Myopic Best-Response Learning in Large-Scale Games

Submitted in partial fulfillment of the requirements for
the degree of

Doctor of Philosophy
in
Electrical and Computer Engineering

Brian Woodbury Swenson

M.S. in Mathematics, Carnegie Mellon University
B.S. in Electrical Engineering, Brigham Young University

Carnegie Mellon University
Pittsburgh, PA

May 2017

Acknowledgements

I am deeply indebted to many who have contributed to this work in a variety of forms, whether intellectual, administrative, emotional, or edible. First and foremost, I must acknowledge the invaluable contributions of my advisers Soummya Kar and João Xavier. Their counsel, guidance, and vast knowledge base has helped me work through innumerable difficult problems. Of no less importance, their constant enthusiasm and support has been a perpetual source of inspiration that has always blown my ship out of the doldrums and made the pursuit of a PhD an immensely enjoyable experience.

I am grateful to the National Science Foundation (NSF grants CCF 1513936 and ECCS 1305979) and to the Fundação para a Ciência e a Tecnologia (FCT scholarship SFRH/BD/51920/2012) for funding this research. I would like to thank my committee members, Soummya Kar, Rohit Negi, Ariel Proccacia, Jeff Shamma, and Mario Figueiredo for their valuable comments, critiques, and insights. I am tremendously grateful to my collaborators, David Leslie, Ceyhun Eksin, Alejandro Ribeiro, and Ryan Murray. David was instrumental in helping us prove the robustness result in Chapter 4. We worked closely with Ceyhun and Alejandro on the study of inertial-based MBR dynamics in Chapter 8. And we worked closely with Ryan on the study of convergence properties of FP in potential games in Chapters 9 and 10. This dissertation would not be possible without their help.

To those at CMU and IST who provided support, friendship, and maybe a little too much good conversation—you are too numerous to list, but you know who you are, and thanks! I am particularly grateful to Carol Patterson, who habitually exceeds the call of duty when aiding work-laden graduate students. This journey started long ago, and a proper acknowledgement would feel incomplete without an appreciative and wistful nod to Nate, Pappy, and the BYU pool tables. I've yet to find a better way to argue about physics and engineering. To my mother—thank you for patiently tolerating and guiding a recalcitrant teen. To Olivia—a PhD student couldn't ask for a more supportive wife. Thank you for everything.

Abstract

This dissertation studies multi-agent algorithms for learning Nash equilibrium strategies in games with many players. We focus our study on a set of learning dynamics in which agents seek to myopically optimize their next-stage utility given some forecast of opponent behavior; i.e., players act according to *myopic best response dynamics.* The prototypical algorithm in this class is the well-known fictitious play (FP) algorithm. FP dynamics are intuitively simple and can be seen as the "natural" learning dynamics associated with the Nash equilibrium concept. Accordingly, FP has received extensive study over the years and has been used in a variety of applications. Our contributions may be divided into two main research areas. First, we study fundamental properties of myopic best response (MBR) dynamics in large-scale games. We have three main contributions in this area. (i) We characterize the robustness of MBR dynamics to a class of perturbations common in real-world applications. (ii) We study FP dynamics in the important class of large-scale games known as potential games. We show that for almost all potential games and for almost all initial conditions, FP converges to a pure-strategy (deterministic) equilibrium. (iii) We develop tools to characterize the rate of convergence of MBR algorithms in potential games. In particular, we show that the rate of convergence of FP is "almost always" exponential in potential games.

Our second research focus concerns implementation of MBR learning dynamics in large-scale games. MBR dynamics can be shown, theoretically, to converge to equilibrium strategies in important classes of large-scale games (e.g., potential games). However, despite theoretical convergence guarantees, MBR dynamics can be extremely impractical to implement in large games due to demanding requirements in terms of computational capacity, information overhead, communication infrastructure, and global synchronization. Using the aforementioned robustness result, we study practical methods to mitigate each of these issues. We place a special emphasis on studying algorithms that may be implemented in a network-based setting, i.e., a setting in which inter-agent communication is restricted to a (possibly sparse) overlaid communication graph. Within the network-based setting, we also study the use of so-called "inertia" in MBR algorithms as a tool for learning pure-strategy NE.

Contents

Chapter 1

Introduction

1.1 Motivating Example

A game is a mathematical object that describes the strategic interaction between a group of agents. Multi-agent systems such as wireless networks [1–5], smart grid infrastructures [6–8], electric vehicle charging networks [9,10], wind farms [6,11], multi-robot systems [12], camera networks [13], mobile sensor networks [14,15], and others [16,17], can be naturally modeled using the framework of game theory. Within this context, a desirable operating condition in the multi-agent system can be modeled as an equilibrium in the associated game [18].

As a motivating example, consider the following traffic routing scenario. Suppose a group of drivers wishes to navigate a traffic grid. Each driver's vehicle is equipped with an onboard computer and is connected to neighboring vehicles in the group via an adhoc vehicular network. Each vehicle is equipped with a model of the traffic grid and knows its starting point and destination. Prior to physically engaging in the commute, the drivers wish to compute an equilibrium routing strategy. In particular, it is desired that, by repeatedly exchanging information with neighboring vehicles in the network, the group of vehicles can negotiate on a equilibrium routing strategy that can be recommend to the drivers.

The strategic problem of routing traffic can be modeled as a game by letting each vehicle be represented as a player. The set of actions for each player is the set of route choices available to the vehicle. The utility function of each player is determined by travel time of the vehicle from starting point to destination, as determined by the onboard traffic grid model. (More precisely, we let the utility be the negative of the travel time so that players prefer to maximize their utility function.) An equilibrium routing strategy defines a route choice for each vehicle that optimizes the utility (travel time) of each player (vehicle).

This particular scenario falls in the intersection of game theory [19, 20] and distributed algorithms [21,22]. In a distributed algorithm, it is desired that a group of agents connected via a sparse communication graph cooperatively compute some network-dependent quantity; e.g., the average temperature in a sensor network [23], the optimal value of some network-dependent function [24], or the likelihood of some event based on observations in a sensor network [25]. In this case, it is desired that agents compute an equilibrium routing strategy of the associated traffic game. A distributed algorithm prescribes the manner in which information is disseminated through the network and the manner in which the disseminated information is used in order to compute the desired quantity.

Returning to the traffic routing example, the challenge is to design distributed algorithms that allow the vehicles to cooperatively compute an equilibrium strategy and are practical to implement when the number of players is large. In particular, the following issues should be taken into consideration:

- The computations that must be performed by each vehicle should be simple enough to be carried out using the limited computational resources of the vehicle's onboard computer. (See Chapters 3 and 5.)
- The amount of information that must be exchanged between vehicles should remain relatively small, even when there are a large number of vehicles in the game. (See Chapter 3.)
- In this setting, it is not possible for agents to physically measure payoff information—accordingly, all information dissemination must be explicitly handled by the learning algorithm using the overlaid communication graph. (See Section 1.3.2, and Chapters 7–8.)
- The interagent communication graph should be permitted to be sparse, e.g., vehicles may only communicate with other nearby vehicles. (See Chapters 7–8.)
- Vehicles should be permitted to process information and communicate with one another in an asynchronous manner (i.e., the algorithm should be able to operate without a global clock). (See Chapter 6.)
- The rate of convergence of an algorithm should be reasonably fast. (See Chapters 3 and 10.)
- An algorithm should be robust to perturbations that may occur in real-world implementations, e.g., due to computational limitations, communication errors, and asynchrony. (See Chapter 4.)
- An algorithm should converge to a desirable equilibrium point, e.g., in some applications, a pure-strategy (deterministic) equilibrium is preferred to a mixed-strategy (probabilistic) equilibrium. (See Chapters 8–10.)

In this dissertation, we will focus on a particular class of natural learning dynamics that are induced by the Nash equilibrium concept. In Section 1.2 we introduce these learning dynamics, and in Section 1.3 we discuss the main contributions of the dissertation.

1.2 Myopic Best-Response Dynamics

Before introducing the learning dynamics studied in this dissertation, we first briefly introduce some necessary notation.

Formally, a game consists of a finite set of players $\mathcal{N} := \{1, \ldots, N\}$, a convex compact set Δ_i representing the set of strategies[1] available to each player $i \in \mathcal{N}$, and a utility function $U_i : \Delta_1 \times \cdots \times \Delta_N \to \mathbb{R}$ establishing the utility received by each player $i \in \mathcal{N}$ for any given (joint) strategy choice.

An equilibrium of a game may be defined using the notion of a *best response*. Informally, a best response for player $i \in \mathcal{N}$ is a strategy choice that maximizes the utility received by player i if the strategies of opponents are held constant. Formally, given a strategy $q_{-i} \in \prod_{j \in \mathcal{N}\setminus\{i\}} \Delta_j$, the set of best responses for player i is given by

$$\mathrm{BR}_i(q_{-i}) := \arg\max_{q_i \in \Delta_i} U(q_i, q_{-i}),$$

and given a strategy $q \in \Delta_1 \times \cdots \times \Delta_N$, the joint best response set is given by

$$\mathrm{BR}(q) := \mathrm{BR}_1(q_{-1}) \times \cdots \times \mathrm{BR}_N(q_{-N}),$$

where $q_{-i} = (q_1, \ldots, q_{i-1}, q_{i+1}, \ldots, q_N)$, $\forall i \in \mathcal{N}$. A strategy $q \in \Delta_1 \times \cdots \times \Delta_N$ is said to be a *Nash equilibrium* (NE) if $U_i(q_i, q_{-i}) \geq U_i(q_i', q_{-i}),\ \forall q_i' \in \Delta_i,\ \forall i \in \mathcal{N}$, or equivalently,

$$q \in BR(q). \tag{1.1}$$

In words, this means that at a Nash equilibrium, no player can improve their utility by unilaterally changing their strategy.

A question of fundamental interest is, how might a group of interacting agents learn an equilibrium strategy over time? In a game-theoretic learning algorithm, players are permitted to repeat-

[1]Throughout this dissertation we will assume that players have finite action spaces and we will use the symbol Δ_i to refer to the associated set of *mixed* strategies available to player $i \in \mathcal{N}$. In order to simplify the notation in the introduction, we refer to the set of mixed strategies Δ_i as the set of strategies available to player i. This is made precise in Section 2.1.

edly face off in some fixed game. Over the course of the repeated interaction, players may revise their strategies based on the history of game play. Ideally, players strategies should adapt to one another's strategy choices, converging to an equilibrium over time.

A popular learning procedure, known as fictitious play (FP), follows naturally from the Nash equilibrium concept (1.1). In particular, consider the (continuous-time) dynamical system

$$\dot{\mathbf{q}} \in \mathrm{BR}(\mathbf{q}) - \mathbf{q}. \tag{1.2}$$

By definition, a strategy is a rest point of (1.2) if and only if it is a NE (1.1). For historical reasons, the dynamical system (1.2) is generally referred to as *continuous-time fictitious play* [26–29].

In order to study the discrete-time analog of this system, assume that players are permitted to repeatedly face off in some fixed game and let $a_i(n) \in \Delta_i$ denote the "action" used by player $i \in \mathcal{N}$ in round $n \in \{1, 2, \ldots\}$ of the repeated interaction. Let $a(n) = (a_1(n), \ldots, a_N(n))$ denote the joint action. The discrete-time analog of (1.2) is given by the process $(q(n))_{n\geq 1}$ defined by the system[2]

$$a(n+1) \in BR(q(n)), \tag{1.3}$$

$$q(n) = \frac{1}{n}\sum_{s=1}^{n} a(s). \tag{1.4}$$

Note that the process $(q(n))_{n\geq 1}$ defined above may be expressed recursively as

$$q(n+1) - q(n) \in \frac{1}{n+1}\left(\mathrm{BR}(q(n)) - q(n)\right). \tag{1.5}$$

This may be seen as an Euler discretization [30] of the differential inclusion (1.2) in which the step size is given by $\frac{1}{n+1}$. As $n \to \infty$, the continuous-time interpolation of (1.5) more closely approximates a solution of (1.2). (This is made precise in [28]; see also Chapter 4 of this dissertation.) In light of this relationship, we sometimes refer to (1.2) as the *mean-field system* associated with (1.3)–(1.4)

The system (1.3)–(1.4) defines a game-theoretic learning algorithm—players repeatedly face off in some fixed game, and after each stage $n \in \{1, 2, \ldots\}$ of the repeated interaction, players choose their next-stage action $a(n+1)$ as a (myopic) best response to the current state $q(n)$. The action $a(n+1)$ is a *myopic* best response in the sense that players each player seeks to optimize the utility received in the upcoming round given the current state $q(n)$. This may be contrasted with a non-myopic learning process, in which agents seek to optimize their received utility in a long-term

[2]The initial action $a(1)$ may be chosen arbitrarily.

sense, e.g., using a discounted sum of future utility values.

The myopic best-response learning algorithm defined by (1.3)–(1.4) is historically known as (discrete-time) FP [26].[3] FP is known to converge to the set of NE in several important classes of games including two-player, zero-sum games [31, 32]; two-player, two-action games [33, 34]; generic two-player m-action games [35]; games solvable by iterated strict dominance [36]; "one-against-all" multi-player games [37]; and potential games [34, 38] (see also Section 2.5).

The FP algorithm has a variety useful properties. For example, while it is known that no (uncoupled) learning algorithm can converge in all games [39], it has been shown that FP can have good payoff performance, even in the absence of convergence [40]. As another example, the paper [41] studies approximation guarantees unique to FP. The FP algorithm has been shown to be useful in solving general extensive form games [42, 43]; in particular, it has been used to compute approximately-optimal strategies in poker [43–46] as well as large-state-space word games [47]. The work [48] showed that FP can be useful as an optimization heuristic in large-scale games. Variants of FP have been studied in a wide range of applications including traffic routing [48, 49], distributed constraint optimization problems [50], control of robotic teams [12, 51], dynamic programming [52, 53], cognitive radio [54–56], and learning in Markov decision processes [57].

In this dissertation we study a more general myopic best-response learning process, defined by the system

$$a(n+1) \in \mathrm{BR}_1(f_1(z(n))) \times \cdots \times \mathrm{BR}_N(f_N(z(n))), \tag{1.6}$$

$$z(n+1) = z(n) + \gamma(n)\big(g(a(n+1)) - z(n)\big), \tag{1.7}$$

where $a(n)$ denotes the joint action taken in stage n of the repeated interaction, $z(n) \in Z$ is the *observation state* and Z is the *observation space* (the observation state $z(n)$ represents the information available to players in stage n), $(\gamma(n))_{n\geq 1}$ is a step-size sequence, f_i, $i \in \mathcal{N}$ is a forecast function that maps the current observation state to a location in the strategy space, and g is the *observation mapping*, which maps back from the strategy space into the observation space.

Intuitively speaking, a MBR algorithm may be described as follows: Each iteration of the algorithm, each player uses the current observation state $z(n)$ to generate a forecast of the strategy that will be used by the other players in the upcoming round. The forecast of player i is denoted by $f_i(z(n))$. Each player chooses their next-stage action $a_i(n+1)$ as a myopic best response given their forecast (see (1.6)). The new action information is then recursively incorporated into the

[3] Unless otherwise stated, throughout the dissertation we will refer to the continuous-time system (1.2) as continuous-time FP and we will refer to the discrete-time system (1.3)–(1.4) simply as FP

observation state using observation mapping g (see (1.7)).

By studying the general MBR dynamics (1.6)–(1.7), we are able to derive general results applying a variety of useful learning processes, including FP (1.3)–(1.4) as a special case. In Chapter 2 we will discuss these general dynamics in more detail after rigorously introducing the notation to be used throughout the dissertation.

1.3 Contributions

The main contributions of the dissertation are broadly divided into two main research areas: (i) fundamental properties of MBR dynamics and (ii) implementation in large-scale games. We discuss both of these below.

1.3.1 Fundamental Properties

As our first research focus, we study fundamental properties of MBR dynamics. We have four broad objectives in this regard, listed below. The first objective deals with MBR dynamics generally. The next two objectives focus on understanding MBR dynamics in the important class of multi-agent games known as potential games [34] (see also Section 2.1.1). The final objective deals with fundamental stability properties of equilibria in potential games.

- **Robustness of MBR dynamics**: Broadly speaking, our goal is to understand the robustness of MBR dynamics to perturbations that may occur in practical applications. We consider a sequence of perturbations at the output of the best response in (1.6) that gradually decay as the repeated play process progresses. More precisely, we suppose that rather than choosing an action that is a precise best response in (1.6), players may choose a action that is ϵ_n-suboptimal, with $\epsilon_n \to 0$ as $n \to \infty$.

  We will study the robustness of MBR dynamics in Chapter 4. Subsequently, in Chapters 5–7 we will see that such perturbations commonly arise in practical applications. In particular, in Chapters 5–7 we will apply the robustness result to develop practical methods for mitigating the problems of high computational complexity, global synchronization, and demanding communication requirements discussed in Section 1.3.2 below.

- **Convergence of FP to pure-strategy equilibria in potential games**: Potential games are an important class of multi-agent games with many applications in engineering and economics (see Section 2.1.1). Furthermore, potential games play a fundamental role within the theory of games in the sense that every game admits a unique decomposition in terms of a potential game component, a harmonic game component, and a non-strategic component [58].

The set of NE may be subdivided into pure-strategy (deterministic) NE and mixed-strategy (probabilistic) NE. Mixed-strategy NE can be problematic for a number of reasons [59]. In engineering applications involving potential games, mixed-strategy NE can be undesirable since they are nondeterministic, have suboptimal expected utility, and do not always have clear physical meaning [60]. Consequently, in applications, preference is generally given to algorithms that are guaranteed to converge to a pure-strategy NE [60–63].

For FP learning dynamics, mixed-strategy NE are problematic at an even more fundamental level (both in and out of potential games), since they can cause non-uniqueness of solutions and make it impossible to establish convergence rate estimates [27].

It has been speculated that, for FP learning dynamics, the difficulties arising due to mixed equilibria rarely occur in practice. For example, in [12] it is noted that "it is generally believed that convergence of [FP] to a mixed (but not pure) Nash equilibrium happens rarely when [players'] utilities are not equivalent to a zero-sum game", and in [64] it is noted that "for bimatrix games it seems this difficulty (finite-time convergence to mixed equilibria and non-uniqueness of solutions) occurs only for a null set of initial conditions. But this needs a proof!". Despite such speculation, there are currently no rigorous results showing that FP behaves in this manner (i.e., only reaching mixed equilibria from a null set of initial conditions) in any particular class of games.

An important benefit of potential games is that they guarantee the existence of pure-strategy NE. Nevertheless, it is well known that FP *can* converge to a mixed-strategy equilibrium in such games. In fact, this deficiency was first noted in the paper where it was originally proven that FP converges to NE in potential games (see [34], Remark (2)).

In this dissertation, we refine the convergence result for FP in potential games in an attempt to redeem it somewhat in this regard and address the fundamental issues noted above. In particular, we study the continuous-time FP dynamics (1.2) and show that in almost all potential games and for almost all initial conditions, (continuous-time) FP converges to a pure-strategy equilibrium.[4]

We note that as byproducts of this result we get that (i) in almost all potential games, FP may only reach mixed equilibria from some null set of initial conditions, (ii) in almost all potential games and for almost all initial conditions, solutions of FP are unique, and (iii) in

[4]When we say that continuous-time FP converges to some set, we mean that solutions of (1.2) converge to the set.

almost all potential games and for almost all initial conditions, the rate of convergence of FP can be characterized (see next bullet point).

We will study convergence properties of FP in potential games in Chapter 10.

- **Rate of convergence of MBR algorithms in potential games**: Convergence rate estimates for MBR algorithms, e.g., FP, are relatively scarce [27]. In particular, there is no characterization of the rate of convergence of FP in potential games.

  In this dissertation we show that in almost all potential games and for almost all initial conditions, the rate of convergence of (continuous-time) FP is exponential. That is, given some fixed game, for almost every initial condition q_0 there exists a constant c (dependent only on the game and initial condition) such that if $(\mathbf{q}(t))_{t\geq 0}$ is a (continuous-time) FP process satisfying $\mathbf{q}(0) = q_0$, then $d(\mathbf{q}(t), NE) \leq ce^{-t}$ for all $t \geq 0$, where NE denotes the set of Nash equilibria of the game.

  We also study the rate of convergence of the empirical centroid FP (ECFP) dynamics (presented in Chapter 3). We will show that if a continuous-time ECFP path converges to a completely mixed strategy equilibrium, then it converges in finite time.

  The rate of convergence of MBR algorithms in potential games is studied in Chapter 10.

- **Non-degenerate potential games** In order to study the properties of FP in potential games, we define a class of non-degenerate potential games. We say a potential game is *non-degenerate* if the derivatives of the potential function satisfy certain non-degeneracy conditions at every equilibrium point. In any potential game satisfying these conditions, FP can be shown to be "well behaved" in the sense that FP converges generically to pure NE (see second bullet point above) and the rate of convergence of FP can be characterized (see third bullet point above). In this dissertation we show that almost all potential games are non-degenerate, and hence FP is "well behaved" in almost all potential games.

  While our non-degeneracy conditions are tailored to studying learning dynamics in games, they are also of broader general interest. In particular, an equilibrium of a potential game can be shown to be non-degenerate in the sense introduced in this dissertation if and only if it is both *quasi-strong* and *regular*, as introduced by Harsanyi [65]. These important equilibrium refinement concepts have received much study over the years [66] and equilibria satisfying these notions have been shown to possess a variety of desirable stability properties, including robustness to payoff perturbations and robustness to strategy perturbations [67]. It is known that in almost all games, all equilibria are both quasi-strong and regular [65]. However, many games of interest (including potential games) belong to measure zero subsets within the space of all games [68]. Finer results are required in order to conclude anything about the relative frequency with which regular and quasi-strong equilibria occur within such classes of games. Except for the class of bi-matrix games [69,70], little is known about the genericness of quasi-strong and regular equilibria in any special classes of games. In this dissertation, we show that in almost all potential games, all equilibria are both quasi-strong and regular.

We will study non-degenerate potential games in Chapter 9.

We note that while the study of these fundamental properties is of independent interest, it is also intimately linked with our second research focus regarding implementablity in large-scale games (see Section 1.3.2 below). The robustness result permits us to develop methods to mitigate a variety of problems encountered in large-scale implementation, including high computational complexity, demanding communication requirements, and global synchronization. The convergence result for continuous-time FP, showing "almost everywhere" convergence to pure-strategy NE in potential games, allows us to better understand the mean-field dynamics associated with FP in an important class of multi-agent games that is frequently utilized in large-scale applications (see Section 2.1.1).

The question of characterizing the convergence rate of learning algorithms is essential to understanding the practicality of such algorithms in large-scale settings. It has been shown that in many games of interest, including potential games, the number of rounds required for an algorithm to reach a NE strategy can grow exponentially as the number of players increases [71].

Understanding the rate of convergence of a learning algorithm aids practitioners in choosing learning dynamics that converge in a reasonable amount of time in a large-scale game of interest. Furthermore, characterizing convergence rates is an important theoretical step that is necessary in order to successfully apply such algorithms in time-varying environments. Our work provides basic convergence rate estimates for FP and ECFP and develops tools that can aid in developing more precise estimates in particular games.

1.3.2 Implementation in Large-Scale Games

Myopic best-response dynamics (1.6)–(1.7) can be difficult to implement in large games due to the following issues:

- **High computational complexity**: In each round of a MBR algorithm, each player must choose a best response action. In general, the complexity of computing the best-response set grows as $O(e^N)$, where N is the number of players. (See Section 2.6 for more details.) Methods for mitigating computational complexity in MBR algorithms are studied in Chapter 5.

- **High information overhead**: In a MBR algorithm players must track the observation state $z(n)$. In games with many players, this can be a large amount of information. For example, in FP, the observation state is given by $q(n) = (q_1(n), \dots, q_N(n))$, the size of which grows linearly in the number of players. Methods for mitigating information overhead in MBR algorithms are studied in Chapter 3.

- **Demanding communication requirements**: After each round $n \in \{1, 2, \dots\}$ of the repeated learning process, the updated observation state $z(n)$ must be communicated to every player. Since the observation state $z(n)$ depends on the action $a_i(n)$ of every player $i \in \mathcal{N}$, this requires that information regarding the action of each player be communicated to every other player between each iteration of the repeated play process. Methods for mitigating demanding communication requirements in MBR algorithms are studied in Chapters 7–8.

- **Global synchronization**: Execution of a MBR process requires a form of global synchronization in the sense that, before any player can take a stage-$(n+1)$ action, every player is assumed to first take their stage-n action. An asynchronous form of the MBR dynamics (1.6)–(1.7) is studied in Chapter 6.

- **Convergence to pure-strategy equilibria**: As noted in Section 1.3.1 above (see second bullet point), in some applications it can be preferable for a learning process to converge to pure-strategy equilibria rather than mixed-strategy equilibria. Methods for mitigating this issue are studied in Chapters 8 and 10.

The second research focus of this dissertation concerns the development of techniques to mitigate these issues. We place a particular emphasis on developing algorithms that are practical to implement in a *network-based* setting.

Consider the traffic routing example discussed in Section 1.1 in which a group of vehicles connected via a sparse communication graph wishes to compute an equilibrium routing strategy. A network-based repeated play learning algorithm in this setting proceeds roughly as follows. Each vehicle makes a route choice and transmits some information regarding their route choice to neighboring vehicles. (The route choice of any individual vehicle is directly visible only to that vehicle.) Based on the information received from neighboring vehicles, players may revise their route choice and again transmit information to neighboring vehicles.

In general, we say a learning algorithm is *network-based* if

Assumption 1.1. *Players are endowed with a preassigned communication graph $G = (V, E)$, where the vertices V represent the players, and the edge set E consists of communication links between pairs of players that can communicate directly. The graph G is connected.*

Assumption 1.2. *Players directly observe only their own actions.*

Assumption 1.3. *A player may exchange information with immediate neighbors, as defined by G, at most once for each iteration or round of the repeated play.*

In this dissertation we develop methods for mitigating each of the problems listed above in a manner that is practical in network-based settings.

We note that an important difference between the network-based setting and traditional learning settings is the lack of payoff information. In many traditional learning frameworks, it is assumed that players physically engage a game, and at the end of each round of game play, players may observe the payoff received. In this setting, one may mitigate the problems of information overhead and inter-agent communication requirements (see above) by designing algorithms that rely on the information implicitly transmitted by physical interaction in the game.

In the network-based setting, players engage in a form of virtual game play in which no payoff measurements are available. In this setting, information cannot be transmitted by the physical interaction in the game—all information dissemination must be carried out directly by the algorithm. The challenge in this setting is to design algorithms which handle information dissemination in an efficient and practical manner.

1.3.3 Dissertation Outline

We now state the main contributions of the dissertation, outlining the contributions by chapter. Further details as well as literature review can be found in the introduction of each chapter.

Chapters 3 and 5–8 study practical methods for implementation of MBR dynamics in large games. Each of these chapters focuses on one particular issue as outlined in Section 1.3.2. Chapters 4 and 9–10 deal with fundamental properties of MBR dynamics.

Chapter 2: Setup and Notation

In this chapter we review relevant concepts in game theory, set up the notation to be used in the rest of the dissertation, and rigorously introduce myopic best response dynamics.

Chapter 3: Empirical Centroid Fictitious Play

In this chapter we study a variant of FP, referred to as *empirical centroid FP*, in which players are grouped into equivalence classes. This variant of FP mitigates the problems of high information overhead and high computational complexity noted in Section 1.3.2. Furthermore, ECFP is shown to converge quickly in the sense that the associated mean-field system converges to (completely) mixed equilibria in finite time. In addition, simulation results suggest that the rate of convergence depends on the number of equivalence classes but is invariant to the number of players.

In classical FP, players monitor collective behavior by tracking the empirical distribution of each individual player. In ECFP, players monitor collective behavior by tracking the *centroid* empirical distribution associated with each class. In classical FP, the amount of information that must be tracked depends linearly on the number of players. In ECFP, the amount of information that must be tracked depends linearly on the number of equivalence classes but is independent of the number of players. Two different learning processes may be studied in ECFP, depending on the definition of the observation space. We show that the tuple of empirical centroids associated with ECFP converges to the set of symmetric NE strategies, and we show that the tuple of individual empirical distributions associated with ECFP converges to an equilibrium set denoted as the set of mean-centric equilibria.

Similar to the continuous-time FP dynamics (1.2), ECFP admits an associated mean-field system, which we refer to as continuous-time ECFP. Analogous to the convergence result for discrete-time ECFP, it is shown that in continuous-time ECFP, the tuple of empirical centroids converges to the set of symmetric NE, and the tuple of empirical distributions converges to the set of mean-centric equilibria.

The rate of convergence of continuous-time ECFP is characterized. It is shown that if ECFP converges to a completely mixed equilibrium (i.e., an equilibrium lying in the interior of the strategy space), then convergence occurs in finite time. In ECFP, convergence to mixed strategy equilibria is typical. It is conjectured that a similar finite-time convergence result may be shown when ECFP converges converging to any mixed equilibrium (not necessarily completely mixed).

Chapter 4: Robustness in Myopic Best-Response Algorithms

The practicality of MBR algorithms depends on their robustness to perturbations that may be encountered in real-world applications. In particular, suppose that the myopic best response is perturbed so that players may sometimes choose suboptimal actions, but that the degree of suboptimality decays to zero over time. (In the spirit of [72], we sometimes call a MBR process that is perturbed in this manner a weakened MBR process.) We show that the fundamental learning property of a MBR algorithm is retained in the presence of such a perturbation. In the case of classical FP, this means that convergence to NE is preserved. More generally, if a MBR algorithm converges to some equilibrium set in the absence of perturbations, then our result can be applied to study convergence to the same equilibrium set in the presence of perturbations. At the core of this result is a technical lemma (see Lemma 4.16) which shows that a sequence of (unwieldy) ϵ-best responses (see (2.2)) may be equivalently viewed as a sequence of more amenable, so-called δ-perturbations (see Definition 4.4). Sequences of δ-perturbations are naturally handled within a stochastic approximation framework using techniques such as those developed in [28].

Robustness results of this kind were first studied in [29] for the case of classical FP. The chapter extends the approach of [29] to demonstrate robustness of general MBR algorithms. The main result of this chapter is Theorem 4.15, which shows that any weakened MBR process will converge to the internally chain recurrent set of the associated mean-field differential inclusion. This result enables a wide range of applications. Some applications, including a sampling-based technique for mitigating complexity in MBR dynamics, an asynchronous form of MBR dynamics, and network-based implementations of MBR dynamics, are considered in the subsequent chapters.

Chapter 5: Mitigating Computational Complexity: A Single Sample Approach

In this chapter we study a sampling-based method for mitigating the computational complexity of FP using stochastic approximation techniques. This method reduces the per-iteration complexity of executing a MBR algorithm from $O(e^N)$ to $O(N)$.

More precisely, the computationally expensive aspect of computing the best-response set is the evaluation of the expected utility function. Previous research [48, 49] has studied methods for approximating the expected utility function using a Monte Carlo approach. Each iteration of the repeated-play process, players draw samples from an underlying probability distribution and use the average as an estimate of the expected utility. As long as the number of samples drawn in round n goes to infinity as $n \to \infty$, then convergence to the equilibrium set is ensured.

In this chapter we consider an alternative sampling-based method in which players are only required to draw *one* sample each round of the learning process. Players approximate the expected utility by averaging the utility of the samples over time using a stochastic approximation rule. We refer to a MBR algorithm implemented using this method as a single sample MBR (SS-MBR) algorithm. Using the robustness result of Chapter 4, it is shown that a SS-MBR algorithm converges to the internally chain recurrent set of the associated mean-field system. By studying the mean-field system of a particular MBR algorithm one can show that this implies convergence to a particular equilibrium set. For example, in classical FP this implies convergence to the set of NE, and in ECFP this implies convergence of the centroid process to the set of symmetric NE.

Chapter 6: Asynchronous Learning

In this chapter we study asynchronous MBR learning processes and show that the assumption of global synchronization can be relaxed in a way that allows a variety of practical asynchronous implementations. We model asynchronous learning in a discrete-time learning process by supposing that agents may be active in some rounds and idle in others. Let $N_i(n)$ denote the number of rounds that agent $i \in \mathcal{N}$ has been active in rounds up to and including round n. We show that an asynchronous MBR learning process converges to the internally chain recurrent set of the associated mean-field system under the relatively mild assumption that $\lim_{n\to\infty} \frac{N_i(n)}{N_j(n)} = 1$ for all $i, j \in \mathcal{N}$. In classical FP this implies convergence to the set of NE and in ECFP this implies convergence to the set of symmetric Nash equilibria or to the set of mean-centric equilibria, depending on the defined observation space.

In order to model practical real-world learning processes, we consider continuous-time implementations of the asynchronous MBR dynamics described above. We refer to such a process as a continuous-time embedded MBR process. Each player is assumed to take a discrete sequence of actions. For $i \in \mathcal{N}$, we let $\tau_i(n) \in [0, \infty)$ denote the time instance at which player i takes an action for the n-th time, and let $N_i(t)$ denote the number of actions that a player has taken up to and including time $t \in [0, \infty)$. Using the convergence result for (discrete-time) asynchronous MBR processes, it is shown that any such learning process will converge to the internally chain recurrent set of the associated mean-field system as long as the rule governing the timing of players actions satisfies $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$.

We give two prototypical examples of action timing rules satisfying the above assumption. The first is entirely stochastic—players choose the timing of their actions according to a homogenous Poisson process. The second is a deterministic rule in which players adaptively adjust the timing of their actions based on high-level information about the overall action timing process.

As a second application of the asynchronous MBR learning model, we study the problem of, so called, "weak convergence" in a MBR algorithm. In particular, due to the lack of lower hemicontinuity in the best-response mapping, it is possible for a MBR learning process to converge to some equilibrium set, but players' period-by-period strategies never approach an equilibrium themselves. For example, Section 6.5.1 discusses an example in classical FP in which the empirical distribution converges to the unique mixed-strategy NE of a game, but in each round of the learning process, the players receive the lowest possible payoffs.

We say an algorithm converges strongly to equilibrium if players' period-by-period strategies converge to the equilibrium. We present a method for constructing a strongly-convergent MBR algorithm given a weakly-convergent MBR algorithm. We prove the algorithm converges strongly to the internally chain recurrent set of the associated mean-field system by showing that the (strong) learning process contains an embedded asynchronous MBR process. The result then follows as a consequence of our main convergence result for asynchronous MBR processes.

Chapter 7: Network-Based Implementation

In this chapter we study the implementation of MBR algorithms in a network-based setting satisfying Assumptions 1.1–1.3. We give a general template for a network-based MBR algorithm and, using the robustness result of Chapter 4, we show that a network-based MBR algorithm converges to the internally chain recurrent set of the associated mean-field system so long as the information dissemination scheme satisfies a mild asymptotic accuracy condition (see Condition 7.1). We then study two particular network-based MBR algorithms: A network-based implementation of ECFP, introduced in Chapter 3, and a network-based implementation of single sample FP (SS-FP), introduced in Chapter 5. We prove that the network-based ECFP algorithm converges to the set of symmetric NE, and we prove that the network-based implementation of SS-FP converges to the set of NE.

Chapter 8: Inertial MBR Algorithms: Incomplete Information and Network-Based Implementation

In this chapter we consider the problem of learning pure-strategy NE in a network-based setting satisfying Assumptions 1.1–1.3. Furthermore, we suppose that players have *incomplete information* in the sense that they have some uncertainty about the precise structure of their utility function. It is assumed that the probability measure modeling players beliefs regarding their utility function converges weakly to some limit. In order to learn pure-strategy equilibria in this setting, we consider a class of algorithms we denote as *inertial MBR dynamics*—such algorithms consist of a best-response component (similar to classical FP) and an additional "inertial term" which ensures that players occasionally repeat actions in consecutive stages. The use of such an inertial term is a common technique that has been used to ensure convergence of MBR processes to pure-strategy NE in more classical settings [73, 74]. We first prove a general convergence result that applies to any inertial MBR algorithm. The general convergence result applies to a broad class of learning dynamics and applies to a wide variety of information dissemination schemes (not just the synchronous graph-based information dissemination schemes that we focus on in the later part of the chapter). Subsequently, as applications of the general result, we study the network-based implementation (under uncertainty) of two important inertial MBR algorithms: network-based FP with inertia (referred to as N-FP), and network-based joint strategy FP (JSFP) with inertia (referred to as N-JSFP).

It should be noted that in Chapter 10 we will address the closely related problem of studying continuous-time FP dynamics in potential games. We prove that in almost all potential games, for almost all initial conditions, continuous-time FP converges to a pure-strategy equilibrium (see Theorem 10.1). This result, while powerful, does not obviate the results of this chapter. In this chapter we study convergence to pure-strategy equilibria in a broader class of games and a broader class of dynamics than considered in Theorem 10.1. In particular, we study the general class of myopic best response dynamics (a set of dynamics containing FP as a special case), and we prove convergence of such dynamics to a pure-strategy NE in the class of weakly-acyclic games (a class of games strictly larger than potential games).

Chapter 9: Non-Degenerate Potential Games

In this chapter we introduce a class of non-degenerate potential games in which the FP dynamics will be shown (in Chapter 10) to be "well behaved". A potential game is said to be non-degenerate if the gradient and Hessian of the potential function satisfy certain non-degeneracy conditions at every equilibrium point. These conditions ensure that best-response dynamics are well-behaved in the neighborhood of equilibrium points, and simplify the analysis of learning dynamics in (non-degenerate) potential games. While these notions of non-degeneracy are tailored to learning dynamics in potential games, it is shown that they are closely related to other well-known notions of non-degeneracy in general N-player games. A potential game is non-degenerate if and only if it is quasi-strong and regular [65, 67]. The main contribution of the chapter is to show that almost all potential games are non-degenerate. As an interesting consequence of our work in this chapter, we also obtain a useful characterization of the set of NE in potential games; namely, in almost all potential games, the number of NE strategies is finite and odd.

Chapter 10: Myopic Best-Response Dynamics in Potential Games

In this chapter we study fundamental properties of MBR dynamics in potential games. We study continuous-time dynamics throughout the chapter (when we refer to a particular MBR algorithm, we mean the continuous-time dynamics associated with the algorithm). First, we prove that in almost all potential games and for almost all initial conditions, FP converges to a pure-strategy NE. In particular, we prove that, in a non-degenerate potential game (see Chapter 9), mixed-strategy NE can only be reached from a null set of initial conditions. Since FP converges to the set of NE in potential games [28], this is equivalent to the desired result.

We then consider the problem of characterizing the rate of convergence of FP in potential games. In [27] Section 8, it was shown that mixed-strategy equilibria can pose a fundamental barrier to establishing convergence rate estimates for FP in potential games. If a FP path touches a mixed-strategy equilibrium, then it can rest there for an indeterminate amount of time before moving elsewhere. We sidestep this difficulty by showing that, within the class of non-degenerate potential games, for almost every initial condition, the associated FP path will never touch a mixed-strategy equilibrium. In [27] it was conjectured that the rate of convergence of FP is pathwise exponential ([27], Conjecture 25). We show that this conjecture holds in almost all potential games and for almost all initial conditions.

Finally, we use the tools developed earlier in the chapter to prove that if ECFP converges to a completely mixed equilibrium in a potential game, then it converges in finite time.

Chapter 11: Conclusions

In this chapter we recapitulate the main contributions of the thesis and discuss future research directions.

1.4 Related Work

In this section we give a brief overview of related work. Further discussion can be found in the introduction of each chapter. We divide our discussion into two sections corresponding to each of our main research areas.

Implementation of MBR Dynamics in Large-Scale Games

A general review of the field of learning in games can be found in [20, 74].

Various modifications of FP have been studied in order to mitigate the issues discussed in Section 1.3.2. The problem of mitigating complexity in FP using sampling-based techniques was first introduced in [48]. In Chapter 5 (see also [75]) we extend the techniques introduced in [48] to general MBR algorithms and improve on them by further reducing complexity using stochastic approximation techniques. In particular, the sampled FP algorithm developed in [48] has a per-iteration complexity that increases without bound as the algorithm progresses. The single-sample MBR algorithms studied in Chapter 5 have a constant complexity over time.

Complexity in FP may also be mitigated by using the reinforcement technique introduced [76]. In κ-exponential FP, players use any given action with a probability that depends on the average utility the action has generated when played in previous rounds. Alternatively, one may also mitigate complexity using a payoff-based variant of FP such as the actor-critic process studied in [29]. In comparison to these, the single sample approach to mitigating complexity applies more broadly to the general class of MBR dynamics and admits a straightforward network-based implementation (see Chapter 7).

A variant of FP known as joint strategy FP (JSFP) that mitigates the issues of computational complexity and information overhead was studied in [73]. In JSFP, players respond to the joint empirical distribution of play. Somewhat counterintuitively, this results in an algorithm with relatively low computational complexity, since agents can track the expected utility using a simple recursion and whatever aggregate statistics are necessary to compute the utility that each of their actions would have generated in a given round. Convergence to NE of the JSFP dynamics has not been shown. However, a form of JSFP in which the players occasionally repeat actions with some small probability has been shown to converge to the set of pure-strategy NE in ordinal potential games.

In [73] it is assumed that players are informed by an oracle of all information needed to compute the expected utility in each round. In Chapter 8 (see also [77]) we study a network-based implementation of JSFP in which all pertinent information is disseminated over a communication graph. The network-based implementation of JSFP studied in Chapter 8 is shown to converge to pure NE within the class of congestion games—a subset of potential games. The restriction to congestion games is a consequence of the manner in which information is gathered and disseminated.

The problem of computing NE in a network-based setting is a relatively new and active area of research. Stated chronologically by year, the work [78] studies a network-based variant of FP for computing NE, [79] studies a gossip-based algorithm for computing NE in aggregative games, [80] studies an algorithm for finding NE in a spatial spectrum access games, [81] studies a network-based algorithm for NE seeking in a two-network zero-sum games, [63] presents a method for designing games with a local a prescribed local dependence, [82] studies a network-based regret-based reinforcement learning algorithm for tracking the polytope of correlated equilibria in time-varying games, and [83] studies a gossip-based algorithm for computing NE in a network-based setting in games with continuous-action spaces. Our results differ from those above primarily in that, rather than focus on a particular learning algorithm and information dissemination scheme, we study the general class of MBR algorithms and show that, so long as the information dissemination scheme satisfies Condition 7.1, then any such algorithm will converge to the associated internally chain recurrent set. The underlying MBR dynamics can then be chosen to meet the system designers needs. (See Chapters 7–8 for more details.)

Traditionally, game-theoretic learning is couched in the framework of repeated play, which implicitly assumes agents act in synchrony [20]. The most popular asynchronous algorithm is log-linear learning, in which players are given the opportunity to revise their strategies at random times according to a Poisson process [84, 85]. The work [63] studies an asynchronous learning algorithm for potential games with continuous action spaces. The work [86] studies a class of payoff-based asynchronous dynamics. In Chapter 6 we show that MBR dynamics can be implemented in an asynchronous manner so long as the asymptotic rate of play of each player satisfies a mild condition. In general, asynchronous learning schemes would usually be analysed using asynchronous stochastic approximation (e.g. [87]) we show in Chapter 6 that asynchronicity can be handled in a more straightforward manner by simply using the robustness result of Chapter 4.

A popular approach used to mitigate communication issues in classical settings is the use of a payoff-based approach [29,61,84,88,89]. As discussed in Section 1.3.2, such algorithms use physical interaction in the game as a means of implicitly transmitting information. In a network-based setting, players do not have access to payoff measurements, and information dissemination must be handled explicitly by some information dissemination scheme.

We emphasize that, although FP in its classical form can be impractical to implement in large games, FP has been shown to possess a variety useful properties. For example, while it is known that no (uncoupled) learning algorithm can converge in all games [39], it has been shown that FP can have good payoff performance, even in the absence of convergence [40]. As another example, the paper [41] studies approximation guarantees unique to FP. The FP algorithm has been shown to be useful in solving general extensive form games [42,43]; in particular, it has been used to compute approximately-optimal strategies in poker [43–46] as well as large-state-space word games [47]. The work [48] showed that FP can be useful as an optimization heuristic in large-scale games. Variants of FP have been studied in a wide range of applications including traffic routing [48, 49], distributed constraint optimization problems [50], control of robotic teams [12, 51], dynamic programming [52, 53], cognitive radio [54–56], and learning in Markov decision processes [57].

Fundamental Properties of MBR Dynamics

The general motivation for studying game-theoretic learning processes may be broadly subdivided into two categories: a *prescriptive* research agenda, and a *descriptive* research agenda [90]. The prescriptive agenda seeks to prescribe behavior rules to agents in a multi-agent system in order to achieve some desirable global outcome (e.g., as in distributed control). In particular, the prescriptive agenda is not concerned with describing the behavior of agents that may be acting in real world economic settings. On the other hand, the descriptive agenda is concerned with studying how “natural agents learn in the context of other learners” [90].

With our first research focus (implementation in large games), we are largely motivated by a prescriptive agenda. However, our second research focus (fundamental properties of MBR dynamics) may be seen as being motivated by both a descriptive and prescriptive agenda.

The second research focus is motivated by a prescriptive agenda in the sense that, in many parts of the dissertation, we employ the fundamental results developed as part of this research focus as a theoretical underpinning in order to construct algorithms that are practical for implementation in large-scale games. The second research focus can be seen as being motivated by a descriptive agenda for two reasons. First, the discrete-time MBR dynamics (1.6)–(1.7) may be seen as a fairly natural set of learning dynamics. Each round of play, agents predict how others will act and choose their next-stage action as a myopic best response given their prediction. Indeed, the prototypical learning dynamics in this class are the FP dynamics, which in continuous time (1.2), may be considered to be the natural learning dynamics associated with the NE concept. Second, we note that the robustness result that we study for MBR dynamics considers learning processes where agents with bounded rationality may occasionally make mistakes in the best-response computation, as might occur with natural learners.

Robustness results of the kind studied in Chapter 4 were first studied in [29] for the case of classical FP. Chapter 4 extends the approach of [29] to demonstrate robustness of general MBR algorithms. In addition to enhancing the applicability of MBR dynamics to real-world control problems, the results of this chapter have required the development of useful new technical tools. For example, Lemma 4.16 studies ϵ-best-response sequences and demonstrates that such sequences may in fact be considered in terms of a more amenable sequence of so called δ-perturbations (see Section 4.2).

The convergence result for FP in potential games (see Theorem 10.1) is relatively novel. Continuous-time best-response dynamics similar to those we consider here have been studied in various works, including [27, 28, 64, 91]. These papers study a variety of convergence properties for FP and replicator dynamics in a wide class of games, but do not consider the question of generic convergence to pure-strategy equilibria. A related result for two player games has been shown in [92], where it was demonstrated that "continuous-time FP almost never converges cyclically to a mixed-strategy equilibrium in which both players use more than two pure-strategies."

The rate of convergence of FP is studied in [27]. While convergence rate estimates are given for zero-sum games, it is shown that it is not possible to establish a general convergence rate estimate for FP in potential games. If a FP path intersects a mixed-strategy equilibrium, it may rest there for an indeterminate amount of time before moving elsewhere. We work around this problem by showing that in almost all potential games, a FP path never intersects a mixed-strategy equilibrium.

Chapter 2

Setup and Notation

2.1 Games in Normal Form

A game (in normal form) consists of three components: A set of players $\mathcal{N} = \{1, \ldots, N\}$, a set of actions Y_i corresponding to each player $i \in \mathcal{N}$, and a set of utility functions $u_i : Y_1 \times \cdots \times Y_N \to \mathbb{R}$, $i \in \mathcal{N}$, representing the preferences of each player.

We denote an instance of a game as $\Gamma := \{\mathcal{N}, (Y_i, u_i)_{i \in \mathcal{N}}\}$. We assume that the action space Y_i is finite for each player and let $Y := Y_1 \times \cdots \times Y_N$ denote the *joint* action space.

In order to ensure the existence of equilibrium strategies in the game it is necessary to consider an extension of the game in which players are permitted to use probabilistic strategies. For a finite set X, let $\triangle(X)$ denote the set of probability distributions over X. Let $\Delta_i := \triangle(Y_i)$ denote the set of *mixed* strategies available to player i, let $\triangle(Y_{-i})$ be the set of *mixed* strategies (possibly correlated) available to all players other than i, and let $\Delta := \triangle(Y)$ denote the set of joint mixed strategies (possibly correlated) available to all players.[1]

In large-scale distributed settings it is often convenient to study mixed strategies where players act independently. Let $\Delta^N := \prod_{i \in N} \Delta_i$ denote the set of *independent* joint strategies. A strategy $\sigma = (\sigma_1, \ldots, \sigma_N) \in \Delta^N$, where σ_i denotes the marginal strategy of player i, may be represented in the space Δ as the product $\sigma_1 \times \cdots \times \sigma_N \in \Delta$. In this context we define $\Delta_{-i} := \prod_{j \neq i} \Delta_j$ to be the set of (independent) mixed strategies of players other than i, when the meaning is clear from the context. When convenient, we represent a mixed strategy $\sigma \in \Delta^N$ by $\sigma = (\sigma_i, \sigma_{-i})$, where $\sigma_i \in \Delta_i$ denotes the marginal strategy of player i and $\sigma_{-i} = (\sigma_1, \ldots, \sigma_{i-1}, \sigma_{i+1}, \ldots, \sigma_N) \in \Delta_{-i} = \prod_{j \neq i} \Delta_j$ denotes the strategies of players other than i.

[1] In order to simplify the notation in the introduction, we used q_i to denote a generic element of Δ_i. However, for the remainder of the dissertation we will prefer to use σ_i to denote a generic element of Δ_i, reserving q to denote the empirical distribution in an FP process (see Section 2.5).

In the context of mixed strategies, we often wish to retain the notion of playing a single deterministic action. For this purpose, for $i \in \mathcal{N}$, let $\mathbf{1}_{y_i}$ denote the mixed strategy placing probability one on the action $y_i \in Y_i$ and let

$$A_i := \{\sigma_i \in \Delta_i : \ \sigma_i = \mathbf{1}_{y_i} \text{ for some } y_i \in Y_i\}.$$

The set A_i contains the vertices of Δ_i and lies in one-to-one correspondence with the set of pure strategies Y_i.

For $x \in \Delta$, the expected utility of player i is given by[2]

$$U_i(x) := \sum_{y \in Y} u_i(y) x(y_1, \ldots, y_n). \tag{2.1}$$

Given a strategy $x_{-i} \in \triangle(Y_{-i})$, the best-response for player i is given by the set-valued function $\mathrm{BR}_i : \triangle(Y_{-i}) \to \Delta_i$,

$$\mathrm{BR}_i(x_{-i}) := \arg\max_{x_i \in \Delta_i} U_i(x_i, x_{-i}).$$

More generally, for $\epsilon > 0$ the ϵ-best-response set is given by $\mathrm{BR}_{i,\epsilon} : \triangle(Y_{-i}) \to \Delta_i$

$$\mathrm{BR}_{i,\epsilon}(x_{-i}) := \{\tilde{x}_i \in \Delta_i : \ U_i(\tilde{x}_i, x_{-i}) \geq \max_{x_i \in \Delta_i} U_i(x_i, x_{-i}) - \epsilon\}. \tag{2.2}$$

The joint best response is given by $\mathrm{BR} : \Delta \to \Delta$

$$\mathrm{BR}(x) := \mathrm{BR}_1(x_{-1}) \times \cdots \times \mathrm{BR}_N(x_{-N}).$$

To keep notation simple, we sometimes employ the following abuses. The notation $y_i \in \mathrm{BR}_{i,\epsilon}(x_{-i})$ means that $\mathbf{1}_{y_i} \in \mathrm{BR}_{i,\epsilon}(x_{-i})$. Similarly, for $y_i \in Y_i$ the notation $U_i(y_i, x_{-i})$ refers to the expected utility $U_i(\mathbf{1}_{y_i}, x_{-i})$, and for $y \in Y$, $U_i(y)$ refers to the pure strategy utility $U_i(\mathbf{1}_y) = u_i(y)$.

We will often need to work directly with strategies $\sigma \in \Delta^N$. To keep notation simple we overload the notation for U and BR as follows. For $\sigma \in \Delta^N$, we let the expected utility of player i be given by

$$U_i(\sigma) := \sum_{y \in Y} u_i(y) \sigma_1(y_1) \cdots \sigma_N(y_N).$$

[2] We note that, in the introduction, we defined the utility function to be a map from Δ^N to $\mathbb{R}$. Here, we define the utility function in a more general manner that permits correlated strategies. A similar comment holds for the best response set, defined below.

For $\sigma_{-i} \in \Delta_{-i}$ we let $\mathrm{BR}_i(\sigma_{-i}) := \mathrm{BR}_i(\sigma_1 \times \cdots \times \sigma_{i-1} \times \sigma_{i+1} \cdots \times \sigma_N)$. Similarly, for $\sigma \in \Delta^N$ we let $\mathrm{BR}(\sigma) := (\mathrm{BR}_1(\sigma_{-1}), \ldots, \mathrm{BR}_N(\sigma_{-N}))$.

A strategy $\sigma \in \Delta^N$ is said to be a *Nash equilibrium* if $\sigma_i \in \mathrm{BR}_i(\sigma_{-i}),\ \forall i \in \mathcal{N}$, or equivalently,

$$\sigma \in \mathrm{BR}(\sigma). \tag{2.3}$$

At a Nash equilibrium strategy, no player can improve their utility by unilaterally deviating from the given strategy. The set of Nash equilibria is given by

$$NE := \{\sigma \in \Delta^N : \ \sigma \in \mathrm{BR}(\sigma)\}. \tag{2.4}$$

2.1.1 Potential Games

In a potential game, there exists an underlying potential function which all players implicitly seek to maximize. Such games are fundamentally cooperative in nature (all players benefit by maximizing the potential), and have many important applications in engineering and economics [38,93]. Along with so called *harmonic games* (which are fundamentally adversarial in nature), potential games may be seen as one of the basic building blocks of general N-player games [58].

Formally, a game Γ is said to be a potential game[3] if there exists a function $u : Y \to \mathbb{R}$ such that for every $i \in \mathcal{N}$

$$u_i(y_i', y_{-i}) - u_i(y_i'', y_{-i}) = u(y_i', y_{-i}) - u(y_i'', y_{-i}), \qquad \forall y_{-i} \in Y_{-i},\ \forall y_i', y_i'' \in Y_i.$$

Potential games have a wide range of applications in the field of multi-agent systems, including control of power systems [8]; cognitive radio and opportunistic spectrum access [5]; cover optimization and collaborative sensing in mobile sensor networks [13, 15]; power allocation, channel allocation, scheduling, and topology in control in wireless adhoc networks [3, 4, 94, 95]; large-scale optimization [48]; and traffic optimization [48, 49], to name a few. More generally, [93] explores connections between cooperative control and potential games. The paper [63] studies methods for designing a potential game so that the Nash equilibria of the designed game optimize a given objective function.

[3]More precisely, this notion of a potential game is sometimes referred to as an *exact* potential game [38]. Other notions of potential games including weighted potential games and ordinal potential games can be found in [38]. Unless otherwise stated, in this dissertation the term potential game refers to an exact potential game in the sense of [38].

2.1.2 Best-Response Equivalence in Potential Games

Let Γ be a potential game, with potential function u. Let U denote the mixed extension of u. Using the definitions of U_i and U it is straightforward to verify that

$$\mathrm{BR}_i(\sigma_{-i}) := \arg\max_{\sigma_i' \in \Delta_i} U_i(\sigma_i, \sigma_{-i}) = \arg\max_{\sigma_i' \in \Delta_i} U(\sigma_i, \sigma_{-i}).$$

Thus, in order to compute the best-response set we only require knowledge of the potential function U, not necessarily the individual utility functions $(U_i)_{i=1,\ldots,N}$. When studying best-response-based learning processes in potential games, this means that the learning dynamics are completely determined if the potential function is specified, even if the individual player's utility functions are not precisely known or specified.

2.2 Other Notation

Other notation as used throughout the dissertation is as follows.

- $|S|$ denotes the number of elements in a finite set S
- $K_i := |Y_i|$
- $K := |Y|$
- $\mathbb{N} := \{1, 2, \ldots\}$
- S^c denotes the complement of a set S
- $\mathring{S}$ denotes the interior of a set S
- $\mathrm{cl}\, S$ denotes the closure of a set S
- ∂S denotes the boundary of a set S
- The distance between a point $x \in \mathbb{R}^m$ and a set $S \subset \mathbb{R}^m$ is given by $d(x, S) := \inf\{\|x - x'\| : x' \in S\}$
- The support of a function $f : \Omega \to \mathbb{R}$ is given by $\mathrm{spt}(f) := \{x \in \Omega : f(x) \neq 0\}$
- Given a function f, $\mathcal{D}(f)$ refers to the domain of f and $\mathcal{R}(f)$ to the range of f
- $\mathcal{L}^k$, $k \in \{1, 2, \ldots\}$ refers to the k-dimensional Lebesgue measure

- $\mathcal{H}^s$, $0 \leq s < \infty$ refers to the s-dimensional Hausdorff measure

- $\|\cdot\|$ denotes the standard Euclidean norm unless otherwise specified

- Given an open set $\Omega \subset \mathbb{R}^n$, and $k \in \mathbb{N} \cup \{\infty\}$, $C_c^k(\Omega)$ denotes the set of k-times differentiable functions with compact support in Ω

- The mapping $\operatorname{sgn} : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}$ is given by

$$(\operatorname{sgn}(\mathbf{A}))_{i,j} := \begin{cases} 1 & \text{if } a_{i,j} > 0 \\ -1 & \text{if } a_{i,j} < 0 \\ 0 & \text{if } a_{i,j} = 0. \end{cases}$$

- Given two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimension, $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard product (i.e., the entrywise product) of $\mathbf{A}$ and $\mathbf{B}$.

- Suppose $m, n, p \in \mathbb{N}$, $F_i : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \ldots, p$. Suppose further that $F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$ is given by $F(w, z) = (F_i(w, z))_{i=1,\ldots,p}$. Then the operator D_w gives the Jacobian of F with respect to the components of $w = (w_k)_{k=1,\ldots,m}$; that is

$$D_w F(w, z) = \begin{pmatrix} \frac{\partial F_1(w,z)}{\partial w_1} & \cdots & \frac{\partial F_1(w,z)}{\partial w_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_p(w,z)}{\partial w_1} & \cdots & \frac{\partial F_p(w,z)}{\partial w_m} \end{pmatrix}.$$

- Throughout, we assume the existence of probability spaces rich enough to carry out the construction of the various random variables required. As a matter of convention, all equalities, inequalities, and set inclusions involving random quantities are interpreted almost surely (a.s.) with respect to the underlying probability measure, unless otherwise stated.

- The symbol σ is often used to denote an element of Δ^N. We note, however, that when discussing a random variable X in an appropriately defined probability space, we use the notation $\sigma(X)$ in the traditional probabilistic sense to denote the σ-algebra generated by X.

2.3 Repeated Play Learning

Let a game Γ be fixed. In a repeated play learning process, players are assumed to repeatedly face off in the game Γ in stages $n \in \mathbb{N}$. Let $y_i(n) \in Y_i$ denote the action taken by player i in stage n of the repeated play. Alternatively, since there is a one-to-one correspondence between Y_i and A_i, we sometimes use $a_i(n) \in A_i$ to denote the action taken by player i in stage n of the repeated play. In this case we have the relationship $a_i(n) = \mathbf{1}_{y_i(n)}$.

A repeated-play learning algorithm is a set of rules which prescribes how players should choose their stage-$(n+1)$ action based on the history of play though stage n. In this dissertation we will focus on a set of behavior rules where each player choose their stage-$(n+1)$ action as a myopic best response given some prediction of opponent behavior based on the action history through stage n. This class of learning processes is formally defined in the following section.

2.4 Myopic Best-Response Dynamics

Let players be engaged in repeated play of a game $\Gamma = \{\mathcal{N}, (Y_i, u_i)_{i\in\mathcal{N}}\}$ as defined in Section 2.3.

Let Z denote a compact subset of $\mathbb{R}^m$ for some $m \in \mathbb{N}$, where the information that players keep track of is assumed to live. We refer to Z as the *observation space*. Let

$$g : \Delta(Y) \to Z$$

be a map from the joint mixed strategy space to the observation space. We assume

Assumption 2.1. *The observation map g is uniformly continuous.*

Let $\{z(n)\}_{n\geq 1}$ be a sequence in Z that is defined recursively by letting $z(1) \in Z$ be arbitrary and for $n \geq 1$

$$z(n+1) = z(n) + \gamma(n)\big(g\left(a(n+1)\right) - z(n)\big), \tag{2.5}$$

where $\{\gamma(n)\}_{n\geq 1}$ is a predefined sequence of weights satisfying

Assumption 2.2. $\lim_{n\to\infty} \gamma(n) = 0$, $\sum_{n\geq 1} \gamma(n) = \infty$.

We refer to $z(n)$ as the *observation state*. In a myopic best-response (MBR) algorithm, each player forms a prediction (or forecast) of the future behavior of opponents as a function of the observation state $z(n)$. In particular, for each player i, let $f_i : Z \to \Delta(Y_{-i})$ be a function mapping from the observation state to a forecast of opponents strategies. We make the following assumption

Assumption 2.3. *The forecast map f_i is continuous for each $i \in \mathcal{N}$.*

Given $f := (f_1, \ldots, f_N)$, we define the best-response function $\text{BR}_f : Z \to \Delta(Y)$ associated with a MBR algorithm as

$$\text{BR}_f(z) := \text{BR}_1(f_1(z)) \times \cdots \times \text{BR}_N(f_N(z)).$$

When players are engaged in repeated play, we say the sequence $\{z(n)\}_{n\geq 1}$ is a MBR process if each player's stage-$(n+1)$ strategy is chosen as a myopic best response given their prediction of opponents strategies. That is, a sequence $(z(n))_{n\geq 1}$ satisfying (2.5) is a MBR process if actions are chosen according to the rule

$$a_i(n+1) \in \text{BR}_i(f_i(z(n))), \ \forall i, \ \forall n, \tag{2.6}$$

or equivalently in recursive form (see (2.5))

$$z(n+1) - z(n) \in \gamma(n)\big(g(\text{BR}_f(z(n))) - z(n)\big). \tag{2.7}$$

We denote an instance of a MBR algorithm as $\Psi = (\{\gamma(n)\}_{n\geq 1}, g, (f_i)_{i=1}^N)$.

2.5 Fictitious Play

The fictitious play (FP) algorithm, introduced in [26], is a prototypical myopic best-response algorithm. Let

$$q_i(n) := \frac{1}{n}\sum_{s=1}^{n} a_i(s) \tag{2.8}$$

denote the *empirical distribution of player* i; $q_i(n) \in \Delta_i$ is a normalized histogram of the action choices of player i through time n. Let the *empirical distribution* be given by the tuple $q(n) := (q_1(n), \ldots, q_N(n)) \in \Delta^N$.

In FP, players choose their next-stage action as a myopic best response to the empirical distribution. In particular, we say the sequence $(q(n))_{n\geq 1}$ is a *FP process* if

$$a_i(n+1) \in \text{BR}_i(q_{-i}(n)), \quad \forall i \in \mathcal{N}. \tag{2.9}$$

The FP algorithm defined by (2.8)–(2.9) may be intuitively understood as follows. Each player naively assumes that all other players are playing according to some (unknown) time-invariant mixed strategy, and that all other players strategies are independent of one another. Accordingly, each player uses the empirical distribution $q_{-i}(n)$ as an estimate of the (mixed) strategy that will be used by opponents in round $n+1$, and chooses an action for stage $n+1$ in order to myopically optimize their utility given this assumption.

Note that (2.8) may be expressed recursively as

$$q_i(n+1) = q_i(n) + \frac{1}{n+1}\big(a_i(n+1) - q_i(n)\big). \tag{2.10}$$

As noted in Section 1.2, by considering the tuples $a(n) = (a_1(n), \ldots, a_N(n))$, and $q(n) = (q_1(n), \ldots, q_N(n))$, and using the recursively relationship (2.10), the process $(q(n))_{n\geq 1}$ defined by (2.8)–(2.9) may be expressed recursively as

$$q(n+1) - q(n) \in \frac{1}{n+1}\left(\mathrm{BR}(q(n)) - q(n)\right). \tag{2.11}$$

Furthermore, (2.11) may be seen as a discretization of the continuous-time dynamical system

$$\dot{\mathbf{q}} \in \mathrm{BR}(\mathbf{q}) - \mathbf{q}. \tag{2.12}$$

The dynamics (2.12) can be considered the "natural" learning dynamics associated with the Nash equilibrium concept (2.3). By definition, the rest points of (2.12) coincide with the set of NE (2.4). Historically, the dynamical system (2.12) is referred to as *continuous-time FP* [27–29].[4]

Since our primary focus through much of the dissertation will be on discrete-time learning processes, unless otherwise stated, we refer to a sequence $(q(n))_{n\geq 1}$ satisfying (2.11) as a FP process and we will refer to appropriate solutions of (2.12) as continuous-time FP processes.[5] More precisely,

Definition 2.4. *A an absolutely-continuous mapping $\mathbf{q} : \mathbb{R} \to \Delta^N$ is said to be a continuous-time fictitious play process with initial condition q_0 if $\mathbf{q}(0) = q_0$ and (2.12) holds for almost all $t \in \mathbb{R}$.*

It can be shown that linear interpolation of a (discrete-time) FP process (2.11) yields a *perturbed solution* of continuous-time FP (2.12) (see Section 4.2 and [28]). Using this relationship, the work [28] presents methods that allow one to derive results about the asymptotic properties of (2.11) by studying the continuous-time process (2.12)

This approach is not restricted to FP. In general, we will find that a useful approach for studying asymptotic properties of discrete-time MBR processes is to consider the associated continuous-time

[4] We note that some authors refer to the non-autonomous system $\dot{\mathbf{q}} \in \frac{1}{t}(\mathrm{BR}(\mathbf{q}) - \mathbf{q})$ as continuous-time FP [27]. Since solutions of these systems are equivalent after a time change [27, 64], we find it convenient to focus on the autonomous system (2.12).

[5] As an exception, in Chapter 10 we will focus exclusively on continuous-time dynamics and we will find it convenient to refer to (2.12) simply as FP rather than continuous-time FP throughout the chapter.

"mean-field" system. In Chapter 4 we will use the tools developed in [28] to study properties of general (discrete-time) MBR dynamics (2.7) via the associated mean-field system.

The FP process (2.11) may be seem as a special case of myopic best-response dynamics (2.7) where the observation space is given by Δ^N, the observation state is given by $q(n)$, and the functions $(f_i)_{i\in\mathcal{N}}$ and g are given by the identity.

2.5.1 Convergence Results for FP

We say that a FP process converges to the set of NE (or briefly, *converges*), if

$$\lim_{n\to\infty} d(q(n), NE) = 0. \tag{2.13}$$

This notion of learning is sometimes referred to as convergence in empirical frequency [20].

FP does not converge in all games; for example, [59, 96, 97] give examples of games where the empirical distribution cycles around a NE strategy without ever converging to it. However, FP is known to converge in several important classes of games including two-player zero-sum games [31,32], two-player two-action games [33,34], generic two-player m-action games [35], games solvable by iterated strict dominance [36], "one-against-all" multi-player games [37], and potential games [34,38].

We summarize these results in the following theorem.

Theorem 2.5. *Suppose Γ is a two-player, zero-sum game, two-player two-action game, generic two-player m-action game, game solvable by iterated strict dominance, one-against-all game, or potential game. Then FP converges to the set of NE in the sense that $\lim_{n\to\infty} d(q(n), NE) = 0$.*

2.6 Computational Complexity in Myopic Best-Response Algorithms

When discussing problems with implementation of MBR algorithms in large-scale games in Section 1.3.2, it was claimed that MBR algorithms can be impractical to deploy in large games due to the high computational complexity of computing a best response strategy. In particular, it was claimed that the complexity of computing the best-response set grows as $O(e^N)$, where N is the number of players. We briefly clarify this claim here.

In a MBR algorithm, each player $i \in \mathcal{N}$ must solve the optimization problem

$$\arg\max_{y_i\in Y_i} U_i(y_i, f_i(z(n)))$$

at each iteration of the algorithm. In order to solve this optimization problem, player i must compute the expected utility for each action $y_i \in Y_i$ given the probability distribution $f_i(z(n))$. In general, the computational complexity of evaluating this expected utility increases exponentially in the number of players N. This is the case, for example, in FP, where $f_i(z(n))$ is given by the $(N-1)$-dimensional probability distribution represented by the tuple

$$q_{-i}(n) = (q_1(n), \ldots, q_{i-1}(n), q_{i+1}(n), \ldots, q_N(n)) \in \Delta_{-i}.$$

Chapter 3

Empirical Centroid Fictitious Play

3.1 Introduction

The FP algorithm, introduced in Section 2.5, is a prototypical MBR algorithm that is guaranteed to converge to NE in several important classes of games. Despite convergence guarantees, FP can be difficult to impractical in games with many players. In this chapter we study a variant of FP which mitigates the problem of high information overhead inherent in large-scale implementations of FP (see Section 1.3.2). Additionally, the methods studied in this chapter aid in reducing the computational complexity of FP, and may be used to mitigate slow convergence rates in games with many players.

In FP (2.8)–(2.9), each player must track the empirical distribution $q(n) = (q_1(n), \ldots, q_N(n))$. This is a vector whose size scales linearly with the number of players, N. In this chapter, we consider a variant of FP in which players are grouped into equivalence classes. Rather than track the empirical distribution of each individual player, players only track the *centroid* empirical distribution corresponding to each equivalence class.

We call this variant of FP *empirical centroid fictitious play* (ECFP). The memory size of the centroid distribution depends on the number of equivalence classes into which players are grouped, but is invariant to the number of players. Hence, in ECFP, the size of the vector that must be tracked is independent of the number of players.

In order to study the properties of ECFP, it can be useful to consider the continuous-time dynamics associated with the algorithm. In the continuous-time setting certain properties of the algorithm, such as convergence to NE, rate of convergence, and robustness to perturbations can be easier to study. Using stochastic approximation techniques, it is often possible to then derive results about the properties of the discrete-time process using the continuous-time process as a type of mean-field approximation.

In this chapter we begin by studying discrete-time ECFP directly. In Section 3.2 we introduce ECFP in its basic form and prove that the algorithm converges to a subset of the set of NE which we refer to as the set of consensus equilibria.

The concept of a consensus equilibrium is closely related to that of a symmetric equilibrium discussed in [98]. The existence of symmetric equilibrium in finite normal form games was first proven by Nash [98] in the same work where the concept of Nash equilibrium was originally presented. In general, a symmetric equilibrium is a Nash equilibrium that is invariant under automorphisms of the game. A consensus equilibrium, on the other hand, is a Nash equilibrium in which all players use an identical strategy. In the case of a symmetric game, the two concepts coincide.

In Section 3.3 we present a more general form of ECFP which allows some of the assumptions of Section 3.2 to be relaxed. The more general form of ECFP will be shown to converge to a subset of the NE, known as the set of symmetric NE (SNE).[1] ECFP converges to the set of SNE in the sense that the tuple of empirical centroid distributions (see (3.1) and (3.6)) converges to the set of SNE.

A more traditional notion of learning is that of *convergence in empirical frequency*, which considers the asymptotic behavior of the tuple of individual empirical distributions (see (2.8) and subsequent discussion), rather than the tuple of centroid distributions [74, 99]. It will be shown that an ECFP process converges to an equilibrium set known as the set of *mean-centric equilibria* (MCE) in terms of convergence in empirical frequency.

In Section 3.4 we present the continuous-time ECFP dynamics. We prove that a continuous-time ECFP process converges to the set of SNE and MCE in a manner similar to the discrete-time process. We also characterize the rate of convergence of the continuous-time process. In particular, it will be shown that if the limit point of ECFP is a completely mixed equilibrium, then the algorithm converges in finite time.

In Section 3.5 we present an illustrative example of ECFP in a traffic routing scenario and provide simulation results.

Before presenting the ECFP algorithm, we briefly discuss related work.

3.1.1 Related Work

The (continuous-time) ECFP dynamics are similar to the best-response dynamics studied in [64] for two-player population games. In [64] it is supposed that there is an infinite population of

[1] We note that our definition of a symmetric NE differs from that of [98] in that explicitly relies on the manner in which players are grouped. See (3.7) for more details.

players and, in each small time interval, a small fraction of players revise their strategies (which are assumed to be pure for each individual player). It is shown that mixed-strategy equilibria are stable under these dynamics in two-player symmetric games. In ECFP, a group of N players responds to the average empirical distribution of the group (or, the average empirical distribution of each class, if multiple equivalence classes are assigned). In the case of two-player symmetric games, this results in dynamics identical to [64]. In this chapter we study the system in more general N-player symmetric potential games.

A variant of FP known as joint strategy FP (JSFP) that mitigates the issues of computational complexity and information overhead was studied in [73]. In JSFP, players respond to the joint empirical distribution of play. Somewhat counterintuitively, this results in an algorithm with relatively low computational complexity, since agents can track the expected utility using a simple recursion and whatever aggregate statistics are necessary to compute the utility that each of their actions would have generated in a given round. Convergence to NE of the JSFP dynamics has not been shown. However, a form of JSFP in which the players occasionally repeat actions with some small probability has been shown to converge to the set of pure-strategy NE in ordinal potential games [73]. In [73], no information gathering scheme is explicitly defined; players are assumed to have full access to the necessary information at all times. In network-based ECFP discussed in Chapter 7, the information gathering scheme is explicitly defined via a preassigned (but arbitrary) communication graph, and convergence results are demonstrated when inter-agent communication is restricted to local neighborhoods conformant to the graph.

Dynamic FP [100] applies principles of dynamic feedback from control theory to improve the convergence properties of a continuous-time version of FP. The algorithm is shown to be stable around some Nash equilibria where traditional FP is unstable. While the results generalize to multi-player games, there is no mitigation of the information gathering problem. In [101], a similar algorithm utilizing only payoff-based dynamics is presented. Similar stability results are shown when the class of games is restricted to games with a pairwise utility structure.

Average strategy fictitious play (ASFP) [102] is a variant of JSFP with inertia applicable within the class of congestion games. This class of games is different from the class of potential games studied for ECFP (see Section 3.3); neither class subsumes the other. In ASFP it is assumed that an oracle provides players with the average congestion profile. Players choose a best response in a manner similar to the JSFP best response rule with inertia. ASFP is shown to converge to a pure-strategy NE rather than a symmetric NE.

3.2 Empirical Centroid FP: Basic Framework

For ease of presentation, in this section we study a basic variant of ECFP in which all players are grouped into a single class. In Section 3.3 we will generalize these ideas to the case where players may be grouped into multiple classes.

Suppose that players do not have the ability to track individual actions of other players, $(a_i(n))_{n\in\mathcal{N}}$, but are able to track the *average* action of the collective, $\bar{a}(n) := \frac{1}{N}\sum_{i=1}^N a_i(n)$. With this information, player do not have knowledge of the empirical distribution $q(n)$ (see (2.8)), but do have knowledge of the *average* empirical distribution,

$$\bar{q}(n) := \frac{1}{N}\sum_{i=1}^{N} q_i(n). \tag{3.1}$$

We say the sequence $(\bar{q}(n))_{n\geq 1}$ is an ECFP process if for all $i \in \mathcal{N}$ and $n \geq 1$, actions are chosen according to the rule,

$$a_i(n+1) \in \arg\max_{\alpha_i \in A_i} U(\alpha_i, \bar{q}_{-i}(n)), \tag{3.2}$$

where $\bar{q}_{-i}(n) = (\bar{q}(n), \ldots, \bar{q}(n))$ is the $(N-1)$-tuple containing $(N-1)$ repeated (identical) copies of $\bar{q}(n)$.

In order to ensure that $\bar{q}(n)$ is well defined, we assume:

Assumption 3.1. *All players use the same strategy space.*

In order to ensure that ECFP converges to a Nash equilibrium strategy we assume

Assumption 3.2. Γ *is a potential game with permutation invariant potential function* u. *That is,* $u([y']_i, [y'']_j, y_{-(i,j)}) = u([y'']_i, [y']_j, y_{-(i,j)})$, *where, for any player* $k \in N$, *the notation* $[y']_k$ *indicates the action* $y' \in Y_k$ *being played by player* k, *and* $y_{-(i,j)}$ *denotes the set of actions being played by all players other than* i *and* j.

Note that because of assumption Assumption 3.1, permutations of the form given in Assumption 3.2 are necessarily well defined. Also note that under these assumptions, the set of consensus equilibria is known to be nonempty [103]. (Both of these assumptions will be weakened in Section 3.3.)

Intuitively speaking, FP may be said to describe a process where players (naively) assume opponents are player mixed strategies which are stationary and independent (see Section 2.5). In a similar vein, intuitively speaking, ECFP (3.1)–(3.2) describes a process where each player (naively) assumes opponents are playing mixed strategies which are stationary, independent, and *identical*—the last assumption being the primary difference between ECFP and FP in this intuitive interpretation. Each player naively assumes $\bar{q}(n)$ accurately represents the (identical) mixed strategy used by each opponent and chooses a best response to myopically optimize her next-stage utility accordingly.

The set of consensus equilibria is given by (cf. (2.4))

$$CNE := \{\sigma \in NE \ : \ \sigma_1 = \sigma_2 = \cdots = \sigma_N\}.$$

Let

$$\bar{q}^N(n) := (\bar{q}(n), \ldots, \bar{q}(n)), \tag{3.3}$$

be the N-tuple containing repeated copies of $\bar{q}(n)$. Players engaged in ECFP asymptotically learn consensus Nash equilibrium strategies in the sense that $d(\bar{q}^N(n), CNE) \to 0$ as $n \to \infty$. This notion of learning, while similar in spirit to the typical notion of convergence in empirical frequency (see Section 2.5.1), differs in that it is the tuple of empirical *centroid* distributions that is converging, rather than the tuple of empirical distributions.

The result is summarized in the following theorem.

Theorem 3.3. *Let $\{q(n)\}_{n\geq 1}$ be an ECFP process. Suppose Assumptions 3.1–3.2 hold. Then players learn a consensus Nash equilibrium strategy in the sense that $d(\bar{q}^N(n), CNE) \to 0$ as $n \to \infty$.*

We prove this result after briefly discussing some properties of the ECFP algorithm.

3.2.1 Information Overhead and Complexity in ECFP

In an ECFP process (3.1)–(3.2), the problem of demanding memory requirements is mitigated by requiring players to track only the centroid distribution, $\bar{q}(n)$, a vector whose memory size is invariant to the number of players in the game.

Likewise, a degree of mitigation in computational complexity is enabled by the introduction of the distribution $\bar{q}_{-i}(n)$ in the best-response computation. In the joint distribution $\bar{q}_{-i}(n)$, all players are assumed to use independent and identical mixed strategies. Analogous to the manner in which independent and identically distributed (i.i.d.) random variables always make life simpler in statistics, the introduction of the "i.i.d." joint distribution $\bar{q}_{-i}(n)$ enables simplifications in the ECFP best-response computation. (See Section 3.5.1 for a detailed illustration.)

In contrast to FP, ECFP, in general, admits reduced complexity best-response computation at the players, which is enabled by the "i.i.d." nature of the $\bar{q}_{-i}(n)$ used in the ECFP best-response computation (see (3.2)). This same factor may enable an explicit characterization of the distribution of certain essential *statistics* involved in the mixed utility computation in terms of parametric statistical families. The FP best response (2.9), in contrast, is based on an optimization involving the collection $q_{-i}(n)$ of individual empirical distributions; since the individual distributions are generally different, the collection $q_{-i}(n)$ does not admit a similar simplification or reduced parameterization as far as the best-response computation is concerned. For more detailed illustrations of the relative complexities of FP and ECFP best-response computations, we refer the reader to Section 3.5.1.

3.2.2 Proof of Convergence

We now prove Theorem 3.3 using relatively simple arguments that work directly with the discrete-time process. The proof technique is similar to that used in [34] for classical FP.[2] An alternative method of proof (and the one that we will resort to frequently in the later parts of the dissertation) is to study a continuous-time variant of the learning process, and prove convergence of the discrete-time process as a corollary of convergence of the continuous-time process. This method of proof requires the introduction of several tools from the theory of stochastic approximation and differential inclusions. We prefer to give a simpler proof of Theorem 3.3 now, and defer the use of these more involved arguments until Chapter 4 where it is necessary to introduce these tools to prove our robustness result for MBR processes. (In particular, see Section 4.5 for a proof of Theorem 3.3 using continuous-time arguments.)

Proof. Let $\bar{a}(n) = \frac{1}{N}\sum_{i=1}^N a_i(n)$, where $a_i(n) \in A_i$. Let $\bar{a}^N(n) \in \Delta^N$ be the N-tuple $(\bar{a}(n), \ldots, \bar{a}(n))$.

For $n \geq 1$ the empirical centroid $\bar{q}(n) = \sum_{s=1}^N a(s)$ may be expressed recursively as $\bar{q}^N(n+1) = \bar{q}^N(n) + \frac{1}{n+1}\left(\bar{a}^N(n+1) - \bar{q}^N(n)\right)$. Using this recursive form we may express the potential at $q(n+1)$ as $U(\bar{q}^N(n+1)) = U\left(\bar{q}^N(n) + \frac{1}{n+1}\left(\bar{a}^N(n+1) - \bar{q}^N(n)\right)\right)$. Using the multilinearity of U this gives

$$\begin{aligned} U(\bar{q}^N(n+1)) =& U(\bar{q}^N(n)) + \frac{1}{n+1}\sum_{i=1}^N U\left(\bar{a}_i(n+1), \bar{q}_{-i}(n)\right) \\ &- \frac{1}{n+1}\sum_{i=1}^N U(\bar{q}_i(n), \bar{q}_{-i}(n)) + \zeta(n+1), \end{aligned}$$

[2] In [104] this same proof technique is used to prove a stronger, "robust" version of this convergence result (see Theorem 1 in [104]) that allows us to develop a network-based implementation of ECFP. However, since the robustness results that will be presented in Chapter 4 will supplant the robustness result proved in [104], we focus here only on presenting convergence results for a simple (non-noisy) setting. Accordingly, we give here a simplified version of the proof in [104] with the aim of making the main ideas more transparent.

where we have explicitly written the first order terms of the expansion and collected the remaining terms in $\zeta(n+1)$. Note that the number of second order terms in the above expansion is finite and the terms are uniformly bounded since $\max_{\sigma \in \Delta^N} |U(\sigma)| < \infty$. Hence, there exists a positive constant M (independent of n) large enough such that $|\zeta(n+1)| \leq M(n+1)^{-2}$ for all n. Thus,

$$\begin{aligned} U(\bar{q}^N(n+1)) \geq & U(\bar{q}^N(n)) + \frac{1}{n+1} \sum_{i=1}^{N} U\left(\bar{a}_i(n+1), \bar{q}_{-i}(n)\right) \\ & - \frac{1}{n+1} \sum_{i=1}^{N} U(\bar{q}_i(n), \bar{q}_{-i}(n)) - \frac{M}{(n+1)^2}. \end{aligned} \tag{3.4}$$

The permutation invariance and multilinearity of $U(\cdot)$ permits the following rearranging of terms:

$$\begin{aligned} \sum_{i=1}^{N} U\left(\bar{a}_i(n+1), \bar{q}_{-i}(n)\right) &= \sum_{i=1}^{N} U\left(\left[\frac{1}{N}\sum_{j=1}^{N} a_j(n+1)\right]_i, \bar{q}_{-i}(n)\right) \\ &= \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{N} U\left(\left[a_j(n+1)\right]_i, \bar{q}_{-i}(n)\right) \\ &= \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{N} U\left(\left[a_j(n+1)\right]_j, \bar{q}_{-j}(n)\right) \\ &= \sum_{j=1}^{N} U\left(a_j(n+1), \bar{q}_{-j}(n)\right), \end{aligned}$$

where the notation $U([a_j(n)]_i, \bar{q}_{-i}(n))$ indicates the expected utility player i would receive were she to use the strategy $a_j(n)$ and all other players use the strategy $\bar{q}(n)$. Substituting this in (3.4) we get

$$\begin{aligned} U\left(\bar{q}^N(n+1)\right) - U\left(\bar{q}^N(n)\right) + \frac{M}{(n+1)^2} &\geq \frac{1}{n+1} \sum_{i=1}^{N} \left(U\left(a_i(n+1), \bar{q}_{-i}(n)\right) - U(\bar{q}_i(n), \bar{q}_{-i}(n))\right) \\ &\geq \frac{\alpha_{n+1}}{n+1} \end{aligned}$$

where $\alpha_{n+1} := \sum_{i=1}^{N} \left(U\left(a_i(n), \bar{q}_{-i}(n)\right) - U\left(\bar{q}_i(n), \bar{q}_{-i}(n)\right)\right)$. Note that $U(\cdot)$ is multilinear and therefore locally Lipschitz continuous. Summing over $1 \leq t \leq T$ above we get

$$U(\bar{q}^N(T+1)) - U(\bar{q}^N(1)) + \sum_{t=1}^{T} \frac{M}{(n+1)^2} \geq \sum_{t=1}^{T} \frac{\alpha_{n+1}}{n+1}.$$

Note that $\sum_{n=1}^{T} \frac{M}{(n+1)^2}$ is summable; therefore all terms on the left hand side are bounded above for all $T \geq 1$, and hence $\lim_{T\to\infty} \sum_{n=2}^{T} \frac{\alpha_n}{n} < \infty$. By Lemma 3.8 in the appendix we get $\lim\limits_{T\to\infty} \frac{\alpha_2+\ldots+\alpha_T}{T} = 0$. Subsequently, by Lemma 3.9 (see appendix), we get $\lim\limits_{T\to\infty} \frac{|\{1\leq t\leq T: \bar{q}^N(n)\notin C_\varepsilon\}|}{T} = 0$ for every $\varepsilon > 0$. By Lemma 3.10 (see appendix), this is equivalent to $\lim\limits_{T\to\infty} \frac{|\{1\leq t\leq T: \bar{q}^N(n)\notin B_\delta(C)\}|}{T} = 0$ for every $\delta > 0$. Finally, by Lemma 3.11 (see appendix), we obtain $d(\bar{q}^N(n), CNE) \to 0$ as $n \to \infty$. $\square$

We emphasize that Theorem 3.3 shows that the N-tuple of the *average* empirical distribution converges to CNE, that is, $d(\bar{q}^N(n), CNE) \to 0$. This is not the same as the more traditional definition of convergence in empirical frequency, (2.13). In the former, the N-tuple containing repeated copies of the empirical centroid converges to equilibrium, in the latter, the N-tuple containing the individual empirical distributions converges to equilibrium.

Note also that the set of limit points of ECFP is restricted to CNE—a subset of the NE. Thus, if Pareto superior Nash equilibria exist outside the set CNE, then ECFP will never reach these points, though an algorithm such as FP may. This may be seen as a tradeoff for the improvements in information overhead, complexity, and convergence rate achieved in ECFP.

3.3 Empirical Centroid FP: General Framework

The ECFP algorithm may be generalized to a setup where players track multiple centroids, each centroid corresponding to a different class of players. This generalization allows for ECFP to be considered in a broader class of games than permitted by Assumptions 3.1–3.2, and permits FP to be considered as special case of ECFP.

In the general version of ECFP, players are grouped into equivalence classes, or "permutation invariant" classes. This grouping allows players to analyze collective behavior by tracking only the statistics associated with each equivalence class, rather than tracking the statistics of every individual player.

Let $m \leq N$, denote the number of classes, let $I = \{1, \ldots, m\}$ be an index set, and let $\mathcal{C} = \{C_1, \ldots, C_m\}$ be a collection of subsets of $\mathcal{N}$; i.e. $C_k \subseteq \mathcal{N},\ \forall k \in I$. A collection $\mathcal{C}$ is said to be a *permutation-invariant partition* of $\mathcal{N}$ if,

(i) $C_k \cap C_\ell = \emptyset$, for $k, \ell \in I$, $k \neq \ell$,

(ii) $\bigcup\limits_{k\in I} C_k = \mathcal{N}$,

(iii) for $k \in I$, $i, j \in C_k$, $Y_i = Y_j$,

(iv) for $k \in I$, $i, j \in C_k$, there holds for any strategy profile $y = (y_i, y_j, y_{-(i,j)}) \in Y$,

$$u(y_i, y_j, y_{-(i,j)}) = u([y_j]_i, [y_i]_j, y_{-(i,j)}),$$

where the notation $([y_i]_j, [y_j]_i, y_{-(i,j)})$ indicates a permutation of (only) the strategies of players i and j in the strategy profile $y = (y_i, y_j, y_{-(i,j)})$.

For a collection $\mathcal{C}$, define $\phi(\cdot) : \mathcal{N} \to I$ to be the unique mapping such that $\phi(i) = k$ if and only if $i \in C_k$.

For $k \in I$, and $\sigma \in \Delta^N$, and permutation-invariant partition $\mathcal{C}$, define

$$\bar{\sigma}^k := |C_k|^{-1} \sum_{i \in C_k} \sigma_i \tag{3.5}$$

to be the *k-th centroid* with respect to $\mathcal{C}$, where $|C_k|$ denotes the cardinality of the set C_k. Likewise for $\sigma \in \Delta^N$ define

$$\bar{\sigma} := (\bar{\sigma}_1, \bar{\sigma}_2, \ldots, \bar{\sigma}_N), \tag{3.6}$$

where $\bar{\sigma}_i = \bar{\sigma}^{\phi(i)}$, to be the *centroid distribution* with respect to $\mathcal{C}$.

Given a permutation-invariant partition $\mathcal{C}$, let the set of symmetric Nash equilibria (relative to $\mathcal{C}$) be given by,

$$SNE := \{\sigma \in NE : \sigma_i = \sigma_j \ \forall i, j \in C_k, \ \forall \ k \in I\}, \tag{3.7}$$

and let the set of mean-centric equilibria (relative to $\mathcal{C}$) be given by,

$$MCE := \{\sigma \in \Delta^N : \ U(\sigma_i, \bar{\sigma}_{-i}) \geq U(\sigma_i', \bar{\sigma}_{-i}), \forall \sigma_i' \in \Delta_i, \ \forall i\}.$$

The set of MCE is neither a strict superset nor subset of the NE set—rather, it is a set of natural equilibrium points tailored to the ECFP dynamics. The set of SNE however, is contained in the set of MCE.

The sets of SNE and MCE relative to a partition $\mathcal{C}$, can be shown to be non-empty in potential games using fixed point arguments similar to [103, 105].

In ECFP, players do not track the empirical distribution of each individual player. Instead, they track only the centroid $\bar{q}^k(n)$ for each $k \in I$ (see (3.5)). Intuitively speaking, in ECFP each player i assumes (perhaps incorrectly) that for each class $C_k \in \mathcal{C}$ the centroid $\bar{q}^k(n)$ accurately represents the mixed strategy for all players $j \in C_k$. Each player i chooses her next-stage action as a myopic best response given this assumption.

Formally, the joint action at time $(n + 1)$ is chosen according to the rule[3]

$$a(n + 1) \in \mathrm{BR}(\bar{q}(n)), \tag{3.8}$$

[3]The action $a(1)$ may be chosen arbitrarily.

where $\bar{q}(n)$ is the centroid distribution associated with $q(n)$ (see (3.6)).

The empirical distribution is recursively defined by $q(n+1) = \left(1 - \frac{1}{n+1}\right) q(n) + \frac{1}{n+1} a(n+1)$. Combining this with (3.8) we get the following difference inclusion governing the behavior of $\{q(n)\}_{n\geq 1}$,

$$q(n+1) \in \left(1 - \frac{1}{n+1}\right) q(n) + \frac{1}{n+1} \mathrm{BR}(\bar{q}(n)). \tag{3.9}$$

Likewise, Lemma 3.14 (see appendix) shows that the sequence of centroid distributions $\{\bar{q}(n)\}_{n\geq 1}$ follows the difference inclusion,

$$\bar{q}(n+1) \in \left(1 - \frac{1}{n+1}\right) \bar{q}(n) + \frac{1}{n+1} \mathrm{BR}(\bar{q}(n)). \tag{3.10}$$

We refer to the sequences $(q(n))_{n\geq 1}$ and $(\bar{q}(n))_{n\geq 1}$ as ECFP processes.[4]

3.3.1 General ECFP Convergence Result

The following theorem is the main convergence result for ECFP—it states that, if Γ is a potential game, then players engaged in ECFP asymptotically learn elements of sets of SNE and MCE. Learning of MCE occurs in the sense that $d(q(n), MCE) \to 0$; this form of learning corresponds to the typical notion of setwise *convergence in empirical distribution* typical in classical FP (see Section 2.5 and [20, 74]). Learning of SNE occurs in the sense that $d(\bar{q}(n), SNE) \to 0$. This notion of learning, while similar in spirit to the typical notion of convergence in empirical distribution, differs in that it is the empirical centroid distribution (see (3.6)) that is converging to the set of SNE, rather than the empirical distribution itself.

Theorem 3.4. *Assume Γ is a potential game. Let $\mathcal{C}$ be a permutation-invariant partition of the player set $\mathcal{N}$. Let $(q(n), \bar{q}(n))_{n\geq 1}$ be an ECFP process. Then,*
(i) players learn a subset of the MCE in the sense that $\lim_{n\to\infty} d(q(n), MCE) = 0$,
(ii) players learn a subset of the SNE in the sense that $\lim_{n\to\infty} d(\bar{q}(n), SNE) = 0$.

Theorem 3.4 can be proved using similar reasoning to the proof of 3.3, and we omit this proof here for brevity. However, this theorem can also be proved by studying a continuous-time version of ECFP, and relating the limit sets of discrete-time ECFP to those of continuous-time ECFP. In Chapter 4 we will prove Theorem 3.4 in this manner after introducing the appropriate tools and notation (see Section 4.5).

[4]Sometimes we also simply refer to $(q(n), \bar{q}(n))_{n\geq 1}$ as an ECFP process.

3.4 Continuous-Time ECFP

In this section we consider continuous-time ECFP dynamics. Using the theory of stochastic approximation, the continuous-time process can be seen as a mean-field approximation of the discrete-time process [28]. In particular, a linear interpolation of a discrete-time ECFP process can be seen as a perturbed continuous-time ECFP path. In Chapter 4 we will show that convergence results for discrete-time ECFP can be proved by studying its continuous-time counterpart.

In Section 3.4.1 we present continuous-time ECFP dynamics. In Section 3.4.2 we briefly discuss the relationship between continuous-time and discrete-time ECFP. In Section 3.4.3 we show that continuous-time ECFP converges to the sets of SNE and MCE in a manner paralleling the convergence of the discrete-time process shown in Theorem 3.4. In Section 3.4.4 we study the rate of convergence of continuous-time ECFP.

3.4.1 Continuous-Time ECFP Process

Let Γ be a potential game and let $\mathcal{C}$ be a permutation-invariant partition of $\mathcal{N}$. In analogy to[5] (3.9), for $t \geq 0$ let

$$\dot{\mathbf{q}}(t) \in \mathrm{BR}(\bar{\mathbf{q}}(t)) - \mathbf{q}(t), \tag{3.11}$$

where where $\bar{\mathbf{q}}(t)$ is the centroid distribution associated with $\mathbf{q}(t)$ (see (3.6)). We refer to the processes $(\mathbf{q}(t))_{t\geq 0}$ and $(\bar{\mathbf{q}}(t))_{t\geq 0}$ as continuous-time ECFP processes.

As our end goal involves studying the limiting behavior of $\{\bar{\mathbf{q}}(t)\}_{t\geq 0}$, note that for $k \in I$, and $\bar{\mathbf{q}}^k(t)$ defined similar to (3.5), there holds

$$\begin{aligned}
\dot{\bar{\mathbf{q}}}^k(t) &= \frac{d}{dt}|C_k|^{-1}\sum_{j\in C_k}\mathbf{q}_j(t) \\
&= |C_k|^{-1}\sum_{j\in C_k}\frac{d}{dt}\mathbf{q}_j(t) \\
&= |C_k|^{-1}\sum_{j\in C_k}\dot{\mathbf{q}}_j(t),
\end{aligned}$$

Let $\mathbf{p}(t) := \dot{\mathbf{q}}(t) + \mathbf{q}(t)$, so that $\bar{\mathbf{p}}(t) = (\bar{\mathbf{p}}_1(t), \ldots, \bar{\mathbf{p}}_N(t))$ with $\bar{\mathbf{p}}_i(t) := \bar{\mathbf{p}}^{\phi(i)}(t)$ for $i \in \mathcal{N}$ (see (3.6)). By the above, and the linearity of differentiation, $\bar{\mathbf{p}}(t) = \dot{\bar{\mathbf{q}}}(t) + \bar{\mathbf{q}}(t)$. Thus, by Lemma 3.13,

[5] Note that (3.9) may be written as $q(n+1) - q(n) \in \frac{1}{n+1}\left(\mathrm{BR}_{\epsilon_n}(\bar{q}(n)) - q(n)\right)$.

(3.11) implies that $\bar{\mathbf{p}}(t) \in \mathrm{BR}(\bar{\mathbf{q}}(t))$, or equivalently,

$$\dot{\bar{\mathbf{q}}}(t) \in \mathrm{BR}(\bar{\mathbf{q}}(t)) - \bar{\mathbf{q}}(t). \tag{3.12}$$

3.4.2 Linear Interpolation of Discrete-Time ECFP

Discrete-time ECFP is closely related to continuous-time ECFP. In particular, linear interpolation of a discrete-time ECFP processes (3.9) (respectively, (3.10)) can be shown to generate a *perturbed solution* of the continuous-time ECFP differential inclusions (3.11) (respectively, (3.12)).

The linearly interpolated process is generated as follows. Set $\tau_0 = 0$ and $\tau_n = \sum_{i=1}^n \gamma(i)$ for $n \geq 1$. The continuous-time interpolation of discrete-time ECFP (3.9) is given by $\mathbf{w} : [0, \infty) \to \mathbb{R}^m$, where

$$\mathbf{w}(\tau_n + s) = q(n) + s\frac{q(n+1) - q(n)}{\tau_{n+1} - \tau_n}, \ s \in [0, \gamma(n+1)).$$

The linear interpolation of (3.10) is generated in a similar manner, substituting $\bar{q}(n)$ for $q(n)$, above.

In Chapter 4 (see also [28]) the notion of a perturbed solution is precisely defined and it is shown that the linear interpolation of (3.9) (respectively, (3.10)) is a perturbed solution of (3.11) (respectively, (3.12)).

3.4.3 Convergence in Continuous Time

We now show that continuous-time ECFP converges to the sets of SNE and MCE in a manner parallelling the convergence of discrete-time ECFP.

For any solution $\mathbf{q}(t)$ of (3.11) and associated centroid process $\bar{\mathbf{q}}(t)$, let $\mathbf{w}(t) := U(\bar{\mathbf{q}}(t))$ and let $\mathbf{v}(t) := \frac{1}{N}\sum_{i=1}^N U(\mathbf{q}_i(t), \bar{\mathbf{q}}_{-i}(t))$. There holds,

$$\begin{aligned}
\dot{\mathbf{w}}(t) &= \sum_{i=1}^N \frac{\partial}{\partial \bar{\mathbf{q}}_i} U(\bar{\mathbf{q}}(t))\dot{\bar{\mathbf{q}}}_i(t) \\
&\geq \sum_{i=1}^N \left[U(\dot{\bar{\mathbf{q}}}_i(t) + \bar{\mathbf{q}}_i(t), \bar{\mathbf{q}}_{-i}(t)) - U(\bar{\mathbf{q}}(t))\right] \\
&= \sum_{i=1}^N \left[\max_{\alpha_i \in \Delta_i} U(\alpha_i, \bar{\mathbf{q}}_{-i}(t)) - U(\bar{\mathbf{q}}(t))\right] \geq 0,
\end{aligned} \tag{3.13}$$

where the second line follows from the concavity of U in terms of σ_i (i.e., the strategy of player i) and the third follows from (3.12).

By Lemma 3.12 we get

$$\frac{1}{N}\sum_{i=1}^{N} U(\mathbf{q}_i(t), \bar{\mathbf{q}}_{-i}(t)) = U(\bar{\mathbf{q}}(t)). \tag{3.14}$$

Hence $\mathbf{v}(t) = \mathbf{w}(t)$, and we have $\dot{\mathbf{v}}(t) \geq 0$. The following expansion is useful in order to study convergence to the set of MCE.

$$\begin{aligned}
\dot{\mathbf{v}}(t) = \dot{\mathbf{w}}(t) &\geq \sum_{i=1}^{N}\left[\max_{\alpha_i\in\Delta_i} U(\alpha_i, \bar{\mathbf{q}}_{-i}(t)) - U(\bar{\mathbf{q}}(t))\right]\\
&= \sum_{i=1}^{N} \max_{\alpha_i\in\Delta_i} U(\alpha_i, \bar{\mathbf{q}}_{-i}(t)) - N\,U(\bar{\mathbf{q}}(t))\\
&= \sum_{i=1}^{N} \max_{\alpha_i\in\Delta_i} U(\alpha_i, \bar{\mathbf{q}}_{-i}(t)) - \sum_{i=1}^{N} U(\mathbf{q}_i(t), \bar{\mathbf{q}}_{-i}(t))\\
&= \sum_{i=1}^{N}\left[\max_{\alpha_i\in\Delta_i} U(\alpha_i, \bar{\mathbf{q}}_{-i}(t)) - U(\mathbf{q}_i(t), \bar{\mathbf{q}}_{-i}(t))\right] \geq 0,
\end{aligned} \tag{3.15}$$

where the inequality follows from (3.13), and the third line follows again from (3.14).

By (3.13), $\mathbf{w}(t)$ is weakly increasing, and is constant in a time interval T if and only if $\max_{\alpha_i\in\Delta_i} U(\alpha_i, \bar{\mathbf{q}}_{-i}(t)) - U(\bar{\mathbf{q}}(t)) = 0,\ \forall i$; i.e., if and only if $\bar{\mathbf{q}}(t) \in SNE$ for all $t \in T$.

By (3.15), $\dot{\mathbf{v}}(t)$ is weakly increasing, and $\dot{\mathbf{v}}(t) = 0$ in some interval T implies $\max_{\alpha_i\in\Delta_i} U(\alpha_i, \bar{\mathbf{q}}_{-i}(t)) - U(\mathbf{q}_i(t), \bar{\mathbf{q}}_{-i}(t)) = 0,\ \forall i \in \mathcal{N},\ t \in T$; i.e., $\mathbf{q}(t) \in MCE$ for all $t \in T$. Moreover, by Lemma 3.13, $\mathbf{q}(t) \in MCE,\ \forall t \in T \implies \bar{\mathbf{q}}(t) \in SNE\ \forall t \in T$, which by the above comments implies $\dot{\mathbf{w}}(t) = 0$ in T, or equivalently $\dot{\mathbf{v}}(t) = 0$ in T. Thus, $\dot{\mathbf{v}}(t)$ is constant in a time interval T if and only if $\mathbf{q}(t) \in MCE$ for all $t \in T$.

Proposition 3.5. *Assume Γ is a potential game. Then,*
(i) The limit set of every solution of (3.12) *is a connected subset of SNE along which U is constant;*
(ii) For $\sigma \in \Delta^N$, let $V(\sigma) := \frac{1}{N}\sum_{i=1}^{N} U(\sigma_i, \bar{\sigma}_{-i})$. The limit set of every solution of (3.11) *is a connected subset of MCE along which V is constant.*

The proof of this proposition follows from the above comments.

3.4.4 Finite-Time Convergence in ECFP

In the above, we showed that the ECFP process $(\bar{\mathbf{q}}(t))_{t\geq 0}$ converges to the set of SNE. This set tends to contain mixed strategy equilibria—for example, Assumption 6.19 gives a condition under which any consensus equilibrium will be a mixed strategy (within the appropriate class of congestion games). The following result shows that whenever ECFP converges to a (completely) mixed equilibrium, the convergence tends to occur in finite time.[6]

Theorem 3.6. *Let* Γ *be a potential game satisfying Assumption 3.1–3.2. Suppose* q^* *is a completely-mixed SNE of* Γ *and that the Hessian of* U *invertible at* q^**. If an ECFP process* $(\bar{\mathbf{q}}(t))_{t\geq 0}$ *converges to* q^**, then it does so in finite time.*

We delay presenting the proof of this result until Chapter 10 where we study fundamental properties of MBR algorithms (see Section 10.6 and, in particular, Proposition 10.23).

We note that in the above theorem we assumed that the Hessian of the potential function is invertible at q^*. In Section 9.3 we show that in almost all potential games, the Hessian is invertible at every completely mixed NE strategy. Using similar arguments, it may be possible to show that the same property holds for almost all games within the subset of potential games with permutation invariant potential function relative to $\mathcal{C}$.

The above theorem shows that *continuous-time* ECFP often achieves finite-time convergence. While continuous-time ECFP always converges to (completely) mixed equilibria in finite time, discrete-time ECFP does not directly inherit this property. The discrete-time process is an perturbed solution of the continuous-time process (see Section 4.2) and hence tends to approach *the neighborhood* of an equilibrium quickly. However, once near the equilibrium, the discrete-time process tends overshoot the equilibrium, oscillating in a neighborhood of it. The size of the oscillations correspond to the step-size of the process (in this case, $\frac{1}{n}$). As we will show in Chapter 4, the step-size sequence in ECFP can be altered, but must converge to zero and constitute a divergent series in order to maintain the convergence result (see Assumption 2.2). Hence, we cannot *asymptotically* decrease the step size much faster than $\frac{1}{n}$. A good heuristic method for approximating equilibria quickly may be to use a quickly diminishing step size in the early stages of the algorithm and switch to a step size of $\frac{1}{n}$ in the later stages. Alternatively, one could use a small constant step size as a heuristic for approximating equilibria quickly.

3.5 Illustrative Example

In this section we consider the implementation of ECFP in a traffic routing example.

Suppose a group of vehicles wishes to navigate from a common starting point to a common destination over a set of parallel routes. Let $\mathcal{R}$ denote the set of routes available to all players. Let

[6]The theorem is stated for *completely* mixed equilibria. However, in Section 10.3 we show that incompletely mixed equilibria can be handled in a similar manner by projecting to a lower dimensional game. For simplicity, we focus here only on the completely mixed case.

this be cast as a game with the set of players $\mathcal{N}$ given by the set of drivers, and the action space of each player i given by $Y_i = \mathcal{R}$.

Each player's objective is to minimize their personal travel time. The cost of travelling on any given road $r \in \mathcal{R}$ is dependent on the number of vehicles on the route. Given an action $y \in Y$, and $r \in \mathcal{R}$, let $\rho_r(y)$ give the number of players on route r under the strategy y. The cost of using route r if players use strategy $y \in Y$ is given by

$$c_r(y) := \alpha_{r,2}\rho(y)^2 + \alpha_{r,1}\rho(y) + \alpha_{r,0},$$

where $\alpha_{r,2}$, $\alpha_{r,1}$, and $\alpha_{r,0}$ are some predetermined coefficients. In the game theoretic framework, we let the utility received by player i under action profile $y = (y_i, y_{-i}) \in Y$ be given by

$$u_i(y) := -c_{y_i}(\rho(y)).$$

This game is an instance of a congestion game, which is known to be a type of potential game [38].

3.5.1 Best-Response Computation

The inherent symmetry in the distribution $\bar{q}_{-i}(n)$ used in the ECFP best-response calculation (3.2) leads to reductions in computational complexity, particularly when compared with the FP best-response calculation (2.9). We contrast the FP and ECFP best-response computations for the case of the cognitive radio game.

ECFP Best-Response Computation

In order to choose a best response in ECFP, a player must compute[7]

$$\arg\max_{y_i \in Y_i} E_{\bar{q}_{-i}(n)}[u(y_i, y_{-i})],$$

where y_{-i} is a random variable with distribution $\bar{q}_{-i}(n)$, and where $\bar{q}_{-i}(n) = (\bar{q}_{-i}(n), \ldots, \bar{q}_{-i}(n)) \in \Delta_{-i}$ is an $(n-1)$-tuple. The symmetry in the game allows for the following simplification,

$$E_{\bar{q}_{-i}(n)}[u(y_i, y_{-i})] = \sum_{k=0}^{N-1} c_{y_i}(k+1)\mathbb{P}_{\bar{q}_{-i}(n)}(\rho_{y_i}(y_{-i}) = k). \tag{3.16}$$

[7] In the ECFP best response (3.2), a player must maximize the mixed utility $\max_{\alpha_i \in A_i} U(\alpha_i, \bar{q}_{-i}(n))$. Recall that the mixed utility (2.1) is the expected value of $u(\cdot)$ given that players are using probabilistic strategies $\bar{q}_{-i}(n)$. Thus, maximizing the mixed utility of (3.2) is equivalent to maximizing the expected value below.

In the above, players only need to compute the probability associated with having k users on each channel rather than computing the probability of every possible configuration of users.

From here, the symmetry in the i.i.d. distribution $\bar{q}_{-i}(n)$ allows for further simplifications. Let $y_i = r \in \mathcal{R}$ and note that[8]

$$\begin{aligned}
&\mathbb{P}_{\bar{q}_{-i}(n)}(\rho_r(y_{-i}) = 0) = (1 - \bar{q}(n,r))^{N-1} \\
&\mathbb{P}_{\bar{q}_{-i}(n)}(\rho_r(y_{-i}) = 1) = (N-1)\bar{q}(n,r)(1 - \bar{q}(n,r))^{N-2} \\
&\mathbb{P}_{\bar{q}_{-i}(n)}(\rho_r(y_{-i}) = 2) = \binom{N-1}{2}\bar{q}(n,r)^2(1 - \bar{q}(n,r))^{N-3}.
\end{aligned}$$

As the above pattern suggests, the probability is binomial—for $0 \leq k \leq (N-1)$,

$$\mathbb{P}_{\bar{q}_{-i}(n)}(\rho_{y_i}(y_{-i}) = k) = \binom{N-1}{k}\bar{q}(n,r)^k(1 - \bar{q}(n,r))^{N-1-k}.$$

Thus, the requisite probability is given by a computationally simple closed form expression, and the expected utility can be easily computed using (3.16). Furthermore, since players compute a best response each iteration, they reap the computational benefits of using this simplified form for $\mathbb{P}_{q_{-i}(n)}(\rho_{y_i}(y_{-i}) = k)$ on each iteration of ECFP.

FP Best-Response Computation

In order to choose a best response in FP, a player must compute

$$\arg\max_{y_i \in Y_i} E_{q_{-i}(n)}[u(y_i, y_{-i})],$$

where y_{-i} is a random variable with distribution $q_{-i}(n)$. As before, the symmetry in the game allows for a simplification to

$$E_{q_{-i}(n)}[u(y_i, y_{-i})] = \sum_{k=0}^{N-1} c_{y_i}(k+1)\mathbb{P}_{q_{-i}(n)}(\rho_{y_i}(y_{-i}) = k).$$

[8] Recall that the notation $\bar{q}(n,r)$ refers to the r-th element of the vector $\bar{q}(n)$. In ECFP, players (incorrectly) assume that all opponents are independently using the identical mixed strategy $\bar{q}(n)$. Under this assumption, the probability of any given opponent using channel r is given by $\bar{q}(n,r)$.

However, due to the lack of structure in $q_{-i}(n)$, no further simplifications are possible. To illustrate the complexity of this, let $y_i = r \in \mathcal{R}$ and note that

$$\mathbb{P}_{q_{-i}(n)}(\rho_r(y_{-i}) = 0) = \prod_{j\neq i}(1 - q_j(n,r))^{N-1}$$

$$\mathbb{P}_{q_{-i}(n)}(\rho_r(y_{-i}) = 1) = \sum_{j_1\neq i} q_{j_1}(n,r) \prod_{j_2\neq i,j_1}(1 - q_{j_2}(n,r))^{N-2}$$

$$\mathbb{P}_{q_{-i}(n)}(\rho_r(y_{-i}) = 2) = \sum_{j_1\neq i}\sum_{j_2>j_1} q_{j_1}(n,r)q_{j_2}(n,r) \prod_{j_3\neq i,j_1,j_2}(1 - q_{j_3}(n,r))^{N-3}.$$

In general, when the $q(n)$ corresponds to a mixed strategy, the complexity of evaluating $\mathbb{P}_{q_{-i}(n)}(\rho_r(y_{-i}) = k)$ grows combinatorially with k—even in this game with symmetric payoffs.

3.5.2 Simulation Results

We first simulated a fixed game with 50 users and randomly chosen cost coefficients. Using a fixed initial condition $q(1) = q_0$, we ran instantiations of ECFP with 1, 2, 3, and 5 player groupings for 10^6 iterations in order to approximate the limiting SNE point for each instantiation of the algorithm.

Figure 3.1 plots the distance from the centroid distribution at iteration n (i.e., $\bar{q}(n)$, see (3.6)) to the limiting SNE point, denoted as q^*, in each instantiation. Since the game satisfies Assumptions 3.1–3.2, the set of consensus NE is nonempty. When players are grouped into a single class, ECFP approaches a consensus equilibrium. Let $\bar{q}^*$ denote the consensus equilibrium approached by the instantiation with one player grouping. Figure 3.2 plots, for each instantiation, the distance from the centroid distribution $\bar{q}(n)$ to the consensus equilibrium $\bar{q}^*$. Figures 3.1 and 3.2 show that in each instantiation, $\bar{q}(n)$ tends to approach the consensus equilibrium $\bar{q}^*$ before breaking away and moving towards the limiting SNE point q^*.

Figure 3.3 plots, for each instantiation, the worst case utility of any player under the mixed strategy $\bar{q}(n)$. The figure shows that the utility tends to decrease along the path of $\bar{q}(n)$. (We note also that the utility of $\bar{q}(n)$ is similar for all groups during the phase in which $\bar{q}(n)$ approaches the consensus equilibrium $\bar{q}^*$.)

As noted in Section 3.1, the behavior of discrete-time ECFP (DT-ECFP) is closely related to that of continuous-time ECFP (CT-ECFP). Due to the discontinuous nature of the best response, simulations of CT-ECFP are prone to numerical instability using traditional Runge-Kutta methods. However, the behavior of CT-ECFP may be approximated by considering the linear interpolation of DT-ECFP (see Section 3.4.2). Letting $(\mathbf{w}(t))_{t\geq 0}$ denote the continuous-time linear interpolation

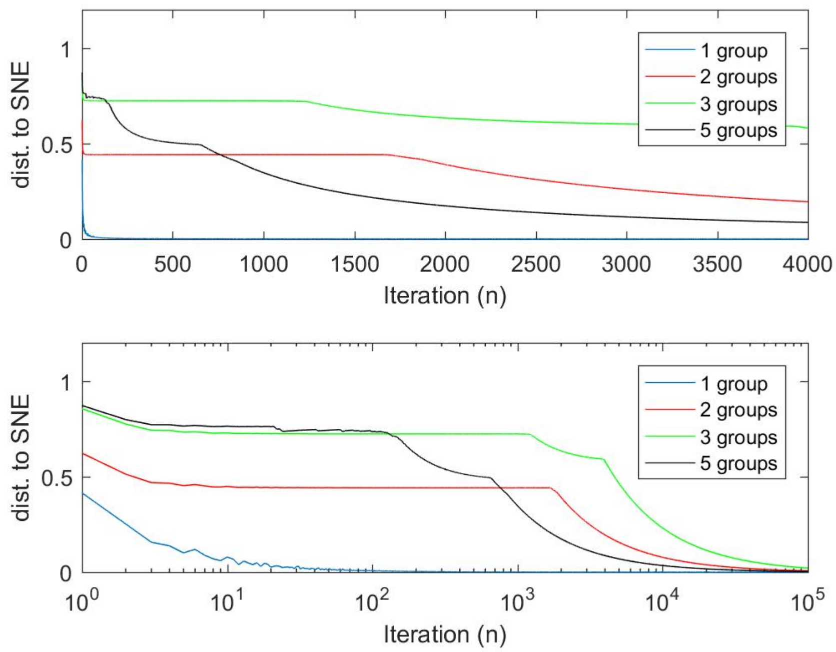


Figure 3.1: The distance from $\bar{q}(n)$ to the limiting SNE point q^*, using a linear time scale and a logarithmic time scale.

of a the discrete-time ECFP process $(\bar{q}(n))_{n\geq 1}$ for any given instantiation, Figure 3.4 shows the distance from $\mathbf{w}(t)$ to the limiting SNE in each instantiation, where we consider the same fixed game and initial condition as before. Figure 3.5 shows the distance from $\mathbf{w}(t)$ to the consensus equilibrium $\bar{q}^*$. In each instantiation, the continuous-time interpolation tends to approach the consensus equilibrium $\bar{q}^*$ and hover nearby $\bar{q}^*$ for a time before moving towards the final limiting SNE point. We hypothesize that the effect of hovering nearby $\bar{q}^*$ is an artifact of the discretization inherent in the interpolation of the discrete-time process, and not typical of the pure continuous-time dynamics.

In order to study how the rate of convergence scales in terms of the number of players we ran instantiations of ECFP with a single player grouping in a game with 50 players, a game with 200 players, and a game with 400 players. Figures 3.6 plots the distance from $\bar{q}(n)$ to the consensus equilibrium $\bar{q}^*$ in each case. The figure shows that the rate of convergence tends to be invariant to the number of players.

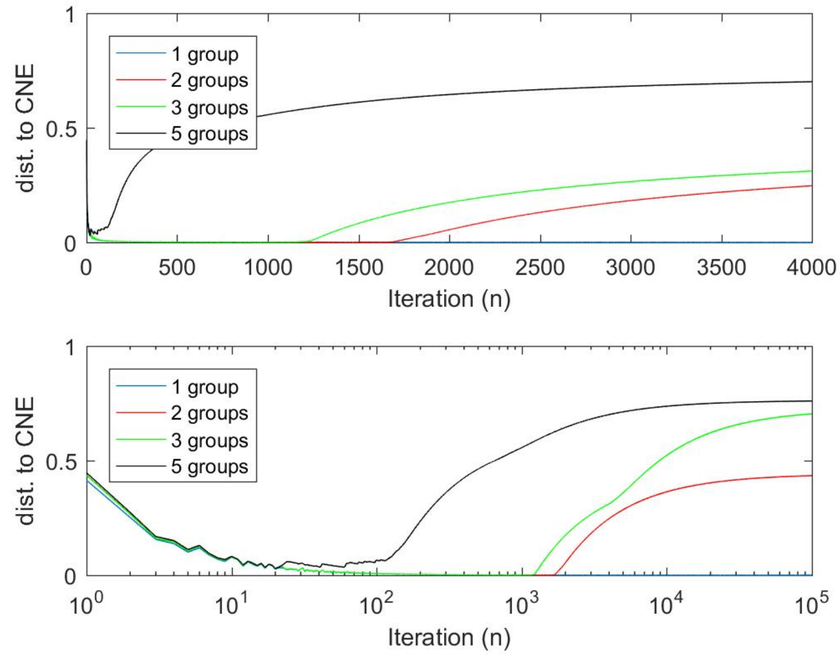


Figure 3.2: The distance from $\bar{q}(n)$ to the $\bar{q}^*$ point using a linear time scale and a logarithmic time scale.

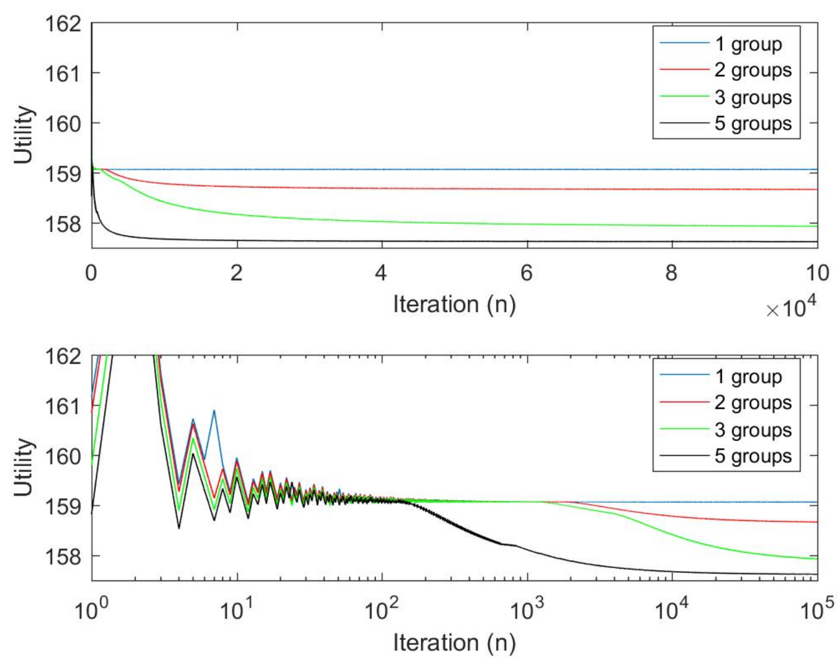


Figure 3.3: Plot of worst case utility in each instantiation of ECFP using a linear time scale and a logarithmic time scale.

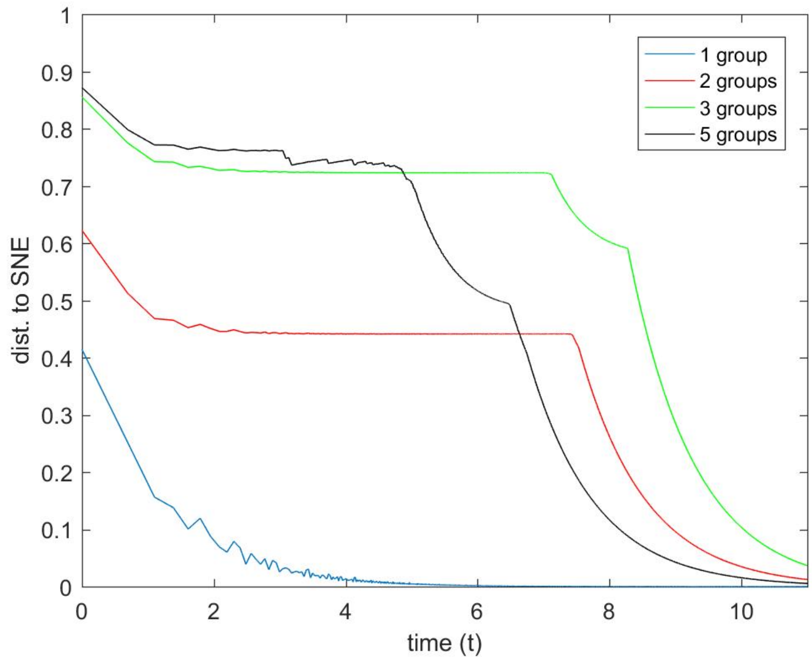


Figure 3.4: The distance from the interpolated process $\mathbf{w}(t)$ to the limiting SNE point.

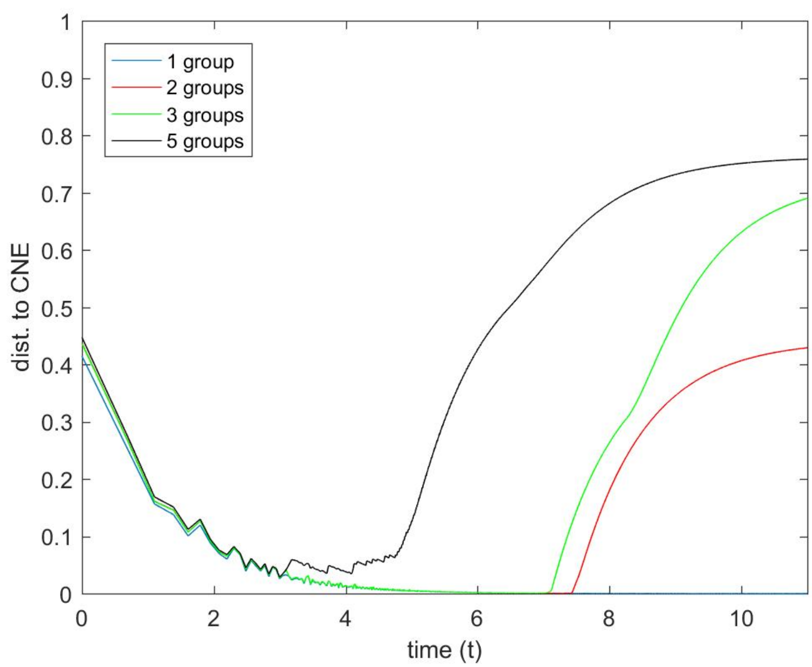


Figure 3.5: The distance from the interpolated process $\mathbf{w}(t)$ to the consensus equilibrium $\bar{q}^*$.

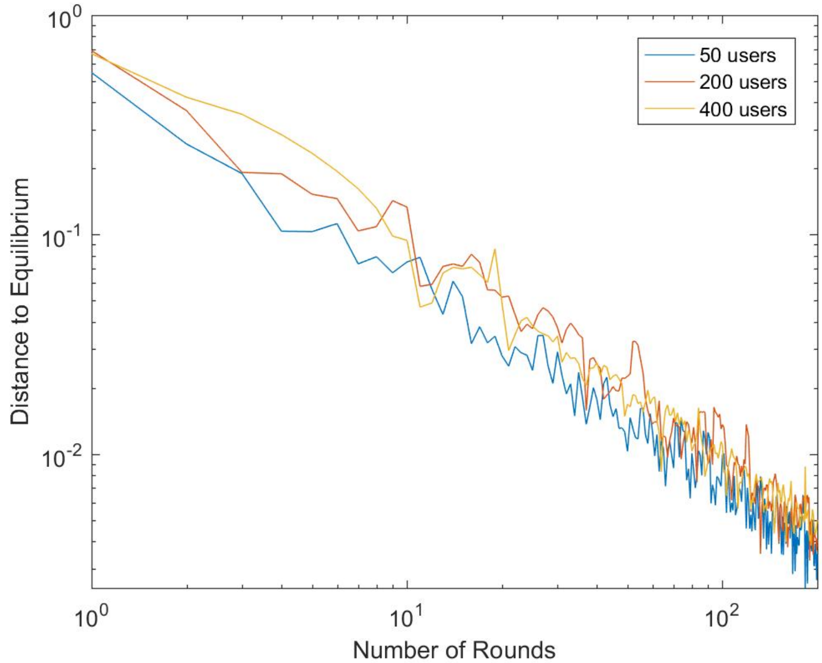


Figure 3.6: The distance from $\bar{q}(n)$ to the consensus equilibrium $\bar{q}^*$ in games with 50, 200, and 400 users.

3.6 Appendix to Chapter 3

Definition 3.7. *For $\epsilon > 0$, the set of ϵ-NE is given by*

$$NE_\epsilon := \{\sigma \in \Delta^N : \ \sigma \in \mathrm{BR}_\epsilon(\sigma)\},$$

and the set of ϵ-consensus NE is given by

$$CNE_\epsilon := \{\sigma \in NE : \sigma_1 = \ldots = \sigma_N\}.$$

Lemma 3.8. *Suppose the sum $-\infty < \sum_{n=1}^{\infty} \frac{a_n}{n} < \infty$ converges, then $\lim_{T\to\infty} \frac{a_1+a_2+\ldots+a_T}{T} = 0$.*

Proof. By Kronecker's Lemma [106], $-\infty < \sum_{k=1}^{\infty} \frac{a_k}{k} < \infty \Rightarrow \lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} k \frac{a_k}{k} = 0$, which implies that $\lim_{T\to\infty} \frac{a_1+\ldots+a_T}{T} = 0$. □

Lemma 3.9. *Let $a_n := \sum_{i=1}^{N} \left[v_i^m(\bar{q}(n)) - U(\bar{q}^N(n))\right]$, then $\lim_{T\to\infty} \frac{a_1+\ldots+a_T}{T} = 0$ implies that, for every*

$\varepsilon > 0$,

$$\lim_{T\to\infty} \frac{|\{1 \le n \le T:\ \bar{q}^N(n) \notin CNE_\varepsilon\}|}{T} = 0.$$

Proof. Let $\varepsilon > 0$ be given. By definition,

$$\bar{q}^N(n) \in CNE_\varepsilon \iff v_i^m(\bar{q}(n)) - U(\bar{q}^N(n)) < \varepsilon \ \forall i. \tag{3.17}$$

The utility function $U(\cdot)$ is assumed to be permutation invariant for all players, so an equivalent statement to (3.17) is, $\bar{q}^N(n) \in CNE_\varepsilon \iff \sum_{i=1}^{N} \left[v_i^m(\bar{q}(n)) - U(\bar{q}^N(n))\right] < N\varepsilon$. For $n \in \mathbb{N}$, let $b_n = 1$ if $a_n \ge n\varepsilon$ and let $b_n = 0$ otherwise. Note that $b_n = 0 \iff \bar{q}^N(n) \in CNE_\varepsilon$ and $b_n = 1 \iff \bar{q}^N(n) \notin CNE_\varepsilon$, thus

$$\frac{|\{1 \le n \le T : \bar{q}^N(n) \notin CNE_\varepsilon\}|}{T} = \frac{b_1 + \ldots + b_T}{T}.$$

Note also that $a_n \ge 0$. Clearly,

$$\frac{b_1 + \ldots + b_T}{T} \le \frac{1}{N\varepsilon} \frac{a_1 + \ldots + a_T}{T},$$

implying $\lim_{T\to\infty} \frac{b_1+\ldots+b_T}{T} = 0$, from which the desired result follows. □

Lemma 3.10. $\lim_{T\to\infty} \frac{|\{1\le n\le T:\bar{q}^N(n)\notin CNE_\varepsilon\}|}{T} = 0$ *for all* $\varepsilon > 0$ *implies that* $\lim_{T\to\infty} \frac{|\{1\le n\le T:\bar{q}^N(n)\notin B_\delta(CNE)\}|}{T} = 0$ *for all* $\delta > 0$.

Proof. Suppose $\lim_{T\to\infty} \frac{|\{1\le n\le T:\bar{q}^N(n)\notin CNE_\varepsilon\}|}{T} = 0$ for all $\varepsilon > 0$, but there exists some $\delta > 0$ such that $\limsup_{T\to\infty} \frac{|\{1\le n\le T:\bar{q}^N(n)\notin B_\delta(CNE)\}|}{T} = \alpha > 0$. Then there exists an $\varepsilon' > 0$ such that

$$\bar{q}^N(n) \notin B_\delta(CNE) \Rightarrow \bar{q}^N(n) \notin CNE_{\varepsilon'},$$

which implies that

$$|\{1 \le n \le T : \bar{q}^N(n) \notin CNE_{\varepsilon'}\}| \ge |\{1 \le n \le T : \bar{q}^N(n) \notin B_\delta(CNE)\}|.$$

This implies that

$$\limsup_{T\to\infty} \frac{|\{1 \le n \le T : \bar{q}^N(n) \notin CNE_{\varepsilon'}\}|}{T} \ge \alpha$$

for some $\varepsilon' > 0$, a contradiction. □

Lemma 3.11. $\lim_{T\to\infty} \frac{|\{1\leq n\leq T: \bar{q}^N(n)\notin B_\delta(CNE)\}|}{T} = 0$ *for all* $\delta > 0$ *implies* $\lim_{n\to\infty} d(\bar{q}^N(n), CNE) = 0.$

The proof of this result closely follows the proof of [34], Lemma 1, and is omitted here for brevity.

Lemma 3.12. *Let* $\mathcal{C}$ *be a partition of* $\mathcal{N}$*, and for* $\sigma \in \Delta^N$ *let* $\bar{\sigma}$ *be as defined in* (3.6). *Then* $\frac{1}{N}\sum_{i=1}^N U(p_i, \bar{p}_{-i}) = U(\bar{\sigma})$.

Proof. Let I be an index set for $\mathcal{C}$ and let m be the cardinality of I. For $k \in I$, and $j \in C_k$ note that

$$\begin{aligned} |C_k|U(\bar{\sigma}) = |C_k|U(\bar{\sigma}_j, \bar{\sigma}_{-j}) &= |C_k|U(\left[|C_k|^{-1}\sum_{i\in C_k}\sigma_i\right]_j, \bar{\sigma}_{-j}) \\ &= \sum_{i\in C_k} U([\sigma_i]_j, \bar{\sigma}_{-j}) \\ &= \sum_{i\in C_k} U(\sigma_i, \bar{\sigma}_{-i}), \end{aligned}$$

where the second line follow from the definition of $\bar{\sigma}_j$ (see (3.6)), the third by multilinearity of U, and the fourth by permutation invariance of elements in C_k. Thus,

$$\frac{1}{N}\sum_{i=1}^N U(\sigma_i, \bar{\sigma}_{-i}) = \frac{1}{N}\sum_{k\in I}\sum_{i\in C_k} U(\sigma_i, \bar{\sigma}_{-i}) = \frac{1}{N}\sum_{k\in I}|C_k|U(\bar{\sigma}) = \frac{1}{N}\sum_{i=1}^N U(\bar{\sigma}) = U(\bar{\sigma}).$$

□

Lemma 3.13. *Let* $q \in \Delta^N$*, let* $\bar{q}$ *be as defined in* (3.6), *and let* $\epsilon \geq 0$. *If* $\sigma \in \mathrm{BR}_\epsilon(\bar{q})$*, then* $\bar{\sigma} \in \mathrm{BR}_\epsilon(\bar{q})$.

Proof. Let $i \in \mathcal{N}$. Recall that $\bar{\sigma} := (\sigma_1, \ldots, \sigma_N)$ with $\sigma_i = \sigma^{\phi(i)}$. There holds

$$\begin{aligned}
U(\bar{\sigma}_i, \bar{q}_{-i}) &= U(\left[|C_{\phi(i)}|^{-1} \sum_{j \in C_{\phi(i)}} \sigma_j\right]_i, \bar{q}_{-i}) \\
&= |C_{\phi(i)}|^{-1} \sum_{j \in C_{\phi(i)}} U([\sigma_j]_i, \bar{q}_{-i}) \\
&= |C_{\phi(i)}|^{-1} \sum_{j \in C_{\phi(i)}} U([\sigma_j]_i, [\bar{q}^{\phi(i)}]_j, \bar{q}_{-(i,j)}) \\
&= |C_{\phi(i)}|^{-1} \sum_{j \in C_{\phi(i)}} U([\sigma_j]_j, [\bar{q}^{\phi(i)}]_i, \bar{q}_{-(i,j)}) \\
&= |C_{\phi(i)}|^{-1} \sum_{j \in C_{\phi(i)}} U([\sigma_j]_j, \bar{q}_{-j}) \\
&\geq |C_{\phi(i)}|^{-1} \sum_{j \in C_{\phi(i)}} \max_{\sigma_j' \in \Delta_j} (U(\alpha_j, \bar{q}_{-j}) - \epsilon) \\
&= |C_{\phi(i)}|^{-1} \sum_{j \in C_{\phi(i)}} \max_{\sigma_i' \in \Delta_i} (U(\alpha_i, \bar{q}_{-i}) - \epsilon) \\
&= \max_{\alpha_i \in \Delta_i} U(\alpha_i, \bar{q}_{-i}) - \epsilon,
\end{aligned}$$

where the first line follows by the definition of $\bar{\sigma}_i$ (see (3.6)), the second from the multilinearity of U, the fourth by permutation invariance of elements of $C_{\phi(i)}$, the sixth by the fact that, by hypothesis, $\sigma_j \in \mathrm{BR}_{j,\epsilon}(\bar{q}_{-j})$, and the seventh by permutation invariance of elements of $C_{\phi(i)}$. Since this holds for all $i \in \mathcal{N}$, it follows that $\bar{\sigma} \in \mathrm{BR}_\epsilon(\bar{q})$. $\square$

Lemma 3.14. *Assume Γ is a potential game and suppose the action sequence $\{a(n)\}_{n \geq 1}$ is chosen according to* (3.8)*. Then centroid process $\{\bar{q}(n)\}_{n \geq 1}$ follows the differential inclusion* (3.10)*.*

Proof. Note that $\bar{q}(n+1)$ may be written recursively as

$$\bar{q}(n+1) = \bar{q}(n) + \frac{1}{n+1} \left(\bar{a}(n+1) - \bar{q}(n) \right).$$

By Lemma 3.13, $a(n+1) \in \mathrm{BR}_{\epsilon_n}(\bar{q}(n))$ implies $\bar{a}(n+1) \in \mathrm{BR}_{\epsilon_n}(\bar{q}(n))$. Substituting this into the above recursion and rearranging terms shows that $\{\bar{q}(n)\}$ follows the difference inclusion (3.10). $\square$

Chapter 4

Robustness in Myopic Best-Response Algorithms

4.1 Introduction

In the canonical MBR algorithm presented in Section 2.4, we assumed that rather idealistic conditions hold. For example, players are assumed to act in perfect synchrony, be capable of perfectly computing the best response in each stage, and are assumed to have instantaneous access to all information required to compute the best response. Such assumptions are often extremely impractical—particularly, in large-scale distributed settings. This motivates the study of the robustness of learning results to perturbations occurring in practical real-world scenarios.

The main theoretical contribution of this chapter is to show that MBR algorithms are robust in the presence of certain perturbations, common in real-world applications [107]. In particular, suppose that the myopic best response is perturbed so that players may sometimes choose suboptimal actions, but that the degree of suboptimality decays to zero over time. (In the spirit of [72], we sometimes call a MBR process that is perturbed in this manner a *weakened MBR process*.) We show that the fundamental learning property of a MBR algorithm is retained in the presence of such perturbations. For example, in the case of classical FP, this means that convergence to NE is preserved in the presence of such perturbations. More generally, if a MBR algorithm converges to some equilibrium set in the absence of perturbations, then our result can be applied to prove convergence to the same equilibrium set in the presence of perturbations.

Robustness results of this kind were first studied in [29] for the case of classical FP. The chapter extends the approach of [29] to demonstrate robustness of general MBR algorithms. This greatly enhances the applicability of game-theoretical learning theory to real-world control problems. Moreover, the results of this chapter have required the development of useful new technical tools. For example, Lemma 4.16 studies ϵ-best-response sequences and demonstrates that such sequences may in fact be considered in terms of a more amenable sequence of, so called, δ-perturbations (see Section 4.2). Sequences of δ-perturbations are naturally handled within a stochastic approximation framework using techniques such as those developed in [28].

In the subsequent chapters we will consider three applications of the robustness result developed in this chapter.

As a first application, in Chapter 5 we study a sampling-based method for reducing the complexity of MBR algorithms using the robustness property. As discussed in Sections 1.3.2 and 2.6, computation of the best response at each iteration of a MBR algorithm can be computationally demanding. Using the robustness result we show that sampling-based MBR algorithms are able to achieve convergence with significantly reduced computational cost.

As a second application, in Chapter 6 we consider the problem of asynchronous implementation of a MBR algorithm. In many game-theoretic learning algorithms, it is assumed that players act in a perfectly synchronous manner. This assumption is unrealistic in large-scale distributed scenarios where players do not have access to a global clock. We study a practical variant of FP where players are permitted to choose actions in an asynchronous manner, and derive a mild condition under which convergence can be shown to occur. The proofs of these results follow as a simple consequence of the robustness result, and do not require the use of additional stochastic approximation techniques.

As a third application, in Chapter 7 we study the problem of implementing a MBR algorithm in a distributed setting. In traditional implementations of MBR algorithms it is assumed that players have instantaneous access to the information required to generate their forecast. However, in practical scenarios this information may often be distributed among the agents and must be disseminated using an overlaid communication graph. We present a generic method for implementing a MBR algorithm in this setting, and we show that convergence of such an algorithm can be ensured as a consequence of the robustness result.[1]

The remainder of the chapter is organized as follows. Section 4.2 sets up the mathematical tools to be used in the proof of the main theoretical result. Section 4.3 presents the notion of a *weakened* myopic best-response process. Section 4.4 presents the robustness result. Section 4.5 uses the robustness result to study the convergence and robustness of the ECFP algorithm presented in Chapter 3.

[1]We remark that while these applications are interesting in and of themselves, additional utility may be gained by considering them in conjunction with one another. For example, the first application allows for *synchronous* distributed implementation and the second allows for generic asynchronous implementation. Together, they allow one to study *asynchronous* distributed implementation of a MBR algorithm, using, for example, asynchronous gossip [21] as a means of disseminating information amongst agents.

4.2 Difference Inclusions and Differential Inclusions

In this section we introduce the mathematical tools necessary to prove our robustness result for MBR processes. In particular, in Section 4.4 we will study the limiting behavior of a discrete-time MBR processes using the stochastic approximation techniques introduced in this section.

Following the approach of [28], let $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ denote a set-valued function mapping each point $\xi \in \mathbb{R}^m$ to a set $F(\xi) \subseteq \mathbb{R}^m$. We assume:

Assumption 4.1.
(i) F is a closed set-valued map; i.e., Graph(F) $:= \{(\xi, \eta) : \eta \in F(\xi)\}$ is a closed subset of $\mathbb{R}^m \times \mathbb{R}^m$.
(ii) $F(\xi)$ is a nonempty compact convex subset of $\mathbb{R}^m$, for all $\xi \in \mathbb{R}^m$.
(iii) For some norm $\|\cdot\|$ on $\mathbb{R}^m$, there exists a $c > 0$ such that for all $\xi \in \mathbb{R}^m$, $\sup_{\eta \in F(\xi)} \|\eta\| \leq c(1 + \|\xi\|)$.

Definition 4.2. *A solution for the differential inclusion*

$$\frac{d\mathbf{x}(t)}{dt} \in F(\mathbf{x}(t)) \tag{4.1}$$

with initial point $\xi \in \mathbb{R}^m$ is an absolutely continuous mapping $\mathbf{x} : \mathbb{R} \to \mathbb{R}^m$ such that $\mathbf{x}(0) = \xi$ and (4.1) *holds for almost every $t \in \mathbb{R}$.*

In order to study the asymptotic behavior of discrete-time processes in this context, it is useful to consider the notion of an interpolated process. Formally, given a step-size sequence $\{\gamma(n)\}_{n\geq 1}$, we define the continuous-time interpolation of a discrete-time process as follows.

Definition 4.3. *Let $\{x(n)\}_{n\geq 1}$ be a sequence in $\mathbb{R}^m$. Set $\tau_0 = 0$ and $\tau_n = \sum_{i=1}^n \gamma(i)$ for $n \geq 1$ and define the continuous-time interpolation of $\{x(n)\}_{n\geq 1}$ to be the process $w : [0, \infty) \to \mathbb{R}^m$ satisfying*

$$w(\tau_n + s) = x(n) + s\frac{x(n+1) - x(n)}{\tau_{n+1} - \tau_n}, \;\; s \in [0, \gamma(n+1)).$$

In general, we will be interested in studying discrete-time processes of the form $x(n+1) - x(n) \in \gamma(n)F(x(n))$, where $(\gamma(n))_{n\geq 0}$ is a step-size sequence converging to zero. The continuous-time interpolation of such a discrete-time process will not itself be a precise solution for the differential inclusion (4.1). However, the interpolated process may be shown to satisfy the more relaxed solution concept—namely, that of a *perturbed solution* to the differential inclusion. We first define the notion of a δ-perturbation which we then use to define the notion of a perturbed solution.

Definition 4.4. *Let $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ be a set-valued map, and let $\delta > 0$. The δ-perturbation of F is given by*

$$F^{\delta}(x) := \{y \in \mathbb{R}^m : \exists z \in \mathbb{R}^m \text{ s.t. } \|z - x\| < \delta,\ d(y, F(z)) < \delta\}.$$

Definition 4.5. *A continuous function $\mathbf{y} : [0, \infty) \to \mathbb{R}^m$ will be called a* perturbed solution *to* (4.1) *if it satisfies the following set of conditions:*
(i) $\mathbf{y}$ is absolutely continuous.
(ii) There exists a locally integrable function $t \mapsto U(t)$ such that

$$\lim_{t\to\infty} \sup_{0\le\nu\le T} \left\| \int_t^{t+\nu} U(s)ds \right\| = 0,$$

and,
(iii) $\frac{d\mathbf{y}(t)}{dt} - U(t) \in F^{\delta(t)}(\mathbf{y}(t))$ for almost every $t > 0$, for some function $\delta : [0, \infty) \to \mathbb{R}$ with $\delta(t) \to 0$ as $t \to \infty$.

The following proposition gives sufficient conditions under which an interpolated process will in fact be a perturbed solution.

Proposition 4.6. *Consider a discrete-time process $\{x(n)\}_{n\ge1}$ such that*

$$\gamma(n)^{-1} (x(n+1) - x(n)) - \gamma(n)^{-1} U(n+1) \in F^{\delta_n}(x(n)),$$

where $\{\gamma(n)\}_{n\ge2}$ is a sequence of positive numbers such that $\gamma(n) \to 0$ and $\sum_{n=1}^{\infty} \gamma(n) = \infty$, $\{U(n)\}_{n\ge1}$ is a sequence of stochastic or deterministic perturbations satisfying

$$\lim_{n\to\infty} \sup_k \left\{ \left\| \sum_{s=n}^{k-1} \gamma(s+1) U(s+1) \right\| : \sum_{s=n}^{k-1} \gamma(s+1) \le T \right\} = 0, \quad \forall T > 0,$$

$\{\delta_n\}_{n\ge1}$ is a sequence of non-negative numbers converging to 0, and $\sup_n \|x(n)\| < \infty$. *Then the continuous-time interpolation of $\{x(n)\}_{n\ge1}$ is a perturbed solution of* (4.1)

The proof of Proposition 4.6 follows similar reasoning to the proof of Proposition 1.3 in [28].

Our end goal is to characterize the set of limit points of the discrete-time process $\{x(n)\}_{n\ge1}$ by characterizing the set of limit points of its continuous-time interpolation. With that end in mind, it is useful to consider the notion of an *internally chain-recurrent set*—a set of natural limit points for perturbed processes.

Definition 4.7. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^m$, and let $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ be a set valued map satisfying Assumption 4.1. Consider the differential inclusion* (4.1)*.*
(a) Given a set $X \subset \mathbb{R}^m$ and points ξ and η, we write $\xi \hookrightarrow \eta$ if for every $\epsilon > 0$ and $T > 0$ there exist an integer $n^* \geq 1$, solutions $x_1, \ldots, x_{n^*}$ to the differential inclusion (4.1)*, and real numbers $t_1, \ldots, t_{n^*}$ greater than T such that*
(i) $\mathbf{x}_i(s) \in X$, for all $0 \leq s \leq t_i$ and for all $i = 1, \ldots, n^*$,
(ii) $\|\mathbf{x}_i(t_i) - \mathbf{x}_{i+1}(0)\| \leq \epsilon$ for all $i = 1, \ldots, n^* - 1$,
(iii) $\|\mathbf{x}_1(0) - \xi\| \leq \epsilon$ and $\|\mathbf{x}_{n^*}(t_{n^*}) - \eta\| \leq \epsilon$.
(b) X is said to be internally chain recurrent if X is compact and $\xi \hookrightarrow \xi'$ for all $\xi, \xi' \in X$.

The following theorem from [28] allows one to relate the set of limit points of a perturbed solution of (4.1) to the internally chain recurrent sets F.

Theorem 4.8 ([28], Theorem 3.6)**.** *Let $\mathbf{y}$ be a bounded perturbed solution to* (4.1)*. Then the limit set of $\mathbf{y}$, $L(\mathbf{y}) = \bigcap_{t \geq 0} \overline{\{\mathbf{y}(s) : \ s \geq t\}}$ is internally chain recurrent.*

In our study of MBR algorithms we will sometimes focus on the class of potential games. The existence of a potential function is useful in that it allows us to infer the existence of a Lyapunov function that may be used to characterize the internally chain recurrent sets.

In particular, the differential inclusion (4.1) induces a set-valued dynamical system $\{\Phi_t\}_{t \in \mathbb{R}}$ defined by

$$\Phi_t(x_0) := \{\mathbf{x}(t) : x \text{ is a solution to (4.1) with } \mathbf{x}(0) = x_0\}.$$

Let Λ be any subset of $\mathbb{R}^m$. A continuous function $V : \mathbb{R}^m \to \mathbb{R}$ is called a Lyapunov function for Λ if $V(y) < V(x_0)$ for all $x_0 \in \mathbb{R}^m \backslash \Lambda$, $y \in \Phi_t(x_0)$, $t > 0$, and $V(y) \leq V(x_0)$ for all $x_0 \in \Lambda$, $y \in \Phi_t(x_0)$ and $t \geq 0$. The following proposition ([28], Proposition 3.27) allows one to relate the internally chain recurrent sets of a differential inclusion to Lyapunov attracting sets.

Proposition 4.9. *Suppose that V is a Lyapunov function for Λ. Assume that $V(\Lambda)$, the image of Λ under V, has empty interior. Then every internally chain recurrent set L is contained in Λ and $V|L$, the restriction of V to the set L, is constant.*

In order to eventually prove the main theoretical result of this chapter (Theorem 4.15) we will show that the continuous-time interpolation of a MBR process is in fact a bounded perturbed solution to the associated differential inclusion (4.4), and hence by Theorem 4.8, the limit set of the MBR process is an internally chain recurrent set of the associated differential inclusion.

4.3 Weakened Myopic Best-Response Process

In a MBR process it is assumed that players actions are always chosen as optimal (best-response) strategies. This is a strong assumption—for example, there are circumstances where computation (or execution) of a precise best response may be difficult to carry out, whereas an approximate best response can be implemented at relatively low cost (see, for example, Chapter 5). Similarly, in large-scale distributed settings it may be difficult to gather all the information necessary to compute a precise best response, but still feasible to gather sufficient information to compute an approximate best response (See Chapter 7).

In the spirit of [29, 72], we wish to study the robustness of the MBR algorithms in a setting where agents may sometimes choose suboptimal actions. As we will see in Chapters 5–7, this relaxation allows for a breadth of practical applications.

Formally, let the ϵ-best response set in this context be given by $\mathrm{BR}_{f,\epsilon} : Z \to \Delta(Y)$, where

$$\mathrm{BR}_{f,\epsilon}(z) := \mathrm{BR}_{1,\epsilon}(f_1(z)) \times \cdots \times \mathrm{BR}_{N,\epsilon}(f_N(z)),$$

and where $\mathrm{BR}_{i,\epsilon}$ is as defined in (2.2).

Let $\sigma_i(n) \in \Delta_i$ denote the strategy used by player i in round n, and let $\sigma(n) = (\sigma_1(n), \ldots, \sigma_N(n)) \in \Delta^N$ denote the joint strategy.[2] Suppose that players choose their next-stage strategies as

$$\sigma(n+1) \in \mathrm{BR}_{f,\epsilon_n}(z(n)), \tag{4.2}$$

where we assume the sequence $\{\epsilon_n\}_{n\geq 1}$ satisfies

Assumption 4.10. $\lim_{n\to\infty} \epsilon_n = 0$,

The associated observation state is updated as

$$z(n+1) - z(n) \in \gamma(n)\big(g(BR_{f,\epsilon_n}(z(n))) - z(n) + M_{n+1}\big). \tag{4.3}$$

where we assume the sequence $(M_n)_{n\geq 1}$ satisfies

Assumption 4.11. *For any $T > 0$ there holds*

$$\lim_{n\to\infty} \sup_k \left\{ \left\| \sum_{s=n}^{k-1} \gamma(s+1) M_{s+1} \right\| : \sum_{s=n}^{k-1} \gamma(s+1) \leq T \right\} = 0.$$

[2] Note that this is a generalization of the repeated play framework discussed in Section 2.3, wherein we allow players to use strategies $\sigma_i(n) \in \Delta_i$, rather than restricting to vertices of Δ_i.

We refer to the sequence $\{\epsilon_n\}_{n\geq 1}$ in (4.2) as a *best-response perturbation.* We refer to a sequence $\{z(n)\}_{n\geq 1}$ satisfying (4.3) and (4.2) as a *weakened MBR process* (cf. [29, 72]).

4.3.1 Examples

Example 4.12. *Classical FP is recovered by letting* $\gamma(n) = \frac{1}{n+1}$*, letting the observation space be given by* $Z = \Delta^N$*, letting* $g : \Delta(Y) \to \Delta^N$ *with* $g(x) = (g_1(x), \ldots, g_N(x))$*, where* $g_i : \Delta(Y) \to \Delta(Y_i)$ *is given by* $g_i(x) = \sum_{y_{-i}\in Y_{-i}} x(y_i, y_{-i})$*, and for each* i *letting* $f_i : \Delta^N \to \Delta(Y_{-i})$ *with* $f_i(z) = (z_1, \ldots, z_{i-1}, z_{i+1} \ldots, z_N)$*.*

Example 4.13. *Joint Strategy FP [73] is recovered by letting* $\gamma(n) = \frac{1}{n+1}$*, setting the observations space to be* $Z = \Delta(Y)$*, letting* $g : \Delta(Y) \to \Delta(Y)$ *to be the identity function and letting* $f_i : \Delta(Y) \to \Delta(Y_{-i})$ *be given by* $f_i(z) = \sum_{y_i\in Y_i} z(y_i, y_{-i})$*.*

Example 4.14. *Suppose all players use an identical action space given by* $Y_i = \bar{Y}$, $\forall i$*. In this case, Empirical Centroid FP (ECFP) [104] is recovered by letting* $\gamma(n) = \frac{1}{n+1}$*, letting the observation space be given by* $Z = \Delta(\bar{Y})$*, letting* $g : \Delta(Y) \to \Delta(\bar{Y})$ *be given by* $g(x) = N^{-1}\sum_{i=1}^N x_i$ *where* $y_i \mapsto x_i(y_i) = \sum_{y_{-i}\in Y_{-i}} x(y_i, y_{-i})$*, and letting* $f_i : \Delta(\bar{Y}) \to \Delta(Y_{-i})$ *be given by* $f_i(z) = (z, \ldots, z)$*, i.e., the* $(N-1)$*-tuple containing repeated copies of* z*.*[3]

4.3.2 Observation of Strategies

In a game-theoretic learning algorithm, it is generally assumed that players observe the "realized actions" played by others, but may not observe the mixed strategies used to generate the realized actions.

In a FP-type algorithm, the question of whether players view "realized actions vs mixed strategies" is more involved due to the fact that the objects that players view are assumed to live in the observation space Z.

Let $\sigma_i(n)$ denote the mixed strategy used by player $i \in \mathcal{N}$ in round $n \geq 1$ and let $y_i(n)$ denote the realized action used by player i in round n. The action $y_i(n)$ is assumed to be drawn as a random sample from $\sigma_i(n)$. Let $\sigma(n) := \sigma_1(n) \times \cdots \times \sigma_N(n)$ be the joint mixed strategy used in round n and let $y(n) = (y_1(n), \ldots, y_N(n))$ be the joint action tuple played in round n.

Suppose that there exists a sequence $\{\epsilon_n\}_{n\geq 1}$ satisfying Assumption 4.10 such that for any $\sigma_i \in BR_{i,\epsilon_n}(f_i(z(n-1)))$, the support of the mixed strategy σ_i contains only pure strategies that are ϵ_n-best responses.

[3]The ECFP algorithm is explored in more depth in Section 4.5 in connection with the robustness result.

In this case, we will assume that the observation state is updated towards the mapping of the realized action $g(\mathbf{1}_{y(n)})$ and that $M_n = 0$ for all $n \geq 1$. That is,

$$z(n+1) - z(n) \in \gamma(n)\big(g(\mathbf{1}_{y(n)}) - z(n)\big).$$

This effectively assumes that players observe the mapping of the realized action $g(\mathbf{1}_{y(n)})$. In the case of classical FP, this is equivalent to assuming that players view the realized actions of others. The applications studied in the later parts of the dissertation (Chapters 5–7) admit such a sequence $\{\epsilon_n\}_{n\geq 1}$.

Suppose that there exists a sequence $\{\epsilon_n\}_{n\geq 1}$ satisfying Assumption 4.10 such that in each stage $n \geq 1$ the mixed strategy of each player $i \in \mathcal{N}$ satisfies $\sigma_i(n) \in BR_{f,\epsilon_n}(z(n))$, but there does not exist a sequence $\{\eta_n\}_{n\geq 1}$ satisfying $\eta_n \to 0$ such that each strategy y_i in the support of $\sigma_i(n)$ is an η_n-best response. We consider two cases.

Case 1. If the observation mapping g is linear, then we assume that the observation state is updated towards the mapping of the realized action $g(\mathbf{1}_{y(n)})$. That is

$$z(n+1) - z(n) \in \gamma(n)\big(g(\sigma(n)) - z(n) + M_n\big)$$

where the random variable M_n is given by

$$M_n = g(\mathbf{1}_{y(n)}) - g(\sigma(n)).$$

As long as $\gamma(n)$ is deterministic and $\gamma(n) = o(\frac{1}{\log n})$, then $\{M_n\}_{n\geq 1}$ is a martingale difference sequence satisfying Assumption 4.11 with probability 1. Note that in classical FP, JSFP, and ECFP, the mapping g is linear.

Case 2. If the observation mapping g is nonlinear, then we assume that players observe the mapping of the mixed strategies into the observation space. That is, the observation state is updated as

$$z(n+1) - z(n) \in \gamma(n)\big(g(\sigma(n)) - z(n)\big).$$

4.4 Robustness Property for MBR Process

The following theorem is the main theoretical result of the chapter. It shows that if Assumptions 2.1–2.3 and 4.10–4.11 are satisfied, then the set of limit points of any weakened (discrete-time)

MBR process are contained in a chain-recurrent set of the associated differential inclusion

$$\dot{\mathbf{z}}(t) \in g\big(\mathrm{BR}_f(\mathbf{z}(t))\big) - \mathbf{z}(t). \tag{4.4}$$

Theorem 4.15. *Let $\Psi = (\{\gamma(n)\}_{n\geq 1}, g, (f_i)_{i=1}^n)$ be a MBR algorithm. Assume that Ψ satisfies Assumptions 2.1–2.3. Assume that $(z(n))_{n\geq 1}$ is a weakened MBR process satisfying Assumptions 4.10–4.11. Then $(z(n))_{n\geq 1}$ converges to the internally chain recurrent set of the associated differential inclusion* (4.4)*.*

The proof of Theorem 4.15 follows directly from the following lemma together with Proposition 4.6 and Theorem 4.8. The lemma shows that for sufficiently small ϵ the ϵ-best responses are contained in the δ-perturbations of BR for all z. While this is clearly true pointwise, the uniformity in z has not previously been shown. This observation was not made in [29] and results in a gap in the proof presented there.

Lemma 4.16. *Let $\epsilon_n \to 0$ as $n \to \infty$. Then there exists a sequence $\delta_n \to 0$ such that $\mathrm{BR}_{f,\epsilon_n}(z) \subseteq \mathrm{BR}_f^{\delta_n}(z)$ uniformly for $z \in Z$.*

Proof. We work with the supremum norm on Z and $\Delta(Y_i)$ throughout the proof.

Fix an arbitrary $\delta > 0$. Following [108], define the "stability set" of a (joint) action $y \in Y$ as

$$St(y) := \{z \in Z : \; y_i \in \mathrm{BR}_i(f_i(z)), \; \forall i\}.$$

Note that the closer that z is to boundary of $St(y)$, the smaller that ϵ must be to ensure that ϵ-best responses place large mass on y, and hence are δ-perturbations of $y = \mathrm{BR}(z)$. To gain the uniform inclusion of the ϵ-best responses in the δ-perturbations we consider the interior of the sets $St(y)$ separately from neighbourhoods of boundaries of the stability sets. To this end, extend the stability set concept to sets of actions $T \subseteq Y$ by defining

$$St(T) := \bigcap_{y\in T} St(y)$$

to be the set of $z \in Z$ such that all actions $y \in T$ are best responses to z. In what follows, we will use the stability sets $St(T)$ to construct a finite cover $\{D(T)\}_{T\subseteq Y}$ of Z such that $\mathrm{BR}(f(z)) \subseteq T$ for each $z \in D(T)$. This allows us to show that ϵ-best responses to elements in $D(T)$ place most of their mass on T, and in particular it can be shown that for each set $D(T) \subseteq Z$ there holds

$$\mathrm{BR}_{f,\epsilon}(z) \subseteq \mathrm{BR}_f^{\delta}(z), \quad \text{for all } z \in D(T) \tag{4.5}$$

for all ϵ sufficiently small. Since the cover is finite, we can show that in fact

$$\mathrm{BR}_{f,\epsilon}(z) \subseteq \mathrm{BR}_f^\delta(z), \quad \text{for all } z \in Z \tag{4.6}$$

holds for all ϵ sufficiently small. (We note, however, that we proceed along a slightly more direct route, showing (4.6) without directly verifying (4.5).)

To this end, note that by the upper hemicontinuity of BR_i and continuity of f_i, we have that $St(y)$ and $St(T)$ are closed sets. For any $\eta > 0$ and any $T \subseteq Y$, let $B(St(T), \eta)$ be the open ball of radius η about $St(T)$ which is empty if $St(T)$ is empty. Let $M = \prod_{i \in \mathcal{N}} |Y_i|$ and for each $k \in \{1, 2, \ldots, M\}$ let $\mathcal{T}^k$ be the collection of all subsets $T \subseteq Y$ such that $|T| = k$. For the tuple $\eta_{>k} = (\eta_{k+1}, \ldots, \eta_M)$ define the "exclusion set"

$$E^k(\eta_{>k}) := \bigcup_{\kappa=k+1}^{M} \bigcup_{T \in \mathcal{T}^\kappa} B(St(T), \eta_\kappa)$$

to be the set of $z \in Z$ that are close to any stability sets $St(T)$ with $|T| > k$, where close is measured by the tuple $\eta_{>k}$.

We now work recursively from $k = M$ down to $k = 1$. Start by letting $\eta_M = \delta$ and let

$$D(Y) := B(St(Y), \eta_M).$$

Now let $k \in \{1, \ldots, M-1\}$ and suppose $\eta_{>k}$ is given. Suppose $T \in \mathcal{T}^k$, and let $\tilde{T} \subseteq Y$ such that $\tilde{T} \not\subseteq T$. Then $|T \cup \tilde{T}| > k$, so by the definition of $E^k(\eta_{>k})$ we have that $St(T) \cap St(\tilde{T}) = St(T \cup \tilde{T}) \subseteq E^k(\eta_{>k})$. Therefore $St(T) \cap St(\tilde{T}) \cap E^k(\eta_{>k})^c = \emptyset$. Since $E^k(\eta_{>k})$ is open by definition, the complement is closed. Therefore the sets $St(T) \cap E^k(\eta_{>k})^c$ and $St(\tilde{T})$ are disjoint compact sets and either have a minimal separating distance or at least one is empty. We can therefore fix an η_k such that, for each $T \in \mathcal{T}^k$,

$$D(T) := B(St(T), \eta_k) \cap E^k(\eta_{>k})^c$$

is separated from

1. $St(\tilde{T})$ for all $\tilde{T} \subseteq Y$ such that $\tilde{T} \not\subseteq T$, and
2. $D(\tilde{T})$ for all $\tilde{T} \in \mathcal{T}^k$ with $\tilde{T} \neq T$.

Iterating this reasoning down to $k = 1$ defines the full set of η_k values as well as $D(T)$ for all $T \subseteq Y$ with $T \neq \emptyset$.

We now show that the sets $\{D(T)\}_{T\subseteq Y}$ partition Z. By definition we have that $D(Y) = B(St(Y),\eta_M)$; using a backwards induction argument one may verify that

$$\bigcup_{k=1}^{M}\bigcup_{T\in\mathcal{T}^k} D(T) = \bigcup_{k=1}^{M}\bigcup_{T\in\mathcal{T}^k} B(St(T),\eta_k). \tag{4.7}$$

Hence, $Z = \bigcup_{T\subseteq Y} St(T) \subseteq \bigcup_{T\subseteq Y} D(T) \subseteq Z$, where the equality holds because there exists a best response to any $z\in Z$, and the first containment holds by (4.7). Furthermore, by property 2) above (and the fact that, by construction, $D(T)\cap D(\tilde{T}) = \emptyset$ for $|T| \neq |\tilde{T}|$) we have that $D(T)\cap D(\tilde{T}) = \emptyset$, $\forall T,\tilde{T}\subseteq Y$, $\tilde{T}\neq T$. Hence the sets $\{D(T)\}_{T\subseteq Y}$ partition Z.

We wish to show that for $z\in D(T)$, the ϵ-best responses place most of their mass on elements in T. To this end, let $T\in\mathcal{T}^k$ for arbitrary $1\le k\le M$, and let $\bar{D}(T)$ be the closure of $D(T)$. We claim that if $z\in\bar{D}(T)$, then all pure strategy best responses to z are contained in T. To see this, suppose contrariwise that $z\in\bar{D}(T)$ has a pure strategy best response not contained in T. Then $z\in St(\tilde{T})$ for some $\tilde{T}\not\subseteq T$, which violates Property 1) above.

Now define, for $z\in Z$, the set $T(z)$ to be the $T\subseteq Y$ such that $z\in D(T)$. Also define $T_i(z) := \{y_i\in Y_i : (y_i,y_{-i})\in T(z) \text{ for some } y_{-i}\in Y_{-i}\}$, so that all of Player i's pure strategy best responses to $z\in Z$ are contained in $T_i(z)$. Thus, for $z\in\bar{D}(T)$, for each i there exists a $\xi_{i,\delta}(z) > 0$ such that

$$\max_{y_i\in Y_i} U_i(\mathbf{1}_{y_i}, f(z)) - \max_{\tilde{y}_i\notin T_i(z)} U_i(\mathbf{1}_{\tilde{y}_i}, f(z)) = \xi_{i,\delta}(z).^4$$

Since $\bar{D}(T)$ is compact and U_i (and hence ξ_i) is continuous, we get $\inf\limits_{p\in\bar{D}(T)} \xi_{i,\delta}(z) > 0$, $\forall i$. Since there are finitely many $T\subseteq Y$ and $i\in\mathcal{N}$, there exists a $\xi_\delta > 0$ such that for each i and $z\in Z$, $\max_{y_i\in Y_i} U_i(\mathbf{1}_{y_i}, f_i(z)) - \max_{\tilde{y}_i\notin T_i(z)} U_i(\mathbf{1}_{\tilde{y}_i}, f_i(z)) \ge \xi_\delta$. We have shown that for any i and any z, any action not in $T_i(z)$ receives utility less than the best response by at least an amount ξ_δ.

Invoking the linearity of $z_i\mapsto U_i(z_i,z_{-i})$, it follows that for $z\in Z$, for each i, an ϵ-best response to z can put probability at most ϵ/ξ_δ on actions not in $T_i(z)$. That is, for any $z\in Z$ and for any $i\in\mathcal{N}$,

$$\mathrm{BR}_{i,\epsilon}(f_i(z)) \subseteq \Big\{x_i\in\Delta(Y_i) : \sum_{y_i\in T_i(z)} z_i(y_i) \ge 1 - \frac{\epsilon}{\xi_\delta}\Big\}.$$

Let $\epsilon \le \min\{\delta\xi_\delta,\delta\}$ and let $x\in \mathrm{BR}_{f,\epsilon}(z)$. By the above, x is a distance at most δ from a strategy x' which places all its mass on $T(z)$. Simultaneously, by the construction of $D(T)$, z is a distance at most δ from the set $St(T(z))$; i.e., there exists a $z'\in St(T(z))$ such that $d(z,z')\le\delta$. By the definition of the stability set, we have $x'\in\mathrm{BR}_f(z')$. This shows that $x\in\mathrm{BR}_f^\delta(z)$. Since

[4] For completeness we emphasize that $\xi_{i,\delta}$ is in fact a function of δ, as well as z.

z was arbitrary, and this holds for any $x \in \mathrm{BR}_{f,\epsilon}(z)$ we have $\mathrm{BR}_{f,\epsilon}(z) \subseteq \mathrm{BR}_f^\delta(z), \quad \text{for all} \quad z \in Z$. Since this holds for any $\epsilon \leq \min\{\delta\xi_\delta, \delta\}$, it follows that for any sequence $\epsilon_n \to 0$ there exists a sequence $\delta_n \to 0$ such that $\mathrm{BR}_{f,\epsilon_n}(z) \subseteq \mathrm{BR}_f^{\delta_n}(z)$ for any $z \in Z$. □

We now prove Theorem 4.15.

Proof. By assumption, players choose their strategies according to (4.2). Applying (4.3) we get the recursive form $\gamma(n)^{-1}(z(n+1) - z(n)) - \gamma(n)^{-1}M_{n+1} \in g(\mathrm{BR}_{f,\epsilon_n}(z(n))) - z(n)$, where $\epsilon_n \to 0$. By Lemma 4.16, we know that $\gamma(n)^{-1}(z(n+1) - z(n)) \in g(\mathrm{BR}_f^{\delta_n}(z(n))) - z(n)$ for some sequence $\delta_n \to 0$. Let $F : Z \rightrightarrows Z$ be given by $F(z) = g(\mathrm{BR}_f(z)) - z$. Since g is uniformly continuous, the previous equation implies that $\gamma(n)^{-1}(z(n+1) - z(n)) \in F^{\eta_n}(z(n))$ for some sequence $\eta_n \to 0$. By Proposition 4.6, the continuous-time interpolation of $\{z(n)\}_{n\geq 1}$ is a bounded perturbed solution to the associated differential inclusion (4.4). The result then follows by Theorem 4.8. □

An important consequence of Theorem 4.15 is that, if one wishes to show convergence of a MBR algorithm to some equilibrium set, one need only verify that the associated internally chain recurrent set is contained in the equilibrium set.

This has been shown, for example, with the set of NE and classical FP in potential games [28], two-player zero-sum games [64], and generic $2 \times m$ games [35]. Thus, the following important result ([29], Corollary 5) may also be seen as a consequence of Theorem 4.15. As this result will arise in the subsequent discussion, we find it convenient to state it here.

Corollary 4.17 ([29], Corollary 5)**.** *Let Γ be a potential game, two-player zero-sum game, or generic $2 \times m$ game. Let $(q(n))_{n\geq 1}$ be a weakened FP process and suppose that Assumptions 4.10–4.11 hold. Then $(q(n))_{n\geq 1}$ converges to the set of NE in the sense that* $\lim_{n\to\infty} d(q(n), NE) = 0$.

4.5 Example: Empirical Centroid Fictitious Play

In this section we use the tools developed in this chapter to prove a robust version of Theorem 3.4. In particular, we will show that under the assumptions of Theorem 3.4, any *weakened* ECFP process converges in the same sense as in Theorem 3.4.

We study the limit sets of discrete-time ECFP (DT-ECFP) by relating them to the limit sets of continuous-time ECFP (CT-ECFP). Lemma 4.18 uses the robustness result for MBR processes (Theorem 4.15) to show that the limit sets of the DT-ECFP processes (3.9) and (3.10) are contained in the internally chain recurrent sets of the CT-ECFP differential inclusions (3.11) and (3.12). Lemma 4.19 shows that the respective internally chain recurrent sets of CT-ECFP, (3.11) and

(3.12), are contained in the SNE and MCE sets. Lemmas 4.18 and 4.19 together prove Theorem 4.20—the robustness result for ECFP processes.

Lemma 4.18. *Suppose Assumptions 3.1–3.2 hold. Let $\{q(n), \bar{q}(n)\}_{n\geq 1}$ be a weakened ECFP process such that the associated sequence $(\epsilon_n)_{n\geq 1}$ satisfies Assumption 4.10. Then,*
(i) The set of limit points of $\{\bar{q}(n)\}_{n\geq 1}$ is an internally chain recurrent set of (3.12)*,*
(ii) The set of limit points of $\{q(n)\}_{n\geq 1}$ is an internally chain recurrent set of (3.11)*.*

Proof. For $q \in \Delta^N$, let $F_1(q) = \mathrm{BR}(\bar{q}) - \bar{q}$ and let $F_2(q) = \mathrm{BR}(\bar{q}) - q$. By Lemma 4.16 and Assumption 4.10, there exist sequences $\delta_{1,n} \to 0$ as $n \to \infty$ and $\delta_{2,n} \to 0$ as $n \to \infty$ such that $\bar{q}(n)$ as given in (3.10) is a $\delta_{1,n}$-perturbation of F_1 for all $n \geq 1$, and $q(n)$ as given in (3.9) is a $\delta_{2,n}$-perturbation of F_2 for all $n \geq 1$. By Proposition 4.6, the continuous-time interpolation of $\bar{q}(n)$ is a perturbed solution of $\frac{d\mathbf{x}}{dt} \in F_1(\mathbf{x})$ and the continuous-time interpolation of $q(n)$ is a perturbed solution of $\frac{d\mathbf{x}}{dt} \in F_2(\mathbf{x})$. By Theorem 4.8, the set of limit points of a bounded perturbed solution of a differential inclusion is an internally chain recurrent set of the differential inclusion, which proves the result. □

Lemma 4.19. *Let Γ be an identical interests game. Let $\mathcal{C}$ be a permutation-invariant partition of the associated player set $\mathcal{N}$. Then,*
(i) Every internally chain recurrent set of (3.12) *is contained in the set of SNE.*
(ii) Every internally chain recurrent set of (3.11) *is contained in the set of MCE.*

Proof. Proof of (i): Let $W := -U$. By Section 3.4.3 (in particular, see (3.13)), W is a Lyapunov function for the set of SNE with $\mathbf{x}(t) := \bar{\mathbf{q}}(t)$. Note that W is multilinear and hence continuously differentiable.

For a differentiable function $f : \mathbb{R}^m \to \mathbb{R}$, we say $x \in \mathbb{R}^m$ is a critical point of f if for $i = 1, \ldots, m$, the partial derivative at x is zero, i.e, $\frac{\partial}{\partial x_j} f(x) = 0$. By Sard's Theorem ([109], p. 69), if CP is the set of critical points of W, then $W(CP)$ contains no intervals. By definition, the set of NE is contained in the set of critical points of U, and hence also contained in the set of critical points of W. Furthermore, by definition, $SNE \subset NE$, and hence the set SNE is contained in the critical points set of W. Thus, by Proposition 4.9, every internally chain recurrent set of (3.12) is contained in the set SNE.

Proof of (ii): Note that, by Lemma 3.13, $\sigma \in \mathrm{BR}(\bar{\sigma}) \implies \bar{\sigma} \in \mathrm{BR}(\bar{\sigma})$. Thus, $\sigma \in MCE$ implies that $\bar{\sigma} \in SNE$. Let $V : \Delta^N \to \mathbb{R}$, with $V(\sigma) := \frac{1}{n}\sum_{i=1}^N U(\sigma_i, \bar{\sigma}_{-i})$, and note that by Lemma 3.12, $V(\sigma) = U(\bar{\sigma})$. Invoking again Sard's Theorem, $U(NE)$ contains no intervals, and hence $U(SNE) \subset U(NE)$ contains no intervals. Since $U(SNE)$ contains no intervals, $V(MCE)$ also contains no intervals.

By Section 3.4.3 (in particular, see (3.15)) the function V is a Lyapunov function for the set of MCE with $\mathbf{x}(t) := \mathbf{q}(t)$. It follows from Proposition 4.9 that every internally chain recurrent set of (3.11) is contained in the set MCE. $\square$

Lemmas 4.18 and 4.19 together prove the following Theorem (cf. Theorem 3.4).

Theorem 4.20. *Assume Γ is a potential game. Let $\mathcal{C}$ be a permutation-invariant partition of the associated player set $\mathcal{N}$. Let $\{q(n), \bar{q}(n)\}_{n\geq 1}$ be a weakened ECFP process such that the associated sequence $(\epsilon_n)_{n\geq 1}$ satisfies Assumption 4.10. Then,*
(i) players learn MCE strategies in the sense that $\lim_{n\to\infty} d(q(n), MCE) = 0$,
(ii) players learn SNE strategies in the sense that $\lim_{n\to\infty} d(\bar{q}(n), SNE) = 0$.

Chapter 5

Mitigating Complexity: A Single Sample Approach

5.1 Introduction

While MBR algorithms have many useful properties, their practicality can be undermined by the enormous computational burden inherent in the best-response computation. For example, in each stage of the FP algorithm, each player $i \in \mathcal{N}$ must compute the expected utility for each of her actions given her current beliefs regarding opponents' strategies. Evaluating this expected utility is a problem whose complexity scales exponentially in the number of players (see Section 5.2).

In this chapter we study a Monte Carlo-based technique for mitigating computational complexity in MBR algorithms. Using this technique, a MBR algorithm can be implemented with a per-iteration complexity that scales linearly (rather than exponentially) in the number of players.

The Sampled FP (SFP) algorithm [48, 49, 52, 53, 57, 110] introduced the idea of mitigating complexity in MBR algorithms using a Monte Carlo (i.e., sampling-based) approach.[1] At each iteration of the SFP algorithm, players approximate the expected utility of each of their actions by drawing several samples from an underlying probability distribution. Players then myopically choose an "optimal" next-stage action using the approximated utility as a surrogate for the true expected utility. The work [48] showed that as long as the number of samples drawn each round grows sufficiently quickly, players learn an equilibrium in the same sense as FP, almost surely.

SFP mitigates complexity by avoiding any direct evaluation of the expected utility. However, SFP has a notable shortcoming: In order to guarantee learning is achieved, the number of samples that must be drawn in each iteration (i.e., round) of the algorithm grows without bound (on the order of $\sqrt{n}$ samples per round, where n is the current round of the repeated play algorithm).

[1]The work [48] introduced these techniques and studied their use in large-scale traffic routing Problems. In particular, in [48], sampled FP was applied to optimize traffic routing using a black-box simulation model of the Troy, Michigan city traffic grid. An interesting future research direction would be to study traffic optimization using single sample MBR dynamics in similar models of a city traffic grid.

In this chapter (see also [75,111,112]) we first extend the techniques proposed in [48] to general MBR algorithms. We call this such an algorithm a *sampled MBR* (S-MBR) algorithm. We then propose a variant of the S-MBR algorithm—which we call the Single Sample MBR (SS-MBR) algorithm—in which only one sample need be drawn each round of the repeated play process. A SS-MBR algorithm achieves the same fundamental computational advantage as a S-MBR algorithm (i.e., direct evaluation of the expected utility is avoided) but does so by drawing only one sample per round (rather than the drawing k_n samples in round n, with $\lim_{n\to\infty} k_n = \infty$ as required in S-MBR).

Intuitively, the reduction in the required number of per-round samples is accomplished by treating the expected-utility process as *quasi static*. Such treatment is possible due to the diminishing incremental step size in the expected-utility process. In a SS-MBR algorithm, the sample data gathered in the current round of repeated play is recursively combined with sample data from previous rounds using a stochastic-approximation-type estimation rule. This may be contrasted with a S-MBR algorithm where, in each round of the repeated play, data gathered from sampling in the previous round is wholly discarded and a fresh set of samples is gathered to approximate the expected utility for the upcoming round. (See Section 5.4.3 for more details.)

Due to the improved efficiency in information handling, SS-MBR algorithms are able to achieve convergence at a rate similar to that of S-MBR algorithm (in terms of repeated-play iterations) despite drawing far fewer samples per-iteration. (See Section 5.5.4 for more details.)

The main contribution of the chapter is the presentation of a prototypical SS-MBR algorithm and proof of convergence to the internally chain recurrent set of the associated differential inclusion. The proof relies on applying the robustness result developed in Chapter 4 to the SS-MBR processes.

Related works have studied approaches for mitigating computational issues arising in large-scale implementations of FP. Joint Strategy FP (JSFP) [73] studies a variant of FP in which players update a utility estimate using a computationally simple recursive procedure and choose next-stage actions using a best-response rule combined with an inertial term. JSFP is shown to converge to pure strategy Nash equilibria (NE) in ordinal potential games but is fundamentally different from SS-FP (and FP) in that the tracked utility corresponds to an empirical distribution taken over joint actions,[2] and convergence may only occur at pure NE. In κ-exponential FP [76], players use any given action with a probability that depends on the average utility the action has generated when played in previous rounds. In effect, this results in a reinforcement technique which relies on a relatively simple recursive update rule. Alternatively, one may also mitigate complexity

[2]In FP, Sampled FP, and SS-FP, players choose a best response to the product of marginal empirical distributions (or an estimate thereof), which implicitly presumes a form of independence among opponents' strategies. Tracking and responding to the empirical distribution of joint actions, as in JSFP, fundamentally alters the dynamics of classical FP. SS-FP achieves computational efficiency while preserving the basic dynamical structure of FP.

using a payoff-based variant of FP such as the actor-critic process studied in [29]. Our focus on SS-MBR dynamics is due in part to its intrinsic value as a means of mitigating complexity, but also stems from our eventual goal to implement the algorithm in a network-based setting. SS-MBR dynamics will admit a simple network-based implementation in which players will form an estimate $\hat{z}(n)$ of the "true" observation state $z(n)$ by sharing information with neighbors, and approximate the utility of each of their actions by drawing samples from the probability distribution $f_i(\hat{z}(n))$.

Other work has considered mitigating complexity by restricting attention to games where the utility functions exhibit only a local dependence on neighboring players as defined by a graph over the player set; e.g., [89, 113].

We note that payoff-based learning algorithms and first-order methods e.g., [29,61,86,88,89,114, 115], tend to be computationally simple and, under certain circumstances, can be a good alternative to the present approach. However, such algorithms often assume that players have access to payoff information, and may not be applicable in settings where this information is costly to obtain or otherwise unavailable. For example, in a network-based setting such as [111], players are engaged in a form of virtual game play without access to physical payoff measurements. In such a setting, SS-FP is shown to achieve learning and may be preferable to the aforementioned approaches.

The remainder of the chapter is organized as follows. Section 5.2 introduces the notion of complexity studied throughout the chapter. Section 5.3 reviews sampled MBR algorithms. Section 5.4 presents the SS-MBR algorithm and proves that such processes converge to the internally chain recurrent set of the associated differential inclusion. As an example application, Section 5.5 applies the single sample approach to classical fictitious play.

5.2 Computational Complexity in MBR Algorithms

In a MBR algorithm, each player must solve the optimization problem

$$\arg\max_{y_i \in Y_i} U_i(y_i, f_i(z(n)))$$

at each iteration of the algorithm. In order to solve this optimization problem, player i must compute the expected utility for each action $y_i \in Y_i$ given the probability distribution $f_i(z(n))$. In general, the computational complexity of evaluating this expected utility increases exponentially in the number of players N. This is the case, for example in FP.

Our focus in this chapter will be on mitigating the *per-iteration* computational complexity of an algorithm. We say a learning algorithm has (per-iteration) complexity $O(f(N, n))$ if the complexity of a single player choosing her next-stage action is $O(f(N, n))$, where N is the number of players and n is the stage of the algorithm. Using this convention, a MBR algorithm has computational complexity (in the worst case) of $O(e^N)$. We will show that the single-sample variant of a MBR algorithm achieves learning in the same sense as its deterministic counterpart, but has a complexity of $O(N)$.

5.3 Sampled Myopic Best-Response Algorithm

In this section we briefly review the sampling-based approach for mitigating complexity introduced in [48] in the context of MBR algorithms. Section 5.3.1 introduces some necessary notation and Section 5.3.2 presents the Sampled MBR (S-MBR) algorithm.

5.3.1 Algorithm Setup

In a S-MBR algorithm, the expected utility $U_i(y_i, f_i(z(n)))$, $y_i \in Y_i$ is approximated by drawing samples from the underlying probability distribution $f_i(z(n))$ and computing the average utility over the sampled actions.

In particular, for each $y_i \in Y_i$, let $\widehat{U}_i(y_i, n)$ denote an estimate that player i forms of the expected utility $U_i(y_i, f_i(z(n)))$. Each round of play, player i draws several "test actions" as random samples from the forecasted strategy $f_i(z(n)) \in \Delta_{-i}$. For each action $y_i \in Y_i$, player i computes the average utility the action y_i would generate given the randomly sampled "test actions." Player i then chooses a next-stage action that is myopically optimal using the estimated utility $(\widehat{U}_i(y_i, n))_{y_i \in Y_i}$ as a surrogate for the true expected utility in the best response computation (2.6).

Let $y_i(n)$ denote the action taken by player i in round n, and let $y(n) := (y_i(n))_{i \in \mathcal{N}}$ be the joint action taken at time n.

Let k_n denote the number of samples drawn by each player in round n of the repeated play. It will be shown that, as long as the number of samples satisfies

Assumption 5.1. $\lim_{n\to\infty} k_n = \infty$,

then any S-MBR process will converge to the internally chain recurrent set of the associated differential inclusion.

We now present the general sampled MBR algorithm.

5.3.2 Sampled MBR Algorithm

Initialize

(i) Each player $i \in \mathcal{N}$ chooses an arbitrary initial action $y_i(1) \in Y_i$. The observation state is initialized as $z(1) = g(\mathbf{1}_{y(1)})$. A sample rate sequence $(k_n)_{n\geq 1}$ is fixed.

Iterate $(n \geq 1)$

(ii) Each player $i \in \mathcal{N}$ draws k_n "test actions" as (conditionally independent given $z(n)$) random samples from the distribution $f_i(z(n))$; let $\tilde{y}^s_{-i}(n)$ denote the s-th random sample drawn by player i in round n. Player i estimates the utility of each of her actions $y_i \in Y_i$ using

$$\widehat{U}_i(y_i, n) = \frac{1}{k_n} \sum_{s=1}^{k_n} U_i(y_i, \tilde{y}^s_{-i}(n)). \tag{5.1}$$

(iii) Each player $i \in \mathcal{N}$ chooses a next-stage action that is a best response given her estimate of the expected utility, i.e.,

$$y_i(n+1) \in \arg\max_{y_i \in Y_i} \widehat{U}_i(y_i, n). \tag{5.2}$$

(iv) The observation state is updated recursively to account for the action just taken: $z(n+1) = z(n) + \frac{1}{n+1}\left(g(\mathbf{1}_{y(n+1)}) - z(n)\right)$.

5.3.3 Discussion

The following theorem shows that convergence of a S-MBR process is ensured so long as the sample rate sequence satisfies Assumption 5.1.

Theorem 5.2. *Let $(z(n))_{n\geq 1}$ be a S-MBR process. Suppose Assumptions 2.1–2.3 and Assumption 5.1 hold. Then $(z(n))_{n\geq 1}$ converges to the internally chain recurrent set of* (4.4)*, almost surely.*

Proof. By the strong law of large numbers, $\widehat{U}(y_i, n) \to U(y_i, f_i(z(n)))$, a.s., as $n \to \infty$. Hence, a S-MBR process is a weakened MBR process, almost surely. The result then follows by Theorem 4.15.

□

We note that in the case of classical FP, the above result combined with Corollary 4.17 implies that $\lim_{n\to\infty} d(q(n), NE) = 0$ for the sampled variant of FP. In the case of ECFP, the above result combined with Lemma 4.19 implies that $\lim_{n\to\infty} d(\bar{q}^N(n), CNE) = 0$ and $\lim_{n\to\infty} d(q(n), MCE) = 0$ for the sampled variant of ECFP.

Computational Complexity in S-MBR

In a sampled MBR algorithm players must drawn k_n samples from the distribution $f_i(z(n))$ in round n and evaluate the pure-strategy utility at each sample. This gives a per-iteration complexity of

$O(Nk_n)$ where N is the number of players and, per Assumption 5.1, $k_n \to \infty$ as $n \to \infty$.

5.4 Single Sample Approach

5.4.1 Algorithm Setup

In a single sample MBR algorithm, players form an estimate of the expected utility using a recursive stochastic-approximation-type rule. Similar to S-MBR, let $\widehat{U}_i(y_i, n)$ be the estimate which player i maintains of the expected utility $U_i(y_i, q_{-i}(n))$ for each of her actions $y_i \in Y_i$ and let $y_i(n)$ be the action taken by player i in round n. Let $\{\rho(n)\}_{n\geq 1}$ be a deterministic sequence of weights to be used in the stochastic-approximation-type procedure, and assume:

Assumption 5.3. *The sequence* $(\rho(n))_{n\geq 1}$ *is such that* $0 < \rho(n) \leq 1$, $\sum_{n\geq 1}(\rho(n))^2 < \infty$, *and* $\sum_{n\geq 1} \rho(n) = \infty$.

We will assume that the step-size sequence $(\gamma(n))_{n\geq 1}$ satisfies Assumption 2.2 and

Assumption 5.4. $\lim_{n\to\infty} \frac{\gamma(n)}{\rho(n)} = 0$.

The SS-MBR algorithm is outlined below.

5.4.2 Single Sample MBR Algorithm

Initialize
(i) For each $i \in \mathcal{N}$, let $y_i(1) \in Y_i$ be arbitrary. Initialize the observation state as $z(1) = g(\mathbf{1}_{y(1)})$ and initialize the utility estimate as $\widehat{U}_i(y_i, 0) = 0,\ \forall y_i \in Y_i,\ \forall i$.

Iterate $(n \geq 1)$
(ii) Each player $i \in \mathcal{N}$ draws a single "test action" $\tilde{y}_{-i}(n)$ as a (conditionally independent, given $z(n)$) random sample $\tilde{y}_{-i}(n) \sim f_i(z(n))$ and updates the estimate $\widehat{U}_i(y_i, n)$ for each action $y_i \in Y_i$ according to the recursion,[3]

$$\widehat{U}_i(y_i, n) = (1 - \rho(n))\widehat{U}_i(y_i, n-1) + \rho(n)U_i(y_i, \tilde{y}_{-i}(n)).$$

(iii) Each player $i \in \mathcal{N}$ chooses a next-stage action using the rule:

$$y_i(n+1) \in \arg\max_{y_i \in Y_i} \widehat{U}_i(y_i, n).$$

[3]Since $\tilde{y}_{-i}(n)$ is a pure strategy, the evaluation of the utility is relatively simple.

(iv) The observation state is updated to reflect the action just taken:

$$z(n+1) = z(n) + \gamma(n+1)(g(\mathbf{1}_{y(n+1)}) - z(n)).$$

5.4.3 Discussion

The main difference between S-MBR and SS-MBR is the manner in which players form estimates of the expected utility sequence $\{U_i(y_i, f_i(z(n)))\}_{n\geq 1},\ \forall y_i \in Y_i$. In S-MBR, players' estimates (see (5.1)) "start afresh" each round of the repeated play—information gathered from sampling in the previous round is discarded, and players draw k_n (with $k_n \to \infty$) new samples in order to form an estimate of the utility for the current round.

This may be considered an inefficient use of information, since the expected utility only changes slightly from one round to the next. In particular, note that the expected utility $U_i(y_i, \cdot)$ is Lipschitz continuous with some Lipschitz constant K, and the increment of the empirical distribution (5.8) is bounded as $\|z(n) - z(n-1)\| \leq M\gamma(n)$ for some constant $M > 0$. Thus, the increment in the expected utility is bounded as

$$|U_i(y_i, f_i(z(n))) - U_i(y_i, f_i(z(n-1)))| \leq KM\gamma(n), \tag{5.3}$$

where, by Assumption 2.2, $\gamma(n) \to 0$. Intuitively speaking, this means that if one has an accurate estimate of the expected utility $U_i(y_i, f_i(z(n-1)))$ in round $(n-1)$, then it is wasteful to wholly discard this information when forming an estimate of the round-n utility, $U_i(y_i, f_i(z(n)))$. The SS-MBR estimation rule leverages the diminishing increment property (5.3) in order to form an accurate estimate using only one sample per round.

The S-MBR estimation rule treats $\{U_i(y_i, f_i(z(n)))\}_{n\geq 1}$ as if it were arbitrarily generated from one round to the next, drawing a completely new set of k_n, $k_n \to \infty$ samples to estimate each $U_i(y_i, f_i(z(n))),\ \forall y_i \in Y_i$ in each round. The SS-MBR estimation rule, on the other hand, treats $\{U_i(y_i, f_i(z(n)))\}_{n\geq 1}$ as if it were *quasi static*, drawing one sample per round, and taking a type of average over time.[4] Because of this, despite drawing only one sample per round, the SS-MBR esti-

[4] Additional insight may be gained by considering the dynamical systems approach to stochastic approximations (e.g., [30]), which allows one to study the behavior of certain discrete-time processes by analyzing an associated differential equation. In such an analysis, an estimation rule such as (5.6) is often considered as a two-time-scale system [116, 117], with the ODE associated with the estimation rule operating at a *faster* rate than ODE associated with the expected-utility process. In the asymptotic analysis of such systems, the slower process may often be treated as effectively static compared to the faster process. We note, however, that the proofs of our results rely on self-contained, martingale-type arguments rather than invoking results from dynamical systems based treatment of stochastic approximation literature.

mate of $U_i(y_i, f_i(z(n)))$ actually utilizes information from n samples, whereas the S-MBR estimate utilizes information from (only) k_n samples. Thus, the SS-MBR estimation procedure actually uses data from *more* samples than SFP when the S-MBR sampling rate is sublinear; i.e., when $\lim_{n\to\infty} k_n/n = 0$. This is reflected in the simulation results (see Section 5.5.4).

Computational Complexity in SS-MBR Algorithms

In a SS-MBR algorithm each player draws one sample from the distribution $f_i(z(n))$ in round n and evaluates the pure-strategy utility at this sample. This gives a complexity of $O(N)$.

5.4.4 SS-MBR: Convergence Result

In this section we prove that a SS-MBR process converges, almost surely, to the internally chain recurrent set of the associated differential inclusion (see Theorem 5.7). We begin by stating the following technical lemma, which follows as a consequence of Toeplitz's lemma [118].

Lemma 5.5. *Let $\{x_n\}_{n\geq 0}$ satisfy $\lim_{n\to\infty} x_n = x$. Let $\{\rho_n\}_{n\geq 0}$ satisfy $0 < \rho_n \leq 1$ and $\sum_{n\geq 0} \rho_n = \infty$. Then, the sequence $\{w_n\}_{n\geq 0}$ given by $w_n = (1-\rho_n)w_{n-1} + \rho_n x_n, \quad n \geq 1$, satisfies $\lim_{n\to\infty} w_n = x$.*

The following lemma characterizes the convergence properties of the estimation procedure used to compute $\widehat{U}(y_i, n)$ in a SS-MBR algorithm.

Lemma 5.6. *Let $\{\rho_n\}_{n\geq 1}$ satisfy $0 < \rho_n \leq 1$, $\sum_{n\geq 1} \rho_n = \infty$ and $\sum_{n\geq 1} \rho_n^2 < \infty$. Let $\{\mathcal{F}_n\}_{n\geq 1}$ be a filtration and let $\{X_n\}_{n\geq 1}$ be a sequence of bounded random variables, adapted to the filtration, say, $|X_n| \leq B$. Let $\mu_n = \mathbb{E}(X_n \,|\, \mathcal{F}_{n-1})$ and assume that $\left(\frac{1}{\rho_n} - 1\right)(\mu_n - \mu_{n-1}) \to 0$ almost surely. Then, the sequence of random variables $\{\widehat{\mu}_n\}_{n\geq 0}$ given by $\widehat{\mu}_n = (1-\rho_n)\widehat{\mu}_{n-1} + \rho_n X_n, \quad n \geq 1$, satisfies $|\widehat{\mu}_n - \mu_n| \to 0$ almost surely.*

Proof. Subtracting μ_n from both sides of $\widehat{\mu}_n = (1-\rho_n)\widehat{\mu}_{n-1} + \rho_n X_n$ gives $E_n = (1-\rho_n)E_{n-1} + \rho_n\left(X_n - \mu_n + \left(\frac{1}{\rho_n} - 1\right)\delta_n\right)$, where $E_n := \widehat{\mu}_n - \mu_n$ and $\delta_n := \mu_{n-1} - \mu_n$, $\delta_0 := 0$.

Introduce the $\mathcal{F}_n$-adapted sequences, for $n \geq 1$

$$\begin{aligned} F_n &= (1-\rho_n)F_{n-1} + \rho_n(X_n - \mu_n), \qquad F_1 = E_1 \\ G_n &= (1-\rho_n)G_{n-1} + \rho_n\left(\frac{1}{\rho_n} - 1\right)\delta_n, \quad G_1 = 0, \end{aligned}$$

and note that $E_n = F_n + G_n$. We will now show that $F_n \to 0$ and $G_n \to 0$ almost surely.

By assumption, $\left(\frac{1}{\rho_n} - 1\right)\delta_n \to 0$ almost surely. Lemma 5.5 applied to $\{G_n\}_{n\geq 1}$ gives $G_n \to 0$ almost surely.

On the other hand,

$$\begin{aligned}\mathbb{E}(F_n^2 \,|\, \mathcal{F}_{n-1}) &= (1-\rho_n)^2 F_{n-1}^2 + \rho_n^2 \mathbb{E}\left((X_n - \mu_n)^2 \,|\, \mathcal{F}_{n-1}\right) \\ &\leq (1-\rho_n)^2 F_{n-1}^2 + \rho_n^2 4B^2 \\ &= \left(1+\rho_n^2\right) F_{n-1}^2 - 2\rho_n F_{n-1}^2 + \rho_n^2 4B^2.\end{aligned}$$

Since $\sum_{n\geq 0} \rho_n^2 < \infty$, the Robbins-Monro Lemma [119] implies that, $\{F_n^2\}_{n\geq 1}$ converges a.s., and $\sum_{n\geq 0} \rho_n F_{n-1}^2 < \infty$. These two properties imply $F_n \to 0$, a.s. $\square$

The following theorem is the main convergence result for SS-MBR learning processes. It shows that if Ψ is a MBR algorithm, then the single sample variant of Ψ achieves learning in the same sense as Ψ.

Theorem 5.7. *Let $\{z(n)\}_{n\geq 1}$ be a SS-MBR process. Assume Assumptions 2.1–2.3 and Assumptions 5.3–5.4 hold. Then $z(n)$ converges to an internally chain recurrent set of* (4.4)*, almost surely.*

Proof. We will prove the result by showing that there exists a sequence $\{\epsilon_n\}_{n\geq 1}$ with $\lim_{n\to\infty} \epsilon_n = 0$ such that $U_i(y_i(n+1), f_i(z(n))) \geq \max_{y_i\in Y_i} U_i(y_i, f_i(z(n))) - \epsilon_n$. By Theorem 4.15, this is sufficient to guarantee that $z(n)$ converges to an internally chain recurrent set of (4.4).

Since, by (5.7), $y_i(n+1) \in \arg\max_{y_i\in Y_i} \widehat{U}_i(y_i, n)$, it is sufficient to show that, for every $i \in N$ and every $y_i \in Y_i$,

$$|\widehat{U}_i(y_i, n) - U_i(y_i, f_i(z(n)))| \to 0 \text{ as } t \to \infty \quad \text{a.s.} \tag{5.4}$$

(Note that the individual action spaces Y_i are finite.) We will show this by invoking the result of Lemma 5.6. To that end, fix $i \in N$ and $y_i \in Y_i$, and let $X(n) := U_i(y_i, \tilde{y}_{-i}(n)),\ t \geq 1$, $\mu(n) := U_i(y_i, f_i(z(n))),\ t \geq 1$, and $\hat{\mu}(n) := \widehat{U}_i(y_i, n),\ n \geq 0$. For $n \geq 0$, let $\mathcal{F}_n := \sigma(\{q_{-i}(s)\}_{s=1}^{n+1})$. Note that $\mu(n)$ is $\mathcal{F}_{n-1}$-measurable and that $E(X(n)|\mathcal{F}_{n-1}) = \mu(n)$.

In order to invoke Lemma 5.6 it is sufficient to show that

$$\left(\frac{1}{\rho(n)} - 1\right)(\mu(n) - \mu(n-1)) \to 0. \tag{5.5}$$

Let $M := \max_{\sigma'_{-i}, \sigma''_{-i}\in\Delta_{-i}} \|\sigma'_{-i} - \sigma''_{-i}\|$, and note that by (5.8), $\|q_{-i}(n) - q_{-i}(n-1)\| \leq M\gamma(n)$.

Since U_i is Lipschitz continuous and f_i is uniformly continuous, there exists a constant K such that

$$\begin{aligned}
&|\mu(n) - \mu(n-1)| = |U_i(y_i, q_{-i}(n)) - U_i(y_i, q_{-i}(n-1))| \\
&\leq K\|q_{-i}(n) - q_{-i}(n-1)\| \leq KM\gamma(n).
\end{aligned}$$

This, together with Assumption 2.2, implies that (5.5) holds.

Thus, $X(n),\ \mu(n),\ \hat{\mu}(n)$, and $\mathcal{F}_n$ as defined above fit the template of Lemma 5.6. By Lemma 5.6, $|\widehat{U}_i(y_i, n) - U_i(y_i, q_{-i}(n))| = |\hat{\mu}(n) - \mu(n)| \to 0$ as $n \to \infty$, verifying that (5.4) holds. □

5.5 Example: Single Sample FP

In this section we use the results of Section 5.4 to study a reduced complexity variant of FP known as *Single Sample FP* (SS-FP).

SS-FP may be applicable as a computationally efficient variant of FP in a variety of settings including large-scale optimization [48], [110]; dynamic programming [52, 53]; traffic routing [49]; cognitive radio [54–56], and learning in Markov decision processes [57]. SS-FP may also be used as a general tool for distributed learning [111] or control [18].

We construct the process using use the notation $q(n)$ to denote the empirical distribution as introduced in Section 2.5, and $\widehat{U}(y_i, n)$, $\rho(n)$, and $\gamma(n)$ as discussed in Section 5.4.1.

We note that the single sample FP process differs from the actorcritic approach of [29] in that the utility estimates for non-played actions are updated as well as those for the played action. [29] also needed to ensure all actions were played infinitely often, but that is not necessary here.

5.5.1 Single Sample FP Algorithm

Initialize

(i) For each $i \in N$, let $y_i(1) \in Y_i$ be arbitrary. Initialize the empirical distribution as $q_i(1) = y_i(1),\ \forall i$, and initialize the utility estimate as $\widehat{U}_i(y_i, 0) = 0,\ \forall y_i \in Y_i,\ \forall i$.

Iterate $(n \geq 1)$

(ii) A single "test action" $\tilde{y}(n)$ is drawn as a (conditionally independent, given $q(n)$) random sample from the distribution $q(n)$, and each player $i \in \mathcal{N}$ updates the estimate $\widehat{U}_i(y_i, n)$ for each action

$y_i \in Y_i$ according to the recursion,[5]

$$\widehat{U}_i(y_i, n) = (1 - \rho(n))\widehat{U}_i(y_i, n-1) + \rho(n)U_i(y_i, \tilde{y}_{-i}(n)). \tag{5.6}$$

(iii) Each player $i \in \mathcal{N}$ chooses a next-stage action using the rule (cf. (2.9), (5.2)):

$$y_i(n+1) \in \arg\max_{y_i \in Y_i} \widehat{U}_i(y_i, n). \tag{5.7}$$

(iv) The empirical distribution for each player $i \in \mathcal{N}$ is updated to reflect the action just taken:

$$q_i(n+1) = q_i(n) + \gamma(n+1)(\mathbf{1}_{y_i(n+1)} - q_i(n)). \tag{5.8}$$

5.5.2 SS-FP Convergence Result

The following result shows that SS-FP achieves learning in the same sense as classical FP (see Section 2.5) and Sampled FP (S-FP) [48].

Corollary 5.8. *Let Γ be a potential game, zero-sum game, or generic $2\times m$ game. Let $\{q(n)\}_{n\geq 1}$ be a SS-FP process, and assume Assumption 2.2 and Assumption 5.3 hold. Then* $\lim_{n\to\infty} d(q(n), NE) = 0$*, almost surely.*

The proof of this result follows from Theorem 5.7 and the discussion immediately preceding Corollary 4.17.

5.5.3 Computational Complexity in SS-FP

In SS-FP, the computation of the next-stage action (see (5.7)) requires players to draw a sample from the distribution $q(n)$ and evaluate the pure-strategy utility of this sample. This computation has a complexity of $O(N)$—a significant reduction from the $O(e^N)$ complexity of classical FP. S-FP, on the other hand, has a complexity of $O(Nn^\beta)$, if we assume the step-size sequence to be of the form $k_n = n^\beta$. The complexity of SS-FP and S-FP scale identically in terms of the number of players. However, due to the dependence on n, S-FP becomes increasingly difficult to implement as the algorithm progresses, whereas, the per-iteration complexity of S-FP is independent of the iteration number.

[5]Since $\tilde{y}_{-i}(n)$ is a pure strategy, the evaluation of the utility is relatively simple. Also note that it is not necessary for each player i to draw a separate "test action" $\tilde{y}_{-i}(n+1)$, although, if desired (for example, in a network-based setting [111]), doing so does not affect the convergence result.

5.5.4 Simulation Results

In order to demonstrate the computational properties of SS-FP and S-FP we simulated both algorithms in a simple traffic routing scenario. Let $\mathcal{N} = \{1, \ldots N\}$ denote a finite set of drivers (or players). Drivers share a common starting point and a common destination and may travel on one of 10 parallel routes. Let the set of routes be denoted by R, and let the action space of player i be given by $Y_i = R,\ \forall i \in \mathcal{N}$. Let $\psi_r(y)$ denote the number of drivers on route r given the joint strategy y. Each route $r \in R$ has an associated cost function $c_r : \mathbb{N} \to \mathbb{R}$ signifying the delay experienced on route r given the number of drivers using the route. Let the utility function of player i be given by $u_i(y) := -c_{y_i}(\psi_{y_i}(y))$. We note that this game is an instance of a congestion game—a known subset of potential games.

We simulated S-FP and SS-FP in this routing scenario with 30 drivers and 10 routes.[6] Figure 5.1 plots the expected utility for SS-FP with parameter $\rho_n = n^{.7}$, and for S-FP with sampling rate given by $k_n = \lfloor n^{\beta} \rfloor$ for several values of β. The plots are averaged over 5 random instantiations. The convergence rate of S-FP tends to be slower than that of SS-FP for values of β less than 1. This is consistent with the observation that SS-FP uses *more* sampled date than S-FP when the S-FP sampling rate is sublinear (see Section 5.4.3).

Figure 5.2 plots the evaluation time for each algorithm. Note that since the complexity of S-FP is dependent on time, the algorithm becomes progressively more difficult to evaluate with each passing iteration.

[6]We note that even in this simple traffic routing scenario, classical FP is extremely impractical to deploy. See, for example, [104] Section VIII-B or [48] for more details.

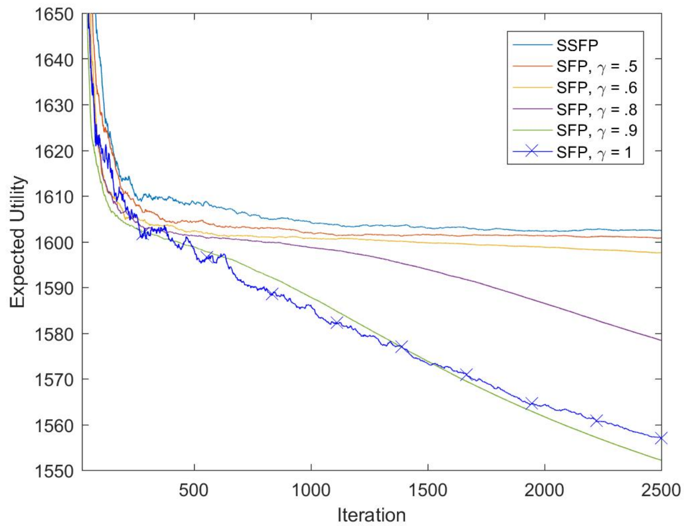


Figure 5.1: Expected utility of $q(n)$.

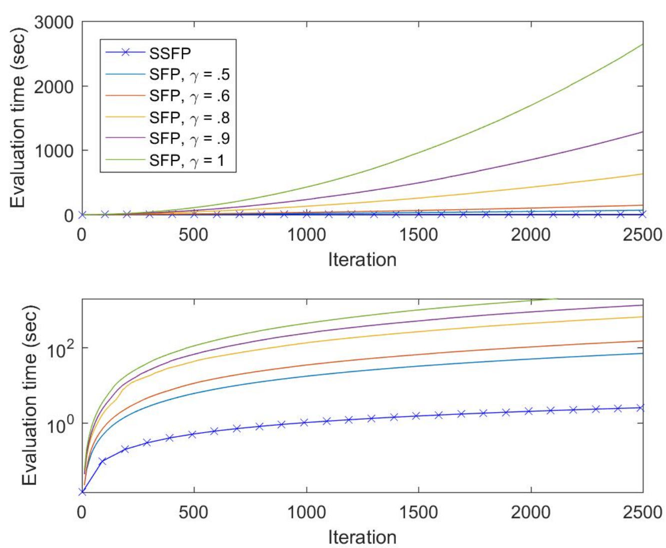


Figure 5.2: Wall-clock evaluation times plotted using linear scale and logarithmic scale.

Chapter 6

Asynchronous Learning

6.1 Introduction

The canonical MBR algorithm presented in Section 2.4 is couched in the framework of synchronous repeated play learning. In this classical learning framework, some game Γ is assumed to be fixed, and players repeatedly face off in the game Γ. At each stage $n \in \mathbb{N}$ of the repeated interaction, players are given the opportunity to revise their strategy.

This learning framework implicitly assumes a form of global synchronization. Each agent must choose their stage-n action before any other agent chooses their stage-$(n+1)$ action. In large scale distributed systems, such global synchronization can be highly impractical.

In this chapter (see also [107, 120], we study a variant of MBR dynamics in which agents are permitted to act in an asynchronous manner. While asynchronous learning schemes would usually be analysed using asynchronous stochastic approximation (e.g. [87]) we show in this chapter that asynchronicity can be handled in a more straightforward manner by simply using our robustness result. In particular, using Theorem 4.15 we develop a mild sufficient condition under which an "asynchronous MBR process" can be shown to converge to the internally chain recurrent set of the associated differential inclusion.

The initial model of asynchronous MBR learning that we study in Section 6.3 is somewhat abstract—it is this feature that allows us to capture a broad range of asynchronous processes. After introducing this model and proving convergence results (Section 6.2–6.3), we then provide two simple examples of highly practical real world models that fall within this framework (Section 6.4).

Subsequently, in Section 6.5 we use the asynchronous MBR learning model to study a variant of MBR dynamics in which players are guaranteed to learn in stronger sense than is typically ensured by MBR convergence results.

Typically, a MBR algorithm is guaranteed to converge in the sense that the observation state $z(n)$ converges to some equilibrium set. This notion of learning is somewhat abstract and does not guarantee that the period-by-period strategies used by players converge to any kind of optimal set themselves. Indeed, there are examples of games where the observation state can converge to equilibrium but the period-by-period strategies used by players always give the lowest possible utility (see Section 6.5.1).

We refer to the form of learning where the observation state $z(n)$ approaches an equilibrium set as *weak learning* and we refer to the form of learning where players period-by-period mixed strategies, denoted here by $g(n)$, converge to an equilibrium set as *strong learning.* In Section 6.5 we present a method that, given a MBR algorithm that achieves weak learning, produces a variant of the same MBR algorithm that is guaranteed to achieve strong learning. The proof of convergence relies on showing that the strongly convergent variant of the algorithm fits the template of an asynchronous MBR algorithm.

We now introduce the notion of asynchronous repeated play learning—a slight modification of classical repeated play learning framework considered in Sections 2.3–2.4.

6.2 Asynchronous Repeated Play Learning

In order to model asynchrony, we consider an extension of the traditional repeated play framework in which players may be "active" in some rounds and "idle" in others.

Let $(X_i(n))_{n\geq 1}$, be a sequence of (deterministic or random) variables $X_i(n) \in \{0,1\}$ indicating the rounds in which player i is active. Let $N_i(n)$ count the number of rounds in which player i has been active up to and including round $n \in \mathbb{N}$; i.e.,

$$N_i(n) := \sum_{s=1}^{n} X_i(s). \tag{6.1}$$

Let $a_i(n) \in A_i$ represent the action chosen by player i in round n. Let the empirical distribution of player i be defined in this setting as

$$q_i(n) := \frac{1}{N_i(n)} \sum_{s=1}^{n} a_i(s) X_i(s), \tag{6.2}$$

and let $q(n) := (q_1(n), \ldots, q_N(n))$.

In order to study asynchronous learning we focus on a special subclass of MBR dynamics in which the observation state is given by the empirical distribution $q(n)$, introduced in Section 2.5. This restriction ensures that the observation state is decoupled across players in a way that allows for asynchronous updating. In particular, $q_i(n)$ need not be updated at the same time (or even rate) as $q_j(n)$ for $i \neq j$.

6.3 Myopic Best-Response Dynamics with Asynchronous Updates

Let $(f_i)_{i=1}^N$ be a family of prediction functions satisfying Assumption 2.3. Within the asynchronous repeated-play framework given above, we say a sequence a $(q(n))_{n\geq 1}$ satisfying (6.2) is a *myopic best-response process with asynchronous updates* (or asynchronous MBR process) if for $n \geq 1$,[1]

$$a_i(n+1) \in \begin{cases} \mathrm{BR}_i(f_i(q_{-i}(n))) & \text{if } X_i(n+1) = 1, \\ a_i(n) & \text{otherwise.} \end{cases} \tag{6.3}$$

This models a scenario in which each player i may update her action in round $(n+1)$ according to traditional myopic best-response dynamics only if $X_i(n+1) = 1$; otherwise, the action of player i persists from the previous round.[2]

We will assume that for each $i \in \mathcal{N}$, the counting process (6.1) satisfies

Assumption 6.1.
(i) For each $i \in \mathcal{N}$ there holds $\lim_{n\to\infty} N_i(n) = \infty$*,*
(ii) For all $i, j \in \mathcal{N}$ there holds, $\lim_{n\to\infty} \frac{N_i(n)}{N_j(n)} = 1$.

Part (i) in the above assumption ensures that players are active in infinitely many rounds. Part (ii) ensures that the number of actions taken by each player remain relatively close; in effect, (ii) ensures that players obtain a weak form of synchronization.

The following theorem is the main theoretical result of this Section. It shows that under the above assumptions, a MBR process with asynchronous updates converges in the same sense as its synchronous counterpart.

Theorem 6.2. *Let $\Psi = (\{\gamma(n)\}_{n\geq 1}, g, (f_i)_{i=1}^N)$ be a MBR algorithm satisfying Assumptions 2.1–2.3. Let $(q(n))_{n\geq 1}$ be an asynchronous MBR process generated from Ψ, and assume that Assumption 6.1 holds. Then $(q(n))_{n\geq 1}$ converges to the chain recurrent set of the associated differential inclusion (4.4).*

[1] Let $X_i(1) = 1$, $\forall i$ and let the initial action $a_i(1)$ be chosen arbitrarily for all i. Moreover, for convenience in notation we have used an inclusion in (6.3). However, if $X_i(n+1) \neq 1$, then the inclusion should be interpreted as an equality: $a_i(n+1) = a_i(n)$.

[2] Note that classical FP may be seen as a special case within this framework with $X_i(n) = 1$, $\forall i, n$.

We will prove this theorem after briefly discussing some examples.

6.3.1 Examples

Fictitious Play

In fictitious play, the prediction functions $(f_i)_{i\in\mathcal{N}}$ and observation map g are given as in Example 4.12 in Section 4.3. In this case, it has been shown that the internally chain recurrent set of the associated differential inclusion is contained in the set of NE for two-player zero-sum games [64], generic $2\times N$ games [35], and potential games [28]. Hence, we get the following corollary to Theorem 6.2.

Corollary 6.3. *Let Γ be a two-player zero-sum game, generic $2\times N$ game, or potential game. Let $(q(n))_{n\geq 1}$ be an asynchronous FP process, and assume that Assumption 6.1 holds. Then* $\lim_{n\to\infty} d(q(n), NE) = 0$.

Empirical Centroid FP

Assume that all players have an identical action space (Assumption 3.1) and Γ is a potential game with permutation invariant potential function (Assumption 3.2). In an ECFP process, the prediction functions $(f_i)_{i\in\mathcal{N}}$ and the observation map g are given as in Example 4.14 in Section 4.3.. In Section 4.5 it was shown that under these assumptions, the internally chain recurrent set of the associated differential inclusion (3.11) is contained in the set of MCE. Hence, we get the following corollary to Theorem 6.2.

Corollary 6.4. *Let Γ be game satisfying Assumptions 3.1–3.2. Let $(q(n))_{n\geq 1}$ be an asynchronous ECFP process, and assume Assumption 6.1 holds. Then* $\lim_{n\to\infty} d(q(n), MCE) = 0$.

Remark 6.5. *Under the assumptions Assumption 3.1 and Assumption 3.2, if $\sigma \in MCE$ then using the permutation invariance of the potential function we get $\bar{\sigma} \in CNE$. Hence, the above theorem also implies that in an asynchronous ECFP process, the centroid distribution converges to the set of consensus NE; that is,* $\lim_{n\to\infty} d(\bar{q}(n), CNE) = 0$.[3]

6.3.2 Proof of Theorem 6.2

In order to prove Theorem 6.2 we will study an underlying (synchronous) MBR process that is embedded in the asynchronous myopic BR process $(q(n))_{n\geq 1}$ defined in (6.2)–(6.3). We begin by

[3]We note that similar results can be proven for the more general version of ECFP with permutation invariant classes. However for simplicity, we focus our examples in this section on the basic formulation of ECFP. See [121] for more details about the general case.

presenting some additional definitions that allow us to study the embedded process.

In particular, for $s \in \mathbb{N}$ define the following terms:

$$\begin{aligned}
\tau_i(s) &:= \sup\{n : \mathbb{N} : \; N_i(n) \leq s\} \\
\tilde{a}_i(s) &:= a_i(\tau_i(s)) \\
\tilde{a}(s) &:= (\tilde{a}_1(s), \ldots, \tilde{a}_n(s)) \\
\tilde{q}_i(s) &:= q_i(\tau_i(s)) \\
\tilde{q}(s) &:= (q_1(s), \ldots, q_n(s)) \\
\hat{q}^i_j(s) &:= q_j(\tau_i(s+1) - 1) \\
\hat{q}^i(s) &:= (\hat{q}^i_1(s), \ldots, \hat{q}^i_n(s)).
\end{aligned}$$

In words, the term $\tau_i(s)$ denotes the round number when player i is active for the s-th time. The terms marked with a $\sim$ correspond to the embedded (synchronous) MBR process that we will study in the proof of Theorem 6.2.

When studying the embedded (synchronous) MBR process $(\tilde{a}(s))_{s\geq 1}$, it will be important to characterize the terms to which players are "best responding". With this in mind, note that per (6.3), the action at time $\tau_i(s+1)$ is chosen as $a_i(\tau_i(s+1)) \in \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, f_i(q(\tau_i(s+1)-1)))$. Thus, by construction, the $(s+1)$-th action of player i in the embedded (synchronous) FP process is chosen as $\tilde{a}_i(s+1) \in \mathrm{BR}_i(f_i(\hat{q}^i(s)))$. In the embedded (synchronous) FP process, the term $\tilde{q}_j(s)$ may be thought of as the "true" empirical distribution of player j, and the term $\hat{q}^i_j(s)$ may be thought of as an estimate which player i maintains of $\tilde{q}_j(s)$, and the term $\hat{q}^i(s)$ (note the superscript) may be thought of as player i's estimate of the joint empirical distribution $\tilde{q}(s)$ at the time of player i's $(s+1)$-th best response. Loosely speaking, if we can show that $\hat{q}^i(s) \to \tilde{q}(s), \; \forall i$, then convergence of the embedded process $(\tilde{q}(s))$ (and eventually the original process $(q(n))$) will follow from the robustness result, Theorem 4.15.

Before proceeding to the proof of Theorem 6.2, we point out a few useful properties that will be helpful in the proof. Note that for $i \in \mathcal{N}$ and $s \in \{1, 2, \ldots\}$, we have

$$N_i(\tau_i(s)) = s, \tag{6.4}$$

and for $i \in \mathcal{N}$ and $t \in \{1, 2, \ldots\}$ we have

$$X_i(n) = 1 \implies \tau_i(N_i(n)) = n. \tag{6.5}$$

Furthermore, note that $X_i(n) = 0$ implies that $N_i(n) = N_i(n-1)$, and in particular,

$$X_i(n) = 0 \implies q_i(n) = q_i(n-1). \tag{6.6}$$

These facts are readily verified by conferring with the definitions of τ_i, N_i, and X_i.

We now prove Theorem 6.2.

Proof. Let $E \subset \Delta^N$ denote the internally chain recurrent set of (4.4). As a first step, we wish to show that $\lim_{s\to\infty} d(\tilde{q}(s),\ E) = 0$. We accomplish this by invoking Theorem 4.15. In particular, we wish to show that there exists a sequence $(\epsilon_s)_{s\geq 1}$ such that $\lim_{s\to\infty} \epsilon_s = 0$ and

$$U_i(a_i(s+1), f_i(\tilde{q}(s))) \geq \max_{y_i \in Y_i} U_i(\alpha_i, f_i(\tilde{q}(s))) - \epsilon_s,\ \forall s \geq 1. \tag{6.7}$$

To that end, for $i \in \mathcal{N}$ define $v_i : \Delta_{-i} \to \mathbb{R}$ by $v(q) := \max_{y_i \in Y_i} U_i(\alpha_i, f_i(q))$, and note that by (6.3), we have $U_i(a_i(\tau_i(s+1)), f_i(q(\tau_i(s+1)-1))) = v_i(q_{-i}(\tau_i(s+1)-1))$, or equivalently by the definitions of $\tilde{a}(s)$ and $\hat{q}^i(s)$,

$$U_i(\tilde{a}_i(s+1)), f_i(\hat{q}^i(s))) = v_i(\hat{q}^i(s)).$$

Using Lemma 6.20 in the appendix, it is straightforward to verify that $\lim_{s\to\infty} \|\hat{q}^i(s) - \tilde{q}(s)\| = 0$. Since U_i is Lipschitz continuous and f_i uniformly continuous, this gives $\lim\limits_{s\to\infty} |U_i(\tilde{a}_i(s+1)), f_i(\tilde{q}(s))) - v_i(f_i(\tilde{q}(s)))| = 0,\ \forall i$; i.e., there exists a sequence $(\epsilon_s)_{s\geq 1}$ such that $\epsilon_s \to 0$ and (6.7) holds. It follows by Theorem 4.15 that

$$\lim_{s\to\infty} d(\tilde{q}(s),\ E) = 0. \tag{6.8}$$

We now show that $\lim_{n\to\infty} d(q(n),\ NE) = 0$. Let $\varepsilon > 0$ be given. By Lemma 6.20 (see appendix), for each $i \in \mathcal{N}$ there exists a time $S_i > 0$ such that $\forall s \geq S_i$, $\|q(\tau_i(s)) - \tilde{q}(s)\| < \frac{\varepsilon}{2}$. Let $S' = \max_i\{S_i\}$. By (6.8) there exists a time S'' such that $\forall s \geq S''$, $d(\tilde{q}(s),\ NE) < \frac{\varepsilon}{2}$. Let $S = \max\{S', S''\}$. Then

$$d(q(\tau_i(s)),\ NE) < \varepsilon,\ \forall i,\ \forall s \geq S. \tag{6.9}$$

Let $T = \max_i\{\tau_i(S)\}$. Note that for some i, $q(T) = q(\tau_i(S))$, and hence by (6.9),

$$d(q(T),\ NE) < \varepsilon. \tag{6.10}$$

Also note that for any $n_0 > T$, it holds that $N_i(n_0) \geq S$ (since $N_i(\tau_i(S)) = S$, and $N_i(n)$ is non-decreasing in n), and moreover

$$\begin{aligned} X_i(n_0) = 1 \text{ for some } i &\implies q(n_0) = q(\tau_i(N_i(n_0))), \\ X_i(n_0) = 0 \text{ for all } i &\implies q(n_0) = q(n_0 - 1), \end{aligned} \tag{6.11}$$

where the first implication holds with $N_i(n_0) \geq S$. In the above, the first line follows from (6.5), and the second line follows from (6.6). Consider $n \geq T$. If for some i, $X_i(n) = 1$, then by (6.11) and (6.9), $d(q(n),\ NE) = d(q(\tau_i(N_i(n))),\ NE) < \varepsilon$. Otherwise, if $X_i(n) = 0\ \forall i$, then $q(n) = q(n-1)$.

Iterate this argument m times until either (i) $X_i(n-m) = 1$ for some i, or (ii), $t - m = T$. In the case of (i), $d(q(n),\ NE) = d(q(n-m),\ NE) = d(q(\tau_i(N_i(n-m))),\ NE) < \varepsilon$, where the inequality again follows from (6.9) and the fact that $n - m > T \implies N_i(n-m) \geq S$. In the case of (ii), $d(q(n),\ NE) = d(q(T),\ NE) < \varepsilon$, where the inequality follows from (6.10). Since $\varepsilon > 0$ was arbitrarily, the result follows. □

6.4 Continuous-Time Embedding of Fictitious Play

The asynchronous MBR algorithm discussed in Section 6.3 is a somewhat abstract discrete-time process. In this section we give a concrete interpretation of the process within a practical setting. In particular, we consider the implementation of the (discrete-time) FP algorithm in a continuous-time setting where agents do not have access to a global clock. Effectively, this results in a discrete-time asynchronous FP process embedded within a continuous-time framework.

For simplicity, we focus here on the FP algorithm. Given an arbitrary MBR algorithm Ψ, identical arguments can be used to prove convergence of the internally chain recurrent set of the associated differential inclusion (4.4).

We first introduce the continuous-time embedding of FP and derive a sufficient condition for convergence using Theorem 6.2. Subsequently, we give two simple and practical implementations that achieve the condition. The example implementations are prototypical in that one uses a synchronization rule that is entirely stochastic, and the other, entirely deterministic.

As in the previous models of repeated play learning (see Sections 2.3 and 6.2), assume each player executes a (countable) sequence of actions (or strategies) $(a_i(n))_{n\geq 1}$. Furthermore, assume that each action is taken at some instant in real time $t \in [0,\infty)$ as measured by some universal clock.[4] In particular, for each player $i \in \mathcal{N}$, let $(\tau_i(n))_{n=1}^{\infty} \subset [0,\infty)$ be an increasing sequence where $\tau_i(n)$ indicates the time (as measured by the universal clock) at which player i chooses an action for the n-th time. Let $a_i(n)$ denote the n-th action taken by player i; i.e., the action taken by player i at time $t = \tau_i(n)$. For $t \in [0,\infty)$, let $N_i(t) = \sup\{n : \ \tau_i(n) \leq t\}$ denote the number of actions taken by player i by time t. For $t \in [0,\infty)$, we define the empirical distribution of player i in this settings as $\mathbf{q}_i(t) := \frac{1}{N_i(t)}\sum_{k=1}^{N_i(t)} a_i(k)$. In particular, for $t \in [0,\infty)$, let $\mathbf{q}_i(t_-) := \lim_{\tilde{t}\uparrow t} \mathbf{q}_i(\tilde{t})$.

In this context, we say the sequence $(a_i(n))_{n\geq 1}$ is an asynchronous FP action process if for $n \geq 1$ each player i chooses their stage-n action according to the rule:[5]

$$a_i(n) \in \mathrm{BR}_i(q_{-i}(\tau_i(n)_-)).$$

We call the sequence $(\tau_i(n))_{n\geq 1}$ the action-timing process for player i, and we refer to any method used to generate $(\tau_i(n))_{n\geq 1}$ (whether deterministic or stochastic) as an action timing rule. Together, we refer to the joint sequence $(\tau_i(n), a_i(n))_{i\in\mathcal{N},n\geq 1}$ as a continuous-time embedded FP process.

The following assumption provides a sufficient condition on the action-timing process in order to ensure convergence of the continuous-time embedded FP process. The assumption is essentially a restatement of Assumption 6.1, but in a continuous-time setting.

Assumption 6.6.
(i) For each i there holds $\lim_{t\to\infty} N_i(t) = \infty$,
(ii) for each i, j there holds $\lim_{t\to\infty} N_i(t)/N_j(t) = 1$.

Part (i) of the above assumption may be satisfied, for instance, as long as the clock skew of each agent stays bounded (with respect to the universal clock), and each agent takes actions infinitely often with respect to their local clock. In order to ensure (ii) is satisfied, slightly more care is needed, as demonstrated by the specific application scenarios below.

The following theorem demonstrates that if the action-timing sequence is chosen to satisfy Assumption 6.6, then the continuous-time embedding of FP will converge to the set of NE.

Theorem 6.7. *Let Γ be a two-player zero-sum game, generic $2 \times N$ game, or potential game. Suppose that $(a_i(n), \tau_i(n))_{i\in\mathcal{N},\ n\geq 1}$ is a continuous-time embedding of FP satisfying Assumption 6.6. Then players learn NE strategies in the sense that $\lim_{t\to\infty} d(\mathbf{q}(t), NE) = 0$.*

[4]We use the term "universal clock" to refer to some reference clock by which we can compare the timing of actions taken by individual players. However, the universal clock is merely an artifice for analyzing the process, and we do not suppose that players have any particular knowledge concerning it.

[5]Let $\tau_i(1) = 0$ for all i, and let the initial action $a_i(1)$ be chosen arbitrarily for all i.

The proof of Theorem 6.7 follows readily from Corollary 6.3.

In the following two subsections, we give two simple examples of action-timing rules that illustrate different methods for satisfying Assumption 6.6 (and hence achieving NE learning in the continuous-time embedded FP process).

6.4.1 Independent Poisson Clocks

Let $w_i(n) = \tau_i(n+1) - \tau_i(n)$ denote the stage-n "waiting time" for player i. Suppose that for each player i and $n \geq 1$, $w_i(n)$ is an independent random variable with distribution $w_i(n) \sim exp(\lambda)$, where $\lambda > 0$ is some parameter that is common among all i. In this case, the action-timing process $(\tau_i(n))_{n\geq 1}$ is said to be a homogenous Poisson process.

The following theorem shows that if the action-timing process is randomly generated in this manner, then players will achieve NE learning.

Theorem 6.8. *Let* Γ *be potential game. Suppose that* $(a_i(n), \tau_i(n))_{i\in\mathcal{N},\ n\geq 1}$ *is a continuous-time embedding of FP with the action-timing sequences* $(\tau_i(n))_{n\geq 1}$ *are generated as independent homogenous Poisson processes with common parameter* λ*. Then players learn NE strategies in the sense that* $\lim_{t\to\infty} d(\mathbf{q}(t), NE) = 0$*, almost surely.*

Proof. By Theorem 6.2 it is sufficient to show that $\lim_{t\to\infty} N_i(t) = \infty,\ \forall i$, and $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$ for all i, j.

First, note that for any i and $n \geq 1$, $w_i(n) < \infty$ almost surely. Hence, $\tau_i(n) = \sum_{k=1}^n w_i(k) < \infty$ for all i, almost surely. Equivalently, for any $M > 0$, almost surely there exists a (random) time $T > 0$ such that $N_i(t) \geq M$ for all $t \geq T$. Hence, $\lim_{t\to\infty} N_i(t) = \infty$, almost surely.

Now we show that $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$ for all i, j. Let $\tau(1) := \min_i \tau_i(1)$ and let $\mathcal{T}_1 := (\tau_i(n))_{i\in\mathcal{N},n\geq 1}\backslash\tau(1)$. For $n \geq 2$, let $\tau(n) := \min \mathcal{T}_{n-1}$ and let $\mathcal{T}_n := \mathcal{T}_{n-1}\backslash\tau(n)$. In this manner, we produce the sequence $(\tau(n))$. For $n \geq 1$, $i \in \mathcal{N}$, define $X_i(n) \in \{0,1\}$ to be an indicator variable with $X_i(n) = 1$ if $\tau(n) \in (\tau_i(k))_{k\geq 1}$ and $X_i(n) = 0$ otherwise.

Let $\mathcal{F}_0 := \emptyset$ and for $n \geq 1$, let $\mathcal{F}_n := \sigma((\tau(k))_{k=1}^n)$. For $n \geq 1$ let $\xi_i(n) := \mathbb{P}(X_i(n) = 1 |\ \mathcal{F}_{n-1})$.

Since for each i, $(\tau_i(n))_{n\geq 1}$ is a Poisson process with common parameter λ, there holds $\xi_i(n) = \frac{1}{N}$ for all i and n.[6] By Levi's extension of the Borel-Cantelli Lemma (see [122], p.124) there holds

$$\lim_{n\to\infty} (\sum_{k=1}^{n} X_i(n))/(\sum_{k=1}^{n} \xi_i(n)) = 1, \text{ a.s.} \tag{6.12}$$

[6]Recall that N denotes the number of players.

Note that for each i, $\sum_{k=1}^{n} X_i(k) = N_i(\tau(n))$ and $\sum_{k=1}^{n} \xi_i(n) = \frac{n}{N}$. Thus by (6.12),

$$\lim_{n\to\infty} \frac{N_i(\tau(n))}{N_j(\tau(n))} = \lim_{n\to\infty} \frac{N_i(\tau(n))}{n/N} \frac{n/N}{N_j(\tau(n))} = 1, \text{ a.s., } \forall i,j.$$

Finally, note that $\lim_{n\to\infty} \tau(n) = \infty$ a.s., and for each i $N_i(t)$ is constant on $[0,\infty)\backslash(\tau(n))_{n\geq 1}$. Thus, $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$, almost surely. □

6.4.2 Adaptive Clock Rates

In this section we consider a scenario in which each player chooses the timing of her actions (deterministically) according to a personal clock with a skew rate that may be different among players.

Let $w_i(n) = \tau_i(n+1) - \tau_i(n)$ again denote the stage-n "waiting time" for player i. For each i, let $w_{i,0}$ denote a base waiting time for player i. The base waiting time of player i may be interpreted as the amount of time which expires according to the universal clock during one unit of time as measured by player i's personal clock. The disparity in the $w_{i,0}$ thus reflects disparate skew rates among players' personal clocks.

Let $N_{min}(t) := \min_i N_i(t)$. At time t, we suppose that player i has knowledge of $N_{min}(s)$ at the time instances $s \in \{kw_{i,0} : k \in \mathbb{N}, kw_{i,0} \leq t\}$. (I.e., player i is aware of the value of N_{min} at instances when her "clock ticks".) For each $i \in \mathcal{N}$, let $B_i \in \mathbb{R}$ be a number satisfying $B_i > \max_i w_{i,0}$.

Suppose that player i adaptively chooses her stage-n waiting time according to the rule:

$$w_i(n) = \min\left\{kw_{i,0} : \ k \in \mathbb{N},\, N_{min}(\tau_i(n) + kw_{i,0}) \geq N_i(\tau_i(n)) - B_i\right\}. \tag{6.13}$$

In words, this rule may be described as follows: Player i periodically observes $N_{min}(t)$. If $N_i(t) - N_{min}(t) \leq B_i$ then player i takes a new action. If $N_i(t) - N_{min}(t) > B_i$ then player i waits for $N_{min}(t)$ to increase sufficiently (satisfying $N_i(t) - N_{min}(t) \leq B_i$) before taking a new action.

Theorem 6.9. *Let Γ be a potential game. Suppose that $(a_i(n), \tau_i(n))_{i\in\mathcal{N},\, n\geq 1}$ is a continuous-time embedding of FP in which the action-timing sequence $(\tau_i(n))_{n\geq 1}$ is generated according to the adaptive rule* (6.13). *Then players learn NE strategies in the sense that* $\lim_{t\to\infty} d(\mathbf{q}(t), NE) = 0$.

Proof. By Theorem 6.2, it is sufficient to show that $\lim_{t\to\infty} N_i(t) = \infty$ for some (and hence all) $i \in \mathcal{N}$, and that $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$.

Note that for $i^* \in \arg\max_i w_{i,0}$, there holds $N_{i^*}(t) = \lfloor \frac{t}{w_{i^*,0}} \rfloor + 1$, and hence $\lim_{t\to\infty} N_{i^*}(t) = \infty$. Furthermore, by construction, $|N_i(t) - N_{i^*}(t)| \leq 2\max_i B_i$ for all i and for all $t \geq 0$. Hence,

$\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$, for all i, j. $\square$

6.5 From Weak to Strong Learning

In a MBR algorithms learning generally occurs in the sense that the observation state $z(n)$ converges to an equilibrium set. This can be a weak notion of learning since it does not guarantee that players period-by-period strategies converge to any kind of equilibrium set themselves.

For example, in FP it can happen that the empirical distribution converges to the set of NE, but the period-by-period strategies used by players in each round always yield the lowest possible utility (see Section 6.5.1).

In this section we study subclass of MBR algorithms where the observation state is given by the empirical distribution $q(n)$ as defined in Section 2.5, and we consider learning within *synchronous* repeated play framework defined in Section 2.3. Indeed, although the focus of this chapter is on asynchronous learning, the processes considered in this chapter all assume the synchronous repeated play framework of Section 2.3. As will be shown in Section 6.5.3, the synchronous learning processes considered in this section contain an *embedded* asynchronous learning process which will be amenable to the convergence results derived in Section 6.3. Hence, we will be able to prove the main result of this section (Theorem 6.16) as an application of our main convergence result for asynchronous MBR processes, Theorem 6.2.

We refer to convergence of the empirical distribution (or some function thereof) to an equilibrium set as weak convergence, and we refer to any form of learning involving weak convergence as weak learning. We refer to the convergence of players' period-by-period strategies to an equilibrium set as strong convergence, and we refer to any form of learning involving strong convergence as strong learning. Intuitively speaking, weak learning means that players learn an equilibrium strategy in some abstract sense (i.e., convergence in empirical distribution) but may never actually implement the strategy they are learning. In strong learning, not only do players *learn* an equilibrium strategy, but they also implement it.

Weak learning is inherently problematic (e.g., [59, 99], [74], ch. 3.5), particularly in the context of distributed control. Strong learning resolves many fundamental issues of weak learning and, in particular, is better suited to control applications. Among other motivations, strong learning of mixed strategies can be important in order to achieve security guarantees [123–125], and to effectively achieve learning in algorithms whose limit points often include mixed strategies, e.g., [104, 126], and [74], ch. 3.5. For example, Section 6.5.4 discusses an example achieving strong convergence in the ECFP algorithm, the limit points of which are often mixed equilibria.

The main contribution of this section is the presentation of a method for taking a weakly

	A	B
A	$\sqrt{2}, 1$	$0, 0$
B	$0, 0$	$1, \sqrt{2}$

Figure 6.1: Payoff matrix in miscoordination game.

convergent MBR learning algorithm, and constructing from it, a strongly convergent variant [127].

Using the convergence result for asynchronous MBR processes (Theorem 6.2), we show that if a MBR achieves weak learning, then a strongly convergent variant can be constructed from it. As an example of how the general result may be applied, we construct the strongly convergent variant of the ECFP algorithm presented in Chapter 3.

The remainder of the section is outlined as follows. Section 6.5.1 presents a motivating example. Section 6.5.2 presents the strongly convergent variant of classical FP. Section 6.5.3 presents the strongly convergent variant of a general MBR algorithm, and Section 6.5.4 presents an example of strong convergence in ECFP.

6.5.1 Weak Convergence in Fictitious Play

The following example (see [74], p. 78), while fairly simple, clearly illustrates the phenomenon of weak convergence in FP and demonstrates why weak convergence can be a unsatisfactory notion of learning.

Consider the two-player asymmetric coordination game shown in Figure 6.1. The game has three Nash equilibria: both players play A, both players play B, and an asymmetric mixed-strategy Nash equilibrium. The game is a potential game [38] and hence falls within the purview of Theorem 2.5—regardless of the initial conditions, players engaged in a FP process will learn an equilibrium in the weak sense that $d(q(n),\ NE) \to 0$ as $n \to \infty$.

Suppose that the players are engaged in a FP process on this game, and in the first round they miscoordinate their actions (e.g., one chooses A, and the other chooses B). Young [74] shows the somewhat counterintuitive result that the FP dynamics will in fact lead players to miscoordinate their action choices in every subsequent round of the learning process. Thus, despite the fact that $\lim_{n\to\infty} d(q(n),\ NE) = 0$, the players' realized action choices are extremely suboptimal—yielding the lowest possible utility in each round of play. Intuitively speaking, this phenomenon occurs when players' actions cycle in such a way as to drive the time-averaged empirical distribution to a mixed-strategy Nash equilibrium, yet player's period-by-period strategies never constitute (nor even approach) a Nash equilibrium themselves.[7]

It may be said that in weak learning players "learn" a NE strategy in some abstract sense, but never actually implement the strategy they are learning. In strong learning, players not only learn a NE strategy, but they also physically implement the strategy that is being learned.

The following section presents a simple modification of FP that achieves strong learning; i.e., players' period-by-period strategies converge to equilibrium in addition to convergence of the empirical distributions.

6.5.2 Strongly Convergent Variant of Fictitious Play

Consider a variant of FP in which the action for player i in stage n is chosen by drawing a random sample from the mixed strategy (i.e., probability distribution) $g_i(n)$, where

$$g_i(n) \in \mathrm{BR}_i(q_{-i}(n-1))\rho_i(n) + q_i(n-1)(1-\rho_i(n)), \tag{6.14}$$

$\rho_i(n) \in [0,1]$, and $\lim_{n\to\infty} \rho_i(n) = 0$. Intuitively, this is similar to the classical FP process (2.9), but rather than playing a deliberate best response each round, players gradually transition toward drawing their stage-n action as a random sample from their own empirical distribution, $q_i(n)$.

The idea is that players will play a best response sufficiently often so that, per FP, the empirical distribution $q(n)$ will be driven toward equilibrium, as in Theorem 2.5. Then, since $\rho_i(n) \to 0$ as $n \to \infty$, the mixed strategy $g_i(n)$ tends towards $q_i(n)$, which is itself tending towards equilibrium. Informally, (6.14) captures the main idea of strongly convergent FP. A formal presentation of the algorithm is given below.

Strongly Convergent Variant of Classical FP

Suppose the action for player i at time n is chosen according to the following randomized rule:

$$a_i(n) \sim \begin{cases} b_i(n-1), & \text{if } X_i(n) = 1, \\ q_i(n-1), & \text{otherwise,} \end{cases} \tag{6.15}$$

[7] It must be noted that this example looses some potency in light of Theorem 10.1 in Chapter 10. In almost all potential games (including this example game), for almost all initial conditions, the mean-field FP dynamics converge to a pure-strategy NE. In this example game, there does exists a low-dimensional (Lebesgue measure zero) set, from which a FP path may converge to a mixed equilibrium. The initial conditions assumed in this example lie on this low dimensional set.

where

$$b_i(n-1) \in \mathrm{BR}_i(q_{-i}(n-1)),$$

for probability mass function h over Y_i, the notation $a_i(n) \sim h$ indicates that the action $a_i(n)$ is drawn as a random sample [8] from h, $X_i(n) \in \{0,1\}$ is a random variable, and $q_i(n)$ is the player's empirical distribution as defined in (6.16) below. Let $\mathcal{F}_n := \sigma(\{a(s), X_1(s), \ldots, X_n(s), b_1(s), \ldots, b_n(s)\}_{s \leq t})$, and note that $b_i(n-1)$ and $q_i(n-1)$ are $\mathcal{F}_n$-measurable. Let

$$\rho_i(n) := \mathbb{P}(X_i(n) = 1 | \ \mathcal{F}_{n-1}),$$

and note that $\rho_i(n)$ is $\mathcal{F}_{n-1}$-measurable. Intuitively speaking, $\rho_i(n)$ represents the probability that player i deliberately chooses to play a best-response strategy in round n given the history of play up through the previous round. We make the following assumptions regarding each player's probability of deliberately choosing a best response:

Assumption 6.10. $\lim_{n\to\infty} \rho_i(n) = 0, \ \forall i \in \mathcal{N}, \ a.s.,$

Assumption 6.11. $\sum_{t \geq 1} \rho_i(n) = \infty, \ \forall i \in \mathcal{N}, \ a.s.,$

Assumption 6.12. $\lim_{n\to\infty} \frac{\sum_{k=1}^n \rho_i(k)}{\sum_{k=1}^n \rho_j(k)} = 1, \ \forall i,j \in \mathcal{N}, \ a.s.$

The first assumption ensures that players eventually transition towards playing their next-stage action as a sample from their empirical distribution rather than playing a deliberate best response. The second assumption ensures that, for each player, a deliberate best response is played infinitely often. The third assumption ensures that the number of deliberate best responses taken by each player remain relatively in sync.[9]–6.12. In practice, players may choose their deliberate best responses completely asynchronously; for example, setting $\rho_i(n) = 1/n^r, \ \forall i$, with $r \in (0,1]$, results in (purely) independent sampling of deliberate best-response rounds and secures Assumptions 6.10–6.12.

Let

$$N_i(n) := \sum_{k=1}^n X_i(k)$$

[8]The action $a_i(n) \in A_i$ is technically a dirac distribution over the finite action space Y_i (see Section 2.1). More precisely, the notation $a_i(n) \sim h$ means that an action $y_i(n)$ is drawn as a random sample from h with $a_i(n) := \delta_{y_i(n)}(y_i)$, where $\delta_{y_i(n)}(y_i) = 1$ if $y_i = y_i(n)$ and $\delta_{y_i(n)}(y_i) = 0$ otherwise.

[9]Note that since $\rho_i(n)$ is only required to be $\mathcal{F}_{n-1}$-measurable, this parameter is in fact adaptively tunable. This is a feature of practical interest since it allows players to adjust their deliberate best-response rates on the fly—possibly adapting to the (initially unknown) deliberate best-response rates of others and to underlying process dynamics—in order to satisfy Assumptions 6.10

count the number of times player i has deliberately played a best response until and including round n. Note that $N_i(n)$ is $\mathcal{F}_n$-measurable. The empirical distribution $q_i(n)$ is defined recursively as[10]

$$q_i(n+1) = q_i(n) + \frac{1}{N_i(n+1)}\left(a_i(n+1) - q_i(n)\right) X_i(n+1). \tag{6.16}$$

Intuitively speaking, the empirical distribution (6.16) is updated only over rounds when a deliberate best response is played. Note that $q_i(n)$ is $\mathcal{F}_n$-measurable.

In order for players to compute $q_i(n)$ it is necessary that in addition to having knowledge of the actions $a_i(n)$, players have knowledge of $X_i(n)$. Formally we assume:

Assumption 6.13. *At time n, each player observes $X_i(n)$ for all $i \in \mathcal{N}$.*

This may be thought of as a (one bit) tag that is attached to the action of each player indicating if the action is taken as a deliberate best response or not.

Finally, let[11]

$$g_i(n) := b_i(n-1)\rho_i(n) + q_i(n-1)(1-\rho_i(n)), \tag{6.17}$$

and note that $g_i(n)$ is $\mathcal{F}_{n-1}$ measurable.[12] More importantly, note that for every $\alpha_i \in A_i$, $g_i(\alpha_i, t) = \mathbb{P}(a_i(n) = \alpha_i | \mathcal{F}_{n-1})$, and thus $g_i(n)$ represents the mixed strategy (conditioned on past play) used by player i in round n. The joint mixed strategy used in round n is given by $g(n) := (g_1(n), \dots, g_n(n))$.

We refer to a process where, for each player i, $a_i(n)$ is selected according to (6.15), $q_i(n)$ is calculated according to (6.16), and $g_i(n)$ is updated according to (6.17) as the strongly convergent variant of (classical) FP (for reasons to be clear soon).

The following result states that in the strongly convergent variant of FP, players' period-by-period mixed strategies converge to the set of Nash equilibria—i.e., strong learning is achieved.[13]

Corollary 6.14. *Let Γ be a two-player zero-sum game, potential game, or generic $2 \times m$ game. Assume Assumptions 6.10–6.13 hold. Then the strongly convergent variant of FP achieves strong learning in the sense that $\lim_{n\to\infty} d(g(n),\ NE) = 0$ almost surely.*

[10]To initialize the process, let the action $a_i(1)$ be chosen arbitrarily, let $q_i(1) = a_i(1)$, and let $X_i(1) = 1$ for all i.

[11]We note that the mixed strategy $g_i(n)$ defined here is identical to that defined originally in (6.14), where we informally introduced the strongly convergent FP process. In order to *precisely* define the strongly convergent FP process (in particular, $q(n)$) it was necessary to define the "intermediate" term b_i, which allowed us to explicitly track deliberate best-response rounds via $X_i(n)$. For clarity, (6.17) redefines $g_i(n)$ in terms of the intermediate variables used in the precise definition of the process.

[12]To see this, note first that $q_i(n-1)$ and $\rho_i(n)$ have been shown to be $\mathcal{F}_{n-1}$ measurable. Furthermore, this implies that $\mathrm{BR}_i(q_i(n-1))$ is $\mathcal{F}_{n-1}$-measurable. Lastly, by construction, $b_i(n) \in \mathrm{BR}_i(q_i(n-1))$ is $\mathcal{F}_{n-1}$-measurable.

[13]In zero-sum games, standard tie-breaking rules should be taken into account when choosing a deliberate best response in order to ensure convergence [33]. In particular, tie-breaking rules ensure weak convergence is achieved in the classical FP process which allows one to apply Theorem 6.16.

	H	T
H	1,-1	-1,1
T	-1,1	1,-1

Figure 6.2: Matching pennies payoff matrix

In order to prove the above result, we first study strong convergence in the more general class of MBR algorithms and then prove the result as a corollary of the general theorem (see Theorem 6.16).

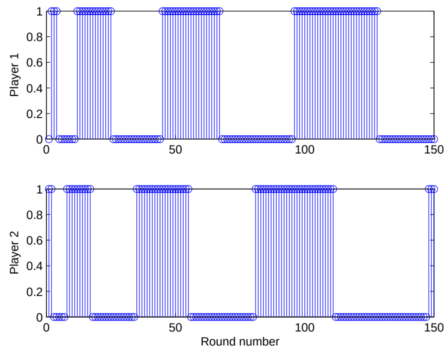
Simulation Example

In order to demonstrate the learning properties of strongly convergent FP, we simulated classical FP and strongly convergent FP in a simple two-player matching pennies game with utility functions as shown in Figure 6.2. The game has a unique (symmetric) mixed-strategy equilibrium in which both players choose either action with probability 1/2. Figure 6.3a shows the period-by-period strategies generated by classical FP. Players' strategies are always pure and progress in continuously lengthening cycles. While the time-averaged empirical distribution is being driven to equilibrium, the period-by-period strategies clearly are not.

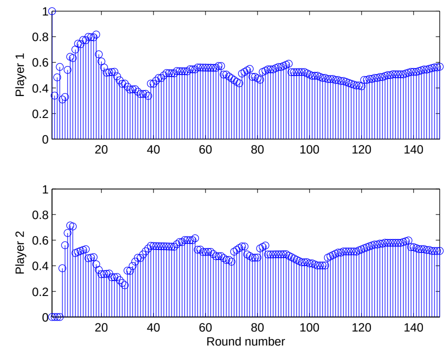
Figure 6.3b shows the period-by-period strategies generated by strongly convergent FP with $\rho(n) = n^{-.35}$. Players' period-by-period strategies are converging to the unique Nash equilibrium of the game.

Figure 6.3c shows the utility received by the realized joint action $a(n)$ in each round of repeated play for both learning algorithms. The received payoffs in classical FP cycle around the value of the game, while the received payoffs in strongly convergent FP converge to the value of the game.

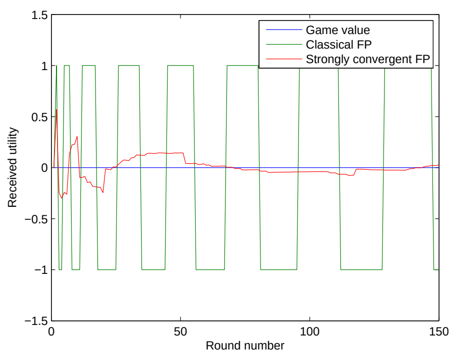
An important tradeoff in strongly convergent FP is that less frequent deliberate best-response actions and less frequent updating of the empirical distribution (see (6.16)) can lead to a slow-down in convergence rate. The empirical distribution processes for player 1 in each algorithm is shown in Figure 6.3d with $\rho(n) = n^{-.35}$.



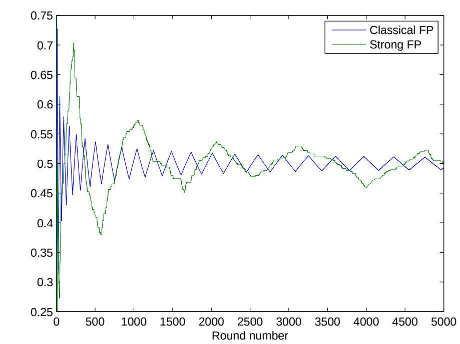
(a)



(b)



(c)



(d)

Figure 6.3: 6.3a: The probability of each player playing heads in round n using the classical FP algorithm, 6.3b: The probability of each player playing heads in round n using the strongly convergent FP algorithm, 6.3c: The received utility in round n given the realized action $a(n)$, 6.3d: The empirical distribution process of the action H (heads) for player 1 in both FP and strongly convergent FP.

6.5.3 Strongly Convergent Variant of a Myopic Best-Response Algorithm

In this section we construct the strongly convergent variant of a MBR algorithm. The construction here is a generalization of that of Section 6.5.2 where we constructed the strongly convergent variant of classical FP.

Let Ψ be a myopic best-response learning algorithm such that the observation state is decoupled across players as in Section 6.2.

For each $i \in \mathcal{N}$, let $(X_i(n))_{t\geq 1}$ be a sequence of random variables with $X_i(n) \in \{0,1\}$. Analogous to Section 6.5.2, $X_i(n) = 1$ will serve to indicate that player i took a deliberate best response in round n. Let

$$N_i(n) := \sum_{s=1}^{n} X_i(s) \tag{6.18}$$

count the number of deliberate best responses taken by player i through n.

In Section 6.5.2 the empirical distribution of player i, (6.16), is a time average taken only over rounds when player i took a deliberate best response. In order to generalize this notion to a MBR algorithm, define the term

$$\tau_i(s) := \inf\{n : N_i(n) = s\}.$$

For $s \geq 1$, $\tau_i(s)$ indicates the round when player i took their s-th deliberate best response,[14] and the sequence $(\tau_i(s))_{s\geq 1}$ gives the subsequence of rounds when player i took a deliberate best response.

Let the empirical distribution of player i at time n be formed as

$$q_i(n) := \frac{1}{N_i(n)} \sum_{s=1}^{n} a_i(s) X_i(s). \tag{6.19}$$

Let the action for player i in round $n \geq 2$ be chosen according to the random rule[15]

$$a_i(n) \sim \begin{cases} b_i(n-1), & \text{if } X_i(n) = 1, \\ q_i(n-1), & \text{otherwise,} \end{cases} \tag{6.20}$$

where $b_i(n-1) \in \mathrm{BR}_i^{\eta_n}(f_i(q(n-1)))$, and assume:[16]

[14] Note that by Assumption 6.11 and Lemma 6.21, $\tau_i(s)$ is finite valued a.s. for any $s \in \{1,\ 2, \ldots\}$.

[15] To initialize the process, let the action $a_i(1)$ be chosen arbitrarily, let $X_i(1) = 1$, and let $\bar{H}(1) = a_i(1)$ for all i.

[16] Note that this assumption subsumes the more typical assumption that $\eta_n = 0,\ \forall n$. By making this more general assumption we are able to handle interesting scenarios that may arise in a practical implementation of the algorithm; e.g., players have some asymptotically decaying error in their knowledge of their utility function or knowledge of opponent's empirical distributions.

Assumption 6.15. *The sequence* $(\eta_n)_{n\geq 1}$ *associated with the strategy sequence* $(b_i(n))_{n\geq 1}$ *of* (6.20) *is such that* $\lim_{n\to\infty} \eta_n = 0$.

Let $\mathcal{F}_n := \sigma(\{a(s), X_1(s), \ldots, X_n(s), b_1(s), \ldots, b_n(s)\}_{s\leq n})$. Let the probability that player i chooses a deliberate best response in round n conditioned on past events be given by $\rho_i(n) := \mathbb{P}(X_i(n) = 1|\mathcal{F}_{n-1})$, and assume that Assumptions 6.10–6.12 are satisfied. Note that $q_i(n)$, $p_i(n)$, $\xi_i(n)$ are $\mathcal{F}_n$–measurable and that by definition, $\rho_i(n)$ is $\mathcal{F}_{n-1}$–measurable.

Finally, let

$$g_i(n) := b_i(n-1)\rho_i(n) + \xi_i(n-1)(1-\rho_i(n)). \tag{6.21}$$

Note that $g_i(n)$ is $\mathcal{F}_{n-1}$–measurable and that $g(\alpha_i, n) = \mathbb{P}(a_i(n) = \alpha_i|\mathcal{F}_{n-1})$; that is, $g_i(n)$ represents the mixed strategy in use by player i in round n (compare with (6.17)). Let $g(n) := (g_1(n), \ldots, g_n(n))$ denote the joint mixed strategy in use at time n.

We refer to a process where, for each player i, $q_i(n)$ is updated according to (6.19), $a_i(n)$ is updated according to (6.20), and $g_i(n)$ is updated according to (6.21) as the strongly convergent variant of Ψ (for reasons to be clear soon—see Theorem 6.16).

The following theorem provides the general result from which the strong convergence of various MBR algorithms can be derived.

Theorem 6.16. *Let* Γ *be a finite normal form game, let* Ψ *be a myopic best-response algorithm. Assume Assumptions 6.10–6.15 hold. Then the strongly convergent variant of* Ψ *achieves strong learning in the sense that the period-by-period strategy tuple* $g(n)$ *converges to the internally chain recurrent set of the differential inclusion* (4.4).

Proof. By *Assumption*6.10 we have $\lim_{n\to\infty} \|g_i(n) - q_i(n)\| = 0$. Hence, to prove the theorem it is sufficient to show that

$$\lim_{n\to\infty} d(q(n), E) = 0. \tag{6.22}$$

The empirical distribution sequence $(q(n))_{n\geq 1}$ generated by the strongly convergent variant of a Ψ is updated according to the asynchronous variant of Ψ given in Section 6.3. By Theorem 6.2, in order to verify that (6.22) holds, it is sufficient to verify that the counting processes (6.18) satisfy Assumption 6.1.

By Lemma 6.21 (see appendix) and Assumption 6.11 we get that

$$\lim_{n\to\infty} \frac{N_i(n)}{\sum_{s=1}^n \rho_i(s)} = 1, \qquad \text{a.s.}, \tag{6.23}$$

which implies part (i) of Assumption 6.1 holds, almost surely. To show that part (ii) of Assumption

6.1 holds almost surely, note that by (6.23) and Assumption 6.12 we have

$$\lim_{n\to\infty}\frac{N_i(n)}{N_j(n)} = \lim_{n\to\infty}\frac{N_i(n)}{\sum_{s=1}^n \rho_i(s)}\frac{\sum_{s=1}^n \rho_i(s)}{\sum_{s=1}^n \rho_j(s)}\frac{\sum_{s=1}^n \rho_j(s)}{N_j(n)} = 1, \qquad \text{a.s.}$$

□

6.5.4 Example: ECFP

Assume that all players have an identical action space (Assumption 3.1) and Γ is a potential game with permutation invariant potential function (Assumption 3.2). In an ECFP process, the prediction function f_i is chosen for each player so that $a_i(n+1) \in \mathrm{BR}_i(\bar{q}_{-i}(n))$. It was shown in Section 4.5 that if Γ is a potential game, the internally chain recurrent set of the associated differential inclusion (3.11) is contained in the set of MCE. Hence, we get the following corollary to Theorem 6.2.

Corollary 6.17. *Let Γ be a potential game satisfying Assumptions 3.1 and 3.2. Suppose that Assumptions 6.10–6.13 hold. Then the strongly convergent variant of ECFP achieves strong learning in the sense that* $\lim_{n\to 0} d(g(n),\ MCE) = 0$.

Remark 6.18. *As noted in Remark 6.5, under the assumptions Assumption 3.1 and Assumption 3.2 if $\sigma \in MCE$ then using the permutation invariance of the potential function we get $\bar{\sigma} \in CNE$. An alternative strongly convergent variant of ECFP can be constructed by replacing $q_i(n)$ in* (6.20) *with the empirical centroid distribution $\bar{q}(n)$. In this case, players' period-by-period strategies converge strongly to the set of consensus NE; that is,* $\lim_{n\to\infty} d(g(n), CNE) = 0$.

As an example illustrating the importance of achieving string learning, we consider ECFP in a simple congestion game with parallel routes. Let $\mathcal{R} = \{1,\dots,R\}$ denote a set of routes available to all players and let $Y_i = R$, $\forall i$. For $r \in \mathcal{R}$, $y \in Y$, let $\rho_r(y)$ denote the number of players (or drivers) utilizing route r under the strategy y. For $r \in \mathcal{R}$ and $k \in \mathbb{N}$, let $c_r(k)$ denote the cost of using route r when k players are using the route r. Let the utility function for player i be given by $u_i(y) = -c_{y_i}(\rho_{y_i}(y))$.

There exist various mild assumptions under which no pure strategy consensus equilibria exist. For example, we assume:

Assumption 6.19. *$c_r(k)$ is strictly increasing in k, and* $\min_{\substack{r\in\mathcal{R},\\ r'\neq r}} c_r(n) - c_{r'}(1) > 0$.

ECFP can be an advantageous algorithm for learning equilibria in games with a large number of players. For example, the convergence rate of ECFP tends to be independent of the number of players (rather, it is more closely related to the manner in which players in ECFP are grouped—see [104]). Moreover, in ECFP the only piece of information that needs to be known among players is the centroid distribution—the memory size of which is invariant to the number of players.

We simulated classical ECFP and the strongly convergent variant of ECFP with parameter $\rho_i(n) = n^{-.5}$ in a congestion game satisfying Assumption 6.19 with 150 users and 7 routes. Figures 6.4a–6.4b and 6.4e–6.4f plot the average travel time in the respective processes. Note that this is a measure of "global welfare" and not a direct measure of NE convergence, however the trend is consistent with convergence to NE.

Note that the classical ECFP process leads players to often choose the same action as all others in stages of the repeated play, thus leading to extremely poor period-by-period utilities, and poor global welfare. On the other hand, in the strongly convergent variant of ECFP, players period-by-period strategies approach NE leading to period-by-period utilities and global welfare that are consistent with the learned NE strategy.

6.6 Appendix to Chapter 6

Lemma 6.20. *Let $i, j \in \mathcal{N}$, let $\tau_i(s)$ and $\tilde{q}_j(s)$ be defined as in Section 6.3, and assume Assumption 6.1 holds. Then* $\lim_{s\to\infty} \|q_j(\tau_i(s)) - \tilde{q}_j(s)\| = 0$.

Proof. Note that by the definitions of τ_j, N_j, and $\tilde{q}_j$ there holds $q_j(n) = q_j(\tau_j(N_j(n))) = \tilde{q}_j(N_j(n))$, for any $n \in \mathbb{N}$ Noting that $\sqrt{2} = \max_{p',p''\in\Delta(Y_j)} \|p'-p''\|$, we also have $\|\tilde{q}_j(s+1)-\tilde{q}_j(s)\| \leq \frac{\sqrt{2}}{s}$, for $s \in \mathbb{N}$, and more generally, for $s_1, s_2 \in \mathbb{N}$, we have $\|\tilde{q}_j(s_1)-\tilde{q}_j(s_2)\| \leq \sum_{s=\min(s_1,s2)}^{\max(s_1,s_2)-1} \|\tilde{q}_j(s+1)-\tilde{q}_j(s)\| \leq \frac{|s_2-s_1|}{\min(s_1,s_2)}\sqrt{2}$. Hence, $|q_j(\tau_i(s))-\tilde{q}_j(s)\| = \|\tilde{q}_j(N_j(\tau_i(s)))-\tilde{q}_j(s)\| = \|\tilde{q}_j(N_j(\tau_i(s)))-\tilde{q}_j(N_i(\tau_i(s)))\| \leq \frac{|N_j(\tau_i(s))-N_i(\tau_i(s))|}{\min(N_i(\tau_i(s)),N_j(\tau_i(s)))}\sqrt{2}$, where the second equality follows from the fact that $N_i(\tau_i(s)) = s$ (see (6.4)). Thus, it suffices to show that $\lim\limits_{s\to\infty} \frac{|N_j(\tau_i(s))-N_i(\tau_i(s))|}{\min(N_i(\tau_i(s)),N_j(\tau_i(s)))} = 0$. But, by Assumption 6.1, for any i, j there holds: $0 = \lim_{n\to\infty} \frac{N_i(n)}{N_j(n)} - 1 = \lim_{s\to\infty} \frac{N_i(\tau_i(s))}{N_j(\tau_i(s))} - 1 = \lim_{s\to\infty} \frac{N_i(\tau_i(s))-N_j(\tau_i(s))}{N_j(\tau_i(s))}$, where the second equality follows from the fact that (again by Assumption 6.1) $\lim_{s\to\infty} \tau_i(s) = \infty$. □

Lemma 6.21. *Let $(X(n))_{n\geq 1}$ be $0-1$ Bernoulli random variables, let $\ell(n) := \sum_{k=1}^n X(k)$ be the associated counting process, let $\mathcal{G}_n := \sigma(\{X(k)\}_{k=1}^n)$, and let $\rho(n) = \mathbb{P}(X(n) = 1|\mathcal{G}_{n-1})$. Assume $\sum_{n\geq 1} \rho(n) = \infty$. Then there holds,* $\lim\limits_{n\to\infty} (\ell(n)) / (\sum_{k=1}^n \rho(n)) = 1, \quad$ *a.s.*

Proof. The result follows from Levi's extension of the Borel-Cantelli Lemmas, [122] p.124. □

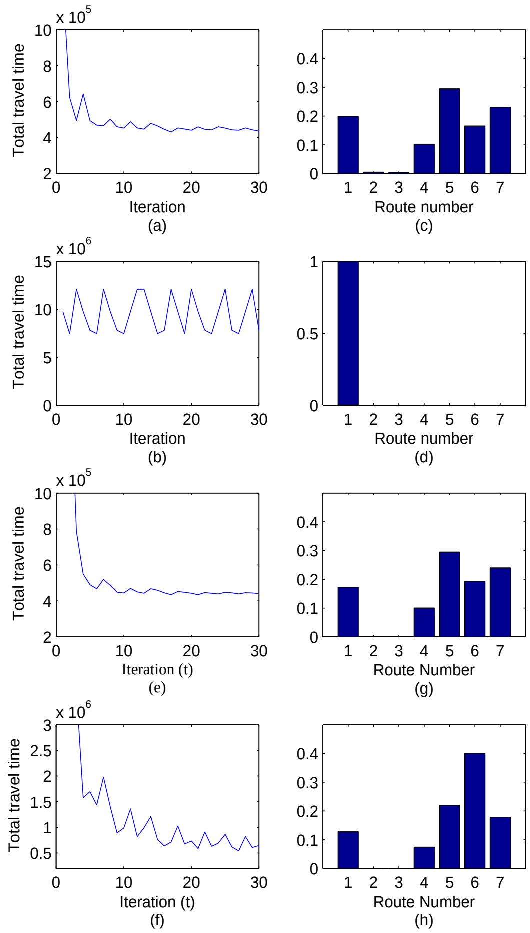


Figure 6.4: 3a: average travel time under $\bar{q}^n(n)$ in ECFP, 3b: average utility under period-by-period mixed strategy $g(n)$ in ECFP, 3c: $\bar{q}(n)$ in ECFP at 50 iterations, 3d: $g_1(n)$ in ECFP at 50 iterations, 3e: average travel time under $\bar{q}^n(n)$ in strong-ECFP, 3f: average travel time under period-by-period mixed strategy $g(n)$ in strong-ECFP, 3g: $\bar{q}(n)$ in strong-ECFP at 50 iterations, 3h: $g(n)$ in strong-ECFP at 50 iterations.

Chapter 7

Network-Based Implementation

7.1 Introduction

In the formulation of MBR dynamics in Section 2.4, it is implicitly assumed that each agent has instantaneous access to all information required to compute her next-stage action. More precisely, in order for an agent to choose her action at stage $n+1$, the agent is assumed to have complete knowledge of the stage-n observation state, $z(n)$.

This assumption can be impractical in large-scale settings where physical limitations may hinder agents' ability to directly observe the actions of others or to directly communicate one with another. For example, in classical FP, players are assumed to have precise knowledge of the stage-n empirical distribution, $q(n)$ (see (2.8)). Note that writing (2.8) recursively, for each $j \in \mathcal{N}$ we have $q_j(n) = (1-\frac{1}{n})q_j(n-1) + \frac{1}{n}a_j(n)$, where $a_j(n)$ is the action taken by player j in stage n. Hence, in order for some player i to compute $q(n)$ given $q(n-1)$ (which in turn allows the player to compute best-response set $BR_i(q(n))$ and choose a stage-$(n+1)$ action), the player must know the stage-n action $a_j(n)$ of all other players $j \in \mathcal{N},\ j \neq i$. In large scale distributed settings it may be difficult for every player to communicate this information to every other player between each iteration of the repeated play.

A more practical assumption in many large-scale settings can be to suppose that players are equipped with an overlaid communication graph through which they may periodically exchange information relevant to the learning process. We refer to this as a *network-based* learning setting.

Consider, for example a traffic routing scenario in which a group of vehicles wish to navigate a traffic grid while minimizing their respective travel times. Suppose that the vehicles are connected via an ad-hoc vehicular communication network. Before physically engaging in the commute, suppose that players wish to compute a Nash equilibrium routing strategy by communicating over the ad-hoc network. In this setting there is no physical feedback from game play—any information related to a repeated play learning process must be disseminated through the communication network.

The problem of computing NE in a network-based setting has been a subject of recent research interest. For example, [81] studies a network-based algorithm for NE seeking in a two-network zero-sum games, [80] studies an algorithm for finding NE in a spatial spectrum access games, [82] studies a network-based regret-based reinforcement learning algorithm for tracking the polytope of correlated equilibria in time-varying games. The work [63] presents a method for designing games with a local a prescribed local dependence. The works [79] and [83] study gossip-based algorithms for computing NE in a network-based setting in games with continuous-action spaces.

In this chapter (see also [78, 104, 111]) we show that MBR algorithms can easily be adapted for deployment in a network-based setting. We present a generic network-based MBR algorithm and prove that as long as the information dissemination scheme satisfies a mild condition (see Condition 7.1), the network-based algorithm will converge to the internally chain recurrent set of the associated differential inclusion. Our results differ from those above primarily in that, rather than focus on a particular learning algorithm and information dissemination scheme, we study the general class of MBR algorithms and show that, so long as the information dissemination scheme satisfies Condition 7.1, then any such algorithm will converge to the associated internally chain recurrent set. The underlying MBR dynamics can then be chosen to meet the system designers needs. For example, ECFP dynamics (Chapter 3) may be used in order to reduce information overhead, SS-MBR dynamics (Chapter 5) may be used in order to reduce complexity, and asynchronous MBR dynamics (Chapter 6) may be used if a global clock is not available.[1]

After introducing the general model of a network-based MBR algorithm, we study the network-based implementation of two example MBR algorithms. First, we consider network-based implementation of the ECFP algorithm introduced in Chapter 3. ECFP mitigates the problems of high information overhead, high computational complexity, and slow convergence rates discussed in Sections 1.3.2–1.3.1. We present a network-based implementation of ECFP and prove that the algorithm converges to the set of consensus NE. In particular, we explicitly specify an information dissemination scheme and prove that it satisfies the necessary conditions for convergence.

Subsequently, we consider network-based implementation of a SS-MBR algorithm. Chapter 5 introduced SS-MBR algorithms, which use a stochastic approximation technique to mitigate computational complexity. SS-FP, discussed in Section 5.5, is the single sample variant of classical FP. This variant of FP reduces complexity from $O(e^N)$ in classical FP (where N is the number of

[1]Note that any combination of these may also be used.

players), to $O(N)$ in SS-FP. In this chapter we present a network-based variant of SS-FP and prove that the algorithm converges to the set of NE in the same sense as classical FP. As in the case of network-based ECFP, we explicitly specify an appropriate information dissemination scheme, and prove that the it satisfies the necessary conditions for convergence.

The remainder of the chapter is organized as follows. In Section 7.2 we present the general model of a network-based myopic best-response algorithm and discuss a sufficient condition for convergence to the internally chain recurrent set of the associated differential inclusion. In Section 7.3 we study the ECFP algorithm in a network-based setting. In Section 7.4 we study the SS-FP algorithm in a network-based setting.

7.2 Network-Based Myopic Best-Response Dynamics

Algorithm Setup

In a network-based MBR algorithm, players do not have perfect knowledge of the observation state $z(n)$. Let $\hat{z}^i(n)$ be an estimate that player i maintains of the observation state $z(n)$.

A prototypical network-based MBR algorithm is given below.

7.2.1 Network-Based MBR Algorithm

Initialize

(i) Initialize the state estimate $\hat{z}^i(1)$ for each $i \in \mathcal{N}$.[2] Let players choose an arbitrary initial action.

Iterate $(n \geq 1)$

(ii) Each agent $i \in \mathcal{N}$ chooses a next-stage action according to the rule

$$a_i(n+1) \in \mathrm{BR}_{i,\epsilon_n}(f_i(\hat{z}^i(n))$$

where $\epsilon_n \to 0$ as $n \to \infty$.

(iii) Each agent $i \in \mathcal{N}$ may engage in one round of information exchange with neighboring agents (as defined by G) and update their estimate $\hat{z}^i(n+1)$ using the information obtained.[3]

[2]The initialization of $\hat{z}^i(n)$ may be subject to some conditions depending on the particular information dissemination scheme used, e.g., [21, 104]. See discussion below for more details.

[3]Note that the "true" observation state at time $(n+1)$ is given by $z(n+1) = z(n) + \gamma(n)\left(g(a_i(n+1)) - z(n)\right)$. However, it is not assumed that players have knowledge of $(z(n))_{n\geq 1}$.

7.2.2 Discussion

The protocol used to form the estimate $\hat{z}^i(n)$ in step (iv) is intentionally crafted to be broad in order to emphasize that a wide variety of information dissemination protocols may be used. We only require that the information dissemination scheme satisfy the following condition:

Condition 7.1. *For all $i \in \mathcal{N}$, there holds $\|\hat{z}^i(n) - z(n)\| \to 0$ as $n \to \infty$.*

The following Theorem gives the general convergence result for network-based MBR algorithms. It follows readily from the robustness result of Theorem 4.15.

Theorem 7.2. *Let $(z(n))_{n\geq 1}$ be a network-based MBR algorithm satisfying Assumptions 2.1–2.3 and Condition 7.1. Then $(z(n))_{n\geq 1}$ converges to the internally chain recurrent set of the associated differential inclusion,* (4.4).

Proof. Using the Lipschitz continuity of $U_i(\cdot)$, Condition 7.1 implies that there exists a sequence $(\eta_n)_{n\geq 1}$, $\eta_n \to 0$ such that $a(n+1) \in \mathrm{BR}_{\eta_n}(z(n))$; i.e., the network-based MBR process fits the template of a weakened MBR process. The result then follows by Theorem 4.15. □

In the following two sections we develop two example implementations of network-based MBR algorithms. We explicitly specify information dissemination schemes and prove that the algorithms converge in the desired sense.

We note that, as time progresses, the observation state may require an infinite number of bits to transmit precisely. For example, in FP players track the empirical distribution. The empirical distribution is composed of rational numbers, which, as the algorithm progresses, require an unbounded number of bits to specify precisely. Realistically, players will only be able to track the observation state up to some finite precision.

In order to apply our convergence results, we require that Condition 7.1 hold. The problem of precise computing in network-based settings with quantized information has been considered in prior work [128]. In order to guarantee Condition 7.1 holds in an environment where players track the observation state with finite precision, it may be possible to use dithering techniques, similar to those used in [128].

For the special case of FP in non-degenerate potential games (Chapters 9–10), we note that Theorem 10.1 implies that FP converges to pure-strategy equilibria, and the first-order non-degeneracy condition implies that pure-strategy equilibria are robust to strategy perturbations. Hence, when running network-based FP in a non-degenerate potential game, if the precision with which agents track $q(n)$ is sufficient to get within a sufficiently small neighborhood of some pure-strategy equilibrium q^*, then the convergence of FP to q^* will be robust to strategy perturbations occurring from imprecise specification of $q(n)$.

7.3 Network-Based ECFP

In this section we present a network-based implementation of the ECFP algorithm presented in Chapter 3. We note that, for simplicity, we focus on the basic formulation of ECFP. We prove the network-based algorithm converges to the set of CNE and present simulation results. We begin in the next section by introducing some definitions that facilitate the presentation of the algorithm.

7.3.1 Algorithm Setup

Define the following two matrices:

$$\mathbf{Q}(n) := (q_1(n)\ q_2(n)\ \ldots\ q_N(n))^T \in \mathbb{R}^{N\times m},$$

$$\hat{\mathbf{Q}}(n) := (\hat{q}_1(n)\ \hat{q}_2(n)\ \ldots\ \hat{q}_N(n))^T \in \mathbb{R}^{N\times m},$$

where $\hat{q}_i(n) \in \mathbb{R}^m$ denotes player i's estimate of $\bar{q}(n) \in \mathbb{R}^m$ (see (3.1)). Let $\hat{q}(n) \in \mathbb{R}^N$ be the N-tuple $(\hat{q}_1(n), \ldots, \hat{q}_N(n))$. The tuple $\hat{q}(n)$ will be important in network-based ECFP; in particular we will prove that $\hat{q}(n)$ converges to the set of consensus equilibria.

Let $\mathbf{W} \in \mathbb{R}^{N\times N}$ be a weighting matrix with entries $w_{i,j}$. We assume $\mathbf{W}$ satisfies the following assumption:

Assumption 7.3. *The weight matrix* $\mathbf{W}$ *is an* $N \times N$ *matrix that is doubly stochastic, aperiodic, and irreducible, with sparsity conforming to the communication graph* G.

Note that given assumption Assumption 1.1 (G is a connected graph), it is always possible to find a matrix $\mathbf{W}$ satisfying these conditions (see [21, 22, 129]).

7.3.2 Network-Based ECFP Algorithm

Initialize

(i) Each player i chooses an arbitrary initial action $a_i(1)$. The initial empirical distribution for player i is given by $q_i(1) = a_i(1)$. Each player $i \in \mathcal{N}$ initializes her local estimate of the empirical distribution as

$$\hat{q}_i(1) = \sum_{j\in\Omega_i\cup\{i\}} w_{ij} q_j(1) \tag{7.1}$$

where Ω_i is the set of neighbors of player i and the w_{ij}'s are constant neighborhood weighting factors.

Iterate $(n \geq 1)$

(ii) Each player $i \in \mathcal{N}$ computes the set of best responses using $\hat{q}_i(n)$ as the assumed mixed strategy for each of the other $(N-1)$ players. The next action

$$a_i(n+1) \in \arg\max_{\alpha_i \in A_i} U(\alpha_i, \hat{q}_{-i}(n)) \tag{7.2}$$

is played according to the best-response calculation. In the event of multiple pure strategy best responses, any of the maximizing actions in (7.2) may be chosen arbitrarily. The local empirical distribution $q_i(n+1)$ is updated to reflect the action taken, i.e.,

$$q_i(n+1) = q_i(n) + \frac{1}{n+1}(a_i(n+1) - q_i(n)).$$

(iii) Subsequently each player i computes a new estimate of the network-averaged empirical distribution using the following update rule:[4]

$$\hat{q}_i(n+1) = \sum_{j \in \Omega_i \cup \{i\}} w_{i,j} \left(\hat{q}_j(n) + q_j(n+1) - q_j(n)\right), \tag{7.3}$$

where Ω_i is the set of neighbors of player i, and $w_{i,j}$ is a weighting constant.[5]
The update in (7.3) is represented in more compact notation as

$$\hat{\mathbf{Q}}(n+1) = \mathbf{W}\left(\hat{\mathbf{Q}}(n) + \mathbf{Q}(n+1) - \mathbf{Q}(n)\right), \tag{7.4}$$

where $\mathbf{W} \in \mathbb{R}^{N \times N}$ satisfies Assumption 7.3

7.3.3 Network-Based ECFP: Main Convergence Result

We will refer the the sequence $(q(n))_{n\geq 1}$ generated by the foregoing algorithm as a *network-based ECFP process*. In a network-based ECFP process, players learn a consensus equilibrium strategy

[4]Note that (7.3) is equivalent to $\hat{q}_i(n+1) = \sum_j w_{i,j}(\hat{q}_j(n) + 1/(n+1)(a_i(n+1) - q_i(n)))$. This is closely related to minimizing an aggregate cost function using a distributed stochastic gradient descent method [130–133]: $J^{glob}(q) = \sum_{i=1}^{N} E\|a_i(n) - q\|^2$ whose minimizer is the average of actions over time and over players, and where the expectation is over the empirical distribution of $a_i(n)$ over time. The key thing to note is that the exact minimizer of this cost is time-varying and, given that we are operating in a distributed environment with only one round of communication allowed per time slot, we can only track this dynamic minimizer using an iterative method as given in (7.3).

[5]Note that the set $\Omega_i \cup \{i\}$ in the summation indicates that player i uses her own (local) information and that of her neighbors to update her estimate. The update rule is clearly distributed as information exchange is restricted to neighboring players only.

in a setting where information exchange is restricted to a local neighborhood of each agent. The result is summarized in the following Theorem.

Theorem 7.4. *Let $(q(n))_{n\geq 1}$ be a network-based ECFP process such that Assumptions 1.1–1.3, 3.1–3.2, and 7.3 hold. Then $d(\hat{q}(n), CNE) \to 0$ as $n \to \infty$. In particular, the agents' estimates $\hat{q}_i(n)$ reach asymptotic consensus, i.e. $d\left(\hat{q}_i(n), \hat{q}_j(n)\right) \to 0$ as $n \to \infty$ for each pair (i, j) of agents.*

In the above result, the N-tuple $\hat{q}_i(n) = (\hat{q}_1(n), \ldots, \hat{q}_N(n))$ converges to the set CNE; since $\hat{q}_i(n)$ is available to player i, player i learns the component of the consensus equilibrium strategy relevant to him.

Proof. By Lemma 7.9 in the appendix, the error in a network-based ECFP process satisfies $\|\hat{q}_i(n) - \bar{q}(n)\| \to 0$ as $n \to \infty$, thus Condition 7.1 is satisfied and the network-based ECFP fits the template of Theorem 7.2. In Lemma 4.19 it was shown that every internally chain recurrent set of the ECFP differential inclusion is contained in the set of SNE. In this setup, the set of CNE coincides with the set of SNE. Applying Theorem 7.2 we get that $d(\bar{q}^N(n), CNE) \to 0$ as $n \to \infty$. By Lemma 7.9 we obtain, $\|\hat{q}_i(n) - \bar{q}(n)\| \to 0$ as $n \to 0$, and the result $d(\hat{q}(n), CNE) \to 0$ as $n \to \infty$ follows. □

Again, we emphasize that this mode of convergence is not the same as the more traditional convergence in empirical frequency (see (2.13) and accompanying discussion).

7.3.4 Simulation Example

Cognitive Radio Setup

In this section we illustrate the operation of network-based ECFP by implementing it in a cognitive radio application.

Let Ch indicate a finite collection of permissible frequency channels. Assume there are two classes of users sharing the allocated set of channels: primary users and secondary users. Assume each primary user has been assigned to a fixed channel from which they may not deviate. Secondary users are free to use any channel they wish. The objective in this setup is for the secondary users to cooperatively learn a channel allocation that is both fair and in some sense optimal.

Cast this setup in the format of a normal form game $\Gamma = (N, (Y_i, u_i(\cdot))_{i\in\mathcal{N}})$ with $\mathcal{N}$ being the set of secondary users, and $Y_i = Ch$ for all i. Let $\psi_r(y)$ (respectively, $\psi_r(y_{-i})$) denote the number of users on channel $r \in Ch$ for the joint strategy $y \in Y^N$ ($y_{-i} \in Y_{-i}$).

The cost associated with channel r when k users are on channel r is given by $c_r(k)$. The utility for player i is given by $u_i(y) = -c_{y_i}(\psi_{y_i}(y))$, and the mixed utility is given by the usual multilinear extension. The game Γ is an instance of a congestion game—a known subset of potential games [38].

Communication Graph Setup

We assume that some small portion of spectrum is allocated for the purpose of transmitting data pertinent to the learning algorithm (i.e., disseminating information about the empirical centroid $\bar{q}(n)$.) Such an assumption is reasonable when the communication overhead associated with the learning algorithm is relatively small compared to the objective data being transmitted, e.g., users objective is to transmit large video files.

We model user-to-user communication using a geometric random graph. In implementing the network-based ECFP algorithm, we assign the weight constants $w_{i,j}$ of (7.3) according to the Metropolis-Hastings rule [134].

Simulation Results

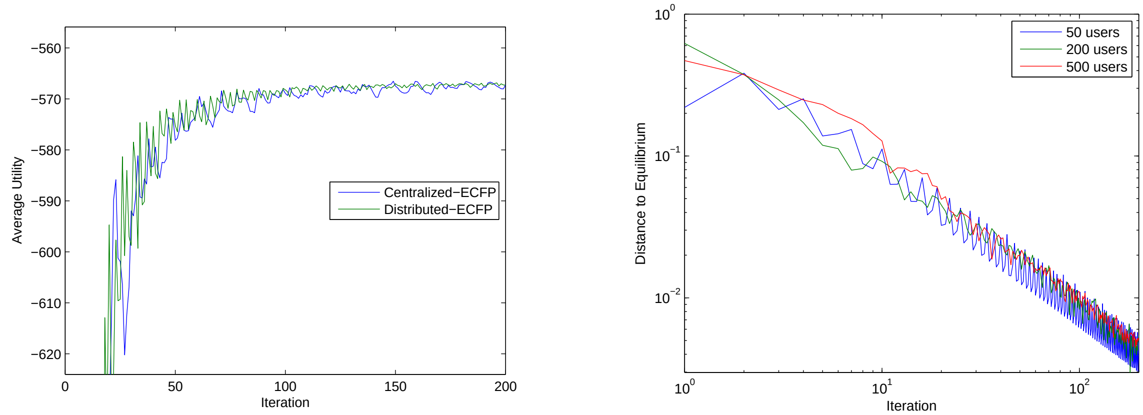
We simulated ECFP in two different cognitive radio scenarios. In the first, there are 10 channels and 400 users, and the cost function for channel $r \in Ch$ is given by a cubic polynomial of the form

$$c_r(k) = a_3 k^3 + a_2 k^2 + a_1 k + a_0,$$

where k is the number of users on channel r and a_j, $0 \leq j \leq 3$ are arbitrary coefficients. Figure 5.1 shows a plot of the utilities $U(\bar{q}^N(n))$ and $U((\hat{q}_1(n))^N)$ in the fully-connected and network-based cases respectively, where $\bar{q}^N(n)$ is as defined in (3.3).[6] The choice of the distribution of player 1, $\hat{q}_1(n)$, to represent the network-based case was arbitrary—$\hat{q}_i(n)$ for any $i \in N$ produces a similar result. In the network-based case, players communicated via a randomly generated geometric graph with average node degree of 8.78.

Both the fully-connected and network-based algorithms were started with identical initial conditions. It is interesting to note that, while multiple NE do exist, both algorithms tend to converge to the same equilibrium, regardless of the communication graph topology. This trend suggests that the basin of attraction for any given NE is similar for both fully-connected ECFP and network-based ECFP. Neither algorithm is noticeably superior in terms of the quality of equilibria attained.

[6] The notation $(\hat{q}_i(n))^N$ signifies the N-tuple containing N repeated copies of $\hat{q}_i(n)$; cf. the definition of $\bar{q}^N(n)$ in (3.3).



(a)

(b)

Figure 7.1: 7.1a: The average utility (taken over the set of players) of the joint empirical distribution, $\bar{q}^N(n)$, 7.1b: The distance from the joint empirical distribution to the set of NE.

A useful feature of consensus NE (CNE) in this setup is their adaptability to players entering or exiting the game. If a new player enters the game after an ECFP learning process has been running for some time, then incumbent players can simply inform the new player of the current empirical distribution $\bar{q}(n)$, and the distribution $\bar{q}^{N+1}(n)$ (meaning the $(N+1)$ tuple which contains repeated copies of $\bar{q}(n)$) will be an approximate CNE in the newly formed $(N+1)$ player game. Similarly, if a player exits the game, the distribution $\bar{q}^{N-1}(n)$ will be an approximate CNE in the newly formed $(N-1)$ player game.

In the second cognitive radio scenario simulated, there are 10 channels, each with a quadratic cost function. This choice of cost functions guarantees the existence of a unique CNE. We simulated network-based ECFP in this scenario for the cases of 50, 200, and 500 users; each case had different randomly generated cost functions. In each case, the communication graph was generated as a random geometric graph. The average node degree for the associated communication graph in each case was 8.04, 8.72, and 8.98 respectively.

Figure 7.1b shows a plot of the normalized distance of $(\hat{q}_1(n))^N$ to the unique NE in each case (the particular choice of $\hat{q}_1(n)$ again being arbitrary). Distance was measured using the Euclidean norm, normalized by $\sqrt{N}$, where N is the number of players. Simulation results suggest that the convergence rate of ECFP is independent of the number of players. Indeed, the analytical properties of ECFP (in general, see Section VII) suggest that the convergence is dependent only on the number of permutation invariant classes into which the player set is partitioned and not the overall number of players. A rigorous characterization of the precise nature of this relationship may be an interesting topic for future research.

7.4 Network-Based Single-Sample FP

In this section we present a network-based implementation of the SS-FP algorithm given in Section 5.5. We prove that the network-based SS-FP algorithm converges to the set of NE (almost surely) and present a simulation example. We begin in the next section by introducing some definitions that facilitate the presentation of the algorithm.

7.4.1 Algorithm Setup

Let $\hat{q}_j^i(n)$ denote an estimate that player $i \in \mathcal{N}$ maintains of the empirical distribution of player $j \in \mathcal{N}$, and let $\hat{q}^i(n) := (\hat{q}_1^i(n), \ldots, \hat{q}_N^i(n)) \in \mathbb{R}^{\sum_{i\in\mathcal{N}} |Y_i|}$ denote an estimate that player i maintains of the tuple of empirical distributions. In network-based SS-FP, each player $i \in \mathcal{N}$ will form their empirical-distribution estimate $\hat{q}^i(n)$ by exchanging information with neighboring players (see steps (i) and (iv) of the network-based SS-FP algorithm below). Let $\mathbf{W} = (w_{ij})_{i,j=1}^N$ be a weighting matrix to be used in the distributed computation of $\hat{q}^i(n)$. We will assume that $\mathbf{W}$ satisfies Assumption 7.3.

Similar to the (centralized) SS-FP algorithm presented in Section 5.5, in network-based SS-FP, each player $i \in \mathcal{N}$ forms an estimate of the utility $U_i(\cdot, q_{-i}(n))$, where $q_{-i}(n) = (q_j(n))_{j\in\mathcal{N}\backslash\{i\}}$ and $q_j(n)$ is as defined in (2.8). Let $\hat{U}_i(\alpha_i, n) \in \mathbb{R}$ denote the estimate player i maintains of $U_i(\alpha_i, q_{-i}(n))$ for each $\alpha_i \in A_i$, $n \in \{0, 1, \ldots\}$.

Some Additional Definitions

The following notation is used to facilitate a compact description of the algorithm. Let $s := \sum_{k\in N} m_k$. Let $q_i'(n)$ be an augmented (zero-stuffed) vector representing the empirical distribution of player i such that

$$q_i'(n) := (0, \ldots 0, Nq_i(n), 0 \ldots, 0) \in \mathbb{R}^s.$$

The augmented vector $q_i'(n)$ matches the general structure of $\hat{q}^i(n)$, but in the place of $q_i^i(n)$ we substitute in $N \cdot q_i(n)$ (a scaled copy of the true empirical distribution) and set all other entries to zero.

For $i \in N$, $\xi = (\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_N)$, $\xi_j \in \mathbb{R}^{m_j}$ let

$$P_{\Delta_{-i}}(\xi) := (P_{\Delta_1}(\xi_1), \ldots, P_{\Delta_N}(\xi_N)),$$

where $P_{\Delta_i}(\xi_i) := \arg\inf_{\sigma_i \in \Delta_i} \|\sigma_i - \xi_i\|$ is the projection of ξ_i onto the set Δ_i.

7.4.2 Network-Based SS-FP Algorithm

Initialize

(i) Each player i takes an arbitrary initial action $a_i(1)$, and the empirical distribution for player i is initialized as $q_i(1) = a_i(1)$. Each player $i \in \mathcal{N}$ initializes their local estimate of the empirical distribution as

$$\hat{q}^i(1) = \sum_{j \in \Omega_i \cup \{i\}} w_{ij} q'_j(1), \tag{7.5}$$

where Ω_i is the set of neighbors of player i. The utility estimate is initialized as $\hat{U}_i(\alpha_i, 0) = 0$ for all $\alpha_i \in A_i$.

Iterate $(n \geq 1)$

(ii) Each player $i \in \mathcal{N}$ draws an action $a^*_{-i}(n)$ as a random sample from the probability mass function[7] $P_{\Delta_{-i}}\left(\hat{q}^i_{-i}(n)\right)$. For each $\alpha_i \in A_i$ player i updates their estimate of the predicted utility according to the rule:

$$\hat{U}_i(\alpha_i, n) = (1 - \rho(n))\hat{U}_i(\alpha_i, n-1) + \rho(n) U_i(\alpha_i, a^*_{-i}(n)).$$

(iii) Each player $i \in \mathcal{N}$ chooses their stage-$(n+1)$ action as a best response to the predicted utility $\hat{U}(\cdot, n)$; i.e.,

$$a_i(n+1) \in \arg\max_{\alpha_i \in A_i} \hat{U}(\alpha_i, n). \tag{7.6}$$

The local empirical distribution of player i is recursively updated to include the action just taken, i.e.,

$$q_i(n+1) = q_i(n) + \frac{1}{n+1}\left(a_i(n+1) - q_i(n)\right).$$

(iv) Each player $i \in \mathcal{N}$ updates their estimate of the joint empirical distribution using the following

[7]Given the distributed update rule used to compute $\hat{q}^i(n)$ in step iv, it is possible that $\hat{q}^i(n)$ may sometimes leave the probability simplex. The projection operation guarantees that a player has a valid probability distribution from which to draw the sample $a^*_{-i}(n)$.

rule:

$$\hat{q}^i(n+1) = \sum_{j \in \Omega_i \cup \{i\}} w_{ij} \left(\hat{q}^j(n) + q_j'(n+1) - q_j'(n) \right). \tag{7.7}$$

7.4.3 Network-Based SS-FP: Main Convergence result

The following result shows that players engaged in a network-based SS-FP process asymptotically learn a Nash equilibrium strategy.

Theorem 7.5. *Let* Γ *be a potential game. Let* $(q(n))_{n\geq 1}$*, where* $q(n) := (q_1(n), \dots, q_N(n))$ *be computed according to the network-based SS-FP algorithm of Section 7.4.2, and suppose that Assumptions 1.1–1.3, 5.3–5.4, and 7.3 hold. Then players learn a Nash equilibrium in the sense that* $\lim_{n\to\infty} d(q(n), NE) = 0$*. Furthermore, each player* $i \in \mathcal{N}$ *achieves asymptotic strategy learning in the sense that* $\lim_{n\to\infty} \|\hat{q}^i_j(n) - q_j(n)\| = 0$ *for all* $j \in \mathcal{N}$*.*

Proof. We will prove the result by showing that there exists a sequence $\{\epsilon_n\}_{n\geq 1}$ such that $\epsilon_n \to 0$ and $U_i(a_i(n+1), q_{-i}(n)) \geq \max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(n)) - \epsilon_n$ for all i. By Corollary 4.17, this is sufficient to ensure $d(q(n), NE) \to 0$ in potential games.

By Lemma 7.10, there holds $\lim_{n\to\infty} \|\hat{q}^i(n) - q(n)\| = 0$. Since the map $P_{\Delta_{-i}}$ is Lipschitz continuous, this implies $\lim_{n\to\infty} \|P_{\Delta_{-i}}\left(\hat{q}^i_{-i}(n)\right) - q_{-i}(n)\| = 0$. By Lipschitz continuity of U_i, this implies that

$$\lim_{n\to\infty} |U_i(\alpha_i, P_{\Delta_{-i}}\left(\hat{q}^i_{-i}(n)\right)) - U_i(\alpha_i, q_{-i}(n))| = 0, \ \forall \alpha_i, \ \forall i. \tag{7.8}$$

Using assumption **A.1** in Lemma 5.6 we get

$$\lim_{n\to\infty} |\hat{U}_i(\alpha_i, n) - U_i(\alpha_i, P_{\Delta_{-i}}\left(\hat{q}^i_{-i}(n)\right))| = 0, \ \forall \alpha_i, \ \forall i.$$

Combining this with (7.8) gives,

$$\lim_{n\to\infty} |\hat{U}_i(\alpha_i, n) - U_i(\alpha_i, q_{-i}(n))| = 0, \ \forall \alpha_i, \ \forall i.$$

Combining this with the network-based SS-FP action rule (7.6), we see that there exists a sequence $\{\epsilon_n\}_{n\geq 1}$ such that $\epsilon_n \to 0$ and

$$U_i(a_i(n+1), q_{-i}(n)) \geq \max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(n)) - \epsilon_n, \ \forall i,$$

and the desired result holds. $\square$

7.4.4 Simulation Results

We simulated network-based SS-FP in a simple cognitive radio example. Let $\mathcal{N} = \{1, \dots N\}$ indicate a set of users (or players), and let Ch indicate a finite collection of permissible frequency channels shared by all users (i.e., $Y_i = Ch$, $\forall i \in N$). For $y \in Y$, and $k \in Ch$, let $\psi_k(y)$ denote the number of users on channel k given the (joint) action tuple y. Further, for $k \in Ch$ and $\ell \in \{0, 1, 2, \dots\}$, let $c_k(\ell)$ denote the cost of using channel k when there are ℓ users occupying the channel. Let the utility for player $i \in \mathcal{N}$ be given by $u_i(y) = -c_{y_i}(\psi(y))$. This game is classified as a congestion game—a known type of potential game [38].

We simulated the network-based SS-FP algorithm presented in Section 7.4.2 in the cognitive radio game with 40 users and 10 channels. The stochastic approximation weight sequence was given by $\rho(n) = n^{-.6}$, $n \geq 1$, and the step-size sequence was given by $\gamma(n) = n^{.65}$, $n \geq 1$. The weight matrix W used to estimate $q(n)$ (see (7.7)) was derived using the Metropolis-Hastings rule [134].

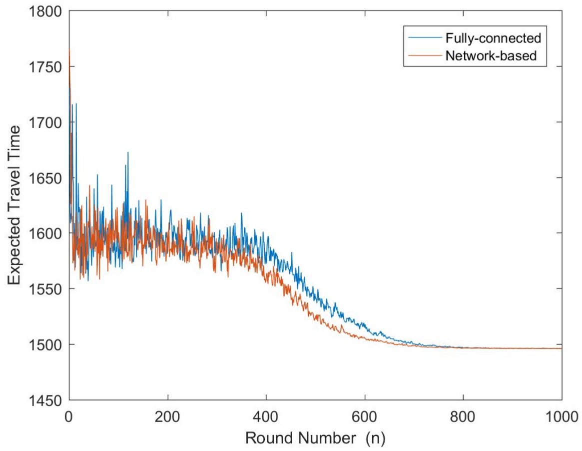
Figure 7.2a shows a plot of the expected utility of player[8] 1 averaged over 5 random instantiations in a fully-connected version of SSFP and a network-based version of SS-FP. In the network-based version of SS-FP, the communication graph, shown in figure 7.2b, was generated as a random geometric graph. Network-based SS-FP and fully-connected SS-FP tended to converge at similar rates.

We note that the fully connected SS-FP algorithm in this section tends to converge more quickly than the "centralized" SS-FP algorithm studied in Section 5.5.4. This is due to two factors. First, by choosing the step-size sequence of the form $\gamma(n) = (n+1)^{-r}$, $r < 1$ we were able to improve the rate of convergence as compared to a step-size sequence of the form of $\gamma(n) = (n + 1)^{-1}$. Second, in Section 5.5.4, in each iteration of the algorithm we drew a single test action that was used to update the utility estimates of all players. In the fully-connected version of SS-FP simulated here, we assume that each player draws a sample test action (independent from the other players) at each iteration. This parallelizes the process of drawing samples, which tends to accelerate the *per-round* convergence rate of the algorithm. However, this does not necessarily improve the *per-sample* convergence rate of the algorithm.

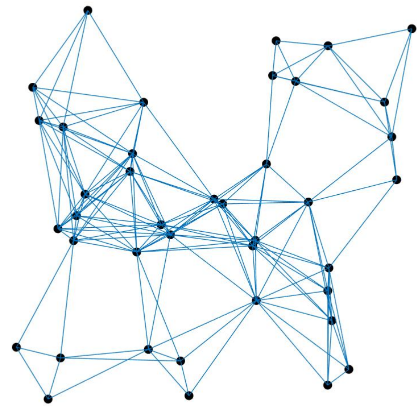
7.5 Appendix to Chapter 7

This appendix is concerned with topics in distributed consensus in networks where node values are dynamic quantities. The results of this section are used to prove convergence of the network-based algorithms presented in section 7.3.2. Results in this section are similar to results on distributed

[8] The utility trend for all players was similar.



(a)



(b)

Figure 7.2: (a) Utility of $q(n)$ in fully connected SS-FP and network-based SS-FP, (b) Communication graph used in network-based SS-FP.

averaging in networks with additive changes in node values and information dynamics in [130, 135, 136]. For a survey of traditional consensus and gossip algorithms, the reader may refer to [21, 22, 137].

Consider a network of N nodes connected through a communication graph $G = (V, E)$. The graph is assumed to be connected. Let $x_i(n) \in \mathbb{R}$ be the value of node i at time n, and let $x(n) \in \mathbb{R}^N$ be the vector of values at all nodes. The goal is for each node to track the instantaneous average $\bar{x}(n) = \frac{1}{n}\sum_{i=1}^{N} x_i(n)$, $\bar{x}(n) \in \mathbb{R}$, given that the value at each node $x_i(n)$ is time varying. Let $\delta_i(n) = x_i(n+1) - x_i(n)$ be the change in the value at node i, and $\delta(n) = x(n+1) - x(n)$ be the vector of changes at all nodes, $\delta(n) \in \mathbb{R}^N$. Suppose the magnitude of the change at time n is bounded by $|\delta_i(n)| = |x_i(n+1) - x_i(n)| \leq \epsilon(n)\ \forall i$. We make the following assumption:

Assumption 7.6. *The sequence* $\{\varepsilon(n)\}_{n=0}^{\infty}$ *is monotone non-increasing.*

Let $\hat{x}_i(n) \in \mathbb{R}$ be the estimate of $\bar{x}(n)$ at node i and let $\hat{x}(n) \in \mathbb{R}^N$ be the vector of estimates. We make the following assumption pertaining to the initial error in players' estimates.

Assumption 7.7. $\hat{x}_i(0) - \bar{x}(0) = 0\ \forall i$.

Let the average be estimated using the update rule

$$\hat{x}(n+1) = \mathbf{W}\left(\hat{x}(n) + x(n+1) - x(n)\right), \tag{7.9}$$

where the matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ is aperiodic, irreducible, and doubly stochastic with sparsity conforming to G. The following Lemma gives a bound for the error in the estimates of $\bar{x}(n)$.

Lemma 7.8. *Let the sequence $\{\hat{x}(n)\}_{n=1}^{\infty}$ be computed according to* (7.9) *where the communication graph G satisfies Assumption 1.1, the weight matrix $\mathbf{W}$ satisfies Assumption 7.3, the incremental change in $x(n)$ is bounded according to assumption Assumption 7.6 and the initialization satisfies Assumption 7.7. Then the error at any time n is bounded by,*

$$\|\hat{x}(n) - \bar{x}(n)\mathbf{1}\| \leq \frac{2\sqrt{N}}{1-\lambda}\epsilon_{avg}(n),$$

where $\lambda = \sup\limits_{y \in \mathbb{R}^N : \sum\limits_i y_i = 0} \frac{\|\mathbf{W}y\|}{\|y\|}$, *and* $\epsilon_{avg}(n) = \frac{1}{n}\sum\limits_{\tau=0}^{n-1}\epsilon(\tau)$ *is the time average of* $\{\epsilon(\tau)\}_{\tau=0}^{n-1}$.

Proof. Let $e(n) = \hat{x}(n) - \bar{x}(n)\mathbf{1}$ be the vector of errors in each players estimate of $\bar{x}(n)$, where $\mathbf{1}$ denotes the $N \times 1$ vector of all ones. Let $\bar{\delta}(n) = \frac{1}{n}\sum_i \delta_i(n), \ \forall t$. Using the relation (7.9) and the properties of doubly stochastic matrices, the vector of errors may be written recursively as,

$$e(n+1) = \mathbf{W}\left(e(n) + \xi(n)\right) \tag{7.10}$$

where $\xi(n) = \delta(n) - \bar{\delta}(n)\mathbf{1}$. Note that

$$|\xi_i(n)| = |\delta_i(n) - \bar{\delta}_i(n)| \leq |\delta_i(n)| + |\bar{\delta}_i(n)| \leq 2\epsilon(n),$$

and,

$$\|\xi(n)\|^2 = \sum_{i=1}^{N}(\xi_i(n))^2 \leq \sum_{i=1}^{N} 4\epsilon(n)^2 = 4n\epsilon(n)^2. \tag{7.11}$$

Using (7.10), the error $e(n)$ can be rewritten as a function of $\xi(n)$ and $e(0)$, that is $e(n+1) = \sum\limits_{r=0}^{n}\mathbf{W}^{r+1}\xi(n-r) + \mathbf{W}^{n+1}e(0)$. Using this relationship we establish an upper bound on the error,

$$\begin{aligned}\|e(n+1)\| &= \|\sum_{r=0}^{n}\mathbf{W}^{r+1}\xi(n-r) + \mathbf{W}^{n+1}e(0)\| \\ &\leq \sum_{r=0}^{n}\lambda^{r+1}\|\xi(n-r)\|,\end{aligned} \tag{7.12}$$

where we have employed assumption Assumption 7.7, $e(0) = 0$. Applying (7.11) in (7.12), we get $\|e(n+1)\| \leq \sum\limits_{r=0}^{n}\lambda^{r+1}2\sqrt{N}\epsilon(n-r)$.

Recall that $\varepsilon_{avg}(n) = \frac{1}{n}\sum_{\tau=0}^{n-1}\epsilon(\tau)$ is the time average of the sequence $\{\varepsilon(n)\}$ up to time n, and note that given our assumptions on $\mathbf{W}$, it holds that $\lambda < 1$ (see [21]). Note that, by Chebychev's sum inequality [138] (p. 43-44),

$$\sum_{r=0}^{n}\lambda^{r+1}2\sqrt{N}\epsilon(n-r) \leq \sum_{r=0}^{n}\lambda^{r+1}2\sqrt{N}\varepsilon_{avg}(n+1),$$

and hence,

$$\begin{aligned}\|e(n+1)\| &\leq \sum_{r=0}^{n}\lambda^{r+1}2\sqrt{N}\varepsilon_{avg}(n+1)\\ &= \left(\lambda\frac{1-\lambda^{n+1}}{1-\lambda}\right)2\sqrt{N}\epsilon_{avg}(n+1) \leq \frac{2\sqrt{N}}{1-\lambda}\epsilon_{avg}(n+1),\end{aligned}$$

giving the desired upper bound for the error. □

Lemma 7.9. *Let $\{a(n)\}_{n\geq 1}$ be a distributed ECFP process as defined in section 7.3.2 (see equations (7.4)-(7.2)). Then $\|\hat{q}_i(n) - \bar{q}(n)\| = O(\frac{\log n}{n})$, where $\bar{q}(n)$ is the average empirical distribution and $\hat{q}_i(n)$ is player i's estimate of $\bar{q}(n)$.*

Proof. We use the second argument, k, to index the components of the vector $q_i(n) \in \mathbb{R}^m$. Noting that

$$q_i(n+1) = q_i(n) + \frac{1}{n+1}\left(a_i(n+1) - q_i(n)\right),$$

it follows that the maximum incremental change for any single value in the vector $q_i(n)$ is bounded by $|q_i(n+1,k) - q_i(n,k)| \leq \frac{1}{n+1} = \varepsilon(n)$, where we let $\epsilon(n) := \frac{1}{n+1}$. Note that the distributed ECFP process (7.4) is updated column-wise (each column corresponds to an action k) using an update rule equivalent to (7.9) of Lemma 7.8. Also note that, column-wise, all necessary conditions of Lemma 7.8 are satisfied,[9] and specifically, we have $\epsilon(n) = \frac{1}{n+1}$. Thus we apply Lemma 7.8 column-wise to $\hat{\mathbf{Q}}$ and $\mathbf{Q}(n)$ of (7.4), where $x(n)$ of Lemma 7.8 corresponds to the k'th column of $\mathbf{Q}(n)$, and $\hat{X}(n)$ of Lemma 7.8 corresponds to the k'th column of $\hat{\mathbf{Q}}(n)$, and obtain

$$\|\hat{q}(n,k) - \bar{q}(n,k)\mathbf{1}\| \leq \frac{2\sqrt{N}}{1-\lambda}\epsilon_{avg}(n) = O\left(\frac{\log n}{n}\right),$$

where $\hat{q}(n,k) = (\hat{q}_1(n,k)\ \hat{q}_2(n,k)\cdots\hat{q}_N(n,k))^T$ and $\epsilon_{avg}(n) = \frac{1}{n}\sum_{s=1}^{n}\frac{1}{n+1} = O\left(\frac{\log n}{n}\right)$. Thus,

[9] The assumption of zero initial error (Assumption 7.7) is satisfied since the initialization of $\hat{q}_i(1)$ in (7.1) is equivalent to letting $\bar{q}_i(0) = 0$, $\hat{q}_i(0) = 0$ for all i in (7.3).

$|\hat{q}_i(n,k) - \bar{q}(n,k)| = O\left(\frac{\log n}{n}\right),\ k = 1,\dots,m,\ \forall i$, and hence, $\|\hat{q}_i(n) - \bar{q}(n)\| = O(\frac{\log n}{n}),\ \forall i.$ $\square$

The following lemma is used to prove convergence of distributed synchronous SS-FP algorithm.

Lemma 7.10. *Assume **A.2**–**A.5**. For $i, j \in N$, let $\hat{q}^i_j(n)$ be as defined in Section VI-B. Let the estimates $\hat{q}^i(n)$ be formed as in Section VI-C. Then $\lim_{n\to\infty} \|\hat{q}^i_j(n) - q_j(n)\| = 0,\ \forall i, j \in N$.*

Proof. Let $\bar{q}'(n) = \frac{1}{n}\sum_{i=1}^{N} q'_i(n)$ be the average of the augmented empirical distributions as defined in Section VI-B. Note that $\bar{q}'(n) \in \mathbb{R}^s$ is, in fact, a vector which stacks the true empirical distributions, $\bar{q}'(n) = (q_1(n),\ \cdots,\ q_N(n))$. Thus, by solving for $\bar{q}'(n)$, players are in fact solving for the true empirical distribution.

Let $k \in \{1,\dots,s\}$, and let $\tilde{q}_k(n) \in \mathbb{R}^N$ with

$$\tilde{q}_k(n) := (\hat{q}^1(n,k),\dots,\hat{q}^N(n,k)),$$

where $\hat{q}^i(n,k)$ is the k-th entry of the vector $\hat{q}^i(n)$. Note that the incremental change in any element of $\tilde{q}_k(n)$ is bounded by $1/n$, and that $\tilde{q}_k(n)$ is updated in a manner fitting the template of Lemma 1 of [104]. Invoking Lemma 1 of [104] (where (7.5) is to be seen as (7.7) for $n = 0$ and $\hat{q}^i(0) = 0$, $q_i(0) = 0$) gives

$$\lim_{n\to\infty} \|\tilde{q}_k(n) - \mathbf{1}\bar{q}'(n,k)\| = 0,$$

or equivalently, $\lim_{n\to\infty} \|\hat{q}^i(n,k) - \bar{q}'(n,k)\| = 0,\ \forall i$. Since this holds for all $k \in \{1,\dots,s\}$ and $\bar{q}'(n) = q(n)$, (where $q(n)$ is considered as a vector), this gives the desired result. $\square$

Chapter 8

Inertial MBR Algorithms: Incomplete Information and Network-Based Implementation

8.1 Introduction

Game theoretic learning processes have gained increasing prominence in the control, signal processing, and communication communities due to their applicability as decentralized algorithms [1, 2, 7, 18, 101, 139]. In these contexts, a multi-agent system is often modeled using the framework of a normal-form game [18, 20]. Pure-strategy Nash equilibria of the game can often be shown to represent desirable operating conditions in the associated multi-agent system (e.g. [12]).

There exist many game-theoretic learning algorithms that ensure agents learn pure NE strategies (e.g., [73, 74]). However, such algorithms generally assume that agents have instantaneous access to information necessary to compute their next-stage action. In many real-world scenarios, global information is not instantly available to all agents—rather, information tends to be distributed throughout the system and must be disseminated to all agents using some overlaid communication infrastructure, e.g., wireless networks [1, 2], multi-robot systems [12], smart grid infrastructures [6, 7].

In this chapter, we consider a setting in which each agent has access to private information about their own action history, but does not have (direct) access to information about the actions of others, nor access to instantaneous payoff information [77, 140]. We assume that agents are equipped with an overlaid communication graph where a vertex represents a player and an edge denotes the ability for two players to exchange information (e.g., see [78, 79, 104, 141, 142]). An agent may only obtain information related to another agent's action history if that information is disseminated through the communication graph.

The assumption that players do not have access to payoff information is relevant in settings where feedback from physical game play is not available, or is delayed. This may occur, for example, in scenarios where agents wish to precompute an equilibrium strategy prior to physical game play, or where the designed utility functions do not readily admit physical measurement. In such a framework, information can be subject to corruption and delay as it is disseminated through the communication graph, and the traditional convergence results of pure-strategy-NE seeking algorithms do not apply.

In addition to assuming that information flow is restricted to an inter-agent communication graph, we suppose that agents may have some uncertainty about the state of the world. We model this uncertainty using a game of incomplete information in which players have some uncertainty about the game utility functions. We assume that each player receives a private signal that is correlated to the true (time-invariant) state of the world. Using this information, each agent forms a personal belief about the state of the world, which they may update as information is shared over the communication graph through the course of the learning process. A learning process in this setting must take into account the joint objectives of uncertainty mitigation and pure-strategy NE learning.

In order to learn pure-strategy NE within the framework described above, we consider a class of MBR learning algorithms and demonstrate convergence results within the class of weakly acyclic games [143].

The problem of learning Nash equilibria (possibly mixed) through network-based implementation of MBR dynamics in multi-agent networks has been addressed in prior work [78, 104, 142] (see also, Chapter 7). In this chapter we focus on network-based equilibria learning in which the agents specifically seek to learn *pure-strategy* Nash equilibria which is not guaranteed by the network-based learning mechanisms presented in [78, 104] (and Chapter 7) and also, in settings with incomplete information [142].

In order to learn pure-strategy equilibria in this setting we consider a class of algorithms we denote as *inertial MBR dynamics*—such algorithms consist of a best-response component (similar to classical FP) and an additional "inertial term" which ensures that players occasionally repeat actions in consecutive stages. The use of an inertial term is a common technique used to ensure convergence to pure-strategy NE [73, 74].

We first prove a general convergence result that applies to any inertial MBR algorithm. The general convergence result applies to a broad class of learning dynamics and applies to a wide variety of information dissemination schemes (not just the synchronous graph-based information dissemination schemes that we focus on in the later part of the chapter).

Subsequently, as applications of the general result, we study the network-based implementation (under uncertainty) of two important inertial MBR algorithms.

As a first application, we study the network-based implementation of classical FP with inertia (referred to as N-FP) (see Section 8.4).

As a second application, we study the network-based implementation of a variant of the Joint Strategy FP (JSFP) algorithm [73] that is applicable within the class of congestion games. (We refer to this network-based algorithm as N-JSFP.) Classical FP can be difficult to implement in large-scale games due to high demands in terms of information overhead and complexity as the number of players grows large. JSFP is a variant of FP that mitigates these issues. We note that while the N-JSFP algorithm of this chapter is able to operate with lower computational and communication burdens than N-FP, it is applicable within a narrower class of games than N-FP. (This is a consequence of the structure of the information that is passed through the network—see Section 8.5 for more details.)

In summary, the chapter investigates the problem of learning pure-strategy NE in games where there is environmental uncertainty and players must communicate information using an overlaid communication graph. As our first main contribution, we study a general class of learning algorithms that we call inertial MBR dynamics, and we prove such algorithms converge to pure-strategy NE under appropriate assumptions. Our second main contribution is an application of this general result to develop variants of the FP and JSFP algorithms that achieve pure-strategy NE learning in network-based settings with uncertainty.

It should be noted that in Chapter 10 we will address the closely related problem of studying continuous-time FP dynamics in potential games. We prove that in almost all potential games, for almost all initial conditions, continuous-time FP converges to a pure-strategy equilibrium (see Theorem 10.1). We suspect that a similar result holds for discrete-time FP in potential games, although no such result has yet been proven.

This result, while powerful, does not obviate the results of this chapter. In this chapter we study convergence to pure-strategy equilibria in a broader class of games and a broader class of dynamics than considered in Theorem 10.1. In particular, we study the general class of myopic best response algorithms (with inertia), and we prove convergence of such algorithms to a pure-strategy NE in the class of weakly acyclic games (a class of games strictly larger than potential games).

The remainder of the chapter is organized as follows. In Section 8.2 we set up the problem formulation. In Section 8.3 we present the general convergence result for inertial MBR dynamics. In Section 8.4 we study distributed implementations of classical FP with inertia. In Section 8.5 we study distributed implementations of the Joint Strategy FP algorithm with inertia. In Section 8.6

we present simulation examples.

8.2 Networked Multi-Agent Systems with Incomplete Information

In this chapter we consider scenarios where players in a game have incomplete information about the state of the world. Formally, let Θ to be the set of possible states of the world and assume that Θ is a metrizable space. We let $\mathcal{B}(\Theta)$ denote the Borel σ-algebra on Θ and let $\triangle(\Theta)$ denote the set of probability measures on the measurable space $(\Theta, \mathcal{B}(\Theta))$.

For each state of the world $\theta \in \Theta$ we consider the associated normal-form game defined by the tuple

$$\Gamma(\theta) = (\mathcal{N}, \{Y_i, u_i(\cdot, \theta)\}_{i \in \mathcal{N}}), \tag{8.1}$$

where $\mathcal{N}$ and $(Y_i)_{i \in \mathcal{N}}$ are as defined in Section 2.1, and $u_i(\cdot, \theta) : Y \to \mathbb{R}$ denotes the utility function of player $i \in \mathcal{N}$ given the state θ. Observe that the set of players and the set of possible actions are assumed to be the same irrespective of the state of the world θ. When player i chooses strategy y_i, the payoff it receives is $u_i(\{y_i\}_{i \in \mathcal{N}}, \theta) = u_i(y, \theta)$. The payoff depends on the joint action profile $y = \{y_i\}_{i \in \mathcal{N}}$ and the state of the world θ. As in the complete-information setup of Section 2.1, we let Δ_i denote the mixed strategy space of player i and let Δ^N denote the set of joint mixed strategies where it is assumed that players use independent strategies.

When a mixed strategy $\sigma \in \Delta$ is played, we are interested in the *expected* payoffs which we write as

$$U_i(\sigma, \theta) \;=\; U_i(\sigma_i, \sigma_{-i}, \theta) \;:=\; \sum_{y \in Y} u_i(y, \theta)\sigma(y). \tag{8.2}$$

The notation $U_i(\sigma_i, \sigma_{-i}, \theta)$ in (8.2) is meant to emphasize that the payoff depends on the strategy σ_i chosen by player i and the strategies $\sigma_{-i} := \{\sigma_j\}_{j \in \mathcal{N}, j \neq i}$ that are chosen by other players.

Since we want to study games in which θ is uncertain, we further define expected payoffs with respect to a probability distribution $\mu \in \triangle(\Theta)$—we refer to μ as a *belief* about the state of the world. The expectation of the payoff in (8.2) with respect to this belief is then given by

$$U_i(\sigma, \mu) \;=\; U_i(\sigma_i, \sigma_{-i}, \mu) \;:=\; \int U_i(\sigma, \theta)\mu(d\theta). \tag{8.3}$$

The notation $U_i(\sigma, \mu)$ in (8.3) is a slight abuse of the notation $U_i(\sigma, \theta)$ in (8.2). This inconsistency is resolved if we use θ to denote a state realization but also as a shorthand for the belief μ with all mass on θ. Similarly, we will use y_i and y to denote actions but also as a shorthand for strategies σ_i and σ with all mass on y_i and y respectively. We adopt this convention henceforth.

The payoffs in (8.3) define a modification of the game $\Gamma(\theta)$ in (8.1) in which the payoffs are given by the expected utilities in (8.3). The equilibria of this modified game are formally defined next.

Definition 8.1. *Given the game $\Gamma(\theta)$ in (8.1) and a belief μ define the game $\Gamma(\mu) := (\mathcal{N}, \{Y_i, u_i(\cdot, \mu)\})$ associated with the utilities in (8.3). The joint strategy $\sigma = \{\sigma_i\}_{i\in\mathcal{N}} \in \Delta$ is a Nash equilibrium of $\Gamma(\mu)$ if the utilities in (8.3) satisfy*

$$U_i(\sigma_i, \sigma_{-i}, \mu) \geq U_i(\sigma_i', \sigma_{-i}, \mu), \quad \textit{for all } \sigma_i' \in \Delta_i,\ i \in \mathcal{N}. \tag{8.4}$$

The equilibrium is said to be a pure-strategy equilibrium if there exists an action tuple $y \in Y$ such that $\sigma = \mathbf{1}_y$.

The equilibria in (8.4) exist because, with μ fixed, the payoffs in (8.3) define a normal form game. In general, pure equilibria may or may not exist, but we will assume here that they do (see Assumption 8.4). In such cases, identifying pure equilibria is important in applications because they result in behaviors that are often more reasonable than the behaviors that result from mixed equilibria—see Section 8.1. Our goal in this chapter is to design an algorithm allowing agents to learn the pure equilibria in Definition 8.1 when players play repeatedly over time and acquire a common belief about the state of the world. We formally state this problem in the following section.

8.2.1 Learning via Repeated Play

Players are permitted to repeatedly face off in the game $\Gamma(\theta)$ in discrete stages $n \in \mathbb{N}$. At each stage n, agent i plays an action $y_i(n) \in Y_i$ that is chosen from the (possibly mixed) strategy $\sigma_i(n) \in \Delta_i$. To select the strategy $\sigma_i(n)$, agent i has access to a belief μ_n^i about the state of the world, and has some knowledge of the past history of game play $\{y_j(s)\}_{s=1}^{n-1}$, $j \in \mathcal{N}$. The beliefs $\mu_n^i \in \triangle(\Theta)$ are time varying and possibly different for different agents, although they are required to converge to a common distribution—see Assumption 8.3. Besides this belief about the state of the world, agent i also has estimates $\hat{\sigma}_j^i(n) \in \Delta_j$ of the strategies $\sigma_j(n) \in \Delta_j$ of other agents. The belief μ_n^i and the estimated strategies $\hat{\sigma}_{-i}^i(n) := \{\hat{\sigma}_j(n)\}_{j\in\mathcal{N}, j\neq i}$ allow agent i to estimate the payoff that it would receive from playing an arbitrary action y_i. These estimated payoffs can be written as [cf. (8.3)]

$$\hat{U}_i(y_i, n) := U_i(y_i, \hat{\sigma}_{-i}^i(n), \mu_n^i), \quad y_i \in Y_i. \tag{8.5}$$

The utility $\hat{U}_i(y_i, n)$ in (8.5) is what agent i predicts he would receive as payoff given the (partial) information he has available on the state of the world (the belief μ_n^i) and the actions of other agents (the strategy estimates $\hat{\sigma}^i_{-i}(n)$).

The action that maximizes the utility estimates in (8.5) is of importance to agent i and defined here as the best response

$$\hat{y}_i(n) \;:=\; \arg\max_{y_i \in Y_i} \hat{U}_i(y_i, n) \;=\; \arg\max_{y_i \in Y_i} U(y_i, \hat{\sigma}^i_{-i}(n), \mu^i_n). \tag{8.6}$$

Observe that there may be multiple arguments that maximize $\hat{U}_i(y_i, n)$. Thus, the best response $\hat{y}_i(n)$ in (8.6) is in general a set, not an individual action.

A learning algorithm in game theory is a collection of behavior rules that dictate how each player should choose the "next-stage action" $y_i(n+1)$ based on the information they have available at time n. Myopic best-response rules in which agents play an action $y_i(n+1) \in \hat{y}_i(n)$ are common; see e.g., [20, 26, 74]. Here, we are interested in a slight modification in which agents are sometimes "reluctant" to modify their action choices from round to round. We refer to this general algorithm (formally stated next) as *inertial MBR dynamics.*[1]

Algorithm 8.2. Let $\rho \in (0, 1)$ be an inertia constant and let $y_i(1)$ be an arbitrary initial action for each i. At time $n > 1$, agent i has access to strategy estimates $\hat{\sigma}^i_j(n)$ and a belief μ^i_n that it uses to compute the best-response set $\hat{y}_i(n)$ in (8.6). Players are said to follow inertial MBR dynamics if they play actions according to

$$\mathbb{P}\left(y_i(n+1) = y_i(n) | \mathcal{F}_{n-1}\right) = \rho, \quad \mathbb{P}\left(y_i(n+1) \in \hat{y}_i(n) | \mathcal{F}_{n-1}\right) = 1 - \rho, \tag{8.7}$$

where $(\mathcal{F}_n)_{n \geq 1}$ is a filtration (sequence of increasing σ-algebras) that contains the information available to players in round n. As per (8.7), an inertial MBR algorithm entails player i sticking to its previous play $y_i(n)$ with some (fixed) probability ρ (this is the inertia component of the algorithm), and playing a best response otherwise. The selection of myopic best responses promotes a general trend of improving utility, while the inertia component of the algorithm—with its occasional repetition of actions—ensures that players eventually lock into a pure strategy equilibrium. Examples of algorithms that use best response with inertia can be found in [73, 74].

Throughout this chapter, we want to study mechanisms to update the beliefs $\hat{\sigma}^i_j(n)$ so that players learn pure equilibria (cf. Definition 8.1). We do so using ideas from fictitious play, so that (roughly speaking) $\hat{\sigma}^i_j(n)$ is built from empirical histograms of the actions that have been observed or action information that has been disseminated through an overlaid communication graph. These algorithms are defined in Sections 8.4 and 8.5 after we introduce some preliminary

[1]We also refer to an algorithm of this form as an *inertial MBR algorithm.*

convergence results for the generic inertial MBR methods given in Algorithm 8.2.

8.3 Inertial Best-Response Algorithm: General Convergence Result

In this section we study convergence results for the inertial best-response learning dynamics presented in the previous section (see Algorithm 8.2). These dynamics can be shown to achieve pure-strategy NE learning so long as the estimated payoffs $\hat{U}_i(n) = (\hat{U}_i(y_i, n))_{y_i \in Y_i}$ satisfy an appropriate asymptotic accuracy condition (see Condition 8.7). We first present some general assumptions regarding game play, and then present the convergence results.

In later sections, these results will be used to prove convergence of two particular instances of inertial best-response dynamics.

8.3.1 General Assumptions

We first introduce some notation necessary for the subsequent development. Suppose S is a metric space with Borel σ-algebra Σ, and let $\{\mu_n\}_{n\geq 1}$ and μ be probability measures on (S, Σ). The sequence $\{\mu_n\}_{n\geq 1}$ converges weakly to μ if $\int g(s)\mu_n(ds) \to \int g(s)\mu(ds)$ as $n \to \infty$ for any bounded continuous function $g : S \to \mathbb{R}$. In this case we write $\mu_n \xrightarrow{w} \mu$.

We assume that players' beliefs about the state of the world converge weakly to a fixed common distribution. Formally,

Assumption 8.3. *There exists a probability measure $\mu \in \Delta(\Theta)$ such that $\mu_n^i \xrightarrow{w} \mu$ for all $i \in \mathcal{N}$.*

This assumption is easily attainable by a variety of belief formation mechanisms. For example, suppose that there exist sequences $(\theta_n^i)_{n\geq 1} \subset \Theta$ satisfying $\lim_{n\to\infty} \theta_n^i = \theta$, $\forall i$. Then the belief sequence where μ_n^i places mass 1 on θ_n^i satisfies Assumption 8.3.

Throughout the chapter, our focus is aimed at studying learning dynamics and not the particular process used to form the beliefs μ_n^i. In an effort to maintain this focus, we will not explicitly specify any particular dynamics used to form the belief μ_n^i—rather, we will study learning results that can be achieved under the broad assumption of weak convergence of beliefs. In Section 8.6 we discuss a simulation example in which the dynamics used to form the belief μ_n^i are explicitly specified and satisfy Assumption 8.3.

Given that there exists a μ such that $\mu_n^i \xrightarrow{w} \mu$, for all i, we assume further that under the belief μ, the game is weakly acyclic. Formally,

Assumption 8.4. *Let μ be as in Assumption 8.3 and define the normal form game Γ associated with the belief μ as $\Gamma(\mu) = (\mathcal{N}, \{Y_i, u_i(\cdot, \mu)\})$. We assume the game $\Gamma(\mu)$ is weakly acyclic; that is, for any $y \in Y$, there exists a best-response path that converges to a pure-strategy Nash equilibrium.*

A review of the topic of weakly acyclic games can be found in [143]. The above assumption ensures that the set of pure-strategy NE (see Definition 8.1) is non-empty, and will ensure that the learning algorithms to be studied can always reach a pure-strategy NE.

Finally, we assume that under the limiting belief μ, players are never indifferent between action choices. Formally,

Assumption 8.5. *For any pair of actions $y_i' \in Y_i$, $y_i'' \in Y_i$, $y_i' \neq y_i''$ and for any $y_{-i} \in Y_{-i}$,*

$$U_i(y_i', y_{-i}, \mu) \neq U_i(y_i'', y_{-i}, \mu)$$

for all $i \in \mathcal{N}$.

Note that Assumption 8.5 (cf. [73] Assumption 2.2) implies that under the belief μ all pure NE are strict. Moreover, the assumption implies that for any $i \in \mathcal{N}$ and $y_{-i} \in Y_{-i}$, the set $\arg\max_{y_i \in Y_i} U_i(y_i, y_{-i}, \mu)$ is a singleton.

We remark that Assumption 8.5 is generic in the sense that if the number of players and actions are fixed, then the set of utility functions for which Assumption 8.5 fails is a closed set of Lebesgue measure zero within the space of all possible utility functions.

Assumption 8.6. *Let $\{\mathcal{F}_n\}_{n\geq 1}$ be a filtration (sequence of increasing σ-algebras) with $\mathcal{F}_n := \sigma(\{y(s)\}_{s=1}^n, \{\mu_s^i\}_{i\in\mathcal{N}, s=1,\dots,n})$. The strategy estimate $\hat{\sigma}_j^i(n) \in \Delta_j$ that agent i has of the strategy $\sigma_j(n)$ of agent j is measurable with respect to $\mathcal{F}_n$.*

Assumption 8.6 means that the strategy estimates of agent i are restricted to be a function of the history of play.

The following condition is sufficient (to be shown) to ensure that a MBR algorithm with inertia converges to NE.

Condition 8.7. *There exist $\bar{n}$, $T \in \mathbb{N}$ such that, if $n \geq \bar{n}$, and if any action $y \in Y$ is repeated consecutively for $\tilde{T} \geq T$ stages (i.e., $y(s) = y$ for $s = n, \dots, n + \tilde{T} - 1$), then $\arg\max_{y_i \in Y_i} \hat{U}_i(y_i, n + \tilde{T} - 1) = \arg\max_{y_i \in Y_i} U_i(y_i, y_{-i}, \mu)$ for all $i \in \mathcal{N}$.*

Throughout this section we assume that this condition holds. (In later sections, when studying particular inertial best-response algorithms, our main task will be to show that Condition 8.7 holds.) The condition states that there exists a finite time $\bar{n}$ after which if any action profile y is repeated a sufficiently large number of times (specifically, more than T times) then player i's optimal response to its predictions of the utility coincides with the optimal response to the actual action profile y_{-i}. This condition is natural within the class of learning algorithms with fading or finite memory [74] where, if players repeat an action a sufficient number of times their prediction of the utility can be brought arbitrarily close to the utility of the repeatedly-played action.

When studying particular implementations of inertial best-response dynamics in the following sections, the condition reduces to tangible assumptions on learning past action profiles of other players—see Conditions 8.17 and 8.24.

8.3.2 General Convergence Analysis

The following two lemmas allow us to prove the main theoretical result of this section. Lemma 8.8 shows that pure-strategy Nash equilibria are absorbing, and Lemma 8.9 shows that the probability of reaching such an absorbing state is uniformly bounded from below. Together they prove Theorem 8.10.

Lemma 8.8 (absorption property)**.** *Let $\bar{n}$ be as in Condition 8.7. Let $\{y(n)\}_{n\geq 1}$ be a sequence of actions generated by an inertial best-response algorithm. Suppose Assumptions 8.3–8.6 and Condition 8.7 hold. There exists a $T_1 \in \mathbb{N}$ such that if $n \geq \bar{n}$, and $y^* \in Y$ is any pure-strategy Nash equilibrium, and if y^* is played in T_1 consecutive stages, i.e., $y(s) = y^*$, $\forall s = n, \ldots, n+T_1-1$, then $y(n+\tau) = y^*$ for all $\tau \geq 0$.*

Proof. Let T be as in Condition 8.7, and let $T_1 \geq T$. Suppose y^* is a pure Nash equilibrium and $y(s) = y^*$ for $s = n, \ldots, n+T_1-1$. Then by Condition 8.7, $\arg\max_{y_i'\in Y_i} \hat{U}_i(y_i', n+\tilde{T}) = \arg\max_{y_i'\in Y_i} U_i(y_i', y_{-i}^*, \mu)$, $\forall i$ (recall that $\hat{U}_i(y_i', n+\tilde{T})$ as defined in Section 8.2.1 is implicitly dependent on $\mu^i_{n+\tilde{T}}$). Moreover, by Assumption 8.5, the set $\arg\max_{y_i'\in Y_i} U_i(y_i', y_{-i}^*, \mu) = \{y_i^*\}$ is a singleton for each i. Thus, the action y^* is repeated in stage $n+T_1$. Inductively, we see that $y(n+\tau) = y^*$ for all $\tau \geq 0$. □

Lemma 8.9 (positive probability of absorption)**.** *Let $\bar{n}$ be as in Condition 8.7. Let $\{y(n)\}_{n\geq 1}$ be a sequence of actions generated by an inertial best-response algorithm. Suppose Assumptions 8.3–8.6 and Condition 8.7 hold. Let T_1 be as in Lemma 8.8, and let $T_2 \geq T_1$ be given. Let n be the current stage of the repeated play. Define the event*

$$
\begin{aligned}
E_n := \{y(\tau) = y^* \textit{ for some pure strategy NE } y^* \\
\textit{for all } \tau \in \{n', n'+1, \ldots, n'+T_2-1\}, \\
\textit{for some } n' \in \{n, \ldots, n+T_2|Y|\}\}.
\end{aligned}
$$

There exists an $\epsilon = \epsilon(T_2) > 0$ such that $\mathbb{P}(E_n|\mathcal{F}_n) > \epsilon$ for all $n \geq \bar{n}$.

Proof. The proof follows along the lines of the proof of Theorem 3.1 in [73]. By Condition 8.7, for any $n' \geq \bar{n}$, if any action $y \in Y$ is repeated consecutively from stage n' to stage $n' + T_2 - 1$, then $\arg\max_{y_i \in Y_i} \hat{U}_i(y_i, n' + T_2 - 1) = \arg\max_{y_i \in Y_i} U_i(y_i, y_{-i}, \mu)$. Let $y^0 = y(n)$. Conditioned on $\mathcal{F}_n$, the action y^0 will be played repeatedly in T_2 consecutive stages with probability at least $\epsilon_1 := \rho^{N(T_2-1)} > 0$. Supposing this occurs, then at stage $\tau = t + T_2 - 1$, $\arg\max_{y_i \in Y_i} \hat{U}_i(y_i, \tau) = \arg\max_{y_i \in Y_i} U_i(y_i, y^0_{-i}, \mu)$. At this point, either no players can improve their utility (in which case we are at a pure NE), or at least one player can improve their utility. If the latter is the case then, conditioned on $\mathcal{F}_{n+T_2-1}$, with probability at least $\epsilon_2 := \rho^{N-1}(1-\rho)$, exactly one player i chooses to take a best response and improves their utility, and all others continue to play y^0_{-i}. Call the new action profile y^1. Continuing in this manner, we can construct a sequence of actions $y^0, y^1, \ldots, y^m$ (terminating with at most $m = |Y|$) such that y^m is a pure-strategy Nash equilibrium. Conditioned on $\mathcal{F}_n$, the probability of this action sequence occurring (and then the final action y^m being played for T_2 consecutive stages) is bounded from below by $\epsilon := (\epsilon_1 \epsilon_2)^{|Y|} \rho^{T_2 - 1}$. $\square$

The lemma establishes that from any state of beliefs and predictions, the probability that a pure strategy Nash equilibrium is reached and repeated until absorption is positive. Next, we state our main result for this section.

Theorem 8.10. *Let $\{y(n)\}_{n \geq 1}$ be a sequence of actions generated by inertial best-response dynamics. Suppose Assumptions 8.3–8.6 and Condition 8.7 hold. Then the action sequence $\{y(n)\}_{n \geq 1}$ converges to a pure-strategy NE of the game $\Gamma(\mu)$, almost surely. Moreover, let $\tau \in \mathbb{N} \cup \{\infty\}$ be a random variable indicating the round number in which the action sequence $y(n)$ is absorbed to a pure-strategy NE. Then $E(\tau) < \infty$.*

Proof. Let $n \geq \bar{n}$ and let T_2 be as in Lemma 8.9. By Lemma 8.8, if a pure NE action y^* is played in T_2 consecutive stages, then y^* will be played in all consecutive stages. By Lemma 8.9, the probability of reaching such an "absorbing state" is uniformly lower bounded by some $\epsilon > 0$. Thus, the process is absorbed to a pure NE almost surely in finite time, and $E(\tau) < \infty$ ([122], p.233). $\square$

In the following sections, we will use this result to study the convergence of various instances of inertial MBR algorithms in settings with environmental uncertainty and limited communication. In particular our main task will be to show that the algorithms under consideration satisfy Condition 8.7.

8.4 Network-Based Implementation of Fictitious Play with Inertia

In this section we study a variation of the classical FP algorithm in which the best response of classical FP is augmented with an inertia term, and inter-agent communication is restricted to a graph. We refer to the class of algorithms studied in this section as network-based FP (N-FP).

The main contributions of this section are twofold. First, we present the general form of the N-FP algorithm (see Section 8.4.3); this general algorithm explicitly specifies the learning dynamics to be used, but does not explicitly specify the dynamics of the inter-agent communication scheme. We show that any inter-agent communication scheme will lead to pure-strategy NE learning so long as it satisfies a fairly mild condition (see Condition 8.17). Subsequently, we present an example implementation of N-FP (see Section 8.4.5) in which the inter-agent communication scheme is explicitly specified and we demonstrate that the example implementation converges to a pure-strategy NE.

We note that in Chapter 10 it will be shown, roughly speaking, that continuous-time FP almost always converges to a pure-strategy NE in potential games. We suspect that a similar result holds for discrete-time FP in potential games, although no such result has yet been proven. The inertial-based FP algorithm studied in this chapter converges to a pure NE in the broader class of weakly-acyclic games—a class of games strictly subsuming potential games. However, if the game is a potential game, it may be possible to ensure convergence to pure-strategy NE using standard FP dynamics if one can prove a discrete-time analog of Theorem 10.

We begin this section by reviewing the centralized FP algorithm with fading memory and inertia.

8.4.1 Fictitious Play with Inertia and Fading Memory

A review of the classical FP algorithm can be found in Section 2.5 and a review of FP with inertia can be found in [74].

The FP with inertia algorithm is defined as follows. Let $q_i(n) \in \mathcal{R}^{|Y_i|}$ denote the weighted empirical distribution (or just *empirical distribution*) of player i. Let $\alpha \in (0,1]$ be a step-size parameter. Let $q_i(n)$ be defined recursively as $q_i(1) = \mathbf{1}_{y_i(1)}$ and for $n \geq 1$,

$$q_i(n+1) = (1-\alpha)q_i(n) + \alpha \mathbf{1}_{y_i(n+1)}. \tag{8.8}$$

Let the joint weighted empirical distribution profile (or joint empirical distribution) be given by $q(n) := (q_1(n), \ldots, q_N(n))$. The weighted empirical distribution is said to have "fading memory" because it places greater weight on recent events.[2]

In fictitious play with fading memory and inertia, each player chooses their next-stage action according to the rule

$$y_i(n+1) \in \begin{cases} \arg\max_{y_i \in Y_i} U_i(y_i, q_{-i}(n), \theta) & \text{with prob. } 1-\rho, \\ y_i(n) & \text{with prob. } \rho, \end{cases}$$

where $\rho \in (0,1)$ is some predefined inertia constant and the probability is conditioned on $\mathcal{F}_{n-1}$ (see Assumption 8.6).[3]

In the network-based setting, player i may not have precise knowledge of the action history of player j, and hence may not have precise knowledge of $q_j(n)$. Let $\hat{q}_j^i(n)$ be an estimate that player i maintains of $q_j(n)$. Let $\hat{q}^i(n) = (\hat{q}_1^i(n), \ldots, \hat{q}_N^i(n))$ be an estimate that player i maintains of the empirical distribution profile $q(n)$.

Next, we present the network-based setup and inter-agent communication scheme.

8.4.2 Network-Based Setup

In a large-scale setting, the physically distributed nature of the system can make it difficult for players to observe the actions taken by every other player. We assume that players may be unable to observe the actions taken by others, but are equipped with an overlaid communication infrastructure through which they may exchange information with neighboring players. Formally, we will assume Assumptions 1.1–1.3 hold.

8.4.3 General N-FP Algorithm

We now state the general N-FP algorithm. The algorithm follows the format of an inertial MBR algorithm, but in this case, the utility prediction of player i, $\hat{U}_i(n) = (\hat{U}_i(y_i, n))_{y_i \in Y_i}$, is formed using player i's estimate of the empirical distribution $\hat{q}_{-i}^i(n)$ (in particular, this is the product distribution of $\{\hat{q}_j^i(n)\}_{j \neq i}$). Intuitively speaking, this may be thought of as player i forming the (possibly incorrect) belief that her estimate of the empirical distribution $\hat{q}_{-i}^i(n)$ accurately represents the mixed strategies of opponents.

[2] While the use of inertia is essential to the structure of the proofs in this chapter, the use of fading memory is less critical. It is possible that the results still hold using a time-varying step size α_t, (e.g., [73], Section II-E); however, the assumption of fading memory simplifies the analysis.

[3] We note that in the classical fictitious play algorithm it is common to omit the state parameter θ from the argument of the utility function, as the implicit assumption is that the players know their own utilities, i.e., $u_i : Y \to \mathbb{R}$. Here we have kept θ to be consistent with our earlier notation.

Algorithm 8.11.

Initialize

(i) Let $\rho \in (0,1)$ be fixed. For each i, let $y_i(1)$ be chosen arbitrarily, and let the estimate $\hat{q}^i(1)$ be initialized as $\hat{q}^i(1) = \mathbf{1}_{y_i(1)}$.

Iterate $(n \geq 1)$

(ii) Each agent i chooses their next-stage action according the rule

$$y_i(n+1) = \begin{cases} \arg\max_{y_i \in Y_i} U_i(y_i, \hat{q}^i_{-i}(n), \mu^i_n) & \text{with prob. } 1-\rho, \\ y_i(n) & \text{with prob. } \rho. \end{cases}$$

(iii) Each player i updates their personal empirical distribution $q_i(n)$ according to (8.8).

(iv) Each player i engages in one round of information exchange with neighboring agents $j \in \mathcal{N}_i$, and updates their estimate of the joint empirical distribution $\hat{q}^i(n)$.

In the following section we analyze convergence of N-FP. We prove convergence to a pure-strategy NE by showing that the algorithm fits the template of Theorem 8.10.

Remark 8.12. Note that the framework assumed by Assumptions 8.3 and 1.1 may is a generalization of the classical repeated play learning framework. The classic framework of centralized and complete information is recovered by letting G be the complete graph and letting $\mu^i_n = \mu = \delta_\theta$ for all $i \in \mathcal{N}$ and $n \geq 1$, where δ_θ is the Dirac delta distribution placing mass 1 on the point $\{\theta\}$.

Remark 8.13. Note that it has not been precisely specified how $\hat{q}^i(n)$ is updated. In the following section we give a general sufficient condition on the update rule under which convergence to pure NE can be guaranteed.

8.4.4 Convergence Analysis for Generic N-FP Algorithm

We first introduce some notation necessary for the subsequent development. Given metric space (X, d_X) and a family of functions $\{g_\alpha\}_{\alpha \in A}$ with $g_\alpha : X \to \mathbb{R}$ for all $\alpha \in A$ we say $\{g_\alpha\}_{\alpha \in A}$ is uniformly equicontinuous if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $|g_\alpha(x) - g_\alpha(y)| < \epsilon$, $\forall \alpha \in A$ whenever $d_X(x,y) < \delta$, $x, y \in X$.

In order to ensure that weak convergence in beliefs results in a meaningful notion of convergence in terms of players' utility functions, we require that the utility functions satisfy the following assumption:

Assumption 8.14. *For each $i \in \mathcal{N}$, the family of functions $\{u_i(\cdot, \theta)\}_{\theta \in \Theta}$ is uniformly equicontinuous.*

The following lemma allows us to establish the existence of a finite time such that after this time players' beliefs μ_n^i are sufficiently "close together" for our subsequent analysis.

Lemma 8.15. *Suppose Assumptions 8.3 and 8.14 hold. Let $\epsilon > 0$. There exists a time $T > 0$ such that $|U_i(\sigma, \mu_n^i) - U_i(\sigma, \mu)| < \epsilon, \ \forall \sigma \in \Delta^N, \ \forall i$, for all $n \geq T$.*

The proof of the lemma follows immediately from Lemma 8.30 in the appendix.

Let

$$\eta := \min_{\substack{y_i', y_i'' \in Y_i, \ y_i' \neq y_i'', \\ y_{-i} \in Y_{-i}, \\ i \in \mathcal{N}}} \frac{1}{4} |U_i(y_i', y_{-i}, \mu) - U_i(y_i'', y_{-i}, \mu)|;$$

the above lemma implies that there exists a time $n_\eta \in \mathbb{N}$ such that, for any $\sigma \in \Delta^N$ and any $i \in \mathcal{N}$ there holds $|U_i(\sigma, \mu_n^i) - U_i(\sigma, \mu)| < \eta$ for all $n \geq n_\eta$. Intuitively, this means that for all $n \geq n_\eta$, players' beliefs are sufficiently close so that any disparities have negligible effect on the utilities. We note that the existence of such a time n_η is not trivially obtained and relies explicitly on the uniform equicontinuity of the family $\{u_i(\cdot, \theta)\}_{\theta \in \Theta}$. The time n_η here plays the role of time $\bar{n}$ in Section 8.3.

Remark 8.16. Suppose $n \geq n_\eta$. Then for any player, the best response to any pure strategy is unique; i.e., $\arg\max_{y_i \in Y_i} U_i(y_i, y_{-i}, \mu_n^i)$ is a singleton for all $y_{-i} \in Y_{-i}$, for all i. This follows immediately from Assumption 8.5 and the definition of n_η.

The general form of N-FP (see Section 8.4.3) explicitly specifies the learning dynamics to be used, but does not explicitly specify the dynamics of the inter-agent communication scheme in step (iii) of the N-FP algorithm. Any protocol for determining $\hat{q}^i(n)$ will be valid so long as the following condition is satisfied:

Condition 8.17. *There exists a $T \in \mathbb{N}$ such that if any action $y^* \in Y$ is repeated in $\tilde{T} \geq T$ consecutive stages (i.e., $y(s) = y^*, \ s = n, \ldots, n + \tilde{T} - 1$), then $\|\hat{q}^i(n + \tilde{T} - 1) - q(n + \tilde{T} - 1)\| < \eta$ for all i.*

We note that the condition is fairly mild and is attainable by a wide variety of distributed tracking algorithms (e.g. [144,145]). In the following section we will present an example implementation of N-FP in which this condition is explicitly satisfied.

The following lemma shows that if Condition 8.17 is satisfied, then there exists a time T such that if players repeatedly play an action in T consecutive stages, then the set of maximizers of $U_i(\cdot, \hat{q}^i(n), \mu_n^i)$ and $U_i(\cdot, q(n), \mu)$ coincide. The time T here is analogous to the time T in Section 8.3 (see Condition 8.7).

Lemma 8.18. *Let $\{y(n)\}_{n\geq 1}$, $\{q(n)\}_{n\geq 1}$, and $\{\hat{q}^1(n), \ldots, \hat{q}^N(n)\}_{n\geq 1}$ be generated according to Algorithm 8.11. Suppose Assumptions 8.3–8.6, 8.14, and Condition 8.17 hold. Assume $n \geq n_\eta$. Then there exists a finite $T \in \mathbb{N}$ such that if any action $y^* \in Y$ is repeatedly played in $\tilde{T} \geq T$ consecutive stages, i.e., $y(s) = y^*$ for $s = n, \ldots, n + \tilde{T} - 1$, then $\arg\max_{y_i \in Y_i} U_i(y_i, \hat{q}^i_{-i}(n + \tilde{T}), \mu^i_{n+\tilde{T}}) = \arg\max_{y_i \in Y_i} U_i(y_i, y^*_{-i}, \mu)$ for all i.*

The proof of Lemma 8.18 is found in the appendix of this chapter. The preceding lemma ensures that the N-FP process satisfies Condition 8.7. This allows us to invoke Theorem 8.10 in order to prove the following theorem, which is the main convergence result for N-FP.

Theorem 8.19. *Let $\{y(n)\}_{n\geq 1}$, $\{q(n)\}_{n\geq 1}$, and $\{\hat{q}^1(n), \ldots, \hat{q}^N(n)\}_{n\geq 1}$ be generated according to Algorithm 8.11. Suppose Assumptions 8.3–8.6, 8.14 and Condition 8.17 hold. Then the action sequence $\{y(n)\}_{n\geq 1}$ converges to a pure-strategy Nash equilibrium, almost surely.*

Proof. The N-FP process fits the template of Condition 8.7 with $\hat{U}_i(y_i, n) = U_i(y_i, \hat{q}_{-i}(n), \mu^i_n)$ for each i and each $\alpha_i \in Y_i$. By Lemma 8.18, the sequence of predictions $\{U_i(y_i, \hat{q}_{-i}(n), \mu^i_n)\}_{n\geq 1}$ satisfies Condition 8.7. Thus, the result follows from Theorem 8.10. $\square$

In the following section, we present an example implementation of N-FP in which the inter-agent communication scheme is explicitly specified. We demonstrate that the example implementation converges to a pure-strategy NE.

8.4.5 An Implementation of N-FP

In this section we provide an example implementation of N-FP in which the mechanism for forming the estimates $\hat{q}^i(n)$ and the network-based information dissemination scheme are explicitly specified.

Algorithm 8.20.

Initialize

(i) Let $\rho \in (0,1)$ be fixed. For each i, let $y_i(1)$ be chosen arbitrarily. Let $q_i(1) = \mathbf{1}_{y_i(1)}$ for all i and let $\hat{q}_i(1)^i = q_i(1)$. For $j \neq i$, let $\hat{q}_j^i(1) = w_{j,j}^i q_j(1)$ if $j \in \mathcal{N}_i$ and $\hat{q}_j^i(1) = 0$ otherwise, where $w_{j,j}^i$ is a weight constant (see step (iv) and Lemma 8.21).

Iterate $(n \geq 1)$

(ii) Each agent i chooses their next-stage action according the rule

$$y_i(n+1) = \begin{cases} \arg\max_{y_i \in Y_i} U_i(y_i, \hat{q}_{-i}^i(n), \mu_n^i) & \text{with prob. } 1-\rho, \\ y_i(n) & \text{with prob. } \rho. \end{cases}$$

(iii) Each player i updates their personal empirical distribution $q_i(n)$ according to (8.8).

(iv) For each i, $\hat{q}_j^i(n)$ is updated as

$$\hat{q}_j^i(n+1) = \sum_{k \in \mathcal{N}_i} w_{j,k}^i \left(\hat{q}_j^k(n) + (q_j(n+1) - q_j(n))\, \chi_{\{k=j\}} \right), \tag{8.9}$$

where $\chi_{\{k=j\}}$ is the characteristic function defined by $\chi_{\{k=j\}} = 1$ if $k = j$ and $\chi_{\{k=j\}} = 0$ otherwise, and where $w_{j,k}^i$ is the weight that player i attributes to k's estimate of j's empirical frequency (see Lemma 8.21.)

8.4.6 An Implementation of N-FP: Convergence Analysis

In this section we analyze the convergence of the example implementation of N-FP presented in Section 8.4.5.

The following Lemma ensures that Condition 8.17 is satisfied. It then follows by Theorem 8.19 that the example implementation of N-FP converges to a pure-strategy NE.

Lemma 8.21. *Let $\{y(n)\}_{n\geq 1}$, $\{q(n)\}_{n\geq 1}$, and $\{\hat{q}^1(n), \ldots, \hat{q}^N(n)\}_{n\geq 1}$ be generated according to Algorithm 8.20. Suppose Assumptions 1.1–1.3 hold. Let $\mathbf{W}_j \in R^{N\times N}$, $j \in \mathcal{N}$ be a* weight matrix *with the i,k-th entry given by $\mathbf{W}_j(i,k) = w_{j,k}^i$. Assume that the matrix $\mathbf{W}_j$ is row stochastic with sparsity conforming to the communication network G. Assume the j-th diagonal entry satisfies $w_{j,j}^j = 1$ for each $j \in \mathcal{N}$. Let P_j be the matrix obtained by removing the j-th row and column from*

$\mathbf{W}_j$. *Assume P_j is irreducible and substochastic in the sense that at least one row sum of P_j is strictly less than* 1.

Then, for any $\epsilon > 0$ there exists $T \in \mathbb{N}$ such that if players repeat any action $y^* \in Y$ for $\tilde{T} \geq T$ consecutive stages (i.e., $y(s) = y^*,\ s = n, \ldots, n + \tilde{T} - 1$) then $\|\hat{q}^i(n + \tilde{T} - 1) - q(n + \tilde{T} - 1)\| < \epsilon$.

The proof of Lemma 8.21 is found in the appendix. The size of the T term in the above lemma depends on the step size parameter α in (8.8), network structure G, and the weight matrices $\{\mathbf{W}_j\}_{j\in\mathcal{N}}$. In general, we expect T to be small if the step size is close to 1, $0 \ll \alpha \leq 1$, and the diameter of the network is small. For instance, in a complete network with $\alpha = 1$ and weights $w^i_{k,k} = 1$ if $k \in \mathcal{N}_i$ we have $T = 1$ time steps. In a star network with $\alpha = 1$ such that the weights associated with the center player i satisfy $w^i_{k,k} = 1$ if $k \in \mathcal{N}_i$ and the weights associated with fringe players satisfy $j \neq i$ $w^j_{k,i} = 1$ for $k \neq j$, then we have $T = 2$ time steps.

In the N-FP implementation presented here, each player i maintains an estimate of the empirical distribution of *every* individual player, and exchanges these estimates with its neighbors $\{\hat{q}^i_j(n)\}_{j\in\mathcal{N}}$. This means that at each time step players exchange $\sum_{j=1}^N |Y_j|$ values, and in general, track $\sum_{j=1}^N |Y_j|$ values.

In the following section we study a network-based inertial MBR algorithm that can alleviate this communication and memory burden within the class of games known as congestion games.

8.5 Network-Based Implementation of Joint Strategy Fictitious Play with Inertia

Joint Strategy FP (JSFP) with inertia, introduced in [73], is a variant of FP developed for large-scale games that has relatively low computational complexity and low information overhead requirements. In this section we study a network-based implementation of JSFP with inertia (referred to hereafter as N-JSFP) for use in settings with environmental uncertainty. The algorithms in this section may be seen as instances of inertial MBR dynamics.[4]

The variant of JSFP that we study is applicable within the class of congestion games—a subset of the more general class of weakly-acyclic games (see Assumption 8.4). This restriction comes as a consequence of the aggregative information that is shared over the communication graph. Thus, while N-JSFP operates with lower complexity and communication overhead than N-FP (Section 8.4), N-JSFP is applicable within a narrower class of games than N-FP.

[4]A related variant of JSFP—termed Average Strategy FP (ASFP) is studied in [102]. However, ASFP differs fundamentally from N-JSFP in that (i) ASFP assumes instantaneous and perfect information dissemination by an oracle, (ii) N-JSFP uses a projection operation to make sense of the notion of players “assuming” that the average congestion profile represents choices taken by agents, and (iii) N-JSFP considers learning under environmental uncertainty.

We begin by introducing the class of congestion games in Section 8.5.1. We then introduce the general N-JSFP algorithm in Sections 8.5.2–8.5.3. The general N-JSFP algorithm may be implemented using a wide range of inter-agent communication schemes. We provide sufficient conditions for the convergence of the general N-JSFP algorithm in Secion 8.5.4, and prove convergence in the same section. Finally, in Sections 8.5.5–8.5.6 we provide an explicit example implementation of N-JSFP in which the inter-agent communication dynamics are explicitly defined, and prove convergence of the algorithm.

8.5.1 Congestion Games

Let $R = \{1, \ldots, m\}$ denote a set of resources. For each $i \in \mathcal{N}$, let $Y_i \subseteq 2^R$, where 2^R denotes the power set of R. In particular, an action choice y_i indicates a subset of resources being utilized by player i.

In a congestion game, the cost associated with using a resource is dependent on the total number of players using the same resource. For each $r \in R$, $y \in Y$, let $\psi_r(y) \in \{0, 1, 2, \ldots\}$ denote the number of players using resource r under the action profile y. More generally, for a subset of players $\mathcal{K} \subseteq \mathcal{N}$, the number of players in $\mathcal{K}$ utilizing resource r given $\{y_j\}_{j\in\mathcal{K}}$, is given by

$$\psi_r(\{y_j\}_{j\in\mathcal{K}}) := \sum_{j\in\mathcal{K}} \mathbf{1}(r \in y_j).$$

where $\mathbf{1}(r \in y_j) = 1$ if $r \in y_j$ and $\mathbf{1}(r \in y_j) = 0$ otherwise. Let

$$N_0 := \{0, 1, 2, \ldots\}$$

denote the set of natural numbers including zero. Given a subset of players $\mathcal{K} \subset \mathcal{N}$, and a corresponding set of actions $\{y_j\}_{j\in\mathcal{K}}$, we represent the number of players using each resource by $\psi(\{y_j\}_{j\in\mathcal{K}})$, where $\psi : \prod_{j\in\mathcal{K}} A_j \to \mathbb{N}_0^m$ is a mapping with the r-th entry in $\psi(\{y_j\}_{j\in\mathcal{K}})$ given by $\psi_r(\{y_j\}_{j\in\mathcal{K}})$.

For $r \in R$ and $k \in \mathbb{N}_0$, let $c_r(k, \theta_r)$ be the cost associated with using resource r given state θ_r, when there are precisely k players simultaneously using the resource. For $y_i \in Y_i$ and $\psi_r(y_{-i}) \in \mathbb{N}_0$ and state parameter $\theta := \{\theta_r\}_{r\in R} \in \Theta$, let the utility of player i be given by

$$U_i(y_i, y_{-i}, \theta) = -\sum_{r\in y_i} c_r(\psi_r(y_{-i}) + \psi_r(y_i), \theta_r)$$

where we have written $\psi_r(y) = \psi_r(y_{-i}) + \psi_r(y_i)$ explicitly to emphasize dependence of the utility on "self action" y_i and actions of other players y_{-i}. Note that within the class of congestion games, players do not need to precisely know the full action profile $y = (y_1, \ldots, y_N) \in Y$ to compute their utility. It is sufficient for each player to have knowledge of $\psi(y_{-i}) \in \mathbb{N}_0^m$ and their own action $y_i \in Y_i$. In this context, we sometimes express the utility function using the abuse of notation $U_i(y_i, \psi(y_{-i}), \theta) := U_i(y_i, y_{-i}, \theta)$.

In the following, we use this property of the utility functions in congestion games to design the N-JSFP algorithm which has a lower communication overhead than N-FP.

8.5.2 N-JSFP Setup

Assume players repeatedly face off in a congestion game. We define $\zeta_i(n,r)$ to be a (fading-memory) weighted average used to track the amount of congestion induced on resource r by the actions of (only) player i. In particular, let $\zeta_i(n,r)$ be defined recursively by $\zeta_{i,1}(r) := \psi_r(y_i(1))$, and for $n \geq 1$,

$$\zeta_i(n+1,r) := (1-\alpha)\zeta_i(n,r) + \alpha\psi_r(y_i(n)), \tag{8.10}$$

where $\alpha \in (0,1]$ is a weight parameter inducing a fading-memory effect (cf. (8.8) and subsequent discussion).

Furthermore, define $\zeta_i(n) \in \mathbb{R}^m$ to be the vector stacking $(\zeta_i(n,r))_{r\in R}$—this is a vectorized representation of the congestion induced by player i on any given resource.

Define $\zeta(n,r) := \sum_{j\in\mathcal{N}} \zeta_j(n,r)$—this represents the congestion induced on resource r by the actions of *all* players. Note this can also be expressed recursively as $\zeta(n,r) = (1-\alpha)\zeta(n,r) + \alpha\psi_r(y(n))$.

Similar to the above, let $\zeta(n)$ be a vector in $\mathbb{R}^m$ stacking $(\zeta(n,r))_{r\in R}$—this is a vectorized representation of the congestion induced by *all* players on any given resource. We refer to $\zeta(n)$ as the empirical congestion distribution.

In the network-based framework, players may not have precise knowledge of $\zeta(n)$. Instead, we assume each player i maintains an estimate of $\zeta(n)$ which we denote by $\hat{\zeta}^i(n) \in \mathbb{R}^m$. The r-th term of player i's estimate, $\hat{\zeta}^i(n,r)$, represents her estimate of the congestion at resource $r \in R$.

Finally, in order to rigorously define N-JSFP, we require the following notion of a projection. For a vector $v \in \mathbb{R}^m$ define $P(v)$ to be a projection of v onto the set of non-negative m-dimensional integer-valued vectors $\mathbb{N}_0^m$; formally, for $1 \leq r \leq m$, let $P(v)(r) := z$ for the unique $z \in \mathbb{N}_0$ satisfying $z - \frac{1}{2} \leq v(r) < z + \frac{1}{2}$. Let $P(v)$ be the vector stacking $\{P(v)(r)\}_{r\in R}$.

8.5.3 General N-JSFP Algorithm

We now state the general N-JSFP algorithm. The algorithm follows the format of an inertial MBR algorithm, but in this case, the utility prediction of player i $\hat{U}_i(n) = (\hat{U}_i(y_i, n))_{y_i \in Y_i}$ is formed using player i's estimate of the empirical congestion distribution $\hat{\zeta}^i(n)$. The algorithm is formally stated below.

Algorithm 8.22.

initialize

(i) Let $\rho \in (0, 1)$ be the inertia probability. Let each player i choose an arbitrary initial action $y_i(1)$. Initialize the estimate $\hat{\zeta}^i(1)$ and belief on state μ_1^i.

iterate $(n \geq 1)$

(ii) Let $\hat{\zeta}^i_{-i}(n) := \hat{\zeta}^i(n) - \zeta_i(n)$. For each player i, the next-stage action is chosen according to the rule

$$y_i(n+1) \in \begin{cases} \arg\max_{y_i \in Y_i} U_i(y_i, P(\hat{\zeta}^i_{-i}(n)), \mu_n^i), & \text{w.p. } 1-\rho \\ y_i(n), & \text{w.p. } \rho \end{cases}$$

(iii) Each player i updates their state belief μ_n^i, exchanges information with neighboring agents $j \in \mathcal{N}_i$ and updates their estimate $\hat{\zeta}^i(n)$.

Remark 8.23. The N-JSFP algorithm is closely related to the JSFP algorithm proposed in [73]. In particular, the vector representing the weighted congestion at each resource $\zeta(n) := (\zeta(n, r))_{r \in \mathcal{N}}$ is closely related to the empirical frequency of the joint action profile used in the JSFP algorithm. In this case, the vectors $\zeta(n)$ and $\zeta_i(n)$ together may be thought of as a sufficient statistic to compute the predicted utility. We note that in N-JSFP, the complexity problem associated with FP in large games is mitigated in a different manner than classical JSFP. In classical JSFP, the expected utility is directly tracked using a simple recursion. In our formulation of JSFP, the utility is not directly tracked; instead, players track some "sufficient statistic" $\zeta_{-i}(n) = \zeta(n) - \zeta_i(n)$ and only evaluate the utility at nearby pure strategies via the projection operation, $U_i(y_i, P(\zeta(n) - \zeta_i(n)), \mu_n^i)$.

8.5.4 Convergence Analysis for General N-JSFP Algorithm

The general form described in Section 8.5.3 explicitly specifies the learning dynamics to be used, but does not explicitly specify the dynamics of the inter-agent communication scheme. In the following we show that any inter-agent communication scheme will lead to convergence so long as it satisfies the following condition (cf. Condition 8.17 for N-FP). The main result of this section is Theorem 8.27.

Condition 8.24. *There exists a $T \geq 1$ such that if any action y^* is repeated in $\tilde{T} \geq T$ consecutive stages then $|\hat{\zeta}^i(n+\tilde{T}, r) - \psi_r(y^*)| < \frac{1}{4}$ for every $r \in R$.*

Given Condition 8.24, we show in the following lemma that if any action y^* is repeated in sufficiently many stages, then each player's estimate $\hat{\zeta}^i_{-i}(n)$ may be brought sufficiently close to the congestion profile $\psi(y^*_{-i})$ to ensure convergence.

Lemma 8.25. *Let $\{y(n)\}_{n\geq 1}$ and $\{\hat{\zeta}^1(n), \ldots, \hat{\zeta}^N(n)\}_{n\geq 1}$ be generated according to a N-JSFP process as defined in Section 8.5.3, and let $\{\zeta_i(n)\}_{i\in\mathcal{N}, n\geq 1}$ be as defined in* (8.10)*. Assume Condition 8.24 holds. There exists a $T \geq 1$ such that if any action y^* is repeated in $\tilde{T} \geq T$ consecutive stages then $|\hat{\zeta}^i_{-i}(n+\tilde{T}, r) - \psi_r(y^*_{-i})| < \frac{1}{2}$ for every $r \in R$.*

Proof. Let $\tilde{T} \geq 0$ and note that

$$\begin{aligned} &|\hat{\zeta}^i_{-i}(n+\tilde{T}, r) - \psi_r(y^*_{-i})| \\ &= |(\hat{\zeta}^i(n+\tilde{T}, r) - \zeta_i(n+\tilde{T}, r)) - (\psi_r(y^*) - \psi_r(y^*_i))| \\ &\leq |\hat{\zeta}^i(n+\tilde{T}, r) - \psi_r(y^*)| + |\zeta_i(n+\tilde{T}, r) - \psi_r(y^*_i)|. \end{aligned} \tag{8.11}$$

By Condition 8.24, we may choose T' such that if y^* is repeated in $\tilde{T} \geq T'$ consecutive stages, there holds $|\hat{\zeta}^i(n+\tilde{T}, r) - \psi_r(y^*)| < \frac{1}{4}$. Note also that $|\zeta_i(n+\tilde{T}, r) - \psi_r(y^*_i)| \to 0$ as $\tilde{T} \to \infty$ (this follows from (8.10)), and thus we may choose T'' such that for $\tilde{T} \geq T''$ there holds $|\zeta_i(n+\tilde{T}, r) - \psi_r(y^*_i)| < \frac{1}{4}$. Letting $T = \max\{T', T''\}$, the desired result follows from (8.11). $\square$

The next lemma sets us up to prove convergence of N-JSFP using Theorem 8.10 and Condition 8.7.

Lemma 8.26. *Let $\{y(n)\}_{n\geq 1}$ and $\{\hat{\zeta}^1(n), \ldots, \hat{\zeta}^N(n)\}_{n\geq 1}$ be generated according to a N-JSFP process as defined in Section 8.5.3, and let $\{\zeta_i(n)\}_{i\in\mathcal{N}, n\geq 1}$ be as defined in* (8.10)*. Suppose Assumptions 1.1–1.3, 8.3–8.6, and Condition 8.24 hold. There exists a $T \geq 1$ such that if any action y^* is repeated in $\tilde{T} \geq T$ consecutive stages then* $\arg\max_{y_i \in Y_i} U_i(y_i, P(\hat{\zeta}^i_{-i}(n+\tilde{T})), \mu^i_{n+\tilde{T}}) = \arg\max_{y_i \in Y_i} U_i(y_i, \psi(y^*_{-i}), \mu)$.

Proof. Let T' be chosen as in Lemma 8.25 so that $|\hat{\zeta}^i_{-i}(n+\tilde{T}, r) - \psi_r(y^*_{-i})| < \frac{1}{2}$ for every $r \in R$, $i \in \mathcal{N}$ and all $\tilde{T} \geq T'$. It follows that $P(\hat{\zeta}^i_{-i}(n+\tilde{T})) = \psi(y^*_{-i})$. Thus,

$$\arg\max_{y_i \in Y_i} U_i(y_i, P(\hat{\zeta}^i_{-i}(n+\tilde{T})), \mu^i_{n+\tilde{T}}) = \arg\max_{y_i \in Y_i} U_i(y_i, \psi(y^*_{-i}), \mu^i_{n+\tilde{T}}).$$

Moreover, there exists a $\bar{n}$ such that for $n \geq \bar{n}$ there holds

$$\arg\max_{y_i \in Y_i} U_i(y_i, y_{-i}, \mu^i_{n+\tilde{T}}) = \arg\max_{y_i \in Y_i} U_i(y_i, y_{-i}, \mu)$$

for all $y_{-i} \in Y_{-i}$. This is a consequence of Assumption 8.3 and the fact that U_i is continuous and bounded, and that Y_{-i} is a finite set. By setting $T = \max\{\bar{n}, T'\}$ we get the desired result. $\square$

In the following theorem we combine the results above with Theorem 8.10 to prove that the N-JSFP algorithm converges to a pure-strategy Nash equilibrium.

Theorem 8.27. *Let* $\{y(n)\}_{n\geq 1}$ *and* $\{\hat{\zeta}^1(n), \ldots, \zeta^N(n)\}_{n\geq 1}$ *be generated according to a N-JSFP process as defined in Section 8.4.3, and let* $\{\zeta(n)\}_{n\geq 1}$ *be as defined in* (8.10). *Suppose Assumptions 1.1–1.3, 8.3–8.6, and Condition 8.24 hold. Then the action sequence* $\{y(n)\}_{n\geq 1}$ *converges to a pure strategy Nash equilibrium, almost surely.*

Proof. The process fits the template of Theorem 8.10, with $U_i(y_i, P(\hat{\zeta}^i_{-i}(n)), \mu^i_n) = \hat{U}_i(y_i, n)$ for each i and each $\alpha_i \in Y_i$. By Lemma 8.26, the sequence $\{U_i(y_i, P(\zeta^i_{-i}(n)), \mu^i_n)\}_{n\geq 1}$ satisfies Condition 8.7. The result then follows from Theorem 8.10. $\square$

8.5.5 An Implementation of N-JSFP

In this section we provide an example implementation of N-JSFP in which the mechanism for forming the estimate $\hat{\zeta}^i(n)$, and the network-based information dissemination scheme are explicitly specified. The algorithm is as follows:

Algorithm 8.28.
initialize
(i) Let $y_i(1)$ be arbitrary for all i. Let $\hat{\zeta}^i(1) = \psi(y_i(1))$ for all i.

iterate $(n \geq 1)$
(ii) Let $\hat{\zeta}^i_{-i}(n) = \hat{\zeta}^i(n) - \zeta_i(n)$. For each player i, the next-stage action is chosen according to the rule

$$y_i(n+1) \in \begin{cases} \arg\max_{y_i \in Y_i} U_i(y_i, P(\hat{\zeta}^i_{-i}(n)), \mu^i_n), & \text{w.p. } 1-\rho \\ y_i(n), & \text{w.p. } \rho. \end{cases}$$

(iii) Update $\zeta_i(n+1)$ according to (8.10).

(iv) Each player i updates their estimate of $\zeta(n)$ as:

$$\hat{\zeta}^i(n+1) := \sum_{k\in\mathcal{N}_i} w_k^i \left(\hat{\zeta}^k(n) + \zeta_k(n+1) - \zeta_k(n)\right)$$

where w_k^i denotes the weight that player i places on the information received from player k (see Lemma 8.29).

We remark that in N-JSFP players only share a vector with m integer values with their neighbors where m is the cardinality of the set of resources R. In comparison, in the N-FP algorithm players share their estimate of each agent's empirical frequency implying that they share $N \times m$ values at each step.

8.5.6 Convergence Analysis of Example Implementation of N-JSFP

The following lemma shows that the distributed information tracking scheme used to form the estimates $\hat{\zeta}^i(n)$ (see step (iii) in Section 8.5.5) satisfies Condition 8.24 given an appropriate assumption on players' weights w_k^i. It then follows by Theorem 8.27 that the given implementation of N-JSFP converges to a pure strategy Nash equilibrium.

Lemma 8.29. *Let $\{y(n)\}_{n\geq 1}$ and $\{\hat{\zeta}^1(n),\ldots,\zeta^N(n)\}_{n\geq 1}$ be generated according to a N-JSFP process as defined in Section 8.5.5, and let $\{\zeta(n)\}_{n\geq 1}$ be as defined in* (8.10)*. Suppose Assumptions 1.1–1.3 and 8.3–8.6 hold. Let $\mathbf{W} \in R^{N\times N}$ be a* weight matrix *with the i,k-th entry given by $\mathbf{W}(i,k) = w_k^i$. Assume that the matrix $\mathbf{W}$ is doubly stochastic, aperiodic, and irreducible.*

There exists a $\tilde{T} \geq 1$ such that if any action y^* is repeated in $\tilde{T} \geq T$ consecutive stages then $|\hat{\zeta}^i_{n+\tilde{T}}(r) - \psi_r(y^*)| < \frac{1}{4}$ for every $r \in R$.

The proof of Lemma 8.29 is similar to the proof of Lemma 8.21, and is omitted here for brevity.

8.6 Numerical examples

8.6.1 Traffic Congestion Game

We consider the congestion game defined in Section 8.5.1. We assume each agent can only pick a single route at each time $Y = R$. The cost of route $r \in R$ is quadratic in the number of players on that route,

$$c_r(N_r(y)) = \phi_r N_r(y)^2 + \beta_r N_r(y) + \kappa_r$$

where $\phi_r, \beta_r, \kappa_r$ are route specific constants which play the role of the unknown state of the world.

In the numerical setup, we let $\phi_r = 1/(1 + 2(r - 1))$, $\beta_r = 1$ and $\kappa_r = 1$ for each route. Agents implement the N-JSFP algorithm defined in Section 8.5.5. We let the fading constant α and the inertia probability ρ be 0.1 and 0.3, respectively. Players are uncertain about κ_r and ϕ_r. Each player receives a single initial signal on κ_r and ϕ_r for each $r \in R$. The signals on κ_r come from the uniform distribution in $[0.75, 1.25]$. The signals on ϕ_r come from the uniform distribution in $[\phi_r - 0.25, \phi_r + 0.25]$. Following each round of game play, players exchange their mean estimates of the state parameters κ_r and ϕ_r with neighbors, and form new estimates using a weighted average consensus recursion.

In this simple example, the μ_n^i are deterministic delta distributions, as is the limit distribution μ. As noted in Section 8.3, this achieves weak convergence of the beliefs, thus satisfying Assumption 8.3. Moreover, Assumption 8.3 could also be satisfied in an almost sure sense by analogously generating a sequence of *random* delta distributions using consensus recursions that ensure almost sure convergence of the estimates [21, 146, 147].

There are $N = 10$ agents and $m = 3$ routes. The connection among agents is determined by a geometric network. Agents are randomly placed on a two dimensional space (5 units $\times$ 5 units) and two agents are connected if their distance is less than a given threshold value (2 units). In Fig. 8.1 we plot the number of vehicles on each route. In Fig. 8.2 we plot the congestion cost at each road and total congestion cost over time. We initialize all agents' actions to be route 1. We observe that at equilibrium there are four agents on roads 2 and 3, and two agents on road 1. Total congestion cost diminishes significantly in the first few time steps—see Fig. 8.2. At equilibrium we observe that congestion cost at road 2 is slightly higher than the other roads.

Next, we analyze convergence times with respect to initial level of uncertainty on the state parameters κ_r and ϕ_r, and network connectivity. In particular we use distance thresholds 2, 2.5, 3 and 3.5 to generate four geometric networks with diameters 4, 3, 2 and 2 and with average lengths 2, 1.3, 1.2 and 1.1, respectively. For each network, we make 50 runs with and without uncertainty on the state parameters. The average convergence time of the N-JSFP algorithm is reported in Table 8.1. We observe convergence time tends to decrease as the connection threshold increases. We do not see a significant difference between the convergence times when there is uncertainty or not about the state parameters.

8.6.2 Target Assignment Game

We consider a team of N robots, and N target objects. Each robot can target one object and the team's goal is to target all of the objects. The action space is the set of objects $\{1, 2 \ldots, N\}$ for each robot. The payoff of a robot i targeting object k is inversely proportional to its distance to

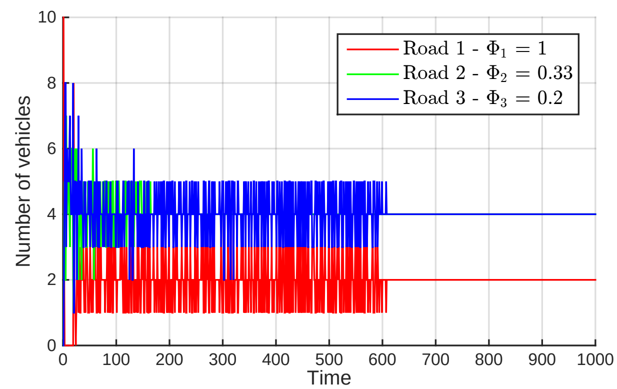


Figure 8.1: Number of vehicles on each route. There are $N = 10$ vehicles each day choosing among $m = 3$ possible routes. Convergence to pure Nash equilibrium happens after time 600.

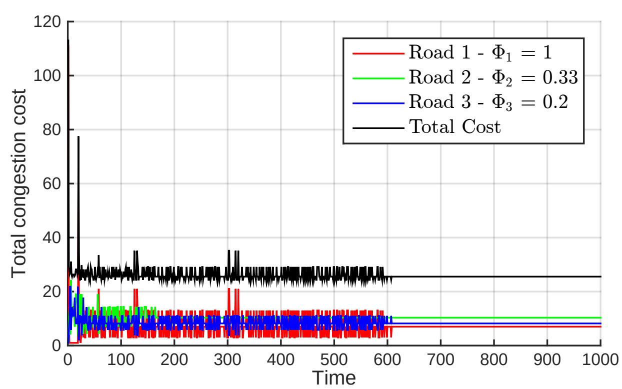


Figure 8.2: Congestion cost at each road and total congestion cost $\sum_r c_r$ at each time. Initial total cost dampens quickly in the first few time steps.

	Connection threshold (units)			
	2	2.5	3	3.5
Certain	361	172	220	104
Uncertain	329	179	212	95

Table 8.1: Average convergence time of the N-JSFP algorithm

the object, represented by $d(i,k)$, if no other robot is targeting object k,

$$U_i(y_i = k, y_{-i}, d(i,k)) = d(i,k)^{-1}\mathbf{1}(\sum_{j=1}^{N}\mathbf{1}(y_j = k) = 1)$$

where $\mathbf{1}(\cdot)$ is the indicator function. The target assignment game with payoffs as above is a congestion game with each object representing a resource. Note that any action profile that covers all the objects is a Nash equilibrium because any unilateral deviation from such profile results in zero payoff for the deviating agent. The optimal Nash equilibrium profile minimizes the total distance while covering all the objects.

We assume robots do not exactly know their distance to each target. Robots receive an initial noisy signal about the location of each object $\tilde{\theta}_i(k)$ distributed according to $\mathcal{N}(\theta(k), \xi_k)$ where $\theta(k)$ is the location of the object and $\xi_k > 0$ is the variance of the noisy signal. Robot i updates its belief about each object by averaging its neighbors' beliefs with initial belief equal to $\tilde{\theta}_i(k)$. The distance to an object is the norm of the difference between the locations of the robot and the object.

In the numerical setup, we consider $N = 5$ robots and $N = 5$ objects. For comparison, we consider line, ring and star communication networks. For each network, we consider 50 runs with random initialization of the signals. In Figure 8.3, we plot sample average welfare normalized by the optimal welfare over time for each network type. Welfare at time N is defined as the sum of utilities of robots given the action profile generated by the N-JSFP process at time N. The expected welfare at time N is the average of welfare values at time N obtained over 50 runs. Optimal welfare is the value of welfare obtained by the action profile that maximizes the welfare. Fig. 8.3 shows that the expected welfare of the Nash equilibrium reached by the algorithm is the same regardless of the communication network. However, the convergence time of the algorithm depends on the network structure where the star network has the slowest convergence time and the ring network has the fastest convergence time.

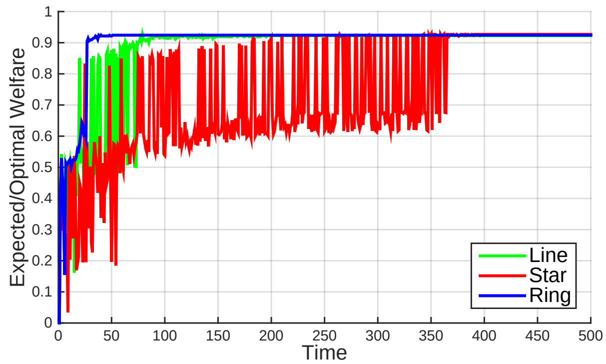


Figure 8.3: Expected welfare normalized by optimal welfare for line, star and ring networks. The expected welfare is the same for all communication networks. Convergence time to a Nash equilibrium is the fastest for the ring network and slowest for the star network.

8.7 Appendix to Chapter 8

The following lemma studies a uniform convergence property for families of equicontiuous functions. The lemma is applied in the text to prove that given weak convergence of beliefs (see Assumption 8.3) and equicontinuity of utilities (see Assumption 8.14), Lemma 8.15 holds.

Lemma 8.30. *Let Θ be as defined in Section 8.2. Let X be a compact subset of a finite-dimensional Euclidean space. Let $g : X \times \Theta \to \mathbb{R}$ be bounded. Assume that for each $x \in X$ the function $g(x, \cdot) : \Theta \to \mathbb{R}$ is Borel measurable. Furthermore, assume that the family of functions $\{g(\cdot, \theta)\}_{\theta\in\Theta}$ is uniformly equicontinuous. Let $(\mu_n)_{n\geq 1}$ and μ be probability measures on $(\Theta, \mathcal{B}(\Theta))$ such that $\mu_n \xrightarrow{w} \mu$. For any $\epsilon > 0$ there exists $T \in \{0, 1, 2, \ldots\}$ such that for all $n \geq T$ there holds*

$$|\int_\Theta g(x, \theta)\mu_n(d\theta) - \int_\Theta g(x, \theta)\mu(d\theta)| < \epsilon, \;\; \forall x \in X.$$

Proof. Let $\epsilon > 0$. Since the family of functions $\{g(\cdot, \theta)\}_{\theta\in\Theta}$ is equicontinuous, there exists a $\delta > 0$ such that if $x, y \in X$ and $\|x - y\| < \delta$ then $|g(x, \theta) - g(y, \theta)| < \frac{\epsilon}{3}$ for any $\theta \in \Theta$. Since X is a compact subset of Euclidean space we may choose a finite set $\{x_n\}_{n=1}^N$ such that $\bigcup_{n=1}^N B_\delta(x_n) \supset X$, where $B_\delta(x_0)$ denotes the open ball of radius δ centered at x_0.

Since μ_n converges weakly to μ, and the set $\{x_n\}_{n=1}^N$ is finite, and the function $g(x', \cdot)$ is measurable for each $x' \in \{x_n\}_{n=1}^N$, there exists a $T \in \{0, 1, 2, \ldots\}$ such that

$$|\int g(x', \theta)\mu_n(d\theta) - \int g(x', \theta)\mu(d\theta)| < \frac{\epsilon}{3}$$

for all $x' \in \{x_n\}_{n=1}^N$.

Choose an arbitrary $x \in X$. There exists a $\tilde{x} \in \{x_n\}_{n=1}^N$ such that $x \in B_\delta(\tilde{x})$. In particular, note that

$$|\int [g(x,\theta) - g(\tilde{x},\theta)]\mu_n(d\theta)| < \frac{\epsilon}{3}\int \mu_n(d\theta) = \frac{\epsilon}{3},$$

and likewise,

$$|\int [g(x,\theta) - g(\tilde{x},\theta)]\mu(d\theta)| < \frac{\epsilon}{3}\int \mu(d\theta) = \frac{\epsilon}{3}.$$

Using the triangle inequality and applying the estimates derived above gives,

$$\begin{aligned}
&|\int g(x,\theta)\mu_n(d\theta) - g(x,\theta)\mu(d\theta)| \\
&\leq |\int [g(x,\theta) - g(\tilde{x},\theta)]\mu_n(d\theta)| \\
&+ |\int g(\tilde{x},\theta)\mu_n(d\theta) - \int g(\tilde{x},\theta)\mu(d\theta)| \\
&+ |\int [g(\tilde{x},\theta) - g(x,\theta)]\mu(d\theta)| < \epsilon,
\end{aligned}$$

for all $n \geq T$. $\square$

The following is a proof of Lemma 8.18 in Section 8.4.4.

(Lemma 8.18). We first show that for every $\xi > 0$, there exists a finite $T' \in \mathbb{N}$ such that if any action $y^* \in Y$ is repeatedly played in $T \geq T'$ consecutive stages, i.e., $y(s) = y^*$ for $s = n, \dots, n+T$, then $\|\hat{q}^i(n+T) - \mathbf{1}_{y^*}\| < \xi, \ \forall i$.
Let $\xi > 0$ be given. By the triangle inequality, $\|\hat{q}^i(n) - \mathbf{1}_{y^*}\| \leq \|\hat{q}^i(n) - q(n)\| + \|q(n) - \mathbf{1}_{y^*}\|$. Thus it is sufficient to choose T' sufficiently large so that

$$\|\hat{q}^i(n) - q(n)\| < \frac{\xi}{2} \text{ and } \|q(n) - \mathbf{1}_{y^*}\| < \frac{\xi}{2}, \ \forall t \geq T'. \tag{8.12}$$

Claim there exists a finite $T_1' > 0$ such that if any action y^* is repeated in T_1' consecutive stages, then $\|q(n+T_1') - \mathbf{1}_{y^*}\| < \frac{\xi}{2}$. This follows immediately from the observation that, for all i, and for any $\tau \in \{1, 2, \dots\}$ there holds $q_i(n+\tau) = \alpha(1-(1-\alpha)^{\tau-1})\mathbf{1}_{y_i^*} + (1-\alpha)^\tau q_i(n)$.

By Condition 8.17 there exists a T_2' such that if any action y^* is repeated for $T \geq T_2'$ consecutive stages, then $\|\hat{q}^i(n+T) - f(n+T)\| < \frac{\xi}{2}$. Let $T' = \max(T_1', T_2')$ and note that for $n \geq T'$, (8.12) holds.

We now move to prove the statement in the Theorem. Note that under Condition 8.17, there exists a compact C such that for each player i the belief $\hat{q}^i(n)$ belongs to C for all n. For each

i, $U_i(\cdot,\mu)$ (as defined in (8.3)) is multilinear in the first argument. Hence, it is locally Lipschitz continuous in the first argument; i.e., given the belief μ there exists a finite $K_i > 0$ such that for any $p', p'' \in C$ there holds $|U_i(p',\mu) - U_i(p'',\mu)| \leq K_i \|p' - p''\|$. Let $K := \max\{K_i : i \in \mathcal{N}\}$.

Since the action space is finite, Assumption 8.5 implies there exists an $\epsilon > 0$ such that for all $i \in \mathcal{N}$ there holds $|U_i(y_i', y_{-i}, \mu) - U_i(y_i'', y_{-i}, \mu)| > \epsilon$ for all $y_i', y_i'' \in Y_i$, $y_i' \neq y_i''$ $y_{-i} \in Y_{-i}$.

Let $\xi < \frac{\epsilon}{4K}$. Applying the claim demonstrated in the beginning of the proof there exists a $T > 0$ such that if any action $y^* \in Y$ is repeated in $\tilde{T} \geq T$ consecutive stages, then $\|\hat{q}^i(n+T) - \mathbf{1}_{y^*}\| < \xi$, $\forall i$.

Let $\{y_i'\} = \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, y_{-i}^*, \mu)$ (by Remark 8.16, the best response to any pure strategy is a singleton), and note that

$$U_i(y_i', y_{-i}^*, \mu) - \epsilon > U_i(y_i'', y_{-i}^*, \mu) \tag{8.13}$$

for all $y_i'' \in Y_i, y_i'' \neq y_i'$. Furthermore, for $\tau = t + \tilde{T}$, $\tilde{T} \geq T$, there holds by Lipschitz continuity of $U_i(\cdot,\mu)$,

$$|U_i(y_i, \hat{q}_{-i}^i(\tau), \mu) - U_i(y_i, y_{-i}^*, \mu)| \leq K\|\hat{q}^i(\tau) - y^*\| < \frac{\epsilon}{4},$$

for all $y_i \in Y_i$. In particular, for $\{y_i'\} = \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, y_{-i}^*, \mu)$ and for any $y_i'' \in Y_i$ satisfying $y_i'' \neq y_i'$ there holds, $|U_i(y_i', \hat{q}_{-i}^i(\tau), \mu) - U_i(y_i', y_{-i}^*, \mu)| < \frac{\epsilon}{4}$ and $|U_i(y_i'', \hat{q}_{-i}^i(\tau), \mu) - U_i(y_i'', y_{-i}^*, \mu)| < \frac{\epsilon}{4}$. Adding and subtracting $U_i(y_i'', \hat{q}_{-i}^i(\tau), \mu)$ to the right hand side of (8.13) and adding and subtracting $U_i(y_i', \hat{q}_{-i}^i(\tau), \mu)$ to the left hand side of (8.13), the preceding two inequalities give

$$U_i(y_i', \hat{q}_{-i}^i(\tau), \mu) - \frac{\epsilon}{2} > U_i(y_i'', \hat{q}_{-i}^i(\tau), \mu).$$

By Assumption 8.3 and Lemma 8.15, $|U_i(y_i', \hat{q}_{-i}^i(\tau), \mu_\tau^i) - U_i(y_i', \hat{q}_{-i}^i(\tau), \mu)| < \frac{\epsilon}{4}$ and $|U_i(y_i'', \hat{q}_{-i}^i(\tau), \mu) - U_i(y_i'', \hat{q}_{-i}^i(\tau), \mu_\tau^i)| < \frac{\epsilon}{4}$ for all $n \geq t_\eta$; adding and subtracting these terms above gives

$$U_i(y_i', \hat{q}_{-i}^i(\tau), \mu_\tau^i) > U_i(y_i'', \hat{q}_{-i}^i(\tau), \mu_\tau^i).$$

Since $\{y_i'\} = \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, y_{-i}^*, \mu)$ and $y_i'' \neq y_i'$ is arbitrary, this implies that $\arg\max_{\alpha_i \in A_i} U_i(\alpha_i, \hat{q}_{-i}^i(\tau), \mu_\tau^i) = \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, y_{-i}^*, \mu)$ for all i. $\square$

Distributed Averaging in Dynamic Networks

Consider a network of N nodes connected through a communication graph $G = (V, E)$. The graph is assumed to be strongly connected. For $n = \{0, 1, 2, \ldots\}$ let $x_1(n) \in \mathbb{R}$ denote a value held by node 1 at time n. The objective is for all nodes to track as closely as possible the value $x_1(n)$.[5] Let $\epsilon(n) := |x_1(n+1) - x_1(n)|$ and assume that:

Assumption 8.31. *There exists a $B > 0$ such that $\epsilon(n) < B$ for all $n \in \{0, 1, 2, \ldots, \}$.*

Let $\hat{x}_i(n)$ be the estimate player i maintains of $x_1(n)$. We make the following assumption pertaining to the initial error in players' estimates:

Assumption 8.32. *$\hat{x}_i(0) - x_1(0) = 0\ \forall i$.*

Let $\hat{x}(n) = (\hat{x}_1(n), \ldots, \hat{x}_N(n)) \in \mathbb{R}^N$ be a vector stacking all players' estimates, with $\hat{x}_1(n) = x_1(n)$ (i.e., node j knows its own value). Suppose the estimates are updated according to the following recursion:

$$\hat{x}(n+1) = \mathbf{W}\left(\hat{x}(n) + e_1(x_1(n+1) - x_1(n))\right), \tag{8.14}$$

where $e_1 \in \mathbb{R}^N$ is the 1-st canonical vector, and where the matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$ satisfies

Assumption 8.33. *$\mathbf{W}$ is row stochastic with sparsity conforming to G. Furthermore, $\mathbf{W}$ may be decomposed as*

$$\mathbf{W} = \begin{pmatrix} 1 & 0 \\ b & \mathbf{P} \end{pmatrix}$$

where $b \in \mathbb{R}^{N-1}$, $b \neq 0$ and $\mathbf{P}$ is irreducible (cf. [148]).

Note that in since $\mathbf{W}$ is row stochastic, $b \neq 0$ if and only if $\mathbf{P}$ is substochastic in the sense that at least one row sum of $\mathbf{P}$ is less than 1.

The following lemma gives a bound on the error in the agents' estimates of $x_1(n)$.

Lemma 8.34. *Suppose Assumptions 8.31–8.32 hold, and let the sequence $\{\hat{x}(n)\}_{n=1}^{\infty}$ be computed according to* (8.14)*. Suppose there exists a $n^* \geq 1$ and $T \geq 1$ such that $\sum_{n=n^*}^{n^*+T-2} \epsilon(n) \leq 1$ and $\{\epsilon(n)\}_{n=n^*}^{n^*+T-2}$ is decreasing. Then the error at time $n^* + T - 1$ is bounded as,*

$$\|\hat{x}(n^* + T - 1) - x_1(n^* + T - 1)\mathbf{1}\| \leq \frac{N+1}{1-\lambda}\left(\frac{1}{T} + B\lambda^T\right),$$

where $\lambda := \sup_{\|y\|=1} \|\mathbf{P}y\| < 1$.

[5] In general, the objective may be to track the value $x_j(n)$ held by an arbitrary node j. Here, we only consider tracking $x_1(n)$, however, the general case is recovered by a permutation of the node labels.

Proof. Let $y(n) := \hat{x}(n) - x_1(n)\mathbf{1}$. Let $\delta(n) := [x_1(n+1) - x_1(n)]e_1 - [x_1(n+1) - x_1(n)]\mathbf{1}$. Subtracting $x_1(n+1)\mathbf{1}$ from both sides of (8.14) we get

$$\begin{aligned} y(n+1) &= \mathbf{W}\left(\hat{x}(n) + [x_1(n+1) - x_1(n)]e_1\right) - x_1(n+1)\mathbf{1} \\ &= \mathbf{W}\left(\hat{x}(n) + [x_1(n+1) - x_1(n)]e_1 - x_1(n+1)\mathbf{1}\right) \end{aligned}$$

where, in the second line, we may bring $x_1(n+1)\mathbf{1}$ inside the matrix multiplication due to the row stochasticity of $\mathbf{W}$. Now we add and subtract $x_1(n)$ and use the definitions of $y(n)$ and $\delta(n)$ to get $y(n+1) = \mathbf{W}\left(y(n) + \delta(n)\right)$. Inductively, this gives $y(n+1) = \sum_{s=0}^{n} \mathbf{W}^{s+1}\delta(n-s) + \mathbf{W}^{n+1}y(0)$. By Assumption 8.32 we have $y(0) = 0$, and hence $y(n+1) = \sum_{s=0}^{n} \mathbf{W}^{s+1}\delta(n-s)$. By the triangle inequality we have

$$\|y(n+1)\| \leq \sum_{s=0}^{n} \|\mathbf{W}^{s+1}\delta(n-s)\|. \tag{8.15}$$

Again using the triangle inequality, we establish a bound on $\|\delta(n)\|$:

$$\begin{aligned} \|\delta(n)\| &\leq \|[x_1(n+1) - x_1(n)]e_1\| + \|[x_1(n+1) - x_1(n)]\mathbf{1}\| \\ &\leq \epsilon(n) + N\epsilon(n) = (N+1)\epsilon(n). \end{aligned} \tag{8.16}$$

Let $\overline{\mathbf{W}} := \mathbf{W} - \mathbf{1}e_1^T$. In block form we have $\overline{\mathbf{W}} = [0 \ \dots \ 0; \ (b - \mathbf{1}) \ \ \mathbf{P}]$. Due to the special block form of $\overline{\mathbf{W}}$, the spectrum $\sigma(\overline{\mathbf{W}})$[6] of $\overline{\mathbf{W}}$ consists precisely of $\{\sigma(\mathbf{P}) \cup \{0\}\}$. Hence, the spectral radius of $\overline{\mathbf{W}}$ coincides with that of $\mathbf{P}$. In particular, the spectral radius of $\overline{\mathbf{W}}$ is given by λ. Since $\mathbf{P}$ is substochastic, we have $\lambda < 1$.

Substituting $\mathbf{W} = \overline{\mathbf{W}} + \mathbf{1}e_1^T$ in (8.15) gives $\|y(n+1)\| \leq \sum_{s=1}^{n} \|(\overline{\mathbf{W}} + \mathbf{1}e_1^T)^{s+1}\delta(n-s)\|$. Since $\overline{\mathbf{W}}\mathbf{1} = \mathbf{0}$, and $e_1^T\overline{\mathbf{W}} = \mathbf{0}$, and $\mathbf{1}e_1^T = (\mathbf{1}e_1^T)^s$ for any $s = 1, 2, \ldots$, an inductive argument shows that $\mathbf{W}^s = \overline{\mathbf{W}}^s + \mathbf{1}e_1^T$ for any $s \in \{1, 2, \ldots\}$. Thus we can upper bound $\|y(n+1)\|$ using the triangle inequality as follows

$$\|y(n+1)\| \leq \sum_{s=0}^{n} \left(\|\overline{\mathbf{W}}^{s+1}\delta(n-s)\| + \|(\mathbf{1}e_1^T)^{s+1}\delta(n-s)\| \right)$$

It is readily verified that for $s = 1, 2, \ldots$ there holds $e_1^T\delta(s) = 0$. Thus, the second term on the

[6]In Section 8.2 the symbol σ was used to represent a mixed strategy. In keeping with standard conventions, we use σ here to denote the spectrum of a matrix, where the distinction is clear from the context.

right hand side above is zero, i.e., $\|y(n+1)\| \leq \sum_{s=0}^{n} \|\overline{\mathbf{W}}^{s+1}\delta(n-s)\|$. As a result we have, $\|y(n+1)\| \leq \sum_{s=0}^{n} \lambda^{s+1}\|\delta(n-s)\|$.

Let $n = n^* + T - 1$. Using the bound in (8.16) gives,

$$\begin{aligned}\|y(n)\| &\leq \sum_{s=0}^{n-1} \lambda^{s+1}(N+1)\epsilon(n-1-s) \\ &= (N+1)\sum_{s=0}^{n-n^*-1} \lambda^{s+1}\epsilon(n-1-s) + (N+1)\sum_{s=t-n^*}^{n-1} \lambda^{s+1}\epsilon(n-1-s).\end{aligned} \tag{8.17}$$

Consider the first term on the right hand side (RHS) above. Let $\epsilon_{avg}(n^*, T) := \frac{1}{T}\sum_{s=n^*}^{n^*+T-2} \epsilon(s)$. By assumption, the sequence $\{\epsilon(n)\}_{n=n^*}^{n^*+T-2}$ is decreasing, hence by Chebychev's sum inequality [138] (p. 43-44), $\sum\limits_{s=0}^{n-n^*-1} \lambda^{s+1}\epsilon(n-1-s) \leq \epsilon_{avg}(n^*, T)\sum\limits_{s=0}^{n-n^*-1} \lambda^{s+1} \leq \epsilon_{avg}(n^*, T)\frac{1}{1-\lambda}$, where the latter inequality follows by taking the closed form of the geometric sum. Furthermore, by assumption we have $\sum_{s=n^*}^{n^*+T-2} \epsilon(s) \leq 1$, and hence $\epsilon_{avg}(n^*, T) \leq \frac{1}{T}$, which gives that $\sum\limits_{s=0}^{n-n^*-1} \lambda^{s+1}\epsilon(n-s) \leq \frac{1}{T}\frac{1}{1-\lambda}$.

Now consider the second term on the RHS of (8.17). By Assumption 8.31, we have $\epsilon(n-s) \leq B$ which allows us to bound the second term as $(N+1)\sum\limits_{s=t-n^*}^{n-1} \lambda^{s+1}\epsilon(n-1-s) \leq (N+1)B\sum\limits_{s=t-n^*}^{n-1} \lambda^{s+1} = (N+1)B\lambda^T\sum\limits_{s=0}^{n^*-1} \lambda^s \leq (N+1)B\lambda^T(1-\lambda)^{-1}$, where the latter inequality again follows by taking the closed form of the geometric sum.

Substituting these two bounds back into (8.17) we get

$$\begin{aligned}\|y(n)\| &\leq (N+1)\frac{1}{T}\frac{1}{1-\lambda} + (N+1)B\lambda^T\frac{1}{1-\lambda} \\ &= (N+1)\frac{1}{1-\lambda}\left(\frac{1}{T} + B\lambda^T\right).\end{aligned}$$

Since we chose $n = n^* + T - 1$, this concludes the proof. □

In order to apply Lemma 8.34, one must show that $\sum_{n=n^*}^{n^*+T-2} \epsilon(n) < 1$. Essentially, this condition states that the variation in the node value $x_1(n)$ during the designated time interval remains bounded. This can be easily ensured, for example, if the value of $x_1(n)$ is monotone. This is the content of the following Lemma.

Lemma 8.35. *Suppose that $x_1(n) \in [0,1]$ for all n. Suppose also there exist $n^*, T \in \mathbb{N}$ such that $\{x_1(n)\}_{n=n^*}^{n^*+T-1}$ is a monotone sequence. Then $\sum_{n=n^*}^{n^*+T-2} \epsilon(n) \leq 1$.*

Proof. Suppose that $\{x_1(n)\}_{n=n^*}^{n^*+T-1}$ is monotone increasing. Then $\epsilon(n) = |x_1(n+1) - x_1(n)| = x_1(n+1) - x_1(n)$. Substituting this into the sum in question gives a telescoping sum $\sum_{n=n^*}^{n^*+T-2} \epsilon(n) =$

$\sum_{n=n^*}^{n^*+T-2} x_1(n+1) - x_1(n) = x_1(n^*+T-1) - x_1(n^*) \leq 1$. The final inequality follows since $0 \leq x_1(n) \leq 1$ for all n. A similar argument handles the monotone decreasing case. $\square$

We now prove Lemma 8.21 of Section 8.4.6.

Proof. Let $\epsilon > 0$ and let $n^* \in \mathbb{N}$ be arbitrary. Our task is to show that under the update rule (8.9), there exists a T such that if starting at time n^* any action is repeated in $\tilde{T} \geq T$ consecutive stages, then $\|\hat{q}^i(n^* + \tilde{T} - 1) - q(n^* + \tilde{T} - 1)\| < \epsilon$. We will accomplish this by showing that the update rule (8.9) fits the template of Lemma 8.34.

Fix a player $j \in \mathcal{N}$ and action $y_j \in Y_j$. Let $q_j(n, y_j)$ denote the weight that the empirical distribution $q_j(n)$ places on y_j, and similarly, let $\hat{q}_j^i(n, y_j)$ denote the weight that $\hat{q}_j^i(n)$ places on y_j. For the purpose of applying Lemma 8.34, let $x_j(0) = 0$, let $x_j(n) := q_j(n, y_j)$, $n \geq 1$ and for $i = 1, \ldots, n$ let $\hat{x}_i(0) = 0$, and let $\hat{x}_i(n) := \hat{q}_j^i(n, y_j)$ for $n \geq 1$. By (8.9) and the initialization condition for Algorithm 8.20 for $n \geq 0$ we have $\hat{x}_i(n+1) = \sum_{k \in \mathcal{N}_i} w_{j,k}^i \left(\hat{x}_k(n) + (x_j(n+1) - x_j(n))\chi_{\{k=j\}}\right)$. Letting $\hat{x}(n) = (\hat{x}_i(n))_{i=1}^N \in \mathbb{R}^N$ we may express the update rule in more compact notation as

$$\hat{x}(n+1) = \mathbf{W}_j \left(\hat{x}(n) + e_j(x_j(n+1) - x_j(n))\right),$$

where $\mathbf{W}_j = (w_{j,k}^i)_{i,k \in \mathcal{N}}$ is the weight matrix as assumed in Lemma 8.21 and e_j is the j-th canonnical vector in $\mathbb{R}^{|Y_j|}$. Note that, after a permutation of the player ordering (which causes no loss in generality), this fits the format of (8.14). Note also that Assumption 8.31 is satisfied since $x_j(n) \in [0,1]$ for all n. Furthermore, Assumption 8.32 is satisfied since, by construction, $\hat{x}_i(0) = x_j(0) = 0$ for all i, and Assumption 8.33 is satisfied by the hypothesis of Lemma 8.21.

Now, let $\epsilon(n) := |x_j(n+1) - x_j(n)|$ and suppose that starting at time n^* some action $y^* = (y_1^*, \ldots, y_N^*) \in Y$ is repeated in T consecutive stages, where $T \in \mathbb{N}$ is arbitrary. Two cases must be considered—the case that $y_j = y_j^*$ (i.e., the action which defines x_j is in fact the action being repeated by player j), and the case that $y_j \neq y_j^*$ (i.e., the action which defines x_j is not being played at all by j during the designated time sequence.) If $y_j = y_j^*$ then $\{x_j(n)\}_{n=n^*}^{n^*+T-1} = \{q_j(n, y_j)\}_{n=n^*}^{n^*+T-1}$ increases monotonically towards 1 (this follows from (8.8)). Otherwise, if $y_j \neq y_j^*$ then $\{x_j(n)\}_{n=n^*}^{n^*+T-1} = \{q_j(n, y_j)\}_{n=n^*}^{n^*+T-1}$ decreases monotonically towards 0. Since, in either case the sequence is monotone, we have by Lemma 8.35 that $\sum_{n=n^*+1}^{n^*+T-2} \epsilon(n) \leq 1$.

Note also that if some action y^* is repeated from time n^* to $n^* + T - 1$, then the difference sequence $\{\epsilon(n)\}_{n=n^*}^{n^*+T-2} = \{|q_j(n+1, y_j) - q_j(n, y_j)|\}_{n=n^*}^{n^*+T-2}$ is decreasing. This follows from (8.8).

We are now in a position to apply Lemma 8.34. By the equivalence of finite dimensional norms, there exist constants c_1 and c_∞ such that $\|\cdot\| \leq c_1 \|\cdot\|_1$ and $c_\infty \|\cdot\|_\infty \leq \|\cdot\|$. Given $j \in \mathcal{N}$ and some action $y_j \in Y_j$ we may choose a constant $T_{y_j} \in \mathbb{N}$ sufficiently large such that

$\frac{N+1}{1-\lambda}\left(\frac{1}{T_{y_j}} + B\lambda^{T_{y_j}}\right) < c_\infty \frac{\epsilon}{c_1 \sum_{i=1}^N |Y_i|}$. Applying Lemma 8.34 we get that if any action y^* is repeated in $T \geq T_{y_j}$ consecutive stages starting at any time n^* then

$$\begin{aligned}
&\max_{i \in \mathcal{N}} |q^i(j, n^* + T - 1, y_j) - q_j(n^* + T - 1, y_j)| \qquad (8.18)\\
&= \|(q^i(j, n^* + T - 1, y_j))_{i=1}^N - q_j(n^* + T - 1, y_j)\mathbf{1}\|_\infty \\
&= \|\hat{x}(n^* + T - 1) - x_j(n^* + T - 1)\mathbf{1}\|_\infty \\
&< \frac{1}{c_\infty}\|\hat{x}(n^* + T - 1) - x_j(n^* + T - 1)\mathbf{1}\| \leq \frac{\epsilon}{c_1 \sum_{i=1}^N |Y_i|}.
\end{aligned}$$

Let $T := \max_{j \in \mathcal{N}, y_j \in Y_j} T_{y_j}$. By (8.18) we have $|q_j^i(n^* + \tilde{T}, y_j) - q_j(n^* + \tilde{T}, y_j)| < \frac{\epsilon}{c_1 | \sum_{i=1}^N |Y_i|}$ for all $\tilde{T} \geq T$ for all $i, j \in \mathcal{N}$ and for all $y_j \in Y_j$.

Now, fix any player $i \in \mathcal{N}$. Observe that for any $\tilde{T} \geq T$ we have $\|\hat{q}^i(n^* + \tilde{T} - 1) - q(n^* + \tilde{T} - 1)\| \leq c_1 \|\hat{q}^i(n^* + \tilde{T} - 1) - q(n^* + \tilde{T} - 1)\|_1 = c_1 \sum_{j \in \mathcal{N}} \sum_{y_j \in Y_j} |\hat{q}_j^i(n^* + \tilde{T} - 1, y_j) - q_j(n^* + \tilde{T} - 1, y_j)| \leq c_1 (\sum_{j=1}^N |Y_j|) \frac{\epsilon}{c_1 \sum_{j=1}^N |Y_j|} = \epsilon$, which is the desired result. □

Chapter 9

Non-Degenerate Potential Games

9.1 Introduction

The class of potential games was introduced in Section 2.1.1. In a potential game, there exists some underlying potential function which all players implicitly seek to optimize. Such games have extensive applications within the field of multi-agent systems [3–5, 8, 13, 15, 48, 48, 49, 63, 93–95].

In this chapter we define the notion of a non-degenerate potential game. Our primary motivation in defining this class of games comes from the perspective learning theory: One might expect that FP dynamics *ought* to converge to a pure-strategy NE in potential games. Yet, there are seemingly "degenerate" cases in which FP dynamics can converge generically to mixed-strategy NE. Furthermore, one might expect that FP dynamics *ought* to always converge at an exponential rate. Yet, there are seeming degeneracies which make it difficult to establish any convergence rate estimates for FP in potential games [27]. We wish to define a class of games in which (i) FP can be shown to generically converge to pure-strategy NE, (ii) convergence rate estimates for FP can be established, and (iii) the game structure allows for relatively simple analysis of the associated learning dynamics.

We note that, in addition to facilitating the study of game-theoretic learning processes, we will see that the non-degenerate games defined here possess several other useful properties of general interest (see, e.g., Theorem 9.1).

We will say that an equilibrium strategy is non-degenerate if the derivatives of the potential function at the equilibrium satisfy certain properties. Roughly speaking, we say an equilibrium $\sigma \in \Delta^N$ is *first-order degenerate* if the directional derivative of the potential function is zero is the direction of some pure-strategy that is *not* in the support of σ. Otherwise, we will say an equilibrium σ is *first-order non-degenerate.*

Intuitively speaking, the notion of first-order degeneracy serves the following purpose. If σ is some equilibrium which does to have full support in terms of players pure-strategies, then σ may be seen as lying in a "face" of the strategy space, rather than in the interior. The first-order non-degeneracy condition ensures that, within a neighborhood of the equilibrium σ, best-response learning dynamics cannot abruptly change direction *orthogonal* to the face of the strategy space containing the σ.

We will say that an equilibrium σ is *second-order degenerate* if the Hessian of the potential function (constrained to the face of the strategy space containing σ) is singular. Otherwise, we will say an equilibrium σ is *second-order non-degenerate*. Second-order non-degeneracy may be seen as serving several purposes. First, it ensures that the potential function is not "flat" in the neighborhood of an equilibrium. While the first-order non-degeneracy condition ensures that the dynamics of a learning process are well behaved in directions that are *orthogonal* to the face of the strategy space containing σ, second-order non-degeneracy ensures that the dynamics are well behaved in directions *tangential* to the face of the strategy space containing σ. The second-order degeneracy condition also serves a secondary, more practical purpose. By ensuring that the Hessian of the potential function is non-singular, we are able to greatly simplify the analysis of dynamics around NE points. In particular, one can understand many fundamental properties of the potential function without needing consider anything higher than second-order terms in the Taylor-series expansion.

While our non-degeneracy conditions are tailored around studying learning dynamics in potential games, they coincide with well-known notions of non-degeneracy in general N-player games. In general, an equilibrium σ is said to be a *quasi-strong* equilibrium if the number of pure-strategy best responses to σ is no greater than the number of strategies in the support of σ [65, 67, 149]. Restricted to the class of potential games, this coincides precisely with our notion of first-order non-degeneracy.

The notion of first-order non-degeneracy is also closely related to the notion of non-degeneracy that is generally considered when studying the Lemke-Howson algorithm for computing NE strategies in finite games [16]. In this context, a game is said to be non-degenerate if the number of pure-strategy best responses to any mixed-strategy σ is no greater than the number of pure strategies in the support of σ (cf. [16], Definition 3.2). Our first-order degeneracy condition is slightly weaker than this, since it only requires this condition to hold at equilibrium points.

An equilibrium σ in a general N-player game is said to be *regular* if the Jacobian of a certain differential map of the players utility functions is non-singular at σ [67, 149]. Restricted to the class of potential games, this coincides with our notion of second-order non-degeneracy.

Regular and quasi-strong equilibria have been well studied [65, 67], and it has been shown that in almost all N-player games, all NE are both regular and quasi-strong [65].[1] Such equilibria are of interest in the general game theory community, where they are considered more "reasonable" notions of equilibrium (e.g., robust to perturbations) than prescribed by the NE concept alone. Furthermore, robust and quasi-strong equilibria are the subject of Harsanyi's celebrated purification theorem [149], where it is shown that such equilibria may be "approached" in an an associated game with incomplete information.

We will say a potential game is degenerate if it contains any first or second-order degenerate equilibria, and we will say that it is non-degenerate otherwise. The main contribution of this chapter will be to show that almost all potential games are non-degenerate (see Theorem 9.9).

Using the first and second-order non-degeneracy conditions, it is fairly straightforward to show that, in a non-degenerate potential game, the set of NE is finite, and all NE are isolated. From the perspective of learning theory, this is a useful property, since it tends to simplify analysis by allowing one to study learning dynamics near each equilibrium in isolation.

However, using the work of Harsanyi [65], we may in fact obtain an even stronger characterization of the set of NE in potential games.

Theorem 9.1. *In almost all potential games, the number of Nash equilibrium strategies is finite and odd.*

This result parallels the well-known oddness theorem for general N-player games [65, 150]. In particular, Harsanyi [65] showed that in any game in which all equilibria are quasi-strong and regular, the number of NE strategies is finite and odd. Since we show that almost all potential games are non-degenerate in this sense, Theorem 9.1 follows as a consequence of the main result of this chapter (Theorem 9.9) and Theorem 1 of [65].

The remainder of the chapter is organized as follows. In Section 9.1.1 we discuss our strategy for proving Theorem 9.9 (the main result of the chapter). In Section 9.2 we set up some notation that will be used in both this chapter and in Chapter 10. In Section 9.3 we rigorously define our notions of non-degeneracy and present the main result of the chapter (Theorem 9.9). In Section 9.4 we prove that almost all potential games are second-order non-degenerate (Proposition 9.10), and in Section 9.5 we prove that almost all potential games are first-order non-degenerate (Proposition 9.15). In Section 9.6 we show that in any non-degenerate potential games, all equilibria quasi-strong and regular.

[1] We note that the class of N-player potential games constitutes a null set in the class of general N-player games. Hence, finer results are required to show that these properties hold in almost all potential games.

9.1.1 Proof Strategy

We will prove Theorem 9.9 in two steps. First, we will prove that almost all potential games are second-order non-degenerate (Proposition 9.10). We follow roughly the approach of Harsanyi [65], setting up an appropriate mapping into the space of N-player games, and proving that degenerate games are contained in the set of critical values of the map. The result then follows from Sard's Theorem [109]. We note, however, that the mapping used by Harsanyi for general games fails to give any useful information when restricted to potential games. Consequently, our construction differs significantly from Harsanyi's in terms of the mapping used and some of the fundamental technical tools used. Must significantly, we require a strong characterization of the rank of the linear mapping that relates equilibrium points of a game and the game utility structure (see (9.16) and Proposition 9.11). This characterization relies on (relatively) recent results on the signsolvability of matrices [151], not available to Harsanyi. The case of potential games also differs from the general case in that it is not possible to construct a single mapping a la Harsanyi [65] whose critical values set contains all degenerate games. We are required to decompose the strategy space into a countable family of subsets and construct an appropriate mapping into the space of potential games from each subset.

As our second step in proving Theorem 9.9, we show that almost all potential games are first-order non-degenerate (Proposition 9.15). In order to show this, we use our characterization of the rank of the linear mapping in (9.16) (see discussion above) to construct a Lipschitz mapping from a set with Low Hausdorff dimension into the space of potential games, such that the range of the map contains all first-order degenerate games. The result then follows from the fact that the graph of a Lipschitz mapping cannot have a higher Hausdorff dimension than its domain [152].

9.2 Notation

Supposing that Γ is a potential game with potential function $u : Y \to \mathbb{R}$, let $U : \Delta^N \to \mathbb{R}$ be the multilinear extension of u defined by

$$U(\sigma_1, \ldots, \sigma_N) = \sum_{y \in Y} u(y)\sigma_1(y_1) \cdots \sigma(y_N). \tag{9.1}$$

The function U may be seen as giving the expected value of u under the mixed strategy σ. Throughout this chapter, we refer to U as the *potential function* and to u as the *pure form of the potential function.*

Given a $\sigma_i \in \Delta_i$, let σ_i^k denote value of the k-th entry in σ_i, so that $\sigma_i = (\sigma_i^k)_{k=1}^{K_i}$. Since the potential function is linear in each σ_i, if we fix any $i = 1, \ldots, N$ we may express it as

$$U(\sigma) = \sum_{k=1}^{K_i} \sigma_i^k U(y_i^k, \sigma_{-i}). \tag{9.2}$$

In order to study learning dynamics without being (directly) encumbered by the hyperplane constraint inherent in Δ_i we define

$$X_i := \{x_i \in \mathbb{R}^{K_i - 1} : \ 0 \leq x_i^k \leq 1 \text{ for } k = 1, \ldots, K_i - 1, \text{ and } \sum_{k=1}^{K_i - 1} x_i^k \leq 1\},$$

where we use the convention that x_i^k denotes the k-th entry in x_i so that $x_i = (x_i^k)_{k=1}^{K_i - 1}$.

Given $x_i \in X_i$ define the bijective mapping $T_i : X_i \to \Delta_i$ as $T_i(x_i) = \sigma_i$ for the unique $\sigma_i \in \Delta_i$ such that $\sigma_i^k = x_i^{k-1}$ for $k = 2, \ldots, K_i$ and $\sigma_i^1 = 1 - \sum_{k=1}^{K_i - 1} x_i^k$. For $k = 1, \ldots, K_i$ let T_i^k be the k-th component map of T_i so that $T_i = (T_i^k)_{i=1}^{K_i}$.

Let $X := X_1 \times \cdots \times X_N$ and let $T : X \to \Delta^N$ be the bijection given by $T = T_1 \times \cdots \times T_N$. In an abuse of terminology, we sometimes refer to X as the *mixed-strategy space* of Γ. When convenient, given an $x \in X$ we use the notation x_{-i} to denote the tuple $(x_j)_{j \neq i}$. Letting $X_{-i} := \prod_{j \neq i} X_j$, we define $T_{-i} : X_{-i} \to \Delta_{-i}$ as $T_{-i} := (T_j)_{j \neq i}$. Let

$$\kappa := \sum_{i=1}^{N} (|Y_i| - 1) \tag{9.3}$$

denote the dimension of X, and note that $\kappa \neq K$, where K, defined earlier, is the cardinality of the joint pure strategy set Y.

Throughout the next two chapters we often find it convenient to work in X rather than Δ^N. In order to keep the notation as simple as possible we overload the definitions of some symbols when the meaning can be clearly derived from the context. In particular, let $\mathrm{BR}_i : X_{-i} \rightrightarrows X_i$ be defined by $\mathrm{BR}_i(x_{-i}) := \{x_i \in X_i : \ \mathrm{BR}_i(\sigma_{-i}) = \sigma_i, \ \sigma_i \in \Delta_i, \ \sigma_{-i} \in \Delta_{-i}, \ \sigma_i = T_i(x_i), \ \sigma_{-i} = T_{-i}(x_{-i})\}$. Similarly, given an $x \in X$ we abuse notation and write $U(x)$ instead of $U(T(x))$.

Given a pure strategy $y_i \in Y_i$, we will write $U(y_i, x_{-i})$ to indicate the value of U when player i uses a mixed strategy placing all weight on the y_i and the remaining players use the strategy $x_{-i} \in X_{-i}$. Similarly, we will say $y_i^k \in \mathrm{BR}_i(x_{-i})$ if there exists an $x_i \in \mathrm{BR}_i(x_{-i})$ such that $T_i(x_i)$ places weight one on y_i^k.

In this context, we let

$$\nabla U(x) := \left(\frac{\partial U(x)}{\partial x_i^k}\right)_{\substack{i=1,\ldots,N\\ k=1,\ldots,K_i-1}},$$

denote the "full" gradient of U, and for $i = 1, \ldots, N$, we let

$$\nabla_{x_i} U(x) := \left(\frac{\partial U(x)}{\partial x_i^k}\right)_{k=1,\ldots,K_i-1}$$

denote the gradient of U with respect to the strategy of player i only.

Applying the definition of T_i to (9.2) we see that $U(x)$ may also be expressed as

$$U(x) = \sum_{k=1}^{K_i-1} x_i^k U(y_i^{k+1}, x_{-i}) + \left(1 - \sum_{k=1}^{K_i-1} x_i^k\right) U(y_i^1, x_{-i}). \tag{9.4}$$

for any $i = 1, \ldots, N$.

We use the following nomenclature to refer to strategies in X.

Definition 9.2.
(i) A strategy $x \in X$ is said to be pure *if $T(x)$ places all its mass on a single action tuple $y \in Y$.*
(ii) A strategy $x \in X$ is said to be completely mixed *if x is in the interior of X.*
(iii) In all other cases, a strategy $x \in X$ is said to be incompletely mixed*.*

9.3 Degenerate Potential Games

In this section we formally define the notions of degeneracy considered throughout the chapter. In Section 9.3.1 we define the notion of a first-order degenerate equilibrium. In Section 9.3.2 we define the notion of a second-order degenerate equilibrium. In Section 9.3.3 we define the notion of a degenerate potential game and present the main result of the chapter, Theorem 9.9.

9.3.1 First-Order Degeneracy

Let Γ be a potential game. Throughout the section we work primarily in the spaces X_i and X. Following Harsanyi [65], we will define the carrier set of an element $x \in X$, a natural modification of a support set to the present context. For $x_i \in X_i$ let

$$\text{carr}_i(x_i) := \text{spt}(T_i(x_i)) \subseteq Y_i \tag{9.5}$$

and for $x = (x_1, \ldots, x_N) \in X$ let $\operatorname{carr}(x) := \operatorname{carr}_1(x_1) \cup \cdots \cup \operatorname{carr}_N(x_N)$.

Let $C = C_1 \cup \cdots \cup C_N$, where for each $i = 1, \ldots, N$, C_i is a nonempty subset of Y_i. We say that C is the carrier for $x = (x_1, \ldots, x_N) \in X$ if $C_i = \operatorname{carr}_i(x_i)$ for $i = 1, \ldots, N$ (or equivalently, if $C = \operatorname{carr}(x)$).

Let $\gamma_i := |C_i|$ and assume that the strategy set Y_i is reordered so that $\{y_i^1, \ldots, y_i^{\gamma_i}\} = C_i$; that is, the first γ_i strategies in Y_i are precisely the strategies in C_i. Under this ordering, the first $\gamma_i - 1$ components of any strategy x_i with $\operatorname{carr}_i(x_i) = C_i$ are free (not constrained to zero by C_i) and the remaining components of x_i are constrained to zero. That is $(x_i^k)_{k=1}^{\gamma_i-1}$ is free under C_i and $(x_i^k)_{k=\gamma_i}^{K_i} = 0$. The set of strategies $\{x \in X : \operatorname{carr}(x) = C\}$ is precisely the interior of the face of X given by

$$\Omega_C := \{x \in X : x_i^k = 0, \; k = \gamma_i, \ldots, K_i - 1, \; i = 1, \ldots, N\}. \tag{9.6}$$

Definition 9.3. *Let x^* be an equilibrium with carrier C. We say that x^* is* first-order degenerate *if there exists a pair (i,k), $i = 1, \ldots, N$, $k = \gamma_i, \ldots, K_i - 1$ such that $\frac{\partial U(x^*)}{\partial x_i^k} = 0$, and we say x^* is* first-order non-degenerate *otherwise.*

Remark 9.4. *We note that, using the multilinearity of U one may show that an equilibrium x^* is first-order degenerate if and only if $\operatorname{carr}_i(x_i^*) \subsetneq \operatorname{BR}_i(x_{-i}^*)$ for some $i = 1, \ldots, N$ (cf. non-degeneracy in [153]).*

Example 9.5. *The 2×2 identical interests game with payoff matrix $M = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}$ has a first-order degenerate equilibrium at the strategy in which the row player plays his second action with probability 1 and the column player mixes between both his actions with equal probability.*

9.3.2 Second-Order Degeneracy

Let C be some carrier set. Let $\tilde{N} := |\{i = 1, \ldots, N : \gamma_i \geq 2\}|$, and assume that the player set is ordered so that $\gamma_i \geq 2$ for $i = 1, \ldots, \tilde{N}$. Under this ordering, for strategies with $\operatorname{carr}(x) = C$, the first $\tilde{N}$ players use mixed strategies and the remaining players use pure strategies. Assume that $\tilde{N} \geq 1$ so that any x with carrier C is a mixed (not pure) strategy.

Let the Hessian of U taken with respect to C be given by

$$\tilde{\mathbf{H}}(x) := \left(\frac{\partial^2 U(x)}{\partial x_i^k \partial x_j^\ell} \right)_{\substack{i,j=1,\ldots,\tilde{N}, \\ k=1,\ldots,\gamma_i-1, \\ \ell=1,\ldots,\gamma_j-1}}. \tag{9.7}$$

Note that this definition of the Hessian restricts attention to the components of x that are free under C

Definition 9.6. *We say an equilibrium* $x^* \in X$ *is* second-order degenerate *if the Hessian* $\tilde{\mathbf{H}}(x^*)$ *taken with respect to* $\mathrm{carr}(x^*)$ *is singular, and we say* x^* *is* second-order non-degenerate *otherwise.*

Remark 9.7. *Note that both forms of degeneracy are concerned with the interaction of the potential function and the "face" of the strategy space containing the equilibrium* x^**. If* x^* *touches one or more constraints, then first-order non-degeneracy ensures that the gradient of the potential function is nonzero normal to the face* $\Omega_{\mathrm{carr}(x^*)}$*, defined in* (9.6)*. Second-order non-degeneracy ensures that, restricting* U *to the face* $\Omega_{\mathrm{carr}(x^*)}$*, the Hessian of* $U\big|_{\Omega_{\mathrm{carr}(x^*)}}$ *is non-singular. If* x^* *is contained within the interior of* X*, then the first-order condition becomes moot and the second-order condition reduces to the standard definition of a non-degenerate critical point.*

9.3.3 Degenerate Games: Main Result

We say that a potential game is *first-order degenerate* (*second-order degenerate*) if it contains a first-order degenerate equilibrium (second-order degenerate equilibrium) and we say it is *first-order non-degenerate* (*second-order non-degenerate*) otherwise. We say a potential game is *degenerate* if it is either first or second-order degenerate, and we say the game is *non-degenerate* otherwise.

The potential function U is uniquely determined by a vector in $\mathbb{R}^K$ designating the utility (or, more accurately, the potential) assigned to each action tuple $y \in Y$. Viewing the potential function as such a vector in $\mathbb{R}^K$, we have the following definition.

Definition 9.8. *We say that a property holds for* almost all *potential games if the complement of the set where the property holds has* $\mathcal{L}^K$*-measure zero.*

The following theorem is the main result of this section.

Theorem 9.9. *Almost all potential games are non-degenerate.*

We will prove this theorem in two steps. We begin by showing that almost all games are second-order non-degenerate (Section 9.4, Proposition 9.10). Subsequently we show that almost all games are first-order non-degenerate (Section 9.5, Proposition 9.15). Propositions 9.10 and 9.15 together prove Theorem 9.9.

9.4 Second-Order Degenerate Games

The goal of this subsection is to prove the following proposition.

Proposition 9.10. *The set of potential games which are second-order degenerate has $\mathcal{L}^K$-measure zero.*

We will prove the proposition using Sard's theorem. Our construction roughly follows that of [65]. We begin by introducing some pertinent notation and giving some preliminary results.

Note that the set of joint pure strategies Y may be expressed as an ordered set $Y = \{y^1, \dots, y^K\}$ where each $y^\tau \in Y$, is an N-tuple of strategies, $\tau \in \{1, \dots, K\}$. We will assume a particular ordering for this set after Proposition 9.11.

For each pure strategy $y^\tau \in Y$, $\tau = 1, \dots, K$ let u^τ denote the pure-strategy potential associated with playing y^τ; that is, $u^\tau := u(y^\tau)$, where u is the pure form of the potential function defined in Section 9.2. A vector of *potential coefficients* $u = (u^\tau)_{\tau=1}^K$ is an element of $\mathbb{R}^K$.

If we consider u as a variable, then by (9.1), U is linear in u.[2] At this point we will express U in a more convenient form.

Let $\tau = 1, \dots, K$, $i = 1, \dots, N$ and $x_i \in X_i$. We define $q_i^\tau : X_i \to [0,1]$ by

$$q_i^\tau(x_i) := T_i^k(x_i) \tag{9.8}$$

where k corresponds to the action played by player i in the tuple y^τ, i.e, $(y^\tau)_i = y_i^k$, and where $T_i^k(x_i)$ is the k-th component of $T_i(x_i)$. In an abuse of notation, given a pure strategy $y_i^k \in Y_i$, we let $q_i^\tau(y_i^k) = 1$ if $(y^\tau)_i = y_i^k$ and $q_i^\tau(y_i^k) = 0$ otherwise.

Given a fixed vector of potential coefficients $u \in \mathbb{R}^K$, the potential function $U : X \to \mathbb{R}$ may be expressed as (see (9.1) and (9.8))

$$U(x) = \sum_{\tau=1}^K u^\tau \left[\prod_{i=1}^N q_i^\tau(x_i) \right]. \tag{9.9}$$

Note that this form makes it clear that U is linear in u.

Now, let $C = C_1 \cup \dots \cup C_N$ be some carrier set. The analysis through the remainder of the section will rely on this carrier set being fixed, and many of the subsequent terms are implicitly dependent on the choice of C. In keeping with our prior convention we let $\gamma_i := |C_i|$, and let $\tilde{N} := |\{i \in \{1, \dots, N\} : \gamma_i \geq 2\}|$.

Any x_i with carrier C_i has precisely $\gamma_i - 1$ free components (i.e., not constrained to zero by C_i).

[2] The potential function U is, of course, a function of both x and u. However, since we will only exploit the dependence on u in this section, we generally stick to the standard game-theoretic convention of writing U as a function of x only [19].

The joint strategy $x = (x_1, \dots, x_N)$ is a vector with

$$\gamma := \sum_{i=1}^{N} (\gamma_i - 1)$$

free components.

By (9.4) we have that

$$\frac{\partial U(x)}{\partial x_i^k} = U(y_i^{k+1}, x_{-i}) - U(y_i^1, x_{-i}) =: F_i^k(x, u) \tag{9.10}$$

for $i = 1, \dots, \tilde{N},\ k = 1, \dots, \gamma_i - 1$. Let

$$F(x, u) := \Big(F_i^k(x, u) \Big)_{\substack{i=1,\dots,\tilde{N} \\ k=1,\dots,\gamma_i-1}} . \tag{9.11}$$

Given an $x \in X$, it is at times useful to decompose it as $x = (x_p, x_m)$, where $x_m = (x_i^k)_{i=1,\dots,\tilde{N},\ k=1,\dots,\gamma_i-1}$ and x_p contains the remaining components of x. (The subscript of x_m is indicative of "mixed strategy components" of x and x_p indicative of "pure strategy components" of x.) In this decomposition, x_m is a γ-dimensional vector containing the free components of x. Taking the Jacobian of F in terms of the components of x_m we find that

$$D_{x_m} F(x_p, x_m, u) = \tilde{\mathbf{H}}(x), \tag{9.12}$$

where $D_{x_m} F(x_p, x_m, u) = \left(\frac{\partial F}{\partial x_i^\ell} \right)_{\substack{i=1,\dots,\tilde{N}, \\ \ell=1,\dots,\gamma_i-1}}$.

Let x^* be a mixed equilibrium with carrier C. Differentiating (9.4) we see that at the equilibrium x^* we have $\frac{\partial U(x^*)}{\partial x_i^k} = 0$ for $i = 1, \dots, \tilde{N},\ k = 1, \dots, \gamma_i - 1$ (see Lemma 10.28 in appendix), or equivalently,

$$F_i^k(x^*, u) = U(y_i^{k+1}, x_{-i}^*) - U(y_i^1, x_{-i}^*) = 0$$

for $i = 1, \dots, \tilde{N},\ k = 1, \dots, \gamma_i - 1$. Using (9.9) in the above we get

$$F_i^k(x^*, u) = \sum_{\tau=1}^{K} u^\tau \left[\Big(q_i^\tau(y_i^{k+1}) - q_i^\tau(y_i^1) \Big) \prod_{j \neq i} q_j^\tau(x_j^*) \right] = 0. \tag{9.13}$$

It will be convenient to be able to relate the ordering of Y with the ordering of $(F_i^k)_{i=1,\dots,\tilde{N},\ k=1,\dots,\gamma_i-1}$.

For this purpose, given $i = 1, \dots, \tilde{N}$, $k = 1, \dots, \gamma_i - 1$, let

$$s^*(i,k) := \begin{cases} k & \text{for } \ i = 1, \\ \sum_{j=1}^{i-1}(\gamma_j - 1) + k & \text{for } \ i \geq 2. \end{cases}$$

Define $i^* : \{1, \dots, \gamma\} \to \{1, \dots, \tilde{N}\}$ and $k^* : \{1, \dots, \gamma\} \to \{1, \dots, \max_i\{\gamma_i - 1\}\}$ to be the inverse of s^*; that is

$$s^*(i^*(s), k^*(s)) = s \tag{9.14}$$

for all $s = 1, \dots, \gamma$.

Given an $x \in X$, let $\mathbf{A}(x) = (a_{s,\tau}(x))_{\substack{s=1,\dots,\gamma, \\ \tau=1,\dots,K}} \in \mathbb{R}^{\gamma \times K}$ be defined as the matrix with entries

$$a_{s,\tau}(x) := \left(q^\tau_{i^*(s)}(y_i^{k^*(s)+1}) - q^\tau_{i^*(s)}(y_i^1)\right) \prod_{j \neq i^*(s)} q^\tau_j(x_j), \tag{9.15}$$

Using this notation, (9.13) is equivalently expressed as

$$\mathbf{A}(x^*)u = 0. \tag{9.16}$$

The following proposition establishes the solvability of (9.16). We note that, in order to prove Proposition 9.10 (the main result of this section), it is sufficient to consider only x such that $\operatorname{carr}(x) = C$ in the following proposition. However, later, when studying first order degenerate games in Section 9.5, we will consider the notion of an "extended carrier set" (see (9.25)), and we will need to characterize the rank of $\mathbf{A}(x)$ for x with $\operatorname{carr}(x) \subset C$.

Proposition 9.11. *For any x such that $\operatorname{carr}(x) \subseteq C$, the matrix $\mathbf{A}(x)$ has full row rank.*

Before proving this proposition we introduce some notation that will permit us to study the structure of $\mathbf{A}(x)$.

We begin by establishing a particular ordering of elements in Y. Let

$$\tilde{K} := \prod_{i=1}^{N} \gamma_i.$$

For $\tau = 1, \dots, \tilde{K}$, let $\alpha_\tau = (\alpha_\tau^1, \dots, \alpha_\tau^{\tilde{N}})$ be a multi-index associated with the τ-th action tuple in

Y, meaning that

$$y^\tau = (y_1^{\alpha_\tau^1}, \ldots, y_{\tilde N}^{\alpha_\tau^{\tilde N}}, y_{\tilde N+1}^1, \ldots, y_N^1),$$

where y_i^1, for $i = \tilde N + 1, \ldots, N$, is, by construction, the pure strategy used by player i at any strategy x with carrier C. Let Y be reordered so that for every $\tau = 1, \ldots, \tilde K$, $i = 1, \ldots, \tilde N$ we have $1 \le \alpha_\tau^i \le \gamma_i$. This ensures that the first $\tilde K$ strategies in Y contain all strategy combinations of the elements of $C_1, \ldots, C_N$, and that for any multi-index $\alpha = (\alpha^1, \ldots, \alpha^{\tilde N})$, $1 \le \alpha^i \le \gamma_i$, there exists a unique $1 \le \tau \le \tilde K$ such that $y^\tau = (y_1^{\alpha^1}, \ldots, y_{\tilde N}^{\alpha^{\tilde N}}, y_{\tilde N+1}^1, \ldots, y_N^1)$.

By definition (9.8) and the ordering we assumed for Y we have that

$$q_i^\tau(y_i^k) = \begin{cases} 1 & \text{if } k = \alpha_\tau^i \\ 0 & \text{otherwise.} \end{cases} \tag{9.17}$$

For $1 \le s \le \gamma$ and $1 \le \tau \le \tilde K$, let

$$r_{s,\tau} := q_i^\tau(y_i^{k+1}) - q_i^\tau(y_i^1) \quad \text{and} \quad p_{s,\tau}(x) := \prod_{j \ne i} q_j^\tau(x_j), \tag{9.18}$$

where $i = i^*(s)$ and $k = k^*(s)$ (see (9.14)), and note that $a_{s,\tau}(x) = r_{s,\tau} p_{s,\tau}(x)$ (see (9.15)). We may write $\mathbf{A}(x) = \mathbf{R} \circ \mathbf{P}(x)$, where $\circ$ is the Hadamard product, and $\mathbf{R}$ and $\mathbf{P}(x)$ have entries $r_{s,\tau}$ and $p_{s,\tau}(x)$ respectively.

Partition $\mathbf{A}(x)$, $\mathbf{R}$ and $\mathbf{P}(x)$ as $\mathbf{A}(x) = [\mathbf{A}_1(x)\ \mathbf{A}_2(x)]$, $\mathbf{R} = [\mathbf{R}_1\ \mathbf{R}_2]$, and $\mathbf{P}(x) = [\mathbf{P}_1(x)\ \mathbf{P}_2(x)]$, where $\mathbf{A}_1(x), \mathbf{R}_1, \mathbf{P}_1(x) \in \mathbb{R}^{\gamma \times \tilde K}$ and $\mathbf{A}_2(x), \mathbf{R}_2, \mathbf{P}_2(x) \in \mathbb{R}^{\gamma \times (K - \tilde K)}$, so we may write

$$\mathbf{A}(x) = [\mathbf{A}_1(x)\ \mathbf{A}_2(x)], \quad \text{with,} \quad \mathbf{A}_1(x) = \mathbf{R}_1 \circ \mathbf{P}_1(x) \ \text{ and } \ \mathbf{A}_2(x) = \mathbf{R}_2 \circ \mathbf{P}_2(x).$$

In order to show that $\mathbf{A}(x)$ has full row rank, it is sufficient to prove that $\mathbf{A}_1(x)$ has full row rank—this is the approach we will take in proving the proposition.

We address this by studying the sign pattern of $\mathbf{A}_1(x)$. Properties of *sign pattern matrices* (i.e., matrices with entries in $\{-1, 0, 1\}$) have been well-studied. We recall the following definition from [154],

Definition 9.12. *A sign pattern matrix* $\mathbf{L} \in \mathbb{R}^{m \times n}$ *with* $n \ge m$ *is said to be an L-matrix if for every matrix* $\mathbf{M}$ *with sgn*$(\mathbf{M}) =$ *sgn*$(\mathbf{L})$*, the matrix* $\mathbf{M}$ *has full row rank.*

The following Lemma characterizes L-matrices [151, 154].

Lemma 9.13. *Let* $\mathbf{L} \in \mathbb{R}^{m\times n}$ *be a sign pattern matrix with* $n \geq m$*. Then* $\mathbf{L}$ *is an L-matrix if and only if for every diagonal sign pattern matrix* $\mathbf{D} \in \mathbb{R}^{m\times m}$*,* $\mathbf{D} \neq 0$ *there is a nonzero column of* $\mathbf{DL}$ *in which each nonzero entry has the same sign.*

In light of Definition 9.12, Proposition 9.11 is equivalent to the following lemma.

Lemma 9.14. *For any* x *such that* $\mathrm{carr}(x) \subseteq C$*, the matrix* $\mathbf{A}_1(x) = (\mathbf{R}_1 \circ sgn(\mathbf{P}_1(x)))$ *is an L-matrix.*

The proof of this lemma relies on showing that $(\mathbf{R}_1 \circ \mathrm{sgn}(\mathbf{P}_1(x)))$ satisfies the L-matrix characterization given in Lemma 9.13.

Before proving this lemma we introduce some definitions that will be useful in the proof.

Given a diagonal matrix $\mathbf{D} \in \mathbb{R}^{\ell\times\ell}$, $\ell \in \mathbb{N}$, let $\mathrm{diag}(\mathbf{D})$ be the vector in $\mathbb{R}^{\ell}$ containing the diagonal elements of $\mathbf{D}$.

Given a diagonal sign pattern matrix $\mathbf{D} \in \mathbb{R}^{\ell\times\ell}$, $\ell \in \mathbb{N}$, define $\mathrm{idx}(\mathbf{D})$ as follows. If $\mathrm{diag}(\mathbf{D})$ does not contain any ones, then let $\mathrm{idx}(\mathbf{D}) = 1$. Otherwise, let $\mathrm{idx}(\mathbf{D})$ be one more than the first index in $\mathrm{diag}(\mathbf{D})$ containing a 1.[3][4]

Given a diagonal matrix $\mathbf{D} \in \mathbb{R}^{\gamma\times\gamma}$, let $\mathbf{D}_i \in \mathbb{R}^{(\gamma_i-1)\times(\gamma_i-1)}$, $i = 1, \ldots, \tilde{N}$ be the (unique) diagonal matrices satisfying $\mathrm{diag}(\mathbf{D}) = (\mathrm{diag}(\mathbf{D}_1), \ldots, \mathrm{diag}(\mathbf{D}_N))$.

Proof. Let x be a strategy satisfying $\mathrm{carr}(x) \subseteq C$. In order to show that $(\mathbf{R}_1 \circ \mathrm{sgn}(\mathbf{P}_1(x)))$ is an L-matrix, it is sufficient (by Lemma 9.13) to show that and for any diagonal sign pattern matrix $\mathbf{D} \neq 0$, there exists a column of $\mathbf{D}(\mathbf{R}_1 \circ \mathrm{sgn}(\mathbf{P}_1(x)))$ which is nonzero and in which every nonzero entry has the same sign. With this in mind, we begin by giving a characterization of the columns of $\mathbf{R}_1$.

Suppose that $i = 1, \ldots, \tilde{N}$ and $k = 1, \ldots, \gamma_i - 1$ are fixed. Note the following:

(i) Suppose $\tau \in \{1, \ldots, \tilde{K}\}$ is such that $\alpha_\tau^i = k + 1$, where α_τ^i is the i-th index of the multi-index α_τ. Since α_τ is used to define the ordering of actions in Y, we have $q_i^\tau(y_i^{k+1}) = 1$ and $q_i^\tau(y_i^1) = 0$ (see (9.17) and preceding discussion). Hence, $q_i^\tau(y_i^{k+1}) - q_i^\tau(y_i^1) = 1$.

(ii) Suppose $\tau \in \{1, \ldots, \tilde{K}\}$ is such that $\alpha_\tau^i = 1$. Then $q_i^\tau(y_i^{k+1}) = 0$, and $q_i^\tau(y_i^1) = 1$. Hence, $q_i^\tau(y_i^{k+1}) - q_i^\tau(y_i^1) = -1$.

[3] Assume indexing starts with one, not zero. For example, if the first time a 1 appears in $\mathrm{diag}(\mathbf{D})$ is at index 2, then $\mathrm{idx}(\mathbf{D}) = 3$.

[4] The awkward offset in this definition is needed in order to handle the indexing offset inherent in the mapping $T_i : X_i \to \Delta_i$, $i = 1, \ldots, N$.

(iii) For all other $\tau \in \{1, \ldots, \tilde{K}\}$ we have $q_i^\tau(y_i^{k+1}) = 0$, and $q_i^\tau(y_i^1) = 0$, and hence $q_i^\tau(y_i^{k+1}) - q_i^\tau(y_i^1) = 0$.

For $1 \le \tau \le \tilde{K}$, let $\mathbf{r}_\tau \in \mathbb{R}^\gamma$ be the τ-th column of $\mathbf{R}_1$. Partition this column as

$$\mathbf{r}_\tau = \begin{pmatrix} \mathbf{r}_\tau^1 \\ \vdots \\ \mathbf{r}_\tau^{\tilde{N}} \end{pmatrix}.$$

where $\mathbf{r}_\tau^i \in \mathbb{R}^{\gamma_i - 1}$. From (9.18) we see that

$$\mathbf{r}_\tau^i = \begin{pmatrix} q_i^\tau(y_i^2) - q_i^\tau(y_i^1) \\ \vdots \\ q_i^\tau(y_i^{\gamma_i}) - q_i^\tau(y_i^1) \end{pmatrix}.$$

Given the observations (i)–(iii) above we see that for each i we have

$$\mathbf{r}_\tau^i = \begin{cases} -\mathbf{1} & \text{if } \alpha_\tau^i = 1 \\ e_{\alpha_\tau^i - 1} & \text{if } 2 \le \alpha_\tau^i \le \gamma_i, \end{cases} \tag{9.19}$$

where the symbol $e_{\alpha_\tau^i - 1}$ refers to the $(\alpha_\tau^i - 1)$-th canonical vector in $\mathbb{R}^{\gamma_i - 1}$ and $\mathbf{1} \in \mathbb{R}^{\gamma_i - 1}$ is the vector of all ones.

We now characterize the columns of $\mathrm{sgn}(\mathbf{P}_1(x))$. For $i = 1, \ldots, \tilde{N}$ we define

$$\mathcal{I}_i(x) := \Big\{k \in \{1, \ldots, \gamma_i\} :\ T_i^k(x_i) > 0\Big\}.$$

Since $\mathrm{carr}(x) = C$, the ordering we assumed for Y_i implies that $T_i^k(x) = 0$ for $k \ge \gamma_i + 1$. By the definition of T_i, it is not possible to have $T_i^k(x_i) = 0$ for all $k = 1, \ldots, K_i$ and hence $\mathcal{I}_i(x) \neq \emptyset$.

Let $\mathbf{p}_\tau$ be the τ-th column of $\mathbf{P}_1(x)$ and let

$$\tilde{\mathbf{p}}_\tau := \mathrm{sgn}(\mathbf{p}_\tau)$$

be the τ-th column of $\mathrm{sgn}(\mathbf{P}_1(x))$.[5] Suppose that $\tau \in \{1, \ldots, \tilde{K}\}$ is such that for the multi-index α_τ we have $\alpha_\tau^i \in \mathcal{I}_i(x)$ for all $i = 1, \ldots, \tilde{N}$. Then for each $s = 1, \ldots, \gamma$ the (s, τ)-th entry of $\mathbf{P}_1(x)$ is strictly positive (see (9.8) and (9.18)), and hence $\mathbf{p}_\tau$ is positive and $\tilde{\mathbf{p}}_\tau = \mathbf{1}$.

Partition the columns $\mathbf{p}_\tau$ and $\tilde{\mathbf{p}}_\tau$ as

$$\mathbf{p}_\tau = \begin{pmatrix} \mathbf{p}_\tau^1 \\ \vdots \\ \mathbf{p}_\tau^{\tilde{N}} \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{p}}_\tau = \begin{pmatrix} \tilde{\mathbf{p}}_\tau^1 \\ \vdots \\ \tilde{\mathbf{p}}_\tau^{\tilde{N}} \end{pmatrix}, \tag{9.20}$$

where $\mathbf{p}_\tau^i, \tilde{\mathbf{p}}_\tau^i \in \mathbb{R}^{\gamma_i - 1}$. Suppose that τ is such that for the multi-index α_τ we have $\alpha_\tau^i \notin \mathcal{I}_i(x)$ for exactly one subindex $i \in \{1, \ldots, \tilde{N}\}$. Then $\mathbf{p}_\tau^i$ is positive (see (9.18)) and $\mathbf{p}_\tau^j$ is zero for any $j \neq i$. Hence, $\tilde{\mathbf{p}}_\tau^i = \mathbf{1}$ and $\tilde{\mathbf{p}}_\tau^j = 0$ for any $j \neq i$.

Now, let $\mathbf{D} \in \mathbb{R}^{\gamma \times \gamma}$ be an non-null diagonal sign pattern matrix. We will show that there is a nonzero column of $\mathbf{D}(\mathbf{R}_1 \circ \mathrm{sgn}(\mathbf{P}_1(x)))$ in which each nonzero entry is a 1.

We now consider two possible cases for the structure of $\mathbf{D}$ and show that in each case there is a nonzero column of $\mathbf{D}(\mathbf{R}_1 \circ \mathrm{sgn}(\mathbf{P}_1(x)))$ in which every nonzero entry is 1.

Case 1: Suppose that for all $i \in \{1, \ldots, \tilde{N}\}$ such that $\mathrm{idx}(\mathbf{D}_i) \notin \mathcal{I}_i(x)$ we have $\mathrm{diag}(\mathbf{D}_i) = 0$. Choose τ such that

$$\begin{cases} \alpha_\tau^i = \mathrm{idx}(\mathbf{D}_i) & \text{if } \ \mathrm{idx}(\mathbf{D}_i) \in \mathcal{I}_i(x), \\ \alpha_\tau^i \in \mathcal{I}_i(x) & \text{if } \ \mathrm{idx}(\mathbf{D}_i) \notin \mathcal{I}_i(x). \end{cases}$$

Note that $\alpha_\tau^i \in \mathcal{I}_i(x)$ for all $i = 1, \ldots, \tilde{N}$, and hence $\tilde{\mathbf{p}}_\tau = \mathbf{1}$ (see discussion preceding (9.20)). The τ-th column of $\mathbf{D}(\mathbf{R}_1 \circ \mathrm{sgn}(\mathbf{P}_1(x)))$ is given by

$$\mathbf{D}(\mathbf{r}_\tau \circ \tilde{\mathbf{p}}_\tau) = \mathrm{diag}(\mathbf{D}) \circ \mathbf{r}_\tau = \big(\mathrm{diag}(\mathbf{D}_i) \circ \mathbf{r}_\tau^i\big)_{i=1}^{\tilde{N}}. \tag{9.21}$$

For any i such that $\mathrm{idx}(\mathbf{D}_i) \notin \mathcal{I}_i(x)$ we have, by assumption, $\mathrm{diag}(\mathbf{D}_i) = 0$ and hence $\mathrm{diag}(\mathbf{D}_i) \circ \mathbf{r}_\tau^i = 0$. Moreover, note that in this case we have $\mathrm{idx}(\mathbf{D}_i) = 1$ since, by the definition of $\mathrm{idx}(\cdot)$, $\mathrm{diag}(\mathbf{D}_i) = 0$ implies $\mathrm{idx}(\mathbf{D}_i) = 1$.

Suppose now that i is such that $\mathrm{idx}(\mathbf{D}_i) \in \mathcal{I}_i(x)$. For $i = 1, \ldots, \tilde{N}$, if $\alpha_\tau^i = 1$ then $\mathbf{r}_\tau^i = -\mathbf{1}$ (by (9.19)) and $\mathrm{diag}(\mathbf{D}_i)$ contains no ones (this is the definition of $\mathrm{idx}(\mathbf{D}_i) = 1$). In fact, $\mathrm{diag}(\mathbf{D}_i)$ contains only entries with value of 0 or -1. Hence, $\mathrm{diag}(\mathbf{D}_i) \circ \mathbf{r}_\tau^i = -\mathrm{diag}(\mathbf{D}_i)$, which is a nonnegative vector.

[5]For convenience in notation, we suppress the argument x when writing the columns of these matrices.

If $2 \le \alpha_\tau^i \le \gamma_i$ then $\mathbf{r}_\tau^i = e_{\alpha_\tau^i - 1}$ (by (9.19)). Recalling the definition of $\mathrm{idx}(\cdot)$, by our choice of $\alpha_\tau^i = \mathrm{idx}(\mathbf{D}_i)$, the $(\alpha_\tau^i - 1)$-th entry of $\mathrm{diag}(\mathbf{D}_i)$ is 1. Hence, $\mathrm{diag}(\mathbf{D}_i) \circ \mathbf{r}_\tau^i = e_{\alpha_\tau^i - 1}$. In particular, this implies that if $\alpha_\tau^i \neq 1$ then $\mathrm{diag}(\mathbf{D}_i) \circ r_\tau^i$ is not identically zero and every nonzero entry of $\mathrm{diag}(\mathbf{D}_i) \circ r_\tau^i$ is 1.

In summary, for $i = 1, \ldots, \tilde{N}$, we have $\mathrm{diag}(\mathbf{D}_i) \circ \mathbf{r}_\tau^i \ge 0$, with equality only when $\mathrm{idx}(\mathbf{D}_i) = 1$ and $\mathbf{D}_i = 0$. Hence, by (9.21), the τ-th column of $\mathbf{D}(\mathbf{R}_1 \circ \mathrm{sgn}(\mathbf{P}_1(x)))$ satisfies $\mathbf{D}(\mathbf{r}_\tau \circ \tilde{\mathbf{p}}_\tau) \ge 0$, with equality only when $\mathrm{idx}(\mathbf{D}_i) = 1$ and $\mathbf{D}_i = 0$ for all i. But, by assumption $\mathbf{D} \neq 0$, so $\mathbf{D}(\mathbf{r}_\tau \circ \tilde{\mathbf{p}}_\tau) \neq 0$.

Case 2: Suppose that for some $i \in \{1, \ldots, \tilde{N}\}$ we have $\mathrm{idx}(\mathbf{D}_i) \notin \mathcal{I}_i(x)$ and $\mathrm{diag}(\mathbf{D}_i) \neq 0$. Let $\tau \in \{1, \ldots, \tilde{K}\}$ be chosen such that $\alpha_\tau^i = \mathrm{idx}(\mathbf{D}_i)$ for exactly one such $i \in \{1, \ldots, \tilde{N}\}$ and for all other $j \neq i$ we have $\alpha_\tau^i \in \mathcal{I}_j$ (this is always possible since $\mathcal{I}_j \neq \emptyset$). Then we have

$$\tilde{\mathbf{p}}_\tau^i = \mathbf{1}, \qquad \text{and} \qquad \tilde{\mathbf{p}}_\tau^j = 0, \quad \text{for all } j \neq i$$

(see discussion following (9.20)).

As shown in Case 1, if $\alpha_\tau^i = 1$, then $\mathbf{D}_i \le 0$ and $r_\tau^i = -1$ which implies that $\mathrm{diag}(\mathbf{D}_i) \circ \mathbf{r}_\tau^i \ge 0$. Moreover, since $\tilde{\mathbf{p}}_\tau^i = \mathbf{1}$ and since, by assumption $\mathrm{diag}(\mathbf{D}_i) \neq 0$ we have $\mathrm{diag}(\mathbf{D}_i) \circ \mathbf{r}_\tau^i \circ \tilde{\mathbf{p}}_\tau^i \neq 0$ and every nonzero entry is 1.

If $2 \le \alpha_\tau^i \le \gamma_i$, then, again using the same reasoning as in Case 1, we see that $\mathrm{diag}(\mathbf{D}_i) \circ \mathbf{r}_\tau^i = e_{\alpha_\tau^i - 1}$. Since $\tilde{\mathbf{p}}_\tau^i = \mathbf{1}$ we get that $\mathrm{diag}(\mathbf{D}_i) \circ \mathbf{r}_\tau^i \circ \mathbf{p}_\tau^i = e_{\alpha_\tau^i - 1}$.

For $j \neq i$ we have $\tilde{\mathbf{p}}_\tau^j = 0$, which implies $\mathrm{diag}(\mathbf{D}_j) \circ \mathbf{r}_\tau^j \circ \tilde{\mathbf{p}}_\tau^j = 0$.

All together, this implies that the τ-th column of $\mathbf{D}(\mathbf{R}_1 \circ \mathrm{sgn}(\mathbf{P}_1(x)))$, given by $(\mathrm{diag}(\mathbf{D}_j) \circ \mathbf{r}_\tau^j \circ \tilde{\mathbf{p}}_\tau^j)_{j=1}^{\tilde{N}}$, is nonzero and every nonzero entry is equal to 1.

Since this holds for arbitrary diagonal sign matrix $\mathbf{D} \neq 0$, Lemma 9.13 implies that $(\mathbf{R}_1 \circ \mathrm{sgn}(\mathbf{P}_1(x)))$ is an L-matrix. Since this holds for any x satisfying $\mathrm{carr}(x) \subseteq C$, we see that the desired result holds. □

Given the carrier C there are $\binom{K}{\gamma}$ possible combinations (of length γ) of the columns of $\mathbf{A}(x)$. For each $r = 1, \ldots, \binom{K}{\gamma}$, let $\mathbf{A}_r(x) \in \mathbb{R}^{\gamma \times \gamma}$ denote a square matrix formed by taking one unique combination of the columns of $\mathbf{A}(x)$. For $r = 1, \ldots, \binom{K}{\gamma}$, let

$$S_r := \{x \in X : \ \mathrm{carr}(x) = C, \ \det \mathbf{A}_r(x) \neq 0\}.$$

By Proposition 9.11, no strategy $x \in X$ with $\mathrm{carr}(x) = C$ may simultaneously be in all S_r^c. Note also that each S_r is open relative to the set $\{x \in X : \ \mathrm{carr}(x) = C\} = \mathring{\Omega}_C$ (see (9.6)). Thus, we may construct a countable family of open balls $(B_\ell)_{\ell \ge 1}$, $B_\ell \subset \Omega_C$ that satisfy:

(i) $\bigcup_{\ell \geq 1} B_\ell = \{x \in X : \text{ carr}(x) = C\}$
(ii) For each $\ell \in \mathbb{N}$ there exists an $r_\ell \in \{1, \ldots, \binom{K}{\gamma}\}$ such that $B_\ell \subseteq S_{r_\ell}$ (i.e., $\mathbf{A}_{r_\ell}(x)$ is invertible for all $x \in B_\ell$).

Fix $\ell \in \{1, 2, \ldots\}$. After reordering, $\mathbf{A}(x)$ may be partitioned as $\mathbf{A}(x) = [\tilde{\mathbf{A}}_{r_\ell}(x)\ \mathbf{A}_{r_\ell}(x)]$, where $\tilde{\mathbf{A}}_{r_\ell}(x)$ is a matrix formed by the columns of $\mathbf{A}(x)$ not used to form $\mathbf{A}_{r_\ell}(x)$. Let the strategy set Y be reordered in the same way as the columns $\mathbf{A}(x)$.[6] Given a vector of potential coefficients $u \in \mathbb{R}^K$, let it be partitioned as $u = (u_1, u_2)$, where $u_1 = (u^1, \ldots, u^{K-\gamma})$ and $u_2 = (u^{K-\gamma+1}, \ldots, u^K)$. Define $\tilde{\rho}_\ell : B_\ell \times \mathbb{R}^{K-\gamma} \to \mathbb{R}^\gamma$ by

$$\tilde{\rho}_\ell(x, u_1) := -\mathbf{A}_{r_\ell}(x)^{-1}\tilde{\mathbf{A}}_{r_\ell}(x)u_1,$$

If $x^* \in B_\ell$ is an equilibrium for some potential game with potential coefficient vector u, then by (9.16) we have $\mathbf{A}(x^*)u = 0$. Since $\mathbf{A}_{r_\ell}(x^*)$ is invertible, this is equivalent to $u_2 = -\mathbf{A}_{r_\ell}(x^*)^{-1}\tilde{\mathbf{A}}_{r_\ell}(x^*)u_1$. Hence, if $x^* \in B_\ell$ is an equilibrium of some potential game with potential coefficient vector $u = (u_1, u_2)$, the function $\tilde{\rho}_\ell$ permits us to recover u_2 given u_1 and x^*.

Conversely, suppose $x \in B_\ell$ and $u_1 \in \mathbb{R}^{K-\gamma}$ are arbitrary. If $u = (u_1, u_2)$ with $u_2 = \tilde{\rho}_\ell(x, u_1)$, then by the definition of $\tilde{\rho}_\ell$ we see that $\mathbf{A}(x)u = 0$. By the definition of $\mathbf{A}(x)$ (see (9.11)–(9.15)) this implies

$$F(x, u_1, \tilde{\rho}_\ell(x, u_1)) = 0, \qquad \text{for all } x \in B_\ell.$$

Thus, taking a partial derivative with respect to x_i^k we get

$$\frac{\partial}{\partial x_i^k} F(x, u_1, \tilde{\rho}_\ell(x, u_1)) = 0, \qquad i = 1, \ldots, \tilde{N},\ k = 1, \ldots, \gamma_i - 1. \tag{9.22}$$

Consider again the decomposition $x = (x_p, x_m)$. Using compact notation, (9.22) is restated as

$$D_{x_m} F(x_p, x_m, u_1, \tilde{\rho}_\ell(x_p, x_m, u_1)) = 0. \tag{9.23}$$

Suppose $x^* \in B_\ell$ is an equilibrium of a potential game with potential coefficients u and $\text{carr}(x^*) = C$. Applying the chain rule in (9.23), using (9.12), and using the fact that $F(x, u) = \mathbf{A}(x)u = \tilde{\mathbf{A}}_{r_\ell}(x)u_1 + \mathbf{A}_{r_\ell}(x)u_2$, we find that at x^* there holds

$$\tilde{\mathbf{H}}(x^*) = -\mathbf{A}_{r_\ell}(x^*)\tilde{\mathbf{J}}_{\rho_\ell}(x^*, u), \tag{9.24}$$

where $\tilde{\mathbf{J}}_{\rho_\ell}(x, u) := D_{\tilde{x}_m}\tilde{\rho}_\ell(x_p, \tilde{x}_m, u)\big|_{\tilde{x}_m = x_m}$ is the Jacobian of $\tilde{\rho}_\ell$ taken with respect to x_m.

[6]Note that we previously assumed a specific ordering for Y. However, this was for the purpose of proving Proposition 9.11, which is unaffected by a reordering of Y at this point.

Since $\mathbf{A}_{r_\ell}(x)$ is invertible for all $x \in B_\ell$, this means that given any equilibrium $x^* \in B_\ell$, the Hessian $\tilde{\mathbf{H}}(x^*)$ is nonsingular if and only if the Jacobian $\tilde{\mathbf{J}}_{\rho_\ell}(x^*, u)$ is nonsingular.

For each $\ell \in \mathbb{N}$, define the function $\rho_\ell : B_\ell \times \mathbb{R}^{K-\gamma} \to \mathbb{R}^K$ by $\rho_\ell(x, u_1) := (u_1, \tilde{\rho}_\ell(x, u_1))$. The function ρ_ℓ is a trivial extension of $\tilde{\rho}_\ell$ that recovers the full vector of potential coefficients $u \in \mathbb{R}^K$ given $u_1 \in \mathbb{R}^{K-\gamma}$ and an equilibrium $x^* \in B_\ell$.

Note that the Jacobian of ρ_ℓ takes the form

$$\mathbf{J}_{\rho_\ell}(x, u) = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{J}_{\tilde{\rho}_\ell} & \mathbf{M} \end{pmatrix},$$

for some matrix $\mathbf{M}$. Clearly, $\det \mathbf{J}_{\rho_\ell} = 0$ if and only if $\det \tilde{\mathbf{J}}_{\rho_\ell} = 0$. Thus, by (9.24) we see that if $x^* \in B_\ell$ is an equilibrium of a game with potential coefficient vector u, then

$$\det \mathbf{J}_{\rho_\ell}(x^*, u) = 0 \quad \iff \quad \det \tilde{\mathbf{H}}(x^*) = 0.$$

We now prove Proposition 9.10.

Proof. Let C be a carrier set. Let $\mathcal{U}(C) \subseteq \mathbb{R}^K$ be the set of games having at least one degenerate equilibrium with carrier set C; that is,

$$\mathcal{U}(C) := \big\{u \in \mathbb{R}^K : \exists \text{ degenerate equilibrium } x^* \in \{x \in X : \text{carr}(x) = C\}\big\}.$$

For $\ell \in \mathbb{N}$, let $\mathcal{U}(C, \ell) \subseteq \mathbb{R}^K$ be the subset of games having at least one degenerate equilibrium $x^* \in B_\ell$, where B_ℓ is defined with respect to C; that is,

$$\mathcal{U}(C, \ell) := \big\{u \in \mathbb{R}^K : \exists \text{ degenerate equilibrium } x^* \in B_\ell\big\}.$$

By construction, we have $\bigcup_{\ell \geq 1} B_\ell = \{x \in X : \text{carr}(x) = C\}$, and hence $\mathcal{U}(C) = \bigcup_{\ell \geq 1} \mathcal{U}(C, \ell)$.

We showed above that for any $(x, u) \in B_\ell \times \mathbb{R}^K$ such that x is an equilibrium of the potential game with potential coefficients u, the Hessian $\tilde{\mathbf{H}}(x)$ (taken with respect to C) is invertible if and only if the Jacobian of $\rho_\ell(x, u)$ is invertible. Thus, the set $\mathcal{U}(C, \ell)$ is contained in the set of critical values of ρ_ℓ. By Sard's theorem, we get that $\mathcal{U}(C, \ell)$ is a set with $\mathcal{L}^K$-measure zero. Since $\mathcal{U}(C)$ is the countable union of sets of $\mathcal{L}^K$-measure zero, it is itself a set with $\mathcal{L}^K$-measure zero.

Let $\mathcal{U} \subset \mathbb{R}^K$ denote the set of games with at least one degenerate equilibrium. The set $\mathcal{U}$ may be expressed as the union $\mathcal{U} = \bigcup_C \mathcal{U}(C)$ taken over all possible support sets C. Since there are a finite number of support sets C, the set $\mathcal{U}$ has $\mathcal{L}^K$-measure zero. $\square$

9.5 First-Order Degenerate Games

The following proposition shows that first-order degenerate games form a null set.

Proposition 9.15. *The set of potential games which are first-order degenerate has $\mathcal{L}^K$-measure zero.*

Proof. Fix some set $C = C_1 \cup \cdots \cup C_N$ where each C_i is a nonempty subset of Y_i. Let $\widehat{C}$ be any *strict* subset of C. In the context of this proof let $\gamma_i := |C_i|$, let $\gamma := \sum_{i=1}^N (\gamma_i - 1)$, let $\tilde{N} := |\{i = 1, \ldots, N : \gamma_i \geq 2\}|$, and assume Y_i is reordered so that $\{y_i^1, \ldots, y_i^{\gamma_i}\} = C_i$. Note that this ordering implies that for any x with $\text{carr}(x) = C$ we have $y_i^1 \in \text{carr}(x)$, $i = 1, \ldots, N$.

Given an equilibrium x^* let the *extended carrier* of x^* be defined as

$$\text{ext carr}(x^*) := \text{carr}(x^*) \cup \left(\bigcup_{i=1}^N \{y_i^k \in Y_i : \ k = 2, \ldots K_i, \ \frac{\partial U(x^*)}{\partial x_i^{k-1}} = 0\} \right). \tag{9.25}$$

Suppose that x^* is an equilibrium with *extended* carrier C. By Lemma 10.28 (see appendix) and the ordering we assumed for Y_i we have

$$F_i^k(x^*, u) = \frac{\partial U(x^*)}{\partial x_i^{k-1}} = 0, \qquad i = 1, \ldots, \tilde{N}, \ k = 1, \ldots, \gamma_i - 1, \tag{9.26}$$

where F_i^k is as defined in (9.10). Thus, if x^* is an equilibrium for some game with potential coefficient vector $u \in \mathbb{R}^K$, and $\text{ext carr}(x^*) = C$, then by the definition of $\mathbf{A}(x)$ (see (9.11)–(9.15)), (9.26) implies that $\mathbf{A}(x^*)u = 0$, or equivalently,

$$u \in \ker \mathbf{A}(x^*),$$

where the matrix $\mathbf{A}(x) \in \mathbb{R}^{\gamma \times K}$ is defined with respect to C, as in (9.15).

Let $\mathcal{U}(C, \widehat{C}) \subseteq \mathbb{R}^K$ be the set of games in which there exists an equilibrium x^* with $\text{carr}(x^*) = \widehat{C}$ and $\text{ext carr}(x^*) = C$. Let

$$\widehat{X} := \{x \in X : \ \text{carr}(x) = \widehat{C}\}.$$

By the above we see that

$$\mathcal{U}(C, \widehat{C}) \subseteq \cup_{x \in \widehat{X}} \ker \mathbf{A}(x). \tag{9.27}$$

For each $x \in \widehat{X}$, let $\operatorname{range} \mathbf{A}(x)^T$ denote the range space of $\mathbf{A}(x)^T$. Each entry of $\mathbf{A}(x)^T$ is a polynomial function in x and hence is Lipschitz continuous over the bounded set $\widehat{X}$. By Proposition 9.11 we have $\operatorname{rank}\mathbf{A}(x)^T = \gamma$ for all $x \in \widehat{X}$. Thus, we may choose a set of γ basis vectors $\{\mathbf{b}_1(x), \ldots, \mathbf{b}_\gamma(x)\}$ spanning $\operatorname{range} \mathbf{A}(x)^T$ such that each $\mathbf{b}_k(x) \in \mathbb{R}^K$, $k = 1, \ldots, \gamma$ is a Lipschitz continuous function in x. Moreover, we may choose a complementary set of $(K - \gamma)$ linearly independent vectors $\{\mathbf{b}_{\gamma+1}(x), \ldots, \mathbf{b}_K(x)\}$ forming a basis for the orthogonal complement $(\operatorname{range} \mathbf{A}(x)^T)^\perp$, with each $\mathbf{b}_k(x) \in \mathbb{R}^K$, $k = \gamma+1, \ldots, K$ being a continuous function in x. Let $\mathbf{B}(x) := (\mathbf{b}_{\gamma+1}(x) \ \cdots \ \mathbf{b}_K(x)) \in \mathbb{R}^{K \times (K-\gamma)}$. Let $f : \widehat{X} \times \mathbb{R}^{K-\gamma} \to \mathbb{R}^K$ be given by $f(x, v) := \mathbf{B}(x)v$. Since $\mathbf{B}(x)$ is Lipschitz continuous in x and $\widehat{X}$ is bounded, f is Lipschitz continuous. By the fundamental theorem of linear algebra, for each $x \in \widehat{X}$, $\ker \mathbf{A}(x) = (\operatorname{range} \mathbf{A}(x)^T)^\perp = \operatorname{range} \mathbf{B}(x)$. Hence,

$$\bigcup_{x \in \widehat{X}} \ker \mathbf{A}(x) = f(\widehat{X} \times \mathbb{R}^{K-\gamma}). \tag{9.28}$$

Since $\widehat{C} \subsetneq C$, the Hausdorff dimension of $\widehat{X}$ is at most $(\gamma - 1)$ and the Hausdorff dimension of $\widehat{X} \times \mathbb{R}^{K-\gamma}$ is at most $K-1$. Since f is Lipschitz continuous, this implies (see [152], Section 2.4) that the Hausdorff dimension of $f(\widehat{X} \times \mathbb{R}^{K-\gamma})$ is at most $K-1$, and in particular, that $f(\widehat{X} \times \mathbb{R}^{K-\gamma})$ has $\mathcal{L}^K$-measure zero. By (9.27) and (9.28), this implies that $\mathcal{U}(C, \widehat{C})$ has $\mathcal{L}^K$-measure zero.

Let $\mathcal{U} \subseteq \mathbb{R}^K$ denote the set of all games containing a first-order degenerate equilibrium. Since we may represent this set as a finite union of $\mathcal{L}^K$-measure zero sets,

$$\mathcal{U} = \bigcup_{\substack{\emptyset \neq C_i \subseteq Y_i, \ i=1,\ldots,N \\ C = C_1 \cup \cdots \cup C_N \\ \widehat{C} \subsetneq C}} \mathcal{U}(C, \widehat{C}),$$

the set $\mathcal{U}$ also has $\mathcal{L}^K$ measure zero. $\square$

9.6 Quasi-Strong and Regular Equilibria

The notions of a *quasi-strong equilibrium* and a *regular equilibrium* were introduced in [65] and have been studied in many subsequent works (see [66,67] and citations therein). Equilibria that are quasi-strong and regular possess a variety of useful properties. Of greatest note, such equilibria have been shown to be "essential" (robust against perturbations in the payoffs) and "strongly proper" (robust against perturbations in players' strategies) [67].

It has been shown [65] that in almost all (finite) N-player games, all equilibria are quasi-strong and regular. In this section we relate our non-degeneracy conditions (which were crafted specifically for potential games and rely explicitly on the potential function) to the notions of quasi-strong and regular equilibria. The main goal of this section will be show that the following two properties hold (see Proposition 9.18):

1. An equilibrium in a potential game is quasi-strong if and only if it first-order non-degenerate.
2. If an equilibrium in a potential game is quasi-strong, then it is regular if and only if it is second-order non-degenerate.

This means that an equilibrium of a potential game is non-degenerate (both first and second-order) if and only if it is quasi-strong and regular. Along with Theorem 9.9, this implies that in almost all potential games, all equilibria are both quasi-strong and regular, and possess all the associated robustness and stability properties.

The work [65] showed that in any game where all equilibria are quasi-strong and regular, the number of equilibrium strategies is finite and odd ([65], Theorem 1). Thus, Proposition 9.18 (the main result of this section), together with Theorem 9.9 of this dissertation and Theorem 1 of [65], implies that Theorem 9.1 holds.

We now formally define the notions of a quasi-strong and regular equilibrium. These notions are traditionally defined within the space Δ^N. We will review the traditional definitions given in [65] and then for convenience, we will recast these definitions to work with elements in X rather than in Δ^N.

An equilibrium $\sigma^* \in \Delta^N$ is said to be quasi-strong if for each $i = 1, \ldots, N$, every pure strategy in the support of σ_i^* is a best response to σ_{-i}^* [65, 67]. This may be stated equivalently as follows.

Definition 9.16. *An equilibrium $x^* \in X$ is said to be quasi-strong if* $\operatorname{carr}_i(x^*) = \{y_i \in Y_i : \ y_i \in \operatorname{BR}_i(x_{-i}^*)\}$ *for all $i = 1, \ldots, N$.*

Given $\sigma \in \Delta^N$ and $u \in \mathbb{R}^K$, let

$$\tilde{F}_i^k(\sigma, u) := \sigma_i^1 \sigma_i^{k+1} [U(y_i^{k+1}, \sigma_{-i}) - U(y_i^1, \sigma_{-i})] \tag{9.29}$$

for $i = 1, \ldots, N$, $k = 1, \ldots, K_i - 1$, and let $F_i^0(\sigma, u) := \sum_{k=1}^{K_i} \sigma_i^k - 1$ for $i = 1, \ldots, N$. Let

$$\tilde{F}(\sigma, u) := \left(\tilde{F}_i^k(\sigma, u)\right)_{\substack{i=1,\ldots,N \\ k=0,\ldots,K_i-1}}$$

An equilibrium $\sigma^* \in \Delta^N$ is said to be *regular* if the Jacobian $D_\sigma F(\sigma, u)$ is non-singular [65, 67]. We note that this definition of a regular equilibrium is slightly different from the one given in [65] in that we have used the potential function U in the place of the individual utility function U_i. Using the definition of a potential game it is straightforward to verify that $U(y_i^{k+1}, x_{-i}) - U(y_i^1, x_{-i}) = U_i(y_i^{k+1}, x_{-i}) - U_i(y_i^1, x_{-i})$ for any $x_{-i} \in X_{-i}$, and hence, this coincides with the definition given in [65] within the class of potential games.

We will now recast this definition so we may work with elements in the space X rather than Δ^N. In an abuse of notation, given an $x \in X$ and $u \in \mathbb{R}^K$, let

$$\begin{aligned} \tilde{F}_i^k(x,u) &:= T_i^1(x_i) x_i^k [U(y_i^{k+1}, x_{-i}) - U(y_i^1, x_{-i})] \\ &= T_i^1(x_i) x_i^k F_i^k(x,u), \end{aligned} \tag{9.30}$$

for $i = 1, \ldots, N$, $k = 1, \ldots, K_i - 1$, where $F_i^k(x,u)$ is defined as in (9.10).[7] Let

$$\tilde{F}(x,u) := \left(\tilde{F}_i^k(x,u)\right)_{\substack{i=1,\ldots,N \\ k=1,\ldots,K_i-1}}.$$

Note that the indexing here differs from that used to define $F(x,u)$ in (9.11), where we restricted to $k \in \{1, \ldots, \gamma_i - 1\}$.

Definition 9.17. *An equilibrium x^* of a potential game with potential coefficient vector u, is said to be* regular *if the Jacobian of $\tilde{F}(x^*,u)$, given by $D_x\tilde{F}(x^*,u)$, is non-singular.*

To see that this definition of a regular equilibrium coincides with the definition given in [65], suppose that $x^* \in X$ is an equilibrium of some potential game with potential coefficient vector u, and let $\sigma^* = T(x^*)$. The Jacobian $D_x\tilde{F}(x^*,u)$ may be formed from $D_\sigma\tilde{F}(\sigma^*,u)$ by removing the rows of $D_\sigma\tilde{F}(\sigma^*,u)$ corresponding to the coordinate maps $\tilde{F}_i^0$, $i = 1, \ldots, N$ and removing the columns of $D_\sigma\tilde{F}(\sigma^*,u)$ in which the partial derivative is taken with respect to σ_i^1, $i = 1, \ldots, N$.

Note that for $i = 1, \ldots, N$, and $k = 1, \ldots, K_i - 1$, we have $\frac{\partial \tilde{F}_i^k(\sigma^*,u)}{\partial \sigma_i^1} = 0$. This follows from (9.29) and the fact that for $k = 1, \ldots, K_i - 1$, either $\frac{\partial U(\sigma)}{\partial \sigma_i^{k+1}} = U(y_i^{k+1}, \sigma_{-i}) - U(y_i^1, \sigma_{-i}) = 0$ or $\sigma_i^{k+1} = 0$. Note also that $\frac{\partial \tilde{F}_i^0(\sigma^*,u)}{\partial \sigma_i^1} = 1$ for $i = 1, \ldots, N$. This means that for each $i = 1, \ldots, N$, the column of $D_\sigma\tilde{F}(\sigma^*,u)$ in which the partial derivative is taken with respect to σ_i^1 is composed of all zeros except for a one in the row corresponding to $\tilde{F}_i^0$. Thus, by the definition of the determinant, removing the above mentioned rows and columns from $D_\sigma\tilde{F}(\sigma^*,u)$ does not change the value of the determinant of the resulting matrix. Consequently, $D_x\tilde{F}(x^*,u)$ is invertible if and only if $D_\sigma\tilde{F}(\sigma^*,u)$ is invertible, and the equilibrium $x^* \in X$ is regular as defined in Definition 9.17 if and

[7] In (9.10), $F_i^k(x,u)$ is only defined for $i = 1, \ldots, \tilde{N}$, $k = 1, \ldots, \gamma_i - 1$; here we trivially extend this definition to $i = 1, \ldots, N$, $k = 1, \ldots, K_i - 1$.

only if σ^* is regular as defined in [65].

The following proposition is the main result of the section.

Proposition 9.18. *Let* Γ *be a potential game. Then,*
(i) An equilibrium x^* *is first-order non-degenerate if and only if it is quasi-strong.*
(ii) If an equilibrium x^* *is quasi-strong, then it is second-order non-degenerate if and only if it is regular.*

In particular, an equilibrium x^* *of a potential game is non-degenerate if and only if it is quasi-strong and regular.*

Proof. In order to simplify notation, throughout the proof we will use the symbol $\hat{x}$ rather than the usual x^*, when referring to an equilibrium. Without loss of generality, given an equilibrium $\hat{x}$, assume that each player's action set Y_i is reordered so that $\text{carr}_i(\hat{x}_i) = \{y_i^1, \dots, y_i^{\gamma_i}\}$. Note that this comports with the ordering assumption implicit in the definitions of both first and second-order degeneracy (see Sections 9.3.1–9.3.2).

Using the multilinearity of U, it is straightforward to verify that, within the class of potential games, an equilibrium $\hat{x}$ is first-order non-degenerate if and only if it is quasi-strong (see Remark 9.4).

We now show that, if an equilibrium $\hat{x}$ is quasi-strong, then $\hat{x}$ is regular if and only if $\hat{x}$ is second-order non-degenerate. Assume henceforth that $\hat{x}$ is a quasi-strong equilibrium.

Our goal is to show that $D_x\tilde{F}(\hat{x}, u)$ is invertible if and only if $\tilde{\mathbf{H}}(\hat{x})$ is invertible (see (9.7)). By (9.12), this is equivalent to showing that $D_x\tilde{F}(\hat{x}, u)$ is invertible if and only if $D_{x_m}F(\hat{x}, u)$ is invertible. With this end in mind, we begin by considering the behavior of the component maps of $\tilde{F}$ and F in two important cases.

Case 1: Suppose (i, k) is such that $k \in \{1, \dots, \gamma_i - 1\}$. Note that in this case we have $\hat{x}_i^k > 0$. Differentiating (9.30) with respect to x_j^ℓ, $(j, \ell) \neq (i, k)$ we get

$$\frac{\partial \tilde{F}_i^k(\hat{x}, u)}{\partial x_j^\ell} = \left(\frac{\partial T_i^1(\hat{x}_i)}{\partial x_j^\ell}\right) \hat{x}_i^k F_i^k(\hat{x}, u) + T_i^1(\hat{x}_i)\hat{x}_i^k \frac{\partial F_i^k(\hat{x}, u)}{\partial x_j^\ell}$$

Since $k \in \{1, \dots, \gamma_i - 1\}$, we have $F_i^k(\hat{x}, u) = 0$ and hence,

$$\frac{\partial \tilde{F}_i^k(\hat{x}, u)}{\partial x_j^\ell} = T_i^1(\hat{x}_i)\hat{x}_i^k \frac{\partial F_i^k(\hat{x}, u)}{\partial x_j^\ell} \tag{9.31}$$

for $(j, \ell) \neq (i, k)$.

Differentiating (9.30) with respect to x_i^k we get

$$\frac{\partial \tilde{F}_i^k(\hat{x},u)}{\partial x_i^k} = -\hat{x}_i^k F_i^k(\hat{x},u) + T_i^1(\hat{x}_i)F_i^k(\hat{x},u) + \hat{x}_i^k T_i^1(\hat{x}_i)\frac{\partial F_i^k(\hat{x},u)}{\partial x_i^k} \tag{9.32}$$

By our choice of (i,k) we have $F_i^k(\hat{x},u)=0$. Also note that $F_i^k(x,u)$ does not depend on x_i^k (see (9.10)), and hence $\frac{\partial F_i^k(\hat{x},u)}{\partial x_i^k}=0$. By (9.32), this implies that $\frac{\partial \tilde{F}_i^k(\hat{x},u)}{\partial x_i^k}=0$. But we just showed $\frac{\partial F_i^k(\hat{x},u)}{\partial x_i^k}=0$, hence

$$\frac{\partial \tilde{F}_i^k(\hat{x},u)}{\partial x_i^k} = \frac{\partial F_i^k(\hat{x},u)}{\partial x_i^k} = 0. \tag{9.33}$$

Together, (9.31) and (9.33) imply that for each (i,k) such that $k=1,\ldots,\gamma_i-1$ we have $D_{x_m}\tilde{F}_i^k(\hat{x},u) = T_i^1(\hat{x}_i)\hat{x}_i^k D_{x_m}F_i^k(\hat{x},u)$, where $\hat{x}_i^k>0$ and $T_i^1(\hat{x}_i)>0$. This implies that

$$D_{x_m}\left(\tilde{F}_i^k(\hat{x},u)\right)_{\substack{i=1,\ldots,N\\ k=1,\ldots,\gamma_i-1}} \text{ is nonsingular} \qquad \iff \qquad D_{x_m}F(\hat{x},u) \text{ is non-singular,} \tag{9.34}$$

where $F(x,u)$ is defined in (9.11).

Case 2: Suppose (i,k) is such that $k\in\{\gamma_i,\ldots,K_i-1\}$. Note that in this case we have $\hat{x}_i^k=0$. Differentiating (9.30) with respect to x_j^ℓ, $(j,\ell)\neq(i,k)$ we get

$$\frac{\partial \tilde{F}_i^k(\hat{x},u)}{\partial x_j^\ell} = T_i^1(\hat{x}_i)\hat{x}_i^k\frac{\partial F_i^k(\hat{x},u)}{\partial x_j^\ell} = 0, \tag{9.35}$$

where the equality to zero holds since $\hat{x}_i^k=0$. Note in particular that this implies that $D_{x_m}\tilde{F}_i^k(\hat{x},u) = 0$.

If we differentiate (9.30) with respect to x_i^k we get

$$\begin{aligned}\frac{\partial \tilde{F}_i^k(\hat{x},u)}{\partial x_i^k} &= T_i^1(\hat{x}_i)F_i^k(\hat{x},u) + \underbrace{\hat{x}_i^k T_i^1(\hat{x}_i)\frac{\partial F_i^k(\hat{x},u)}{\partial x_i^k}}_{=0}\\ &= T_i^1(\hat{x}_i)[U(y_i^{k+1},\hat{x}_{-i}) - U(y_i^1,\hat{x}_{-i})],\end{aligned}$$

where the equality to zero in the first line holds since $\hat{x}_i^k=0$, and the second line follows from the definition of F_i^k (see (9.10)).

Since $\hat{x}$ is a first-order non-degenerate equilibrium, $U(y_i^1,\hat{x}_{-i}) > U(y_i^{\ell+1},\hat{x}_{-i})$ for all $\ell=\gamma_i,\ldots,K_i-1$. Also, by our ordering of Y_i we have $T_i^1(\hat{x}_i)>0$. Hence, $T_i^1(\hat{x}_i)[U(y_i^{k+1},\hat{x}_{-i}) -$

$U(y_i^1, \hat{x}_{-i})] < 0$. This, along with (9.35), implies that $D_{x_p}\left(\tilde{F}_i^k(\hat{x},u)\right)_{\substack{i=1,\dots,N\\ k=\gamma_i,\dots,K_i-1}}$ is a diagonal matrix with non-zero diagonal.

We now consider the Jacobian $D_x\tilde{F}(\hat{x},u)$. This may be expressed as

$$D_x\tilde{F}(\hat{x},m) = \begin{pmatrix} D_{x_m}\left(\tilde{F}_i^k(\hat{x},u)\right)_{\substack{i=1,\dots,N\\ k=1,\dots,\gamma_i-1}} & D_{x_p}\left(\tilde{F}_i^k(\hat{x},u)\right)_{\substack{i=1,\dots,N\\ k=1,\dots,\gamma_i-1}} \\ D_{x_m}\left(\tilde{F}_i^k(\hat{x},u)\right)_{\substack{i=1,\dots,N\\ k=\gamma_i,\dots,K_i-1}} & D_{x_p}\left(\tilde{F}_i^k(\hat{x},u)\right)_{\substack{i=1,\dots,N\\ k=\gamma_i,\dots,K_i-1}} \end{pmatrix}$$

By the above we see that this matrix has the form

$$D_x\tilde{F}(\hat{x},m) = \begin{pmatrix} D_{x_m}\left(\tilde{F}_i^k(\hat{x},u)\right)_{\substack{i=1,\dots,N\\ k=1,\dots,\gamma_i-1}} & \mathbf{M} \\ 0 & \mathbf{D} \end{pmatrix},$$

where $\mathbf{M} \in \mathbb{R}^{\gamma\times(\kappa-\gamma)}$ and $\mathbf{D} \in \mathbb{R}^{(\kappa-\gamma)\times(\kappa-\gamma)}$ is an invertible diagonal matrix. Given this block form we see that $D_x\tilde{F}(\hat{x},m)$ is invertible if and only if $D_{x_m}\left(\tilde{F}_i^k(\hat{x},u)\right)_{\substack{i=1,\dots,N\\ k=1,\dots,\gamma_i-1}}$ is invertible. By (9.34) we see that $D_x\tilde{F}(\hat{x},u)$ is invertible if and only if $D_{x_m}F(\hat{x},u)$ is invertible, which proves the desired result.

$\square$

Chapter 10

Myopic Best-Response Dynamics in Potential Games

10.1 Introduction

In this chapter we study fundamental convergence properties of continuous-time MBR dynamics. We focus our attention on the important class of multi-agent games known as potential games (see Section 2.1.1). This chapter has three main contributions: (i) We show that in almost all potential games and for almost all initial conditions, FP converges to a pure-strategy equilibrium, (ii) We characterize the rate of convergence of FP in potential games, and (iii) We characterize the rate of convergence of ECFP in potential games.

We briefly discuss each of these contributions below.

10.1.1 Fictitious Play in Potential Games

The set of NE may be subdivided into pure-strategy (deterministic) NE and mixed-strategy (probabilistic) NE. Mixed-strategy NE can be problematic for a number of reasons [59]. In engineering applications, mixed-strategy NE can be undesirable since they are nondeterministic, have suboptimal expected utility, and do not always have clear physical meaning [60]. Thus, in applications, preference is often given to algorithms that are guaranteed to converge to a pure-strategy NE [60–63].

In potential games, pure-strategy NE are guaranteed to exist. It has been shown that FP will converge to the *set* of NE in potential games [28,34,38]; however, it is well-known that FP need not converge to a *pure-strategy* equilibrium in such games [34,74]. In fact, this deficiency was first noted in the paper where FP was first shown to converge to the set of NE in potential games (see [34], Remark (2)). In this chapter we attempt to redeem FP somewhat in this regard. We show that, while convergence to mixed-strategy equilibria is possible, it is an exceptional occurrence. The following theorem is the first main result of this chapter.

Theorem 10.1. *In almost all potential games, and for almost all initial conditions, FP converges to a pure-strategy NE.*

We now discuss our strategy for proving Theorem 10.1.

FP in Potential Games: Proof Strategy

In Chapter 9 we introduced the notion of a non-degenerate potential game and showed that almost every potential game is non-degenerate. The basic strategy we will use to prove Theorem 10.1 is to leverage two noteworthy properties satisfied in every non-degenerate potential game:

1. The FP vector field cannot concentrate mass in finite time, meaning it cannot map a set of positive (Lebesgue) measure to a set of zero measure in finite time (see Section 10.4.1).

2. In a neighborhood of an *interior* Nash equilibrium (i.e., a *completely mixed equilibrium*), the magnitude of the time derivative of the potential along paths grows linearly in the distance to the Nash equilibrium, while the value of the potential varies only quadratically; that is,

$$\frac{d}{dt}U(\mathbf{x}(t)) \geq d(\mathbf{x}(t), x^*) \geq \sqrt{|U(\mathbf{x}(t)) - U(x^*)|}.$$

   where U denotes the potential function and x^* is the equilibrium point.

Using Markov's inequality, property 2 immediately implies that if a path converges to an interior NE then it must do so in finite time (see Section 10.4.2). Hence, properties 1 and 2 together imply that the set of points from which FP converges to an interior NE must have Lebesgue measure zero.

In order to handle mixed NE that are *not* in the interior of the strategy space (i.e., *incompletely mixed equilibria*), we consider a projection that maps incompletely mixed equilibria to the *interior* of the strategy space of a lower dimensional game. Using the techniques described above, we are then able to handle completely and incompletely mixed equilibria in a unified manner.

In particular, we see that the set of points from which FP converges to the set of mixed-strategy NE has Lebesgue measure zero. Since any FP path must converge to a NE [28], this implies Theorem 10.1.

FP in Potential Games: Intuition and Comparison with Classical Techniques

It is well known that FP converges to the set of NE in potential games [28]. Hence, to prove Theorem 10.1 we need only prove that each mixed-strategy NE can only be reached from a set of measure zero.

Given a classical ODE, such a result could be proved via linearization. In particular, given some equilibrium point, consider the linearization of the dynamics near the equilibrium point. Assuming all eigenvalues associated with the linearized system are non-zero, the dimension of the stable manifold (i.e., the set of initial conditions from which the equilibrium can be reached) is equal to the dimension of the stable eigenspace of the linearized system [155].

In FP, the vector field is discontinuous—hence, it is not possible to linearize around an equilibrium point, and such classical techniques cannot be applied directly. However, to a certain extent, the gradient field of the potential function may be seen as an approximation of the FP vector field. In particular, it can be shown that the direction of the FP vector field approximately coincides with the direction of the gradient field of the potential function (see Lemma 10.3).

Unlike the FP vector field, the gradient field of the potential function *can* be linearized. In a non-degenerate game, any completely mixed-strategy NE is a non-degenerate saddle point of the potential function. Hence, at least one eigenvalue of the linearized system must lie in the right-half plane. This implies that, for the gradient dynamics of the potential function, the stable manifold associated with an equilibrium point has dimension at most $\kappa - 1$, where κ is the dimension of the strategy space.

Since the FP vector field approximates the gradient field of the potential function, intuition suggests that for FP dynamics, each mixed equilibrium should also admit a similar low-dimensional stable manifold.

While this provides an intuitive explanation for why one might expect Theorem 10.1 to hold, we did not use any such linearization arguments in the proof. We found that studying the rate of potential production near mixed equilibria (e.g., as discussed in the "proof strategy" section above) led to shorter and simpler proofs. However, we hypothesize that an alternate proof of Theorem 10.1 could be obtained by considering smooth approximations to FP and linearizing near mixed-strategy equilibria.

10.1.2 Rate of Convergence of FP in Potential Games

The second main contribution of this chapter is a characterization of the rate of convergence of FP in potential games.

The problem of characterizing the rate of convergence in FP is intimately related to the problem of characterizing the dynamics of FP near mixed-strategy equilibria. If an FP path *passes through* a mixed-strategy NE, it can rest there for an indeterminate amount of time before moving elsewhere ([27], Section 8). In this way, the mere existence of mixed-strategy equilibria presents a fundamental barrier to establishing general convergence rate estimates in FP.

In [27] it was conjectured that within the class of weighted potential games, the rate of convergence of every FP path is asymptotically exponential ([27], Conjecture 25). In Section 10.5, we partially resolve this conjecture by showing that it holds in almost all (exact) potential games. In particular, we show that it holds for almost all initial conditions within the class of non-degenerate potential games introduced in Chapter 9.

In addition to resolving this conjecture, our techniques are able to cast some light on the manner in which convergence occurs. We show that, for almost all games and almost all initial conditions, FP never passes through a mixed-strategy equilibrium. Hence, the problem of sojourning at mixed-strategy equilibria pointed out in [27] almost never occurs.

In future work we plan to study methods for sharpening this estimate using the techniques developed in this chapter.

10.1.3 Rate of Convergence of ECFP in Potential Games

The third main contribution of the chapter concerns the rate of convergence of ECFP in potential games.

In Section 3.4.3 we showed that ECFP converges to the set of SNE (3.7). At a SNE strategy, players in the same grouping are required to use identical strategies, and such equilibria are often mixed, rather than pure. For example, within the class of congestion games, Assumption 6.19 gives a condition under which any consensus equilibrium will be a mixed strategy.

As the third main contribution of this chapter, in Section 10.6 we show that if ECFP converges to a *completely* mixed-strategy equilibrium in a non-degenerate potential game satisfying Assumptions 3.1–3.2, then convergence occurs in finite time.

We note that, in this dissertation we do not consider the rate of convergence of ECFP to *incompletely* mixed equilibria. Studying convergence to incompletely mixed equilibria is fundamentally more delicate than studying convergence to completely mixed equilibria (see Section 10.3). We only consider the simpler case of convergence of ECFP to completely mixed equilibria. However, we conjecture that using the projection techniques developed in Section 10.3, it can be shown that if ECFP converges to any mixed equilibrium (not necessarily completely mixed), then it converges in finite time.

10.1.4 Related Work

A related result for two player games has been shown in [92], where it was demonstrated that FP almost never converges cyclically to a mixed-strategy equilibrium in which both players use more than two pure-strategies.

Continuous-time best-response dynamics similar to those we consider here have been studied in various works, including [27, 28, 64, 91]. These papers study a variety of convergence properties for FP and replicator dynamics in a wide class of games, but do not consider the question of generic convergence to pure-strategy equilibria.

The work [27] studies the rate of convergence of FP in zero-sum games. It is proven that FP converges to the set of NE in (weighted) potential games, and it is conjectured that the rate of convergence in this case is exponential.

In [153] it is shown that in potential games, FP can take an arbitrarily long time to converge to a neighborhood of a pure-strategy equilibrium. This is accomplished by choosing a "nearly degenerate" game in which a FP path passes arbitrarily close to a some pure strategy that is "almost" a pure-strategy equilibrium. Using the tools developed in this dissertation, in future work we hope to show that this type of slow convergence is exceptional in the sense that if one removes a small set of "almost degenerate" potential games, then the time needed to converge to a neighborhood of a pure-strategy NE can be uniformly bounded.

The remainder of the chapter is organized as follows. Section 10.2 discusses notation to be used throughout the chapter. Section 10.3 establishes the two key inequalities used to prove Theorem 10.1. Section 10.4 proves Theorem 10.1. Section 10.5 studies the rate of convergence of FP in non-degenerate potential games. Section 10.6 studies the rate of convergence of ECFP to completely mixed equilibria.

10.2 Notation

Throughout the chapter we will find it convenient to use the notation introduced in Section 9.2. In particular, we will often work in the space X, defined Section 9.2, rather than in Δ^N. This will free us from being perpetually encumbered by the hyperplane constraints implicit in Δ_i, $i \in \mathcal{N}$, and Δ^N. In this context, a (continuous-time) fictitious play process is defined as follows (cf. Definition (2.4)).

Definition 10.2. *A mapping* $\mathbf{x} : \mathbb{R} \to X$ *is said to be a (continuous-time)* fictitious play process *with initial condition $x_0 \in X$ if $\mathbf{x}$ is an absolutely continuous mapping such that $\mathbf{x}(0) = x_0$ and*

$$\dot{\mathbf{x}} \in \mathrm{BR}(\mathbf{x}) - \mathbf{x} \tag{10.1}$$

holds for almost all $t \in \mathbb{R}$.

10.3 Potential Production Inequalities

In this section we prove two key inequalities ((10.3) and (10.4)) that are the backbone of our proof of Theorem 10.1.

We note that in proving Theorem 10.1 there is a fundamental dichotomy between studying completely mixed equilibria and incompletely mixed equilibria. Completely mixed equilibria lie in the interior of the strategy space. At these points the gradient of the potential function is zero and the Hessian is non-singular; local analysis of the dynamics is relatively easy. On the other hand, incompletely mixed equilibria necessarily lie on the boundary of X and the potential function may have a nonzero gradient at these points.[1] Analysis of the dynamics around these points is fundamentally more delicate.

In order to handle incompletely mixed equilibria we construct a nonlinear projection whose range is a lower dimensional game in which the image of the equilibrium under consideration is completely mixed. This allows us to handle both types of mixed equilibria in a unified manner.

10.3.1 Projection to a Lower-Dimensional Game

Let x^* be a mixed equilibrium.[2]

For $x \in X_i$, let $\mathrm{carr}_i(x_i)$ be as defined in (9.5), and for $x \in X$, let $\mathrm{carr}(x)$ be as defined in the discussion following (9.5). Let $C_i := \mathrm{carr}_i(x_i^*)$, where x_i^* is the player-i component of x^*, let $C := C_1 \cup \cdots \cup C_N = \mathrm{carr}(x^*)$. Let $\gamma_i := |C_i|$ and assume that Y_i is ordered so that $\{y_i^1, \ldots, y_i^{\gamma_i}\} = C_i$; that is, the first γ_i strategies in Y_i are precisely the strategies in C_i. Let $\tilde{N} := \big|\{i \in \{1, \ldots, N\} : \gamma_i \geq 2\}\big|$, and assume that the player set is ordered so that $\gamma_i \geq 2$ for $i = 1, \ldots, \tilde{N}$. Since x^* is assumed to be a mixed-strategy equilibrium, we have $\tilde{N} \geq 1$.

Given an $x \in X$, we will frequently use the decomposition $x = (x_p, x_m)$, where $x_m :=$

[1] We note that in games that are first-order non-degenerate, the gradient is always non-zero at incompletely mixed equilibria.

[2] We note that x^* is assumed to be fixed throughout the section and many of the subsequently defined terms are implicitly dependent on x^*.

$(x_i^k)_{i=1,\dots,\tilde{N},\ k=1,\dots,\gamma_i-1}$ and x_p contains the remaining components of x.[3] Let $\gamma := \sum_{i=1}^N(\gamma_i - 1)$. Recalling that κ is the dimension of X (see (9.3)), note that for $x \in X$ we have $x \in \mathbb{R}^\kappa$, $x_m \in \mathbb{R}^\gamma$, and $x_p \in \mathbb{R}^{\kappa-\gamma}$.

The set of joint pure strategies Y may be expressed as an ordered set $Y = \{y^1, \dots, y^K\}$ where each element $y^\tau \in Y$, $\tau \in \{1, \dots, K\}$ is an N-tuple of strategies. For each pure strategy $y^\tau \in Y$, $\tau = 1, \dots, K$, let u^τ denote the pure-strategy potential associated with playing y^τ; that is, $u^\tau := u(y^\tau)$, where u is the pure form of the potential function defined in Section 9.2. A vector of *potential coefficients* $u = (u^\tau)_{\tau=1}^K$ is an element of $\mathbb{R}^K$.

Given a vector of potential coefficients $u \in \mathbb{R}^K$ and a strategy $x \in X$, let[4]

$$F_i^k(x,u) := \frac{\partial U(x)}{\partial x_i^k},$$

for $i = 1, \dots, \tilde{N},\ k = 1, \dots, \gamma_i - 1$, and let

$$F(x,u) := \Big(F_i^k(x,u)\Big)_{\substack{i=1,\dots,\tilde{N}\\ k=1,\dots,\gamma_i-1}} = \left(\frac{\partial U(x)}{\partial x_i^k}\right)_{\substack{i=1,\dots,\tilde{N}\\ k=1,\dots,\gamma_i-1}}.$$

Differentiating (9.4) we see that at the equilibrium x^* we have $\frac{\partial U(x^*)}{\partial x_i^k} = 0$ for $i = 1, \dots, \tilde{N}$, $k = 1, \dots, \gamma_i - 1$ (see Lemma 10.28 in appendix), or equivalently,

$$F(x^*, u) = F(x_p^*, x_m^*, u) = 0.$$

By Definition 9.2, the (mixed) equilibrium x^* is completely mixed if $\gamma = \kappa$, and is incompletely mixed otherwise . Suppose $\gamma < \kappa$ so that x^* is incompletely mixed. Let $\mathbf{J}(x) := D_{x_m}F(x_p, x_m, u)$ and let $\tilde{\mathbf{H}}(x)$ be as defined in (9.7). Note that by (9.12) we have $\mathbf{J}(x^*) = \tilde{\mathbf{H}}(x^*)$. Since Γ is assumed to be a non-degenerate game, $\mathbf{J}(x^*)$ is invertible. By the implicit function theorem, there exists a function $g : \mathcal{D}(g) \to \mathbb{R}^\gamma$ such that $F(x_p, g(x_p), u) = 0$ for all x_p in a neighborhood of x_p^*, where $\mathcal{D}(g) \subset \mathbb{R}^{\kappa-\gamma}$ denotes the domain of g, $x_p^* \in \mathcal{D}(g)$, and $\mathcal{D}(g)$ is open.

The graph of g is given by

$$\mathrm{Graph}(g) := \{x \in X :\ x = (x_p, x_m),\ x_p \in \mathcal{D}(g),\ x_m = g(x_p)\}.$$

[3] The subscript in x_m is suggestive of "mixed-strategy components" and the subscript in x_p is suggestive of "pure-strategy components".

[4] We note that the functions F_i^k and F defined below are identical to those defined in (9.10) and (9.11). In order to keep the chapter relatively self contained, we redefine these terms here.

Note that Graph(g) is a smooth manifold with Hausdorff dimension $(\kappa - \gamma)$ [152]. An intuitive interpretation of Graph(g) is given in Remark 10.6.

If Γ is a non-degenerate potential game then, using the multilinearity of U, we see that $\gamma \geq 2$ (see Lemma 10.25 in appendix). This implies that

$$\text{Graph}(g) \text{ has Hausdorff dimension at most } (\kappa - 2).$$

Let $\Omega := \Omega_C$, where Ω_C is defined in (9.6), denote the face of X containing x^*. Define the mapping $\tilde{\mathcal{P}} : \mathcal{D}(\tilde{\mathcal{P}}) \to \Omega$, with domain $\mathcal{D}(\tilde{\mathcal{P}}) := \{x = (x_p, x_m) \in X : \ x_p \in \mathcal{D}(g)\}$, as follows. If x^* is completely mixed then let $\tilde{\mathcal{P}}(x) := x$ be the identity. Otherwise, let

$$\tilde{\mathcal{P}}(x) := x^* + (x - (x_p, g(x_p))). \tag{10.2}$$

Let $\tilde{\mathcal{P}}_i^k(x)$ be the (i,k)-th coordinate map of $\tilde{\mathcal{P}}$, so that $\tilde{\mathcal{P}} = (\tilde{\mathcal{P}}_i^k)_{\substack{i=1,\dots,N \\ k=1,\dots,K_i-1}}$. Following the definitions, it is simple to verify that for $x \in \mathcal{D}(\tilde{\mathcal{P}})$ we have $\tilde{\mathcal{P}}_i^k(x) = 0$ for all (i,k) with $k \geq \gamma_i$, and hence $\tilde{\mathcal{P}}$ indeed maps into Ω.

Let $\tilde{X}_i := \{\tilde{x}_i \in \mathbb{R}^{\gamma_i - 1} : \ \tilde{x}_i^k \geq 0, k = 1, \dots, \gamma_i - 1, \ \sum_{k=1}^{\gamma_i - 1} \tilde{x}_i^k \leq 1\}$, $i = 1, \dots, \tilde{N}$, and let $\tilde{X} := \tilde{X}_1 \times \cdots \times \tilde{X}_{\tilde{N}}$. Let $\mathcal{P} : \mathcal{D}(\mathcal{P}) \to \tilde{X}$ with domain $\mathcal{D}(\mathcal{P}) = \mathcal{D}(\tilde{\mathcal{P}}) \subset X$ be given by

$$\mathcal{P} := (\tilde{\mathcal{P}}_i^k)_{i=1,\dots,\tilde{N}, \ k=1,\dots,\gamma_i - 1}.$$

Note that $\mathcal{P}$ contains the components of $\tilde{\mathcal{P}}$ not constrained to zero. As we will see in the following section, $\mathcal{P}$ may be interpreted as a projection into a lower dimensional game in which $\mathcal{P}(x^*)$ is a completely mixed equilibrium.

10.3.2 Inequalities

Let $\tilde{U} : \tilde{X} \to \mathbb{R}$ be given by

$$\tilde{U}(\tilde{x}) := U(x_p^*, \tilde{x}),$$

where $x^* = (x_p^*, x_m^*)$ is the mixed equilibrium fixed in the beginning of the section. Let $\tilde{\Gamma}$ be a potential game with player set $\{1, \dots, \tilde{N}\}$, mixed-strategy space $\tilde{X}_i$, $i = 1, \dots, \tilde{N}$, and potential function $\tilde{U}$. By construction, $\mathcal{P}(x^*)$ is a completely mixed equilibrium of $\tilde{\Gamma}$. Moreover, by the definition of a non-degenerate equilibrium, the Hessian of $\tilde{U}$ is invertible at $\mathcal{P}(x^*)$.

We are interested in studying the projection $\mathcal{P}(\mathbf{x}(t))$ of a FP process into the lower dimensional game $\tilde{\Gamma}$.[5] We wish to show that the following two inequalities hold:

(i) For x in a neighborhood of x^*

$$\left|\tilde{U}(\mathcal{P}(x^*)) - \tilde{U}(\mathcal{P}(x))\right| \leq c_1 d^2(\mathcal{P}(x), \mathcal{P}(x^*)), \tag{10.3}$$

for some constant $c_1 > 0$.

(ii) Suppose $(\mathbf{x}(t))_{t\geq 0}$ is a FP process. For $\mathbf{x}(t)$ residing in a neighborhood of x^*

$$\frac{d}{dt}\tilde{U}(\mathcal{P}(\mathbf{x}(t))) \geq c_2 d(\mathcal{P}(\mathbf{x}(t)), \mathcal{P}(x^*)), \tag{10.4}$$

for some constant $c_2 > 0$.

The first inequality follows from Taylor's theorem and the fact that $\nabla\tilde{U}(\mathcal{P}(x^*)) = 0$. The following section is devoted to proving (10.4).[6]

10.3.3 Proving the Differential Inequality

We begin with Lemma 10.3 which shows—roughly speaking—that within the interior of the action space, the FP vector field approximates the gradient field of the potential function.

The following definitions are useful in the lemma. For $B \subseteq X$, let $P_{X_i}(B) := \{x_i \in X_i : (x_i, x_{-i}) \in B \text{ for some } x_{-i} \in X_{-i}\}$ be the projection of B onto X_i. Given an $x_i \in X_i$, let

$$d(x_i, \partial X_i) := \min\{x_i^1, \ldots, x_i^{K_i-1}, 1 - \sum_{k=1}^{K_i-1} x_i^k\}$$

denote the distance from x_i to the boundary of X_i. Let

$$d(P_{X_i}(B), \partial X_i) := \inf_{x_i \in P_{X_i}(B)} d(x_i, \partial X_i)$$

denote the distance between the set $P_{X_i}(B)$ and the boundary of X_i.

Since we will eventually be interested in studying a lower-dimensional game derived from Γ, in the lemma we consider an alternative game $\hat{\Gamma}$ of arbitrary size.

[5] In the lower dimensional game $\tilde{\Gamma}$, the dynamics of the projected process are not precisely fictitious play dynamics. However, the behave *nearly* like fictitious play dynamics, which is what allows us to establish these inequalities.

[6] We note that when we write these inequalities, we mean they are satisfied in an integrated sense (e.g., as used in (10.24)–(10.25)). In this section, we treat all of these as pointwise inequalities. A rigorous argument could be constructed using the chain rule in Sobolev spaces (see, for example, [156]).

Lemma 10.3. *Let $\hat{\Gamma}$ be a potential game with player set $\{1,\ldots,\hat{N}\}$, action sets $\hat{Y}_i$, $i=1,\ldots,\hat{N}$, with cardinality $\hat{K}_i := |\hat{Y}_i|$, and potential function $\hat{U}$. Let $\hat{X} = \hat{X}_1 \times \cdots \times \hat{X}_{\hat{N}}$ denote the mixed strategy space.*

Let $B \subset \hat{X}$ and fix $i \in \{1,\ldots,\hat{N}\}$. Then for all $x \in B$ there holds

$$z_i \cdot \nabla_{x_i}\hat{U}(x) \geq c\|\nabla_{x_i}\hat{U}(x)\|_1, \qquad \forall\ z_i \in \mathrm{BR}_i(x_{-i}) - x_i$$

where the constant c is given by $c = d(P_{X_i}(B), \partial\hat{X}_i)$.

Proof. Let $x \in B$. If $\|\nabla_{x_i}\hat{U}(x)\|_1 = 0$, then $\nabla_{x_i}\hat{U}(x) = 0$, and the inequality is trivially satisfied. Suppose from now on that $\|\nabla_{x_i}\hat{U}(x)\|_1 > 0$.

Without loss of generality, assume that Y_i is ordered so that

$$y_i^1 \in \mathrm{BR}_i(x_{-i}). \tag{10.5}$$

Differentiating (9.4) we find that[7]

$$\frac{\partial \hat{U}(x)}{\partial x_i^k} = \hat{U}(y_i^{k+1}, x_{-i}) - \hat{U}(y_i^1, x_{-i}). \tag{10.6}$$

Together with (10.5), this implies that for $k = 1,\ldots,\hat{K}_i - 1$ we have

$$y_i^{k+1} \in \mathrm{BR}_i(x_{-i}) \iff \frac{\partial \hat{U}(x)}{\partial x_i^k} = 0. \tag{10.7}$$

Using the multlinearity of $\hat{U}$ we see that if $\xi_i \in \mathrm{BR}_i(x_{-i})$ and $\xi_i^k > 0$ then $y_i^{k+1} \in \mathrm{BR}_i(x_{-i})$. But, by (10.7) this implies that if $\xi_i \in \mathrm{BR}_i(x_{-i})$ and $\xi_i^k > 0$ then $\frac{\partial \hat{U}(x)}{\partial x_i^k} = 0$. Noting that any $\xi_i \in \mathrm{BR}_i(x_{-i})$ is necessarily coordinatewise nonnegative, this gives

$$(\xi_i - x_i)\cdot\nabla_{x_i}\hat{U}(x) = \underbrace{\sum_{k=1}^{\hat{K}_i-1} \xi_i^k \frac{\partial \hat{U}(x)}{\partial x_i^k}}_{=0} - \sum_{k=1}^{\hat{K}_i-1} x_i^k \frac{\partial \hat{U}(x)}{\partial x_i^k}, \qquad \xi_i \in \mathrm{BR}_i(x_{-i}) \tag{10.8}$$

Since we assume $x \in B$, we have $x_i^k \geq d(P_{\hat{X}_i}(B), \partial\hat{X}_i)$, for all $k = 1,\ldots,\hat{K}_i - 1$. Since we assume

[7]Note that the domain of (expected) potential function $\hat{U}$ may be trivially extended to an open neighborhood around $\hat{X}$ (see Section 9.2). Using this extension we see that the derivative is well defined for x lying on the boundary of $\hat{X}$.

$y_i^1 \in \mathrm{BR}_i(x_{-i})$, from (10.6) we get that $\frac{\partial \hat{U}(x)}{\partial x_i^k} \le 0$ for all $k = 1, \ldots, \hat{K}_i - 1$. Substituting into (10.8), this gives

$$(\xi_i - x_i) \cdot \nabla_{x_i} \hat{U}(x) \ge d(P_{\hat{X}_i}(B), \partial \hat{X}_i) \sum_{k=1}^{\hat{K}_i - 1} \left(-\frac{\partial \hat{U}(x)}{\partial x_i^k} \right), \qquad \xi_i \in \mathrm{BR}_i(x_{-i}).$$

But since $\frac{\partial \hat{U}(x)}{\partial x_i^k} \le 0$ for all k we have $\sum_{k=1}^{\hat{K}_i - 1} \left(-\frac{\partial \hat{U}(x)}{\partial x_i^k} \right) = \|\nabla_{x_i} \hat{U}(x)\|_1$, and hence

$$(\xi_i - x_i) \cdot \nabla_{x_i} \hat{U}(x) \ge d(P_{\hat{X}_i}(B), \partial \hat{X}_i) \|\nabla_{x_i} \hat{U}(x)\|_1, \qquad \xi_i \in \mathrm{BR}_i(x_{-i}),$$

which is the desired result. □

Remark 10.4. *Since the space X_i in Lemma 10.3 is finite dimensional, given any norm $\|\cdot\|$, there exists a constant $\tilde{c} > 0$ such that*

$$z_i \cdot \nabla_{x_i} U(x) \ge c \|\nabla_{x_i} U(x)\|, \qquad \forall\ z_i \in \mathrm{BR}_i(x_{-i}) - x_i$$

with $c = \tilde{c} d(P_{X_i}(B), \partial X_i)$.

For each $x = (x_p, x_m) \in X$ near to x^*, the following lemma allows us to define an additional lower dimensional game Γ_{x_p} associated with x_p in which the best-response set is closely related to the best response set for the original game Γ. The lemma is a straightforward consequence of the definition of the best response correspondence and the continuity of ∇U.

Lemma 10.5. *For x in a neighborhood of x^*, the best response set satisfies*

$$\mathrm{BR}_i(x_{-i}) \subseteq \mathrm{BR}_i(x^*_{-i}), \qquad \forall\ i = 1, \ldots, \tilde{N}.$$

Given any $x = (x_p, x_m) \in X$ we define $\tilde{U}_{x_p} : \tilde{X} \to \mathbb{R}$ and $\widetilde{BR}_{x_p,i}$: $\tilde{X}_{-i} \rightrightarrows \tilde{X}_i$ as follows. For $\tilde{x} \in \tilde{X}$ let

$$\tilde{U}_{x_p}(\tilde{x}) := U(x_p, \tilde{x}),$$

and for $\tilde{x}_{-i} \in \tilde{X}_{-i}$ let

$$\widetilde{BR}_{x_p,i}\ (\tilde{x}_{-i}) := \arg\max_{\tilde{x}_i \in \tilde{X}_i} \tilde{U}_{x_p}(\tilde{x}_i, \tilde{x}_{-i})$$

Let Γ_{x_p} be the potential game with player set $\{1, \ldots, \tilde{N}\}$, mixed strategy space $\tilde{X}$ and potential function $\tilde{U}_{x_p}$. Note that since U is continuous and $\tilde{X}$ is compact, $\tilde{U}_{x_p}$ converges uniformly to $\tilde{U}_{x_p^*} =: \tilde{U}$ as $x_p \to x_p^*$. In this sense the game Γ_{x_p} can be seen as converging to $\tilde{\Gamma}$ as $x_p \to x_p^*$.

Remark 10.6. *The function g defined in Section 10.3.1 admits the following interpretation. Suppose we fix some $x_p = (x_i^k)_{i=1,\ldots,N,\ k=\gamma_i,\ldots,K_i-1}$. Then $g(x_p)$ is a completely mixed Nash equilibrium of Γ_{x_p}. Moreover, if we let $x_p \to x_p^*$, then the corresponding equilibrium of the reduced game Γ_{x_p} converges to x^*, i.e., $(x_p, g(x_p)) \to (x_p^*, g(x_p^*)) = x^*$, precisely along* Graph$(g)$.

Remark 10.7. *Suppose x^* is a first-order non-degenerate equilibrium. Using the multilinearity of U we see that for any $x \in X$ we have $\mathrm{carr}_i(x_i) \subseteq \mathrm{BR}_i(x_{-i})$. By Remark 9.4, at x^* we have $\mathrm{carr}_i(x_i^*) = \mathrm{BR}_i(x_{-i}^*)$. Due to the ordering we assumed on Y_i, this implies that $y_i^k \in \mathrm{BR}_i(x_{-i}^*) \iff 1 \leq k \leq \gamma_i$. Moving to the X domain, this means that if $\hat{x}_i \in \mathrm{BR}_i(x_{-i}^*)$, then $\hat{x}_i^k = 0$ for all $k = \gamma_i, \ldots, K_i - 1$. By Lemma 10.5, this implies that for all x in a neighborhood of x^* and for $\hat{x}_i \in \mathrm{BR}_i(x_{-i})$ we have $(\hat{x}_i^k)_{k=\gamma_i}^{K_i-1} = 0$.*

The following lemma extends the result of Lemma 10.3 so it applies in a useful way to the potential function $\tilde{U}$ under the projection $\mathcal{P}$.

Lemma 10.8. *There exists a constant $c > 0$ such that for all $x = (x_p, x_m)$ in a neighborhood of x^* and all $\eta \in \mathbb{R}^{\gamma_i - 1}$ with $\|\eta\|$ sufficiently small we have*

$$(z_i + \eta) \cdot \nabla_{x_i} \tilde{U}(\mathcal{P}(x)) \geq c \|\nabla_{x_i} \tilde{U}(\mathcal{P}(x))\|,$$

for all $z_i \in \widetilde{BR}_{i,x_p}([x_m]_{-i}) - [x_m]_i$, where $[x_m]_i := (x_i^k)_{k=1,\ldots,\gamma_i - 1}$ refers to the player-i component of x_m and $[x_m]_{-i}$ contains the components of x_m corresponding to the remaining players.

Proof. By construction, the projection $\mathcal{P}(x^*)$ maps x^* into the interior of $\tilde{X}$. Choose $\epsilon > 0$ such that the ball $B(\mathcal{P}(x^*), \epsilon) \subset \tilde{X}$ is separated by a positive distance from the boundary of $\tilde{X}$. Applying Lemma 10.3 (and Remark 10.4) to the game Γ_{x_p} we see that there exists a constant $c > 0$ such that for any $x = (x_p, x_m) \in X$ with $x_m \in B(\mathcal{P}(x^*), \epsilon)$ there holds

$$z_i \cdot \nabla_{x_i} \tilde{U}_{x_p}(x_m) \geq 4c \|\nabla_{x_i} \tilde{U}_{x_p}(x_m)\| \tag{10.9}$$

for all $z_i \in \widetilde{BR}_{x_p,i}([x_m]_{-i}) - [x_m]_i$. Note that the constant in Lemma 10.3 is only dependent on the distance from the set B (in this case, $B(\mathcal{P}(x^*), \epsilon)$) to the boundary of the strategy space (in this case, $\tilde{X}$), and is independent of the particular potential function under consideration—this permits the choice of $c > 0$ in (10.9) that holds uniformly for all x_p.

By the continuity of U and $\mathcal{P}$ we have

$$\left\| \frac{\nabla \tilde{U}_{x_p}(\mathcal{P}(x))}{\|\nabla \tilde{U}_{x_p}(\mathcal{P}(x))\|} - \frac{\nabla \tilde{U}(\mathcal{P}(x))}{\|\nabla \tilde{U}(\mathcal{P}(x))\|} \right\| < \frac{2c}{\sqrt{2}} \tag{10.10}$$

and $\mathcal{P}(x) \in B(x^*, \epsilon)$, for all x in a sufficiently small neighborhood of x^* and $x \notin \text{Graph}(g)$. Note that $\text{diam}\, \tilde{X}_i := \max_{x_i, x_i' \in \tilde{X}_i} \|x_i - x_i'\| = \sqrt{2}$, and hence, $\|z_i\| \leq \sqrt{2}$ for any $z_i \in \widetilde{BR}_{x_p,i}\,([x_m]_{-i}) - [x_m]_i$, $x_m \in \tilde{X}$. Thus, (10.9) and (10.10) give

$$z_i \cdot \nabla \tilde{U}(\mathcal{P}(x)) \geq 2c \|\nabla \tilde{U}(\mathcal{P}(x))\|,$$

for all $z_i \in \widetilde{BR}_{x_p,i}\,([x_m]_{-i}) - [x_m]_i$ and all x in a neighborhood of x^*, $x \notin \text{Graph}(g)$. As long as $\|\eta\| \leq c$, the desired result holds for x in a neighborhood of x^*, $x \notin \text{Graph}(g)$. But $x \in \text{Graph}(g) \implies \mathcal{P}(x) = \mathcal{P}(x^*) \implies \nabla_{x_i} \tilde{U}(\mathcal{P}(x)) = 0$, in which case the inequality is trivially satisfied. $\square$

Finally, the following lemma shows that the differential inequality (10.4) holds.

Lemma 10.9. *Let Γ be a non-degenerate potential game with mixed equilibrium x^*, and let $(\mathbf{x}(t))_{t\geq 0}$ be a FP process. Then the inequality* (10.4) *holds for $\mathbf{x}(t)$ in a neighborhood of x^*.*

Proof. Let

$$\mathbf{P}(x) := \left(\frac{\partial \tilde{\mathcal{P}}_i^k}{\partial x_j^\ell} \right)_{\substack{i=1,\ldots,\tilde{N}, k=1,\ldots,\gamma_i - 1 \\ j=1,\ldots,N,\ \ell=\gamma_i,\ldots,K_i-1}}$$

where the partial derivatives are evaluated at x. The Jacobian of $\tilde{\mathcal{P}}$ evaluated at x is given by

$$\left(\frac{\partial \tilde{\mathcal{P}}_i^k}{\partial x_j^\ell} \right)_{\substack{i,j=1,\ldots,N, \\ k,\ell=1,\ldots,K_i-1}} = \begin{pmatrix} I & \mathbf{P}(x) \\ 0 & 0 \end{pmatrix}.$$

Using the chain rule we may express the time derivative of the potential along the path $\tilde{\mathcal{P}}(\mathbf{x}(t))$ as

$$\frac{d}{dt} U(\tilde{\mathcal{P}}(\mathbf{x}(t))) = \nabla U(\tilde{\mathcal{P}}(\mathbf{x}(t))) \begin{pmatrix} I & \mathbf{P}(x) \\ 0 & 0 \end{pmatrix} \dot{x} = \nabla_{x_m} U(\tilde{\mathcal{P}}(\mathbf{x}(t)))(\mathbf{I} \;\; \mathbf{P}(x))\dot{x}.$$

For $i = 1, \ldots, \tilde{N}$, $k = 1, \ldots, \gamma_i - 1$ let $\eta_i^k(t) := \sum_{j=1}^N \sum_{\ell=\gamma_j}^{K_i-1} \frac{\partial \tilde{\mathcal{P}}_i^k}{\partial x_j^\ell} \dot{x}_i^\ell$, let $\eta_i(t) := (\eta_i^k(t))_{k=1}^{\gamma_i-1}$, and let

$\eta(t) = (\eta_i(t))_{i=1}^{\tilde{N}}$. Multiplying out the right two terms above we get

$$\frac{d}{dt}U(\tilde{\mathcal{P}}(\mathbf{x}(t))) = \nabla_{x_m}U(\tilde{\mathcal{P}}(\mathbf{x}(t)))\,(\dot{x}_m + \eta(t)) \tag{10.11}$$

By Lemma 10.5 and Remark 10.7, if we restrict $\mathbf{x}(t)$ to a sufficiently small neighborhood of x^* then for any $z_i = (z_i', z_i'') \in \mathrm{BR}_i(\mathbf{x}_{-i}(t))$, $z_i' = (z_i^k)_{k=1}^{\gamma_i-1}$, $z_i'' = (z_i^k)_{k=\gamma_i}^{K_i-1}$, we have $z_i' \in \widetilde{BR}_{x_p,i}\left([x_m]_{-i}\right)$ and $z_i'' = 0$. We note two important consequences of this:
(i) If we restrict $\mathbf{x}(t)$ to a sufficiently small neighborhood of x^* and note that
$U(\tilde{\mathcal{P}}(\mathbf{x}(t))) = \tilde{U}(\mathcal{P}(\mathbf{x}(t))$, then by (10.11) we have

$$\begin{aligned}\frac{d}{dt}\tilde{U}(\mathcal{P}(\mathbf{x}(t))) &= \nabla\tilde{U}(\mathcal{P}(\mathbf{x}(t)))\cdot\begin{pmatrix} z_1(t)+\eta_1(t)\\ \vdots \\ z_{\tilde{N}}(t)+\eta_{\tilde{N}}(t)\end{pmatrix} \\ &= \sum_{i=1}^{\tilde{N}} \nabla_{x_i}\tilde{U}(\mathcal{P}(\mathbf{x}(t)))\cdot(z_i(t)+\eta_i(t)),\end{aligned} \tag{10.12}$$

where $z_i(t) \in \widetilde{BR}_{x_p(t),i}\left([x_m(t)]_{-i}\right) - [x_m(t)]_i$.
(ii) We may force $\max_{i=1,\ldots,\tilde{N}}\|\eta_i\|$ to be arbitrarily small by restricting $\mathbf{x}(t)$ to a neighborhood of x^*.

Consequence (i) follows readily by using the definition of FP (10.1). To show consequence (ii), note that by (10.1) we have $\dot{\mathbf{x}}_i^k = z_i^k - x_i^k$ for all $i = 1,\ldots,N$, $k = 1,\ldots,K_i$, for some $z_i \in \mathrm{BR}_i(x_{-i})$. But, for x in a neighborhood of x^* and $k \geq \gamma_i$, we have shown above that $z_i^k = 0$, and hence $\dot{x}_i^k = -x_i^k$.[8] Due the ordering we assumed for Y_i, we have $[x^*]_i^k = 0$ for any (i,k) such that $k \geq \gamma_i$. Hence, $x_i^k \to 0$ as $x \to x^*$, for any (i,k) such that $k \geq \gamma_i$.

Furthermore, there exists a $c > 0$ such that $|\frac{\partial\tilde{\mathcal{P}}_i^k(x)}{\partial x_j^\ell}| < c$, $i = 1,\ldots,\tilde{N}$, $k = 1,\ldots,\gamma_i - 1$, $j = 1,\ldots,N$, $\ell \geq \gamma_j$ uniformly for x in a neighborhood of x^* (see Lemma 10.29 in appendix). By the definition of η_i, this implies that $\max_{i=1,\ldots,\tilde{N}}\|\eta_i\|$ may be made arbitrarily small by restricting $\mathbf{x}(t)$ to a sufficiently small neighborhood of x^*.

Now, let $\mathbf{x}(t)$ be restricted to a sufficiently small neighborhood of x^* so that $\|\eta_i(t)\|$ is small enough to apply Lemma 10.8 for each i. Applying Lemma 10.8 to (10.12) we get $\frac{d}{dt}\tilde{U}(\mathcal{P}(\mathbf{x}(t))) \geq \sum_{i=1}^{\tilde{N}} c\|\nabla_{x_i}\tilde{U}(\mathcal{P}(\mathbf{x}(t)))\|$ for $\mathbf{x}(t)$ in a neighborhood of x^*. By the equivalence of finite-dimensional norms, there exists a constant c_1 such that $\frac{d}{dt}\tilde{U}(\mathcal{P}(\mathbf{x}(t))) \geq c_1\|\nabla\tilde{U}(\mathcal{P}(\mathbf{x}(t)))\|$ for $\mathbf{x}(t)$ in a neigh-

[8] We note that this particular step depends crucially on the assumption of first-order non-degeneracy (see Remark 10.7).

borhood of x^*.

Since Γ is assumed to be (second-order) non-degenerate, $\mathcal{P}(x^*)$ is a non-degenerate critical point of $\tilde{U}$. By Lemma 10.30 (see appendix) there exists a constant c_2 such that $c_1\|\nabla\tilde{U}(\tilde{x})\| \geq c_2 d(\tilde{x},\mathcal{P}(x^*))$ for all $\tilde{x} \in \tilde{X}$ in a neighborhood of $\mathcal{P}(x^*)$. Since $\mathcal{P}$ is continuous we have $\frac{d}{dt}\tilde{U}(\mathcal{P}(\mathbf{x}(t))) \geq c_2 d(\mathcal{P}(\mathbf{x}(t)),\mathcal{P}(x^*))$, for $\mathbf{x}(t)$ in a neighborhood of x^*. □

10.4 Fictitious Play in Potential Games: Proof of Main Result

For each mixed equilibrium x^*, let the set $\Lambda(x^*) \subset X$ be defined as

$$\Lambda(x^*) := \begin{cases} \{x^*\} & \text{if } x^* \text{ is completely mixed,} \\ \text{Graph}(g) & \text{otherwise,} \end{cases}$$

where g is defined with respect to x^* as in Section 10.3.1.

In this section we will prove Theorem 10.1 in two steps. First, we will show that for each mixed equilibrium x^*, the set $\Lambda(x^*)$ can only be reached in finite time from an $\mathcal{L}^\kappa$-null set of initial conditions (see Proposition 10.10), where κ, defined in (9.3), is the dimension of X. Second, we will show that if a FP process converges to the set $\Lambda(x^*)$, then it must do so in finite time (see Proposition 10.17). Since $x^* \in \Lambda(x^*)$, Propositions 10.10 and 10.17 together show that that for any mixed equilibrium x^*, the set of initial conditions from which FP converges to x^* has $\mathcal{L}^\kappa$-measure zero.

By Theorem 9.1 we see that, in a non-degenerate potential game, the set of NE is finite. Hence, Propositions 10.10 and 10.17 imply that FP can only converge to *set* of mixed strategy equilibria from a $\mathcal{L}^\kappa$-null set of initial conditions. Since a FP process must converge to the set of NE in a potential game ([28], Theorem 5.5), this implies that Theorem 10.1 holds.

10.4.1 Finite-Time Convergence

The goal of this subsection is to prove the following proposition.

Proposition 10.10. *Let Γ be a non-degenerate game and let x^* be a mixed-strategy NE of Γ. The set $\Lambda(x^*)$ can only be reached by a FP process in finite time from a set of initial conditions with*

$\mathcal{L}^\kappa$*-measure zero. That is,*

$$\mathcal{L}^\kappa(\{x_0 \in X : \mathbf{x}(0) = x_0,\ \mathbf{x}(t) \textit{ is a FP process},$$
$$\mathbf{x}(t) \in \Lambda(x^*) \textit{ for some } t \in [0,\infty)\}) = 0.$$

Before proving the proposition we present some definitions and preliminary results. Let

$$\mathcal{I}_{i,k,\ell} := \{(x_i, x_{-i}) \in X :\ U(y_i^k, x_{-i}) = U(y_i^\ell, x_{-i})\}, \tag{10.13}$$

for $i = 1, \ldots, N$, $k, \ell = 1, \ldots, K_i$, $\ell \neq k$, be the set in which player i is indifferent between his k-th and ℓ-th actions.

If the game Γ is non-degenerate, then each $\mathcal{I}_{i,k,\ell}$ is the union of smooth surfaces with Hausdorff dimension at most $(k-1)$ (see Lemma 10.34 in appendix). In particular, for each $x \in \mathcal{I}_{i,k,\ell}$ there exists a vector $\nu \in \mathbb{R}^\kappa$ that is normal to $\mathcal{I}_{i,k,\ell}$ at x. We refer to the set $\mathcal{I}_{i,k,\ell}$ as an *indifference surface* of player i.

Let $\tilde{Q}$ be defined as the set of points where two or more indifference surfaces intersect and their normal vectors do not coincide. Furthermore, if an indifference surface $\mathcal{I}$ has a component $\hat{\mathcal{I}} \subseteq \mathcal{I}$ with Hausdorff dimension less than $\kappa - 1$, then we put any points where $\hat{\mathcal{I}}$ intersects with another decision surface into $\tilde{Q}$. Since each indifference surface is smooth with dimension at most $\kappa - 1$, $\tilde{Q}$ has Hausdorff dimension at most $\kappa - 2$. Let

$$Q := \tilde{Q} \cup \Lambda(x^*).$$

As shown in Section 10.3.1, if x^* is non-degenerate, then the set $\mathrm{Graph}(g)$ (and hence $\Lambda(x^*)$) has Hausdorff dimension at most $\kappa - 2$. Thus Q has Hausdorff dimension at most $\kappa - 2$.[9]

The FP vector field (see (10.1)) is given by $\mathrm{FP} : X \rightrightarrows X$, where

$$\mathrm{FP}(x) := \mathrm{BR}(x) - x. \tag{10.14}$$

[9]Proposition 10.10 can easily be generalized to say that any set $A \subset X$ such that $\mathrm{cl}\, A$ has Hausdorff dimension at most $\kappa - 2$, can only be reached in finite time from a set of $\mathcal{L}^\kappa$-measure zero by substituting A for $\Lambda(x^*)$ throughout the section.

Let

$$\mathcal{Z} := \big\{x \in X\backslash Q : \ x \in \mathcal{I}_{i,k,\ell} \ \text{ for some } i,k,\ell \text{ with normal } \nu, \qquad \text{and } \nu \cdot z = 0 \text{ for some } z \in \mathrm{FP}(x)\big\}. \tag{10.15}$$

Since each $\mathcal{I}_{i,k,\ell}$ has Hausdorff dimension $\kappa - 1$, $\mathcal{Z}$ has Hausdorff dimension at most $\kappa - 1$. We define the *relative boundary* of $\mathcal{Z}$, denoted here as $\partial\mathcal{Z}$ as follows. If $\mathcal{Z}$ has Hausdorff dimension $\kappa-2$ or less, then let $\partial\mathcal{Z} := \mathcal{Z}$. If $\mathcal{Z}$ has Hausdorff dimension $\kappa-1$ then it may be expressed as the union of a finite number of smooth $(\kappa-1)$-dimensional surfaces, denoted here as $(\mathcal{Z}_s)_{s=1}^{N_z}$, $1 \leq N_z < \infty$, and a component with Hausdorff dimension at most $\kappa-2$, denoted here as $\mathcal{Z}'$. That is, $\mathcal{Z} = (\bigcup_{s=1}^{N_z} \mathcal{Z}_s) \cup \mathcal{Z}'$. Each $\mathcal{Z}_s$, $s = 1, \ldots, N_z$ is contained in some indifference surface, which we denote here as $\mathcal{I}_s$. Define the relative interior of $\mathcal{Z}_s$ (with respect to $\mathcal{I}_s$) as $\mathrm{ri}\,\mathcal{Z}_s := \{x \in Z_s : \ \exists \epsilon > 0 \text{ s.t. } B(x,\epsilon) \cap \mathcal{I}_s \subset \mathcal{Z}_s\}$, and define the relative boundary of $\mathcal{Z}_s$ as $\partial\mathcal{Z}_s := \mathrm{cl}\,\mathcal{Z}_s \backslash \mathrm{ri}\,\mathcal{Z}_s$. We then define the relative boundary of $\mathcal{Z}$ as

$$\partial\mathcal{Z} := \left(\bigcup_{s=1}^{N_z} \partial\mathcal{Z}_s\right) \cup \mathcal{Z}'.$$

Note that $\partial\mathcal{Z}$ is a set with Hausdorff dimension at most $\kappa - 2$. By Lemma 10.37 in the appendix, the FP vector field is oriented tangentially along $\mathcal{Z}$, in the sense that for any $x \in \mathcal{Z}$ there holds $\nu \cdot y = 0$ for any vector ν normal to $\mathcal{Z}$ at x, and any $y \in \mathrm{FP}(x)$. This implies that FP paths can only enter or exit $\mathcal{Z}$ through $\partial\mathcal{Z}$.

Let

$$X^* := X\backslash\left(Q \cup \mathcal{Z}\right)$$

The following technical lemma will be used to show that FP is well posed within X^* (see Lemma 10.12). It is a consequence of the fact that the FP vector field can only have jumps that are tangential to indifference surfaces.

Lemma 10.11. *Suppose $x \in X^*$ is in some indifference surface $\mathcal{I}_{i,k,\ell}$. Then there exists a constant $c > 0$ and a vector ν that is normal to $\mathcal{I}_{i,k,\ell}$ at x, such that*

$$\nu \cdot z \geq c, \quad \forall\ z \in \mathrm{FP}(\tilde{x})$$

for all $\tilde{x} \in X^*$ in a neighborhood of x.

Proof. By the definition of $\mathcal{I}_{i,k,\ell}$, if $x \in \mathcal{I}_{i,k,\ell}$ then for all $\hat{x} \in X$ such that $\hat{x}_{-i} = x_{-i}$ we have $\hat{x} \in \mathcal{I}_{i,k,\ell}$. This implies that for any vector ν that is normal to $\mathcal{I}_{i,k,\ell}$, the (i,m)-th component of ν

must be zero for all $m = 1, \ldots, K_i - 1$.

Suppose that $x \in X^* \cap \mathcal{I}_{i,k,\ell}$. Since $x \notin Q$, there is a neighborhood of x in which no indifference surface intersects with $\mathcal{I}_{i,k,\ell}$. This implies that for $\tilde{x}$ within a neighborhood of x, $\mathrm{BR}_{-i}(\tilde{x}) = a_{-i}$ for some a_{-i} that is a vertex of X_{-i}.

Together, these two facts imply that for all $\tilde{x}$ in a neighborhood of x, we have $\nu \cdot z' = \nu \cdot z''$ for all $z' \in \mathrm{BR}(x)$, $z'' \in \mathrm{BR}(\tilde{x})$, for any vector ν that is normal to $\mathcal{I}_{i,k,\ell}$ at x. Since $x \notin \mathcal{Z}$, recalling the form of FP (10.14), this means we can choose a vector ν that is normal to $\mathcal{I}_{i,k,\ell}$ at x and a constant $c > 0$ such that $\nu \cdot z > c$ for $z \in \mathrm{FP}(\tilde{x})$ for all $\tilde{x}$ in a neighborhood of x. ☐

The following lemma gives a well-posedness result for FP inside X^*.

Lemma 10.12. *For any $x_0 \in X^*$, there exists a $T \in (0, \infty]$ and a unique absolutely-continuous function $\mathbf{x} : [0, T] \to X^*$, with $\mathbf{x}(0) = x_0$, solving the differential inclusion $\frac{d}{dt}\mathbf{x}(t) \in \mathrm{FP}(\mathbf{x}(t))$ for almost all t.*

Proof. If $x \in X^*$ is not on any indifference surface, then FP is single valued in a neighborhood of x, and (10.1) is (locally) a Lipschitz differential equation with unique local solution.

Suppose that $x_0 \in X^*$ is on an indifference surface $\mathcal{I}$. By Lemma 10.11 there exists a constant $c > 0$ such that for all $\tilde{x}$ in a neighborhood of x we have $\mathrm{FP}(\tilde{x}) \cdot \nu > c$, where ν is a normal vector to $\mathcal{I}$ at x. This implies that for $\delta > 0$ sufficiently small we have $\{t \in [-\delta, \delta] : \ \mathbf{x}(t) \in \mathcal{I}\} = \{0\}$. Furthermore, since $x \notin Q$, for $\delta > 0$ sufficiently small we have

$$\{t \in [-\delta, \delta] : \ \mathbf{x}(t) \in \mathcal{I}_{i,k,\ell}, \text{ for any } i, k, \ell\} = \{0\}. \tag{10.16}$$

Now, let $x_0 \in X^*$ and let $(\mathbf{x}(t))_{t \geq 0}$ and $(\mathbf{z}(t))_{t \geq 0}$ be two solutions to (10.1) with $\mathbf{x}(0) = \mathbf{z}(0) = x_0$. If $(\mathbf{x}(t))_{t \geq 0}$ never crosses an indifference surface, then the flow is always classical and the two solutions always coincide; i.e., $\mathbf{x}(t) = \mathbf{z}(t), \ t \geq 0$. Suppose that $(\mathbf{x}(t))_{t \geq 0}$ does cross an indifference surface and let $t^* \geq 0$ be first time when such a crossing occurs. For $t < t^*$, the flow is classical and we have $\mathbf{x}(t) = \mathbf{z}(t)$ for $t \leq t^*$.

By (10.16) we see that for $\delta > 0$ sufficiently small, $\mathbf{x}(t)$ is not in any indifference surface for $t \in [t^* - \delta, t^* + \delta] \backslash \{t^*\}$. Suppose that at time $t = t^* + \delta$ we have $\mathbf{x}(t) = \hat{x} \neq \hat{z} = \mathbf{z}(t)$. Let $(\tilde{\mathbf{x}}(\tau))_{\tau \geq 0}$ and $(\tilde{\mathbf{z}}(\tau))_{\tau \geq 0}$ be solutions to the time-reversed FP flow with $\tilde{\mathbf{x}}(0) = \hat{x}$ and $\tilde{\mathbf{z}}(0) = \hat{z}$.

Since $\hat{x} \neq \hat{z}$, and since the time-reversed flow is classical for $0 \leq \tau < \delta$ (in particular, of the form $\dot{x} = a + x$ for some constant a), we get $\tilde{\mathbf{x}}(\delta) \neq \tilde{\mathbf{z}}(\delta)$. But this is impossible because the paths $(\mathbf{x}(t))_{t \geq 0}$ and $(\mathbf{z}(t))_{t \geq 0}$ are absolutely continuous and we already established that $\tilde{\mathbf{x}}(\delta) = \mathbf{x}(t^*) = \mathbf{z}(t^*) = \tilde{\mathbf{z}}(\delta)$. ☐

As a matter of notation, we say that λ is a signed measure on $\mathbb{R}^\kappa$ if there exists a Radon measure μ on $\mathbb{R}^\kappa$ and a μ-measurable function $\sigma : \mathbb{R}^\kappa \to \{-1, 1\}$ such that

$$\lambda(K) = \int_K \sigma d\mu \tag{10.17}$$

for all compact sets $K \subset \mathbb{R}^\kappa$. When convenient, we write $\sigma\mu$ to denote the signed measure λ in (10.17).

Letting elements $x \in X$ be written componentwise as $(x_s)_{s=1}^\kappa$, we recall [152] that a function $u \in L^1(\Omega)$ (with $\Omega \subseteq \mathbb{R}^\kappa$, Ω open) is a function of bounded variation (i.e., a BV function) if there exist finite signed Radon measures $D_s u$ such that the integration by parts formula

$$\int_\Omega u \frac{\partial \phi}{\partial x_s}\, dx = -\int_\Omega \phi\, dD_s u \tag{10.18}$$

holds for all $\phi \in C_c^\infty(\Omega)$. The measure $D_s u$ is called the *weak*, or *distributional*, *partial derivative of* u with respect to x_s. We let $Du := (D_s u)_{s=1,\ldots,\kappa}$.

The measure Du can be uniquely decomposed into three parts [157]

$$Du = \nabla u \mathcal{L}^\kappa + Cu + Ju. \tag{10.19}$$

Here Ju is supported on a set J_u with Hausdorff dimension $\kappa - 1$, and Cu is singular with respect to $\mathcal{L}^\kappa$ and satisfies $Cu(E) = 0$ for all sets E with finite $\mathcal{H}^{\kappa-1}$ measure.

The L^1 function ∇u is analogous to a classical derivative, and in particular if u is differentiable on an open set V then $Du = \nabla u \mathcal{L}^\kappa$ on that set, with ∇u matching the classical derivative. Furthermore, if u jumps across a smooth $(\kappa-1)$-dimensional hypersurface, then for x on the hypersurface we have

$$Du = Ju = (u^+ - u^-)\nu d\mathcal{H}^{\kappa-1}, \tag{10.20}$$

where u^+ is the value of u on one side of the surface, u^- is the value on the other, and ν is the normal vector pointing from u^- to u^+ [157].

A vector-valued function $f \in L^1(\Omega : \mathbb{R}^\kappa)$ is a function of bounded variation if each of its components is also of bounded variation. Letting f be written componentwise as $f = (f^s)_{s=1}^\kappa$, we write $Df := (D_j f^s)_{j,s=i,\ldots,\kappa}$.

Next we define the divergence of a function $f \in L^1(\Omega : \mathbb{R}^\kappa)$, denoted by $D \cdot f$, as the measure

$$D \cdot f := \sum_{s=1}^{\kappa} D_s f^s.$$

Given a constant $c \in \mathbb{R}$, we say that $D \cdot f = c$ if $D \cdot f = \frac{dD\cdot f}{d\mathcal{L}^\kappa}\mathcal{L}^\kappa$, and $\frac{dD\cdot f}{d\mathcal{L}^\kappa} = c$, where $\frac{dD\cdot f}{d\mathcal{L}^\kappa}$ denotes the Radon-Nikodym derivative. The following lemma characterizes the divergence of the FP vector field. As a matter of notation, if a function $f : X \to X$ satisfies $f(x) \in \mathrm{FP}(x)$ for all $x \in X$ then we say $f \in \mathrm{FP}$.

Lemma 10.13. *For every $f \in \mathrm{FP}$, the vector field f satisfies $D \cdot f = -1$.*

The proof of this lemma follows from the fact that FP is piecewise linear, and any jumps in FP are tangential to indifference surfaces.

Proof. Suppose $f \in \mathrm{FP}$, and let f be written componentwise as $f = (f_i^k)_{i=1,\ldots,N,\ k=1,\ldots,K_i-1}$. Let $i \in \{1, \ldots, N\}$ and $k \in \{1, \ldots, K_i - 1\}$. Let $D_{j,\ell} f_i^k$ denote the weak partial derivative of f_i^k with respect to x_j^ℓ, $j = 1, \ldots, N$, $\ell = 1, \ldots, K_j - 1$, and let $Df_i^k = \left(D_{j,\ell} f_i^k\right)_{j=1,\ldots,N,\ \ell=1,\ldots,K_j-1}$. Let $Jf_i^k = \left(J_{j,\ell} f_i^k\right)_{j=1,\ldots,N,\ \ell=1,\ldots,K_j-1}$ denote the jump component associated with Df_i^k (see (10.19)).

The vector field f is piecewise linear. Breaking up f over regions in which it is linear we see that $\frac{dD\cdot f}{d\mathcal{L}^\kappa} = -1$. It remains to show that $D \cdot f$ has no singular component; i.e., under the decomposition (10.19), the measure $D \cdot f$ has zero Cantor component and zero jump component.

Since f_i^k is piecewise linear and only jumps on the set $\bigcup_{\ell=1,\ \ell\neq k}^{K_i} \mathcal{I}_{i,k,\ell}$ which has finite $\kappa - 1$ measure, f_i^k has no Cantor part; that is, $Cf_i^k = (C_{j,\ell} f_i^k)_{j=1,\ldots,N,\ \ell=1,\ldots,\gamma_j} = 0$ (see (10.19)). Hence, the singular component of $D \cdot f$, which we denote here as S, has no Cantor part and is given by $S := \sum_{i=1}^N \sum_{k=1}^{K_i-1} J_{i,k} f_i^k$.

Suppose that $x \in \mathcal{I}_{i,k,\ell}$ for some ℓ (recall $\ell \neq k$). Suppose ν is a vector that is normal to $\mathcal{I}_{i,k,\ell}$ at x. By the definition of $\mathcal{I}_{i,k,\ell}$, if $x \in \mathcal{I}_{i,k,\ell}$ then for all $\hat{x} \in X$ such that $\hat{x}_{-i} = x_{-i}$ we have $\hat{x} \in \mathcal{I}_{i,k,\ell}$. This implies that the (i,k)-th component of ν must be zero. Since $Jf_i^k = ((f_i^k)^+ - (f_i^k)^-)\nu\mathcal{H}^{\kappa-1}$ for x on $\bigcup_{\ell=1,\ell\neq k}^{K_i} \mathcal{I}_{i,k,\ell}$ (see (10.20)), taking the (i,k)-th component we get $J_{i,k} f_i^k \mathcal{H}^{\kappa-1} = 0$.

Since this is true for every pair (i,k) we see that $S = 0$, and hence $D \cdot f = -1$ in the interior of X. An identical argument holds on the boundary of X, and hence, $S = 0$ and $D \cdot f = -1$. $\square$

The following lemma shows that for sets $E \subseteq X^*$ with relatively smooth boundary, the surface integral of FP over the boundary of E is well defined.

Lemma 10.14. *Let E be a subset of X^* with piecewise smooth boundary. For any functions $f, g \in \mathrm{FP}$ we have*

$$\int_{\partial E} f \cdot \nu_E d\mathcal{H}^{\kappa-1} = \int_{\partial E} g \cdot \nu_E d\mathcal{H}^{\kappa-1} =: \int_{\partial E} \mathrm{FP} \cdot \nu_E d\mathcal{H}^{\kappa-1},$$

where ν_E denotes the outer normal vector of E.

Proof. Suppose $x \in X^*$ is not on any indifference surface $\mathcal{I}_{i,k,\ell}$. Then $\mathrm{FP}(x)$ maps to a singleton and $f(x) = g(x)$.

Suppose $x \in X^*$ is on an indifference surface $\mathcal{I}_{i,k,\ell}$. Let $\nu_{\mathcal{I}}$ denote a normal vector to $\mathcal{I}_{i,k,\ell}$. Since $x \in X^*$, the vector field FP can only jump tangentially to $\nu_{\mathcal{I}}$. Using similar reasoning to the proof of Lemma 10.11, this implies that for any $a, b \in \mathrm{FP}(x)$ we have $a \cdot \nu_{\mathcal{I}}(x) = b \cdot \nu_{\mathcal{I}}(x)$. Hence $\mathrm{FP}(x) \cdot \nu := a \cdot \nu,\ a \in \mathrm{FP}(x)$ is well defined for such x.

In particular, note that if $x \in X^*$ is on some indifference surface $\mathcal{I}$ and $\nu_{\mathcal{I}} = \nu_E$ at x, then $f(x) \cdot \nu_E = \mathrm{FP}(x) \cdot \nu_{\mathcal{I}}$ for any function $f \in \mathrm{FP}$.

Let $\widehat{\mathcal{I}}$ be the union of all indifference surfaces. Since ∂E is piecewise continuous and the indifference surfaces are smooth, the set $S := \{x \in X^* : \ x \in \widehat{\mathcal{I}} \cap \partial E,\ \nu_{\widehat{\mathcal{I}}}(x) \neq \nu_{\partial E}(x)\}$ has $\mathcal{H}^{\kappa-1}$-measure zero, where $\nu_{\widehat{\mathcal{I}}}(x)$ and $\nu_{\partial E}(x)$ denote the normal vectors to $\widehat{\mathcal{I}}$ and ∂E at x.

We have shown that $f\big|_{(\partial E)\setminus S} = g\big|_{(\partial E)\setminus S}$ for all $f, g \in \mathrm{FP}$, and $\mathcal{H}^{\kappa-1}(S) = 0$, and hence,

$$\int_{\partial E} f \cdot \nu_E d\mathcal{H}^{\kappa-1} = \int_{\partial E} g \cdot \nu_E d\mathcal{H}^{\kappa-1}, \qquad \text{for all } f, g \in \mathrm{FP}.$$

□

The following lemma shows that, within X^*, the FP vector field compresses mass at a rate of -1. In particular, this implies that, within X^*, FP cannot map a set of positive measure to a set of zero measure in finite time.[10]

Lemma 10.15. *Let E be a compact subset of X^* with piecewise smooth boundary and finite perimeter. Then*

$$\int_{\partial E} \mathrm{FP} \cdot \nu_E \, d\mathcal{H}^{\kappa-1} = -\mathcal{L}^{\kappa}(E),$$

where ν_E denotes the outer normal vector of E.

Proof. We first note that by Lemma 10.13 for every $f \in \mathrm{FP}$ we have $\int_E dD \cdot f = -\mathcal{L}^{\kappa}(E)$.

[10] We note that this result can also be derived as a consequence of Lemma 3.1 in [158]. For the sake of completeness and to simplify the presentation, we give a proof of the result here using the notation and tools introduced in the chapter.

Let $(f_n)_{n\geq 1}$, $f_n : X^* \to X^*$ be a sequence of uniformly bounded C^1 functions such that $f_n \to f$ a.e. for some function $f : X^* \to X^*$ satisfying $f(x) \in \mathrm{FP}(x)$ for all $x \in X^*$. (Such a sequence can be explicitly constructed by smoothing the FP vector field, e.g., [99].)

Let f and each f_n be written componentwise as $f = (f^s)_{s=1}^{\kappa}$ and $f_n = (f_n^s)_{s=1}^{\kappa}$. Let $D \cdot f_n = \sum_{s=1}^{\kappa} D_s f_n^s$ be the divergence measure associated with f_n and $D \cdot f = \sum_{s=1}^{\kappa} D_s f^s$ the divergence measure associated with f. Since f and f_n are BV functions, by (10.18) we have

$$-\int_{X^*} f_n^s \frac{\partial \phi}{\partial x_s}\, dx = \int_{X^*} \phi\, D_s f_n^s, \qquad \text{and} \qquad -\int_{X^*} f^s \frac{\partial \phi}{\partial x_s}\, dx = \int_{X^*} \phi\, D_s f^s$$

for $n \in \mathbb{N}$, $s = 1, \ldots, \kappa$, for any $\phi \in C_c^1(X^*)$.

For a function $\phi \in C_c^1(X^*)$, there exists a constant $c > 0$ such that $|\frac{\partial \phi(x)}{\partial x_s}| < c$ for all $x \in X^*$. Since $(f_n)_{n\geq 1}$ is uniformly bounded, $|f_n(x)\frac{\partial \phi(x)}{\partial x_s}|$ is bounded by some constant $c > 0$ for all $x \in X^*$, and since X^* is a bounded set, the constant function $c\chi_{X^*}$ (which dominates $|f_n \frac{\partial \phi}{\partial x_s}|$ on X^*) is integrable. Noting that $f_n \frac{\partial \phi}{\partial x_s} \to f \frac{\partial \phi}{\partial x_s}$ pointwise, the dominated convergence theorem gives

$$\lim_{n\to\infty} \int_{X^*} \phi\, D_s f_n^s = -\lim_{n\to\infty} \int_{X^*} f_n^s \frac{\partial \phi}{\partial x_s}\, dx = -\int_{X^*} f^s \frac{\partial \phi}{\partial x_s}\, dx = \int_{X^*} \phi\, D_s f^s.$$

for $n \in \mathbb{N}$, $s = 1, \ldots, \kappa$. This implies that the sequence of measures $(D \cdot f_n)_{n\geq 1}$ converges weakly to $D \cdot f$ in the sense that for any $\phi \in C_c^1(X^*)$ there holds $\lim_{n\to\infty} \int_{X^*} \phi\, dD \cdot f_n = \int_{X^*} \phi\, dD \cdot f$. Letting ϕ approximate the characteristic function χ_E, and noting that by Lemma 10.13 we have $(D \cdot f)(\partial E) = 0$, we see that $\lim_{n\to\infty} \int_E dD \cdot f_n = \int_E dD \cdot f$. Hence,

$$\begin{aligned}
-\mathcal{L}^{\kappa}(E) &= \int_E dD \cdot f \\
&= \lim_{n\to\infty} \int_E dD \cdot f_n \\
&= \lim_{n\to\infty} \int_{\partial E} f_n \cdot \nu_E d\mathcal{H}^{\kappa-1} \\
&= \int_{\partial E} f \cdot \nu_E d\mathcal{H}^{\kappa-1} \\
&= \int_{\partial E} \mathrm{FP} \cdot \nu_E d\mathcal{H}^{\kappa-1},
\end{aligned}$$

where the third line follows from the Gauss-Green theorem [152], the fourth line follows from the dominated convergence theorem (by assumption, E has finite perimeter and a piecewise smooth boundary, and f is bounded), and the fifth line follows from Lemma 10.14. □

We now prove Proposition 10.10.

Proof. We begin by noting that, by Lemma 10.36 in the appendix, $\operatorname{cl} Q$, has Hausdorff dimension at most $\kappa - 2$.

Let $\epsilon > 0$. By the definition of the Hausdorff measure ([152], Chapter 2), there exists a countable collection of balls $(B_\epsilon^j)_{j\geq 1}$, each with diameter less than ϵ, such that $\operatorname{cl} Q \cup \partial\mathcal{Z} \subset \bigcup_{j\geq 1} B_\epsilon^j$ and $\sum_{j=1}^\infty c\left(\frac{\operatorname{diam} B_\epsilon^j}{2}\right)^{\kappa-2} < 2\mathcal{H}^{\kappa-2}(\operatorname{cl} Q \cup \partial\mathcal{Z})$, where $c := \frac{\pi^{\kappa-2}}{\Gamma(\frac{\kappa-2}{2})+1}$, and where Γ in this context denotes the standard Γ function.

Since $\partial\mathcal{Z}$ is closed, $\operatorname{cl} Q \cup \partial Z$ is closed, and hence there exists a finite subcover $(B_\epsilon^j)_{j=1}^{N_\epsilon}$ such that $\operatorname{cl} Q \cup \partial\mathcal{Z} \subset \bigcup_{j=1}^{N_\epsilon} B_\epsilon^j$. Let $B_\epsilon := \bigcup_{j=1}^{N_\epsilon} B_\epsilon^j$, and let

$$X_\epsilon^* := X \setminus (B_\epsilon \cup \mathcal{Z}).$$

Note that, by Lemma 10.35 in the appendix we have

$$\lim_{\epsilon\to 0} \mathcal{H}^{\kappa-1}(\partial B_\epsilon) = 0. \tag{10.21}$$

Fix some time $T > 0$, and for $0 < t \leq T$, let

$$\begin{aligned} E_\epsilon(T-t) := \{x_0 \in X_\epsilon^* :\ & \mathbf{x}(0) = x_0, \mathbf{x}(t) \text{ is a FP process,} \\ & \mathbf{x}(s) \in B_\epsilon, \text{ for some } 0 < s \leq t\} \end{aligned}$$

and note that the boundary $\partial E_\epsilon(T-t)$ is piecewise smooth. The set $E_\epsilon(T-t)$ may be thought of as the set obtained by tracing paths backwards out of B_ϵ from time T back to time $T-t$. Let

$$V_\epsilon(t) := \mathcal{L}^\kappa(E_\epsilon(T-t)).$$

Letting R_ϵ denote the flux through ∂B_ϵ into $E_\epsilon(T-t)$ and again letting ν denote the outer normal to $\partial E_\epsilon(T-t)$, for $t>0$ we have

$$\begin{aligned}\frac{d}{dt}V_\epsilon(t) &= \int_{\partial E_\epsilon(T-t)\backslash \partial B_\epsilon} -\mathrm{FP}\cdot \nu \, dx \\ &\leq R_\epsilon + \int_{\partial E_\epsilon(T-t)} -\mathrm{FP}\cdot\nu\, dx \\ &\leq R_\epsilon + \mathcal{L}^\kappa(E_\epsilon(T-t)) \\ &= R_\epsilon + V_\epsilon(t),\end{aligned} \tag{10.22}$$

where the third line follows by Lemma 10.15.

Noting that $\|\mathrm{FP}\|_\infty < \infty$, the flux through ∂B_ϵ is bounded by

$$R_\epsilon \leq \mathcal{H}^{\kappa-1}(\partial B_\epsilon)\|\mathrm{FP}\|_\infty =: \bar{R}_\epsilon. \tag{10.23}$$

By (10.21) we have $\mathcal{H}^{\kappa-1}(\partial B_\epsilon)\to 0$ as $\epsilon\to 0$, and hence $\bar{R}_\epsilon \to 0$ as $\epsilon \to 0$.

Using the integral form of Gronwall's inequality, (10.22) and (10.23) give $V_\epsilon(t)\leq t\bar{R}_\epsilon e^t$, $0<t\leq T$. In particular, this means that

$$\mathcal{L}^\kappa(E_\epsilon(0)) \leq \bar{R}_\epsilon e^T,$$

where the right hand side goes to zero as $\epsilon\to 0$. Sending $\epsilon \to 0$, we see that the set $W(T) := \{x_0 \in X^* : \ \mathbf{x}(0) = x_0,\ \mathbf{x}(t) \text{ is a FP process, } \mathbf{x}(s)\in Q\cup\partial\mathcal{Z} \text{ for some } 0<s\leq T\}$ has $\mathcal{L}^\kappa$-measure zero.

Since paths may only enter $\mathcal{Z}$ through the boundary $\partial\mathcal{Z}$, this means that the set of points in X from which $\mathcal{Z}$ can be reached within time T is contained in $W(T)\cup\mathcal{Z}$. Furthermore, the set of points from which $Q\cup\mathcal{Z}$ can be reached within time T is contained in $W(T)\cup\mathcal{Z}\cup Q$, which is a $\mathcal{L}^\kappa$-measure zero set. Since this is true for every $T>0$, we get the desired result. □

Remark 10.16. *We note that the proof given above implies that $Q\cup\mathcal{Z}$ can only be reached in finite time from a $\mathcal{L}^\kappa$-measure zero subset of X^*. Since $X^* := X\backslash(Q\cup\mathcal{Z})$, this, combined with Lemma 10.12, implies that for almost every initial condition $x_0\in X^*$, there exists a unique FP process $\mathbf{x}$ with $\mathbf{x}(0) = x_0$. Since $\mathcal{L}^\kappa(X\backslash X^*) = 0$, this implies that for almost every $x_0 \in X$, there exists a unique FP process $\mathbf{x}$ with $\mathbf{x}(0) = x_0$.*

10.4.2 Infinite-Time Convergence

The following proposition shows that it is not possible to converge to $\Lambda(x^*)$ in infinite time.

Proposition 10.17. *Let Γ be a non-degenerate potential game and let x^* be a mixed-strategy equilibrium. Suppose $(\mathbf{x}(t))_{t\geq 0}$ is a FP process and $\mathbf{x}(t) \to x^*$. Then $\mathbf{x}(t)$ converges to $\Lambda(x^*)$ in finite time.*

Proof. From the definitions of $\Lambda(x^*)$ and $\mathcal{P}$ we see that

$$\mathbf{x}(t) \to \Lambda(x^*) \qquad \iff \qquad \mathcal{P}(\mathbf{x}(t)) \to \mathcal{P}(x^*).$$

If we integrate (10.4), use the fact that $\mathcal{P}(\mathbf{x}(t)) \to \mathcal{P}(x^*)$ and set $e(t) := d(\mathcal{P}(\mathbf{x}(t)), \mathcal{P}(x^*))$, then we find that

$$\tilde{U}(\mathcal{P}(x^*)) - \tilde{U}(\mathcal{P}(\mathbf{x}(t))) \geq c_2 \int_t^\infty e(s)ds. \tag{10.24}$$

Using (10.3) above we get

$$ce^2(t) \geq \int_t^\infty e(s)\, ds, \tag{10.25}$$

with $c = c_1/c_2$. Now, using Markov's inequality we have that

$$\mathcal{L}^1\left(\{s : e(s) > e(t)/2, s > t\}\right) \leq \frac{2}{e(t)} \int_t^\infty e(s)\, ds \leq 2ce(t).$$

Let t be fixed. Recursively applying the above inequality we find that

$$\mathcal{L}^1\left(\{s : e(s) > 0, s > t\}\right) \leq 4ce(t).$$

Thus if $\mathcal{P}(\mathbf{x}(t))$ converges to $\mathcal{P}(x^*)$, it must reach it for the first time in finite time.

By construction $\mathcal{P}(x) = \mathcal{P}(x^*)$ if and only if $x \in \Lambda(x^*)$. Hence, if $\mathbf{x}(t)$ converges to $\Lambda(x^*)$ it must reach it for the first time in finite time.

By (10.4) we have $\frac{d}{dt}\tilde{U}(\mathcal{P}(\mathbf{x}(t))) \geq 0$ in a neighborhood of $\mathcal{P}(x^*)$. Since Γ is non-degenerate, the Hessian of $\tilde{U}$ is invertible at $\mathcal{P}(x^*)$, and for all $\tilde{x} \in \tilde{X}$ in a punctured ball around $\mathcal{P}(x^*)$ we have $\tilde{U}(\tilde{x}) \neq \tilde{U}(\mathcal{P}(x^*))$. Thus, if $\mathbf{x}(t) \to x^*$ and $\mathcal{P}(\mathbf{x}(T)) = \mathcal{P}(x^*)$ (i.e., $\mathbf{x}(T) \in \Lambda(x^*)$) for some $T \geq 0$, then we must have $\mathcal{P}(\mathbf{x}(t)) = \mathcal{P}(x^*)$ (i.e., $x(t) \in \Lambda(x^*)$) for all $t \geq T$. Contrariwise, we would have $\tilde{U}(\mathcal{P}(x^*)) = \tilde{U}(\mathcal{P}(\mathbf{x}(T))) < \lim_{s\to\infty} \tilde{U}(\mathcal{P}(\mathbf{x}(s))) = \tilde{U}(\mathcal{P}(x^*))$, which is a contradiction. □

10.5 Rate of Convergence of FP in Potential Games

In this section we characterize the rate of convergence of FP in potential games.

The following result shows that the rate of convergence of FP in non-degenerate potential games is exponential for almost all initial conditions. Since almost all potential games are non-degenerate, this shows that Conjecture 25 of [27] holds for almost all initial conditions in almost all (exact) potential games (see Remark 10.19 below).

Theorem 10.18. *Let Γ be a non-degenerate potential game. For almost every initial condition x_0, there exist a constant $c = c(\Gamma, x_0)$ such that if $\mathbf{x}$ is a FP process associated with Γ and $\mathbf{x}(0) = x_0$, then*

$$d(\mathbf{x}(t), NE) \leq ce^{-t}. \tag{10.26}$$

Proof. Note the following:

(i) By Remark 10.16, for almost every initial condition $x_0 \in X$, there is a unique FP process with $\mathbf{x}(0) = x_0$.

(ii) By Theorem 10.1, for almost every initial condition, FP converges to a pure-strategy NE.

This implies that there exists a set $\Omega \subset X$ satisfying the following properties: (a) $\mathcal{L}^\kappa(X \backslash \Omega) = 0$, (b) for every FP process $\mathbf{x}$ with initial condition $x_0 \in \Omega$, $\mathbf{x}$ is the unique FP process satisfying $\mathbf{x}(0) = x_0$, and $\mathbf{x}$ converges to a pure-strategy NE.

Let $x_0 \in \Omega$, let $\mathbf{x}$ be a FP process with $\mathbf{x}(0) = x_0$, and let x^* be the pure-strategy NE to which $\mathbf{x}$ converges. Without loss of generality, assume that the pure-strategy set Y is reordered so that $x^* = 0$ (i.e., $T_i^1(x_i^*) = 1$ for all $i = 1, \ldots, N$, where T_i^k is defined in Section 9.2).

Since Γ is (first-order) non-degenerate, there exists a neighborhood of x^* through which no indifference surface passes. Since x^* is a NE, this implies that for all x in a neighborhood of x^* we have $\text{BR}(x) = x^*$. Since $\mathbf{x}(t) \to x^*$, this, along with (10.1), implies that there exists a time $\tau = \tau(\Gamma, x_0) > 0$ such that for all $t \geq \tau$, we have $\dot{\mathbf{x}}(t) = -\mathbf{x}(t)$. Hence, for $t \geq \tau$ we have $\|\mathbf{x}(t)\| = \|\mathbf{x}(\tau)\| e^{\tau - t}$. Letting $c := \sup_{t \in [0,\tau]} \|\mathbf{x}(t)\| e^\tau$ we get $\|\mathbf{x}(t)\| \leq ce^{-t}$ for all $t \geq 0$.

□

Remark 10.19. *Conjecture 25 of [27] posited that within the class of weighted potential games, the rate of convergence of any FP process $\mathbf{x}$ is exponential with the constant c in* (10.26) *depending on the game Γ and the particular FP process $\mathbf{x}$; i.e., $c = c(\Gamma, \mathbf{x})$. In Theorem 10.18 we showed this is true in almost all exact potential games and for almost all initial conditions, and furthermore, we showed that, for almost all initial conditions, the constant c in* (10.26) *can be determined by the initial condition alone, rather than depending on the full path $\mathbf{x}$; i.e., $c = c(\Gamma, x_0)$.*

10.6 Finite-Time Convergence in ECFP

In this section we study the rate of convergence of ECFP. The main result of this section is Proposition 10.23, which shows that when ECFP converges to an interior equilibrium, it generally does so in finite time. In order to simplify notation in the proofs, we will study the behavior of ECFP within the space X rather than Δ^N (see Section 9.2).

Let $\mathcal{C} = \{C_1, \ldots, C_m\}$ be a permutation invariant partition of the player set, and let $I = \{1, \ldots, m\}$ be the associated index set (see Section 3.3). Given an $x \in X$, let the k-th partial centroid (relative to $\mathcal{C}$) be given by the mapping

$$f^k(x) := |C_k|^{-1} \sum_{i \in C_k} x_i, \tag{10.27}$$

and let the centroid distribution (relative to $\mathcal{C}$) be given by the mapping $f(x) := (f^{\phi(i)})_{i=1}^N$, where $\phi : \mathcal{N} \to I$ is as defined in Section 3.3. Let $f_i(x) = (f^{\phi(j)})_{j \in \mathcal{N},\, j \neq i}$ give the centroid distribution for players other than i.

In this context, we say $(\mathbf{x}(t))_{t \geq 0}$ is an ECFP process if $\mathbf{x} : [0, \infty) \to X$ is an absolutely continuous mapping and

$$\dot{\mathbf{x}}_i(t) \in \mathrm{BR}_i(f_i(\mathbf{x}(t))) - \mathbf{x}_i(t), \quad \forall i \in \mathcal{N} \tag{10.28}$$

holds for almost all $t \in [0, \infty)$.

In this section we will focus on studying the convergence of ECFP to *interior* equilibrium points and we will assume that at an interior equilibrium point x^*, the following holds.

Assumption 10.20. *The Hessian of U is invertible at x^*.*

In Section 9.3 we showed that almost all potential games satisfy this assumption. In studying ECFP, we focus on the subset of games the are permutation invariant relative to a given player-set partition $\mathcal{C}$. Using similar techniques to those of Section 9.3 it may be possible to show that the almost all games that are permutation invariant with respect to a given partition $\mathcal{C}$ satisfy Assumption 10.20.

Suppose x^* is a completely mixed SNE satisfying Assumption 10.20. The following two inequalities will be instrumental in proving the main result of this section (Proposition 10.23).
(1) For all x in a neighborhood of x^* we have

$$\big| U(x^*) - U(f(x)) \big| \leq c_1 d^2(f(x), x^*). \tag{10.29}$$

(2) Suppose $(\mathbf{x}(t))_{t \geq 0}$ is an ECFP process with $f(\mathbf{x}(t)) \to x^*$. Then for all $\mathbf{x}(t)$ in a neighborhood of x^* there holds

$$\frac{d}{dt} U(f(\mathbf{x}(t))) \geq c_2 d(f(\mathbf{x}(t)), x^*). \tag{10.30}$$

The first inequality is a straightforward consequence of Taylor's theorem and the fact that $\nabla U(x^*) = 0$. The second inequality is proved using the following two lemmas.

Lemma 10.21. *Let x^* be a completely mixed SNE. There exists a constant $c > 0$ such that for each $i = 1, \ldots, N$ there holds*

$$z_i \cdot \nabla_{x_i} U(f(x)) \geq c \|\nabla_{x_i} U(f(x))\|, \qquad \forall z_i \in \mathrm{BR}_i(f_i(x)) - [f(x)]_i$$

for all x in a neighborhood of x^*.

Proof. Let B_ϵ be an ϵ-ball around x^* and let ϵ be chosen so that B_ϵ is separated by a positive distance from the boundary of X.

Let $x \in B_\epsilon$. If $\|\nabla_{x_i} U(f(x))\|_1 = 0$, then the inequality is trivially satisfied. Suppose from now on that $\|\nabla_{x_i} U(f(x))\|_1 > 0$.

Without loss of generality, assume that Y_i is ordered so that

$$y_i^1 \in \mathrm{BR}_i(f_i(x)).$$

Differentiating (9.4) at $f(x)$ we find that

$$\frac{\partial U(f(x))}{\partial x_i^k} = U(y_i^{k+1}, f_i(x)) - U(y_i^1, f_i(x)), \tag{10.31}$$

so that for $k = 1, \ldots, K_i - 1$ we have

$$y_i^{k+1} \in \mathrm{BR}_i(f_i(x)) \iff \frac{\partial U(f(x))}{\partial x_i^k} = 0. \tag{10.32}$$

Using the multlinearity of U we see that if $\xi_i \in \mathrm{BR}_i(f_i(x))$ and $\xi_i^k > 0$ then $y_i^{k+1} \in \mathrm{BR}_i(f_i(x))$. But, by (10.32) this implies that if $\xi_i \in \mathrm{BR}_i(f_i(x))$ and $\xi_i^k > 0$ then $\frac{\partial U(x)}{\partial x_i^k} = 0$. Noting that any $\xi_i \in \mathrm{BR}_i(f_i(x))$ is necessarily coordinatewise nonnegative, this gives

$$(\xi_i - [f(x)]_i) \cdot \nabla_{x_i} U(f(x)) = \underbrace{\sum_{k=1}^{K_i - 1} \xi_i^k \frac{\partial U(f(x))}{\partial x_i^k}}_{=0} - \sum_{k=1}^{K_i - 1} [f(x)]_i^k \frac{\partial U(f(x))}{\partial x_i^k}, \tag{10.33}$$

for $\xi_i \in \mathrm{BR}_i(x_{-i})$. Since we assume $x \in B_\epsilon$, we have $x_i^k \geq d(B_\epsilon, \partial X)$, for all $k = 1, \ldots, K_i - 1$, which implies that the each component of $f(x)$ also satisfies $[f(x)]_i^k \geq d(B_\epsilon, \partial X)$ (see (10.27)).

Since we assume $y_i^1 \in \mathrm{BR}_i(f_i(x))$, from (10.31) we get that $\frac{\partial U(f(x))}{\partial x_i^k} \leq 0$ for all $k = 1, \ldots, K_i - 1$. Substituting into (10.33), this gives

$$(\xi_i - [f(x)]_i) \cdot \nabla_{x_i} U(f(x)) \geq d(B_\epsilon, \partial X) \sum_{k=1}^{K_i-1} \left(-\frac{\partial U(f(x))}{\partial x_i^k} \right), \qquad \xi_i \in \mathrm{BR}_i(f_i(x)_{-i}).$$

But since $\frac{\partial U(f(x))}{\partial x_i^k} \leq 0$ for all k we have $\sum_{k=1}^{K_i-1} \left(-\frac{\partial U(f(x))}{\partial x_i^k} \right) = \|\nabla_{x_i} U(f(x))\|_1$, and hence

$$(\xi_i - [f(x)]_i) \cdot \nabla_{x_i} U(f(x)) \geq d(B_\epsilon, \partial X) \|\nabla_{x_i} U(f(x))\|_1, \quad \xi_i \in \mathrm{BR}_i(f_i(x)).$$

Since X is finite dimensional, the desired result follows by the equivalence of finite dimensional norms. □

Lemma 10.22. *Let x^* be a completely mixed SNE satisfying Assumption 10.20 and assume $(\mathbf{x}(t))_{t \geq 0}$ is an ECFP process satisfying $f(\mathbf{x}(t)) \to x^*$. Then for $\mathbf{x}(t)$ in a neighborhood of x^*, the differential inequality* (10.30) *holds.*

Proof. By the chain rule we have

$$\frac{d}{dt} U(f(\mathbf{x}(t))) = \nabla U(f(\mathbf{x}(t))) \cdot \frac{d}{dt} f(\mathbf{x}(t))$$

Using the permutation invariance of U we see that (10.28) implies that $\frac{d}{dt} f(\mathbf{x}(t)) \in \mathrm{BR}(f(\mathbf{x}(t)) - f(x)$. Restricting to the strategy component of player i, this gives $\frac{d}{dt}[f(\mathbf{x}(t))]_i \in \mathrm{BR}_i(f_i(x)) - [f(x)]_i$. Thus, we may use Lemma 10.21, which combined with the above gives

$$\frac{d}{dt} U(f(\mathbf{x}(t))) \geq c \|\nabla U(f(\mathbf{x}(t)))\|$$

for some constant $c > 0$, and for $f(\mathbf{x}(t))$ in a neighborhood of x^*. Since the Hessian of U is non-degenerate, by Lemma 10.30 in the appendix we have $\|U(f(\mathbf{x}(t)))\| \geq c d(f(\mathbf{x}(t)), x^*)$ for some constant $c > 0$ and for $f(\mathbf{x}(t))$ in a neighborhood of x^*. Combined with the above this gives

$$\frac{d}{dt} U(f(\mathbf{x}(t))) \geq c d(f(\mathbf{x}(t)), x^*), \tag{10.34}$$

for some constant $c > 0$, and for $f(\mathbf{x}(t))$ in a neighborhood of x^*. Let Ω denote this neighborhood of x^*. Since f is continuous and $f(x^*) = x^*$, there exists a neighborhood Ω' of x^* such that $f(\Omega') = \Omega$. Hence, (10.34) holds for all $\mathbf{x}(t)$ in some neighborhood Ω' of x^*, which is the desired result. □

The following proposition characterizes the convergence rate of ECFP when converging to interior equilibria satisfying Assumption 10.20. The proof technique is similar to that of Proposition 10.17.

Proposition 10.23. *Let Γ be potential game satisfying Assumptions 3.1–3.2. Let x^* be a completely mixed SNE satisfying Assumption 10.20. If an ECFP process converges to x^* in the sense that $f(\mathbf{x}(t)) \to x^*$, then it converges in finite time.*

Proof. If we integrate (10.30), use the fact that $f(\mathbf{x}(t)) \to x^*$, and set $E(t) := d(f(\mathbf{x}(t)), x^*)$ then we find that

$$U(x^*) - U(f(\mathbf{x}(t)))) \geq c_2 \int_t^\infty E(s)ds.$$

Using (10.29) above we get

$$cE^2(t) \geq \int_t^\infty E(s)\, ds,$$

with $c = c_1/c_2$. Now, using Markov's inequality we have that

$$\mathcal{L}^1\left(\{s : E(s) > E(t)/2, s > t\}\right) \leq \frac{2}{E(t)} \int_t^\infty E(s)\, ds \leq 2cE(t).$$

Let t be fixed. Recursively applying the above inequality we find that

$$\mathcal{L}^1\left(\{s : E(s) > 0, s > t\}\right) \leq 4cE(t).$$

Thus if $f(\mathbf{x}(t))$ converges to x^*, it must reach it for the first time in finite time.

By (10.30) we have $\frac{d}{dt}U(f(\mathbf{x}(t))) \geq 0$ in a neighborhood of x^*. By assumption, the Hessian of U is invertible at x^*, and for all $\tilde{x} \in \tilde{X}$ in a punctured ball around x^* we have $\tilde{U}(f(\tilde{x})) \neq \tilde{U}(x^*)$. Thus, if $f(\mathbf{x}(t)) \to x^*$ and $f(\mathbf{x}(T)) = x^*$ for some $T \geq 0$, then we must have $f(\mathbf{x}(t)) = x^*$ for all $t \geq T$. Contrariwise, we would have $U(x^*) = \tilde{U}(f(\mathbf{x}(T))) < \lim_{s\to\infty} U(f(\mathbf{x}(s))) = U(x^*)$, which is a contradiction. □

Remark 10.24. *Proposition 10.23 shows that convergence of ECFP to* completely *mixed equilibria generally occurs in finite time. However, using the projection techniques developed in Section 10.3, it may be possible to extend this result to show that convergence of ECFP to* incompletely *mixed equilibria also occurs in finite time.*

Appendix

Lemma 10.25. *Suppose Γ is a degenerate game. At any mixed equilibrium there are at least two players using mixed strategies.*

Proof. Suppose that x^* is an equilibrium in which only one player uses a mixed strategy–say, player 1. Let $C_i = \text{carr}_i(x^*)$ and $\gamma_i = |C_i|$. Then the mixed strategy Hessian is given by $\nabla U(x) = (\frac{\partial^2 U(x)}{\partial x_1^k \partial x_1^\ell})_{k,\ell=1,\ldots,\gamma_i} = 0$, where the equality to zero follows since U is linear in x_1. $\square$

Lemma 10.26. *Let $x \in X$ and $i = 1, \ldots, N$. Assume Y_i is ordered so that $y_i^1 \in BR_i(x_{-i})$. Then:*
(i) For $k = 1, \ldots, K_i - 1$ we have $\frac{\partial U(x)}{\partial x_i^k} \leq 0$.
(ii) For $k = 1, \ldots, K_i - 1$, we have $y_i^{k+1} \in BR_i(x_{-i})$ if and only if $\frac{\partial U(x)}{\partial x_i^k} = 0$. In particular, combined with (i) this implies that $y_i^{k+1} \notin BR_i(x_{-i}) \iff \frac{\partial U(x)}{\partial x_i^k} < 0$.

Proof. (i) Differentiating (9.4) we find that

$$\frac{\partial U(x)}{\partial x_i^k} = U(y_i^{k+1}, x_{-i}) - U(y_i^1, x_{-i}). \tag{10.35}$$

(i) Since y_i^1 is a best response, we must have $U(y_i^1, x_{-i}) \geq U(y_i^{k+1}, x_{-i})$ for any $k = 1, \ldots, K_i - 1$. Hence $\frac{\partial U(x)}{\partial x_i^k} \leq 0$.
(ii) Follows readily from (9.4). $\square$

Lemma 10.27. *Let $x \in X$. If $y_i^k \in BR_i(x_{-i})$ then $\frac{\partial U(x)}{\partial x_i^k} \geq 0$.*

Proof. The result follows readily from (10.35). $\square$

Lemma 10.28. *Suppose x^* is an equilibrium and $y_i^k \in \text{carr}(x^*)$, $k \geq 2$. Then $\frac{\partial U(x^*)}{\partial x_i^k} = 0$.*

Proof. Since U is multilinear, y_i^k must be a pure-strategy best response to x_{-i}^*. The result then follows from Lemma 10.26. $\square$

Lemma 10.29. *There exists a $c > 0$ such that $|\frac{\partial \tilde{\mathcal{P}}_i^k(x)}{\partial x_j^\ell}| < c$, $i = 1, \ldots, \tilde{N}$, $k = 1, \ldots, \gamma_i - 1$, $j = 1, \ldots, N$, $\ell \geq \gamma_j$ for x in a neighborhood of x^*.*

Proof. Differentiating (10.2) we see that $\frac{\partial \tilde{\mathcal{P}}_i^k(x)}{\partial x_j^\ell} = -\frac{\partial g_i^k(x_p)}{\partial x_j^\ell}$, $i = 1, \ldots, \tilde{N}$, $k = 1, \ldots, \gamma_i - 1$, $j = 1, \ldots, N$, $\ell \geq \gamma_j$, $x = (x_p, x_m)$.

By the definition of g we have $F(x_p, g(x_p), u) = 0$ for all x_p in a neighborhood of x_p^*. Hence,

$$\begin{aligned}0 &= D_{x_p} F(x_p, g(x_p), u) \\ &= D_{x_p} F(x_p, x_m', u)\big|_{x_m' = g(x_p)} + D_{x_m} F(x_p, x_m, u) D_{x_p} g(x_p),\end{aligned}$$

By (9.12) we have $D_{x_m} F(x_p, x_m, u) = \mathbf{H}(x)$. Since the equilibrium x^* is assumed to be non-degenerate, $\mathbf{H}(x^*)$ is invertible and the above implies that

$$D_{x_p} g(x_p^*) = \mathbf{H}(x^*)^{-1} D_{x_p} F(x_p^*, x_m^*).$$

Using (9.10) and the multilinearity of U, one may readily verify that $D_{x_p} F(x_p^*, x_m^*, u)$ is entrywise finite. Since g is continuously differentiable, it follows that each entry of

$$\left(\frac{\partial \tilde{\mathcal{P}}_i^k(x)}{\partial x_j^\ell}\right)_{\substack{i=1,\dots,\tilde{N}, k=1,\dots,\gamma_i-1 \\ j=1,\dots,N,\ \ell \geq \gamma_j}} = \left(-\frac{\partial g_i^k(x_p)}{\partial x_j^\ell}\right)_{\substack{i=1,\dots,\tilde{N}, k=1,\dots,\gamma_i-1 \\ j=1,\dots,N,\ \ell \geq \gamma_j}} = -D_{x_p} g(x_p)$$

is uniformly bounded for $x = (x_p, x_m)$ in a neighborhood of x^*. □

Lemma 10.30. *Suppose $V : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. Suppose x^* is a critical point of V and the Hessian of V at x^*, denoted by $\mathbf{H}(x^*)$, is invertible. Then there exists a constant c such that $\|\nabla V(x)\| \geq c d(x^*, x)$ for all x in a neighborhood of x^*.*

Proof. Suppose the claim is false. Then for any $\epsilon > 0$ there exists a sequence $(x_k^\epsilon)_{k\geq 1} \subset B(x^*, \epsilon)$ such that $\|\nabla V(x_k)\| < \frac{1}{k} d(x_k, x^*)$. Let $(x_k)_{k\geq 1}$ be such a sequence that furthermore satisfies $\lim_{k\to\infty} d(x_k, x^*) = 0$. Let $y_k \in \mathbb{R}^n$, $t_k \in \mathbb{R}$ be such that $x_k = x^* + t_k y_k$, $\|y_k\| = 1$. Since $(y_k)_{k\geq 1}$ is a sequence on the unit sphere in $\mathbb{R}^n$ it has a convergent subsequence; say, $y_{k_j} \to y$ as $j \to \infty$. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(t) := V(x^* + ty)$.

Using the continuity of ∇V we see that for any $c > 0$ we have $|f'(t)| < ct$ for all t sufficiently small. Since x^* is a critical point of V we have $f'(0) = 0$. Hence

$$f''(0) = \lim_{t\to 0} \left|\frac{f'(t) - f'(0)}{t}\right| = \lim_{t\to 0} \frac{|f'(t)|}{t} < c.$$

Letting $c \to 0$ we see that $f''(0) = 0$. But this means $0 = f''(0) = y^T \mathbf{H}(x^*) y$, implying the Hessian is singular, which is a contradiction. □

The following lemma characterizes the level sets of polynomial functions. Before presenting the lemma we require the following definition.

Definition 10.31. *Given a polynomial $p : \mathbb{R}^n \to \mathbb{R}$, $n \geq 1$, let*

$$Z(p) := \{x \in \mathbb{R}^n : \ p(x) = 0\}$$

be the zero-level set of p.

Lemma 10.32. *Let $p(x) : \mathbb{R}^n \to \mathbb{R}$, $n \geq 1$ be a polynomial that is not identically zero. Then $\mathcal{L}^n(Z(p)) = 0$.*

Proof. We will prove the result using an inductive argument.

Suppose first that $n = 1$ so that $p : \mathbb{R} \to \mathbb{R}$. Let k denote the degree of p. Since p is not identically zero, the fundamental theorem of algebra implies that p has at most k zeros. Hence $\mathcal{L}^1(Z(p)) = 0$.

Now, suppose that $n \geq 2$ and for any polynomial $\tilde{p} : \mathbb{R}^{n-1} \to \mathbb{R}$ there holds $\mathcal{L}^{n-1}(Z(\tilde{p})) = 0$. We may write

$$p(x, x_n) = \sum_{j=0}^{k} p_j(x) x_n^j,$$

where k is the degree of p in the variable x_n, $x = (x_1, \ldots, x_{n-1})$, the functions p_j, $j = 0, \ldots, k$ are polynomials in $n-1$ variables, and where at least one p_j is not identically zero.

If (x, x_n) is such that $p(x, x_n) = 0$ then there are two possibilities: Either (i) $p_0(x) = \ldots = p_k(x) = 0$, or (ii) x_n is the root of the one-variable polynomial $p_x(t) := \sum_{j=1}^{k} p_j(x) t^j$.

Let A and B be the subsets of $\mathbb{R}^n$ where (i) and (ii) hold respectively, so that $Z(p) = A \cup B$. For any $x_n \in \mathbb{R}$ we have $(x, x_n) \in A \iff x \in Z(p_j), \ \forall j = 1, \ldots, k$. By the induction hypothesis, we have $Z(p_j) = 0$ for at least one j, and hence $\int_{\mathbb{R}^{n-1}} \chi_A(x, x_n)\, dx = 0$ for any $x_n \in \mathbb{R}$, where we include the argument in the characteristic function χ_A, in order to emphasize the dependence on both x and x_n. This implies that $x_n \mapsto \int_{\mathbb{R}^{n-1}} \chi_{(x,x_n) \in A}\, dx$ is a measurable function (it's identically zero) and

$$\mathcal{L}^n(A) = \int_{\mathbb{R}} \int_{\mathbb{R}^{n-1}} \chi_A(x, x_n)\, dx\, dx_n = 0.$$

By the fundamental theorem of algebra, for any $x \in \mathbb{R}^{n-1}$ there are at most k values $t \in \mathbb{R}$ such that $(x, t) \in B$, and hence $\int_{\mathbb{R}} \chi_B(x, x_n)\, dx_n = 0$. As before, this implies that $x \mapsto \int_{\mathbb{R}} \chi_B(x, x_n)\, dx_n$ is a measurable function and

$$\mathcal{L}^n(B) = \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \chi_B(x, x_n)\, dx_n\, dx = 0.$$

Since $Z(p) = B \cup A$, this proves the desired result. $\square$

Remark 10.33. *Note that if $p \equiv 0$, then $Z(p) = \mathbb{R}^n$. Thus, in general, if $p : \mathbb{R}^n \to \mathbb{R}$ is a polynomial, then Lemma 10.32 implies that either $Z(p) = \mathbb{R}^n$ or $\mathcal{L}^n(Z(p)) = 0$.*

Lemma 10.34. *Suppose Γ is a non-degenerate potential game. Then each indifference surface $\mathcal{I}_{i,k,\ell}$, as defined in (10.13), is a union of smooth surfaces with Hausdorff dimension at most $\kappa - 1$.*

Proof. Throughout the proof, when we refer to the dimension of a set we mean the Hausdorff dimension. Let $i \in \{1, \dots, N\}$, $k, \ell \in \{1, \dots, K_i\}$, $k \neq \ell$ and let $\mathcal{I} := \mathcal{I}_{i,k,\ell}$, where $\mathcal{I}_{i,k,\ell}$ is as defined in (10.13). Note that $\mathcal{I}$ is the zero-level set of the polynomial $p(x) := U(y_i^k, x_{-i}) - U(y_i^\ell, x_{-i})$. By Lemma 10.32 and Remark 10.33 we see that either $\mathcal{L}^\kappa(\mathcal{I}) = 0$, or $\mathcal{I} = X$. Being the level set of a polynomial, if $\mathcal{L}^\kappa(\mathcal{I}) = 0$, then $\mathcal{I}$ is the union of smooth surfaces with dimension at most $\kappa - 1$.

Suppose that $\mathcal{I}$ has dimension greater than $\kappa - 1$. Then by the above, we see that $\mathcal{I} = X$. Since Γ is a finite normal-form game, there exists at least one equilibrium $x^* \in X$. Letting x^* be written componentwise as $x^* = ([x^*]_j^m)_{j=1,\dots,N,\ m=1,\dots,K_i-1}$ we see that if $[x^*]_i^k > 0$, then $x^* \in \mathcal{I} = X$ implies that x^* is a second-order degenerate equilibrium. Otherwise, if $[x^*]_i^k = 0$, then $x^* \in \mathcal{I} = X$ implies that x^* is a first-order degenerate equilibrium. In either case we see that x^* is a degenerate equilibrium, and hence Γ is a degenerate game, which is a contradiction.

Since $\mathcal{I}$ was an arbitrary indifference surface, we see that if Γ is a non-degenerate game, then every indifference surface has dimension at most $\kappa - 1$. $\square$

Lemma 10.35. *Let B_ϵ be as defined in the proof of Proposition 10.10. Then,*

$$\mathcal{H}^{\kappa-1}(\partial B_\epsilon) \to 0 \quad \textit{as} \quad \epsilon \to 0.$$

Proof. Following standard notation (see [152], Chapter 2), for $0 \le s < \infty$, $0 < \delta \le \infty$, and $A \subset \mathbb{R}^\kappa$, let

$$\mathcal{H}^s_\delta(A) := \inf\left\{ \sum_{j=1}^\infty \alpha(s) \left(\frac{\operatorname{diam} C_j}{2}\right)^s :\ A \subset \bigcup_{j=1}^\infty C_j,\ \operatorname{diam} C_j \le \delta \right\},$$

where $\alpha(s) := \frac{\pi^s}{\Gamma(\frac{s}{2})+1}$, and where Γ in this context denotes the Γ function

By our construction of $(B_\epsilon^j)_{j\ge 1}$, for every $\epsilon > 0$ we have

$$\sum_{j=1}^\infty \alpha(s) \left(\frac{\operatorname{diam} B_\epsilon^j}{2}\right)^{\kappa-2} < 2\mathcal{H}^{\kappa-2}(Q \cup \partial\mathcal{Z}) < \infty.$$

Since $\operatorname{diam} B_\epsilon^j \le \epsilon$ for every $\epsilon > 0$, $j \in \mathbb{N}$, this gives

$$\begin{aligned}
\lim_{\epsilon\to 0} \mathcal{H}_\epsilon^{\kappa-1}(\partial B_\epsilon) &\le \lim_{\epsilon\to 0} \mathcal{H}_\epsilon^{\kappa-1}(B_\epsilon) \\
&\le \lim_{\epsilon\to 0} \sum_{j=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} B_\epsilon^j}{2}\right)^{\kappa-1} \\
&\le \lim_{\epsilon\to 0} \epsilon \sum_{j=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} B_\epsilon^j}{2}\right)^{\kappa-2} \\
&\le \lim_{\epsilon\to 0} \epsilon 2\mathcal{H}^{\kappa-2}(Q \cup \partial\mathcal{Z}) \ = 0.
\end{aligned}$$

By the definition of the Hausdorff measure we have $\mathcal{H}^{\kappa-1}(\partial B_\epsilon) := \sup_{\delta>0} \mathcal{H}_\delta^{\kappa-1}(\partial B_\epsilon)$. Hence, the above implies $\lim_{\epsilon\to 0} \mathcal{H}^{\kappa-1}(\partial B_\epsilon) = 0$. $\square$

Lemma 10.36. *Let Q be defined as in Section 10.4.1. Then* $\operatorname{cl} Q$ *has Hausdorff dimension at most* $\kappa - 2$.

Proof. Let A be the subset of X where two or more decision surfaces intersect. Let $N \subset A$ be the subset of X where two or more decision surfaces intersect and their normal vectors coincide. Define the *relative interior* of N with respect to A as

$$\operatorname{ri} N := \{x \in N : \ \exists \epsilon > 0 \text{ s.t. } B(x,\epsilon) \cap A \subset N\},$$

and define the *relative boundary* of N with respect to A as

$$\partial N := \operatorname{cl} N \backslash \operatorname{ri} N.$$

Since each indifference surface has Hausdorff dimension $\kappa - 1$, N has Hausdorff dimension at most $\kappa - 1$. In particular, N is the union of a finite number of smooth $\kappa - 1$ dimensional surfaces and a component with Hausdorff dimension at most $\kappa - 2$. This implies that the relative boundary of N has Hausdorff dimension at most $\kappa - 2$.

Let $\tilde{Q}$ be as defined in Section 10.4.1. Note that the closure of $\tilde{Q}$ satisfies $\operatorname{cl} \tilde{Q} \subseteq \tilde{Q} \cup \partial N$. Since the sets $\tilde{Q}$ and ∂N have Hausdorff dimension at most $\kappa - 2$, the set $\operatorname{cl} \tilde{Q}$ also has Hausdorff dimension at most $\kappa - 2$.

Let $\Lambda(x^*)$ be as defined in Section 10.4. If $\Lambda(x^*) = \{x^*\}$, then $\Lambda(x^*)$ is closed and has Hausdorff dimension 0. Otherwise, $\Lambda(x^*)$ is defined as the graph of g. In Section 10.3.1 it was shown that $\operatorname{Graph}(g)$ has Hausdorff dimension at most $\kappa - 2$. Since g is a smooth function, the closure of

Graph(g) has Hausdorff dimension at most $\kappa - 2$.

Recall that Q is defined as $Q = \tilde{Q} \cup \Lambda(x^*)$ and hence $\operatorname{cl} Q = \operatorname{cl} \tilde{Q} \cup \operatorname{cl} \Lambda(x^*)$. Since $\operatorname{cl} \tilde{Q}$ and $\operatorname{cl} \Lambda(x^*)$ each have Hausdorff dimension at most $\kappa - 2$, $\operatorname{cl} Q$ also has Hausdorff dimension at most $\kappa - 2$. □

Lemma 10.37. *Let $\mathcal{Z}$ be as defined* (10.15). *Then for any $x \in \mathcal{Z}$ there holds $\nu \cdot y = 0$ for any vector ν normal to $\mathcal{Z}$ at x, and any $y \in FP(x)$*

Proof. Suppose $x \in X$ and x is in some indifference surface $\mathcal{I}_{i,k,\ell}$. Suppose ν is a vector that is normal to $\mathcal{I}_{i,k,\ell}$ at x. By the definition of $\mathcal{I}_{i,k,\ell}$, if $x \in \mathcal{I}_{i,k,\ell}$ then for all $\hat{x} \in X$ such that $\hat{x}_{-i} = x_{-i}$ we have $\hat{x} \in \mathcal{I}_{i,k,\ell}$. This implies that the $(i, \tilde{k})$-th component of ν must be zero for every $\tilde{k} = 1, \ldots, K_i - 1$.

For $x \in X$, let $\mathcal{N}(x) := \{(i,k) : \; i \in \{1, \ldots, N\}, \; k \in \{1, \ldots, K_i - 1\}, \; x \in \mathcal{I}_{i,k,\ell} \text{ for some } \; \ell = 1, \ldots, K_i - 1, \; \ell \neq k\}$. Letting FP_i^k be the (i,k)-th component map of FP, note that by the definition of an indifference surface, $\mathrm{FP}_i^k(x)$ is single valued for every pair $(i,k) \notin \mathcal{N}(x)$.

Suppose $x \in X \backslash Q$ is in at least one decision surface $\mathcal{I}$ and let ν be a vector that is normal to $\mathcal{I}$ at x. Note that $x \notin Q$ implies that if x is contained in any other decision surface $\hat{\mathcal{I}} \neq \mathcal{I}$, then ν is also normal to $\hat{\mathcal{I}}$ at x. Letting ν be written componentwise as $\nu = (\nu_i^k)_{i=1,\ldots N, \; k=1,\ldots,K_i-1}$, the above discussion implies that $\nu_i^k = 0$ for every pair $(i,k) \in \mathcal{N}(x)$.

Now suppose $x \in \mathcal{Z}$. By the definition of $\mathcal{Z}$ we have $x \notin Q$ and x is in at least one decision surface $\mathcal{I}$. Let ν be a vector that is normal to $\mathcal{I}$ at x. By the definition of $\mathcal{Z}$, there exists some $y \in \mathrm{FP}(x)$ such that $y \cdot \nu = 0$. Breaking this down in terms of components in $\mathcal{N}(x)$ we have

$$0 = y \cdot \nu = \sum_{(i,k) \in \mathcal{N}(x)} y_i^k \nu_i^k + \sum_{(i,k) \notin \mathcal{N}(x)} y_i^k \nu_i^k.$$

The first sum is zero since $\nu_i^k = 0$ for all $(i,k) \in \mathcal{N}(x)$. Consequently, the second sum must also be zero. But we have shown above that $F_i^k(x)$ is single valued for any $(i,k) \notin \mathcal{N}(x)$. Hence, for any $\tilde{y} \in \mathrm{FP}(x)$ we have $\tilde{y}_i^k = y_i^k$ for all $(i,k) \notin \mathcal{N}(x)$, and in particular, $\sum_{(i,k) \notin \mathcal{N}(x)} \tilde{y}_i^k \nu_i^k = \sum_{(i,k) \notin \mathcal{N}(x)} y_i^k \nu_i^k = 0$. Moreover, since $\nu_i^k = 0$ for all $(i,k) \in \mathcal{N}(x)$ we have $\sum_{(i,k) \in \mathcal{N}(x)} \tilde{y}_i^k \nu_i^k = 0$, which implies

$$\tilde{y} \cdot \nu = \sum_{(i,k) \in \mathcal{N}(x)} \tilde{y}_i^k \nu_i^k + \sum_{(i,k) \notin \mathcal{N}(x)} \tilde{y}_i^k \nu_i^k = 0.$$

Since $\tilde{y} \in \mathrm{FP}(x)$ was arbitrary, this proves the desired result. □

Chapter 11

Conclusions

In this dissertation we studied myopic best-response dynamics in large-scale games. Our first main research focus was implementation of MBR dynamics in large games. MBR algorithms can be difficult to implement in large games due to intensive requirements in terms of computational capacity, information overhead, inter-agent communication, and global synchronization. Furthermore, it is sometimes desired for algorithms to converge to pure-strategy equilibria. We studied methods for mitigating each of these issues with a particular focus on network-based settings in which all inter-agent communication is restricted to an overlaid communication graph. In this non-classical setting, information cannot be implicitly transmitted by game play—all information dissemination must be explicitly handled by the learning algorithm.

Our second main research focus concerned fundamental properties of MBR algorithms. We studied the robustness of MBR dynamics to an important class of perturbations that occur in practical real-world applications. We then studied convergence properties of MBR dynamics in potential games. Potential games [34] constitute an important class of multi-agent games with a wide variety of applications in large-scale engineering settings (e.g., [3–5, 8, 13, 15, 48, 48, 49, 63, 93–95]). FP is a highly prototypical MBR algorithm that may be considered the "natural" learning dynamics associated with the NE concept. As noted above, in applications it can be desirable to have an algorithm converge to pure-strategy NE. We showed that in almost all potential games, and for almost all initial conditions, FP converges to a pure-strategy NE. We also developed tools to study the rate of convergence of MBR dynamics in potential games. We showed that the rate of convergence of FP is exponential in almost all potential games and for almost all initial conditions. We also characterized the convergence rate of ECFP in potential games.

We now recapitulate, by chapter, the major contributions of the dissertation.

Chapter 3: Empirical Centroid Fictitious Play

In order to mitigate the problem of high information overhead, we considered a form of MBR learning in which the dimension of the observation space is not dependent on the number of players, but instead, is dependent on the number of equivalence classes into which players are grouped. The resultant algorithm, termed ECFP, mitigates the problem of information overhead in MBR algorithms. Additionally, the algorithm can be used to mitigate slow convergence in large games—the rate of convergence of the algorithm is shown empirically to be independent of the number of players, and the continuous-time ECFP dynamics are shown theoretically to converge in finite time to completely mixed equilibria. Discrete and continuous-time ECFP dynamics are studied. It is shown that both dynamics converge to the set of symmetric NE.

Chapter 4: Robustness Properties in Myopic Best-Response Algorithms

In this chapter we studied the robustness of MBR dynamics. We considereed a sequence of asymptotically decaying perturbations at the output of the best response. We show that MBR dynamics are robust in the sense that, if the MBR dynamics converged to some equilibrium set in the absence of perturbations, then they will also converge in the presence of such perturbations. In subsequent chapters, we studied methods to mitigate the issues of complexity, information overhead, global synchronization, and demanding communication requirements using this robustness result.

Chapter 5: A Single Sample Approach

In this chapter we studied a sampling-based method for mitigating complexity in MBR algorithms. The computationally expensive aspect of computing the best response is evaluating the expected utility. We considered a method for implementing MBR dynamics in which players approximate the expected utility by drawing one sample from an underlying distribution in each round and averaging over time using stochastic-approximation techniques. It is shown that a single sample MBR process converges to the internally chain recurrent set of the associated differential inclusion. The single sample approach reduces the per-iteration computational complexity of implementing a MBR algorithm from $O(e^N)$ to $O(N)$, where N is the number of players.

Chapter 6: Asynchronous Implementation

In the classical setup of repeated play learning, it is assumed that agents act in a synchronous manner. In large-scale distributed settings, global synchronization can be impractical. We show that MBR dynamics can be practically implemented in an asynchronous fashion, and, using the robustness result of Chapter 4, we derive conditions under which an asynchronous MBR will converge to the internally chain recurrent set of the associated differential inclusion. We define a generic notion of asynchronous repeated play learning in which player may be "active" in some rounds and "idle" in others. We first prove convergence of asynchronous MBR dynamics in this setting. Subsequently, we study continuous-time embedded MBR processes. Such processes model real-world implementations in which agents choose a countable sequence of actions, but the timing of their actions occurs in continuous time. Convergence under a mild condition was proven in this setting, and practical prototypical implementations using stochastic and deterministic action timing rules were given.

Chapter 7: Network-Based Implementation

In this chapter we studied the problem of implementing a MBR algorithm in a network-based setting. In a network-based setting, physical payoff measurements are not available, and it is not possible to implicitly transmit information via game play. The task of information dissemination must be handled directly by the learning algorithm. A generic network-based MBR algorithm was presented and, using the robustness result, we derived a condition under which such an algorithm can be shown to converge. Subsequently, we developed network-based implementations of the ECFP algorithm presented in Chapter 3 and the single sample FP algorithm presented in Chapter 5. The network-based ECFP algorithm was shown to converge to the set of symmetric NE and the SSFP algorithm was shown to converge to the set of NE.

Chapter 8: Inertial MBR Algorithms: Incomplete Information and Network-Based Implementation

The use of inertia is a common technique used to ensure convergence to pure-strategy NE in classical settings where players have complete information and may observe the actions of other players or possibly their own payoffs. In this chapter we show that the use of so-called inertial techniques can be used to ensure convergence of MBR dynamics to pure-strategy NE in network-based settings and settings where players have incomplete information about the state of the world. We studied general MBR dynamics with inertia and proved that under a mild assumption, any such algorithm converges to a pure-strategy NE. We then studied MBR algorithms with inertia in a network-based setting with incomplete information. We focused on two particular MBR algorithms: FP with inertia and JSFP with inertia. We showed that so long as the information dissemination scheme satisfies a mild assumption, each algorithm will converge to a pure strategy NE. We then explicitly defined an information dissemination scheme and proved convergence of each algorithm to a pure strategy NE. Network-based FP with inertia and network-based JSFP with inertia differ primarily in terms of their information requirements. Network-based JSFP with inertia operates using less information than network-based FP with inertia. However, due to the manner in which information is disseminated in network-based JSFP with inertia, it is applicable within a narrower class of games than network-based FP with inertia.

Chapter 9: Non-Degenerate Potential Games

In this chapter we presented a notion of non-degeneracy for potential games. In a non-degenerate potential game, FP dynamics are "well behaved" (shown in Chapter 10) in the sense that they almost always converge to pure-strategy NE, and the rate of convergence can be estimated. It was shown that the notion of a non-degenerate potential game coincides with other well known notions of non-degeneracy in general N-player games. The main contribution of the chapter was a proof that almost all potential games are non-degenerate. It was shown that in almost all potential games, the number of NE is finite and odd.

Chapter 10: Myopic Best-Response Dynamics in Potential Games

The chapter studied continuous-time MBR processes in potential games. The Chapter has three main results. First, it is shown that in almost all potential games, and for almost all initial conditions, FP converges to a pure-strategy NE. Second, it is shown that in almost all potential games and for almost all initial conditions, the rate of convergence of FP is exponential. Third, it is shown that if a MBR process converges to a completely mixed-strategy equilibrium (a relatively typical occurrence), then convergence occurs in finite time.

11.1 Future Work

We now briefly highlight some promising future research directions.

In studying ECFP, we assumed that the utility function satisfies a symmetry condition amongst players in a given group. It would be interesting to study ECFP dynamics in games that *nearly* satisfy this symmetry condition. In particular, in games non-symmetric games, symmetric NE may not exist—it would be interesting to study if the centroid tuple $\bar{q}(n)$ converges to a *nearly* optimal (symmetric) strategy in such games. It may also be interesting to study methods to optimally cluster players in order to nearly satisfy the symmetry condition. This research direction may permit ECFP to be used as a method for reducing information overhead and accelerating convergence in a broader class of games.

ECFP may be seen as solving a type of generalized NE problem [159], in which it is desired to find an equilibrium within a constrained subset of the strategy space. It may be interesting to study MBR dynamics as a more general tool for solving these types of problems.

Simulation results show that in the case where players are grouped into a single group, the convergence rate tends to be independent of the number of players. A potential future research direction may include a rigorous characterization of the rate of convergence of ECFP in terms of both the number of players and the number of groups.

The convergence rate estimate for ECFP (see Section 10.6) is only shown to hold when the process converges to a completely mixed-strategy equilibrium. Using the projection techniques developed in Section 10.3, it may be possible to show a similar convergence result for any (possibly incompletely) mixed-strategy equilibrium.

In this dissertation we studied asynchronous MBR dynamics and network-based learning algorithms. However, we did not study asynchronous MBR dynamics in a network-based setting. This is an important future research direction since the assumption of global synchronization tends to be impractical in a large-scale network-based setting. In an asynchronous network-based MBR algorithm, players could choose the timing of their actions and the timing of communication according to a homogenous Poisson process, as discussed in Section 6.4.1. Using such a timing rule, information could be disseminated in an asynchronous manner using an asynchronous gossip algorithm [21]. In future research, this could be made explicit and rigorous.

The study of fundamental properties of MBR dynamics in Chapter 10 sews fertile ground for several future research topics. In [153] it was shown that the rate of convergence in FP can be made arbitrarily slow in potential games by choosing an appropriate (nearly generate) game. In effect, the coefficient in the exponential convergence rate estimate for FP (see Section 10.5) can be made arbitrarily large by choosing the potential game to be nearly degenerate. In future research, we plan to investigate establishing uniform convergence rate estimates by removing small sets of "nearly degenerate" games.

It was shown that continuous-time FP almost always converges to a pure-strategy NE. It would be interesting to study an extension of this result to discrete-time FP. The extension is not straightforward. In certain games, the interpolated discrete-time process can "overshoot" sets that would normally draw the process away from a mixed equilibrium and towards a pure equilibrium. This overshooting phenomenon must be accounted for when characterizing the class of games in which discrete-time FP converges generically to a pure-equilibrium. Based on simulation results, we hypothesize that discrete-time FP also converges to pure-strategy equilibria in almost all potential games.

Another worthwhile future research direction may be the study of discrete-time FP with a constant step-size. In particular, we hypothesize that constant step-size FP converges to a neighborhood of a pure NE in almost all potential games. Constant step-size FP does not not slow down over time as does traditional FP in which the step size is diminishing. Consequently, constant step-size FP may be useful for learning and control in time-varying environments.

Bibliography

[1] S. Lasaulce, M. Debbah, and E. Altman, "Methodologies for analyzing equilibria in wireless games," *IEEE Signal Processing Magazine*, vol. 5, no. 26, pp. 41–52, 2009.

[2] L. Rose, S. Lasaulce, S. M. Perlaza, and M. Debbah, "Learning equilibria with partial information in decentralized wireless networks," *IEEE Communications Magazine*, vol. 49, no. 8, pp. 136–142, 2011.

[3] X. Chu and H. Sethu, "Cooperative topology control with adaptation for improved lifetime in wireless ad hoc networks," in *Proceedings of IEEE International Conference on Computer Communications*. IEEE, 2012, pp. 262–270.

[4] J. Zheng, Y. Cai, Y. Liu, Y. Xu, B. Duan, and X. S. Shen, "Optimal power allocation and user scheduling in multicell networks: Base station cooperation using a game-theoretic approach," *IEEE Transactions on Wireless Communications*, vol. 13, no. 12, pp. 6928–6942, 2014.

[5] Y. Xu, A. Anpalagan, Q. Wu, L. Shen, Z. Gao, and J. Wang, "Decision-theoretic distributed channel selection for opportunistic spectrum access: Strategies, challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1689–1713, 2013.

[6] J. R. Marden, S. D. Ruben, and L. Y. Pao, "A model-free approach to wind farm control using game theoretic methods," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 4, pp. 1207–1214, 2013.

[7] W. Saad, Z. Han, H. V. Poor, and T. Basar, "Smart meters for power grid: Challenges, issues, advantages and status," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 86–105, 2012.

[8] G. Scutari, S. Barbarossa, and D. P. Palomar, "Potential games: A framework for vector power control problems with coupled constraints," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4. IEEE, 2006, p. IV.

[9] Z. Ma, D. S. Callaway, and I. A. Hiskens, "Decentralized charging control of large populations of plug-in electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 1, pp. 67–78, 2013.

[10] W. Tushar, W. Saad, H. V. Poor, and D. B. Smith, "Economics of electric vehicle charging: A game theoretic approach," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1767–1778, 2012.

[11] C. Wu, H. Mohsenian-Rad, J. Huang, and A. Y. Wang, "Demand side management for wind power integration in microgrid using dynamic potential game theory," in *Global Communications Conference Workshops*. IEEE, 2011, pp. 1199–1204.

[12] G. Arslan, J. Marden, and J. Shamma, "Autonomous vehicle-target assignment: a game-theoretical formulation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 129, no. 5, pp. 584–596, 2007.

[13] C. Ding, B. Song, A. Morye, J. A. Farrell, and A. K. Roy-Chowdhury, "Collaborative sensing in a distributed PTZ camera network," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3282–3295, 2012.

[14] M. S. Stankovic, K. H. Johansson, and D. M. Stipanovic, "Distributed seeking of Nash equilibria with applications to mobile sensor networks," *IEEE Transactions on Automatic Control*, vol. 57, no. 4, pp. 904–919, 2012.

[15] M. Zhu and S. Martínez, "Distributed coverage games for energy-aware mobile sensor networks," *SIAM Journal on Control and Optimization*, vol. 51, no. 1, pp. 1–27, 2013.

[16] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press, 2007, vol. 1.

[17] M. Wooldridge, *An introduction to multiagent systems*. John Wiley & Sons, 2009.

[18] J. R. Marden and J. S. Shamma, "Game theory and distributed control," *Handbook of Game Theory*, vol. 4, 2012.

[19] D. Fudenberg and J. Tirole, *Game theory*. The MIT Press, 1991.

[20] D. Fudenberg and D. K. Levine, *The theory of learning in games*. MIT press, 1998, vol. 2.

[21] A. G. Dimakis, S. Kar, J. M. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[22] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[23] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3375–3380, 2008.

[24] J. F. Mota, J. M. Xavier, P. M. Aguiar, and M. Puschel, "D-ADMM: A communication-efficient distributed algorithm for separable optimization," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2718–2723, 2013.

[25] D. Bajovic, J. Xavier, B. Sinopoli, J. M. Moura, and D. Jakovetic, "Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4381–4396, 2011.

[26] G. W. Brown, "Iterative solutions of games by fictitious play," in *Activity Analysis of Production and Allocation*. Wiley, 1951, pp. 374–376.

[27] C. Harris, "On the rate of convergence of continuous-time fictitious play," *Games and Economic Behavior*, vol. 22, no. 2, pp. 238–259, 1998.

[28] M. Benaïm, J. Hofbauer, and S. Sorin, "Stochastic approximations and differential inclusions," *SIAM Journal on Control and Optimization*, vol. 44, no. 1, pp. 328–348, 2005.

[29] D. S. Leslie and E. Collins, "Generalised weakened fictitious play," *Games and Economic Behavior*, vol. 56, no. 2, pp. 285–298, Aug. 2006.

[30] V. S. Borkar, *Stochastic approximation: A dynamical systems viewpoint*. Cambridge Univeristy Press, 2008.

[31] J. Robinson, "An iterative method of solving a game," *The Annals of Mathematics*, vol. 54, no. 2, pp. 296–301, Sep. 1951.

[32] J. Hofbauer and S. Sorin, "Best response dynamics for continuous zero-sum games," *Discrete and Continuous Dynamical Systems Series B*, vol. 6, no. 1, p. 215, 2006.

[33] K. Miyasawa, "On the convergence of the learning process in a 2×2 non-zero-sum two-person game," Economic Memorandum No. 33, Economic Research Program, Princeton University, Tech. Rep., 1961.

[34] D. Monderer and L. S. Shapley, "Fictitious play property for games with identical interests," *Journal of Economic Theory*, vol. 68, no. 1, pp. 258–265, Jan. 1996.

[35] U. Berger, "Fictitious play in 2×n games," *Journal of Economic Theory*, vol. 120, no. 2, pp. 139–154, 2005.

[36] J. R. P. Milgrom, "Adaptive and sophisticated learning in normal form games," *Games and Economic behavior*, vol. 3, pp. 82–100, 1991.

[37] A. Sela, "Fictitious play in 'one-against-all' multi-player games," *Economic Theory*, vol. 14, no. 3, pp. 635–651, 1999.

[38] D. Monderer and L. S. Shapley, "Potential Games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, May 1996.

[39] S. Hart and A. Mas-Colell, "Uncoupled dynamics do not lead to Nash equilibrium," *The American Economic Review*, vol. 93, no. 5, pp. 1830–1836, 2003.

[40] G. Ostrovski and S. van Strien, "Payoff performance of fictitious play," 2013, submitted for journal publication, https://arxiv.org/abs/1308.4049.

[41] V. Conitzer, "Approximation guarantees for fictitious play," in *47th Annual Allerton Conference on Communication, Control, and Computing.* IEEE, 2009, pp. 636–643.

[42] E. Hendon, H. J. Jacobsen, and B. Sloth, "Fictitious play in extensive form games," *Games and Economic Behavior*, vol. 15, no. 2, pp. 177–202, 1996.

[43] J. Heinrich, M. Lanctot, and D. Silver, "Fictitious self-play in extensive-form games," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, pp. 805–813.

[44] S. Ganzfried and T. Sandholm, "Computing an approximate jam/fold equilibrium for 3-player no-limit Texas Hold'em tournaments," in *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, vol. 2. International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 919–925.

[45] T. Sandholm, "The state of solving large incomplete-information games, and application to poker," *AI Magazine*, vol. 31, no. 4, pp. 13–32, 2010.

[46] J. Shi and M. L. Littman, "Abstraction methods for game theoretic poker," in *International Conference on Computers and Games.* Springer, 2000, pp. 333–345.

[47] S. Ganzfried, "Computing strong game-theoretic strategies in jotto," 2016, submitted for journal publication, https://arxiv.org/pdf/1107.3342v3.pdf.

[48] T. J. Lambert, M. A. Epelman, and R. L. Smith, "A fictitious play approach to large-scale optimization," *Operations Research*, vol. 53, no. 3, pp. 477–489, May 2005.

[49] A. Garcia, D. Reaume, and R. L. Smith, "Fictitious play for finding system optimal routings in dynamic traffic networks," *Transportation Research Part B: Methodological*, vol. 34, no. 2, pp. 147–156, 2000.

[50] A. Chapman, A. Rogers, N. R. Jennings, and D. Leslie, "A unifying framework for iterative approximate best–response algorithms for distributed constraint optimisation problems," *The Knowledge Engineering Review*, vol. 26, no. 4, pp. 411–444, 2011.

[51] M. Smyrnakis and S. M. Veres, "Fictitious play for cooperative action selection in robot teams," *Engineering Applications of Artificial Intelligence*, vol. 56, pp. 14–29, 2016.

[52] M. Epelman, A. Ghate, and R. L. Smith, "Sampled fictitious play for approximate dynamic programming," *Computers & Operations Research*, vol. 38, no. 12, pp. 1705–1718, 2011.

[53] A. Ghate, S.-F. Cheng, S. Baumert, D. Reaume, D. Sharma, and R. L. Smith, "Sampled fictitious play for multi-action stochastic dynamic programs," *IIE Transactions*, vol. 46, no. 7, pp. 742–756, 2014.

[54] K. Dabcevic, A. Betancourt, L. Marcenaro, and C. S. Regazzoni, "A fictitious play-based game-theoretical approach to alleviating jamming attacks for cognitive radios," in *IEEE International Conference on Acoustics, Speech and Signal Prcessing*, 2014, pp. 8158–8162.

[55] ——, "Intelligent cognitive radio jamming—a game-theoretical approach," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1–18, 2014.

[56] B. Wang, Y. Wu, and K. R. Liu, "Game theory for cognitive radio networks: An overview," *Computer Networks*, vol. 54, no. 14, pp. 2537–2561, 2010.

[57] E. Sisikoglu, M. A. Epelman, and R. L. Smith, "A sampled fictitious play based learning algorithm for infinite horizon Markov decision processes," in *Proceedings of the Winter Simulation Conference*, 2011, pp. 4086–4097.

[58] O. Candogan, A. Oxdaglar, and P. Parrilo, "Flows and decompositions of games: Harmonic and potential games," *Mathematics of Operations Research*, vol. 36, no. 3, pp. 474–503, 2011.

[59] J. S. Jordan, "Three problems in learning mixed-strategy Nash equilibria," *Games and Economic Behavior*, vol. 5, no. 3, pp. 368–386, Jul. 1993.

[60] G. Arslan, J. R. Marden, and J. S. Shamma, "Autonomous vehicle-target assignment: A game-theoretical formulation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 129, no. 5, pp. 584–596, 2007.

[61] J. R. Marden, H. P. Young, G. Arslan, and J. S. Shamma, "Payoff-based dynamics for multi-player weakly acyclic games," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 373–396, 2009.

[62] J. R. Marden and A. Wierman, "Overcoming limitations of game-theoretic distributed control," in *Proceedings of the 48th IEEE Conference on Decision and Control, held jointly with the 28th Chinese Control Conference.* IEEE, 2009, pp. 6466–6471.

[63] N. Li and J. R. Marden, "Designing games for distributed optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 230–242, 2013.

[64] J. Hofbauer, "Stability for the best response dynamics," Institut für Mathematik, Universität Wien, Tech. Rep., 1995.

[65] J. C. Harsanyi, "Oddness of the number of equilibrium points: A new proof," *International Journal of Game Theory*, vol. 2, no. 1, pp. 235–250, 1973.

[66] E. Van Damme, *Stability and perfection of Nash equilibria.* Springer, 1991, vol. 339.

[67] ——, *Refinements of the Nash equilibrium concept.* Springer Science & Business Media, 2012, vol. 219.

[68] F. Brandt and F. Fischer, "On the hardness and existence of quasi-strict equilibria," in *International Symposium on Algorithmic Game Theory.* Springer, 2008, pp. 291–302.

[69] H. Norde, "Bimatrix games have quasi-strict equilibria," *Mathematical Programming*, vol. 85, no. 1, pp. 35–49, 1999.

[70] M. Jansen, "Regularity and stability of equilibrium points of bimatrix games," *Mathematics of Operations Research*, vol. 6, no. 4, pp. 530–550, 1981.

[71] S. Hart and Y. Mansour, "How long to equilibrium? The communication complexity of uncoupled equilibrium procedures," *Games and Economic Behavior*, vol. 69, no. 1, pp. 107–126, 2010.

[72] B. Van der Genugten, "A weakened form of fictitious play in two-person zero-sum games," *International Game Theory Review*, vol. 2, no. 04, pp. 307–328, 2000.

[73] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 208–220, Feb. 2009.

[74] H. Young, *Strategic learning and its limits.* Oxford University Press, 2004.

[75] B. Swenson, S. Kar, and J. Xavier, "Single simple fictitious play," 2017, IEEE Transactions on Automatic Control. Scheduled for publication, November 2017. [Online]. Available: https://arxiv.org/abs/1506.04320.

[76] D. Fudenberg and D. K. Levine, "Consistency and cautious fictitious play," *Journal of Economic Dynamics and Control*, vol. 19, no. 5, pp. 1065–1089, Jul. 1995.

[77] B. Swenson, C. Eksin, S. Kar, and A. Ribeiro, "Fictitious play with inertia learns pure equilibria in distributed games with incomplete information," 2017, IEEE Transactions on Automatic Control. Submitted. Initial Submission: Jan. 2017. [Online]. Available: https://arxiv.org/abs/1605.00601.

[78] B. Swenson, S. Kar, and J. Xavier, "Distributed learning in large-scale multi-agent games: A modified fictitious play approach," in *46th Asilomar Conference on Signals, Systems, and Computers*, 2012, pp. 1490–1495.

[79] J. Koshal, A. Nedic, and U. V. Shanbhag, "A gossip algorithm for aggregative games on graphs," in *Proceedings of the 51st IEEE Conference on Decision and Control*, 2012, pp. 4840–4845.

[80] X. Chen and J. Huang, "Spatial spectrum access game: Nash equilibria and distributed learning," in *Proceedings of the 13th ACM International Symposium on Mobile Ad Hoc Networking and Computing.* ACM, 2012, pp. 205–214.

[81] B. Gharesifard and J. Cortes, "Distributed convergence to Nash equilibria in two-network zero-sum games," *Automatica*, vol. 49, no. 6, pp. 1683–1692, 2013.

[82] O. N. Gharehshiran, V. Krishnamurthy, and G. Yin, "Distributed tracking of correlated equilibria in regime switching noncooperative games," *IEEE Transactions on Automatic Control*, vol. 58, no. 10, pp. 2435–2450, 2013.

[83] F. Salehisadaghiani and L. Pavel, "Distributed Nash equilibrium seeking: A gossip-based algorithm," *Automatica*, vol. 72, pp. 209–216, 2016.

[84] J. R. Marden and J. S. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation," *Games and Economic Behavior*, vol. 75, no. 2, pp. 788–808, Jul. 2012.

[85] D. Shah and J. Shin, "Dynamics in congestion games," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 38, no. 1. ACM, 2010, pp. 107–118.

[86] P. Coucheney, B. Gaujal, and P. Mertikopoulos, "Penalty-regulated dynamics and robust learning procedures in games," *Mathematics of Operations Research*, vol. 40, no. 3, pp. 611–633, 2014.

[87] S. Perkins and D. S. Leslie, "Asynchronous stochastic approximation with differential inclusions," *Stochastic Systems*, vol. 2, no. 2, pp. 409–446, 2012.

[88] B. S. Pradelski and H. P. Young, "Learning efficient Nash equilibria in distributed systems," *Games and Economic Behavior*, vol. 75, no. 2, pp. 882–879, Jul. 2012.

[89] A. C. Chapman, D. S. Leslie, A. Rogers, and N. R. Jennings, "Convergent learning algorithms for unknown reward games," *SIAM Journal on Control and Optimization*, vol. 51, no. 4, pp. 3154–3180, Aug. 2013.

[90] Y. Shoham, R. Powers, and T. Grenager, "If multi-agent learning is the answer, what is the question?" *Artificial Intelligence*, vol. 171, no. 7, pp. 365–377, 2007.

[91] A. Matsui, "Best response dynamics and socially stable strategies," *Journal of Economic Theory*, vol. 57, no. 2, pp. 343–362, 1992.

[92] V. Krishna and T. Sjöström, "On the convergence of fictitious play," *Mathematics of Operations Research*, vol. 23, no. 2, pp. 479–511, 1998.

[93] J. R. Marden, G. Arslan, and J. S. Shamma, "Connections between cooperative control and potential games," *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, vol. 39, pp. 1393–1407, Dec. 2009.

[94] N. Nie and C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," *Mobile Networks and Applications*, vol. 11, no. 6, pp. 779–797, 2006.

[95] V. Srivastava, J. O. Neel, A. B. MacKenzie, R. Menon, L. A. DaSilva, J. E. Hicks, J. H. Reed, and R. P. Gilles, "Using game theory to analyze wireless ad hoc networks." *IEEE Communications Surveys and Tutorials*, vol. 7, no. 4, pp. 46–56, 2005.

[96] L. S. Shapley, "Some topics in two-person games," *Advances in Game Theory*, vol. 52, pp. 1–29, 1964.

[97] D. P. Foster and H. P. Young, "On the nonconvergence of fictitious play in coordination games," *Games and Economic Behavior*, vol. 25, no. 1, pp. 79–96, Oct. 1998.

[98] J. Nash, "Equilibrium points in n-person games," *Proceedings of The National Academy of Sciences*, vol. 36, no. 1, pp. 48–49, Jan. 1950.

[99] D. Fudenberg, "Learning Mixed Equilibria," *Games and Economic Behavior*, vol. 5, pp. 320–367, 1993.

[100] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 312–327, Mar. 2005.

[101] G. Arslan and J. S. Shamma, "Distributed convergence to Nash equilibria with local utility measurements," in *Proceedings of the 43rd IEEE Conference on Decision and Control*, vol. 2, 2004, pp. 1538–1543.

[102] N. Xiao, X. Wang, T. Wongpiromsarn, K. You, L. Xie, E. Frazzoli, and D. Rus, "Average strategy fictitious play with application to road pricing," in *American Control Conference*. IEEE, 2013, pp. 1920–1925.

[103] S.-F. Cheng, D. M. Reeves, Y. Vorobeychik, and M. P. Wellman, "Notes on equilibria in symmetric games," in *Proceedings of Workshop on Game Theory and Decision Theory*, 2004.

[104] B. Swenson, S. Kar, and J. Xavier, "Empirical centroid fictitious play: An approach for distributed learning in multi-agent games," *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3888–3901, Aug 2015.

[105] ——, "Mean-centric equilibrium: An equilibrium concept for learning in large-scale games," in *Proceedings of IEEE Global Conference on Signal and Information Processing*, 2013, pp. 571–574.

[106] A. N. Shiryaev, *Probability*, ser. Graduate Texts in Mathematics. Springer-Verlag, 1996, vol. 95.

[107] B. Swenson, S. Kar, J. Xavier, and D. S. Leslie, "Robustness properties in fictitious-play-type algorithms," 2016, SIAM Journal on Control and Optimization. Submitted, in 2nd round of review. Initial Submission: Apr. 2016. [Online]. Available: https://arxiv.org/abs/1609.08830.

[108] S. Hurkens, "Learning by forgetful players," *Games and Economic Behavior*, vol. 11, no. 2, pp. 304–329, 1995.

[109] M. W. Hirsch, *Differential topology.* Springer, 1976.

[110] T. J. Lambert and H. Wang, "Fictitious play approach to a mobile unit situation awareness problem," *Univ. Michigan, Tech. Rep*, 2003.

[111] B. Swenson, S. Kar, and J. Xavier, "A computationally efficient implementation of fictitious play in a distributed setting," in *Proceedings of European Signal Processing Conference*, 2015.

[112] ——, "Computationally efficient learning in large-scale games: Sampled fictitious play revisited," in *50th Asilomar Conference on Signals, Systems and Computers.* IEEE, 2016, pp. 1212–1215.

[113] M. Kearns, M. L. Littman, and S. Singh, "Graphical models for game theory," in *Proceedings of the 17th conference on Uncertainty in Artificial Intelligence.* Morgan Kaufmann Publishers Inc., 2001, pp. 253–260.

[114] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games.* Cambridge University Press, 2006.

[115] F. Germano and G. Lugosi, "Global Nash convergence of Foster and Young's regret testing," *Games and Econonomic Behavior*, vol. 60, no. 1, pp. 135–154, 2007.

[116] V. S. Borkar, "Stochastic approximation with two time scales," *Systems & Control Letters*, vol. 29, no. 5, pp. 291–294, 1997.

[117] D. S. Leslie and E. Collins, "Convergent multiple-timescales reinforcement learning algorithms in normal form games," *The Annals of Applied Probability*, vol. 13, no. 4, pp. 1231–1251, 2003.

[118] M. Loève, *Probability Theory I.* Springer-Verlag, 1977.

[119] H. Robbins and D. Siegmund, "A convergence theorem for non negative almost supermartingales and some applications," in *Herbert Robbins Selected Papers.* Springer, 1985, pp. 111–135.

[120] B. Swenson, S. Kar, and J. Xavier, "On asynchronous implementations of fictitious play for distributed learning," in *49th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2015, pp. 1119–1124.

[121] ——, "On robustness properties in empirical centroid fictitious play," in *IEEE Conference on Decision and Control*. IEEE, 2015, pp. 3324–3330.

[122] D. Williams, *Probability with martingales*. Cambridge University Press, 1991.

[123] C. Kiekintveld, M. Jain, J. Tsai, J. Pita, F. Ordóñez, and M. Tambe, "Computing optimal randomized resource allocations for massive security games," in *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, vol. 1. International Foundation for Autonomous Agents and Multiagent Systems, 2009, pp. 689–696.

[124] T. Alpcan and T. Basar, *Network security: A decision and game-theoretic approach*. Cambridge University Press, 2010.

[125] M. Voorneveld, "Pareto-optimal security strategies as minimax strategies of a standard matrix game," *Journal of Optimization Theory and Applications*, vol. 102, no. 1, pp. 203–210, 1999.

[126] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, 2000.

[127] B. Swenson, S. Kar, and J. Xavier, "Strong convergence to mixed equilibria in fictitious play," in *48th Annual Conference on Information Sciences and Systems*. IEEE, 2014, pp. 1–6.

[128] S. Kar and J. M. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1383–1400, 2010.

[129] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, June 2003.

[130] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4289–4305, Aug. 2012.

[131] ——, "Distributed pareto optimization via diffusion strategies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 205–220, Apr. 2013.

[132] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, Dec. 2010.

[133] K. Srivastava and A. Nedic, "Distributed asynchronous constrained stochastic optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 772–790, Aug. 2011.

[134] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[135] S. Rajagopalan and D. Shah, "Distributed averaging in dynamic networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 845–854, Aug. 2011.

[136] S. Kar and J. M. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 674–690, Aug. 2011.

[137] D. Shah, "Gossip algorithms," *Foundations and Trends in Networking*, vol. 3, no. 1, pp. 1–125, 2008.

[138] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*. Cambridge University Press, 1988.

[139] G. Bacci, S. Lasaulce, W. Saad, and L. Sanguinetti, "Game theory for networks: A tutorial on game-theoretic tools for emerging signal processing applications," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 94–119, 2016.

[140] C. Eksin, B. Swenson, S. Kar, and A. Ribeiro, "Learning pure-strategy Nash equilibria in networked multi-agent systems with uncertainty," in *IEEE 55th Conference on Decision and Control*. IEEE, 2016, pp. 5292–5297.

[141] F. Salehisadaghiani and L. Pavel, "Nash equilibrium seeking by a gossip-based algorithm," in *IEEE 53rd Annual Conference on Decision and Control*. IEEE, 2014, pp. 1155–1160.

[142] C. Eksin and A. Ribeiro, "Distributed fictitious play in potential games of incomplete information," in *IEEE 54th Annual Conference on Decision and Control*, 2015, pp. 5190–5196.

[143] A. Fabrikant, A. D. Jaggard, and M. Schapira, "On the structure of weakly acyclic games," in *Algorithmic Game Theory*. Springer, 2010, pp. 126–137.

[144] R. Olfati-Saber and J. S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *44th IEEE Conference on Decision and Control, and European Control Conference.* IEEE, 2005, pp. 6698–6703.

[145] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 622–633, 2008.

[146] S. Kar and J. M. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 355–369, 2009.

[147] A. H. Sayed, "Diffusion adaptation over networks," *Academic Press Library in Signal Processing*, vol. 3, pp. 323–454, 2013.

[148] U. A. Khan, S. Kar, and J. M. Moura, "Higher dimensional consensus: Learning in large-scale networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2836–2849, 2010.

[149] J. C. Harsanyi, "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points," *International Journal of Game Theory*, vol. 2, no. 1, pp. 1–23, 1973.

[150] R. Wilson, "Computing equilibria of n-person games," *SIAM Journal on Applied Mathematics*, vol. 21, no. 1, pp. 80–87, 1971.

[151] V. Klee, R. Ladner, and R. Manber, "Signsolvability revisited," *Linear Algebra and its Applications*, vol. 59, pp. 131–157, 1984. [Online]. Available: http://dx.doi.org/10.1016/0024-3795(84)90164-2

[152] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, revised ed., ser. Textbooks in Mathematics. CRC Press, 2015.

[153] F. Brandt, F. Fischer, and P. Harrenstein, "On the rate of convergence of fictitious play," in *International Symposium on Algorithmic Game Theory.* Springer, 2010, pp. 102–113.

[154] R. A. Brualdi, K. L. Chavey, and B. L. Shader, "Rectangular L-matrices," *Linear Algebra Appl.*, vol. 196, pp. 37–61, 1994. [Online]. Available: http://dx.doi.org/10.1016/0024-3795(94)90314-X

[155] C. Chicone, *Ordinary differential equations with applications.* Springer, 2006, no. 34.

[156] G. Leoni and M. Morini, "Necessary and sufficient conditions for the chain rule in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$ and $\mathrm{BV}_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$," *J. Eur. Math. Soc. (JEMS)*, vol. 9, no. 2, pp. 219–252, 2007. [Online]. Available: http://dx.doi.org/10.4171/JEMS/78

[157] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems.* Oxford University Press, 2000.

[158] G.-Q. Chen, W. P. Ziemer, and M. Torres, "Gauss-Green theorem for weakly differentiable vector fields, sets of finite perimeter, and balance laws," *Communications on Pure and Applied Mathematics*, vol. 62, no. 2, pp. 242–304, 2009.

[159] F. Facchinei and C. Kanzow, "Generalized nash equilibrium problems," *4OR: A Quarterly Journal of Operations Research*, vol. 5, no. 3, pp. 173–210, 2007.