D_\psi(x, \hat{x}_u^t) \right\} \quad (9)$$

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  $u$  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  $\{x^1, \dots, x^T\}$  are the same between both formulations.

#### B.4 Proof of Lemma 4

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

$$\hat{x}^\top \cdot R \cdot \hat{x} - (x')^\top \cdot R \cdot \hat{x} \leq \epsilon$$

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

$$-\sum_{v:(u,v) \in E} (x'_u)^\top \cdot A^{uv} \cdot \hat{x}_v + \sum_{v:(u,v) \in E} (\hat{x}_u)^\top \cdot A^{uv} \cdot \hat{x}_v \geq -\epsilon \quad \text{for all } x_u \in \mathcal{X}_u$$

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  $x = y \in \mathcal{X}$ .

**Lemma 7.**  $x^\top \cdot R \cdot x = 0$  for all  $x \in \mathcal{X}$ .

*Proof.* Consider a vector  $x \in \mathcal{X}$ . To simplify notation let  $x := (x_1, \dots, x_n)$  where each vector  $x_u \in \mathcal{X}_u$ . Let  $\sigma_u^x \in \Sigma$  denote the behavioral plan for agent  $u$  constructed from the vector  $x_u \in \mathcal{X}_u$  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  $U^u(\sigma^x) = \sum_{v:(u,v) \in E} U_u^{uv}(\sigma_u^x, \sigma_v^x) = \sum_{v:(u,v) \in E} x_u^\top \cdot A^{uv} \cdot x_v$  meaning that

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

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

We will also utilize the following result:

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

$$\sum_{v \in \mathcal{N}_u} x_v^\top \cdot A^{vu} \cdot x_u + \sum_{v \in \mathcal{N}_u} x_u^\top \cdot A^{uv} \cdot x_v$$

remains constant over the range of  $x_u$ .

*Proof.* For any vector  $x := (x_1, \dots, x_n) \in \mathcal{X}$ , consider the vector  $x' \in \mathcal{X}$  such that  $x'_v = x_v$  for all  $v \neq u$ . By Lemma 7 we get that

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

The latter directly implies that

$$\sum_{v \in \mathcal{N}_u} x_v^\top \cdot A^{vu} \cdot x_u + \sum_{v \in \mathcal{N}_u} x_u^\top \cdot A^{uv} \cdot x_v = \sum_{v \in \mathcal{N}_u} x_v^\top \cdot A^{vu} \cdot x'_u + \sum_{v \in \mathcal{N}_u} (x'_u)^\top \cdot A^{uv} \cdot x_v$$

for all  $x_u, x'_u \in \mathcal{X}_u$ . □

*Proof of Lemma 5.* Consider vectors  $x, y \in \mathcal{X}$ . Consider the vector  $y' \in \mathcal{X}$  such that  $y'_v = y'_v$  for all  $v \neq u$ . We first show that

$$x^\top \cdot R \cdot y + y^\top \cdot R \cdot x = x^\top \cdot R \cdot y' + (y')^\top \cdot R \cdot x$$



Let  $\mathcal{N}_u$  denote the neighbors of agent  $u \in V$ ,

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

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

## B.6 Proof of Lemma 6

Applying Lemma 1 of [29] to our setting, we obtain:

**Lemma 9** ([29]). *Let  $\{x^t, \hat{x}^t\}$  be the sequences produced by Equations (6),(7). Then,*

$$\begin{aligned}
 \sum_{t=1}^T (x^t)^\top \cdot R \cdot x^t - \min_{x \in \mathcal{X}} \sum_{t=1}^T x^\top \cdot R \cdot x^t & \leq \frac{\mathcal{D}^2}{\eta} + \frac{1}{2} \sum_{t=1}^T \|R \cdot x^t - R \cdot x^{t-1}\|^2 \\
 & + \frac{1}{2} \sum_{t=1}^T \|x^t - \hat{x}^t\|^2 - \frac{1}{2\eta} \sum_{t=1}^T [\|\hat{x}^t - x^t\|^2 + \|\hat{x}^t - x^{t+1}\|^2]
 \end{aligned}$$

where  $\mathcal{D}$  is the diameter of the treplex polytope  $\mathcal{X}$ .

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

$$\begin{aligned}
 & \sum_{t=1}^T (x^t)^\top \cdot R \cdot x^t - \min_{x \in \mathcal{X}} \sum_{t=1}^T x^\top \cdot R \cdot x^t \\
 \leq & \frac{\mathcal{D}^2}{\eta} + \frac{1}{2} \sum_{t=1}^T \|R \cdot x^t - R \cdot x^{t-1}\|^2 - \frac{1}{4\eta} \cdot \sum_{t=1}^T [\|\hat{x}^t - x^t\|^2 + \|\hat{x}^t - x^{t+1}\|^2] \\
 \leq & \frac{\mathcal{D}^2}{\eta} + \frac{1}{2} \sum_{t=1}^T \|R \cdot x^t - R \cdot x^{t-1}\|^2 - 2\|R\|^2 \cdot \sum_{t=1}^T [\|\hat{x}^t - x^t\|^2 + \|\hat{x}^t - x^{t+1}\|^2] \\
 \leq & \frac{\mathcal{D}^2}{\eta} + \frac{\|R\|^2}{2} \sum_{t=1}^T \|x^t - x^{t-1}\|^2 - \|R\|^2 \cdot \sum_{t=1}^T \|x^t - x^{t-1}\|^2 \\
 \leq & \frac{\mathcal{D}^2}{\eta}
 \end{aligned}$$

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

## B.7 Proof of Theorem 1

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

$$\min_{x \in \mathcal{X}} x^\top \cdot R \cdot \hat{x} \geq -\Theta \left( \frac{\mathcal{D}^2 \|R\|^2}{T} \right)$$

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

$$\hat{x}^\top \cdot R \cdot \hat{x} \leq \min_{x \in \mathcal{X}} x^\top \cdot R \cdot \hat{x} + \Theta \left( \frac{\mathcal{D}^2 \|R\|^2}{T} \right)$$

meaning that  $(\hat{x}, \hat{x})$  is a  $\Theta \left( \frac{\mathcal{D}^2 \|R\|^2}{T} \right)$ -approximate symmetric Nash Equilibrium of the symmetric game  $(R, R)$ . By Lemma 4 we get that  $\hat{x}$  is a  $\Theta \left( \frac{\mathcal{D}^2 \|R\|^2}{T} \right)$ -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  $\mathcal{X}^* := \{x^* \in \mathcal{X} : \min_{x \in \mathcal{X}} x^\top \cdot R \cdot x^* = 0\}$ , which describes the set of symmetric Nash Equilibria.

In order to establish Theorem 3, we follow the approach and notation of [35], with minor modifications along the way to apply the steps to our setting. Applying Lemma 1 of [35]) to the Equations (6), (7) we get the following lemma:

**Lemma 10** ([35]). *Let  $\{x^t, \hat{x}^t\}_{t \geq 1}$  be the sequence of strategy vectors produced by Equations (6), (7) for  $\eta \leq 1/8\|R\|^2$ . Then,*

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

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

$$2\eta(R \cdot x^t)^\top (x^t - x) \leq \|\hat{x}^t - x\|^2 - \|\hat{x}^{t+1} - x\|^2 - \|\hat{x}^{t+1} - x^t\|^2 - \frac{15}{16}\|x^t - \hat{x}^t\|^2 + \frac{1}{16}\|\hat{x}^t - x^{t-1}\|^2 \quad (10)$$

To simplify notation let  $x^* := \Pi_{\mathcal{X}^*}(\hat{x}^t) \in \mathcal{X}^*$  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

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

where the last inequality follows by the fact that  $(x^*, x^*)$  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,

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

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

$$\text{dist}^2(\hat{x}^{t+1}, \mathcal{X}^*) \leq \text{dist}^2(\hat{x}^t, \mathcal{X}^*) - \|\hat{x}^{t+1} - x^t\|^2 - \frac{15}{16}\|x^t - \hat{x}^t\|^2 + \frac{1}{16}\|\hat{x}^t - x^{t-1}\|^2 \quad (11)$$

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

$$\Theta^{t+1} \leq \Theta^t - \frac{15}{16}\xi^t \quad (12)$$

As in [35], we now lower bound  $\xi^t$  by a quantity related to  $\text{dist}^2(\hat{x}^{t+1}, \mathcal{X}^*)$  which will then give us a convergence rate for  $\Theta^t$ . To do so we need to establish a property that is known as saddle-point metric subregularity ([35]).

**Lemma 11.** (*Saddle-Point Metric Subregularity (SP-MS)*) *For any  $x, x' \in \mathcal{X} \setminus \mathcal{X}^*$ ,*

$$\sup_{x' \in \mathcal{X}} \frac{(R \cdot x)^\top (x - x')}{\|x - x'\|} \geq c \cdot \|x - \Pi_{\mathcal{X}^*}(x)\|$$

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 [35]. For the sake of completeness, we conclude the section with this analysis.

**Lemma 12** ([35]). *If the parameter  $\eta$  in Equations (6), (7) is selected less than  $1/8\|R\|^2$  then for any  $t \geq 0$  and  $x' \in \mathcal{X}$  with  $x' \neq \hat{x}^{t+1}$ ,*

$$\|\hat{x}^{t+1} - x^t\|^2 + \|x^t - \hat{x}^t\|^2 \geq \frac{32}{81}\eta^2 \frac{[(R \cdot \hat{x}^{t+1})^\top (\hat{x}^{t+1} - x')]^2_+}{\|\hat{x}^{t+1} - x'\|^2}$$

where  $[a]_+ := \max\{a, 0\}$ , and similarly, for  $x' \neq x^{t+1}$ ,

$$\|\hat{x}^{t+1} - x^{t+1}\|^2 + \|x^t - \hat{x}^{t+1}\|^2 \geq \frac{32}{81}\eta^2 \frac{[(R \cdot x^{t+1})^\top (x^{t+1} - x')]^2_+}{\|x^{t+1} - x'\|^2}$$

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

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

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  $\|x^{t-1} - x^t\| \rightarrow 0$  as  $T \rightarrow \infty$ . Thus,  $x^t$  converges to a point as  $T \rightarrow \infty$ . 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:

$$\|\hat{x}^{t+1} - x^t\|^2 \leq \xi^t \leq \frac{16}{15} \Theta^t \leq \dots \leq \frac{16}{15} \Theta^1 \quad (13)$$

Then we have:

$$\begin{aligned} \xi^t &\geq \frac{1}{2} \|\hat{x}^{t+1} - x^t\|^2 + \frac{1}{2} (\|\hat{x}^{t+1} - x^t\|^2 + \|x^t - \hat{x}^t\|^2) \\ &\geq \frac{1}{2} \|\hat{x}^{t+1} - x^t\|^2 + \frac{16\eta^2}{81} \sup_{x' \in \mathcal{X}} \frac{[(R \cdot \hat{x}^{t+1})^\top (\hat{x}^{t+1} - x')]^2_+}{\|\hat{x}^{t+1} - x'\|^2} && \text{(Lemma 12)} \\ &\geq \frac{1}{2} \|\hat{x}^{t+1} - x^t\|^2 + \frac{16\eta^2 C^2}{81} \|\hat{x}^{t+1} - \Pi_{\mathcal{X}^*}(\hat{x}^{t+1})\|^2 && \text{(SP-MS condition)} \\ &\geq \min \left\{ \frac{16\eta^2 C^2}{81}, \frac{1}{2} \right\} (\|\hat{x}^{t+1} - x^t\|^2 + \|\hat{x}^{t+1} - \Pi_{\mathcal{X}^*}(\hat{x}^{t+1})\|^2) && \text{(Equation 13)} \\ &= C_2 \Theta^{t+1} \end{aligned}$$

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

$$\Theta^{t+1} \leq \Theta^t - C_2 \Theta^{t+1} \quad (14)$$

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

$$\Theta^t \leq (1 + C_2)^{-t+1} \Theta^1 \leq 2\Theta^1 (1 + C_2)^{-t} \quad (15)$$

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

$$\begin{aligned} \text{dist}^2(x^t, \mathcal{X}^*) &\leq \|x^t - \Pi_{\mathcal{X}^*}(\hat{x}^{t+1})\|^2 \\ &\leq 2\|\hat{x}^{t+1} - \Pi_{\mathcal{X}^*}(\hat{x}^{t+1})\|^2 + 2\|\hat{x}^{t+1} - x^t\|^2 \\ &\leq 32\Theta^{t+1} \leq 32\Theta^t \end{aligned}$$

Combining this observation with Equation 15 we get that

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

where  $C_2 = \min \left\{ \frac{16\eta^2 C^2}{81}, \frac{1}{2} \right\}$ , 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  $x \in \mathcal{X}$  the following holds:*

$$- \min_{x' \in \mathcal{X}} x'^\top R \cdot x \geq c \cdot \|x - \Pi_{\mathcal{X}^*}(x)\|. \quad (16)$$

for some game-dependent constant  $c \in (0, 1)$ .

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

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

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

□

### B.10 Proof of Lemma 13

The proof of this lemma follows the basic steps in the proof of Theorem 5 in [35], with some necessary modifications. We remind the reader that for the purposes of the proof, we defined the set of symmetric Nash Equilibria as  $\mathcal{X}^* = \{x^* : \min_{x \in \mathcal{X}} x^{\top} \cdot R \cdot x^* = 0\}$ . 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  $\min_{x \in \mathcal{X}} x^{\top} \cdot R \cdot x^* = 0$ . 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

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

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

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

$$\mathcal{X} := \{x : \alpha_i^{\top} \cdot x \leq \beta_i \text{ for } i = 1, \dots, L\}$$

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

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

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

Now fix 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

- $\alpha_i^{\top} \cdot x^* = \beta_i$  for  $i = 1, \dots, \ell$
- $c_i^{\top} \cdot x^* = 0$  for  $i = 1, \dots, k$

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

*Proof.* Let assume that  $\{c_i^{\top} \cdot x = 0 \text{ for } i = 1, \dots, k\}$ . Since  $x \notin \mathcal{X}^*$  there exists at least one vertex  $v \in \mathcal{V}(\mathcal{X})$  such that  $v^{\top} \cdot R \cdot x < 0$ . The latter implies that there exists  $x' \in \mathcal{X}$  lying in line segment between  $x$  and  $x^*$  such that  $v^{\top} \cdot R \cdot x \geq 0$  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  $\mathcal{X}^*$  at  $x^*$  is

$$\mathcal{N}_{x^*} = \{x' - x^* : x^* = \Pi_{\mathcal{X}^*}(x')\}$$

From a standard result in linear programming literature [35], we know that the normal cone can be written in the following form:

$$\mathcal{N}_{x^*} = \left\{ \sum_{i=1}^{\ell} p_i \cdot \alpha_i + \sum_{i=1}^k q_i \cdot c_i \text{ for some } p_i, q_i \geq 0 \right\}$$

Again, following the steps of [35], we have the following claim:

**Claim 1.** For any  $x \in \mathcal{X}$  such that  $x^* = \Pi_{x \in \mathcal{X}^*}(x)$  the vector  $x - x^*$  belongs in the set

$$\mathcal{M}_{x^*} = \left\{ \sum_{i=1}^{\ell} p_i \cdot \alpha_i + \sum_{i=1}^k q_i \cdot c_i : p_i, q_i \geq 0, \alpha_j^\top \cdot \left( \sum_{i=1}^{\ell} p_i \cdot \alpha_i + \sum_{i=1}^k q_i \cdot c_i \right) \leq 0 \right\}$$

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

$$\alpha_i^\top (x - x^*) \leq 0, \quad \forall i \in 1, \dots, \ell$$

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

$$\begin{aligned} \alpha_i^\top (x - x^*) &= (\alpha_i^\top x^* - b_i) + \alpha_i^\top (x - x^*) && (i\text{-th constraint is tight at } x^*) \\ &= \alpha_i^\top (x^* + x - x^*) - b_i \\ &= \alpha_i^\top x - b_i \leq 0 && (x \in \mathcal{X}) \end{aligned}$$

□

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

*Proof.* Note that  $\frac{x-x^*}{\|x-x^*\|} \in \mathcal{M}_{x^*}$  because  $0 \neq x - x^* \in \mathcal{M}_{x^*}$  and  $\mathcal{M}_{x^*}$  is a cone. Furthermore,  $\frac{x-x^*}{\|x-x^*\|} \in \{v \in \mathbb{R}^M : \|v\|_\infty \leq 1\}$ . Thus,  $\frac{x-x^*}{\|x-x^*\|} \in \mathcal{M}_{x^*} \cap \{v \in \mathbb{R}^M : \|v\|_\infty \leq 1\}$ , which is a bounded subset of the cone  $\mathcal{M}_{x^*}$ .

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

$$\left\{ \sum_{i=1}^{\ell} p_i \cdot \alpha_i + \sum_{i=1}^k q_i \cdot c_i : 0 \leq p_i, q_i \leq C', \forall i \right\} \supseteq \mathcal{M}_{x^*} \cap \{v \in \mathbb{R}^M : \|v\|_\infty \leq 1\} := \mathcal{P}.$$

First note that  $\mathcal{P}$  is a polytope. For every vertex  $\hat{v}$  of  $\mathcal{P}$ , the smallest  $C'$  such that  $\hat{v}$  belongs to the left-hand side set above is the solution to the following linear program:

$$\begin{aligned} \min_{p_i, q_i, C'_\hat{v}} \quad & C'_\hat{v} \\ \text{s.t.} \quad & \hat{v} = \sum_{i=1}^{\ell} p_i \cdot \alpha_i + \sum_{i=1}^k q_i \cdot c_i, \quad 0 \leq p_i, q_i \leq C'_\hat{v}. \end{aligned}$$

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

□

Now, again following [35], we can piece together all of the auxiliary results to show Lemma 13. Define  $A_i := \alpha_i^\top (x - x^*)$  and  $C_i := c_i^\top (x - x^*)$ . By Claim 2, we can write  $x - x^*$  as  $\sum_{i=1}^{\ell} p_i \cdot \alpha_i + \sum_{i=1}^k q_i \cdot c_i$  where  $0 \leq p_i, q_i \leq C' \|x - x^*\|$ . Thus:

$$\sum_{i=1}^{\ell} p_i \cdot A_i + \sum_{i=1}^k q_i \cdot C_i = \left( \sum_{i=1}^{\ell} p_i \cdot \alpha_i + \sum_{i=1}^k q_i \cdot c_i \right)^\top (x - x^*) = \|x - x^*\|^2$$

Moreover, since  $x - x^* \in \mathcal{M}_{x^*}$  by Claim 1, we have

$$\sum_{i=1}^{\ell} p_i \cdot A_i = \sum_{i=1}^{\ell} p_i \cdot \alpha_i \leq 0$$

and

$$\sum_{i=1}^k q_i \cdot C_i \leq \left( \max_{i \in \{1, \dots, k\}} C_i \right) \sum_{i=1}^k q_i \leq \left( \max_{i \in \{1, \dots, k\}} C_i \right) k C' \|x - x^*\|$$

The first inequality follows because  $p_i \geq 0$ . The second inequality follows because  $\max_{i \in \{1, \dots, k\}} C_i > 0$  (by Lemma 15) and  $0 \leq q_i \leq C' \|x - x^*\|$ .

Combining the above, we obtain:

$$\max_{i \in \{1, \dots, k\}} C_i \geq \frac{1}{k C'} \|x - x^*\|$$

Now, note that:

$$\max_{i \in \{1, \dots, k\}} C_i = \max_{i \in \{1, \dots, k\}} (c_i^\top x - d_i) \leq \max_{i \in \{1, \dots, |\mathcal{V}(\mathcal{X})\}} (c_i^\top x - d_i) = \max_{x' \in \mathcal{X}} (x'^\top R x)$$

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

$$-\min_{x' \in \mathcal{X}} (x'^\top R x) \geq \frac{1}{k C'} \|x - x^*\|.$$

Here  $k$  and  $C'$  only depend on the set of tight constraints at  $x^*$ . There are only finitely many sets of tight constraints, so there exists a constant  $C > 0$  such that  $-\min_{x' \in \mathcal{X}} (x'^\top R x) \geq \frac{1}{k C'} \|x - x^*\|$  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 first 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 dense 4-node graph (shown in Figure 3). Specifically, each game is characterized by a  $3 \times 3$  symmetric matrix which represents the sequence form of an extensive form game written as a matrix. For each run of the simulation, we first create the games which are to be played, randomly generating matrices with elements in  $[0, 1]$ . Then, we optimize for the choice of stepsize  $\eta$ , 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  $\text{dist}^2(x^t, \mathcal{X}^*)$  in the plots. It is worth noting that the 4-node graph takes significantly longer to arrive at the last iterate compared to the 3-node graph.

**Kuhn Poker.** Kuhn poker is a simplified version of poker proposed by [18]. 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 dimension  $|\mathcal{X}| \times |\mathcal{X}| = 13 \times 13$  and the corresponding payoff matrix can be easily computed by hand. For the simulation we show in Figure 2, 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.

**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.

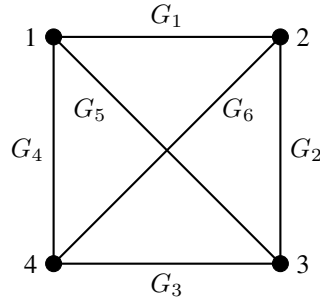


Figure 3: 4-node graph for randomized EFGs. Each node represents a player and each edge represents a game  $G_i$  between the corresponding players.

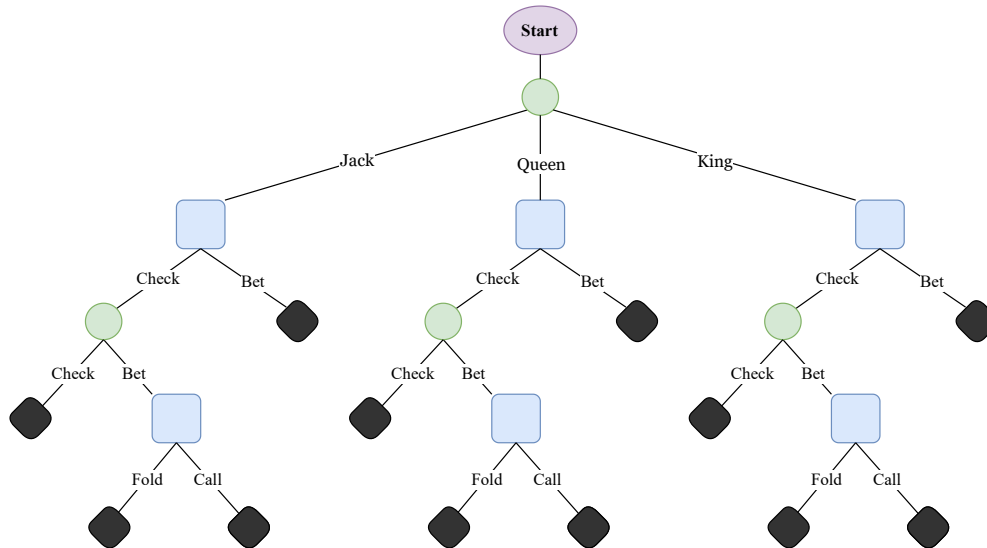


Figure 4: Extensive form representation of Kuhn poker from the perspective of one player. The blue nodes represent decision points for the player, green nodes represent observation points (either the player observes their card or the other player takes an action) and finally the black nodes denote the terminal states of the game.

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.

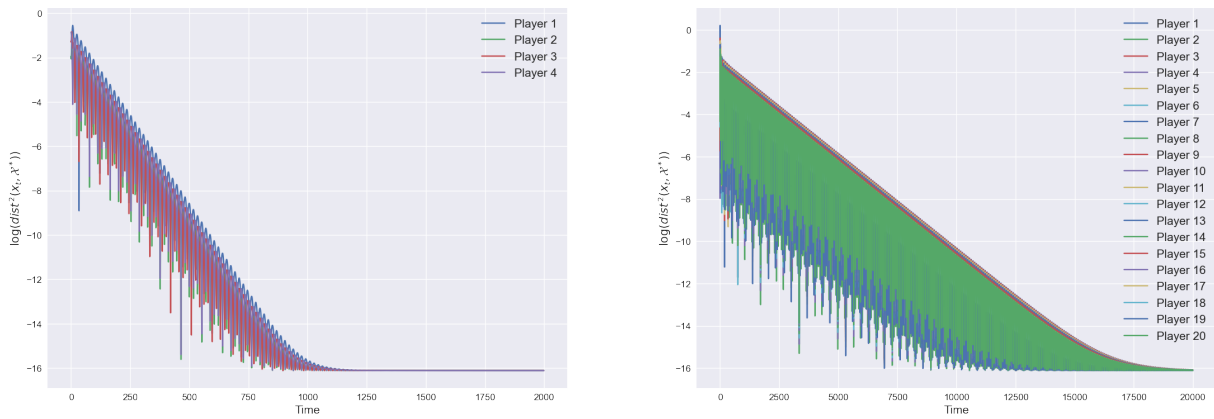


Figure 5: Simulations using OGA in network Matching Pennies games. (Left) Convergence times for 4-player game; (Right) Convergence times for 20-player game.