Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 152 (2014) 179–213

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Stochastic fictitious play with continuous action sets

S. Perkins, D.S. Leslie *

School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, United Kingdom

Received 5 October 2012; final version received 31 March 2014; accepted 16 April 2014

Available online 25 April 2014

Abstract

Continuous action space games are ubiquitous in economics. However, whilst learning dynamics in normal form games with finite action sets are now well studied, it is not until recently that their continuous action space counterparts have been examined. We extend stochastic fictitious play to the continuous action space framework. In normal form games with finite action sets the limiting behaviour of a discrete time learning process is often studied using its continuous time counterpart via stochastic approximation. In this paper we study stochastic fictitious play in games with continuous action spaces using the same method. This requires the asymptotic pseudo-trajectory approach to stochastic approximation to be extended to Banach spaces. In particular the limiting behaviour of stochastic fictitious play is studied using the associated smooth best response dynamics on the space of finite signed measures. Using this approach, stochastic fictitious play is shown to converge to an equilibrium point in two-player zero-sum games and a stochastic fictitious play-like process is shown to converge to an equilibrium in negative definite single population games.

JEL classification: C72; C73; D83; C14

Keywords: Stochastic fictitious play; Learning in games; Continuous action set games; Abstract stochastic approximation

* Corresponding author.
E-mail address: david.leslie@bristol.ac.uk (D.S. Leslie).

http://dx.doi.org/10.1016/j.jet.2014.04.008

1. Introduction

Continuous action space games are ubiquitous in economics. However whilst learning dynamics in normal form games with finite action sets are now well studied (e.g. [18]) it is not until recently that their continuous action space counterparts have been examined. Oechssler and Riedel [39] and Lahkar and Riedel [31] provide existence and uniqueness results for two of the most commonly studied evolutionary dynamics: the replicator dynamics and logit best response dynamics, in the single population scenario. Further results along similar lines are given by Oechssler and Riedel [40], Seymour [41], Cressman [12], Cressman et al. [13] and Hofbauer et al. [23].

Although these dynamics have been studied in continuous time for games with continuous action spaces there are few existing convergence results for discrete time learning. Hofbauer and Sorin [22] study a fictitious play-like process in which each player best responds to the average action played by their opponent(s) rather than their opponent(s) empirical distribution(s). They show that this fictitious play-like process converges to a global attractor of the associated best response dynamic in two-player zero-sum continuous action games. Alternatively, Chen and White [11] investigate a stochastic fictitious play-like model which would be difficult to implement (see Section 3.2 for a discussion) which assumes each player uses a probability density function to represent their beliefs. This rules out placing positive mass on a particular observed action, which would correspond to a Dirac measure, and therefore does not have an associated density.

We study a stochastic fictitious play process in continuous action space games, in which beliefs are the empirically observed action distribution. This requires development of new stochastic approximation tools to contend with the resulting measure-valued process.

Dynamical systems results of Lahkar and Riedel [31], combined with our stochastic approximation theory, allow us to analyse a variant of stochastic fictitious play in single-population games. We prove convergence of the stochastic fictitious play-like process in negative definite single population games with a continuous action set and bounded, Lipschitz continuous rewards.

Furthermore we extend the results of Lahkar and Riedel [31] to the N-player case. This allows us to analyse stochastic fictitious play in N-player continuous action space games. We prove the global convergence of the logit best response dynamics for two-player zero-sum games with continuous action spaces and bounded, Lipschitz continuous rewards. Convergence of stochastic fictitious play follows by applying our stochastic approximation results. This extends the previous results of Fudenberg and Kreps [17], Benaïm and Hirsch [2], Hofbauer and Sandholm [21] and Hofbauer and Hopkins [20] for stochastic fictitious play in normal form games with finite action sets.

This paper is organised in the following manner. Section 2 introduces the formal model which we consider. Section 3 contains essential background material from Benaïm [1] and an extension to this asymptotic pseudo-trajectory approach to stochastic approximation on a Banach space. In Section 4 we analyse the stochastic fictitious play process. We develop the logit best response dynamic for N-player, continuous action space games, and show convergence to the set of logit equilibria in two-player zero-sum games with continuous action sets and bounded, Lipschitz continuous rewards. Using the stochastic approximation framework of Section 3 we show that stochastic fictitious play converges to an equilibrium for two-player zero-sum games with continuous action sets and bounded, Lipschitz continuous rewards. In Section 5 the work of Lahkar and Riedel [31] covering the logit best response dynamic for single population, continuous action games is reviewed and a learning variant of this, similar to stochastic fictitious play, is studied. As in the N-player case, this stochastic fictitious play variant is shown to converge in single population games which are negative definite. In Section 6 some examples of this convergence are presented to demonstrate that stochastic fictitious play can be implemented in the continuous action case. Throughout this work many of the proofs are relegated to Appendices A, B, C and D.

2. Stochastic fictitious play with continuous action spaces

Consider an N-player continuous action space game. We use $i = 1, \dots, N$ to denote the players. In a standard abuse of notation we use $-i$ to denote all the players other than i. Let $A^i \subset \mathbb{R}$ be a compact, convex action set and let $\mathcal{B}^i$ denote the Borel σ-algebra on A^i. Let $\underline{A} := A^1 \times \dots \times A^N$.

In discrete-action games a mixed strategy is described by a probability mass function on the action set. In continuous action space games this approach must be extended. Let $\mathcal{B}^i$ denote the Borel σ-algebra on A^i and let $\mathcal{P}(A^i, \mathcal{B}^i)$ denote the set of all probability measures on $(A^i, \mathcal{B}^i)$. In a continuous action space game a mixed strategy is a probability measure in $\mathcal{P}(A^i, \mathcal{B}^i)$. For a mixed strategy $\pi^i \in \mathcal{P}(A^i, \mathcal{B}^i)$ of Player i, $\pi^i(B^i)$ denotes the probability of Player i selecting an action in the set $B^i \in \mathcal{B}^i$.

As with finite action set games, if the population interpretation is being used then $P^i \in \mathcal{P}(A^i, \mathcal{B}^i)$ is a particular population and selecting an action $x^i \in A^i$ using P^i is interpreted as selecting a member in the population who plays pure strategy x^i. Alternatively, the strategy interpretation can be used meaning that $\pi^i \in \mathcal{P}(A^i, \mathcal{B}^i)$ is the strategy of a particular player and an action $x^i \in A^i$ is randomly selected using π^i. The focus of this paper is on classical stochastic fictitious play using the strategy interpretation in N-player games; however, in Section 5 a variant of stochastic fictitious play is presented which is designed for learning under the single population framework, where A is the continuous action set and $\mathcal{P}(A, \mathcal{B})$ represents the set of all possible populations.

Each player, i, has a reward function $r^i(\cdot) : \underline{A} \to \mathbb{R}$. We assume throughout that $r^i(\underline{x})$ is bounded and Lipschitz continuous. In a standard abuse of notation, if we have strategies $\pi^i \in \mathcal{P}(A^i, \mathcal{B}^i)$ then the expected reward to Player i is

$$r^i(\pi^1, \dots, \pi^N) = r^i(\underline{\pi}) = \int_{A^1} \dots \int_{A^N} r^i(\underline{x}) \pi^1(\mathrm{d}x^1) \dots \pi^N(\mathrm{d}x^N). \tag{2.1}$$

Then, for $x^i \in A^i$, let $r^i(x^i, \pi^{-i}) = r^i(\delta_{x^i}, \pi^{-i})$.

Definition 2.1. *A continuous action space game is a* two-player zero-sum game *if $N = 2$ and for every $\underline{x} \in \underline{A}$*

$$r^1(\underline{x}) = -r^2(\underline{x}).$$

Our main object of study is stochastic fictitious play, in which players repeatedly play the game and respond to the history of play. The belief about Player i at iteration n, denoted by σ_n^i, is the empirical distribution of actions played previously; this belief is independent of the beliefs about the other players. On each play of the repeated game a mixed strategy is then chosen which maximises a perturbation of the expected reward when playing against this empirical distribution. Fudenberg and Levine [18] formulate this microfoundation of the logit best response in discrete action games, and Lahkar and Riedel [31] extend the definition to continuous action games.

Definition 2.2. *The logit best response of Player i to beliefs σ^{-i}, with parameter $\eta > 0$, is given by*

$$L^i_\eta(\sigma^{-i}) := \operatorname*{arg\,max}_{\pi^i \in \mathcal{P}(A^i, \mathcal{B}^i)} \{ r^i(\pi^i, \sigma^{-i}) + \eta \nu^i(\pi^i) \}, \tag{2.2}$$

where $\nu^i(\pi^i)$ is the entropy of π^i. If π^i does not admit a density then $\nu^i(\pi^i) = -\infty$, whereas if π^i has density p^i then $\nu^i(\pi^i) := -\int_{A^i} p^i(x)\log(p^i(x))\,\mathrm{d}x$.

Following the logic of Lahkar and Riedel [31], the logit best response is given by

$$L^i_\eta(\sigma^{-i})(B^i) := \frac{\int_{B^i} \exp\{\eta^{-1} r^i(x^i, \sigma^{-i})\}\,\mathrm{d}x^i}{\int_{A^i} \exp\{\eta^{-1} r^i(y^i, \sigma^{-i})\}\,\mathrm{d}y^i}. \tag{2.3}$$

For convenience, for $\underline{\sigma} \in \Delta$ we take $L_\eta(\underline{\sigma}) := (L^1_\eta(\sigma^{-1}), \dots, L^N_\eta(\sigma^{-N}))$. Again following Lahkar and Riedel [31], the logit best response for Player i as defined in (2.3) has an associated density function

$$l^i_\eta(\sigma^{-i})(x^i) := \frac{\exp\{\eta^{-1} r^i(x^i, \sigma^{-i})\}}{\int_{A^i} \exp\{\eta^{-1} r^i(y^i, \sigma^{-i})\}\,\mathrm{d}y^i}. \tag{2.4}$$

Stochastic fictitious play in an N-player game can therefore be formulated as follows. For each player i, the beliefs follow the recursion

$$\sigma^i_{n+1} = \sigma^i_n + \frac{1}{n+c}\left[\delta_{x^i_{n+1}} - \sigma^i_n\right], \tag{2.5}$$

where x^i_{n+1} is the action selected by Player i on iteration $n+1$, δ_x is a Dirac measure at $x \in \mathbb{R}$ and $c > 0$ is a constant. The action x^i_{n+1} is sampled according to the logit best response $L^i_\eta(\sigma^{-i}_n)$. Let $\underline{\sigma}_n := (\sigma^1_n, \dots, \sigma^N_n) \in \Delta$, and take $\underline{\sigma}_0$ to be any joint probability measure in Δ. Stochastic fictitious play is therefore defined as

$$\underline{\sigma}_{n+1} = \underline{\sigma}_n + \frac{1}{n+c}\left[(\delta_{x^1_{n+1}}, \dots, \delta_{x^N_{n+1}}) - \underline{\sigma}_n\right], \tag{2.6}$$

where each x^i_{n+1} is sampled from the distribution with density $l^i_\eta(\sigma^{-i}_n)$.

Note that (2.6) can be written as

$$\underline{\sigma}_{n+1} = \underline{\sigma}_n + \alpha_{n+1}\left[\left(L_\eta(\underline{\sigma}_n) - \underline{\sigma}_n\right) + U_{n+1}\right], \tag{2.7}$$

where $\alpha_{n=1} = (n+1)^{-1}$, $U_{n+1} = (U^1_{n+1}, \dots, U^N_{n+1})$ and

$$U^i_{n+1} = \delta_{x^i_{n+1}} - L^i_\eta(\sigma^{-i}_n). \tag{2.8}$$

This formulation lends itself to the study of stochastic fictitious play using the tools of stochastic approximation.

3. Stochastic approximation on a Banach space

In normal form games with finite action sets, stochastic approximation is used to study the limiting behaviour of stochastic fictitious play. The discrete time learning process is studied via the continuous time smooth best response dynamics on the probability simplex, which is embedded in standard Euclidean space. We follow the same approach for continuous action space games: the evolving beliefs are probability measures in sets $\mathcal{P}(A^i, \mathcal{B}^i)$, and it is judicious to consider $\mathcal{P}(A^i, \mathcal{B}^i)$ as a subset of the vector space of signed measures $\mathcal{M}(A^i, \mathcal{B}^i)$, to make use of general theory applicable to vector spaces.[1] When associated with an appropriate distance metric, $\mathcal{M}(A^i, \mathcal{B}^i)$ is a Banach space. Therefore, in this section, we extend the standard stochastic approximation framework to the Banach space setting. This will allow us to study the asymptotic behaviour of stochastic fictitious play in continuous action space games.

In general, a stochastic approximation process in a space Θ is an iterative process $\{\theta_n\}_{n\in\mathbb{N}}$ such that

$$\theta_{n+1} = \theta_n + \alpha_{n+1}\big[F(\theta_n) + U_{n+1}\big], \tag{3.1}$$

where $F(\cdot) : \Theta \to \Theta$ is a continuous map, $\{U_n\}_{n\in\mathbb{N}}$ is a noise sequence and $\{\alpha_n\}_{n\in\mathbb{N}}$ is a sequence of learning rates in $\mathbb{R}$ such that $\alpha_n \to 0$ as $n \to \infty$. If $\Theta = \mathbb{R}^K$, standard stochastic approximation results (e.g. [1]) show that the limiting behaviour of (3.1) can be studied using the ordinary differential equation (ODE)

$$\frac{d\theta}{dt} = F(\theta). \tag{3.2}$$

This is commonly known as the ODE method of stochastic approximation, originally proposed by Ljung [36] and extended by many authors including Kushner and Clark [28], Kushner and Yin [29,30], Borkar [7–9], Benaïm [1] and Benaïm et al. [3].

Luenberger [37] provides the functional analysis mechanisms to investigate (3.2) in a Banach space setting; if Θ is a Banach space and $F(\cdot) : \Theta \to \Theta$ is a uniformly continuous map then (3.2) is well defined. Similarly it is possible to consider random variables on Banach spaces (for example, see [33]).

This theory has allowed stochastic approximation to be extended beyond the Euclidean space setting, with Walk [44], Berger [4], Walk and Zsidó [45], Shwartz and Berman [43], Koval [27] and Dippon and Walk [14] producing convergence results for discrete time stochastic processes on general Hilbert or Banach spaces. We provide an update of this earlier work (often called abstract stochastic approximation) to the now common asymptotic pseudo-trajectory approach of Benaïm [1]. We show that, under assumptions that correspond to those used in the Euclidean setting, a linear interpolation of the $\{\theta_n\}_{n\in\mathbb{N}}$ process is an asymptotic pseudo-trajectory of the ordinary differential equation on the Banach space Θ given by (3.2).

3.1. Asymptotic pseudo-trajectory approach

The ideas in this section build on a rich history of work on stochastic approximation. In particular we make use of the asymptotic pseudo-trajectory framework of Benaïm [1]. To begin we review the key results on asymptotic pseudo-trajectories (see [1] for a more extensive review), and verify that these also hold when Θ is a Banach space. Note that a Banach space is a metric space, and we define d to be the distance induced by the Banach space norm $\|\cdot\|_\Theta$.

Definition 3.1. *A* semiflow Φ *on* Θ *is a continuous map* $\Phi : \mathbb{R}^+ \times \Theta \to \Theta$, $(t, \theta) \to \Phi_t(\theta)$, *such that* $\Phi_0(\theta) = \theta$ *and* $\Phi_{t+s}(\theta) = \Phi_t(\Phi_s(\theta))$, *for any* $t, s \geq 0$.

[1] The space $\mathcal{M}(A^i, \mathcal{B}^i)$ consists of all $\mu : \mathcal{B}^i \to \mathbb{R}$ such that there exists two finite measures on $(A^i, \mathcal{B}^i)$, ν_1 and ν_2, such that for all $B \in \mathcal{B}^i$, $\mu(B) = \nu_1(B) - \nu_2(B)$.

If the differential equation (3.2) has unique solution trajectories then it defines a flow. In particular, let $\Phi_t(\theta)$ be the position at time t of the trajectory of (3.2) that passes through θ at time 0.

Definition 3.2. *A continuous function* $\psi(\cdot) : \mathbb{R}^+ \to \Theta$ *is an* asymptotic pseudo-trajectory *for a semiflow* Φ *if for any* $T > 0$,

$$\lim_{t\to\infty} \sup_{0\le s\le T} d\big(\psi(t+s), \Phi_s\big(\psi(t)\big)\big) = 0.$$

Following Benaïm [1], we define an interpolating process of the stochastic approximation process (3.1). Let $\tau_0 := 0$, $\tau_n := \sum_{k=1}^{n} \alpha_k$ and $m(t) := \sup\{k \ge 0 : t \ge \tau_k\}$. Define the continuous time interpolation of $\{\theta_n\}_{n\in\mathbb{N}}$ such that for $s \in [0, \alpha_{n+1})$,

$$\bar{\theta}(\tau_n + s) := \theta_n + \frac{s}{\alpha_{n+1}}[\theta_{n+1} - \theta_n]. \tag{3.3}$$

We need the following assumptions:

(A1) There exists a compact $\Lambda \subset \Theta$ such that $\{\theta_n\}_{n\in\mathbb{N}} \subset \Lambda$.
(A2) $F(\cdot) : \Lambda \to \Lambda$ is a uniformly continuous map and there exists a $C < \infty$ such that $\|F(\theta)\|_\Theta < C$ for all $\theta \in \Lambda$.
(A3) For all $T > 0$

$$\lim_{n\to\infty} \sup_{k} \left\{ \left\| \sum_{i=n}^{k-1} \alpha_{i+1} U_{i+1} \right\|_\Theta : k = n+1, \dots, m(\tau_n + T) \right\} = 0.$$

(A4) Given any initial choice of $\theta_0 \in \Lambda$ there exists a unique solution flow in Λ to the differential equation (3.2).

Theorem 3.3. *Assume assumptions* (A1)–(A4) *hold and that* $\sum_{n=1}^{\infty} \alpha_n = \infty$. *Then* $\bar{\theta}(\cdot) : \mathbb{R}^+ \to \Theta$*, defined in* (3.3)*, is an asymptotic pseudo-trajectory to the semiflow* Φ *induced by the differential equation* (3.2).

Proof. The proof is omitted as, other than the Euclidean norm $\|\cdot\|$ being replaced with the norm $\|\cdot\|_\Theta$, the proof is identical to Benaïm [1, Proposition 4.1]. □

We note that assumptions (A1)–(A4) are extensions to those used by Benaïm [1] for standard stochastic approximation and are similar to those given by Shwartz and Berman [43]. However, verifying these for a general Banach space can be difficult. In particular, the noise assumption (A3) has caused great difficulty. For example Koval [27] considers the simple case when $\{U_n\}_{n\in\mathbb{N}}$ is an i.i.d. noise process whilst Shwartz and Berman [43] prove a very weak convergence result for a particular process which again uses independent noise. In Section 3.2 we provide conditions, applicable to stochastic fictitious play, which allow us to verify the challenging assumption (A3) for martingale noise in a Banach space.

Many further results can be taken from Benaïm [1] to characterise the behaviour of asymptotic pseudo-trajectories. Here we present the result which is used in Theorems 4.4 and 5.4 to prove the convergence of stochastic fictitious play.

Definition 3.4. *A compact, non-empty set $A \subset M$ is called an* attractor *for the semiflow Φ if*

(i) $\Phi_t(A) = A$ *for all* $t \in \mathbb{R}^+$ *(i.e. A is invariant)*
(ii) *A has a neighborhood* $W \subset M$ *such that* $d(\Phi_t(x), A) \to 0$ *as* $t \to \infty$*, uniformly for* $x \in W$.

Theorem 3.5. *Let A be an attractor for the semiflow Φ with attracting neighborhood W, assume that $\{\theta_n\}_{n\in\mathbb{N}}$, defined in (3.1), satisfies* (A1)–(A4) *and that $\sum_{n=1}^{\infty} \alpha_n = \infty$. If $\Lambda \subset W$, where $\{\theta_n\}_{n\in\mathbb{N}} \subset \Lambda$, then $\theta_n \to A$ as $n \to \infty$.*

Theorem 3.5 is produced by combining Proposition 4.1 and Theorem 6.10 of Benaïm [1] and gives us a method of showing that a stochastic approximation process on Θ converges to a limiting set.

3.2. Noise criteria: the space of finite signed measures

We now focus on the space $\mathcal{M}(X, \mathcal{B})$ of finite signed measures on a compact set $X \subset \mathbb{R}$. By using an appropriate norm we can consider this as a Banach space. A greater discussion is given to the choice of norm in Section 4.1. Here we introduce the bounded Lipschitz norm on $\mathcal{M}(X, \mathcal{B})$. For a bounded, Lipschitz continuous function $g(\cdot) : X \to \mathbb{R}$ define

$$\|g\|_{BL} := \sup_{x\in X}\left|g(x)\right| + \sup_{x\neq y} \frac{|g(x) - g(y)|}{|x - y|}. \tag{3.4}$$

Now let

$$\mathcal{BL} := \left\{g : g \text{ bounded \& Lipschitz continuous with } \|g\|_{BL} \le 1\right\} \tag{3.5}$$

be the set of bounded Lipschitz continuous functions with BL-norm bounded by 1. The dual BL^*-norm on $\mathcal{M}(X, \mathcal{B})$ is defined for $\mu \in \mathcal{M}(X, \mathcal{B})$ as

$$\|\mu\|_{BL^*} := \sup_{g\in\mathcal{BL}}\left|\int_X g(x)\mu(\mathrm{d}x)\right|. \tag{3.6}$$

Here we consider the space $(\mathcal{M}(X, \mathcal{B}), \|\cdot\|_{BL^*})$. As discussed by Dudley [15] and Hofbauer et al. [23], $\mathcal{M}(X, \mathcal{B})$ is a Banach space when paired with $\|\cdot\|_{BL^*}$.

We would like to show that assumption (A3) is satisfied for all martingale noise sequences $\{U_n\}$. The additional complications of working in a Banach space do not allow us to do so. However note that for stochastic fictitious play the U_n take a particular form (see (2.8)). In particular, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ a filtration of $\mathcal{F}$, where $\mathcal{F}_n$ represents the information available at the point the $(n+1)$th action is selected. We are interested in random measures of the form $U_{n+1} = \delta_{x_{n+1}} - P_n$ where the δ_x are Dirac masses at the actually selected actions, and the P_n are the (absolutely continuous) smooth best responses used to select the actions. So P_n is measurable with respect to $\mathcal{F}_n$, whereas x_{n+1} (and hence U_{n+1}) is measurable with respect to $\mathcal{F}_{n+1}$. We prove (A3) holds for this particular form of random measure, although see Appendix A for more general results.

Proposition 3.6. *Consider the Banach space $(\mathcal{M}(X, \mathcal{B}), \|\cdot\|_{BL^*})$, where X is a compact, convex subset of $\mathbb{R}$. Assume that $\{\alpha_n\}_{n\in\mathbb{N}}$ is a deterministic sequence such that $\alpha_n \to 0$, $\sum_{n=1}^{\infty} \alpha_n = 0$ and $\sum_{n=1}^{\infty} \alpha_n^2 < \infty$. Also assume that*

$$U_{n+1} = \delta_{x_{n+1}} - P_n,$$

where for all n

- *U_n is measurable with respect to $\mathcal{F}_n$,*
- *$P_n \in \mathcal{P}(X, \mathcal{B})$ is a bounded, absolutely continuous measure which is measurable with respect to $\mathcal{F}_n$ and has density p_n,*
- *$x_{n+1} \in X$ is drawn from the distribution with probability density function p_n.*

Then

$$\lim_{n\to\infty} \sup_k \left\{ \left\| \sum_{j=n}^{k-1} \alpha_{j+1} U_{j+1} \right\|_{BL^*} : k = n+1, \ldots, m(\tau_n + T) \right\} = 0, \quad \text{w.p. 1.}$$

The proof of this result comes from approximating the Dirac delta in U_n with a spike centred on the Dirac measure. This spike and the density function p_n can then be studied in L^2. By applying the convergence result in Proposition A.1 for L^2 we conclude. This proof is given in Appendix B. We note that the earlier result of Chen and White [11] assumes a compact, convex action set and that L^2-valued beliefs P_n are updated towards a symmetric, absolutely continuous distribution centred on the observed action $x_{n+1} \in X$ to ensure that $P_n \in L^2$ for all $n \in \mathbb{N}$. However, no such distribution exists if x_{n+1} is on the boundary of X, so their process is actually impossible. Instead we consider distributional beliefs, and approximate these with L^2 spikes only in the proof of Proposition 3.6.

4. Convergence of stochastic fictitious play

We are now in a position to prove convergence of stochastic fictitious play in two-player zero-sum continuous action games. Recall from (2.7) that stochastic approximation can be written as

$$\underline{\sigma}_{n+1} = \underline{\sigma}_n + \alpha_{n+1}\left[\left(L_\eta(\underline{\sigma}_n) - \underline{\sigma}_n\right) + U_{n+1}\right].$$

It is clear that this process fits the general stochastic approximation form (3.1). The theory of Section 3 relates (2.7) to the logit best response dynamics defined as

$$\dot{\underline{\pi}} = L_\eta(\underline{\pi}) - \underline{\pi}. \tag{4.1}$$

We therefore investigate properties of this dynamical system. First however we need to revisit the choice of norm for the Banach spaces under consideration.

4.1. Topological considerations

In discrete action games a player's strategy is a probability mass function which can be studied on Euclidean space. Since all norms are equivalent on $\mathbb{R}^K$ the choice of distance metric does not affect the type of convergence that occurs. In the continuous action space framework a player's strategy is a probability measure, as already discussed. In the infinite dimensional space $\mathcal{M}(A^i, \mathcal{B}^i)$ the choice of norm directly affects the type of convergence. Therefore it is important to make a careful choice of norm. We consider two norms on $\mathcal{M}(A^i, \mathcal{B}^i)$, under either of which $\mathcal{M}(A^i, \mathcal{B}^i)$ is a Banach space [15,42,23].

Firstly, as in [39] and [41], the total variation norm is used. If $F := \{f(\cdot) : A \to \mathbb{R} : \sup_{x\in A} |f(x)| \leq 1\}$ then for $\mu^i \in \mathcal{M}(A^i, \mathcal{B}^i)$ let

$$\|\mu^i\|_{TV_i} := \sup_{f\in F}\left|\int_{A^i} f\left(\mathrm{d}\mu^i\right)\right|.$$

The total variation norm induces the strong topology on $\mathcal{M}(A^i, \mathcal{B}^i)$ [42]. Under this norm continuity is natural in many cases, and so it is convenient when considering continuity and convergence of dynamical systems. On the other hand compactness is less natural under the strong topology, and in particular $\mathcal{P}(A^i, \mathcal{B}^i)$ is not a compact subset of $\mathcal{M}(A^i, \mathcal{B}^i)$ under the strong topology, causing difficulty for stochastic approximation theory. Therefore the total variation norm is used in Section 4.2 when studying the properties of the logit best response dynamics (4.1).

An alternative norm on $\mathcal{M}(A^i, \mathcal{B}^i)$, which has already been discussed in Section 3.2, is the bounded Lipschitz norm. This norm is discussed by Oechssler and Riedel [40] and used by Hofbauer et al. [23] and Lahkar and Riedel [31]. For $\mu^i \in \mathcal{M}(A^i, \mathcal{B}^i)$ let $\|\mu^i\|_{BL_i^*}$ be defined as in (3.6) with the integrals now over A^i. The bounded Lipschitz norm induces the BL^* topology on $\mathcal{M}(A^i, \mathcal{B}^i)$ [15]. This coincides with the weak-* topology (often refered to as the weak topology—see [5]) on $\mathcal{M}(A^i, \mathcal{B}^i)$ whenever this is metrizable [15]. Although the weak topology cannot in general be metrized on the whole of $\mathcal{M}(A^i, \mathcal{B}^i)$, the weak topology and BL^* topology coincide on $\mathcal{P}(A^i, \mathcal{B}^i)$. Convergence under the weak topology on the set of probability measures is commonly referred to as the weak convergence of probability measures (e.g. [5]). More restrictive conditions are required to obtain continuity under the weak topology when compared to the strong topology, which highlights an important difference between these topologies. However, we show in Section 4.3 that $\mathcal{P}(A^i, \mathcal{B}^i)$ is a compact subset of $\mathcal{M}(A^i, \mathcal{B}^i)$ under the weak topology, and the noise criterion (A3) for stochastic approximation can be verified using Proposition 3.6. We therefore use the bounded Lipschitz norm in Section 4.3 when considering the convergence of stochastic fictitious play using stochastic approximation.

There is a natural relationship between the strong and weak topologies, as discussed by Oechssler and Riedel [40]. Huber [24] shows that for $\pi^i, \rho^i \in \mathcal{P}(A^i, \mathcal{B}^i)$

$$\|\pi^i - \rho^i\|_{BL_i^*} \leq 2\|\pi^i - \rho^i\|_{TV_i}, \tag{4.2}$$

so that convergence of probability measures under the total variation norm implies convergence under the bounded Lipschitz norm.

Since the players do not evolve in isolation, we need to consider the belief processes in parallel. Using the Cartesian product let $\Delta := \mathcal{P}(A^1, \mathcal{B}^1) \times \ldots \times \mathcal{P}(A^N, \mathcal{B}^N)$ and $\Sigma := \mathcal{M}(A^1, \mathcal{B}^1) \times \ldots \times \mathcal{M}(A^N, \mathcal{B}^N)$. Thus stochastic fictitious play is a process in Δ, which is a subspace of the vector space Σ. When we take the Cartesian product $\Sigma := \mathcal{M}(A^1, \mathcal{B}^1) \times \ldots \times \mathcal{M}(A^N, \mathcal{B}^N)$ then it is natural to use the product topology. For $\underline{\mu} := (\mu^1, \ldots, \mu^N) \in \Sigma$ we have total variation norm

$$\|\underline{\mu}\|_{\Sigma TV} = \max\{\|\mu^1\|_{TV_1}, \ldots, \|\mu^N\|_{TV_N}\}$$

and bounded Lipschitz norm

$$\|\underline{\mu}\|_{\Sigma BL^*} = \max\{\|\mu^1\|_{BL_1^*}, \ldots, \|\mu^N\|_{BL_N^*}\}.$$

Under either of these norms Σ is a Banach space as the finite Cartesian product of Banach spaces with the product topology is also a Banach space.

4.2. N-Player, logit best response dynamic

In this section the N-player logit best response dynamics (4.1) are studied using the total variation norm to induce the strong topology on Σ. We aim to obtain convergence to a fixed point and therefore it is important to show that such a fixed point of the logit best response exists.

Definition 4.1. *The set of logit equilibria is given by*

$$\mathcal{LE}_\eta = \left\{ \underline{\pi} : L^i_\eta\left(\pi^{-i}\right) = \pi^i, \text{ for all } i = 1, 2, \ldots, N \right\}. \tag{4.3}$$

Rest points of the logit best response dynamics (4.1) correspond to elements of $\mathcal{LE}_\eta$.

Proposition 4.2. *If the rewards are bounded and Lipschitz continuous then $\mathcal{LE}_\eta$ is non-empty.*

The proof of Proposition 4.2 is presented in Appendix C.

Proposition 4.3. *If the rewards are bounded and Lipschitz continuous, then for any initial strategy $\underline{\pi}(0) \in \Delta$ there exists a unique solution in Δ to the logit best response differential equation (4.1). In addition, the solution is continuous with respect to the initial conditions under $\|\cdot\|_{\Sigma TV}$, meaning that for some $k > 0$ if $\underline{\pi}(0), \underline{\rho}(0) \in \Delta$ then*

$$\left\| \underline{\pi}(t) - \underline{\rho}(t) \right\|_{\Sigma TV} \le e^{kt} \left\| \underline{\pi}(0) - \underline{\rho}(0) \right\|_{\Sigma TV}.$$

The proof of Proposition 4.3 is an extension to N players of results of Lahkar and Riedel [31], and is also presented in Appendix C. It remains to present a global convergence result for the logit best response dynamics (4.1) in two-player zero-sum games with continuous action spaces. This is the natural extension to continuous action spaces of the discrete action case given by Hofbauer and Sandholm [21] and Hofbauer and Hopkins [20].

Theorem 4.4. *If a continuous action set game is a two-player zero-sum game with bounded, Lipschitz continuous rewards then,*

$$\inf_{\underline{\tilde{\pi}} \in \mathcal{LE}_\eta} \left\| \underline{\pi}(t) - \underline{\tilde{\pi}} \right\|_{\Sigma TV} \to 0, \quad \text{as } t \to \infty,$$

uniformly for $\underline{\pi}(0) \in \Delta$.

The proof of Theorem 4.4 is presented in Appendix D. To understand the shape of the proof we need to introduce the set $\Delta_D \subset \Delta$ defined as

$$\begin{aligned} \Delta_D := \Big\{ \underline{\pi} \in \Delta :\ & \forall i = 1, \ldots, N, \pi^i \text{ is absolutely continuous with density } p^i \text{ such that} \\ & D^{-1} \le p^i\left(x^i\right) \le D, \text{ for all } x^i \in A^i \text{ and } p^i(\cdot) \text{ is Lipschitz continuous} \\ & \text{with constant } D \Big\}. \end{aligned}$$

We show in Appendix D that Δ_D is a compact subset of Δ with respect to the total variation norm $\|\cdot\|_{\Sigma TV}$. The Lyapunov function used by Hofbauer and Sandholm [21] and Hofbauer and Hopkins [20] for two-player zero-sum finite action set games can be extended to Δ_D, and the compactness results in global attractivity of $\mathcal{LE}_\eta$ within Δ_D. To extend this to being a global convergence result on Δ it is shown that starting from an arbitrary point in Δ the trajectory gets "close" to the set Δ_D after a time τ_1. Therefore at some finite time in the future, $\tau_1 + \tau_2$, the original trajectory is "close" to a trajectory in Δ_D using the continuity of the solution flow from Proposition 4.3. And the new trajectory in Δ_D will converge to $\mathcal{LE}_\eta$ and hence be close to $\mathcal{LE}_\eta$ at time $\tau_1 + \tau_2$.

Theorem 4.4 provides the global convergence of the logit best response in two-player zero-sum continuous action games under the total variation norm. In Section 4.3 stochastic fictitious play is studied under the bounded Lipschitz norm, therefore it is important that the global convergence result of Theorem 4.4 is presented for the weak topology on Δ.

Corollary 4.5. *If the rewards are bounded and Lipschitz continuous, then, with respect to the weak topology on Δ, for any initial strategy $\underline{\pi}(0) \in \Delta$ there exists a unique solution to the logit best response differential equation (4.1) and $\underline{\pi}(t) \in \Delta$ for all $t \in \mathbb{R}^+$. In addition, if the game is two-player zero-sum then $\mathcal{LE}_\eta$ is an attractor for the logit best response dynamics for all initial strategies in Δ.*

Proof. It follows immediately from Propositions 4.2 and 4.3 that when the rewards are bounded a unique solution exists and that $\mathcal{LE}_\eta$ is non-empty. When the game is two-player zero-sum, Theorem 4.4 gives that, for a solution to the logit best response differential equation (4.1), $\underline{\pi}(\cdot) : \mathbb{R}^+ \to \Delta$,

$$\lim_{t\to\infty} \inf_{\underline{\tilde{\pi}}\in\mathcal{LE}_\eta} \left\|\underline{\pi}(t) - \underline{\tilde{\pi}}\right\|_{\Sigma BL^*} \le 2 \lim_{t\to\infty} \inf_{\underline{\tilde{\pi}}\in\mathcal{LE}_\eta} \left\|\underline{\pi}(t) - \underline{\tilde{\pi}}\right\|_{\Sigma TV} = 0.$$

Since $\|\underline{\pi}(t) - \underline{\tilde{\pi}}\|_{\Sigma TV}$, converges uniformly for $\underline{\pi}(0) \in \Delta$ it follows that $\|\underline{\pi}(t) - \underline{\tilde{\pi}}\|_{\Sigma BL^*}$ converges uniformly as well. Now it is well known that the fixed points of a continuous map from a metric space (more generally, a Hausdorff space) to itself lie in a closed set [25]. Therefore, since $L_\eta(\cdot) : \Delta \to \Delta$ is Lipschitz continuous with respect to the BL^*-norm when the rewards are bounded and Lipschitz continuous (see Lemma C.2), $\mathcal{LE}_\eta$ is a closed set. Then under the weak topology on Δ it is a compact set since a closed subset of a compact set is compact, and from the definition $\mathcal{LE}_\eta$ is strongly invariant. Therefore all the conditions of Definition 3.4 are satisfied, so $\mathcal{LE}_\eta$ is an attractor for the logit best response on Δ under the weak topology on Δ. □

4.3. Stochastic fictitious play

With new results on stochastic approximation (Section 3) and convergence results for the logit best response dynamics (Section 4.2) we are finally in a position to prove the convergence of stochastic fictitious play.

Proposition 4.6. *Assume that, for $i = 1, \ldots, N$, A^i is a compact, convex subset of $\mathbb{R}$ and the rewards are bounded and Lipschitz continuous. Then a linear interpolation of the stochastic fictitious play process (2.6) is almost surely an asymptotic pseudo-trajectory to the N-player logit best response dynamic (4.1) under $\|\cdot\|_{\Sigma BL^*}$.*

Proof. We verify assumptions (A1)–(A4) from page 184 for the stochastic fictitious play process (2.7).

Let $\mathcal{P}(X, \mathcal{B})$ be the subset of $\mathcal{M}(X, \mathcal{B})$ consisting of the probability measures. For any probability measure $\mu \in \mathcal{P}(X, \mathcal{B})$

$$\|\mu\|_{BL^*} = 1.$$

The space of finite signed measures is the topological dual of the space of bounded Lipschitz continuous functions, as illustrated in Section 3.2. Using the Banach–Alaoglu theorem (see e.g. [32]) the subset $\mathcal{P}(X, \mathcal{B})$ is compact under $\| \cdot \|_{BL^*}$. This means that, under the bounded Lipschitz norm, assumption (A1) is satisfied for a process which remains in $\mathcal{P}(X, \mathcal{B})$.

When the rewards are bounded and Lipschitz continuous it is straightforward to show that $L_\eta(\underline{\sigma}_n)$ is Lipschitz continuous, and hence uniformly continuous, with respect to the bounded Lipschitz norm (see Lemma C.2). Therefore, (A2) holds since the mean field in (2.7), $L_\eta(\underline{\sigma}_n) - \underline{\sigma}_n$, is uniformly continuous.

When the rewards are bounded and Lipschitz continuous, the existence of a unique solution to the logit best response differential equation is given by Corollary 4.5, which shows that (A4) holds.

The noise assumption, (A3), is shown to hold by using Proposition 3.6 on $\{U_n^i\}_{n\in\mathbb{N}}$, for $i = 1, \ldots, N$. Let $\mathcal{F}_n$ represent the history of the iterative process (2.6) up to iteration $n \in \mathbb{N}$. The learning rate assumption of Proposition 3.6 is true since $\alpha_n = (n + c)^{-1}$. Clearly for all $n \in \mathbb{N}$, U_n^i is measurable with respect to $\mathcal{F}_n$, and since $L_\eta^i(\underline{\sigma}_n)$ is a bounded and absolutely continuous measure on Σ for all $n \in \mathbb{N}$ [31] the noise term is in the form given in Proposition 3.6. Applying Proposition 3.6 gives that for $i = 1, \ldots, N$,

$$\lim_{n\to\infty} \sup\left\{\left\|\sum_{j=n}^{k-1} \alpha_{j+1} U_{j+1}^i\right\|_{BL_i^*} : k = n + 1, \ldots, m(\tau_n + T)\right\} = 0, \quad \text{w.p. } 1.$$

Since

$$\left\|\sum_{j=n}^{k-1} \alpha_{j+1} U_{j+1}\right\|_{\Sigma BL^*} = \max\left\{\left\|\sum_{j=n}^{k-1} \alpha_{j+1} U_{j+1}^1\right\|_{BL_1^*}, \ldots, \left\|\sum_{j=n}^{k-1} \alpha_{j+1} U_{j+1}^N\right\|_{BL_N^*}\right\},$$

it follows that (A3) holds. Applying Theorem 3.3 concludes the proof. □

The following theorem provides a global convergence result for the logit variant of stochastic fictitious play in (2.6) for two-player zero-sum continuous action space games with bounded, Lipschitz continuous rewards. This extends the well known result for stochastic fictitious play in finite action two-player zero-sum games, originally presented by Fudenberg and Kreps [17], Benaïm and Hirsch [2], Hofbauer and Sandholm [21] and Hofbauer and Hopkins [20].

Theorem 4.7. *Consider a two-player zero-sum game, where the action sets are compact, convex subsets of $\mathbb{R}$ and the rewards are bounded and Lipschitz continuous. The beliefs in (2.6) almost surely converge weakly to $\mathcal{LE}_\eta$.*

Proof. These results follow immediately by combining Proposition 4.6 and Corollary 4.5 with Theorem 3.5. □

5. Single population, continuous action space games

Whilst stochastic fictitious play is generally framed for N-player games, as in Section 4, it is also possible to describe a discrete time process akin to stochastic fictitious play for a single population framework. A continuous action single population game has an action set, $A \subset \mathbb{R}$. Throughout this section A is assumed to be compact and convex and $\mathcal{B}$ is the Borel σ-algebra over A. A population is given by the combined play of its members and is a probability measure defined over the measurable space $(A, \mathcal{B})$. So for $B \in \mathcal{B}$, $P(B)$ is the proportion of players using a strategy in B. $\mathcal{P}(A, \mathcal{B})$ represents the set of all populations on A. The game has a payoff function $r(x, y)$ for $x, y \in A$. As in (2.1) the expected payoff of the population P against a population Q is

$$r(P, Q) = \int_A \int_A r(x, y) P(\mathrm{d}x) Q(\mathrm{d}y).$$

For $B \in \mathcal{B}$ and fixed $\eta > 0$ the logit best response to a population P is given by

$$L_\eta(P)(B) := \frac{\int_B \exp\{\eta^{-1} r(z, P)\}\,\mathrm{d}z}{\int_A \exp\{\eta^{-1} r(u, P)\}\,\mathrm{d}u}.$$

Lahkar and Riedel [31] show that $L_\eta(P)$ has an associated density function which, for $x \in A$, is given by

$$l_\eta(P)(x) := \frac{\exp\{\eta^{-1} r(x, P)\}}{\int_A \exp\{\eta^{-1} r(y, P)\}\,\mathrm{d}y}. \tag{5.1}$$

In a similar manner to the N-player logit best response dynamic from (4.1), the single population logit best response dynamic on $\mathcal{M}(A, \mathcal{B})$ is given by the differential equation,

$$\dot{P} = L_\eta(P) - P. \tag{5.2}$$

In the single population context let $\mathcal{LE}_\eta$ be the set of rest points for (5.2). $\mathcal{LE}_\eta$ contains the fixed points of the logit best response map, i.e. the logit equilibria.

Definition 5.1. *A linear, single population game is* negative definite *if for all $P, Q \in \mathcal{P}(A, \mathcal{B})$, with $P \neq Q$,*

$$r(P - Q, P - Q) < 0.$$

The game is negative semi-definite *if the inequality is replaced with a non-strict inequality.*

As discussed by Hofbauer et al. [23] the class of negative semi-definite games includes many common games including, for example, symmetric zero-sum games. The following result is largely taken from Lahkar and Riedel [31] although identical techniques used in Appendices C and D are required to prove the result. For this reason the full proof is omitted.

Theorem 5.2. *If the rewards are bounded and Lipschitz continuous, then, with respect to the bounded Lipschitz norm $\mathcal{M}(A, \mathcal{B})$, for any initial population $P(0) \in \mathcal{P}(A, \mathcal{B})$ there exists a unique solution to the logit best response differential equation (5.2) and $P(t) \in \mathcal{P}(A, \mathcal{B})$ for all $t \in \mathbb{R}^+$. In addition, if the game is negative definite then $\mathcal{LE}_\eta$ is an attractor for the logit best response dynamics for all initial populations in $\mathcal{P}(A, \mathcal{B})$.*

Sketch Proof. As in Section 4, the logit best response (5.2) is studied using the total variation norm on $\mathcal{M}(A, \mathcal{B})$. An identical approach to the proofs of Propositions 4.2 and 4.3 proves a unique solution to (5.2) exists and $\mathcal{LE}_\eta$ is non-empty when the rewards are bounded and Lipschitz continuous.

Then the Lyapunov function used by Lahkar and Riedel [31] can be shown to produce a result as in Proposition D.4 for initial populations in a compact set Δ_D. The convergence result on $\mathcal{P}(A, \mathcal{B})$ for negative-definite single population games follows the proof of Theorem 4.4. Finally, this result is produced for the weak topology on $\mathcal{P}(A, \mathcal{B})$ as in Corollary 4.5. □

Now we study a stochastic fictitious play-like process for single population games which is similar to the population interpretation of stochastic fictitious play used by Ellison and Fudenberg [16]. At each iteration every member of the population knows the true population P_n. A strategy, $x_{n+1} \in A$, is randomly generated from the distribution with density $l_\eta(P_n)$. All the members of the population observe x_{n+1} and a proportion, α_{n+1}, of the population then revise their strategy by abandoning their previous pure strategy in favour of x_{n+1}. Alternative interpretations of the learning process are available (see [21] for details). This discrete time logit best response learning process is therefore given by

$$P_{n+1} = P_n + \alpha_{n+1}[\delta_{x_{n+1}} - P_n], \tag{5.3}$$

where the action x_{n+1} is an action selected randomly from the logit best response density, $l_\eta(P_n)$, and P_0 can be any probability measure in $\mathcal{P}(A, \mathcal{B})$. The following theorem states the convergence result for the iterative process in (5.3) which is based upon the logit best response dynamic.

Proposition 5.3. *Assume that A is a compact, convex subset of $\mathbb{R}$, the rewards are bounded and Lipschitz continuous and*

$$\sum_{n\in\mathbb{N}} \alpha_n = \infty, \qquad \sum_{n\in\mathbb{N}} \alpha_n^2 < \infty.$$

Then a linear interpolation of the stochastic fictitious play-like process (5.3) is almost surely an asymptotic pseudo-trajectory to the single population logit best response dynamics (5.2).

Proof. The proof is omitted as it has an identical structure to the proof of Proposition 4.6. □

Theorem 5.4. *Assume that A is a compact, convex subset of $\mathbb{R}$, the rewards are bounded and Lipschitz continuous and*

$$\sum_{n\in\mathbb{N}} \alpha_n = \infty, \qquad \sum_{n\in\mathbb{N}} \alpha_n^2 < \infty.$$

If the game is single population negative definite then the stochastic fictitious play-like process (5.3) almost surely converges to $\mathcal{LE}_\eta$ under the bounded Lipschitz norm on $\mathcal{P}(A, \mathcal{B})$.

Proof. These results follow immediately by combining Theorem 5.2 and Proposition 5.3 with Theorem 3.5. □

6. Examples

An important feature of any learning algorithm is that it can easily be implemented. This is demonstrated in this section for extensions of two well-studied classical games. These extensions both have reward functions which are linear in the players' actions which allows the cdf-inversion method to be used to simulate actions x^i_{n+1} from $L^i_\eta(\sigma^{-i}_n)$. If the reward function is quadratic then

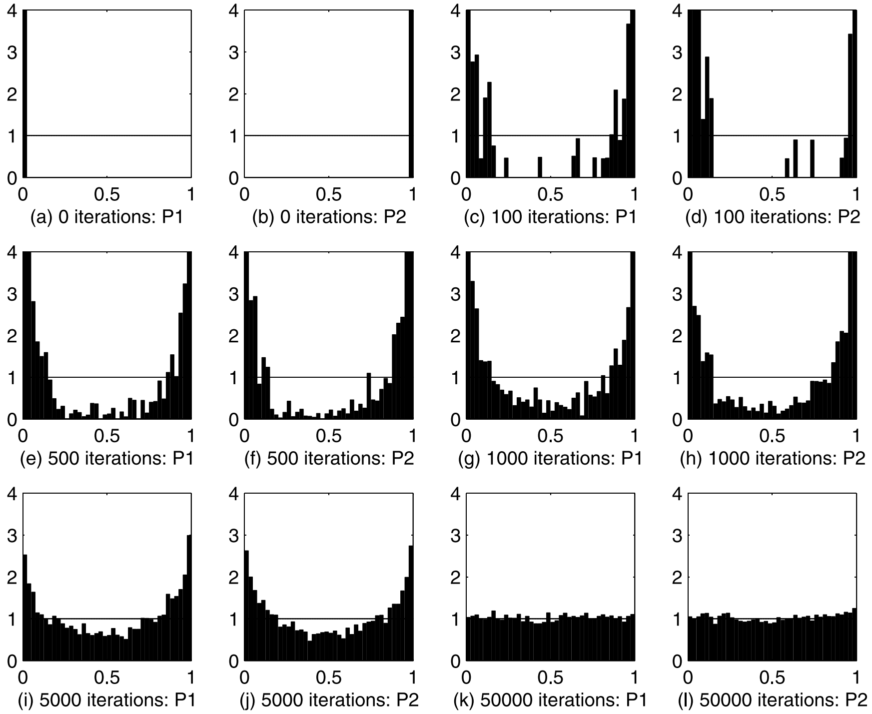

Fig. 1. Evolution of stochastic fictitious play for the continuous "matching pennies" game with $\sigma_0^1 = \delta_0$, $\sigma_0^2 = \delta_1$, $\alpha_n = (n+20)^{-1}$ and $\eta = 0.005$. In each pair of plots a sample from σ_n^1 is in the left-hand plot and a sample from σ_n^2 is in the right-hand plot and the logit equilibrium of the game is also displayed.

$L_\eta^i(\sigma_n^{-i})$ corresponds to a truncated normal distribution. If the reward function is more complex then more sophisticated methods are required to produce a simulation from $L_\eta^i(\sigma_n^{-i})$.

In the first example the stochastic fictitious play algorithm of Section 5 is studied in a two-player zero-sum extension to the standard matching pennies game. Let $r^1(x, y) = (x - 1/2) \times (y - 1/2)$ and $r^2(x, y) = -r^1(x, y)$, where $A^i = [0, 1]$ for $i = 1, 2$. For any $\varepsilon \in [0, 1/2]$, $\tilde{\sigma}^i = 1/2(\delta_\varepsilon + \delta_{1-\varepsilon})$, for $i = 1, 2$, is a Nash equilibrium, as is the joint strategy for both players to play an action in [0, 1] uniformly at random. However, there is a unique logit equilibrium in which $\tilde{\sigma}^i$ is the uniform distribution on [0, 1] for $i = 1, 2$. Theorem 4.7 shows that the beliefs in the stochastic fictitious play process (2.6) converge to this logit equilibrium. This convergence is demonstrated in Fig. 1.

The second example is a single population game, which extends the classic hawk-dove game to the continuous action framework. For $C > V > 0$ let

$$r(x, y) = \left[1 - (y - x)\right]\frac{V}{2} - xy\frac{C}{2},$$

and $A = [0, 1]$. An action in A corresponds to an aggression level, so $x = 1$ corresponds to playing 'hawk' in the classic game and $x = 0$ corresponds to playing 'dove'. The first term in the reward function represents the likelihood of obtaining a resource of value V and the second term is the cost of 'injuries' sustained in contesting the resource. Both of these terms depend linearly on how aggressively each of the players contests the resource. It is straightforward to

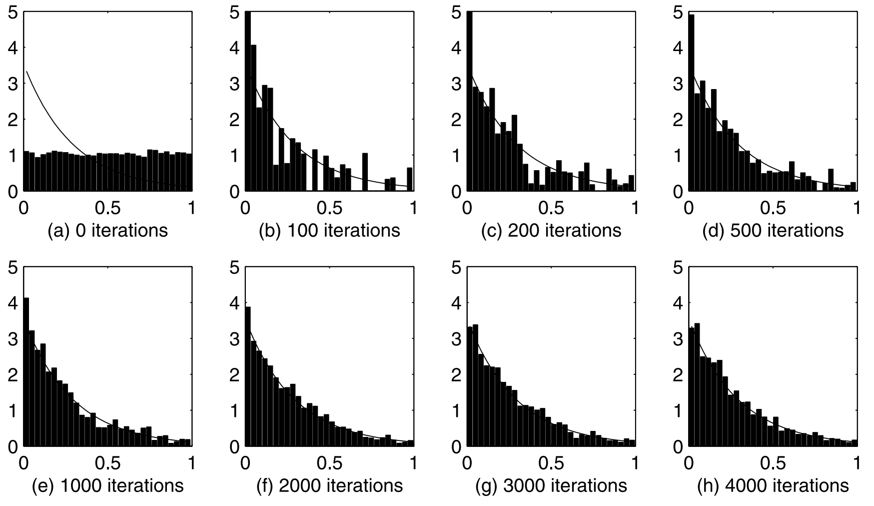

Fig. 2. Evolution of single population stochastic fictitious play-like process for the continuous "Hawk–Dove" game with $V = 1$, $C = 4$, $\alpha_n = (n + 20)^{-1}$ and $\eta = 0.005$. In each plot a sample from the population is shown along with the logit equilibrium of the game.

show that this continuous hawk-dove game is negative definite (see Definition 5.1), and so Theorem 5.4 guarantees the stochastic fictitious play-like process in (5.3) converges to the set of logit equilibria.

This game has infinitely many Nash equilibria as $P = 1/2\delta_{V/C+\varepsilon} + 1/2\delta_{V/C-\varepsilon}$ is a Nash equilibrium for any $\varepsilon \in [0, \min\{V/C, 1 - V/C\}]$, but it has a unique logit equilibrium, as we show. Let $\tilde{P} \in \mathcal{P}(A, \mathcal{B})$ be a logit equilibrium, with associated density $\tilde{p}$. Because the reward function $r(x, y)$ for this continuous hawk-dove game is linear in x it follows that for any $P \in \mathcal{P}(A, \mathcal{B})$, $l_\eta(P)(x) \propto e^{kx}$, for some $k \in \mathbb{R}$. In particular, if $\tilde{P} \in \mathcal{P}(A, \mathcal{B})$ is a logit equilibrium then $\tilde{p}(x) = l_\eta(\tilde{P})(x) \propto e^{kx}$. Normalisation[2] implies that

$$\tilde{p}(x) = \frac{ke^{kx}}{e^k - 1}. \tag{6.1}$$

Knowing that $\tilde{p}(x) = l_\eta(\tilde{P})(x)$ for all $x \in A$ and combining it with (6.1) and the definition of $l_\eta(\tilde{P})$ from (5.1) allows the exact value of k to be calculated. A simulation of this game with $V = 1$, $C = 4$ and $\eta = 0.005$ is shown in Fig. 2. In the discrete action hawk-dove game with $V = 1$, $C = 4$ the Nash equilibrium is for $3/4$ of the population to play 'dove'. With these same parameters the logit equilibrium here is also skewed towards the 'dove' action, capturing this particular feature of the traditional hawk-dove game.

7. Discussion

In this work we present a method for studying the limiting behaviour of iterative learning processes with an uncountably infinite action space. This extends the work of Fudenberg and

[2] When $C = 2V$, $l_\eta(P)(x) \propto e^{kx}$ remains true, but $k = 0$ meaning the normalisation in (6.1) is not valid and the logit equilibrium is a uniform distribution.

Kreps [17], Benaïm and Hirsch [2], Hofbauer and Sandholm [21] and Hofbauer and Hopkins [20] to games with actions selected from a continuous set.

To achieve this we have developed new tools for stochastic approximation. These build on the asymptotic pseudo-trajectory approach to stochastic approximation of Benaïm [1] and extend the abstract stochastic approximation approach presented by Shwartz and Berman [43] to this now more common framework. As a suitable space for continuous strategies we study the space of finite signed measures, $\mathcal{M}(X, \mathcal{B})$, using the bounded Lipschitz norm. Unlike Shwartz and Berman [43] we provide simple conditions in the spirit of Benaïm [1] as to when the difficult martingale noise assumption holds for the bounded Lipschitz norm on $\mathcal{M}(X, \mathcal{B})$.

In Section 4 we present the key application of this framework to stochastic fictitious play in continuous action games. We extend the existence and uniqueness results of Lahkar and Riedel [31] to the N-player case. As a consequence we can study a logit variant of stochastic fictitious play for continuous action space games. We prove the convergence of stochastic fictitious play for two-player zero-sum games with continuous action spaces and bounded, Lipschitz continuous rewards. This extends the previous results of Fudenberg and Kreps [17], Benaïm and Hirsch [2], Hofbauer and Sandholm [21] and Hofbauer and Hopkins [20] for stochastic fictitious play in normal form games with finite action sets.

In Section 5 a stochastic fictitious play-like process in single population games is investigated. The continuous time convergence results of Lahkar and Riedel [31] are extended to this discrete time process to show convergence of a population for negative definite games with a continuous action set and bounded, Lipschitz continuous rewards.

In Section 6 two examples are provided to demonstrate the convergence of these processes in continuous action space games. This demonstrates that stochastic fictitious play can be implemented in these games.

Our abstract stochastic approximation framework could also be used to study learning variants of the replicator dynamics, such as in [38,6] and [34, Chapter 2]. Although the replicator dynamics are studied more frequently, especially in the continuous action space literature (see [39–41,12]), the framework has limitations. In particular, the replicator dynamics do not generate new strategies, so that the limit point of a trajectory always depends on the initial conditions. Even when exploration of the state space can be guaranteed convergence of an associated learning process is generally very slow. In contrast, the logit best response is absolutely continuous and assigns some probability to every part of the action space, which makes it more straightforward to study the associated learning processes. This is true of the discrete action case, where fictitious play and stochastic fictitious play are more frequently studied than similar discrete time variants of the replicator dynamics, and remains true for the continuous action case.

Acknowledgments

We gratefully acknowledge significant contributions arising from discussions with Profs. Frank Riedel and Ratul Lahkar, and Dr Nick Whiteley. We also received extremely helpful comments from anonymous reviewers and editors. Steve Perkins' Ph.D. research was funded by grant number EP/D063485/1 from the United Kingdom Engineering and Physical Sciences Research Council.

Appendix A. Noise criteria: L^2

It is important for us to be able to verify that the noise condition (A3) holds for martingales on $\mathcal{M}(X,\mathcal{B})$ so that stochastic approximation can be performed on this Banach space. Doing so is not straightforward and requires us to use an intermediate result, proving that (A3) holds for martingale noise on L^2, presented in this appendix. The proof that (A3) holds for martingales in Euclidean space [1, Proposition 4.2] relies on the Burkholder–Davis–Gundy inequality to study the martingale difference sequence and this inequality can be extended for certain Banach spaces, notably L^2. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ a filtration on $\mathcal{F}$. Further discussion of Banach space valued random variables is available in [33].

Proposition A.1. *Consider the stochastic approximation process given in (3.1) on the Banach space of L^2 functions with associated norm $\|\cdot\|_{L^2}$. If for some $q\geq 2$,*

(1) $\{\alpha_n\}_{n\in\mathbb{N}}$ *is deterministic with* $\sum_{n\in\mathbb{N}}\alpha_n^{1+q/2}<\infty$,
(2) $\{U_n\}_{n\in\mathbb{N}}$ *is adapted and measurable with respect to $\mathcal{F}_n$ for all n such that*

$$\mathbb{E}[U_{n+1}|\mathcal{F}_n]=0,\qquad \sup_n \mathbb{E}\big[\|U_n\|_{L^2}^q\big]<\infty,$$

then

$$\lim_{n\to\infty}\sup\left\{\left\|\sum_{i=n}^{k-1}\alpha_{i+1}U_{i+1}\right\|_{L^2} : k=n+1,\dots,m(\tau_n+T)\right\}=0,\quad \textit{w.p. 1}.$$

Because the Burkholder–Davis–Gundy inequality can be extended for L^2-valued martingales the proof of this result follows a similar pattern to Benaïm [1, Proposition 4.2] for martingales in Euclidean space.

Proof. The first issue is to ensure that the term of interest,

$$\sup_k\left\{\left\|\sum_{j=n}^{k-1}\alpha_{j+1}U_{j+1}\right\|_{L^2} : k=n+1,\dots,m(\tau_n+T)\right\},\tag{A.1}$$

is measurable with respect to the probability space $(\Omega,\mathcal{F},\mathbb{P})$. Firstly, a topology, $\mathcal{T}$, is produced using the open subsets generated by $\|\cdot\|_{L^2}$. $\mathcal{T}$ is the collection of all Borel sets on the space of L^2 functions so that $\mathcal{T}$ forms the Borel σ-algebra on L^2. Let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra over $\mathbb{R}$. $\|\cdot\|_{L^2}:L^2\to\mathbb{R}$ is a continuous function from $(L^2,\mathcal{T})$ to $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ and therefore an inverse $\|\ \|_{L^2}^{-1}(\cdot):\mathbb{R}\rightrightarrows L^2$ exists. Now for any open set $A\in\mathcal{B}(\mathbb{R})$, $\|\ \|_{L^2}^{-1}(A)=\{U\in L^2:\|U\|_{L^2}\in A\}$. Since A is an open set it follows from the continuity of $\|\cdot\|_{L^2}$ that $\{U\in L^2:\|U\|_{L^2}\in A\}$ is an open set. Since $\mathcal{T}$ contains all such open sets it follows that

$$\|\ \|_{L^2}^{-1}(A)=\big\{U\in L^2:\|U\|_{L^2}\in A\big\}\in\mathcal{T},$$

and therefore $\|\cdot\|_{L^2}$ is a measurable function from $(L^2,\mathcal{T})$ to $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Now, fix $T>0$. Since U_n, which maps from $(\Omega,\mathcal{F},\mathbb{P})$ to L^2, is measurable with respect to the probability space $(\Omega,\mathcal{F},\mathbb{P})$ then for $k\in\{n+1,\dots,m(\tau_n+T)\}$, the finite (countable) weighted summation, $\sum_{j=n}^{k-1}\alpha_{j+1}U_{j+1}$ is also measurable with respect to $(\Omega,\mathcal{F},\mathbb{P})$. Then

$$\left\| \sum_{j=n}^{k-1} \alpha_{j+1} U_{j+1} \right\|_{L^2},$$

is measurable with respect to $(\Omega, \mathcal{F}, \mathbb{P})$ as it is the composition of the two measurable functions $\sum_{j=n}^{k-1} \alpha_{j+1} U_{j+1} : (\Omega, \mathcal{F}, \mathbb{P}) \to L^2$ and $\| \cdot \|_{L^2} : L^2 \to \mathbb{R}$. Finally, (A.1) must be measurable with respect to $(\Omega, \mathcal{F}, \mathbb{P})$ since a finite (countable) supremum of measurable terms remains measurable.

Brzeźniak [10] shows that the Burkholder–Davis–Gundy inequality can be extended for a class of Banach spaces which includes L^2. This means that for any $q \in (1, \infty)$ there exists a $C_q > 0$ such that for an L^2-valued martingale $\{Y_n\}_{n \in \mathbb{N}}$ and stopping time $N > 0$ then

$$\mathbb{E}\Big[\sup_{n \le N} \|Y_n\|_{L^2}^q\Big] \le C_q \mathbb{E}\left[\left(\sum_{n=1}^{N} \|Y_n - Y_{n-1}\|_{L^2}^2\right)^{q/2}\right]. \tag{A.2}$$

The remainder of the proof is an extension to the proof of Benaïm [1, Proposition 4.2] for L^2 using (A.2) rather than the original Burkholder–Davis–Gundy inequality. Let $W_{n,m} := \sum_{i=n}^{m} \alpha_{i+1} U_{i+1}$ for $m \ge n$. Since $U_{i+1} \in L^2$, it follows that $W_{n,m} \in L^2$. Now fixing $n \in \mathbb{N}$, we have that

$$\mathbb{E}[W_{n,m} | \mathcal{F}_m] = W_{n,m-1},$$

and hence $W_{n,m}$ is a martingale in L^2. Using (A.2) there exists some constant $C_q > 0$ such that

$$\mathbb{E}\Big[\sup_{n \le k \le m(\tau_n + T)} \|W_{n,k}\|_{L^2}^q\Big] \le C_q \mathbb{E}\left[\left(\sum_{i=n+1}^{m(\tau_n+T)} \|W_{n,i} - W_{n,i-1}\|_{L^2}^2\right)^{q/2}\right].$$

Now by noticing that $\|W_{n,i} - W_{n,i-1}\|_{L^2} = \alpha_{i+1} \|U_{i+1}\|_{L^2}$ and using the original definition of $W_{n,m}$ gives

$$\mathbb{E}\left[\sup_{n \le k \le m(\tau_n + T)} \left\| \sum_{i=n}^{k} \alpha_{i+1} U_{i+1} \right\|_{L^2}^q\right] \le C_q \mathbb{E}\left[\left(\sum_{i=n}^{m(\tau_n+T)-1} \alpha_{i+1}^2 \|U_{i+1}\|_{L^2}^2\right)^{q/2}\right].$$

The remainder of the proof continues exactly as in [1, Proposition 4.2] with the only difference being the norm used here is $\| \cdot \|_{L^2}$ and in [1, Proposition 4.2] this is the standard Euclidean norm. □

Remark A.2. Brzeźniak [10] shows that the Burkholder–Davis–Gundy inequality can be extended for a class of Banach spaces which includes L^p-spaces for any $p \in [2, \infty)$. The result in Proposition A.1 for L^2 can be extended to any Banach space in the class for which Brzeźniak's result holds with no alteration to the proof.

Appendix B. Proof of Proposition 3.6

If we could show that $\mathcal{M}(X, \mathcal{B})$, with an appropriate norm, is in the class of Banach spaces for which Brzeźniak [10] proves the Burkholder–Davis–Gundy inequality holds then the proof of Proposition 3.6 would be identical to the proof of Proposition A.1. However, we have been unable to show this and so we are forced to take a different approach. Here we approximate all Dirac measures with a spike of fixed width. This allows us to consider functions with proper density functions in L^2. We are then able to use Proposition A.1 to show the convergence, while also showing that the additional error term introduced in this approach is not significant.

Proof of Proposition 3.6. Firstly, the measurability of

$$\sup_k \left\{ \left\| \sum_{j=n}^{k-1} \alpha_{j+1} U_{j+1} \right\|_{BL^*} : k = n+1, \ldots, m(\tau_n + T) \right\},$$

with respect to the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ follows as in the proof of Proposition A.1.

Fix $\gamma > 0$. We approximate atoms in X by measures that have a spike density with base width 2γ. Hence we need to consider an expanded space $\bar{X}$ to accommodate spikes near the boundary of X.[3] Suppose $X \subseteq [a, b] \subset \mathbb{R}$, and define $\bar{X} := [a - \gamma, b + \gamma]$. Let $\| \cdot \|_{\bar{BL}}$, $\bar{\mathcal{BL}}$ and $\| \cdot \|_{\bar{BL}^*}$ be as defined in (3.4)–(3.6) but with integrals over $\bar{X}$ and $\bar{\mathcal{B}}$ the Borel σ-algebra over $\bar{X}$. $\mathcal{M}(\bar{X}, \bar{\mathcal{B}})$ is the space of finite signed measures on $\bar{X}$ and is equipped with the BL^* topology using $\| \cdot \|_{\bar{BL}^*}$.

Consider arbitrary $\tilde{z} \in X \subset \bar{X}$, let $\bar{h}$ be a spike density on $\bar{X}$ centered on $\tilde{z}$ defined as

$$\bar{h}(z, \tilde{z}) := \begin{cases} \frac{1}{\gamma^2}(z - (\tilde{z} - \gamma)), & z \in [\tilde{z} - \gamma, \tilde{z}] \\ \frac{1}{\gamma}(1 - \frac{1}{\gamma}(z - \tilde{z})), & z \in (\tilde{z}, \tilde{z} + \gamma] \\ 0, & \text{otherwise} \end{cases} \tag{B.1}$$

and let $\bar{H}(\tilde{z})$ be the measure in $\mathcal{M}(\bar{X}, \bar{\mathcal{B}})$ associated with density $\bar{h}(\cdot, \tilde{z})$. We firstly show that

$$\left\| \delta_{\tilde{z}} - \bar{H}(\tilde{z}) \right\|_{\bar{BL}^*} \le \gamma$$

First note that if $\bar{g} \in \bar{\mathcal{BL}}$ then $\bar{g}$ has a Lipschitz constant that is not more than 1, so $|z - y| \le \gamma \Rightarrow |\bar{g}(z) - \bar{g}(y)| \le \gamma$. Conversely if $|z - y| \ge \gamma$ then $\bar{h}(z, y) = 0$. Hence

$$\begin{aligned} \left| \int_{\bar{X}} \bar{g}(z)\big(\delta_{\tilde{z}} - H(\tilde{z})\big)\, \mathrm{d}z \right| &= \left| \bar{g}(\tilde{z}) - \int_{\bar{X}} \bar{g}(z)\bar{h}(z, \tilde{z})\, \mathrm{d}z \right| \\ &= \left| \int_{\bar{X}} \big[\bar{g}(\tilde{z}) - \bar{g}(z)\big]\bar{h}(z, \tilde{z})\, \mathrm{d}z \right| \\ &\le \int_{\tilde{z}-\gamma}^{\tilde{z}+\gamma} \big|\bar{g}(\tilde{z}) - \bar{g}(z)\big|\bar{h}(z, \tilde{z})\, \mathrm{d}z \\ &= \int_{\tilde{z}-\gamma}^{\tilde{z}+\gamma} \gamma \bar{h}(z, \tilde{z})\, \mathrm{d}z \\ &= \gamma. \end{aligned} \tag{B.2}$$

Hence

[3] This is a particular issue which Chen and White [11] did not address when using probability densities on L^2.

$$\left\|\delta_{\tilde{z}} - \bar{H}(\tilde{z})\right\|_{\bar{BL}^*} = \sup_{\bar{g}\in\bar{\mathcal{BL}}}\left|\int_{\bar{X}} \bar{g}(z)\big(\delta_{\tilde{z}} - \bar{H}(\tilde{z})\big)(\mathrm{d}z)\right| \le \gamma.$$

To use this within our stochastic approximation process we define $\bar{H}_n := \bar{H}(x_n)$ so that for all $n \in \mathbb{N}$,

$$\|\delta_{x_n} - \bar{H}_n\|_{\bar{BL}^*} \le \gamma. \tag{B.3}$$

To examine the convergence of $\{U_n\}_{n\in\mathbb{N}}$ we also need to consider the convolution

$$\bar{q}_n(z) := \int_X \bar{h}(z, y) p_n(y)\,\mathrm{d}y, \tag{B.4}$$

where $\bar{h}$ is the spike density defined above and p_n is the density of measure P_n and $z \in \bar{X}$. Let $\bar{Q}_n$ be the measure on $\mathcal{M}(\bar{X}, \bar{\mathcal{B}})$ associated with $\bar{q}_n$. This is useful since $\bar{h}_{n+1}$ can be viewed as an L^2-valued random variable and importantly we have that

$$\mathbb{E}[\bar{h}_{n+1}|\mathcal{F}_n] = \bar{q}_n. \tag{B.5}$$

Since both $\bar{h}_{n+1}$ and $\bar{q}_n$ are in $L^2(\bar{X})$, (B.5) means that $\bar{h}_{n+1} - \bar{q}_n$ is an L^2-valued martingale.

Finally, we need to define $\bar{U}_{n+1} = \delta_{x_{n+1}} - \bar{P}_n$, which is the extension of U_{n+1} to $\bar{X}$, where

$$\bar{P}_n := \begin{cases} P_n, & \text{on } X, \\ 0, & \text{on } \bar{X}\backslash X, \end{cases}$$

and $\bar{P}_n$ has a density function $\bar{p}_n$. For fixed $T > 0$ it is clear for $k = n+1, \dots, m(\tau_n + T)$ that

$$\begin{aligned}\left\|\sum_{i=n}^{k-1}\alpha_{i+1}U_{i+1}\right\|_{BL^*} &= \left\|\sum_{i=n}^{k-1}\alpha_{i+1}\bar{U}_{i+1}\right\|_{\bar{BL}^*} \\ &\le \left\|\sum_{i=n}^{k-1}\alpha_{i+1}(\delta_{x_{i+1}} - \bar{H}_{i+1})\right\|_{\bar{BL}^*} + \left\|\sum_{i=n}^{k-1}\alpha_{i+1}(\bar{H}_{i+1} - \bar{Q}_i)\right\|_{\bar{BL}^*} \\ &\quad + \left\|\sum_{i=n}^{k-1}\alpha_{i+1}(\bar{Q}_i - \bar{P}_i)\right\|_{\bar{BL}^*}.\end{aligned} \tag{B.6}$$

We address each of these terms in turn. Using (B.3) we see that

$$\left\|\sum_{i=n}^{k-1}\alpha_{i+1}(\delta_{x_{i+1}} - \bar{H}_{i+1})\right\|_{\bar{BL}^*} \le \sum_{i=n}^{k-1}\alpha_{i+1}\gamma \approx T\gamma. \tag{B.7}$$

The definition of $\|\cdot\|_{\bar{BL}^*}$ implies that $\|\cdot\|_{\bar{BL}^*} \le \|\cdot\|_{L^1}$, and a standard result for the L^1-norm, which follows from Hölder's inequality, is that if $\bar{X}$ is a compact, convex subset of $\mathbb{R}$ then $\|\cdot\|_{L^1} \le |\bar{X}|^{1/2}\|\cdot\|_{L^2}$. Hence

$$\left\|\sum_{i=n}^{k-1}\alpha_{i+1}(\bar{H}_{i+1} - \bar{Q}_i)\right\|_{\bar{BL}^*} \le |\bar{X}|^{1/2}\left\|\sum_{i=n}^{k-1}\alpha_{i+1}(\bar{h}_{i+1} - \bar{q}_i)\right\|_{L^2}. \tag{B.8}$$

We have already observed that $\bar{h}_{n+1} - \bar{q}_n$ is an L^2-valued martingale sequence, and under the assumptions of Proposition 3.6, since p_n is bounded $\bar{q}_n$ is also bounded, and there exists $C > 0$ such that

$$\sup_n \mathbb{E}\big[\|\bar{h}_{n+1}-\bar{q}_n\|_{L^2}^2\big] \le \sup_n\big\{\mathbb{E}\big[\|\bar{h}_{n+1}\|_{L^2}^2\big]\big\} + \sup_n\big\{\mathbb{E}\big[\|\bar{q}_n\|_{L^2}^2\big]\big\} < \frac{2}{\gamma}|\bar{X}| + C^2|\bar{X}| < \infty$$

Using Proposition A.1 immediately gives

$$\lim_{n\to\infty}\sup\left\{\left\|\sum_{i=n}^{k-1}\alpha_{i+1}(\bar{h}_{i+1}-\bar{q}_i)\right\|_{L^2} : k=n+1,\dots,m(\tau_n+T)\right\}=0. \tag{B.9}$$

Finally, following a similar approach to that used in obtaining (B.2), the definition of $\bar{q}_n(\cdot)$ in (B.4) tells us that for any $\bar{g}\in \bar{BL}$

$$\begin{aligned}\left|\int_{\bar{X}}\bar{g}(z)\big[\bar{q}_n(z)-\bar{p}_n(z)\big]\,\mathrm{d}z\right| &= \left|\int_{\bar{X}}\bar{g}(z)\left[\int_{\bar{X}}\bar{h}(z,y)\bar{p}_n(y)\,\mathrm{d}y-\bar{p}_n(z)\right]\mathrm{d}z\right| \\ &= \left|\int_{\bar{X}}\left[\int_{\bar{X}}\bar{g}(z)\bar{h}(z,y)\,\mathrm{d}z-\bar{g}(y)\right]\bar{p}_n(y)\,\mathrm{d}y\right| \\ &\le \int_{\bar{X}}\left|\int_{\bar{X}}\bar{g}(z)\bar{h}(z,y)\,\mathrm{d}z-\bar{g}(y)\right|\bar{p}_n(y)\,\mathrm{d}y \\ &= \int_{\bar{X}}\left|\int_{\bar{X}}\big[\bar{g}(z)-\bar{g}(y)\big]\bar{h}(z,y)\,\mathrm{d}z\right|\bar{p}_n(y)\,\mathrm{d}y \\ &\le \int_{\bar{X}}\gamma\,\bar{p}_n(y)\,\mathrm{d}y \\ &= \gamma.\end{aligned}$$

It then follows that

$$\|\bar{Q}_n-\bar{P}_n\|_{\bar{BL}^*}\le\gamma, \tag{B.10}$$

and from this

$$\left\|\sum_{i=n}^{k-1}\alpha_{i+1}(\bar{Q}_i-\bar{P}_i)\right\|_{\bar{BL}^*}\le\sum_{i=n}^{k-1}\alpha_{i+1}\|\bar{Q}_i-\bar{P}_i\|_{\bar{BL}^*}\le T\gamma. \tag{B.11}$$

Taking the appropriate limit and supremum of (B.6) and substituting (B.7)–(B.9) and (B.11) gives,

$$\begin{aligned}&\lim_{n\to\infty}\sup\left\{\left\|\sum_{i=n}^{k-1}\alpha_{i+1}U_{i+1}\right\|_{BL^*} : k=n+1,\dots,m(\tau_n+T)\right\}\\ &\quad\le 2T\gamma+|\bar{X}|^{1/2}\lim_{n\to\infty}\sup\left\{\left\|\sum_{i=n}^{k-1}\alpha_{i+1}(\bar{h}_{i+1}-\bar{q}_i)\right\|_{L^2} : k=n+1,\dots,m(\tau_n+T)\right\}\\ &\quad=2T\gamma.\end{aligned}$$

Noting that the initial choice of $\gamma>0$ was arbitrary completes the proof. □

Appendix C. N-player logit BR dynamics

This section is used to present the proofs of Propositions 4.2 and 4.3. Lahkar and Riedel [31] prove the existence of a unique solution flow for the logit best response dynamics in single population continuous action games and so many of the results in this section are natural extensions to these previous results.

The logit best response dynamic, (4.1), is studied under the strong topology on Σ using $\|\cdot\|_{\Sigma TV}$. Recall our compact subset of Δ, defined as

$$\Delta_D := \big\{\underline{\pi} \in \Delta : \forall i = 1, \dots, N, \pi^i \text{ is absolutely continuous with density } p^i \text{ such that } D^{-1} \le p^i(x^i) \le D, \text{ for all } x^i \in A^i \text{ and } p^i(\cdot) \text{ is Lipschitz continuous with constant } D\big\}.$$

Lemma C.1. *If the rewards are bounded and Lipschitz continuous then there exists a $D < \infty$ such that for all $\underline{\pi} \in \Delta$ then $L_\eta(\underline{\pi}) \in \Delta_D$. In addition, both Δ and Δ_D are forward invariant under logit best response dynamics.*

Proof. Let $r_{\max} := \max_{i,\underline{x}} r^i(\underline{x})$ and $r_{\min} := \min_{i,\underline{x}} r^i(\underline{x})$. There exists a $\bar{D} < \infty$ such that for every $\underline{\pi} \in \Delta$,

$$l_\eta^i(\pi^{-i})(x^i) \le \frac{\exp\{\eta^{-1} r_{\max}\}}{|A^i| \exp\{\eta^{-1} r_{\min}\}} = \bar{D} < \infty.$$

Similarly there exists a $\underline{D} > 0$ such that for every $\underline{\pi} \in \Delta$

$$l_\eta^i(\pi^{-i})(x^i) \ge \frac{\exp\{\eta^{-1} r_{\min}\}}{|A^i| \exp\{\eta^{-1} r_{\max}\}} = \underline{D} > 0.$$

Furthermore, let L be the Lipschitz constant for all of the reward functions r^i. Then

$$\big|l_\eta^i(\pi^{-i})(x^i) - l_\eta^i(\pi^{-i})(y^i)\big| \le \frac{2L|x^i - y^i| \exp\{\eta^{-1} r_{\max}\}}{\eta |A^i| \exp\{\eta^{-1} r_{\min}\}} = \tilde{L}\big|x^i - y^i\big|.$$

Taking $D := \max\{\bar{D}, \underline{D}^{-1}, \tilde{L}\}$ proves the first claim.

Note that Δ is forward invariant under the logit best response dynamics if for $i = 1, \dots, N$, $\int_{A^i} \dot{\pi}^i(\mathrm{d}x) = 0$. Now, by definition, for all $\underline{\pi} \in \Delta$, $L_\eta(\underline{\pi}) \in \Delta$. Therefore for $\underline{\pi} \in \Delta$

$$\dot{\pi}^i(A^i) = L_\eta^i(\pi^{-i})(A^i) - \pi^i(A^i) = 1 - 1 = 0.$$

Hence Δ is forward invariant under the logit best response dynamics.

Similarly, because $L_\eta(\underline{\pi}) \in \Delta_D$ it must also be that if $\underline{\pi}(0) \in \Delta_D$ then $\underline{\pi}(t) \in \Delta_D$ for all $t \in \mathbb{R}^+$, so that Δ_D is forward invariant under the logit best response dynamics. □

Lemma C.2. *If the rewards are bounded then the logit best response map $L_\eta(\cdot) : \Sigma \to \Delta$ is Lipschitz continuous with respect to the total variation norm on Δ, meaning that there exists a $K > 0$ such that for all $\underline{\pi}, \underline{\rho} \in \Delta$*

$$\big\|L_\eta(\underline{\pi}) - L_\eta(\underline{\rho})\big\|_{\Sigma TV} \le K \|\underline{\pi} - \underline{\rho}\|_{\Sigma TV}.$$

Similarly, if the rewards are bounded and Lipschitz continuous then logit best response map $L_\eta(\cdot) : \Sigma \to \Delta$ is Lipschitz continuous with respect to the bounded Lipschitz norm on Δ.

Note that this is continuity of the map L_η from Σ to Δ, in contrast to the Lipschitz continuity of the density function $l^i_\eta(\pi^{-i})$ for fixed π^{-i} that is considered in Lemma C.1.

Proof. Let $\bar{r} := \max_{i,\underline{x}} |r^i(\underline{x})|$ and take $\underline{\pi}, \underline{\rho} \in \Sigma$. It follows that for every $x^i \in A^i$,

$$\begin{aligned}
\left|r^i\left(x^i, \pi^{-i}\right) - r^i\left(x^i, \rho^{-i}\right)\right| &\le \bar{r} \prod_{j\neq i}^{N} \int_{A^j} \left|\pi^j - \rho^j\right|\left(\mathrm{d}x^j\right) \\
&\le \bar{r} \prod_{j\neq i}^{N} \left|A^j\right| \left\|\pi^j - \rho^j\right\|_{TV_j} \\
&\le \bar{r} \prod_{j\neq i}^{N} \left|A^j\right| \|\pi - \rho\|_{\Sigma TV}.
\end{aligned}$$

This means that $r^i(x^i, \pi^{-i})$ is Lipschitz continuous in π^{-i} with respect to the total variation norm on Σ. Then, since A^i is compact for $i = 1, \ldots, N$, $\int_{A^i} \exp\{\eta^{-1} r^i(x^i, \pi^{-i})\}\,\mathrm{d}x^i$ is Lipschitz continuous in π^{-i} and clearly for any $B^i \subseteq A^i$ $\int_{B^i} \exp\{\eta^{-1} r^i(x^i, \pi^{-i})\}\,\mathrm{d}x^i$ is also Lipschitz continuous with respect to π^{-i}. The ratio of two Lipschitz continuous functions is also Lipschitz continuous given that the denominator is such that

$$\int_{A^i} \exp\left\{\eta^{-1} r^i\left(x^i, \pi^{-i}\right)\right\} \mathrm{d}x^i \ge \left|A^i\right| \exp\left\{\eta^{-1} r_{\min}\right\} > 0.$$

Therefore as the ratio of two Lipschitz continuous functions it must be that $L^i_\eta(\pi^{-i})$ is Lipschitz continuous in π^{-i} with respect to the total variation norm on Σ. It follows then that $L_\eta(\cdot) : \Sigma \to \Delta$ is Lipschitz continuous with respect to the total variation norm on Δ, as required.

Now we prove the Lipschitz continuity with respect to the bounded Lipschitz norm. Fix $x^i \in A^i$. $r^i(\cdot,\cdot) : \underline{A} \to \mathbb{R}$ is Lipschitz continuous, so there exists a Lipschitz constant C such that

$$\left|r^i\left(x^i, x^{-i}\right) - r^i\left(x^i, y^{-i}\right)\right| \le C\left|x^{-i} - y^{-i}\right|.$$

Let $\tilde{r}^i(x^i, x^{-i}) := r^i(x^i, x^{-i})/(C + \bar{r})$. It follows that $\tilde{r}^i(x^i, \cdot) : A^{-i} \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant $C/(C + \bar{r})$ and maximum value $\bar{r}/(C + \bar{r})$.

Take $\underline{\pi}, \underline{\rho} \in \Delta$. From the definition it follows that

$$\begin{aligned}
&\left|r^i\left(x^i, \pi^{-i}\right) - r^i\left(x^i, \rho^{-i}\right)\right| \\
&\quad= \left|\int_{A^1} \ldots \int_{A^N} r^i\left(z^1, \ldots, z^N\right) \delta_{x^i}\left(\mathrm{d}z^i\right) \prod_{j\neq i}^{N} \left(\pi^j - \rho^j\right)\left(\mathrm{d}z^j\right)\right| \\
&\quad= (C + \bar{r}) \left|\int_{A^1} \ldots \int_{A^N} \tilde{r}^i\left(z^1, \ldots, z^N\right) \delta_{x^i}\left(\mathrm{d}z^i\right) \prod_{j\neq i}^{N} \left(\pi^j - \rho^j\right)\left(\mathrm{d}z^j\right)\right| \\
&\quad\le (C + \bar{r}) \prod_{j\neq i}^{N} \left|A^j\right| \|\underline{\pi} - \underline{\rho}\|_{\Sigma BL^*},
\end{aligned}$$

where the final line holds because for each fixed z^{-j}, $\tilde{r}^i(z^1,\dots,z^N)$ is a Lipschitz continuous function mapping from A^j to $\mathbb{R}$ which has absolute value and Lipschitz constant less than or equal to 1. The Lipschitz continuity of $L_\eta(\cdot):\Sigma\to\Delta$ with respect to $\|\cdot\|_{\Sigma BL^*}$ follows using identical arguments to those which were used for the total variation norm. □

Proof of Proposition 4.2. Firstly, Δ_D is shown to be a compact subset of Δ under the strong topology. Let $\Delta_D^i := \{\pi^i : \underline{\pi}\in\Delta_D\}$ be the closed set of all absolutely continuous distributions in $\mathcal{P}(A^i,\mathcal{B}^i)$ whose densities are bounded above and below by D and are Lipschitz continuous with constant D. Appendix A.1 of Oechssler and Riedel [39] demonstrates that the topology on Δ_D^i under the total variation norm is identical to the topology on the set of L^1 density functions on A^i under the standard L^1 norm. Thus Δ_D^i is compact under the total variation norm if and only if the set of densities functions of measures in Δ_D^i is compact under the L^1 norm. This set of density functions of elements of Δ_D^i is equicontinuous, since all of the densities have Lipschitz constant D. The Kolmogorov–Riesz compactness theorem (see [26,19]) therefore implies that this set of densities is compact. Hence Δ_D^i is compact under $\|\cdot\|_{TV_i}$. Finally, Δ_D is compact as it is the Cartesian product of N compact sets.

Now, by Lemmas C.1 and C.2, when the rewards are bounded and Lipschitz continuous $L_\eta(\cdot):\Delta\to\Delta_D$ is continuous with respect to the strong topology on Δ and Δ is a convex set. Therefore, when the rewards are bounded and Lipschitz continuous, the map $L_\eta(\cdot):\Delta\to\Delta_D$ is a continuous map (with respect to the total variation norm) from a convex set to a compact subset of Σ. Applying Schauder's fixed point theorem completes the proof. □

The infinite dimensional Picard–Lindelöf theorem (see [46]) gives a method for verifying (A4), the uniqueness of the solution to the differential equation on $\mathcal{M}(X,\mathcal{B})$ corresponding to (3.2). This can be directly applied if $F(\cdot):\mathcal{M}(X,\mathcal{B})\to\mathcal{M}(X,\mathcal{B})$ is Lipschitz continuous with respect to $\|\cdot\|_{\Sigma BL^*}$ and $\|F(\cdot)\|_{\Sigma BL^*}<\infty$, however, it is possible to show that (A4) holds even when this is not the case, as is demonstrated in the following proof.

Proof of Proposition 4.3. This proof extends the proof of Lahkar and Riedel [31, Theorem 5.2] to the N-player case and has a similar form to proofs from [39] and [23]. Throughout this proof $B:=(B^1,\dots,B^N)$, where $B^i\in\mathcal{B}^i$ for $i=1,\dots,N$. Firstly, since $L_\eta(\underline{\pi})(B)-\underline{\pi}(B)$ is neither bounded nor globally Lipschitz continuous on the space of signed measures the logit best response dynamics is studied using a secondary dynamical system. This alternative differential equation is bounded and Lipschitz continuous on Σ and coincides with (4.1) on Δ. Since Δ is forward invariant under the logit best response dynamics, showing that this alternative system has a unique solution flow implies the same result for (4.1) in the case of interest (for any initial joint strategy in Δ). For $\underline{\pi}\in\Sigma$, denote the mean field of the logit best response by

$$F(\underline{\pi})(B) := L_\eta(\underline{\pi})(B)-\underline{\pi}(B), \tag{C.1}$$

and let

$$F^i(\underline{\pi})\big(B^i\big) := L_\eta^i\big(\pi^{-i}\big)\big(B^i\big)-\pi^i\big(B^i\big).$$

Similarly, for $\underline{\pi}\in\Sigma$ let

$$\tilde{F}(\underline{\pi})(B) := \max\big\{\big(2-\|\underline{\pi}\|_{\Sigma TV}\big),0\big\}F(\underline{\pi})(B)$$

Now the logit best response is studied using the secondary differential equation on Σ,

$$\dot{\underline{\pi}}(B) = \tilde{F}(\underline{\pi})(B). \tag{C.2}$$

If $\tilde{F}(\cdot)$ is shown to be Lipschitz continuous then the infinite dimensional Picard–Lindelöf theorem [46, Corollary 3.9] can be used to prove the existence of a unique solution to (C.2). So the aim is to show that for all $\underline{\pi}, \underline{\rho} \in \Sigma$ there exists a $K > 0$ such that,

$$\|\tilde{F}(\underline{\pi}) - \tilde{F}(\underline{\rho})\|_{\Sigma TV} \leq K\|\underline{\pi} - \underline{\rho}\|_{\Sigma TV}. \tag{C.3}$$

As in the similar proof by Oechssler and Riedel [39], (C.3) is proved for three different cases.

In the first case consider $\|\underline{\pi}\|_{\Sigma TV}, \|\underline{\rho}\|_{\Sigma TV} \geq 2$ then $\tilde{F}(\underline{\pi}) = \tilde{F}(\underline{\rho}) = 0$ and (C.3) is trivially satisfied.

In the second case assume $\|\underline{\pi}\|_{\Sigma TV} \geq 2 > \|\underline{\rho}\|_{\Sigma TV}$. Since $\tilde{F}(\underline{\pi}) = 0$, the left-hand side of (C.3) becomes

$$\|\tilde{F}(\underline{\rho})\|_{\Sigma TV} = \big(2 - \|\underline{\rho}\|_{\Sigma TV}\big)\|F(\underline{\rho})\|_{\Sigma TV}. \tag{C.4}$$

Now it is straightforward to show that

$$\|F^i(\underline{\rho})\|_{TV_i} \leq 1 + \|\rho^i\|_{TV_i} \leq 3. \tag{C.5}$$

Where the first inequality follows because $\|L^i_\eta(\rho)\|_{TV_i} = 1$ and the second follows from the assumption $\|\underline{\rho}\|_{\Sigma TV} < 2$. Combining (C.4) and (C.5) gives

$$\begin{aligned}\|\tilde{F}(\underline{\pi}) - \tilde{F}(\underline{\rho})\|_{\Sigma TV} &\leq 3\big(2 - \|\underline{\rho}\|_{\Sigma TV}\big) \leq 3\big(\|\underline{\pi}\|_{\Sigma TV} - \|\underline{\rho}\|_{\Sigma TV}\big)\\ &\leq 3\|\underline{\pi} - \underline{\rho}\|_{\Sigma TV}.\end{aligned}$$

In the third case assume $\|\underline{\pi}\|_{\Sigma TV}, \|\underline{\rho}\|_{\Sigma TV} < 2$ then the left-hand side of (C.3) becomes

$$\begin{aligned}&\|\tilde{F}(\underline{\pi}) - \tilde{F}(\underline{\rho})\|_{\Sigma TV}\\ &\quad= \big\|\big(2 - \|\underline{\pi}\|_{\Sigma TV}\big)\|F(\underline{\pi})\|_{\Sigma TV} - \big(2 - \|\underline{\rho}\|_{\Sigma TV}\big)\|F(\underline{\rho})\|_{\Sigma TV}\big\|_{\Sigma TV}\\ &\quad\leq \big(2 - \|\underline{\pi}\|_{\Sigma TV}\big)\|F(\underline{\pi}) - F(\underline{\rho})\|_{\Sigma TV}\\ &\qquad+ \|F(\underline{\rho})\|_{\Sigma TV}\big|\|\underline{\pi}\|_{\Sigma TV} - \|\underline{\rho}\|_{\Sigma TV}\big|\\ &\quad\leq 2\|F(\underline{\pi}) - F(\underline{\rho})\|_{\Sigma TV} + 3\|\underline{\pi} - \underline{\rho}\|_{\Sigma TV},\end{aligned} \tag{C.6}$$

where the final inequality holds because $2 - \|\underline{\pi}\|_{\Sigma TV} < 2$ and using (C.5). Showing that $F(\cdot)$ is Lipschitz continuous for $\|\underline{\pi}\|_{\Sigma TV}, \|\underline{\rho}\|_{\Sigma TV} < 2$ will prove the claim for the third case. Clearly,

$$\|F^i(\underline{\pi}) - F^i(\underline{\rho})\|_{TV_i} \leq \|L^i_\eta(\pi^{-i}) - L^i_\eta(\rho^{-i})\|_{TV_i} + \|\pi^i - \rho^i\|_{TV_i}. \tag{C.7}$$

By Lemma C.2 $L^i_\eta(\cdot)$ is also Lipschitz continuous when the rewards are bounded and so by combining this with (C.6) and (C.7) implies that for some $K > 3$ (C.3) must hold.

Therefore using the result of Zeidler [46, Corollary 3.9] the secondary differential equation, (C.2), has a unique solution flow. Since Δ is forward invariant under (C.2) and coincides with (4.1) for all points in Δ it must be the logit best response differential equation also has a unique solution flow on Δ.

Continuity of the semiflow $\Phi_t(\underline{\pi}_0)$ follows by applying Gronwall's lemma (see Propositions 3.10 and 3.11 [46]) in the same manner as in [31, Theorem 5.2] or the similar proof from Hofbauer et al. [23]. This implies that for some $k > 0$ if $\underline{\pi}(0), \underline{\rho}(0) \in \Delta$ then

$$\|\underline{\pi}(t) - \underline{\rho}(t)\|_{\Sigma TV} \leq e^{kt}\|\underline{\pi}(0) - \underline{\rho}(0)\|_{\Sigma TV}. \qquad \square$$

Appendix D. Convergence of the logit best response dynamic

Propositions 4.2 and 4.3 show that the logit best response differential equation satisfies the basic properties required to allow further study of the dynamical system associated with this differential equation. The obvious next step is to explore the behaviour of this well defined dynamical system. In particular, under what conditions convergence to a fixed point should be expected. Therefore it is desirable to produce a global convergence result for the logit best response differential equation to a set of equilibrium strategies.

As in the discrete action case the logit best response can be expressed as a maximisation of an expected reward function and a perturbation term using the entropy, $v^i(\cdot)$ as presented in (2.2). In order to prove the global convergence of the logit best response differential equation the following intermediate lemma is required.

Lemma D.1. *If $\underline{\pi}, \underline{\rho} \in \Delta_D$ with associated densities $\underline{p}, \underline{q}$ respectively then for $i = 1, \ldots, N$,*

$$\int_{A^i} \left[r^i\left(x^i, \pi^{-i}\right) - \eta \log\left(q^i\left(x^i\right)\right)\right]\dot{\pi}^i\left(\mathrm{d}x^i\right) = \eta \int_{A^i} \left[\log\left(l^i_\eta\left(\pi^{-i}\right)\left(x^i\right)\right) - \log\left(q^i\left(x^i\right)\right)\right]\dot{\pi}^i\left(\mathrm{d}x^i\right).$$

Proof. Rearranging the definition of $l^i_\eta(\pi^{-i})$ from (2.4) gives,

$$r^i\left(x^i, \pi^{-i}\right) = \eta \log\left(l^i_\eta\left(\pi^{-i}\right)\left(x^i\right)\right) + \eta \log\left(\int_{A^i} \exp\left\{\eta^{-1} r^i\left(y^i, \pi^{-i}\right)\right\} \mathrm{d}y^i\right).$$

It then follows by substitution that

$$\begin{aligned}\int_{A^i} &\left[r^i\left(x^i, \pi^{-i}\right) - \eta \log\left(q^i\left(x^i\right)\right)\right]\dot{\pi}^i\left(\mathrm{d}x^i\right) \\ &= \int_{A^i} \eta \log\left(l^i_\eta\left(\pi^{-i}\right)\left(x^i\right)\right)\dot{\pi}^i\left(\mathrm{d}x^i\right) \\ &\quad + \int_{A^i} \left[\eta \log\left(\int_{A^i} \exp\left\{\eta^{-1} r^i\left(y^i, \pi^{-i}\right)\right\} \mathrm{d}y^i\right) - \eta \log\left(q^i\left(x^i\right)\right)\right]\dot{\pi}^i\left(\mathrm{d}x^i\right).\end{aligned}$$

Consequently, by rearranging these terms and noting that $\int_{A^i} \dot{\pi}^i(\mathrm{d}x^i) = 0$

$$\int_{A^i} \left[r^i\left(x^i, \pi^{-i}\right) - \eta \log\left(q^i\left(x^i\right)\right)\right]\dot{\pi}^i\left(\mathrm{d}x^i\right) = \eta \int_{A^i} \left[\log\left(l^i_\eta\left(\pi^{-i}\right)\left(x^i\right)\right) - \log\left(q^i\left(x^i\right)\right)\right]\dot{\pi}^i\left(\mathrm{d}x^i\right)$$

$$+\eta\log\left(\int_{A^i}\exp\{\eta^{-1}r^i(y^i,\pi^{-i})\}\,\mathrm{d}y^i\right)\int_{A^i}\dot{\pi}^i(\mathrm{d}x^i)$$
$$=\eta\int_{A^i}\left[\log\left(l^i_\eta(\pi^{-i})(x^i)\right)-\log\left(q^i(x^i)\right)\right]\dot{\pi}^i(\mathrm{d}x^i). \qquad \square$$

Two final concepts are needed before the global convergence result can be presented:

Definition D.2. On a Banach spaces X, Y the *Gateaux derivative* of a function $F(\cdot): X \to Y$ at the point $x \in X$ in the direction of $u \in X$ is defined as

$$DF(x)u := \lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\left[F(x+\varepsilon u)-F(x)\right].$$

The Gateaux derivative extends the definition of a directional derivative in Euclidean space to the more general Banach space framework. If X is a separable Banach space, Y is such that every Lipschitz continuous map $f(\cdot): \mathbb{R} \to Y$ is differential almost everywhere and the map $F(\cdot): X \to Y$ is Lipschitz continuous, then the Gateaux derivative of $F(\cdot)$ exists almost everywhere [35]. The Gateaux derivative satisfies the usual rules associated with differentiation, such as the product rule and the chain rule (see [46, page 138]).

Definition D.3. Let $\Lambda \subset M$ be a compact invariant set for the semiflow Φ. A continuous function $V(\cdot): M \to \mathbb{R}$ is called a *Lyapunov function* for Λ if for all $t \in \mathbb{R}^+$ the function $V(\Phi_t(\theta))$ is constant for all $\theta \in \Lambda$ and strictly decreasing for all $\theta \in M\backslash\Lambda$.

Recall that

$$\begin{aligned}\Delta_D := \big\{\underline{\pi} \in \Delta : \forall i = 1,\dots,N, \pi^i &\text{ is absolutely continuous with density } p^i \text{ such that}\\ &D^{-1} \le p^i(x^i) \le D, \text{ for all } x^i \in A^i \text{ and } p^i(\cdot) \text{ is Lipschitz continuous}\\ &\text{with constant } D\big\}.\end{aligned}$$

For the remainder of the article, assume that D is large enough such that $L_\eta(\pi) \in \Delta_D$ for all $\pi \in \Delta$ (see Lemma C.1). The following result extends the Lyapunov function used by Lahkar and Riedel [31] to the two-player zero-sum case and confirms that this is sufficient to obtain a convergence result on Δ_D.

Proposition D.4. *If a continuous action set game is a two-player zero-sum game with bounded and Lipschitz continuous rewards then, with respect to the strong topology on Δ_D,*

$$V_\eta(\underline{\pi}) := \sum_{i=1}^{2}\left[r^i\left(L^i_\eta(\pi^{-i}),\pi^{-i}\right)-\eta v^i\left(L^i_\eta(\pi^{-i})\right)\right]-\sum_{i=1}^{2}\left[r^i\left(\pi^i,\pi^{-i}\right)-\eta v^i\left(\pi^i\right)\right],$$

is a Lyapunov function for $\mathcal{LE}_\eta$ under the logit best response dynamics on Δ_D.

Proof. This proof follows a similar form to that of Lahkar and Riedel [31, Theorem 7.3]. Fix $\eta > 0$. Notice that for all $\underline{\pi} \in \Delta_D$, $V_\eta(\underline{\pi}) \ge 0$. Also notice that by (2.4)

$$\begin{aligned}\eta\nu^i\big(L^i_\eta(\pi^{-i})\big) &= \eta\int_{A^i}\log\big(l^i_\eta(\pi^{-i})(x^i)\big)l^i_\eta(\pi^{-i})(x^i)\,\mathrm{d}x^i\\ &= \int_{A^i} r^i(x^i,\pi^{-i})l^i_\eta(\pi^{-i})(x^i)\,\mathrm{d}x^i - \eta\log\bigg(\int_{A^i}\exp\{\eta^{-1}r^i(x^i,\pi^{-i})\}\,\mathrm{d}x^i\bigg)\\ &= r^i\big(L^i_\eta(\pi^{-i}),\pi^{-i}\big) - \eta\log\bigg(\int_{A^i}\exp\{\eta^{-1}r^i(x^i,\pi^{-i})\}\,\mathrm{d}x^i\bigg).\end{aligned}$$

From this it follows that

$$V_\eta(\underline{\pi}) := \sum_{i=1}^{2}\eta\log\bigg(\int_{A^i}\exp\{\eta^{-1}r^i(x^i,\pi^{-i})\}\,\mathrm{d}x^i\bigg) - r^i(\pi^i,\pi^{-i}) + \eta\nu^i(\pi^i).$$

The Gateaux derivation (see Definition D.2) is used to check that $V_\eta(\cdot)$ is a decreasing function. It follows that

$$\begin{aligned}\dot V_\eta(\underline{\pi}) &= DV_\eta(\underline{\pi})\underline{\dot\pi}\\ &= \sum_{i=1}^{2}\eta D\log\bigg(\int_{A^i}\exp\{\eta^{-1}r^i(x,\pi^{-i})\}\,\mathrm{d}x\bigg)\dot\pi^{-i} + \eta D\nu^i(\pi^i)\dot\pi^i.\end{aligned}\tag{D.1}$$

Each of these derivatives are evaluated separately. Firstly, note that $D\pi^i(x)\dot\pi^i = \dot\pi^i(x)$ [31]. Using the chain rule gives

$$\begin{aligned}&\eta D\log\bigg(\int_{A^i}\exp\{\eta^{-1}r^i(x^i,\pi^{-i})\}\,\mathrm{d}x^i\bigg)\dot\pi^{-i}\\ &= \frac{\eta}{\int_{A^i}\exp\{\eta^{-1}r^i(y^i,\pi^{-i})\}\,\mathrm{d}y^i}\bigg[\int_{A^i} D\exp\{\eta^{-1}r^i(x^i,\pi^{-i})\}\dot\pi^{-i}\,\mathrm{d}x^i\bigg]\\ &= \frac{\eta\int_{A^i} D\eta^{-1}r^i(x^i,\pi^{-i})\dot\pi^{-i}\exp\{\eta^{-1}r^i(x^i,\pi^{-i})\}\,\mathrm{d}x^i}{\int_{A^i}\exp\{\eta^{-1}r^i(y^i,\pi^{-i})\}\,\mathrm{d}y^i}\\ &= \int_{A^i} Dr^i(x^i,\pi^{-i})\dot\pi^{-i}\bigg(\frac{\exp\{\eta^{-1}r^i(x^i,\pi^{-i})\}}{\int_{A^i}\exp\{\eta^{-1}r^i(y^i,\pi^{-i})\}\,\mathrm{d}y^i}\bigg)\,\mathrm{d}x^i\\ &= r^i\big(L^i_\eta(\pi^{-i}),\dot\pi^{-i}\big).\end{aligned}\tag{D.2}$$

Since $\underline{\pi}\in\Delta_D$ it has an associated joint density $\underline{p}$. Using the product rule gives that the second term of (D.1) is

$$\begin{aligned}\eta D\nu^i(\pi^i)\dot\pi^i &= \eta D\int_{A^i}\log\big(p^i(x)\big)\pi^i(\mathrm{d}x)\dot\pi^i\\ &= \eta\int_{A^i} D\log\big(p^i(x)\big)\dot\pi^i\pi^i(\mathrm{d}x) + \eta\int_{A^i}\log\big(p^i(x)\big)D\pi^i(x)\dot\pi^i\,\mathrm{d}x\end{aligned}$$

$$= \eta \int_{A^i} \frac{1}{p^i(x)} D\pi^i(x)\dot{\pi}^i p^i(x)\,\mathrm{d}x + \eta \int_{A^i} \log\bigl(p^i(x)\bigr)\dot{\pi}^i(\mathrm{d}x)$$
$$= \eta \int_{A^i} \dot{\pi}^i(\mathrm{d}x) + \eta \int_{A^i} \log\bigl(p^i(x)\bigr)\dot{\pi}^i(\mathrm{d}x)$$
$$= \eta \int_{A^i} \log\bigl(p^i(x)\bigr)\dot{\pi}^i(\mathrm{d}x). \tag{D.3}$$

Where the last line holds because $\int_{A^i} \dot{\pi}^i(\mathrm{d}x^i) = 0$. Now substituting (D.2) and (D.3) into (D.1) gives,

$$\dot{V}_\eta(\underline{\pi}) = \sum_{i=1}^{2} r^i\bigl(L^i_\eta(\pi^{-i}), \dot{\pi}^{-i}\bigr) + \eta \int_{A^i} \log\bigl(p^i(x)\bigr)\dot{\pi}^i(\mathrm{d}x)$$
$$= -\sum_{i=1}^{2} \left[r^i\bigl(\dot{\pi}^i, \pi^{-i}\bigr) - \eta \int_{A^i} \log\bigl(p^i(x)\bigr)\dot{\pi}^i(\mathrm{d}x) \right]$$
$$= -\sum_{i=1}^{2} \int_{A^i} \bigl[r^i\bigl(x, \pi^{-i}\bigr) - \eta \log\bigl(p^i(x)\bigr)\bigr]\dot{\pi}^i(\mathrm{d}x).$$

The second line follows from Definition 2.1 because the game is two-player zero-sum. Applying Lemma D.1 gives that

$$\dot{V}_\eta(\underline{\pi}) = -\eta \sum_{i=1}^{2} \int_{A^i} \bigl[\log\bigl(l^i_\eta\bigl(\pi^{-i}\bigr)\bigl(x^i\bigr)\bigr) - \log\bigl(p^i\bigl(x^i\bigr)\bigr)\bigr]\dot{\pi}^i\bigl(\mathrm{d}x^i\bigr)$$
$$= -\eta \sum_{i=1}^{2} \int_{A^i} \bigl[\log\bigl(l^i_\eta\bigl(\pi^{-i}\bigr)\bigl(x^i\bigr)\bigr) - \log\bigl(p^i\bigl(x^i\bigr)\bigr)\bigr]\bigl[l^i_\eta\bigl(\pi^{-i}\bigr)\bigl(x^i\bigr) - p^i\bigl(x^i\bigr)\bigr]\,\mathrm{d}x^i$$
$$= -\eta \sum_{i=1}^{2} \bigl[v^i\bigl(L^i_\eta\bigl(\pi^{-i}\bigr)\bigr) + v^i\bigl(\pi^{-i}\bigr)\bigr]$$
$$+ \eta \sum_{i=1}^{2} \left[\int_{A^i} \log\bigl(l^i_\eta\bigl(\pi^{-i}\bigr)\bigl(x^i\bigr)\bigr)\pi^i\bigl(\mathrm{d}x^i\bigr) + \int_{A^i} \log\bigl(p^i\bigl(x^i\bigr)\bigr) l^i_\eta\bigl(\pi^{-i}\bigr)\bigl(x^i\bigr)\,\mathrm{d}x^i \right]. \tag{D.4}$$

Now by rearranging the definition of $V_\eta(\underline{\pi})$

$$-\eta \sum_{i=1}^{2} v^i\bigl(\pi^i\bigr) = -V_\eta(\underline{\pi}) + \sum_{i=1}^{2} \bigl[r^i\bigl(\dot{\pi}^i, \pi^{-i}\bigr) - \eta v^i\bigl(L^i_\eta\bigl(\pi^{-i}\bigr)\bigr)\bigr] \tag{D.5}$$

It follows by substituting (D.5) into (D.4) that

$$\dot{V}_\eta(\underline{\pi}) = -V_\eta(\underline{\pi}) + \sum_{i=1}^{2} \bigl[r^i\bigl(\dot{\pi}^i, \pi^{-i}\bigr) - 2\eta v^i\bigl(L^i_\eta\bigl(\pi^{-i}\bigr)\bigr)\bigr]$$

$$+\eta\Bigg[\sum_{i=1}^{2}\int_{A^i}\log\big(l^i_\eta\big(\pi^{-i}\big)\big(x^i\big)\big)\pi^i\big(\mathrm{d}x^i\big)+\int_{A^i}\log\big(p^i\big(x^i\big)\big)l^i_\eta\big(\pi^{-i}\big)\big(x^i\big)\,\mathrm{d}x^i\Bigg]$$

$$=-V_\eta(\underline{\pi})+\sum_{i=1}^{2}\Bigg[r^i\big(\dot{\pi}^i,\pi^{-i}\big)-\eta\int_{A^i}\log\big(l^i_\eta\big(\pi^{-i}\big)\big(x^i\big)\big)\dot{\pi}^i\big(\mathrm{d}x^i\big)\Bigg]$$

$$-\eta\sum_{i=1}^{2}\int_{A^i}\big[\log\big(l^i_\eta\big(\pi^{-i}\big)\big(x^i\big)\big)-\log\big(p^i\big(x^i\big)\big)\big]l^i_\eta\big(\pi^{-i}\big)\big(x^i\big)\,\mathrm{d}x^i.$$

Now by first applying Lemma D.1 (with $q^i = l^i_\eta(\pi^{-i})$) to the first sum and then cancelling gives

$$\dot{V}_\eta(\underline{\pi})=-V_\eta(\underline{\pi})-\eta\sum_{i=1}^{2}\int_{A^i}\big[\log\big(l^i_\eta\big(\pi^{-i}\big)\big(x^i\big)\big)-\log\big(p^i\big(x^i\big)\big)\big]l^i_\eta\big(\pi^{-i}\big)\big(x^i\big)\,\mathrm{d}x$$

$$=-V_\eta(\underline{\pi})-\eta\sum_{i=1}^{2}\int_{A^i}\log\bigg(\frac{l^i_\eta(\pi^{-i})(x^i)}{p^i(x^i)}\bigg)l^i_\eta\big(\pi^{-i}\big)\big(x^i\big)\,\mathrm{d}x^i.$$

The remaining summation is the negative Kullback–Liebler divergence measure of p^i from $l^i_\eta(\pi^{-i})$, which must always be negative. This implies

$$\dot{V}_\eta(\underline{\pi})\le -V_\eta(\underline{\pi}). \tag{D.6}$$

It then follows that under the strong topology $V_\eta(\cdot)$ is a Lyapunov function for $\mathcal{LE}_\eta$ under the logit best response on Δ_D. □

Although we have shown the existence of a Lyapunov function on Δ_D, we must show that convergence occurs uniformly on all of Δ. Firstly, for a set $S\subset\Delta$, define an open ball under the strong topology on Δ by letting

$$B_{\Sigma TV}(S,\delta):=\Big\{\underline{\pi}\in\Delta\ :\ \inf_{\underline{\sigma}\in S}\|\underline{\sigma}-\underline{\pi}\|_{\Sigma TV}<\delta\Big\}. \tag{D.7}$$

This represents all strategies which are within δ of the set A under the total variation norm on Δ.

For $k\in\mathbb{R}$, let

$$S_k:=\big\{\underline{\pi}\in\Delta_D\ :\ V_\eta(\underline{\pi})\le k\big\} \tag{D.8}$$

be the level sets in Δ_D of the Lyapunov function $V_\eta(\cdot)$, meaning all strategies in Δ_D which make the Lyapunov function less than some threshold k. The following lemma shows that strategies with small values of $V_\eta(\underline{\pi})$ are close to the equilibrium set $\mathcal{LE}_\eta$.

Lemma D.5. *For all $\varepsilon>0$ there exists a $\kappa>0$ such that*

$$S_\kappa\subseteq B_{\Sigma TV}(\mathcal{LE}_\eta,\varepsilon).$$

Proof. Start by noting that if $\Delta_D\subseteq B_{\Sigma TV}(\mathcal{LE}_\eta,\varepsilon)$ then the result is immediate since $S_\kappa\subseteq\Delta_D$ by definition. Hence assume there exists $\underline{\pi}\in\Delta_D$ such that $\underline{\pi}\notin B_{\Sigma TV}(\mathcal{LE}_\eta,\varepsilon)$. Using the compactness of Δ_D under the strong topology on Δ (see the proof of Proposition 4.2 on page 203), there exists a $\kappa>0$ such that

$$\inf\left\{V_\eta(\underline{\pi}) : \underline{\pi} \in \Delta_D \cap B_{\Sigma TV}(\mathcal{LE}_\eta, \varepsilon)^c\right\} = 2\kappa. \tag{D.9}$$

Thus

$$S_\kappa \subseteq \Delta_D \cap B_{\Sigma TV}(\mathcal{LE}_\eta, \varepsilon),$$

since S_κ consists of elements of Δ_D with V_η values bounded above by κ, and (D.9) ensures that any elements of Δ_D not in $B_{\Sigma TV}(\mathcal{LE}_\eta, \varepsilon)$ have V_η values at least 2κ. It follows immediately that

$$S_\kappa \subseteq B_{\Sigma TV}(\mathcal{LE}_\eta, \varepsilon). \qquad \square$$

It will also be convenient to bound the Lyapunov function uniformly across all elements of Δ_D.

Lemma D.6. *There exists a $C < \infty$ such that for all $\underline{\pi} \in \Delta_D$, $V_\eta(\underline{\pi}) \le C$.*

Proof. Firstly, notice that if $r_{\max} := \max_{i,\underline{a}} r^i(\underline{a})$ and $r_{\min} := \min_{i,\underline{a}} r^i(\underline{a})$ then it must be that for all $\underline{\pi} \in \Delta_D$,

$$l^i_\eta(\underline{\pi})(x^i) \le \frac{\exp\{\eta^{-1} r_{\max}\}}{|A^i| \exp\{\eta^{-1} r_{\min}\}} = D' < \infty.$$

Recall

$$V_\eta(\underline{\pi}) = \sum_{i=1}^{2} r^i\left(L^i_\eta(\pi^{-i}), \pi^{-i}\right) - \eta\nu^i\left(L^i_\eta(\pi^{-i})\right) + \eta\nu^i\left(\pi^i\right).$$

It follows that

$$V_\eta(\underline{\pi}) \le 2r_{\max} + \sum_{i=1}^{2} \eta\nu^i\left(L^i_\eta(\pi^{-i})\right) + \eta\nu^i\left(\pi^i\right).$$

All that remains is to show that the two entropy terms are bounded for each player. Now $p(x) \le D$ for all $x \in A^i$ means that $p(x)\log(p(x)) \le D\log(D)$ for all $x \in A^i$ and $\int_{A^i} p(x) \times \log(p(x))\, dx \le |A^i| D \log(D)$. Therefore, by letting $\bar{D} = \max\{D, D'\}$, it follows that if

$$C := 2r_{\max} + 2\eta \sum_{i=1}^{2} \left|A^i\right| \bar{D} \log(\bar{D}),$$

then for any $\underline{\pi} \in \Delta_D$, $V_\eta(\underline{\pi}) \le C$. $\square$

We are now in a position to prove the uniform global convergence of the logit best response dynamics on Δ.

Proof of Theorem 4.4. Take $\underline{\pi}(0) \in \Delta$ and let $\underline{\pi}(t)$ be the solution trajectory to the logit best response differential equation (4.1), so that

$$\dot{\underline{\pi}}(t) = L_\eta\left(\underline{\pi}(t)\right) - \underline{\pi}(t).$$

It follows that $(\dot{\overline{\underline{\pi}(t)e^t}}) = e^t L_\eta(\underline{\pi}(t))$ and subsequently that

$$\underline{\pi}(t) = e^{-t}\underline{\pi}(0) + \int_0^t e^{s-t} L_\eta\big(\underline{\pi}(s)\big)\,ds. \tag{D.10}$$

Define

$$\underline{\sigma}(\tau_1) = \frac{\int_0^{\tau_1} e^{s-\tau_1} L_\eta(\underline{\pi}(s))\,ds}{1-e^{-\tau_1}}. \tag{D.11}$$

It is immediate that $\underline{\sigma}(\tau_1) \in \Delta_D$, since $L_\eta(\underline{\pi}) \in \Delta_D$ for all $\underline{\pi}$. Furthermore, by (D.10),

$$\big\|\underline{\pi}(\tau_1) - \underline{\sigma}(\tau_1)\big\|_{\Sigma TV} < 2e^{-\tau_1}. \tag{D.12}$$

So it has been shown that at time τ_1 there exists a $\underline{\sigma}(\tau_1) \in \Delta_D$ which is close to the original trajectory $\underline{\pi}(\tau_1)$. (Note that this initial construction is unnecessary if $\underline{\pi}(0) \in \Delta_D$, but allows for a simpler proof of uniform convergence.)

Let $\underline{\sigma}(\tau_1)(\cdot) : \mathbb{R}^+ \to \Delta$ be a solution to the logit best response differential equation starting at $\underline{\sigma}(\tau_1)$ described in (D.11). From Proposition 4.3 it follows that there exists a $c > 0$ (not depending on $\underline{\pi}(0)$ or τ_1) such that for any $\tau_2 > 0$

$$\big\|\underline{\pi}(\tau_1+\tau_2) - \underline{\sigma}(\tau_1)(\tau_2)\big\|_{\Sigma TV} \le e^{c\tau_2}\big\|\underline{\pi}(\tau_1) - \underline{\sigma}(\tau_1)\big\|_{\Sigma TV}. \tag{D.13}$$

Defining $\delta(\tau_1, \tau_2) := 2e^{c\tau_2 - \tau_1}$ it follows from (D.12) and (D.13) that

$$\big\|\underline{\pi}(\tau_1+\tau_2) - \underline{\sigma}(\tau_1)(\tau_2)\big\|_{\Sigma TV} < \delta(\tau_1, \tau_2).$$

This implies that

$$\underline{\pi}(\tau_1+\tau_2) \in B_{\Sigma TV}\big(\underline{\sigma}(\tau_1)(\tau_2), \delta(\tau_1, \tau_2)\big). \tag{D.14}$$

Now note that, since $\underline{\sigma}(\tau_1) \in \Delta_D$, Lemma D.6 shows that there exists a $C < \infty$, that does not depend on $\underline{\pi}(0)$, such that $V_\eta(\underline{\sigma}(\tau_1)) \le C$. By (D.6), we see that

$$V_\eta\big(\underline{\sigma}(\tau_1)(\tau_2)\big) \le Ce^{-\tau_2}.$$

This in turn implies that

$$\underline{\sigma}(\tau_1)(\tau_2) \in S_{Ce^{-\tau_2}}, \tag{D.15}$$

where $S_{Ce^{-\tau_2}}$ is the level set of the Lyapunov function defined in (D.8).

Given arbitrary $\varepsilon > 0$, let τ_2 be sufficiently large that $S_{Ce^{-\tau_2}} \subseteq B_{\Sigma TV}(\mathcal{LE}_\eta, \varepsilon/2)$, which is possible by Lemma D.5. By (D.15) we have that

$$\underline{\sigma}(\tau_1)(\tau_2) \in B_{\Sigma TV}\left(\mathcal{LE}_\eta, \frac{\varepsilon}{2}\right) \tag{D.16}$$

for all τ_1. Then choose τ_1 sufficiently large such that $\delta(\tau_1, \tau_2) < \varepsilon/2$. By (D.14) we have that

$$\underline{\pi}(\tau_1+\tau_2) \in B_{\Sigma TV}\left(\underline{\sigma}(\tau_1)(\tau_2), \frac{\varepsilon}{2}\right). \tag{D.17}$$

Hence by (D.16), (D.17) and the triangle inequality we see that

$$\underline{\pi}(\tau_1+\tau_2) \in B_{\Sigma TV}(\mathcal{LE}_\eta, \varepsilon).$$

This shows the uniform convergence of trajectories starting in Δ to the equilibrium set $\mathcal{LE}_\eta$. □

References

[1] M. Benaïm, Dynamics of stochastic approximation algorithms, Semin. Probab. Lect. Notes 1709 (1999) 1–68.
[2] M. Benaïm, M. Hirsch, Mixed equilibria and dynamical systems arising from fictitious play in perturbed games, Games Econ. Behav. 29 (1999) 36–72.
[3] M. Benaïm, J. Hofbauer, S. Sorin, Stochastic approximations and differential inclusions, SIAM J. Control Optim. 44 (2003) 328–348.
[4] E. Berger, Asymptotic behaviour of a class of stochastic approximation procedures, Probab. Theory Relat. Fields 71 (1986) 517–552.
[5] P. Billingsley, Convergence of Probability Measures, John Wiley & Sons, New York, 1999.
[6] T. Börgers, R. Sarin, Learning through reinforcement and replicator dynamics, J. Econ. Theory 77 (1997) 1–14.
[7] V. Borkar, Stochastic approximation with two time scales, Syst. Control Lett. 29 (1997) 291–294.
[8] V. Borkar, Asynchronous stochastic approximations, SIAM J. Control Optim. 36 (1998) 840–851.
[9] V. Borkar, Stochastic Approximation: A Dynamical Systems Viewpoint, Cambridge University Press, 2008.
[10] Z. Brzeźniak, On stochastic convolution in Banach spaces and applications, Stoch. Stoch. Rep. 61 (1997) 245–295.
[11] X. Chen, H. White, Nonparametric adaptive learning with feedback, J. Econ. Theory 82 (1998) 190–222.
[12] R. Cressman, Stability of the replicator equation with continuous strategy space, Math. Soc. Sci. 50 (2005) 127–147.
[13] R. Cressman, J. Hofbauer, F. Riedel, Stability of the replicator equation for a single species with a multi-dimensional continuous trait space, J. Theor. Biol. 239 (2006) 273–288.
[14] J. Dippon, H. Walk, The averaged Robbins–Monro method for linear problems in a Banach space, J. Theor. Probab. 19 (2006) 166–189.
[15] R. Dudley, Convergence of Baire measures, Stud. Math. 27 (1966) 251–268.
[16] G. Ellison, D. Fudenberg, Learning puified mixed equilibria, J. Econ. Theory 90 (2000) 84–115.
[17] D. Fudenberg, D. Kreps, Learning mixed equilibria, Games Econ. Behav. 5 (1993) 320–367.
[18] D. Fudenberg, D. Levine, Theory of Learning in Games, MIT Press, Cambridge, 1998.
[19] H. Hanche-Olsen, H. Holden, The Kolmogorov–Riesz compactness theorem, Expo. Math. 28 (2010) 385–394.
[20] J. Hofbauer, E. Hopkins, Learning in perturbed asymmetric games, Games Econ. Behav. 52 (2005) 133–152.
[21] J. Hofbauer, W. Sandholm, On the global convergence of stochastic fictitious play, Econometrica 70 (2002) 2265–2294.
[22] J. Hofbauer, S. Sorin, Best response dynamics for continuous zero-sum games, Discrete Contin. Dyn. Syst. 6 (2006) 215–224.
[23] J. Hofbauer, J. Oechssler, F. Riedel, Brown–von Neumann–Nash dynamics: the continuous strategy case, Games Econ. Behav. 65 (2009) 406–429.
[24] P. Huber, Robust Statistics, John Wiley & Sons, 1981.
[25] V. Istrăţescu, Fixed Point Theory: An Introduction, Springer-Verlag, 1981.
[26] A. Kolmogorov, Über kompaktheit der funktionenmengen bei der konvergenz im mittel, Nachr. Ges. Wiss. Gött. 9 (1931) 60–63.
[27] V. Koval, Rate of convergence of stochastic approximation procedures in a Banach space, Cybern. Syst. Anal. 34 (1998) 386–394.
[28] H. Kushner, D. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems, Springer-Verlag, 1978.
[29] H. Kushner, G. Yin, Asymptotic properties of distributed and communicating stochastic approximation algorithms, SIAM J. Control Optim. 25 (1987) 1266–1290.
[30] H. Kushner, G. Yin, Stochastic approximation algorithms for parallel and distributed processing, Stochastics 22 (1987) 219–250.
[31] R. Lahkar, F. Riedel, The continuous logit dynamic and price dispersion, preprint, 2013.
[32] R. Larsen, Functional Analysis, Marcel Dekker, Inc., 1973.
[33] M. Ledoux, M. Talagrand, Probability in Banach Spaces, Springer-Verlag, 1991.
[34] D. Leslie, Reinforcement learning in games, Ph.D. thesis, University of Bristol, 2003.
[35] J. Lindenstrauss, D. Preiss, J. Tišer, Fréchet differentiability of Lipschitz functions and porous sets, in: Banach Spaces, Princeton University Press, 2012.
[36] L. Ljung, Analysis of recursive stochastic algorithms, IEEE Trans. Autom. Control 22 (1977) 551–575.
[37] D. Luenberger, Optimization by Vector Space Methods, Prentice Hall, 1969.
[38] K. Narendra, M. Thathachar, Learning Automata: An Introduction, John Wiley & Sons, New York, 1989.
[39] J. Oechssler, F. Riedel, Evolutionary dynamics on infinite strategy spaces, J. Econ. Theory 17 (2001) 141–162.
[40] J. Oechssler, F. Riedel, On the dynamic foundation of evolutionary stability in continuous models, J. Econ. Theory 107 (2002) 223–252.

[41] R. Seymour, Dynamics for infinite dimensional games, University College London, mimeo, 2002.
[42] A. Shiryaev, Probability, Springer, 1995.
[43] A. Shwartz, N. Berman, Abstract stochastic approximations and applications, Stoch. Process. Appl. 31 (1989) 133–149.
[44] H. Walk, An invariance principle for the Robbins–Monro process in a Hilbert space, Probab. Theory Relat. Fields 39 (1977) 135–150.
[45] H. Walk, L. Zsidó, Convergence of Robbins–Monro method for linear problems in Banach space, J. Math. Anal. Appl. 139 (1989) 152–177.
[46] E. Zeidler, Nonlinear Functional Analysis and Its Applications, Springer-Verlag, 1986.