Equilibria and Convergence of Fictitious Play on Network Aggregative Games

Anonymous Author(s)
Affiliation
Address
email

Abstract

Understanding the long term behaviour of learning algorithms is an important object of study if we are to design AI systems that are safe. In this work, we contribute towards the study of learning with multiple agents. We do this through the introduction of learning on Network Aggregative (NA) games, in which each player's reward depends only on its own strategy and a convex combination of its neighbours. In particular, we present a continuous time analysis of the Fictitious Play learning dynamic on NA games. We show that Fictitious Play reaches a fixed point when the game is zero-sum, and provide conditions under which this fixed point corresponds to a Nash equilibrium. In addition, we show that agents learning through Fictitious Play achieve no-regret, regardless of the choice of game. Finally, we present experimental evidence of a family of games for which Fictitious Play reaches a limit cycle and evidence that the introduction of noise has the potential to break this cyclic behaviour and allow agents to reach the Nash equilibrium.

1 Introduction

Multi-agent learning Schwartz [2014] requires a number of agents to adapt in an environment, where each agent responds to the behaviour of the other agents. This feature leads to a fundamentally non-stationary problem, which presents a challenge to designing effective learning policies. Even for a small number of agents in the game, learning has been shown to lead to non-stationary, and even chaotic behaviour Sato *et al.* [2002], a problem that becomes even more pronounced as the number of agents increases Sanders *et al.* [2018]. In this light, it would seem that, when there is a large population of agents, it is extremely difficult to understand the long-term behaviour of multi-agent learning Chotibut *et al.* [2021]. This difficulty, of course, poses a problem for AI safety, for which it is desirable that learning systems are designed so that their ultimate behaviour is guaranteed to satisfy some predetermined goals.

To solve this problem, a promising approach is to reduce the many-player game to something that is tractable. A number of reductions have been proposed, most notably *Mean Field* games and *Aggregative* games. The former make the assumption of an infinite number of agents, so that the population can be represented through a distribution over players' states Caines *et al.* [2006]. Every agent then updates their action profiles depending on this distribution. In the latter, each agent considers a real valued function which is a convex combination of the states of the other agents. Both approaches allow for a many-player game to be reduced to a set of two-player games. Further, both approaches have been the object of rigorous study, which has shown that agents reach an equilibrium when learning on such games (cf. Perrin *et al.* [2020] for Mean Field Games and Jensen [2010]; De Persis and Grammatico [2020]; Parise *et al.* [2020] for Aggregative Games). However, both also present a fundamental limitation. Namely, they both require that agents have access to the action profiles of the entire population. This could be through communication with all other agents, or

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

through the intervention of a central coordinator who is able access the entire population. Whilst recent work aims to relax this assumption through the introduction of noise Perrin *et al.* [2020] or partial observability Elie *et al.* [2020], the requirement that each agent updates their actions based on the entire population is rather strong and not always supported by empirical evidence.

In this study, we investigate a variant of aggregative games: *Network Aggregative* (NA) games. This framework assumes that each agent updates their actions according only to those agents with whom they are connected on an underlying network. This assumption significantly relaxes the communication load on each agent and lifts the need for a central coordinator. Recent work on NA games has shown that it is possible for agents to reach an equilibrium strategy in an entirely distributed manner Koshal *et al.* [2016]; Shokri and Kebriaei [2020, 2021]; Parise *et al.* [2015]. We contribute in this direction by analysing the long-term behaviour of multi-agent learning on NA games. In particular, we analyse the Fictitious Play learning algorithm Brown P [1949]; Harris [1998], in which agents are assumed to be myopic, in that they react solely to the past behaviour of the others.

Contributions. The main contribution of this work is to introduce and study convergence of learning on Network Aggregative games through the action of Fictitious Play (FP). In particular, we aim to establish guarantees of convergence for learning under certain classes of games. Such guarantees allow for games to be designed so that the long-term behaviour of learning is predictable, regardless of how many agents are engaged in the game. As such, we are able to avoid the pitfalls of chaotic dynamics in games with many players (also called 'agents'), as shownin Sanders *et al.* [2018].

We first show that, in NA games played with FP, a Nash equilibrium exists and that FP admits solutions in this setting. Specifically, we study zero-sum games and show that FP converges to a fixed point which, for a network without any self-loops (i.e., agents are not 'connected' to themselves), corresponds to a Nash equilibrium. In addition, we find that, for games which are not zero-sum, agents following FP are able to achieve no regret.

Further, we explore FP through numerical simulations to check whether it always converges. We answer this question negatively, by finding a family of games in which action profiles cycle around the Nash equilibrium. Finally, our experiments document how noise affects the convergence of FP, suggesting that, under the presence of noise, the algorithm still reaches a fixed point, but perhaps not the Nash equilibrium. This presents an interesting avenue for future research.

To the best of our knowledge, this contribution is the first time that a learning algorithm, which stems from game-theoretic literature, has been studied on Network Aggregative games, typically considered in the context of control and optimisation.

Related Work. **Network Aggregative Games** are a recent extension Parise *et al.* [2015] of aggregative games, obtained by adding an underlying structure to the population. Since its introduction, distributed algorithms have been built with the aim of finding NE in NA games. In particular, Parise *et al.* [2015, 2020] consider the case in which payoffs are given by Lipschitz functions with unique minimisers and apply standard topological fixed-point arguments towards designing algorithms that converge to the NE. Another approach for searching for distributed NE the projected gradient (resp. subgradient) dynamics, which is explored in Zhu *et al.* [2021] (resp. Shokri and Kebriaei [2020, 2021]). In all these works, the cost function is assumed to be convex, and therefore has a unique minimiser. In fact, this is a common assumption in works about NA games Zhu *et al.* [2021]; Lei *et al.* [2020] which, we believe, is due to its ubiquity in control settings. We have not yet come across works which consider NA games from the point of view of payoff matrices, which are more common in multi-agent learning settings. Furthermore, to the best of our knowledge, this is the first work which introduces the application of a learning algorithm in NA games.

Fictitious Play was introduced as a 'natural' way to approximate Nash equilibria in zero-sum games Brown P [1949]. Since then, a number of results on convergence have been proved for two-player games Robinson [1951]; Miyasawa [1961]; Metrick and Polak [1994]; Berger [2007]; Monderer *et al.* [1997]; Monderer and Shapley [1996]. However, works looking at FP with more than two agents is sparse. In Sela [1999] multi-player games are decomposed in two-plpayer games between each pair of players in the game. Each agent's payoff is given by the sum of payoffs in all of these subgames. It was found that, if this game is zero-sum, then FP converges. Similar results for more than two players were found for games where all agents share the same payoff in Monderer and Shapley [1996]. In

Perrin *et al.* [2020], the action of FP was considered in a Mean Field game, with convergence in zero-sum games. The most general result, and the one most similar to our own, appears in Ewerhart and Valkanova [2020], in which the authors show that FP converges in network games, where each agent is engaged in a two-player game with each of their neighbours Our work extends the analysis of FP in multi-player games by considering its action in NA games, so that the agents do not play individual games against each of their neighbours, but rather a single game against the aggregate of their neighbours. We also go beyond the zero-sum requirement by exteding a result for two-player games in Ostrovski and van Strien [2014] which showed that FP achieves no-regret in the multi-player setting.

2 Preliminaries

In this section we introduce the Network Aggregative game framework, as well as defining Fictitious Play on such games.

2.1 Network Aggregative Games

The model we consider consists of a set $\mathcal{N} = \{1, \ldots, N\}$ of agents, who are connected through an underlying interaction graph. More formally:

Definition 1 (Interaction Graph)**.** Given a set $\mathcal{N}$ of agents, an *interaction graph* $I = (\mathcal{N}, (\mathcal{E}, W))$ is such that

- $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$. Then, the set of neighbours of agent μ is denoted as $N^\mu = \{\nu \in \mathcal{N} \mid (\mu, \nu) \in \mathcal{E}\}$.
- $W \in M_N([0, 1])$ is the weight adjacency matrix, whose elements $w^{\mu\nu} \in [0, 1]$ expresses the importance that agent μ places on agent ν. If $(\mu, \nu) \notin \mathcal{E}$ then $w^{\mu\nu} = 0$; $w^{\mu\nu} \in (0, 1]$ otherwise.

Definition 2 (NA Game)**.** A *Network Aggregative game* is a tuple $\Gamma = (I, (S^\mu, u^\mu)_{\mu \in \mathcal{N}})$, where I is an interaction graph, and for every agent $\mu \in \mathcal{N}$, S^μ and u^μ are μ's set of actions (with cardinality $|S^\mu| = n$) and utility function respectively.

We define the *state* of agent μ to be the probability vector $x^\mu \in \mathbb{R}^n$, where x_i^μ is the probability with which agent μ plays action i. This probability vector is often referred to as μ's *mixed strategy*. With this in mind, we can construct, as their state space, the *unit simplex* Δ_μ on agent μ's action set, which is defined as $\Delta_\mu := \{x^\mu \in \mathbb{R}_+^n \mid \sum_i x_i^\mu = 1\}$.

Also associated with each agent is a utility function. For each agent μ and action profile $(x^\mu, x^{-\mu})$, the utility is given as $u^\mu(x^\mu, x^{-\mu})$ in which we use the standard notation $-\mu$ to refer to all agents other than μ. Notice that this requires that each agent plays the same strategy against all of their neighbours.

What is unique about NA games is the structure of the payoffs themselves. Each agent μ receives a *reference* $\sigma^\mu = \sum_{\nu \in N^\mu} w^{\mu\nu} x^\nu$, which is a convex combination of each of their neighbours' state. Agents must optimise their payoff with respect to this reference vector. Thus, instead of considering the actions of the entire population, or playing individual games against each of their neighbours, the agent only considers σ^μ as a 'measurement' of the local aggregate state and optimises with respect to this measurement. This allows us to make the reduction $u^\mu(x^\mu, x^{-\mu}) = u^\mu(x^\mu, \sigma^\mu)$. In particular, we consider that the agent is engaged in a matrix game against the reference vector so that $u^\mu(x^\mu, \sigma^\mu) = x^\mu \cdot A^\mu \sigma^\mu = x^\mu \cdot A^\mu \sum_{\nu \in N^\mu} w^{\mu\nu} x^\nu$, where A^μ is the payoff matrix associated with agent μ. In particular, this means we can rewrite the game Γ with the payoff matrices A^μ in place of the utility functions u^μ.

The agent's goal is to maximise their payoff u^μ with respect to their reference vector σ^μ. As such, we define the best response correspondence BR^μ, which maps every σ^μ to the set $\arg\max_{y \in \Delta_\mu} u^\mu(y, \sigma^\mu)$ Ostrovski and van Strien [2014]. Through the best response function, we can define the Nash equilibrium (NE), a central concept of game theory. The NE condition requires that no rational agent has an incentive to deviate from their current state, as long as the other agents continue to play the NE strategy. This can be formalised by saying that all agents are playing their best response to each other. This leads naturally to the definition of a Nash equilibrium in an NA game as

Definition 3. (NE) The set of vectors $\{\bar{x}^\mu\}_{\mu\in\mathcal{N}}$ is a *Nash equilibrium* if, for all agents μ,

$$\bar{x}^\mu \in BR^\mu(\sigma^\mu) = \arg\max_{x\in\Delta_\mu} u^\mu(x, w^{\mu\mu}x + \sum_{\nu\in N^\mu\setminus\{\bar{\mu}\}} w^{\mu\nu}\bar{x}^\nu).$$

Remark. The notion of Nash equilibrium in NA games is a natural extension of the NE in bimatrix games. In particular, if we consider an NA game with only two players and no self-loops, then Def. 3 yields that $\bar{x}^1$ is an NE iff $\bar{x}^1 \in BR^1(\sigma^1) = \arg\max_{x\in\Delta_1} u^1(x, \bar{x}^2)$, and similarly for $\bar{x}^2$. This is precisely the definition of NE in a two-player game .

In Section 3 we show that NE exist for NA games.

2.2 Continuous Time Fictitious Play

Fictitious Play requires that, at the current time, each agent considers the average state of their opponent in the past, and responds optimally (i.e., play a best response) to this state. In the case of an NA game, each agent considers their reference vector σ^μ to be their 'opponent'. As such, each agent μ must update their state according to the time-average of σ^μ. To formalise this we define α^μ_σ as the time average of agent μ's reference σ^μ up until time t.

$$\alpha^\mu_\sigma = \frac{1}{t}\int_0^t \sigma^\mu(s)\, ds.$$

Using this idea, we follow in the footsteps of Ewerhart and Valkanova [2020] and Harris [1998] to define Fictitious Play in continuous time, but with a slight adaptation for NA games.

Definition 4 (Fictitious Play on Network Aggregative Games)**.** We define Continuous Time Fictitious Play (CTFP) on NA games as a measurable map m with components m^μ such that for all agents μ, $m^\mu : [0,\infty) \to \Delta_\mu$ satisfies $m^\mu(t) \in BR^\mu(\alpha^\mu_\sigma)$ for almost all time $t \geq 1$. Henceforth, m will be called an *NA-CTFP*.

We can think of Def. 4 as saying that the player plays some arbitrary strategy before $t = 1$, but beyond this they must play a best response to the time average of its reference signal. We also refer to this measurable map as a 'path'. In Section 3, we prove that a path m which satisfies Def. 4 exists.

Remark. As an illustration, consider NA games with two players, in which $\mathcal{E} = \{(1,2),(2,1)\}$ and W is a 2×2 matrix with zeros on its leading diagonal and ones on the off diagonal. We write the time-average of both agents' state as

$$\alpha^\mu(t;x) = \frac{1}{t}\int_0^t x^\mu(s)\, ds \text{ for } \mu \in \{1,2\}$$

In this manner, $\alpha^\mu(t;m)$ denotes the time average of the strategies played by agent μ up to time t when the strategies are given by $x^\mu(t)$. Note that we often reduce the notation to $\alpha^\mu(t)$. Then, fictitious play requires that the agents update their strategy as $x^1(t) \in BR^1(\alpha^2(t))$ and $x^2(t) \in BR^2(\alpha^1(t))$. It can be seen, therefore, that NA-CTFP is a natural extension of CTFP in the classic two-player setting Josef Hofbauer [2006].

2.3 Assumptions

With the above preliminaries in place, we can state the assumptions that we make in this study.

Assumption 1. The weighted adjacency matrix W is constant and *row stochastic* meaning that the sum of elements in each row of W is equal to one. This assumption is made to ensure that the analysis of NA games can be derived as a natural extension of the classic setting of two-player games. We can think of the row stochastic condition as the ability of each agent to prioritise the state information it receives from each of its neighbours. It is also a standard assumption made in the analysis of networks and is straightforward to implement Mai and Abed [2019].

Assumption 2. The payoffs are given through matrix games and, therefore, are bilinear. Payoff matrices have a rich history in game theory and allow for the design of multi-agent systems in computational settings, particularly in the case of task and resource allocation Nisan *et al.* [2007]. It should be noted, however, that game-theoretic analysis is starting to consider various other forms of utility functions, including monotone and convex Parise *et al.* [2020]. We believe that the analysis of Fictitious Play should follow in these developments and we consider it as an important area of future work.

Assumption 3. The cardinality of each action set $|S^\mu|$ is equal for all agents. This is another standard assumption that is made in most game-theoretic settings. However, it should be noted that, in Ewerhart and Valkanova [2020], CTFP is analysed without this assumption.

Assumption 4. The NA game is zero-sum in the sense that $\sum_\mu u^\mu(x^\mu, \sum_{\nu \in N^\mu} w^{\mu\nu} x^\nu) = 0$ for any set $(x^\mu)_{\mu \in \mathcal{N}}$ of states. This is, perhaps, one of the strongest assumptions in our analysis, which is required for the fixed-point analysis. However, in Sec. 4, we perform a regret analysis that considers the long-term behaviour of NA-CTFP without this assumption.

3 Convergence of Fictitious Play in Network Aggregative Games

In this section we present our main results. First, we establish the existence of a Nash equilibrium in NA games, as defined in Def. 3. Then we show that an NA-CTFP (i.e. a path m which satisfies Def. 4) exists. Finally, we show that any NA-CTFP reaches a fixed point when NA Games are zero-sum and that, when the network has no self-loops (i.e., $w^{\mu\mu} = 0$ for all agents μ), NA-CTFP reaches a Nash Equilibrium. For the sake of brevity, we defer the proofs of our statements, as well as the standard topological arguments used to derive them, to the supplementary material (Sections (S4 - S6)).

As a reminder, Def. 3 states that $\bar{x}^\mu$ is an NE iff

$$\bar{x}^\mu \in \arg\max_{x \in \Delta_\mu} u^\mu(x, w^{\mu\mu}x + \sum_{\nu \in N^\mu} w^{\mu\nu}\bar{x}^\nu) = \arg\max_{x \in \Delta_i} \bar{u}^\mu(x, \sum_{\nu \in N^\mu} w^{\mu\nu}\bar{x}^\nu)$$

where we have introduced the surrogate function $\bar{u}$ which keeps x in the first argument and all other agent states x^ν in the second argument. We can find $\bar{u}_i$ through the following argument

$$\begin{aligned} u^\mu(x, w^{\mu\mu}x + \sum_{\nu \in N^\nu} w^{\mu\nu}\bar{x}^\nu) &= x \cdot A^\mu(w^{\mu\mu}x + \sum_{\nu \in N^\mu} w_{\mu\nu}\bar{x}^\nu) && (1) \\ &= x \cdot (w^{\mu\mu}A^\mu)x + \sum_{\nu \in N^\mu} u^{\mu\nu}(x, \bar{x}^\nu) && (2) \\ &\stackrel{\text{def}}{=} \bar{u}^\mu(x, \sum_{\nu \in N^\mu} w^{\mu\nu}\bar{x}^\nu), \end{aligned}$$

where $u^{\mu\nu}(x^\mu, x^\nu) = x^\mu \cdot (w^{\mu\nu}A^\mu)x^\nu$.

Note that, in order to get this formulation, we had to use Assumption 2 to move from (1) to (2).

With these in place, we can build towards our main result, namely the convergence of NA-CTFP to an NE. In order to do this, we first need to establish that the NE, and an NA-CTFP (i.e. a path which satisfies Def. 4) exists.

Lemma 1 (Existence of NE)**.** Under assumption (II), namely that the payoff function achieves a bilinear property, a Nash equilibrium $\{\bar{x}^\mu\}_{\mu \in \mathcal{N}}$ exists.

Lemma 2. There exists a path $m(t)$ which satisfies the property that, for all agents μ, $m^\mu(t) \in BR^\mu(\alpha^\mu_\sigma(t))$ for almost all times $t \geq 1$.

With these results in place, we can show that NA-CTFP converges to a fixed point. In particular, let $\Omega(\alpha)$ be the set of all limit points for $\alpha(t)$. Then, a NA-CTFP path is said to *converge* if $\Omega(\alpha)$ is contained within the set of Nash Equilibria of the game. We adapt the techniques of Ewerhart and Valkanova [2020] to prove that this is the case.

Theorem 1. Any zero-sum NA game (Assumption 4) has the property that, for any NA-CTFP path m, $\alpha(t; m)$ (i.e. the time-averaged state) converges to a set of fixed points.

In the proof of Theorem 1 it can be seen that, if we choose $w^{\mu\mu}$ to be zero for all agents μ, then the fixed point corresponds to an NE. In particular Eqn. (26) in the supplementary materal corresponds to the NE condition when $w^{\mu\mu} = 0$. This leads to the next result.

Corollary 1. With the additional assumption that $w^{\mu\mu} = 0$ for all agents μ, all zero-sum NA games have the property that any NA-CTFP path converges to the set of Nash Equilibria.

4 Fictitious Play in Network Aggregative Games Achieves No Regret

In this section we aim to understand the long term behaviour of NA-CTFP for the case in which NA games are not necessarily zero-sum. To do this, we first introduce the coarse correlated equilibria (CCE) Nisan *et al.* [2007] in the context of NA games as a natural extension to of the two-player case. Then, we show that the NA-CTFP process converges to the set of CCE.

Definition 5 (CCE)**.** A distribution $\mathcal{D}$ over the set $S = \times_\mu S^\mu$ of joint actions is called a *coarse correlated equilibrium* if, for all agents μ and all actions $j \in S^\mu$, we have $\mathbb{E}_{s\sim\mathcal{D}}[u^\mu(s^\mu, s^{-\mu})] \geq \mathbb{E}_{s\sim\mathcal{D}}[u^\mu(j, s^{-\mu})]$.

In words, the above definition says that, if the agents are given a probability distribution with which they can play their actions, then the expected payoff, for all agents, is greater than or equal to the payoff that they would get by playing any of their other available actions, assuming that the other agents keep to the distribution.

For an NA game, a set of actions $s = (s^1, \dots, s^N)$ which is drawn from a joint probability distribution $\mathcal{D}$, also generates a corresponding set of reference vectors $\sigma = (\sigma^1, \dots, \sigma^N)$, where $\sigma^\mu = \sum_{\nu \in N^\mu} w^{\mu\nu} s^\nu$. That is, if we draw action s from $\mathcal{D}$, then we have also drawn σ, which means our CCE condition, Def. 5, can be written as $\mathbb{E}_{s\sim\mathcal{D}}[u^\mu(s^\mu, \sigma^\mu)] \geq \mathbb{E}_{s\sim\mathcal{D}}[u^\mu(j, \sigma^\mu)]$, for all agents μ and actions $j \in S^\mu$.

Now, if by playing with NA-CTFP, the agents reach state $(x^\mu)_{i=1}^N$ with references $(\sigma^\mu)_{i=1}^N$, then we can define a distribution $\mathcal{D} = (\mathcal{D}^1, ..., \mathcal{D}^N)$ so that $(\mathcal{D}^\mu)_{ij} = x_i^\mu \sigma_j^\mu$. Then, the expected payoff that the agent would receive for playing this strategy is

$$\mathbb{E}_{s\sim\mathcal{D}}[u^\mu(s^\mu, \sigma^\mu)] = u^\mu(x^\mu, \sigma^\mu) = x^\mu \cdot A^\mu \sigma^\mu = \sum_{i,j} (A^\mu)_{ij} x_i^\mu \sigma_j^\mu$$

As such, we would say that NA-CTFP has converged to the set of CCE if, in the limit of $t \to \infty$, we have that, for all agents μ and all actions $j \in S^\mu$, $u^\mu(x^\mu, \sigma^\mu) \geq u^\mu(j, \sigma^\mu)$

Remark. As usual, the notion of CCE in an NA game is a natural extension of the CCE for two-player games. In fact, if we consider the NA game to be a two-player game with no self-loops, then we recover exactly the definition of the CCE set in two-player games Ostrovski and van Strien [2014].

Remark. The notion of the CCE set is related to the idea of *average regret* Nisan *et al.* [2007]. Here, we will present what is meant by average regret and state that if at some time t all agents' average regret is non-positive, then the game is said to have reached the CCE set. The reader should consult Ostrovski and van Strien [2014] for an excellent exposition regarding the link between the CCE set and average regret in two-player games which, of course, extends naturally to NA Games.

Average regret, for agent μ is defined as

$$R^\mu = \max_{i' \in S^\mu} \Big\{ \frac{1}{t} \int_0^t u^\mu(e_{i'}^\mu, \sigma(s)) - u^\mu(m^\mu(s), \sigma(s))\, ds \Big\},$$

in which e_i^μ denotes the probability vector in Δ_μ with 1 in the slot i' and 0 everywhere else. Note, this is the *average regret* for the agent μ and, of course, can be related to the *cumulative regret* which is used for analysis in Cesa-Bianchi and Orabona [2021]. To illustrate the average regret, let us consider the case where each agent has only two actions. Then $u^\mu(x^\mu(t), \sigma(t))$ is given by

$$u^\mu(x^\mu(t), \sigma(t)) = \sum_{ij} a_{ij} x_i^\mu \sigma_j^\mu = a_{11} x_1^\mu \sigma_1^\mu + a_{12} x_1^\mu \sigma_2^\mu + a_{21} x_2^\mu \sigma_1^\mu + a_{22} x_2^\mu \sigma_2^\mu \tag{3}$$

On the other hand, let us consider that agent μ's first strategy maximises $u^\mu(e_1^\mu, \sigma(t))$, then

$$u^\mu(e_1^\mu, \sigma(t)) = \sum_{ij} a_{1j} x_i^\mu \sigma_j^\mu = a_{11} x_1^\mu \sigma_1^\mu + a_{12} x_1^\mu \sigma_2^\mu + a_{11} x_2^\mu \sigma_1^\mu + a_{12} x_2^\mu \sigma_2^\mu \tag{4}$$

By comparing equations (3) and (4), we can see that the latter gives the reward that agent μ would have received had they played action 1 throughout the entire play, assuming that the behaviour of the other agents (encoded in σ) does not change. As such, this is a measure of agent μ's regret, in hindsight, for not playing action 1 the entire time. An agent achieves *no regret* if R^μ is non-positive.

Theorem 2. Assuming that $w^{\mu\mu} = 0$, then for any choice of payoff matrix, agents following the NA-CTFP process achieve *no regret* in the limit $t \to \infty$, i.e.

$$\lim_{t\to\infty} \max_{x^{\mu}_{i'} \in S^{\mu}} \left\{ \frac{1}{t} \int_0^t u^{\mu}(x^{\mu}_{i'}(s), \sigma(s)) - u^{\mu}(m^{\mu}(s), \sigma(s))\, ds \right\} = 0 \tag{5}$$

In particular, due to the relation between regret and CCE (Remark 4), NA-CTFP converges to the set of CCE.

We note at this point that a related result was found in Ewerhart and Valkanova [2020]. In particular, the authors showed that, when playing on a zero-sum network game, agents learning through Fictitious Play achieve non-positive regret, regardless of the behaviour of the other agents. This is a slightly stronger condition than the CCE, in which agents achieve non-positive regret if all other agents do not deviate from the distribution $\mathcal{D}$. However, the result in Ewerhart and Valkanova [2020] applies only under the zero-sum condition, whereas Theorem 2 applies in all NA games.

5 Experimental Evaluation

In this section, we investigate NA-CTFP through numerical experiments. In particular, we look beyond zero-sum NA games and show that learning on an NA game can lead to periodic behaviour, rather than convergence to a fixed point. In addition, we aim to understand the behaviour of agents learning through NA-CTFP, when the measurements on their reference signal σ^{μ} is corrupted with noise. The code required to reproduce these simulations is provided in the Supplementary Material.

5.1 Non-convergence of General Two-player Games under NA-CTFP

The purpose of this section is to show that, whilst we proved in Sec. 3 that it converges in zero-sum games, NA-CTFP is not guaranteed to converge in general games, and can in fact give rise to a rich variety of dynamics.

As a example of non-convergence we consider the Shapley family of games Shapley [2016]. In van Strien and Sparrow [2011] this family was shown to contain games for which FP gives periodic and even chaotic behaviour.

As an adaptation, we take the example of a three-player chain, in which player 2 is connected to 1 and 3. The aggregation matrix can be given as

$$W = \begin{bmatrix} 0 & 1 & 0 \\ w & 0 & 1-w \\ 0 & 1 & 0 \end{bmatrix}, \; w \in (0, 1).$$

We first consider the zero-sum case to show that it does indeed converge to an equilibrium as expected. Note that the zero-sum condition (Assumption 4) given for the three-player chain is given as

$$x \cdot Ay + y \cdot B(wx + (1-w)z) + z \cdot Cy = 0. \quad \forall x, y, z \in \Delta_1 \times \Delta_2 \times \Delta_3 \tag{6}$$

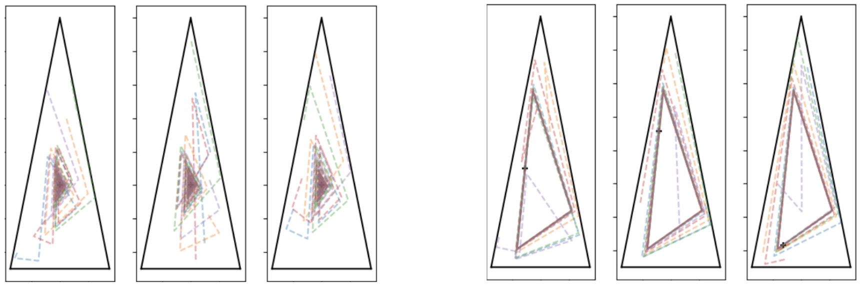
in which we use the notation that x, y, z (resp. A, B, C) denote the strategies (resp. payoffs) of agents 1, 2 and 3 respectively. This condition is satisfied if we fix B and choose

$$\begin{aligned} A &= -wB^T \\ C &= -(1-w)B^T. \end{aligned} \tag{7}$$

As such in the following example, we will set

$$B = \begin{bmatrix} -\beta & 1 & 0 \\ 0 & -\beta & 1 \\ 1 & 0 & -\beta \end{bmatrix} \tag{8}$$

with the choice $\beta \approx 0.576$ and set A and C according to the above with the choice $w \approx 0.288$. These choices are arbitrary and, as we discuss below, the results of this Section were found to hold for a



(a) (b)

Figure 1: Orbits of the Fictitious Play in the Three-Player Chain for which the NE lies at the centre of the simplex. (a) Payoffs are given by (7). The plot shows NA-CTFP yields convergence to the NE (b) NA-CTFP showing cycles around the NE when payoffs are modified slightly (c.f. Sec. 5.1).

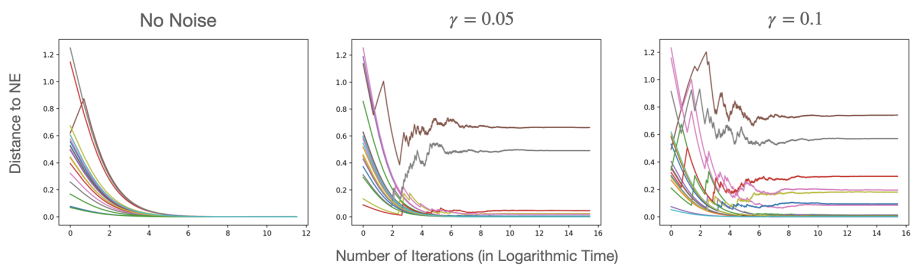


Figure 2: Trajectories of NA-CTFP in a 20-player game with additive noise (Left) No noise is introduced and learning converges directly to an NE. (Middle) $\gamma = 0.05$, the trajectories converge to a fixed point but removed from the NE. (Right) $\gamma = 0.1$, the trajectories converge to a fixed point which is even further away from the NE.

range of choices of β and w. The resulting orbits can be seen in Figure 1a, in which, for each player, they converge to the Nash Equilibrium which lies in the centre of the simplex.

Let us now make the slight modification in the definition of C so that $C = -(1-w)B$, with no alteration to A. The modification itself is small, however it results in the zero-sum assumption being violated. With the same choices of β and w, this results in the periodic orbit seen in Figure 1b. Here, the orbits reach a stable limit cycle which to be centred around the interior NE.

As such, we can see that convergent behaviour is not necessarily the norm in the NA-CTFP dynamics. In fact, for the family of games discussed above, we were unable to find non-periodic behaviour for any choice of β strictly between 0.5 and 1 for any w between 0.2 and 0.8 (so that the influence of player 1 and player 3 on player 2 is not negligible). This suggests that, far from being rare, in fact NA-CTFP lends itself to an incredibly rich variety of dynamics which can be explored as future work.

5.2 Convergence under the Addition of Noise

The Fictitious Play process in NA games requires that, at each time step, an agent takes a 'measurement' of the aggregate strategy of its neighbours. It is on this measurement that they update their own strategy. It stands to reason then, that in real environments this measurement may be corrupted by noise. As such, we investigate the effect that introducing additive noise has on NA-CTFP in a zero-sum NA game. We do this in the following manner: at each time step, the reference signal $\sigma^{\mu}(t)$ is adjusted to $\sigma^{\mu} + \gamma\xi$ where ξ is drawn from the standard normal distribution (zero mean and unit variance). By varying γ, we vary the strength of the noise. We vary γ up to 0.5 since, above this value, noisy measurements are likely to lie outside of the simplex. Since σ^{μ} is constrained to lie within Δ, we can consider the range $\gamma \in [0, 0.5]$ to be the *physical region*, in which noise is meaningful.

In Figure 2, we consider a zero-sum NA game with 20 players. When there is no noise, it can be seen that FP reaches a fixed point which, since we set $w^{\mu\mu} = 0$, corresponds to an NE. After increasing γ,

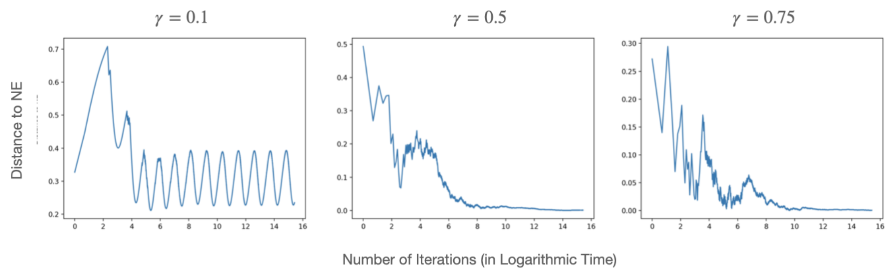


Figure 3: Trajectories of NA-CTFP on the Three Player Chain of Section 5.1 with additive noise. (Left) $\gamma = 0.1$ leads to a decrease in the size of the cyclic orbit (Middle) $\gamma = 0.5$, no periodicity is seen but the trajectory converges to the NE (Right) $\gamma = 0.75$, NA-CTFP still converges, though after a greater amount of time has elapsed.

however, we find that the agents no longer converge to this NE, but rather shift away from it. What is interesting, however, is that the orbits do still reach a stationary state in the long run which suggests that FP is still able to converge with the introduction of noise.

In Figure 3 we revisit the Three-Player Chain of Section 5.1, now under the influence of additive noise. For the sake of brevity, we only display the distance to the Nash Equilibrium of the first player's action, since the other agents behave in the same way. It can be seen that a small amount of noise has the effect of decreasing the size of the periodic orbit. However, as γ is increased to 0.5, the algorithm seems to exhibit convergence to the NE. The implication is that the addition of noise may cause periodic behaviour to break and lead to the Nash Equilibrium. An interesting point to note is that this behaviour is in stark contrast to the replicator dynamic (RD) Smith [1982], another adaptive algorithm linked to multi-agent learning Mertikopoulos *et al.* [2018]. In Imhof *et al.* [2005] and Galla [2011], it was found that the introduction of random mutations can remove convergent behaviour and instead lead to periodicity.

6 Concluding Remarks

In this work, we have considered the action of the Fictitious Play learning algorithm in Network Aggregative Games and investigated its long term behaviour through a continuous time analysis. We find that, under a zero-sum condition, NA-CTFP converges to a fixed point (Theorem 1). However, we find experimentally that this is not always the case. In fact, we find a family of NA games, based on the Shapley family, for which FP cycles about the NE. For these cases, we also perform a regret analysis which shows that, regardless of the type of game, the FP algorithm achieves no regret. We also investigate the influence of noise on the algorithm and find that even with the introduction of additive noise, FP converges to a fixed point. In fact, for our cyclic family of games, we find that the introduction of noise can actually remove the periodicity, resulting in FP converging to a fixed point.

Our work opens a number of lines for future work. Most notable is the effect of noise. It would be prudent to analyse this theoretically, as was done in Perrin *et al.* [2020], and consider the conditions under which FP will still converge to a fixed point. Furthermore, it would be interesting to investigate the phenomenon we report experimentally in a theoretical framework. Namely, the question of why noise breaks periodicity in FP and results in convergence to an NE should be investigated and, indeed, this is a line which we are currently pursuing. In addition, we note that the *Mann Iteration*, a method of approximating fixed points which is investigated in Parise *et al.* [2020], shares a remarkably similar structure to the discrete variant of FP. This may present an avenue by which NA-CTFP may be analysed in the case of convex cost functions. Finally, we note that in recent years FP in two-player games has shown a remarkable variety of dynamical behaviours, including periodicity and chaos. In our work we have shown convergence to a fixed point and, through experiments, periodicity. It stands to reason, therefore, that a greater variety of dynamical behaviours exist for NA-CTFP for certain classes of games. It would be important to determine what these classes are. Short from being merely a curiosity, this would allow for the identification of games in which NA-CTFP leads to inherently unpredictable behaviour, an important question from the point of view of building Safe and Trusted AI.

Broader Impact

The dynamics of learning is an important consideration for all practitioners. In particular, it has been shown a number of times Mertikopoulos *et al.* [2018] that convergence of learning cannot always be assumed. Rather, learning generally presents much more complex dynamics Galla and Farmer [2013], which only increases as the number of players increases Sanders *et al.* [2018]. Our work presents practitioners who applies Fictitious Play with a case in which rigorous stability properties may be guaranteed. We also elucidate the behaviour of the algorithm under more general assumptions, both by understanding its regret properties as well as through an experimental study of the impact of noise.

As regards FP itself, the learning strategy has strong applications in robotic control Smyrnakis and Veres [2016]; Hernández *et al.* [2013]; Sharma and Gopal [2015] as well as economic modelling von Neumann *et al.* [1944]. As such we Furthermore, the algorithm has links to other learning protocols including the replicator dynamic Benaïm *et al.* [2006] and reinforcement learning Leslie and Collins [2006]. Therefore, we believe that an understanding of FP has subsequent impacts on a number of fields.

Finally, our work has a strong impact on the study of the Network Aggregative Game, which has strong applications in multi-agent control Bianchi and Grammatico [2020]; De Persis and Grammatico [2020]. We believe that our work makes a strong step towards ensuring that systems which learn and adapt on NA games maintain stability and, therefore, can be considered safe.

References

Michel Benaïm, Josef Hofbauer, and Sylvain Sorin. Stochastic approximations and differential inclusions, Part II: Applications. *Mathematics of Operations Research*, 31(4):673–695, 11 2006.

Ulrich Berger. Brown's original fictitious play. *Journal of Economic Theory*, 135(1):572–578, 7 2007.

Mattia Bianchi and Sergio Grammatico. A continuous-time distributed generalized nash equilibrium seeking algorithm over networks for double-integrator agents. In *2020 European Control Conference (ECC)*, pages 1474–1479, 2020.

George W Brown P. Some Notes on the Computation of Games Solutions. Technical report, 4 1949.

Peter E. Caines, Minyi Huang, and Roland P. Malhamé. Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the nash certainty equivalence principle. *Communications in Information and Systems*, 6:221–252, 2006.

Nicolò Cesa-Bianchi and Francesco Orabona. Online learning algorithms. *Annual Review of Statistics and Its Application*, 8(1):165–190, 2021.

Thiparat Chotibut, Fryderyk Falniowski, Michał Misiurewicz, and Georgios Piliouras. Family of chaotic maps from game theory. *Dynamical Systems*, 36(1):48–63, 1 2021.

Claudio De Persis and Sergio Grammatico. Continuous-Time Integral Dynamics for a Class of Aggregative Games with Coupling Constraints. *IEEE Transactions on Automatic Control*, 65(5):2171–2176, 5 2020.

Romuald Elie, Julien Pérolat, Mathieu Laurière, Matthieu Geist, and Olivier Pietquin. On the Convergence of Model Free Learning in Mean Field Games. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):7143–7150, 4 2020.

Christian Ewerhart and Kremena Valkanova. Fictitious play in networks. *Games and Economic Behavior*, 123:182–206, 9 2020.

Tobias Galla and J. Doyne Farmer. Complex dynamics in learning complicated games. *Proceedings of the National Academy of Sciences of the United States of America*, 110(4):1232–1236, 2013.

Tobias Galla. Cycles of cooperation and defection in imperfect learning. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(8), 8 2011.

Christopher Harris. On the Rate of Convergence of Continuous-Time Fictitious Play. *Games and Economic Behavior*, 22(2):238–259, 2 1998.

Erik Hernández, Antonio Barrientos, Jaime Del Cerro, and Claudio Rossi. A multi-robot system for patrolling task via Stochastic Fictitious Play. In *ICAART 2013 - Proceedings of the 5th International Conference on Agents and Artificial Intelligence*, volume 1, pages 407–410, 2013.

Lorens A. Imhof, Drew Fudenberg, and Martin A. Nowak. Evolutionary cycles of cooperation and defection. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 102, pages 10797–10800. 8 2005.

Martin Kaae Jensen. Aggregative games and best-reply potentials. *Econ Theory*, 43:45–66, 2010.

Sylvain Sorin Josef Hofbauer. Best response dynamics for continuous zero–sum games. *Discrete and Continuous Dynamical Systems - B*, 6(1):215–224, 2006.

Jayash Koshal, Angelia Nediæ, and Uday V Shanbhag. Distributed algorithms for aggregative games on graphs. *Operations Research*, 64(3):680–704, 2016.

Jinlong Lei, Uday V. Shanbhag, and Jie Chen. Distributed Computation of Nash Equilibria for Monotone Aggregative Games via Iterative Regularization. In *Proceedings of the IEEE Conference on Decision and Control*, volume 2020-December, pages 2285–2290. Institute of Electrical and Electronics Engineers Inc., 12 2020.

David S Leslie and E J Collins. Individual Q-learning in normal form games. *SIAM Journal on Control and Optimization*, 44(2):495–514, 2006.

Van Sy Mai and Eyad H. Abed. Distributed optimization over directed graphs with row stochasticity and constraint regularity. *Automatica*, 102:94–104, 4 2019.

Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. pages 2703–2717, 2018.

Andrew Metrick and Ben Polak. Fictitious play in 2x2 games: A geometric proof of convergence. *Economic Theory*, 4(6):923–933, 11 1994.

K. Miyasawa. On the convergence of the learning process in a 2x2 non-zero-sum two person game. 1961.

Dov Monderer and Lloyd S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124–143, 5 1996.

Dov Monderer, Dov Samet, and Aner Sela. Belief affirming in learning processes. *Journal of Economic Theory*, 73(2):438–452, 4 1997.

Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani. *Algorithmic game theory*, volume 9780521872829. Cambridge University Press, 1 2007.

Georg Ostrovski and Sebastian van Strien. Payoff performance of fictitious play. *Journal of Dynamics and Games*, 1(4):621–638, 8 2014.

Francesca Parise, Basilio Gentile, Sergio Grammatico, and John Lygeros. Network aggregative games: Distributed convergence to nash equilibria. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 2295–2300, 2015.

Francesca Parise, Sergio Grammatico, Basilio Gentile, and John Lygeros. Distributed convergence to Nash equilibria in network and average aggregative games. *Automatica*, 117:108959, 2020.

Sarah Perrin, Julien Perolat, Mathieu Lauriere, Matthieu Geist, Romuald Elie, and Olivier Pietquin. Fictitious play for mean field games: Continuous time analysis and applications. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 13199–13213. Curran Associates, Inc., 2020.

Julia Robinson. An Iterative Method of Solving a Game. *The Annals of Mathematics*, 54(2):296, 9 1951.

James B T Sanders, J Doyne Farmer, and Tobias Galla. The prevalence of chaotic dynamics in games with many players. *Scientific Reports*, 8(1):4902, 2018.

Yuzuru Sato, Eizo Akiyama, and J. Doyne Farmer. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences of the United States of America*, 99(7):4748–4751, 4 2002.

Howard M. Schwartz. *Multi-Agent Machine Learning: A Reinforcement Approach*. 2014.

Aner Sela. Fictitious play in 'one-against-all' multi-player games. Technical report, 1999.

L. S. Shapley. Some Topics in Two-Person Games. In *Advances in Game Theory. (AM-52)*, pages 1–28. Princeton University Press, 5 2016.

Rajneesh Sharma and M. Gopal. Fictitious play based Markov game control for robotic arm manipulator. In *Proceedings - 2014 3rd International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions, ICRITO 2014*. Institute of Electrical and Electronics Engineers Inc., 1 2015.

Mohammad Shokri and Hamed Kebriaei. Leader-Follower Network Aggregative Game with Stochastic Agents' Communication and Activeness. *IEEE Transactions on Automatic Control*, 65(12):5496–5502, 12 2020.

Mohammad Shokri and Hamed Kebriaei. Network Aggregative Game in Unknown Dynamic Environment with Myopic Agents and Delay. *IEEE Transactions on Automatic Control*, 2021.

John Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, 10 1982.

M. Smyrnakis and S. M. Veres. Fictitious play for cooperative action selection in robot teams. *Engineering Applications of Artificial Intelligence*, 56:14–29, 11 2016.

Sebastian van Strien and Colin Sparrow. Fictitious play in 3x3 games: Chaos and dithering behaviour. *Games and Economic Behavior*, 73(1):262–286, 2011.

John von Neumann, Oskar Morgenstern, and Ariel Rubinstein. *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton University Press, 1 1944.

Rongping Zhu, Jiaqi Zhang, Keyou You, and Tamer Başar. Asynchronous networked aggregative games. 2021.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] All contributions stated in the abstract are included in Sections 3-5 of the paper.
   (b) Did you describe the limitations of your work? [Yes] Sec. 2.3 discusses the assumptions placed on the theoretical results, as well as their implications. We provide numerical simulations to provide insight into the behaviour when some of these assumptions are lifted
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] In Sec. 5, we caution against the over-reliance of convergence in learning algorithms and provide experiments to show that this will not always be the case.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] Full Proofs, as well as the arguments used to derive them, are provided in the Supplementary Material
3. If you ran experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Code is provided in the supplementary material
(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] We provide explanations as to the choice for our range of strength of noise in Sec 5.2.
(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A] Our experiments do not concern statistical results but are rather illustrative of the behaviour of a deterministic algorithm
(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]