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 systems which can be considered safe.In this study, we contribute towards the study of learning with multiple agents. We do this through the introduction of learning on network aggregative 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 such a game. 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 (6) requires a number of agents to adapt in an environment where each agent must adapt in response to the behaviour of the other agents. This leads to a fundamentally non-stationary problem, which presents a challenge to designing effective learning policies. Even when there are a small number of agents in the game, learning has been shown to lead to non-stationary, and even chaotic behaviour (5), and this problem becomes even more pronounced as the number of agents increase (4). In this light, it would seem that, when there is a large population of agents, it is almost impossible to understand the long term behaviour of multi-agent learning (1). This, 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 deal with this problem, one must attempt to reduce the many player game to something which is tractable. A number of reductions have been proposed, most notably *Mean Field* Games and *Aggregative* Games. The former makes the assumption of an infinite number of agents, so that the population can be represented through a distribution over players' states (2) . 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 formulation allows for a many player game to be reduced to a set of two-player games.

Both formulations have been the object of rigorous study and it has been shown that agents reach an equilibrium when learning on such games (c.f. (3) for Mean Field Games and (**? ? ?**) for Aggregative Games). However, both also present a fundamental limitation. Namely, they both require that agents

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

have access to the action profiles of the entire population. This could be through communication with all other agents, or through the intervention of a central coordinator who is able view the entire population. Whilst recent work aims to relax this assumption through the introduction of noise (3) or partial observability (**?**), 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 the aggregative game known as the *Network Aggregative* (NA) Game. This format assumes that there is an underlying communication network through which agents interact. Then each agent updates their actions according only to those agents with whom they communicate. This significantly relaxes the communication load on each agent and lifts the need for a central coordinator. Recent work on the NA game has shown that it is possible for agents to reach an equilibrium strategy in an entirely distributed manner (**? ? ? ?**). We contribute in this direction by analysing the long term behaviour of learning on an NA game. In particular, we analyse the Fictitious Play Learning Algorithm (**? ?**), in which agents are assumed to be myopic, in that they react solely to the previous behaviour of the others.

1.1 Contributions

The main contribution of this work is to introduce learning on the Network Aggregative Game through the action of Fictitious Play (FP).

We first show that, in aggregative games played with FP, a Nash Equilibrium exists and that FP admits solutions in this setting. In particular, we study zero-sum games and show that FP converges to a fixed point, which for a network without any self-loops 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.

Finally, we explore FP through numerical simulations to consider the question of whether it always converges. We answer this in the negative 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. Our experiments suggest 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.

1.2 Related Work

Network Aggregative Games The NA game is a recent introduction (**?**) which presents a variant of aggregative games by adding an underlying structure to the population. Since its introduction, distributed algorithms have been built with the aim of finding NE. In particular, (**? ?**) consider the case in which payoffs are given by Lipschitz functions with unique minimisers and apply standard topological fixed point towards designing algorithms which will converge to the NE. Another approach for distributed NE seeking is the projected gradient (resp. subgradient) dynamics which is explored in (**?**) (resp. (**? ?**)). In all of these works, the cost function is assumed to be convex, and therefore have a unique minimiser. In fact this is a common assumption in works concerning NA Games (**? ?**), 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 the NA setting.

Fictitious Play Fictitious Play was introduced in 1951 (**?**) as a 'natural' way in which to approximate the Nash Equilibrium in zero sum games. It was known at the time that a discrete variant of the algorithm would converge to the NE in two player zero-sum games (**?**). Since then a number of proofs of convergence were determined in 2 player games (**? ? ? ? ?**). Many results hold for both the discrete and continuous variants of FP. However, the work looking at FP where there are more than two agents is sparse. (**?**) considers a game in which a multi-player game is decomposed by requiring that each agent engage in a two person game against each of the opponent. Their 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 will converge. Similar results for more than two players were found for games where all agents share the same payoff in (**?**). In (3), 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, was found by Ewerhart and Valkanova in 2020 (**?**) in which it was found 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 multiplayer 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. In particular we also extend a result for two player games found by Ostrovski and van Strien (**?**) which showed that FP achieves no-regret to the multi-player setting.

2 Preliminaries

In this section we introduce the network aggregative game framework, as well as defining FP on such a game.

2.1 Network Aggregative Games

The model we consider consists of a set $\mathcal{N} = \{1, \dots, N\}$ 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

- $\mathcal{E} \subseteq \mathcal{N}^2$. Then, the set of neighbours of agent μ is denoted as $N^\mu = \{\nu \in \mathcal{N} \mid (\mu, \nu) \in \mathcal{E}\}$.
- $W \in M_N(\mathbb{R})$ 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.

Given Def. 1 of interaction graph, we introduce network aggregative (NA) games:

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 $\Delta_\mu := \{x^\mu \in \mathbb{R}^n \,:\, \sum_i x_i^\mu = 1\}$. Also associated with each agent is a utility function

The utility, for each agent μ and action profile $(x^\mu, x^{-\mu})$, 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. In this format, each agent μ receives a reference $\sigma^\mu = \sum_{\nu\in N^\mu} w^{\mu\nu} x^\nu$, which is a weighted sum of each of their neighbours state.

Then, the agent must optimise their payoff with respect to the reference vector σ^μ. Thus, instead of considering the actions of the entire population, or play 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

$$u^\mu(x^\mu, \sigma^\mu) = x^\mu \cdot A^\mu \sigma^\mu = x^\mu \cdot A^\mu \sum_{\nu\in N^\mu} w^{\mu\nu} x^\nu. \tag{1}$$

where A^μ is the payoff matrix associated with agent μ. In particular, this means we can write the game Γ with the payoff matrices A^μ in place of the utility functions u^μ. Hence, the NA game allows for the reduction of a multi-player game into a series of two-player games. The agent's goal is to maximise their payoff with respect to the reference vector. As such, we define the best response correspondence BR^μ, which maps every σ^μ to the set $\arg\max_{y\in\Delta_\mu} u^\mu(y, \sigma^\mu)$.

Finally, a central concept of game theory is that of the Nash Equilibrium, in which no rational agent has the incentive to deviate from their current state. 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 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 the NA game is a natural extension of the NE in a bimatrix game. In particular, if we consider an NA game with only two players and no self-loops then the above definition yields that $\bar{x}^1$ is an NE iff $\bar{x}^1 \in BR^1(\sigma^1) = \arg\max_{xin\Delta_1} u^1(x, \bar{x}^2)$, and similarly for $\bar{x}^2$. This is precisely the definition of an NE in a two player game ().

We will show that the NE exists for an NA game in Section 3.

Finally, we note that an NA game is *zero-sum* if the utilities of each agent sum to zero for any strategy set $\{x^\mu\}_{\mu\in\mathcal{N}}$. Formally, $\sum_\mu u^\mu(x^\mu, \sum_{\nu\in N^\mu} w^{\mu\nu}x^\nu) = 0$.

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 respond optimally (i.e. play a best response) to this state. In the case of an NA game, each agent considers their reference vector 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.

$$\alpha^\mu_\sigma = \frac{1}{t}\int_0^t \sigma^\mu(s)\,ds. \tag{2}$$

Using this, we follow in the footsteps of Ewerhart (**?**) and Harris (**?**) to define Fictitious Play in continuous time, but with a slight adaptation for the NA game.

Definition 4 (Network Aggregative - Continuous Time Fictitious Play (NA-CTFP))**.** An NA-CTFP is defined as a measurable map m with components m^μ such that for all μ and all $t \geq 1$, m^μ : $[0,\infty) \to \Delta_\mu$ satisfies $m^\mu(t) \in BR^\mu(\alpha^\mu_\sigma)$ for almost all $t \geq 1$.

We can think of this definition as saying that the player plays some arbitrary strategy before $t = 1$, but beyond this it must play a best response to the time average of its reference signal.

Remark. As an illustration, consider the NA game with two players, in which $\mathcal{E} = \{(1,2),(2,1)\}$ and W is a 2x2 matrix with zeros on its leading diagonal and ones on the off diagonal. We write the time-average of both agents' state as

$$\alpha^\mu(t;x) = \frac{1}{t}\int_0^T x^\mu(t)\,dt \text{ for } \mu\in\{1,2\} \tag{3}$$

In this manner, the $\alpha^\mu(t;m)$ 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 the NA-CTFP is a natural extension of CTFP in the classical two-player setting (**?**).

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 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 to the classical 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 classical assumption made in the analysis of networks and is straightforward to implement (**?**).

Assumption 2. The payoffs are given through matrix games and, therefore, are bilinear. Payoff matrices have a rich history in game theory and has generated a number of prototypical examples for economic study including the Prisoner's Dilemma game (see (**?**) for an interesting implementation of this). They also allow for the design of multi-agent systems in computational settings, particularly in the case of task and resource allocation (**?**). It should be noted, however, that game theoretic analysis is starting to consider various other forms of utility functions, including monotone (**?**) and convex (**?**). 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 a classical assumption which is made in most game theoretic settings and includes, as a special case. However, it should be noted that, in (**?**), CTFP was analysed without this requirement.

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

3 Convergence of Fictitious Play

In this section we present our main results. First, we establish the existence of a Nash Equilibrium in the NA game, as in Def. Referencesdef::NE. Then we show that the Fictitious Play dynamic on the NA game admits solutions (i.e. the NA-CTFP exists). Finally, we show that NA-CTFP reaches a fixed point when the NA game is zero-sum and that, when the network has no self-loops (i.e. is simple), 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, the NE condition (Def. 3) states

$$\begin{aligned} \bar{x}^\mu &\in \arg\max_{x \in \Delta_\mu} u^\mu(x, w^{\mu\mu} x + \sum_{\nu \in N^\mu} w^{\mu\nu} \bar{x}^\nu) \\ &=: \arg\max_{x \in \Delta_i} \bar{u}^\mu(x, \sum_{\nu \in N^\mu} w^{\mu\nu} \bar{x}^\nu) \end{aligned} \tag{4}$$

where we can find $\bar{u}_i$ through the following argument

$$\begin{aligned} u^\mu(x, w^{\mu\mu} x + \sum_{\nu \in N^\nu} w_{\mu\nu} \bar{x}^\nu) &= x \cdot A^\mu (w^{\mu\mu} x + \sum_{\nu \in N^\mu} w_{\mu\nu} \bar{x}^\nu) && (5) \\ &= x \cdot (w^{\mu\mu} A^\mu) x + \sum_{\nu \in N^\mu} u^{\mu\nu}(x, \bar{x}^\nu) && (6) \\ &=: \bar{u}^\mu(x, \sum_{\nu \in N^\mu} w^{\mu\nu} \bar{x}^\nu), \end{aligned}$$

where $u^{\mu\nu}(x^\mu, x^\nu) = x^\mu \cdot A^\mu x^\nu$. Note that, in order to get this formulation, we had to use the assumption of payoffs being bilinear so that we could separate out the term in the weighted sum involving x from $\bar{x}^\nu$.

With these in place, we can state our main results.

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

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

With the existence of the NA-CTFP in place, we can show that it 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 have converged if $\Omega(\alpha)$ is contained within the set of Nash Equilibria of the game. If this is for any such NA-CTFP path, then the game is said to have the *CTFP property*. We adapt the techniques of (Ewerhart 2020) to prove that zero-sum NA games have the CTFP property.

Theorem 3. Any zero-sum NA game has the property that, for any NA-CTFP path m, the corresponding $\alpha(t; m)$ converges to a set of fixed points.

We now point out that, if we choose $w^{\mu\mu}$ to be zero for all μ, then the final inequality yields that, for all μ

$$u^\mu(y, \sum_{\nu \in N^\mu} w^{\mu\nu} x_\infty^\nu) \leq u^\mu(x_\infty^\mu, \sum_{\nu \in N^\mu} w^{\mu\nu} x_\infty^\nu) \tag{7}$$

which is precisely the Nash Equilibrium condition in the NA game. This leads to the next result

Corollary 1. With the additional assumption that, for all agents μ, all zero-sum NA games have the CTFP property

4 NA-CTFP achieves no regret

In this section we aim to find some convergence structure for the case in which the NA game is not necessarily zero-sum. In particular, we show that the NA-CTFP process converges to the set of coarse correlated equilibria (**?**).

Definition 5. A distribution $\mathcal{D}$ over the joint action set $S = \times_\mu S^\mu$ is called a *coarse correlated equilibrium* (CCE) if, for all μ

$$\mathbb{E}_{s \sim \mathcal{D}}[u^\mu(s^\mu, s^{-\mu})] \geq \mathbb{E}_{x \sim \mathcal{D}}[u^\mu(j, s^{-\mu})] \qquad \forall j \in S^\mu. \tag{8}$$

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.

FB: *perhaps what is below could go into a theorem/lemma.*

For an NA game, if a set of actions $s = (s^1, \ldots, s^N)$ FB: *s is normally used for actions* is drawn from a joint probability distribution $\mathcal{D}$, then there is a corresponding set of reference vectors $\sigma = (\sigma^1, \ldots, \sigma^N)$, where $\sigma^\mu = \sum_{\nu \in N^\mu} w^{\mu\nu} s^\nu$. Therefore, $\mathcal{D}$ is a probability distribution over all actions and all references. Consider, then, a vector of joint distributions over actions and references which is generated by NA-CTFP. In particular, if by playing with NA-CTFP, the agents reach the state $(x^\mu)_{i=1}^N$ with references $(\sigma^\mu)_{i=1}^N$, this is the vector $\mathcal{D} = (\mathcal{D}^1, ..., \mathcal{D}^N)$ such 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

$$\begin{aligned} u^\mu(x^\mu, \sigma^\mu) &= x^\mu \cdot A^\mu \sigma^\mu \\ &= \sum_{i,j} (A^\mu)_{ij} x_i^\mu \sigma_j^\mu \end{aligned}$$

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 μ

$$u^{\mu}(x^{\mu}, \sigma^{\mu}) \geq u^{\mu}(i', \sigma^{\mu}) \qquad \forall i' \in S^{\mu}. \tag{9}$$

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 (**?**).

Remark. The notion of the CCE set is related to the idea of *average external regret* (for the sake of brevity we drop the term 'external' and refer the reader to (**?**)). 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 (**?**) for an excellent exposition regarding the link between the CCE set and average regret in two player games which, of course, extends naturally to the NA game.

Average regret, for agent μ is defined as

$$R^{\mu} = \max_{i' \in S^{\mu}} \left\{ \frac{1}{T} \int_0^T u^{\mu}(e^{\mu}_{i'}, \sigma(t)) - u^{\mu}(m^{\mu}(t), \sigma(t))\, dt \right\}, \tag{10}$$

in which e^{μ}_{i} 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 (**? ?**). 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

$$u^{\mu}(x^{\mu}(t), \sigma(t)) = \sum_{ij} a_{ij} x^{\mu}_i \sigma^{\mu}_j = a_{11} x^{\mu}_1 \sigma^{\mu}_1 + a_{12} x^{\mu}_1 \sigma^{\mu}_2 + a_{21} x^{\mu}_2 \sigma^{\mu}_1 + a_{22} x^{\mu}_2 \sigma^{\mu}_2 \tag{11}$$

On the other hand, let us consider that agent μ's first strategy maximises $u^{\mu}(e^{\mu}_1, \sigma(t))$, then

$$u^{\mu}(e^{\mu}_1, \sigma(t)) = \sum_{ij} a_{1j} x^{\mu}_i \sigma^{\mu}_j = a_{11} x^{\mu}_1 \sigma^{\mu}_1 + a_{12} x^{\mu}_1 \sigma^{\mu}_2 + a_{11} x^{\mu}_2 \sigma^{\mu}_1 + a_{12} x^{\mu}_2 \sigma^{\mu}_2 \tag{12}$$

By comparing the two expanded expressions, 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 4. 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.

$$\lim_{T \to \infty} \max_{x^{\mu}_{i'} \in S^{\mu}} \left\{ \frac{1}{T} \int_0^T u^{\mu}(x^{\mu}_{i'}(t), \sigma(t)) - u^{\mu}(m^{\mu}(t), \sigma(t))\, dt \right\} = 0 \tag{13}$$

In particular, NA-CTFP converges to the set of CCE.

5 Numerical Experiments

5.1 Non-convergent examples

The purpose of this section is to show that, whilst wwe have shown 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 nonconvergence we consider the Shapley family of games (**?**). In a two player bimatrix game (A_{β}, B_{β}) this is given by

$$A_{\beta} = \begin{bmatrix} 1 & 0 & \beta \\ \beta & 1 & 0 \\ 0 & \beta & 1 \end{bmatrix}, \; B_{\beta} = \begin{bmatrix} -\beta & 1 & 0 \\ 0 & -\beta & 1 \\ 1 & 0 & -\beta \end{bmatrix} \tag{14}$$

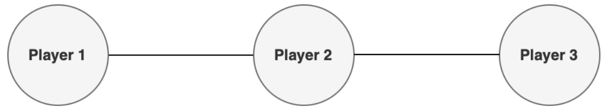


Figure 1: A NA game where the network is defined by a chain of three players, as described in Section 5.1.

where $\beta \in (0,1)$. In (**?**) this was shown to produce both periodic and chaotic behaviour for different choices of β.

As an adaptation, we take the example of a three player chain, as depicted in Figure 1. In this example, we first assume that the network is simple (i.e. there are no self-loops and $w^{\mu\mu} = 0$). The aggregation matrix can be given as

$$W = \begin{bmatrix} 0 & 1 & 0 \\ w & 0 & 1-w \\ 0 & 1 & 0 \end{bmatrix}, \; w \in (0,1). \tag{15}$$

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 given for the three player chain is given as

$$x \cdot Ay + y \cdot B(wx + (1-w)z) + z \cdot Cy = 0. \quad \forall x, y, z \in \Delta_1 \times \Delta_2 \times \Delta_3 \tag{16}$$

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

$$\begin{aligned} A &= -wB^T \\ C &= -(1-w)B^T. \end{aligned} \tag{17}$$

As such in the following example, we will set $B = B_\beta$ with the choice $\beta \approx 0.576$ and set A and C according to the above with the choice $w \approx 0.288$. The orbits that these payoff matrices generate can be seen in Figure 2a, in which, for each player, they converge to the Nash Equilibrium which, for each player, lies in the centre of the simplex.

Let us now make the slight modification in the definition of C so that

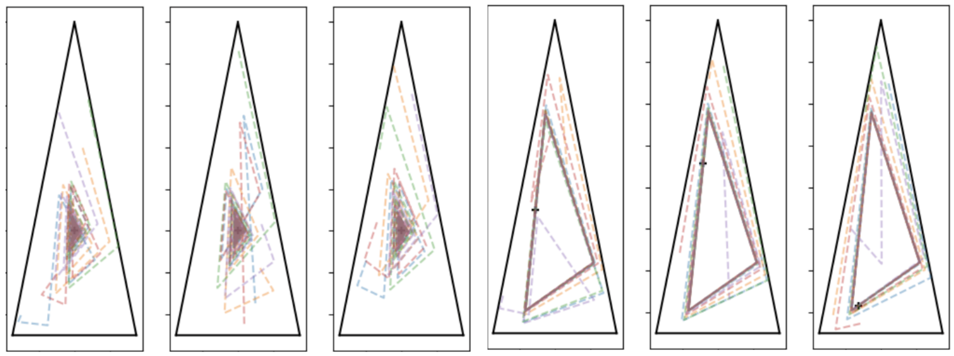
$$C = -(1-w)B, \tag{18}$$

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 Figures 2b. 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 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



(a) (b)

Figure 2: Orbits of the Fictitious Play in the Three Player Chain (c.f. Figure 1) with payoff matrices given by (a) (17) showing convergence to the interior NE (b) (18) showing cycles around the interior NE.

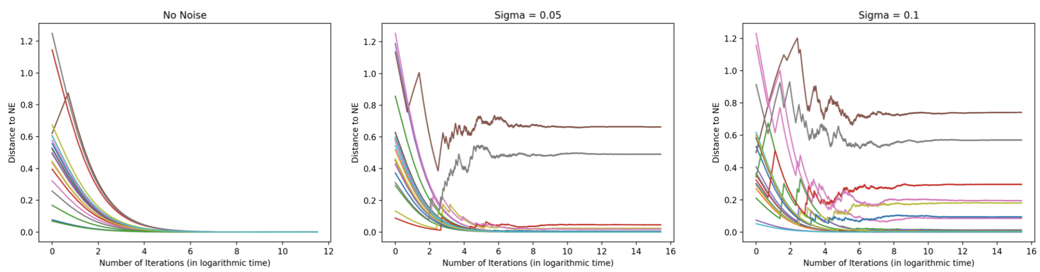


Figure 3: Trajectories of NA-CTFP in a 20 player game with additive noise (Left) No noise is introduced and learning converges directly to an NE. (Middle) $\gamma = 0.05$, the trajectories converge to a fixed point but removed from the NE. (Right) $\gamma = 0.1$, the trajectories converge to a fixed point which is even further away from the NE.

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 3, 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, 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 4 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) (**?**), another adaptive algorithm

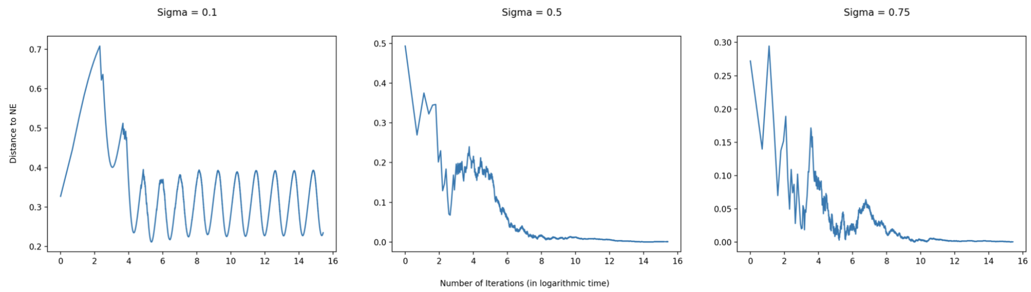


Figure 4: Trajectories of NA-CTFP on the Three Player Chain of Section 5.1 with additive noise. (Left) $\gamma = 0.1$ leads to a decrease in the size of the cyclic orbit (Middle) $\gamma = 0.5$, no periodicity is seen but the trajectory converges to the NE (Right) $\gamma = 0.75$, NA-CTFP still converges, though after a greater amount of time has elapsed.

linked to multi-agent learning (**?**). In (**?**) and (**?**), 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, the algorithm converges to a fixed point. 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, though not necessarily the NE. 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 (3), 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 (**?**), 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 (**?**) that convergence of learning cannot always be assumed. Rather, learning generally presents much more complex dynamics (**?**), which only increases as the number of players increases (4). 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 (**? ? ?**) as well as economic modelling (**?**). As such we Furthermore, the algorithm has links to other learning protocols including the replicator dynamic (**?**) and reinforcement learning (**?**). 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 (**? ?**). 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

[1] CHOTIBUT, T., FALNIOWSKI, F., MISIUREWICZ, M., AND PILIOURAS, G. Family of chaotic maps from game theory. *Dynamical Systems 36*, 1 (1 2021), 48–63.

[2] HUANG, M., MALHAMÉ, R. P., MALHAMÉ, M., AND CAINES, P. E. LARGE POPULATION STOCHASTIC DYNAMIC GAMES: CLOSED-LOOP MCKEAN-VLASOV SYSTEMS AND THE NASH CERTAINTY EQUIVALENCE PRINCIPLE *. 221–252.

[3] PERRIN, S., PEROLAT, J., LAURIÈRE, M., GEIST, M., ELIE, R., AND PIETQUIN, O. Fictitious Play for Mean Field Games: Continuous Time Analysis and Applications.

[4] SANDERS, J. B. T., FARMER, J. D., AND GALLA, T. The prevalence of chaotic dynamics in games with many players. *Scientific Reports 8*, 1 (2018), 4902.

[5] SATO, Y., AKIYAMA, E., AND FARMER, J. D. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences of the United States of America 99*, 7 (4 2002), 4748–4751.

[6] SCHWARTZ, H. M. *Multi-Agent Machine Learning: A Reinforcement Approach*. 2014.