# A Learning Theoretic Approach to Algorithmic Game Theory

by Georgios Piliouras

---

This thesis/dissertation document has been electronically approved by the following individuals:

Tardos,Eva (Chairperson)

Easley,David Alan (Minor Member)

Birman,Kenneth Paul (Minor Member)

Kleinberg,Robert David (Additional Member)

# A LEARNING THEORETIC APPROACH TO ALGORITHMIC GAME THEORY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Georgios Piliouras

August 2010

A LEARNING THEORETIC APPROACH TO ALGORITHMIC GAME THEORY

Georgios Piliouras, Ph.D.

Cornell University 2010

Algorithmic game theory attempts to mathematically capture behavior in strategic situations, in which an individual's success depends on the choices of others. Research in this area tends to focus on one of the following challenges:

- *Price of Anarchy:* Characterize the inefficiency of equilibria vs the global optimum.
- *Computational Complexity of Equilibria*
- *Algorithmic Mechanism Design:* Design games with desirable properties that are efficiently implementable.

Many strategic interactions are in their nature recurrent (e.g. financial markets) with the agents participating in them repeatedly. These agents *learn* over time to adapt to their environment as is defined by the game and the dynamic behavior of the other agents[96]. We show that incorporating the assumption of adaptive agents can lead to exciting new insights to long standing questions in all areas of algorithmic game theory.

**A Learning Theoretic Refinement of the Price of Anarchy:** It is well understood that competitive environments can result in losses to social welfare due to the lack of a central coordinator who could enforce the global optimum solution. Defined as the ratio of the cost of the worst possible (Nash) equilibrium over the cost of the optimum, price of anarchy (and its variants) are commonly

used metrics for these inefficiencies. However, such worst case analysis can be rather misleading as worst case equilibria may never arise in practice. *Can natural learning algorithms beat the price of anarchy?*

We show that in the class of atomic congestion games natural learning algorithms can indeed learn to navigate away from such worst case equilibria. In some cases the implied performance bounds are exponentially better than the previously known worst case guarantees. Furthermore, such positive performance results are shown to be robust even when we impose restrictions on availability of accurate up-to-date information on which agents can base their decisions.

**Learning Inspired Equilibria and Computation:** The plausibility of the Nash equilibrium, as a universal solution concept has been under attack, due to recent negative complexity results (e.g. for normal form games with constant number of players[27, 24]). To make matters worse, there exist games of constant overall size, for which Nash equilibria are unstable for most reasonable dynamics[88].

In this section we will look into learning inspired solution concepts such as *strong CURB (Closed Under Rational Behavior) and CUBR (Closed Under weakly Better Replies) sets*. A strong CURB (CUBR) set is a cartesian product of pure strategy sets such that each player's component contains all best (weakly better) responses to itself given any joint probability distributions of its opponents over the set. A strong CURB (CUBR) set is said to be minimal if it does not contain any proper subset that is also a strong CURB (CUBR) set. Such minimal sets exist in all finite games and are asymptotically stable for a great number of natural learning dynamics. Furthermore, we prove that we can also compute all minimal strong CURB (as well as CUBR) sets for any normal form game with a constant number of players in polynomial time.

**Mechanism Evolution:** The goal of mechanism design is to design games that are uniquely solvable by reasonably self-interested players and which lead to socially optimal outcomes. Such guarantees are of critical importance in the case of economic interactions which are key revenue producers, such as ad-auctions. However, in many cases of social and economic interactions, centralized design is impractical and sometimes even undesirable (i.e. internet). Nevertheless, such mechanisms seem to be working fairly well in practise. How is that possible?

Most mechanisms that we participate in are not the product of "intelligent design". Instead *mechanisms evolve*. All games (mechanisms) that we participate in are in essence social contracts. The rules of these interactions change over time as the result of actions of the same individuals who participate in them. These changes are not random. On the contrary the involved agents are trying to bring on changes to the structure of the game, so as to improve their experience from participating in the game.

We formalize these intuitions and we analyze their implications in the context of oligopolistic markets. Specifically, we allow players one extra dimension in the pursuit of their strategic goals, the possibility of forming coalitions with other agents. We focus on the class of linear and symmetric Cournot games and we study the nature of stable coalition structures. We show than even the worst stable coalition can lead to a significant increase in market prices and profits for the firms.

## BIOGRAPHICAL SKETCH

Georgios Piliouras was born in Athens, Greece on September 28th, 1981. He received his diploma from the Electrical Engineering and Computer Science Department of the National Technical University of Athens, Greece in May 2004. He received a M.S. in Logic and Algorithms from the National and Kapodistrian University of Athens, Greece in August 2006. He received a M.S in Computer Science from Cornell University in August 2009 and expects to receive a PhD in Computer Science from Cornell University in August 2010.

To my parents

## ACKNOWLEDGEMENTS

First and foremost I wish to thank my advisor Éva Tardos. Her unwavering support and encouragement in the pursuit of deep theoretical questions have been invaluable to me.

I was very fortunate to have Bobby Kleinberg join Cornell and introduce me into the world of learning algorithms. His arrival here has been a blessing to me in more ways than I care to count. I would like to thank Ken Birman and David Easley for serving on my committee.

I would also like to thank my advisors during my years in Greece, Stathis Zachos and Elias Koutsoupias. I will always cherish my first explorations in the areas of theoretical computer science under your guidance.

I would like to thank my research partners including Nicole Immorlica, Bobby Kleinberg, Vangelis Markakis, Uri Nadav, Eva Tardos and Ymir Vigfusson.

I would like to thank my friends here at Cornell, in Greece and in every other place where the winds of fate have blown you. Wherever you are you are never far from my thoughts and my heart.

I would like to thank Kimberly for putting a 'we' where my 'I' used to be.

Finally, I would like to thank my parents for their love which made everything possible, my sister Anna for keeping me busy during childhood (and hopefully for many many years to come) and my aunt Peggy for always being there for me.

## TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

CHAPTER 1

# INTRODUCTION

## 1.1 Overview

The central goal of this thesis is to explore the impact of combining learning algorithms and game theory. Game theory attempts to mathematically capture behavior in strategic situations, in which an individual's success depends on the choices of others. In practice, the interacting entities may be numerous and entangled via complex networks of interdependencies. Over the last decade, the prevalence of these issues has risen dramatically following a number of paradigm-shifting events such as the cataclysmic rise of the Internet as a social networking tool, the painful realization of the extent of inter-connectivity of the global economy as well as the necessity of international cooperation for addressing global sustainability concerns. As a result, there has been recorded a swift increase of the interest for a more detailed, realistic and quantitative understanding of such networked interactions.

*Algorithmic game theory* (AGT) employs analytical tools from computer science, such as worst case analysis and complexity theory, to characterize behavioral solutions to strategic situations prescribed by (classical) game theory. Research in this area tends to focus on one of the following challenges:

- *Price of Anarchy:* Characterize the inefficiency of equilibria vs the global optimum
- *Computational Complexity of Equilibria*
- *Algorithmic Mechanism Design:* Design games with desirable properties that are efficiently implementable.

Many strategic interactions are in their nature recurrent (e.g. financial markets) with the agents participating in them repeatedly. These agents *learn* over time to adapt to their environment as is defined by the game and the dynamic behavior of the other agents. Understanding how agents can learn in the presence of other agents that are simultaneously learning constitutes a research problem that is as expansive as it is challenging. Such questions have fueled research endeavors both in economics as well as within computer science (multi-agent learning)[96].

This thesis concentrates on the intersection of algorithmic game theory and learning. Incorporating the rather natural assumption that agents learn to adapt to their environments can lead to exciting new insights to long standing questions. In this chapter, we will provide an informal discussion of the results of this research agenda. The results are organized in three categories, one for each subarea of AGT. Along with their presentation we will briefly discuss how they extend our current understanding of each subarea.

## 1.2 A Learning Theoretic Refinement of the Price of Anarchy

The price of anarchy of a game captures the loss in system performance due to the lack of a centralized authority that could enforce the globally optimal solution. Specifically, the price of anarchy [62] is defined as the ratio of the cost of the worst possible (Nash) equilibrium over the cost of the optimum. Given the assumption that the system is in equilibrium, a low price of anarchy implies that the system performance is close to optimal. In their seminal work, Blum etal [15] showed that some systems with low price of anarchy can be shown to perform well, even under the weaker assumption of learning agents. Surprisingly, in recent papers Blum etal [16] and specifically Roughgarden [85] demonstrate that the above intuition is the norm. If a system provably has low price of anarchy, learning works regardless of whether we reach an equilibrium or not.

On the other hand, a high price of anarchy seems to indicate that such a system is in need of a central coordinator. However, such worst case analysis can be rather misleading as worst case equilibria may never arise in practice. We tackle this issue by modeling players' behavior with learning algorithms and pursuing the following questions: *Can natural learning algorithms beat the price of anarchy? Can they learn to distinguish between good and bad equilibria and converge to the good ones?*

In chapter 3, we investigate these questions for the class of atomic congestion games. These games are rather widespread and capture settings where the experienced performance of an action depends only on the number of other agents using it (e.g. internet/traffic routing, machine scheduling, ...). Unfortunately, the price of anarchy can be arbitrarily bad even for simple congestion games. Surprisingly, we show that under natural behavioral assumptions, agents learn to navigate away from such worst case equilibria. We analyze system behavior under the assumption that agents use the well-known multiplicative updates algorithm (and variants thereof). We use dynamical systems theory to prove that the system converges to a rather restraint subset of equilibria that we call "weakly stable equilibria". Pure Nash equilibria are weakly stable, and we show using techniques from algebraic geometry that the converse is true with probability 1 when congestion costs are selected at random independently on each edge (from any monotonically parameterized distribution). As a result, in many cases the system performance can be shown to beat the price of anarchy and actually be close to optimal.

One point of departure between the theoretical perspective of game theory (or even learning theory) and the reality of distributed systems is the availability of accurate up-to-date information on which agents can base their decisions. In chapter 4, we focus on a concrete model of limited information access and show that the positive performance results extend to this case. Our motivating example was to examine whether the processes of a distributed system can learn to automatically fine-tune the distribution of the system workload. Unlike in traditional game theory, here agents can query the current state of the other parts of the network but cannot reliably predict the effects of their actions on them. In order to predict the evolution of this multi-agent system, we analyze a novel class of ”random” walks which has a negative drift when we are far from the origin. The implied performance bounds are exponentially better than the previously known worst case guarantees.

The common theme throughout this work is that certain networked systems exhibit impressive capabilities in terms efficient self-organization. In such systems placing a ”grain of intelligence” on each node, gives rise to a collective intelligence that guides the system to almost optimal solutions. Understanding the capabilities and harnessing the potential of such systems is an exciting area for future work.

## 1.3 Learning Inspired Equilibria and Computation

The main goal of game theory is to predict the behavior of self-interested interacting agents. Usually solution concepts come in the form of equilibrium notions with Nash equilibrium constituting the predominant such concept. It is widely accepted (at least inside the CS community) that any solution concept should be convincing, universal [1] and efficiently computable[77]. In recent years there has been a lot of work on characterizing the computational complexity of finding Nash equilibria in games. Unfortunately, most results in this area tend to be negative [27] and currently the attention of the research community is mostly focused on approximation algorithms and identifying tractable special cases. Although the importance of these current approaches is beyond doubt, if we take a step back we see that incorporating learning theory into this framework can lead to new insights.

Looking back at the desirable properties of a solution concept, we see that universality and computational efficiency are well-defined, whereas what it means for a solution concept to be convincing is largely open to interpretation. Learning theory could actually provide us with an elegant solution to this predicament. The legitimacy of a solution concept can be characterized by its robustness to multiple learning dynamics. Under this interpretation Nash equilibria can be cast into doubt since most simple dynamics (i.e. reinforcement learning) tend to converge not to Nash equilibria[88] but towards weaker notions of correlated equilibria in general games. On the other hand, not all correlated equilibria can interpreted as the result of natural learning behavior.

A more promising approach would be to consider the set of all (coarse) correlated equilibria, since they are closely connected with the behavior of no-regret dynamics (see Chapter 2). Unfortunately, correlated equilibria offer no insights on the behavior of some rather common dynamics such as best-response dynamics. Sink equilibria[46] on the other hand are specifically tailored to capture the long-run behavior of best response dynamics, but are inconsistent with any other learning behavior (e.g. AMS-dynamics).

[1] exist for all games

In chapter 5, we will look into learning inspired solution concepts such as *strong CURB (Closed Under Rational Behavior) and CUBR (Closed Under weakly Better Replies) sets*. A strong CURB (CUBR) set is a cartesian product of pure strategy sets such that each player's component contains all best (weakly better) responses to itself given any joint probability distributions of its opponents over the set. A strong CURB (CUBR) set is said to be minimal if it does not contain any proper subset that is also a strong CURB (CUBR) set. Such minimal sets exist in all finite games and are asymptotically stable for a great number of natural learning dynamics. Furthermore, we prove that we can also compute all minimal strong CURB (as well as CUBR) sets for any normal form game with a constant number of players in polynomial time. As a result, these concepts are shown to be *universal, robust to different learning dynamics and efficiently computable* for any normal form game with a constant number of players.

These solution concepts are closely related to the notion of CURB (Closed Under Rational Behavior) sets, introduced by Basu and Weibull [11] and CUBR (Closed Under Better Responses) sets [81], introduced by Ritzberger and Weibull. A CURB (CUBR) set is a cartesian product of pure strategy sets such that each player's component contains all best (weakly better replies) replies to itself given any product of mixed actions of his opponents with support within their CURB (CUBR) sets. CURB sets can be computed in polynomial time in the case of extensive games[79] as well as two player normal forms games [13]. In the case of finite extensive forms games, the notion of primitive formations (Harsanyi and Selten[50]) captures the property of strong CURB sets. Finally, in independent working papers Klimm and Weibull[59] as well as Zapechelnyck [108] concurrently define the same notion of strong CURB sets under the names sCURB sets and strict curb sets, respectively. In [59], Klimm and Weibull also present an algorithm for finding all minimal sCURB sets of finite games. In finite games, the notion of strong CUBR sets is equivalent to that of CUBR sets, so our results extend to CUBR sets as well. The advantage of the strong CUBR set definition is that will allow for a unified treatment of strong CURB and CUBR sets.

This line of research shows that by focusing on identifying attractors of natural learning dynamics, we can derive solutions concepts which have better computational properties than Nash. In fact, given that Nash equilibria can be rather poor predictors of the players' behavior[88], by following learning-oriented solution concepts, we can improve simultaneously both the quality of our predictions as well as their computational tractability. In order to further increase the precision of our solutions beyond these set-valued concepts, we need to focus on specific classes of learning algorithms. Exploring the tradeoffs between the size of set of allowable learning algorithms and the precision of the corresponding attractors while maintaining computational tractability is the natural next step in this line of work.

## 1.4 Mechanism Evolution

The goal of mechanism design is to design games that are uniquely solvable by reasonably self-interested players and which lead to socially optimal outcomes. Such guarantees are of critical importance in the case of economic interactions which are key revenue producers, such as ad-auctions. However, in many cases of social and economic interactions, centralized design is impractical and sometimes even undesirable (i.e. internet). Nevertheless, such mechanisms seem to be working fairly well in practise. How is that possible?

Most mechanisms that we participate in are not the product of "intelligent design". Instead *mechanisms evolve.* All games (mechanisms) that we participate in are in essence social contracts. The rules of these interactions change over time as the result of actions of the same individuals who participate in such games. These changes are not random. On the contrary the involved agents are trying to bring on changes to the structure of the game, so as to improve their experience from participating in the game.

We can formalize these intuitions by looking at mechanism evolution as a meta-game. The players in this game are the same players that participate in the current iteration of the evolving mechanism. The utility of a player in the current iteration of the game, is naturally her expectation of profit for participating in the game. Finally, the allowable actions in the meta-game express the power of each player to bring on changes to the game structure. This power depends on the nature of the social interaction. For example, in the case of economic markets the participating firms can form (or destroy) coalitions, so as to change the current balance of powers.

Once we have captured mechanism evolution as a meta-game, we can now apply the usual game theoretic thinking to this context. Specifically, we can conclude that not all games are stable[2]. In fact, if we model accurately the mechanism evolution, we would expect that the stable instances in this meta-game should capture the characteristics of the games that are employed in practice. A natural question that arises is how do these games fair in comparison to an arbitrary (unstable) game of the same class. We answer such questions for the case of oligopolistic markets. Specifically, we address how dynamic coalition formation affects the social welfare. In order to answer this question however, first of all we need to be able to define the meta-game and specifically how the producers evaluate a specific market that they participate in.

[2]For any notion of stability (equilibrium) that we choose.

Cournot and Bertrand oligopolies constitute the two most prevalent models of firm competition. The analysis of Nash equilibria in each model reveals, as desired, a unique prediction about the system. Quite alarmingly, despite that both models are plausible interpretations of the function of economic markets, their projections expose a stark dichotomy. Under the Cournot model, where firms compete by strategically managing their output quantity, firms enjoy positive profits as the resulting market prices exceed that of the marginal costs. On the contrary, the Bertrand model, in which firms compete on price, predicts that a duopoly is enough to push prices down to the marginal cost level. This suggestion that duopoly will result in perfect competition, is commonly referred to in the economics literature as the "Bertrand paradox".

In Chapter 6 we analyze these models in disequilibrium under minimal behavioral hypotheses. Specifically, we assume that firms adapt their strategies over time, so that in hindsight their average payoffs are not exceeded by any single deviating strategy. Given this no-regret guarantee, we show that in the case of Cournot oligopolies, the unique Nash equilibrium fully captures the emergent behavior. Notably, we prove that under natural assumptions the daily market characteristics converge to the unique Nash. In contrast, in the case of Bertrand oligopolies, a wide range of positive average payoff profiles can be sustained. Hence, under the assumption of self-interested adapting agents, the Bertrand paradox is resolved and both models arrive to the same conclusion that increased competition is necessary in order to achieve perfect pricing.

In the case of Cournot markets we have seen that players can reasonably estimate their expected average payoffs from participating in the market. We can now define (chapter 7) a model of a coalition formation process that can be applied on top of such a market. This defines an evolutionary mechanism where agents can explore one extra dimension in the pursuit of their strategic goals, the possibility of forming coalitions with other agents. We focus on the class of linear and symmetric Cournot oligopoly games, and we study the nature of stable coalition structures, i.e., partitions where no profitable deviation exists according to the rules of our coalition formation process. We prove that the ratio between the social welfare of the worst stable partition and the optimum social welfare is $\Theta(n^{2/5})$, where $n$ is the number of firms that participate in the market. We denote this ratio as the price of anarchy of the Cournot coalition formation game and we note that this implies a significant improvement of the actual price of anarchy of Cournot oligopolies which is known to be $\Theta(n)$. Notably, we show that all results are robust even under weak (no-regret) behavioral assumptions.

## 1.5 Chapter Organization & Dependencies

Chapter 2 contains some basic facts about learning in games. Chapters 3 and 4 focus on a learning theoretic approach to the price of anarchy. Specifically, in these chapters we analyze the performance of variants of multiplicative weights algorithms in families of congestion games. In chapter 5 we consider the notions of strong CURB (as well as CUBR) sets and provide positive computational results for these concepts. Chapters 6 and 7 focus on the setting of oligopolistic markets. Specifically, in chapter 6 we analyze both Cournot and Bertrand markets under no-regret behavior. In chapter 7 we look into issues that arise in Cournot markets when players can learn to form coalitions.

CHAPTER 2

# LEARNING IN GAMES

## 2.1 Basic Definitions

A *strategic game* is a triple $(N; (S_i)_{i \in N}; (u_i)_{i \in N})$ where $N$ is the set of players and for every player $i \in N$, $S_i$ is the set of *(pure) strategies (or actions)* of player $i$, and the *utility function* $u_i$ is a real valued function defined on $S = \times_{i \in N} S_i$ . For every strategy profile $s \in S$, $u_i(s)$ represents the payoff (positive utility) to player $i$. For any strategy profile $s \in S$ and any strategy $s'_i$ of player $i$ we use $(s_{-i}, s'_i)$ to denote the strategy profile that we derive by substituting the $i$-th coordinate of the strategy profile $s$ with $s'_i$.

A strategy profile $s$ is a *Nash equilibrium* if $u_i(s) \geq u_i(s_{-i}, s'_i)$ for every $s'_i$ and every $i \in N$. Analogously, a *Nash $\epsilon$-equilibrium* is defined as a strategy profile $s$ such that $u_i(s) \geq u_i(s_{-i}, s'_i) - \epsilon$ for every $s'_i$, and every $i \in N$. These notions are extended to *randomized or mixed strategies* by using the expected playoff.

A *correlated equilibrium* (CE) [9] of a game $(N; (s_i)_{i \in N}; (u_i)_{i \in N})$ is a probability distribution $\pi$ on $S = \times_{i \in N} S_i$ such that for all players $i$ and strategies $s_i, s'_i \in S_i$,

$$\sum_{s_{-i} \in S_{-i}} u_i(s_{-i}, s'_i)\pi(s_{-i}, s_i) \leq \sum_{s_{-i} \in S_{-i}} u_i(s_{-i}, s_i)\pi(s_{-i}, s_i)$$

Less formally, a distribution of strategies is a correlated equilibrium (CE), if after a player's part of the strategy profile has been announced, she prefers to play it instead of something else, assuming that the other players also play their part of the strategy profile.

A *coarse correlated equilibrium* (CCE)[106] of a game $(N; (s_i)_{i \in N}; (u_i)_{i \in N})$ is a probability distribution $\pi$ on $S = \times_{i \in N} S_i$ such that for all players $i$ and strategies $s'_i \in S_i$,

$$\sum_{s_{-i} \in S_{-i}} u_i(s'_i, s_{-i}) \pi_i(s_{-i}) \leq \sum_{s \in S} u_i(s) \pi(s),$$

where $\pi_i(s_{-i}) = \sum_{s'_i \in S_i} \pi(s'_i, s_i)$ is the marginal probability that the strategy tuple $s_{-i} \in S_{-i} = \Pi_{j \neq i} S_j$ will be played.

The notion of the coarse correlated equilibrium is a straightforward generalization of Nash equilibrium, where we allow for correlated behavior amongst the players. It is straightforward to show that the coarse correlated equilibria form a convex set that contains the set of correlated equilibria.

## 2.2 Online Learning and Regret Minimization

An online learning algorithm is an online algorithm for choosing a sequence of elements of some fixed set of *actions*, in response to an observed sequence of cost functions mapping actions to real numbers. The $t$-th action chosen by the algorithm may depend on the first $t-1$ observations but not on any later observations; thus the algorithm must choose an action at time $t$ without knowing the payoffs of any actions at that time. More formally,

**Definition:** An online sequential problem consists of a feasible set $F \in \mathbb{R}^m$, and an infinite sequence of functions $\{f^1, f^2 \ldots, \}$, where $f^t : \mathbb{R}^m \to \mathbb{R}$.

At each time step $t$, an online algorithm selects a vector $x^t \in \mathbb{R}^m$. After the vector is selected, the algorithm receives $f^t$, and collects a payoff of $f^t(x^t)$. All decisions must be made *online*, in the sense that an algorithm does not know $f^t$ before selecting $x^t$, i.e., at each time $t$, a (possibly randomized) algorithm can be thought of as a mapping from a history of functions up to time $t$, $f^1, \ldots, f^{t-1}$, to the set $F$.

Given an algorithm $\mathcal{A}$ and an online sequential problem $(F, \{f^1, f^2, \ldots\})$, if $\{x^1, x^2, \ldots\}$ are the vectors selected by $\mathcal{A}$, then the payoff of $\mathcal{A}$ until time $T$ is $\sum_{t=1}^{T} f^t(x^t)$. The payoff of a static feasible vector $x \in F$, is $\sum_{t=1}^{T} f^t(x)$. Regret compares the performance of an algorithm with the best static action in hindsight:

**Definition:** The *regret*[1] of algorithm $\mathcal{A}$, at time $T$ is defined as

$$\mathcal{R}(T) = \max_{x \in F} \sum_{t=1}^{T} f^t(x) - \sum_{t=1}^{T} f^t(x^t).$$

An algorithm is said to have no regret or that it is *Hannan consistent* [48], if for every online sequential problem, its regret at time $T$ is $o(T)$.

## 2.3 Multiplicative Updates Learning Algorithm

In our work, we will be focusing on specific families on regret-minimizing algorithms. Of particular importance will be the weighted majority algorithm introduced by Littlestone and Warmuth [63] as well as the Hedge algorithm of Freund and Schapire [43]. These algorithms maintain a vector of $n$ probabilities for the $n$ available actions, and at each round they choose an action according to this

[1]Sometimes referred to as external-regret

probability. (Initially, the probabilities are all equal.) At the end of each round, they update the weights multiplicatively, favoring actions that exhibit low cost.

To define the multiplicative updates learning algorithm, let $\varepsilon$ be a small positive number. We will be using costs (i.e. negative payoffs) in our definitions, since this will agree with most of our applications. Let $c_i[1:t] = \sum_{r=1}^{t} c_i[r]$ be the cumulative cost of action $i$ and $Z(t) = \sum_{i \in S_i} \exp(-\varepsilon c_i[1:t-1])$. In period $t$, each player samples action $i$ with probability

$$P(i,t) = \frac{\exp(-\varepsilon c_i[1:t-1])}{Z(t)}, \tag{2.1}$$

i.e., to obtain $P(i,t)$ from $P(i,t-1)$ we multiply it by $\exp(-\varepsilon c_i[t-1])$ and then renormalize all probabilities so that they sum to 1.

An algorithm is said to exhibit $\varepsilon$-regret, if its average performance may be at most $\varepsilon$ worse that than of the best fixed action in hindsight, as time goes to infinity. The multiplicative updates algorithms exhibit $\varepsilon$-regret, but can still in practice offer no-regret guarantees by iteratively reducing their $\varepsilon$ performance gap to zero.

## 2.4 Replicator Dynamics and Evolutionary Game Theory

Evolutionary game theory (EGT) constitutes one of the most well-studied settings of dynamic agent behavior. EGT is the brainchild of John Maynard Smith and G.R. Price [67],[97], who combined ideas from game theory and evolutionary biology in attempt to explain behavioral patterns observed in animal populations.

The prototypical model of dynamic behavior in this context is the replicator dynamics. The replicator dynamics constitute a rather simplified approximation of biological evolution. According to it, the likelihood that a specific behavior (gene) will be replicated grows in proportion to how well it performs relative to the mean performance of the currently adopted behaviors (genes) in the population. In the usual case of a homogeneous population, if we denote by $x_i$ the fraction of the population using strategy $i$ and $f_i(x)$ the payoff of using strategy $i$ given the state $x$ of the population, the replicator dynamics has as follows:

$$\dot{x}_i = x_i[f_i(x) - \bar{f}(x)], \quad \bar{f}(x) = \sum_{i=1}^{n} x_i f_i(x)$$

Players are assumed to be matched randomly inside the population, and their utilities are captured via a symmetric two-player game $G$ with actions $(x_1, x_2, \ldots, x_n)$ and utility function $u(x_i, x_j)$. Hence the payoff of strategy $i$, $f_i(x) = \sum_{j=1}^{n} x_j u(x_i, x_j)$ is merely the expected utility of the player given a random matching to another player in the population[2].

Given the underline game $G$, the replicator equation defines a system of differential equations. The main goal of evolutionary game theory is to characterize the (asymptotically or Lyapunov) stable fixed points of such systems. A Nash equilibrium is a fixed point of the replicator dynamic, and every stable fixed point is a Nash equilibrium. The challenge however lies on deciding whether the set of stable fixed points, defines a refinement of Nash equilibria, that allows for more accurate predictions that the set of Nash Equilibria.

*Evolutionarily stable strategies (ESS)* is a static solution concept that sheds some light into this issues. Informally, ESS requires that a strategy should be robust to small mutations of the population structure. Specifically, if $x$ is the

[2]The population is assumed to be infinite, so the distribution appears identical to all players regardless of their strategies

current population structure and is invaded by a mutant population with strategy $x'$, then ESS requires that for all sufficiently small $\varepsilon$:

$$u(x,(1-\varepsilon)x+\varepsilon x') > u(x',(1-\varepsilon)x+\varepsilon x')$$

This definition implies that any invading mutation is performing strictly worse than the current strategy and hence it will quickly wither out. By linearity of expectation, this definition can be shown to be equivalently to the following:

$$u(x,x) \geq u(x',x) \quad \text{or}$$

$$u(x,x) = u(x',x) \quad \text{and} \quad u(x,x') > u(x',x')$$

The last definition implies that ESS is a refinement of Nash equilibrium. In terms of our goal of characterizing stable states of the replicator dynamics ESS are particularly useful since all ESS are asymptotically stable fixed points of the replicator dynamics. Unfortunately, not all asymptotically stable fixed points are ESS.

A weaker notion of evolutionary stability is the notion of *neutrally stable strategies (NSS)*[97]. NSS requires that the invading mutants should not outperform the incumbent strategy, but unlike the ESS allows for the mutants to earn exactly as much as it. More formally, a strategy is a NSS if for all sufficiently small $\varepsilon$:

$$u(x,(1-\varepsilon)x+\varepsilon x') \geq u(x',(1-\varepsilon)x+\varepsilon x')$$

Equivalently, NSS can be defined via the following properties:

$$u(x,x) \geq u(x',x) \quad \text{or}$$

$$u(x,x) = u(x',x) \quad \text{and} \quad u(x,x') \geq u(x',x')$$

In a similar fashion to the ESS, all NSS are Lyapunov stable fixed points of the replicator dynamics, but not all Lyapunov stable fixed points are NSS. Therefore, NSS (ESS) are sufficient conditions for (asymptotic) stability. However, in order to characterize worst-case stable fixed points we need necessary conditions.

The most general positive in this direction, gives a sufficient and necessary condition for asymptotic stability for the *heterogeneous (multi-population) replicator dynamics*. In this case, each individual in role $i$ can be divided in $n_i$ subpopulations one for each pure strategy $s_i \in S_i$ available to the agents in role $i$. The population state is a vector $(x_1, x_2, \ldots, x_n)$ with each $x_i$ representing the state of the subpopulation corresponding to role $i$. Again, each population is assumed to be infinite. Finally, $n$ players one from each population are picked uniformly at random and interact via an underlying game $G$. This is a game of n players, each with strategies $S_i$, and with utility functions $u_i : \times_i S_i \to \mathcal{R}$. Finally, we extend the utility functions over mixed strategies in the usual multi-linear fashion. The replicator equation in this case, has as follows:

$$\dot{x_i^{s_i}} = x_i^{s_i}[u_i(s_i, x_{-i}) - u_i(x)]$$

In the case of multi-population replicator dynamics a strategy profile is asymptotically stable if and only if it is a strict Nash [55], [103]. Unfortunately, such general results do not carry over to (Lyapunov) stable states.

Although the replicator dynamic was introduced in the context of evolutionary biology, it has gained considerable popularity amongst economists. Part of this success lies on the fact that the analysis of the replicator dynamics in many cases seems to be in agreement with dynamics which have a more concrete economic foundation. In chapter 3, we provide some justification behind such phenomena by showing that the replicator dynamics is the continuous time analogue of the multiplicative updates learning algorithm. The versatility of the multiplicative update algorithm [6] indicates that the replicator dynamic has numerable reasonable economic interpretations.

## 2.5 No-Regret Learning and Coarse Correlated Equilibria

It is well known that the long-run average outcome of repeated play using no-external regret algorithm converges to the set of coarse correlated equilibria[106]. Similarly one can use a somewhat more restrictive notion of no-internal-regret[3] algorithm, whose long-run average outcome of repeated play converges to the set of correlated equilibria.

Here we prove that any coarse correlated equilibrium can arise as the limiting result of regret-minimizing play and, similarly, that any correlated equilibrium can arise as the limiting result of internal-regret-minimizing play.

In [38], Foster and Vohra prove that in the case of calibrated forecasting rules

[3]Integral regret is a refinement of external regret according to which instead of comparing the performance of the algorithm with that of the best action in hindsight, we compare it against the performance of the best algorithm from the following class of learning algorithms. These algorithms take as input the actions of the original learning procedure and consistently map each action to some other. A more detailed discussion of the relationship between no-external regret and no-internal regret algorithms can be found here [17].

in almost every game (excluding a set of games having measure zero in the space of all payoff matrices), every correlated equilibrium can be attained as the limit point of the empirical distribution of play where the players respond myopically to some calibrated forecasting rule. Our results are analogous, but do not have the measure zero exception set.

**Theorem 2.5.1** *Given any coarse correlated equilibrium $C$ of a normal form game with a finite number of players $n$ and finite number of strategies, there exist a set of $n$ no regret processes such that their interplay converges to the correlated equilibrium $C$.*

**Proof:** Suppose that we are given a coarse correlated equilibrium $C$ of a $n$-player game. We will start with the case when these probabilities are rational. In this case, there exists a natural number $K$, such that all probabilities are multiples of $1/K$. We can create a sequence of outcomes $S$ of length $K$, such that their sequential and cyclical play is statistically equivalent to the given coarse correlated equilibrium $C$. Let's denote the j-th element of this sequence as $< x_1^j, x_2^j, \ldots, x_N^j >$, where $0 \leq j \leq K-1$. Each element of this sequence will act as a recommendation vector for the no regret algorithm.

Given the sequence, above we are ready to define for each of the $N$ players a no regret algorithm, such that their interplay converges to the given coarse correlated equilibrium $C$. The algorithm for the $i$-th player is as follows: at time zero she plays the $i$-th coordinate of the first element in $S$ and as long as the other players' responses up to any point in time $t$ are in unison with $S$, that is for every $t' < t$ and $j \neq i$ the strategy implemented by player $j$ at time $t'$ was $x_j^{t' \bmod K}$, then the $i$-player will follow the recommendation of the $S$ sequence playing $x_i^{t \bmod K}$.

However, as soon as the player recognizes any sort of deviation from $S$ by another player, then the player will just disregard any following recommendations coming from $S$ and will merely follow from that point on a no regret algorithm of her liking.

Given the algorithms we have defined above, it is quite straightforward to verify that indeed their interplay will lead to the given coarse correlated equilibrium $C$.

We need to also prove that all of these algorithms are no regret algorithms. When analyzing any of the algorithms above, we have to take into account two cases a) while the other players strategies are such that no deviation is ever recorded from the recommendation provided by $C$, in which case the definition of coarse correlated equilibrium implies that the players have bounded total regret (only corresponding to the partial sequence they played when $t \mod K$ is not 0). Once a first deviation is witnessed by the player in question, she turns to her no regret algorithm of choice and the no regret property then follows from this algorithm.

The only issue that is left to be addressed is to argue how this proof can be generalized in the case the probabilities that appear in the correlated equilibrium $C$ are not all rational. In the case of games with rational payoffs, which are really the games we are interested in, any such correlated equilibrium can be approximated with arbitrarily high precision by a correlated equilibrium with rational probabilities. If we want to converge to a given coarse correlated equilibrium $C$ with irrational probabilities, that can be achieved as long we can express the probability distribution of $C$ in an efficient manner. In this case, instead of having a sequence $S$ of bounded length whose repeated play converges to the correlated equilibrium, we will have sequences $S_i$ of increasing length. Every player will be able to reproduce them and hence derive her recommendation and check for deviations from the other players as before. One way of achieving this is to define $S_i$'s such that for every $i$ the sequence $S_1, S_2, \ldots, S_i$ when played sequentially are statistically indifferent from the $i$-decimal digit approximation of the correlated equilibrium. A specific ordering on each player's strategies can be agreed upon and the sequences $S_i$ can de uniquely defined as the ordering of those outcomes in a lexicographically manner. The rest of the proof holds as is. $\square$

An analogous statement holds for the correlated equilibria and the no-internal regret algorithms.

**Proposition 2.5.2** *Given any correlated equilibrium $C$ of a normal form game with a finite number of players $n$ and finite number of strategies, there exist a set of $n$ no-internal regret processes such that their interplay converges to the coarse correlated equilibrium $C$.*

CHAPTER 3

# MULTIPLICATIVE UPDATES LEARNING IN CONGESTION GAMES

## 3.1 Introduction

Congestion games have been studied extensively in computer science, often from the standpoint of analyzing the price of anarchy: the ratio of solution quality achieved by the worst-case Nash equilibrium versus the optimal solution. Koutsoupias and Papadimitriou [62] introduced the price of anarchy in the context of a load-balancing game studying the makespan objective function. Congestion games both with makespan and with social welfare objective function are well understood; see the surveys by Vöcking [102] and Roughgarden [84]. Analyzing the inefficiency of Nash equilibria provides useful information about the solution quality achieved by selfish players once they reach an equilibrium, but does not provide a model of how selfish players behave, and it says little about whether selfish players will coordinate on an equilibrium, nor which equilibria they are likely to coordinate on if the game has more than one.

Learning has been suggested as a natural model of players' behavior in games. No-regret learning algorithms suggest simple and plausible adaptive procedures where players do not have regrets about their past behavior in some precise sense, which makes them natural candidates to model selfish play. In the theory of learning in games, many have studied the limit of repeated play when all players use such no-regret learning strategies. The resulting equilibrium concepts (variants of correlated equilibrium) typically have worst-case equilibria that fall short of the solution quality achieved by Nash equilibrium. Thus, although researchers studying certain restricted classes of games have proven that the outcome of no-regret learning matches the price-of-anarchy bound [15, 16, 18], there are broad classes of games in which there is a large gap between the predictions arising from analysis of Nash equilibria versus analysis of learning processes (correlated equilibria).

To illustrate this point, consider the load balancing game introduced by Koutsoupias and Papadimitriou [62]. In this game, there are $n$ balls and $n$ bins. Each ball chooses a bin and experiences a cost equal to the number of balls that choose the same bin. If the objective function is the makespan (i.e. maximum load in a bin) then an optimal solution places each ball in a separate bin. These solutions coincide with the pure Nash equilibria of the game. However, there are many other mixed Nash equilibria, including the fully mixed equilibrium in which each ball chooses a bin uniformly at random and the expected makespan is $\Theta(\log n / \log\log n)$. The same game with a non-linear congestion cost results in an arbitrarily high price of anarchy both in the makespan and in the average congestion cost models: the same symmetric fully mixed equilibria are expected to have some bin with congestion $\Theta(\log n / \log\log n)$, which can have arbitrarily high congestion cost. Worse yet, the game can have correlated equilibria that are exponentially worse than the worst mixed Nash equilibrium. In the simple case of linear edge costs the expected makespan is $\Theta(\sqrt{n})$. The ratio is even worse if the congestion cost is decreasing, as in the cost-sharing games where a bin with $x$ players costs $1/x$ to each player. As before, pure equilibria coincide with the social optimal solution, which in this case has a total cost of 1, while the fully mixed Nash equilibrium is expected to use a constant fraction of all bins, and hence to cost $\Theta(n)$.

**Our question.** We focus on understanding the quality of outcomes reached by players using "realistic" learning algorithms. Restricting attention to realistic learning algorithms is consistent with our goal of modeling realistic player behavior, and it is also necessary because within the class of *all* no-regret learning algorithms one can find contrived algorithms whose distribution of play converges to an arbitrary (e.g. worst-case) correlated equilibrium of any game[1], as well as contrived algorithms[2] whose distribution of play converges into the set of Nash equilibria of any game.

**Our results.** We consider a class of learning dynamics, called the aggregate monotonic selection (AMS) dynamics, that extends the multiplicative weights learning algorithm [6, 63] (also known as Hedge [43]) to players whose learning rates may differ and may vary over the strategy space. We show that if players use AMS dynamics to adjust their strategies, then game play converges to a subset of Nash equilibria, which we call *weakly stable equilibria*. These are mixed Nash equilibria $(\sigma_1, \dots, \sigma_n)$ with the additional property that each player $i$ remains indifferent[3] between the strategies in the support of $\sigma_i$ whenever *any other single player* $j$ modifies its mixed strategy to any pure strategy in the support of $\sigma_j$. Pure Nash equilibria are weakly stable, and we show that the converse is true with probability 1 when congestion costs are selected at random independently on each edge (from any monotonically parameterized distribution). Thus, our results imply that in congestion games, learning via AMS dynamics surpasses the Price of Total Anarchy (as defined by Blum et al. [16]) and also the Price of Anarchy

[1] See section 2.5 for a proof. A well-known result of a similar flavor, but using calibrated learning rather than no-regret play, is due to Foster and Vohra [38].

[2] For examples, see the discussion of related work below.

[3] Note that the definition does not require each of the strategies in the support of $\sigma_i$ to remain a best response after player $j$ modifies $\sigma_j$. This modification may cause player $i$ to prefer a strategy lying outside the support of $\sigma_i$.

for mixed Nash equilibria.

Intuitively, players using this learning algorithm are able to steer clear of undesirable mixed Nash equilibria because of symmetry-breaking properties resulting from the inherent randomness in the algorithm. In a load-balancing game, for instance, when one player randomly chooses a machine it causes others to reduce their probability of choosing that machine in the future, and this asymmetry is self-reinforcing. We justify this intuition by showing that the symmetry-breaking is implied by spectral properties of a matrix that is defined at each mixed equilibrium of the game, and that these spectral properties in turn imply weak stability in the sense defined earlier.

In section 3.5 we show that a discrete version of the process with a small amount of added noise at each step follows the solution of the differential equation closely enough to converge to the set of $\nu$-stable equilibria, a generalization of weakly stable equilibria in which the Jacobian is allowed to have eigenvalues whose real part is at most $\nu$, an arbitrarily small positive real number.

**Our techniques.** Our technique is based on analyzing a differential equation expressing a continuum limit of the multiplicative-weights update process, as the multiplicative factor approaches 1 and time is renormalized accordingly. For the case of the Hedge algorithm this differential equation turns out to be identical to the asymmetric replicator dynamic studied in evolutionary game theory. More generally, the differential equation is the extension of the replicator dynamic called aggregate monotonic selection (AMS) dynamics introduced by Samuelson and Zhang [89].

As a first step in analyzing the dynamical system, we need to show that every flow line of the differential equation converges to the set of fixed points. As in prior work on replicator dynamics in potential games (e.g. [3]) we do this by proving that the potential function [74] associated with the congestion game is a *Lyapunov function* for any AMS dynamics; that is, it is non-increasing along flow lines of the differential equation and is strictly decreasing except at fixed points.

The set of fixed points of the differential equation includes all the mixed Nash equilibria (not just the weakly stable or pure ones) as well as some mixed strategy profiles that are not Nash equilibria at all. To see which fixed points arise as limit points of the flow starting from a generic initial condition, we need to distinguish between stable and unstable fixed points. For a fixed point $p$ that is not a Nash equilibrium, it is not hard to argue that $p$ is unstable. For Nash equilibria, we prove that the dynamical-systems' notion of stability — a fixed point where the Jacobian matrix has no complex eigenvalues in the open right half-plane — implies our game-theoretic notion of weakly stable equilibria. To do this, we prove that when the Jacobian matrix is restricted to the subspace of strategies played with positive probability, this submatrix $\mathfrak{J}$ is nilpotent, i.e. its only eigenvalue is 0. It is easy to see that $\mathrm{Tr}(\mathfrak{J}) = 0$. The difficult part here lies in showing that we also have $\mathrm{Tr}(\mathfrak{J}^2) \geq 0$, which uses Steele's [98] non-symmetric version of the Efron-Stein inequality (Lemma 3.3.7). The fact that $\mathrm{Tr}(\mathfrak{J}) = 0$ and $\mathrm{Tr}(\mathfrak{J}^2) \geq 0$, together with the absence of complex eigenvalues in the right half-plane, implies that all eigenvalues are 0. This in turn entails a linear relation on "two-player marginal cost terms" that implies our game-theoretic notion of weak stability.

Clearly all pure equilibria are weakly stable. To show the opposite is true with probability 1 when congestion costs are selected at random independently on each edge, we in fact prove a stronger statement — the existence of a non-pure weakly stable equilibrium implies the vanishing of a non-zero polynomial function of the edge costs — using techniques from algebraic geometry. To illustrate the idea, consider the special case of load balancing games with monotonic cost functions, where weakly stable Nash equilibria are "almost pure" in the sense that each machine has at most one randomizing player using it. (For example, in a load-balancing game with one player and 2 identical machines, any mixed strategy of the one player is stable.) When congestion costs are selected at random, such almost-pure mixed Nash equilibria cannot exist: the probability that two machines have the same cost is 0. To extend this reasoning to general congestion games we need to use more sophisticated techniques, as weakly stable Nash equilibria need not be almost-pure. However, a weakly stable mixed Nash equilibrium must satisfy many polynomial constraints (Nash constraints, insensitivity to one player's change). Using an algebraic-geometric version of Sard's Theorem, we show that congestion costs that have stable mixed Nash equilibria satisfy a nontrivial polynomial equation. In the case of load-balancing games this is a linear relation (two machines having equal cost), but in general it will be a higher-degree polynomial.

**Related work.** No-regret learning algorithms have long been studied in the context of adaptive game playing. There are a number of simple and natural families of no-regret learning algorithms such as the regret matching of Hart and Mas-Colell [51] and the multiplicative weights or Hedge algorithm, introduced by Freund and Schapire [43], which generalizes the well-known weighted majority algorithm of Littlestone and Warmuth [63]. In general games these algorithms converge into the set of coarse correlated equilibria, but not necessarily into the set of Nash equilibria as we prove here for congestion games and the Hedge algorithm. Our decision to analyze the Hedge algorithm in this work was motivated by the algorithm's ubiquitousness in learning theory and other areas of theoretical computer science [6]; the issue of whether similar results can be obtained for other algorithms such as regret matching [51] is an interesting open question.

There are a number of other learning-like processes, such as fictitious play [73], calibrated forecasting [40, 37], regret testing [41, 45], and others (see [52]) whose play is known to converge to Nash equilibria in some games. Some of these results use simple stochastic processes that are tantamount to stochastic search for a pure Nash equilibrium (e.g. [52]) but are not regret-minimizing, while others use complicated adaptive procedures satisfying calibration properties that are closely related to the no-regret property (e.g. [37]). Unlike some of these works, our goal is not to discover uncoupled adaptive procedures for finding a Nash equilibrium but specifically to analyze the behavior of a particular simple, realistic, and well-known adaptive procedure.

The dynamics of repeated play in congestion games has been studied in non-learning-theoretic models such as sink equilibria and selfish rerouting [30, 46, 71]. For *nonatomic* congestion games (i.e., games with infinitesimal players) Fischer, Räcke and Vöcking [34, 32] and Blum, Even-Dar, and Ligett [15] considered learning dynamics in the setting of multicommodity flow (or more generally congestion games) and showed that the dynamics converges to a Wardrop equilibrium. Fischer, Räcke and Vöcking [34, 32] consider a replication-exploration protocol, while Blum, Even-Dar, and Ligett [15] show that in this setting, if each player uses any no-regret strategy the behavior will approach the Wardrop equilibrium. The setting of atomic congestion games studied here is much more intricate because the game typically has many Nash equilibria forming a disconnected set with many components, and we need to distinguish between the stable and unstable ones.

More recently, Blum, Hajiaghayi, Ligett, and Roth [16] defined the price of total anarchy and showed that in a number of games, the known bounds on the price of anarchy extend also to the price of total anarchy: the worst case bound on the quality of coarse correlated equilibria to the optimum outcome is already achieved over Nash equilibria. Roughgarden [85] extends this to a wider class of games. Our results complement these by showing that if one assumes players use a specific, standard no-regret algorithm (namely, Hedge, or AMS) rather than an arbitrarily bad set of no-regret algorithms, one obtains much stronger guarantees about the distribution of play in atomic congestion games: it converges to a weakly stable Nash equilibrium.

The replicator dynamic and other differential equations are studied in evolutionary game theory [97, 92], which also considers associated notions of stability such as evolutionarily stable states (ESS) and neutrally stable states (NSS). The book of Fudenberg and Levine [44] provides an excellent survey on these topics.[4] For each of the three main steps in our analysis, we identify here the most closely related work in evolutionary game theory. **In congestion games, the replicator dynamic converges to its fixed point set.** Proofs of this theorem and generalizations, using the game's potential function as a Lyapunov function, have appeared in many prior works, e.g. [3, 55, 73, 91]. The short proof that we present here conveniently provides a quantitative bound on the rate of decrease of the potential. **A weakly stable fixed point of the replicator dynamic is a weakly stable equilibrium.** Our notion of weakly stable equilibrium is similar to, but weaker than, the notion of neutrally stable states. It is known that neutrally stable states of a game are Lyapunov stable points of the replicator dynamic (hence weakly stable), but our line of attack requires the converse, which is not true in general [103]. Here, we are able to deduce this converse by introducing a weaker

[4]For more general theory of differential equations see the book by Perko [78].

game-theoretic notion of stability. **In almost every game, the weakly stable equilibria coincide with the pure Nash equilibria.** The most closely-related result in evolutionary game theory is Ritzberger and Weibull's theorem [81] that every *asymptotically stable* fixed point set of an AMS dynamic is not contained in the relative interior of any face of the mixed strategy polytope, and hence contains a pure equilibrium. Although their result applies to general games and not just congestion games, it is weaker than ours in two key respects: it assumes a stronger stability property (that need not hold at the weakly stable fixed points considered in our analysis) and derives a weaker conclusion (the asymptotically stable set contains a pure Nash equilibrium but may also contain other equilibria).

## 3.2 Model and Preliminaries

**Congestion games.** *Congestion games* [83] are non-cooperative games in which the utility of each player depends only on the player's strategy and the number of other players that either choose the same strategy, or some strategy that "overlaps" with it. Formally, a *congestion game* is defined by the tuple $(N; E; (\mathcal{S}_i)_{i\in N}; (c_e)_{e\in E})$ where $N$ is the set of *players*, $E$ is a set of *facilities* (also known as *edges* or *bins*), and each player $i$ has a set $\mathcal{S}_i$ of subsets of $E$ ($\mathcal{S}_i \subseteq 2^E$). Each pure strategy $S_i \in \mathcal{S}_i$ is a set of edges (a *path*), and $c_e$ is a cost (negative utility) function associated with facility $e$. Given a pure strategy profile $S = (S_1, S_2, \ldots, S_N)$, the cost of player $i$ is given by $c_i(S) = \sum_{e\in S_i} c_e(k_e(S))$, where $k_e(S)$ is the number of players using $e$ in $S$. Congestion games admit a *potential function* $\Phi(S) = \sum_{e\in E}\sum_{j=1}^{k_e(S)} c_e(j)$, which captures each player's incentive to change his strategy [83]. Specifically, given pure strategy profile $S = (S_1, S_2, \ldots, S_N)$, and strategy $S_i'$ of player $i$: $c_i(S_i', S_{-i}) - c_i(S) = \Phi(S_i', S_{-i}) - \Phi(S)$. The set of pure Nash equilibria can be found by simply locating the local optima of $\Phi(S)$.

**Multiplicative updates algorithm (Hedge) and the AMS dynamic.** An online learning algorithm is an online algorithm for choosing a sequence of elements of some fixed set of *actions*, in response to an observed sequence of cost functions mapping actions to real numbers. The $t$-th action chosen by the algorithm may depend on the first $t-1$ observations but not on any later observations; thus the algorithm must choose an action at time $t$ without knowing the costs of any actions at that time. The regret of an online learning algorithm is defined as the maximum over all input instances of the expected difference in payoff between the algorithm's actions and the best action. If this difference grows sublinearly with time, we say it is a *no-regret learning algorithm* or that it is *Hannan consistent* [48].

The family of regret minimizing algorithms that we study in this paper are called *aggregate monotonic selection (AMS) dynamics* [81, 89], and they constitute a generalization of the weighted majority algorithm introduced by Littlestone and Warmuth [63] and the Hedge algorithm of Freund and Schapire [43]. These algorithms maintain a vector of $n$ probabilities for the $n$ available actions, and at each round they choose an action according to this probability. (Initially, the probabilities are all equal.) At the end of each round, AMS dynamics updates the weights multiplicatively, favoring actions that exhibit low cost. The update is governed by a global parameter $\varepsilon > 0$, and a state-dependent parameter $\beta = \beta(i, p)$ which is determined by the player, $i$, and by the profile of mixed strategies, $p$, representing the current state of *every* player's learning algorithm. We can interpret the parameter $\varepsilon\beta(i, p)$ as the learning rate of player $i$. If the cost of action $a$ at time $t$ is $c^a(t)$ then the weights are updated by multiplying with $(1-\varepsilon\beta)^{c^a(t)}$ and then normalizing to keep the probability distribution:

$$p_a^{t+1} = \frac{p_a^t(1-\varepsilon\beta)^{c^a(t)}}{\sum_j p_j^t(1-\varepsilon\beta)^{c^j(t)}} \tag{3.1}$$

We will assume throughout that $\forall i, p \ 0 < \beta(i,p) \leq 1$. For example, the algorithm Hedge($\varepsilon$) is obtained by setting $\beta(i,p) = 1$ for all $i, p$. Hedge($\epsilon$) is not a no-regret algorithm, but it is an $\varepsilon$-regret algorithm, and well known tricks (halving $\varepsilon$ as the algorithm proceeds) can be employed so that it becomes Hannan consistent.

## 3.3 Analysis of the Continuous-Time Process

In this section we define and analyze the continuous version of the Hedge (multiplicative weights) algorithm. We start with understanding the algorithm's update rule for probabilities, and we derive the limit as $\varepsilon \to 0$, a first-order differential equation (ODE) known as the replicator dynamic.

**Setting up the differential equation.** The continuous-time process we want to analyze in this section arises as the limit of the update rule (3.1) as $\varepsilon \to 0$. The cost of a strategy $S_i$ for a player $i$ is a random variable $C^i(S_i) = \sum_{e\in S_i} c_e(1+K_i(e))$, where $K_i(e)$ denotes $|\{j : j \neq i, e \in S_j\}|$, the number of players other than $i$ that use edge $e$. Taking the derivative of the above update with respect to $\varepsilon$, and substituting $\varepsilon = 0$, we get a differential equation using the random variables $C^i(S_i)$. Taking expectation, and using the notation $c^i(e) = \mathbb{E}(c_e(1+K_i(e))$ for the expected cost of edge $e$ for player $i$, we get that the expected cost is $c^i(R) = \mathbb{E}(C^i(R)) = \sum_{e\in R} c^i(e)$ and the expected update in the probabilities is the following differential equation, $\dot{p} = \xi(P)$, where:

$$\xi_{iR} \quad = \quad \beta(i,p)p_{iR}\left(\sum_{\bar{R}} p_{i\bar{R}}(c^i(\bar{R}) - c^i(R))\right). \tag{3.2}$$

**Fixed points and convergence to fixed points.** It is not hard to establish what are the fixed points of the differential equation. They are all the probability distributions where all players are mixing between equal (but not necessarily smallest) cost options.

**Theorem 3.3.1** *Probability distributions $p_i$ for $i \in N$ form a fixed point of the ODE (3.2) if and only if for all players $i$, all of the strategies $R \in \mathcal{S}_i$ with $p_{iR} > 0$ have the same expected cost $c^i(R)$ for player $i$.*

Before we study fixed points further, we want to establish that the solutions of the differential equation converge to fixed points (e.g., do not cycle). To do this we will consider the standard potential function $\Phi$ of the congestion game. This is analogous to the proof of Monderer and Shapley [73] who show that in games when players have identical interest (and also in potential games that are strategically equivalent to such games for fictitious play and also for the AMS dynamic), repeated play with fictitious play converges to the set of Nash equilibria. Recall that the standard potential function of a congestion game is defined as $\sum_e \sum_{k=1}^{K_e} c_e(k)$, where we use the notation $K_e = |\{j : e \in S_j\}|$ for a pure strategy profile $S = (S_1, \ldots, S_n)$ We will use the expected value of this function as our potential function. Using linearity of expectation, this expectation is

$$\Psi \;=\; \mathbb{E}(\Phi) = \sum_e \sum_{k=1}^{\infty} c_e(k) \Pr(K_e \geq k)$$

**Theorem 3.3.2** *The time derivative $\dot{\Psi}$ of the potential function is 0 at fixed points, and negative at all other points. In fact, $-\dot{\Psi} \geq \frac{||\xi||_1^2}{2\beta_{1..i}(p)}$, where $\beta_{1..i}(p)$ denotes $\sum_{i=1}^{n} \beta(i,p)$.*

**Proof:** Let $\Phi_{-i}$ denote the potential function of the game without player $i$, i.e.

$\Phi_{-i}(S_{-i}) = \sum_{e \in E} \sum_{j=1}^{k_e(S_{-i})} c_e(j)$, where $k_e(S_{-i})$ denotes the number of players using edge $e$ in strategy profile $S_{-i}$. It is well known that the actual potential function $\Phi$ satisfies $\Phi(S) = \Phi_{-i}(S_{-i}) + C^i(S_i)$ for all strategy profiles $S$. Note that

$$\Psi = \mathbb{E}[\Phi_{-i}(S_{-i})] + \mathbb{E}[C^i(S_i)] = \mathbb{E}[\Phi_{-i}(S_{-i})] + \sum_R p_{iR} c^i(R).$$

The terms $\mathbb{E}[\Phi_{-i}(S_{-i})]$ and $c^i(R)$ don't depend on player $i$'s mixed strategy, so $\partial\Psi/\partial p_{iR} = c^i(R)$. Now,

$$\begin{aligned}
\dot{\Psi} &= \sum_{i,R} \left(\frac{\partial \Psi}{\partial p_{iR}}\right) \dot{p}_{iR} \\
&= \sum_i \beta(i,p) \sum_{R,\bar{R}} p_{iR} p_{i\bar{R}} \left(c^i(R) c^i(\bar{R}) - c^i(R)^2\right) \\
&= -\sum_i \beta(i,p) \sum_{R<\bar{R}} p_{iR} p_{i\bar{R}} \left(c^i(R) - c^i(\bar{R})\right)^2.
\end{aligned}$$

The second line is produced by substituting $\frac{\partial\Psi}{\partial p_{iR}}$ and $\dot{p}_{iR}$ and rearranging the resulting terms. In deriving the last line, we have assumed that the strategy set of player $i$ is totally ordered, and we have paired the terms on the preceding line corresponding to $R, \bar{R}$ and $\bar{R}, R$. Finally, to bound $-\dot{\Psi}$ from below in terms of $\|\xi\|_1$ (which is needed for the discrete-time analysis) we use the Cauchy Schwartz inequality:

$$\begin{aligned}
&(-2\beta_{1..i}(p))\dot{\Psi} = \beta_{1..i}(p) \sum_{i,R,\bar{R}} \beta(i,p) p_{iR} p_{i\bar{R}} (c^i(R) - c^i(\bar{R}))^2 \\
&= \left[\sum_{i,R,\bar{R}} \beta(i,p) p_{iR} p_{i\bar{R}}\right] \left[\sum_{i,R,\bar{R}} \beta(i,p) p_{iR} p_{i\bar{R}} (c^i(R) - c^i(\bar{R}))^2\right] \\
&\geq \left[\sum_{i,R,\bar{R}} \beta(i,p) p_{iR} p_{i\bar{R}} \left|c^i(R) - c^i(\bar{R})\right|\right]^2 \geq \|\xi\|_1^2 \square
\end{aligned}$$

$\square$

**Unstable fixed points and the Jacobian** We will use the notion of stability from dynamical systems. In the neighborhood of a fixed point $p_0$ the ODE can be approximated by $\dot{p} \approx J(p - p_0)$, where $J$ is the matrix of partial derivatives, the Jacobian. A fixed point of a dynamical system is said to be unstable (see in [78]) if the Jacobian matrix has an eigenvalue with positive real part.

For an ODE represented by a vector field $\xi(x)$ the entry $J_{ij}$ of the Jacobian is the partial derivative of the $i$-th coordinate $\xi_i(x)$ in the direction of $x_j$. Our ODE has coordinates $p_{iR}$ corresponding to a player $i$ and a strategy $R$, and our vector field $\xi$ is defined in (3.2). Observe in (3.2) that $\xi_{iR}$ is the product of $\beta(i,p)$ with a term that vanishes at a fixed point. When we take the partial derivative in any direction, using the product rule we find that the derivative of $\beta(i,p)$ vanishes, as it is multiplied by 0 at a fixed point. Now let us examine the entries of the Jacobian matrix case by case. For directions corresponding to the same player we get

$$\frac{\partial \xi_{iR}}{\partial p_{iR}} = \beta(i,p) \sum_{\bar{R}} p_{i\bar{R}}(c^i(\bar{R}) - c^i(R)); \qquad \frac{\partial \xi_{iR}}{\partial p_{i\bar{R}}} = \beta(i,p) p_{iR}(c^i(\bar{R}) - c^i(R));$$

where the sum in the partial derivative should be for $\bar{R} \neq R$, but can equally well be understood to be for all $\bar{R}$. The second expression is for $\bar{R} \neq R$.

Finally, taking a derivative in the direction $p_{jQ}$ for $j \neq i$ involves understanding how the cost $c^i(R)$ depends on the probability $p_{jQ}$. We get that $mc_e^{ij} := \mathbb{E}(c_e(2 + K_{ij}(e))) - \mathbb{E}(c_e(1 + K_{ij}(e)))$ is the coefficient of $q$ in the cost $c^i(e)$. Using the notation $mc^{ij}(A) = \sum_{e \in A} mc_e^{ij}$ we get that the marginal cost for player $i$ for increasing probability $p_{jQ}$ is

$$\frac{\partial \xi_{iR}}{\partial p_{jQ}} = \beta(i,p) p_{iR} \sum_{\bar{R}} p_{i\bar{R}}(mc^{ij}(\bar{R} \cap Q) - mc^{ij}(R \cap Q))$$

Note that the marginal cost depends on the probability distributions of players other than $i$ and $j$, but does not depend on those two players. In particular

$mc_e^{ij} = mc_e^{ji}$ for every edge $e$.

**Theorem 3.3.3** *The Jacobian matrix is expressed by the equations above.*

**Lemma 3.3.4** *A fixed point $p$ that is stable corresponds to a Nash equilibrium.*

**Proof:** Consider a fixed point $p$ of the ODE that is not a Nash equilibrium. If a fixed point is not a Nash equilibrium, than there is a player $i$ and a strategy $R$ with $p_{iR} = 0$ that has $\beta(i,p)\sum_{\bar{R}} p_{i\bar{R}}(c^i(\bar{R}) - c^i(R)) = \lambda > 0$. The unit vector $w^{iR}$ with a 1 in the $(i,R)$ coordinate is a left eigenvector of $J$ with $w^{iR}J = \lambda w^{iR}$, hence $J$ has a positive eigenvalue. □

At a fixed point $p$ that is a Nash equilibrium, let $\mathfrak{J}$ be the submatrix of the Jacobian restricted to the subset of strategies played with positive probability ($p_{iR} > 0$).

**Lemma 3.3.5** *For any right eigenvector $w$ of the matrix $\mathfrak{J}$ that satisfies $\mathfrak{J}w = \lambda w$, the vector $w^\circ$ extending $w$ with 0 values to the remaining coordinates is an eigenvector of the full Jacobian, $J$, with eigenvalue $\lambda$.*

First note that the trace of $\mathfrak{J}$ is 0, which follows directly from the definition of fixed point, and definition of $\mathfrak{J}$, as all diagonal entries are 0. To help establish that the submatrix $\mathfrak{J}$ has an eigenvalue with positive real part we will prove that the trace of $\mathfrak{J}^2$ is nonnegative. This will follow from a matrix inequality of independent interest, that can be derived from Steele's non-symmetric version of the Efron-Stein inequality [98].

**Theorem 3.3.6** *Consider a fixed point of the ODE, and let $\mathfrak{J}$ be the submatrix of the Jacobian defined above. Then* $\mathrm{Tr}(\mathfrak{J}^2) \geq 0$*, and in fact it is equal to*

$$\sum_{i,j} \beta(i,p)\beta(j,p) \sum_{R<\bar{R},Q<\bar{Q}} p_{iR}p_{i\bar{R}}p_{jQ}p_{j\bar{Q}}(M_{i,j}^{R,\bar{R},Q,\bar{Q}})^2,$$

*where* $M_{i,j}^{R,\bar{R},Q,\bar{Q}}$ *is defined to be* $mc^{ij}(R\cap Q) - mc^{ij}(R\cap\bar{Q}) - mc^{ij}(\bar{R}\cap Q) + mc^{ij}(\bar{R}\cap\bar{Q})$.

**Proof of Theorem** 3.3.6 To prove that the trace of $\mathfrak{J}^2$ is nonnegative we use a matrix inequality. For any matrix $A$ we use $A\cdot A$ as matrix obtained by taking the square of each term of $A$. The following inequality compares the square of the expectation of $M$ and the expectation of the square with taking expectation on rows (or columns) squaring the result and then taking expectation. It may be useful to think of this inequality as a variational inequality on the expectation of squares.

**Lemma 3.3.7** *For any matrix $M$ and any probability distributions $p_i$ and $p_j$ on the rows and columns respectively, we have the following inequality*

$$0 \quad \leq \quad (p_i^T M p_j)^2 - p_i^T((Mp_j)\cdot(Mp_j)) - ((p_i^T M)\cdot(p_i^T M))p_j + p_i^T(M\cdot M)p_j.$$

*In fact this value equals* $\sum_{R,\bar{R},Q,\bar{Q}} p_{iR}p_{i\bar{R}}p_{jQ}p_{j\bar{Q}}(m_{QR} - m_{\bar{Q}R} - m_{Q\bar{R}} + m_{\bar{Q}\bar{R}})^2$

To prove this one may simply check that each term of the form $m_{QR}m_{\bar{Q}\bar{R}}$ occurs in the two expression with the same multiplier. Alternatively, the inequality stated in the lemma follows from Steele's non-symmetric version of the Efron-Stein inequality [98].

Theorem 3.3.6 follows by applying this inequality with $m_{RQ} = mc^{ij}(R\cap Q)$ to establish that $\mathrm{Tr}(\mathfrak{J}^2) \geq 0$. $\square$

We are ready to show that if a fixed point is stable, then $\mathfrak{J}$ must have 0 as its only eigenvalue.

**Theorem 3.3.8** *For a stable fixed point $p$, all eigenvalues of the submatrix $\mathfrak{J}$ of the Jacobian corresponding to the coordinates with $p_{iR} > 0$ are zero. Also, for all players $i, j$ and all strategies $R, \bar{R}, Q, \bar{Q}$ played with positive probability by the players $i$ and $j$ respectively, we must have $mc_{ij}(R \cap Q) - mc_{ij}(R \cap \bar{Q}) - mc_{ij}(\bar{R} \cap Q) + mc_{ij}(\bar{R} \cap \bar{Q}) = 0$.*

**Proof:** For any fixed point the sum of the eigenvalues, $\mathrm{Tr}(\mathfrak{J})$, is zero, hence if $\mathfrak{J}$ has no eigenvalues with positive real part then all eigenvalues must be pure imaginary. But in this case $\mathrm{Tr}(\mathfrak{J}^2)$ is nonpositive, as it is the sum of squares of pure imaginary numbers. We know that $\mathrm{Tr}(\mathfrak{J}^2) \geq 0$ and hence it must equal zero. Hence all eigenvalues of $\mathfrak{J}$ must equal zero, as claimed. $\square$

Using the condition derived in the above theorem for stable fixed points of the ODE, we can connect the notion of stable for the dynamical system to our game theoretic notion of weakly stable, in the sense defined in the introduction, that each player $i$ remains indifferent between the strategies in the support of $\sigma_i$ whenever *any other single player $j$* modifies its mixed strategy to any pure strategy in the support of $\sigma_j$.

**Theorem 3.3.9** *If a Nash equilibrium is stable for the dynamical system then it is a weakly stable Nash equilibrium.*

**Proof:** We have proved that if a Nash equilibrium is stable for the dynamical system, then for all players $i, j$ and all strategies $R, \bar{R}, Q, \bar{Q}$ played with positive probability by the players $i$ and $j$, respectively, $mc_{ij}(R \cap Q) - mc_{ij}(\bar{R} \cap Q) = mc_{ij}(R \cap \bar{Q}) - mc_{ij}(\bar{R} \cap \bar{Q})$. Using $j$'s mixed strategy $\sigma_j$ to take a weighted average over all $\bar{Q}$, we get the claim that if $i$ is indifferent between two of its strategies $R, R'$ when $j$ randomizes, he remains indifferent when $j$ plays purely strategy $Q$. $\square$

We have seen the solution of the ODE converges, and that fixed points of the dynamic system that are not weakly stable Nash equilibria are unstable for the dynamic system. Using the theory of differential equations [78], we can conclude that starting from a generic initial condition, the ODE converges to weakly stable Nash equilibria.

**Theorem 3.3.10** *From all but a measure 0 set of starting points, the solution of the ODE* (3.2) *converges to weakly stable Nash equilibria.*

## 3.4 Weakly Stable Equilibria

If one fixes the set of players and facilities, and the strategy sets of each player of a congestion game — which we collectively denote as the game's "combinatorial structure" — the game itself is determined by the vector of edge costs $\vec{c}$, i.e., the vector whose components are the numbers $c_e(k)$ for every edge $e$ and every possible load value $k$ on that edge. One can thus identify the set of congestion games having a fixed combinatorial structure with the vector space $\mathbb{R}^N$ where $N$ is equal to the number of pairs $(e, k)$ for which $c_e(k)$ is defined. If one imposes other constraints such as non-negativity and monotonicity on the edge costs, then the set of games is identified with a convex subset of $\mathbb{R}^N$ rather than $\mathbb{R}^N$ itself.

Our goal in this section is to prove the following theorem.

**Theorem 3.4.1** *For almost every congestion game, every weakly stable equilibrium is a pure Nash equilibrium. In other words, the set of congestion games having non-pure weakly stable equilibria is a measure-zero subset of* $\mathbb{R}^N$.

In fact, we will prove the following stronger version of Theorem 3.4.1.

**Theorem 3.4.2** *There is a non-zero multivariate polynomial* $W$*, defined on* $\mathbb{R}^N$*, such that for every game with a non-pure weakly stable equilibrium, its edge costs satisfy the equation* $W(\vec{c}) = 0$.

This strengthening implies, for example, that among all the congestion games with a fixed combinatorial structure and with cost functions taking integer values between 0 and $B$, the fraction of such games having a non-pure weakly stable equilibrium tends to zero as $B \to \infty$.

Define an equilibrium to be *fully mixed* if it satisfies $p_{iR} > 0$ for every player $i$ and every strategy $R$ in that player's strategy set. Every mixed equilibrium $\vec{p}$ of a game is a fully mixed equilibrium of the subgame obtained by deleting the strategies that satisfy $p_{iR} = 0$. Since there are only finitely many such subgames, we can establish Theorems 3.4.1 and 3.4.2 by proving the corresponding statements about fully mixed weakly stable equilibria. Theorem 3.4.1 then follows because a finite union of measure-zero sets has measure zero, and Theorem 3.4.2 follows because the union of the zero-sets of polynomials $W_1, W_2, \ldots, W_k$ is the zero-set of their product $W_1 W_2 \ldots W_k$.

Let $X$ be the set of pairs $(\vec{p}, \vec{c})$ such that $\vec{p}$ is a fully mixed weakly stable equilibrium of the game with edge costs $\vec{c}$, let $f : X \to \mathbb{R}^N$ be the function that projects such a pair $(\vec{p}, \vec{c})$ to its second component, $\vec{c}$, and let $Y \subseteq \mathbb{R}^N$ be the set $f(X)$, i.e. the set of games having a fully mixed weakly stable equilibrium. To prove that $Y$ has measure zero and is contained in the zero-set of a nontrivial polynomial, we will first prove a "local, linearized version" of the same statement. Lemma 3.4.3 below asserts, roughly[5], that for every point $x \in X$, with tangent space $T_x X$, the projection of $T_x X$ to $\mathbb{R}^N$ has dimension strictly less than $N$. (And thus, the image of $T_x X$ in $\mathbb{R}^N$ has measure zero and is contained in the zero-set of a nontrivial linear function.) Theorems 3.4.1 and 3.4.2 are then obtained using general theorems that allow global conclusions to be deduced from these local criteria. To obtain Theorem 3.4.1 we work in the category of differentiable manifolds and apply Sard's Theorem [70]: the set of critical values of a differentiable function has measure zero. To obtain Theorem 3.4.2 we work in the category of algebraic varieties and apply an "algebraic geometry version" of Sard's Theorem ([93], Lemma II.6.2.2): if $X \xrightarrow{f} Y$ is a regular map of varieties defined over a field of characteristic 0, and $f$ is surjective, then there exists a nonempty open set $V \subseteq X$ such that the differential $d_x f$ is surjective for all $x \in V$.[6] As an aid to the reader unfamiliar with algebraic geometry we present a concise exposition of standard definitions in section 3.6.

**Linearized version of Theorems 3.4.1 and 3.4.2.** Each of the expressions $c^i(R), mc^{ij}(R \cap Q)$ used in Section 3.3 actually refers to a polynomial — in fact, a multilinear polynomial — in the variables $p_{**}$ and $c_*(*)$, because the probability

[5]The actual statement is more complicated because $X$ may have singularities, so the tangent space $T_x X$ may be ill-defined. To deal with this, what we actually show is that $X$ can be partitioned into finitely many nonsingular subsets $X_1, X_2, \ldots, X_k$ — possibly of different dimensions — such that for every $x \in X_i$ $(1 \leq i \leq k)$, the projection of $T_x X_i$ to $\mathbb{R}^N$ has dimension strictly less than $N$.

[6]Actually the Lemma as stated in [93] requires the field to be algebraically closed, and it requires the variety $X$ to be nonsingular. We describe how to work around these technical difficulties in the proof of Theorem 3.4.2.

of any given pure strategy profile being sampled is a multilinear polynomial in the $p_{**}$ variables, and the cost of any edge, in any given pure strategy profile, is one of the variables $c_*(*)$. Let the polynomial equation $A_i^{R,R'} = 0$ express the fact that player $i$ is indifferent between strategies $R$ and $R'$, i.e. $c^i(R) - c^i(R') = 0$. By definition of a weakly stable equilibrium we must have $mc^{ij}(R \cap Q) - mc^{ij}(R' \cap Q) = \sum_{\bar{Q}} p_{j\bar{Q}}(mc^{ij}(R \cap \bar{Q}) - mc^{ij}(R' \cap \bar{Q}))$ for any $Q$, hence we get $M_{i,j}^{R,R',Q,Q'} = mc^{ij}(R \cap Q) - mc^{ij}(R' \cap Q) - mc^{ij}(R \cap Q') + mc^{ij}(R' \cap Q') = 0$ for all $R, R', Q, Q'$. Finally let $P_i = 0$ encode $\sum_{R \in \mathcal{S}_i} p_{iR} = 1$. In earlier sections of this paper, we have seen that all of these equations must hold when $\vec{p}$ is a fully mixed weakly stable equilibrium of the game with edge costs $\vec{c}$. In other words, if $I$ denotes the polynomial ideal generated by $\{A_i^{R,R'}\} \cup \{M_{i,j}^{R,R',Q,Q'}\} \cup \{P_i\}$, then the set of fully mixed weakly stable equilibria, $X$, is contained in the algebraic variety $V(I)$ defined by the vanishing of all the polynomials in $I$.

**Lemma 3.4.3** *The ideal $I$ contains a polynomial $F \in \mathbb{R}[\vec{p}, \vec{c}]$ that satisfies:*

1. *$\partial F / \partial p_{iR} \in I$ for all variables $p_{iR}$.*
2. *$\partial F / \partial c_e(k) \notin I$ for at least one variable $c_e(k)$. In fact, there exists an edge $e$ such that the sum of the partial derivatives in directions $c_e(k)$, for $k = 1, \ldots, n$, is in $1 + I$.*

**Proof:** Fix any player $i$, and fix any two strategies $R, R'$ for that player. For all players $j \neq i$ fix a strategy $Q_j^0 \in \mathcal{S}_j$. Consider the polynomial

$$F = A_i^{R,R'} + \sum_{j \neq i} \left[ mc^{ij}(R' \cap Q_j^0) - mc^{ij}(R \cap Q_j^0) \right] P_j.$$

We will show that $\partial F / \partial p_{jQ} = M_{i,j}^{R,R',Q,Q_j^0}$ if $j \neq i$ and $Q \neq Q_j^0$, and otherwise $\partial F / \partial p_{jQ} = 0$. This confirms property (1).

As before, $c^i(e), c^i(R)$ denote the expected cost for $i$ for using $e, R$ respectively, whereas $K_i(e)$ is a random variable expressing the number of players other than $i$ using $e$. Similarly, $K_{ij}(e)$ is a random variable expressing the number of player other than $i, j$ using $e$. We can express $c^i(e), c^i(R)$ as follows:

$$
\begin{aligned}
c^i(e) &= \mathcal{E}[c_e(1 + K_i(e))] \\
&= \mathcal{E}[p_{je}c_e(2 + K_{ij}(e)) + (1 - p_{je})c_e(1 + K_{ij}(e))] \\
&= p_{je}\mathcal{E}[c_e(2 + K_{ij}(e)) - c_e(1 + K_{ij}(e))] + \mathcal{E}[c_e(1 + K_{ij}(e))] \\
&= p_{je}mc^{ij}(e) + \mathcal{E}[c_e(1 + K_{ij}(e))]
\end{aligned}
$$

where as before $mc^{ij}(e) = \mathcal{E}[c_e(2 + K_{ij}(e)) - c_e(1 + K_{ij}(e))]$.

$$
\begin{aligned}
c^i(R) &= \sum_{e \in R}[p_{je}mc^{ij}(e) + \mathcal{E}[c_e(1 + K_{ij}(e))]] \\
&= \sum_{e \in R}\sum_{Q \ni e} p_{jQ}mc^{ij}(e) + \sum_{e \in R}\mathcal{E}[c_e(1 + K_{ij}(e))] \\
&= \sum_{Q \in \mathcal{S}_j} p_{jQ}mc^{ij}(R \cap Q) + \sum_{e \in R}\mathcal{E}[c_e(1 + K_{ij}(e))]
\end{aligned}
$$

The Nash equilibrium condition $A_i^{R,R'}$ for player $i$ and strategies $R, R'$ can be expressed as:

$$
\begin{aligned}
A_i^{R,R'} &= \sum_{Q \in \mathcal{S}_j} p_{jQ}[mc^{ij}(R \cap Q) - mc^{ij}(R' \cap Q)] + \sum_{e \in R}\mathcal{E}[c_e(1 + K_{ij}(e))] - \\
&- \sum_{e \in R'}\mathcal{E}[c_e(1 + K_{ij}(e))]
\end{aligned}
$$

Finally, by taking partial derivatives we get:

$$\frac{\partial A_i^{R,R'}}{\partial p_{jQ}} = mc^{ij}(R \cap Q) - mc^{ij}(R' \cap Q) \tag{3.3}$$

The equation $P_i = 0$ can be written as $\sum_{Q \in \mathcal{S}_j} p_{iR} - 1 = 0$. By taking partial derivatives we get:

$$\frac{\partial P_i}{\partial p_{jQ}} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

Recall that $F = A_i^{R,R'} - \sum_{j \neq i} \left[ mc^{ij}(R' \cap Q_j^0) - mc^{ij}(R \cap Q_j^0) \right] P_j$. The product rule for partial derivatives implies that if $A, B$ are two polynomials and $A \in I$ then

$$\frac{\partial(AB)}{\partial p_{jQ}} = \frac{\partial A}{\partial p_{jQ}} B + A \left( \frac{\partial B}{\partial p_{jQ}} \right) \equiv \left( \frac{\partial A}{\partial p_{jQ}} \right) B \qquad \pmod{I}.$$

We may apply this rule (using the fact that $P_\ell \in I$ for every player $\ell$) to compute the partial derivative of $F$ with respect to $p_{jQ}$ modulo $I$:

$$\begin{aligned} \frac{\partial F}{\partial p_{jQ}} &\equiv \frac{\partial A}{\partial p_{jQ}} - \sum_{\ell \neq i} \left[ mc^{ij}(R' \cap Q_j^0) - mc^{ij}(R \cap Q_j^0) \right] \frac{\partial P_\ell}{\partial p_{jQ}} \qquad \pmod{I} \\ &= mc^{ij}(R \cap Q) - mc^{ij}(R' \cap Q) - mc^{ij}(R' \cap Q_j^0) + mc^{ij}(R \cap Q_j^0) \\ &= M_{ij}^{R,R',Q,Q_j^0} \equiv 0 \qquad \pmod{I}. \end{aligned}$$

Property (2) follows from the formula

$$\begin{aligned} \forall e \in R \setminus R' \quad \sum_{k=1}^{n} \frac{\partial F}{\partial c_e(k)} &= \prod_{j \neq i} (P_j + 1) \; + \\ + \sum_{j \neq i} &\left[ \sum_{k=1}^{n} \left( \frac{\partial mc^{ij}(R \cap Q_j^0)}{\partial c_e(k)} - \frac{\partial mc^{ij}(R' \cap Q_j^0)}{\partial c_e(k)} \right) \right] P_j \end{aligned} \tag{3.5}$$

This equation follows from a direct calculation.

$$\begin{aligned} \sum_{k=1}^{n} \frac{\partial F}{\partial c_e(k)} &= \sum_{k=1}^{n} \frac{\partial A_i^{R,R'}}{\partial c_e(k)} + \sum_{k=1}^{n} \frac{\partial}{\partial c_e(k)} \left\{ \sum_{j \neq i} \left[ mc^{ij}(R' \cap Q_j^0) - mc^{ij}(R \cap Q_j^0) \right] P_j \right\} \\ &= \sum_{k=1}^{n} \frac{\partial A_i^{R,R'}}{\partial c_e(k)} + \sum_{j \neq i} \left[ \sum_{k=1}^{n} \left( \frac{\partial mc^{ij}(R \cap Q_j^0)}{\partial c_e(k)} - \frac{\partial mc^{ij}(R' \cap Q_j^0)}{\partial c_e(k)} \right) \right] P_j, \end{aligned}$$

where the last line follows because $\partial P_j / \partial c_e(k) = 0$ for all $j, e, k$.

Now it remains to show that $\sum_{k=1}^{n} \frac{\partial A_i^{R,R'}}{\partial c_e(k)} = \prod_{j \neq i}(P_j + 1)$. We compute:

$$\frac{\partial A_i^{R,R'}}{\partial c_e(k)} = \begin{cases} \sum_{\sigma_{-i}} p_{\sigma_{-i}} \mathbf{1}[K_i(e) = k-1] & \text{if } e \in R \setminus R' \\ -\sum_{\sigma_{-i}} p_{\sigma_{-i}} \mathbf{1}[K_i(e) = k-1] & \text{if } e \in R' \setminus R \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

and then sum over $k$ for any $e \in R \setminus R'$:

$$\begin{aligned} \sum_{k=1}^{n} \frac{\partial A_i^{R,R'}}{\partial c_e(k)} &= \sum_{k=1}^{n} \sum_{\sigma_{-i}} p_{\sigma_{-i}} \mathbf{1}[K_i(e) = k-1] \\ &= \sum_{\sigma_{-i}} \left( p_{\sigma_{-i}} \sum_{k=1}^{n} \mathbf{1}[K_i(e) = k-1] \right) \\ &= \sum_{\sigma_{-i}} p_{\sigma_{-i}} \mathbf{1}[0 \leq K_i(e) \leq n-1] \\ &= \sum_{\sigma_{-i}} p_{\sigma_{-i}} \\ &= \sum_{\sigma_{-i}} \left( \prod_{j \neq i} p_{j,\sigma(j)} \right) = \prod_{j \neq i} \left( \sum_{Q} p_{jQ} \right) = \prod_{j \neq i} (P_j + 1). \end{aligned}$$

□

**Measure-theoretic and algebraic conclusions.** In order to apply Sard's Theorem, we need to work with a smooth manifold, whereas the set $X$ of fully mixed weakly stable equilibria may have singularities. However, we know that $X$ is contained in the affine algebraic variety $V(I)$, so we may use the following standard fact whose proof appears in section 3.6.

**Lemma 3.4.4** *If $I \subseteq \mathbb{R}[x_1, x_2, \ldots, x_n]$ then the variety $Z = V(I)$ is the union of finitely many subsets $Z_1, Z_2, \ldots, Z_m$, each of which is a nonsingular quasi-affine* *algebraic variety*[7]*, and therefore also a smooth manifold. For any polynomial* $P \in I$ *and any point* $z \in Z_j \, (1 \le j \le m)$*, the gradient vector* $\nabla P$ *at* $z$ *is orthogonal to the entire tangent space* $T_z Z_j$ *of the manifold* $Z_j$*.*

**Proof of Theorem 3.4.1** By Lemma 3.4.4, the set $X$ of fully mixed weakly stable equilibria can be covered by finitely many smooth manifolds $X_1, \ldots, X_m$, so it suffices to prove that each of them projects to a measure-zero subset of $\mathbb{R}^N$. If $F$ is the polynomial defined in Lemma 3.4.3, then for every $x \in X_j \, (1 \le j \le m)$ and every tangent vector $v \in T_x X_j$, we have $\nabla F(x) \cdot v = 0$, by Lemma 3.4.4. Recalling that $X$ is a subset of the vector space $\mathbb{R}^{M+N}$, where $M$ is the combined number of strategies in all players' strategy sets, and $N$ is the combined number of pairs $(e, k)$ such that the edge cost $c_e(k)$ is well-defined, then we may write $v = (v_p, v_c)$, where $v_p$ denotes the first $M$ components of $v$ (corresponding to the "probability coordinates") and $v_c$ denotes the last $N$ components of $v$ (corresponding to the "edge cost coordinates"). Recalling that every polynomial in $I$ vanishes at $x$, properties (1)-(2) of Lemma 3.4.3 imply that the first $M$ coordinates of $\nabla F(x)$ vanish whereas the last $N$ coordinates do not. Thus, the equation $\nabla F(x) \cdot v = 0$ imposes a nontrivial linear constraint on the vector $v_c$. This implies that the tangent space $T_x X_j$ projects to a proper linear subspace of $\mathbb{R}^N$. Since $x$ was arbitrary, we have proven that the differential of the projection map $X_j \to \mathbb{R}^N$ has rank less than $N$ at every point of $X_j$. By Sard's Theorem, the image of $X_j$ in $\mathbb{R}^N$ has measure zero. □

In fact, the technique used to prove Theorem 3.4.1 actually allows us to establish a stronger theorem, in which we consider congestion games whose edge

[7] A quasi-affine algebraic variety is any variety isomorphic to an open subset of an affine algebraic variety

cost functions are drawn from a specified class of cost functions. Let us define a *smooth, monotonically parameterized class of cost functions* to be a collection of functions $c_\gamma : \{1, 2, \ldots, n\} \to \mathbb{R}$ parameterized by a vector of real numbers $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_d) \in U$ for some open subset $U \subseteq \mathbb{R}^{d+1}$, satisfying the following two properties: (1) for all $k \in \{1, 2, \ldots, n\}$, the function $\gamma \mapsto c_\gamma(k)$ is a smooth function, which we will denote by $h_k(\gamma)$; (2) for all $k \in \{1, 2, \ldots, n\}$ and all $\gamma \in U$, the partial derivative $\partial c_\gamma(k)/\partial \gamma_0$ is strictly positive. For example, edge costs that are specified by degree-$d$ polynomials $c_e(k) = \gamma_0 + \gamma_1 k + \ldots + \gamma_d k^d$ constitute a smooth, monotonically parameterized class of cost functions. The following generalization of Theorem 3.4.1 is proven using essentially the same technique.

**Theorem 3.4.5** *Suppose that for each edge $e$ of a congestion game we are given a smooth, monotonically parameterized class of cost functions with parameter $\gamma^e \in U_e$. The set of congestion games having non-pure weakly stable equilibria is a measure-zero subset of $\prod_e U_e$.*

**Proof of Theorem 3.4.5** There is a natural smooth function $h : U_* \to \mathbb{R}^N$ which maps the parameter vector $\Gamma$ to the vector of edge costs $(c_e(k))_{e,k}$ determined by these parameters. Letting $\mathbb{R}^M$ denote the vector space spanned by the probability variable $p_{iR}$, we have the map $H : \mathbb{R}^M \times U_* \to \mathbb{R}^M \times \mathbb{R}^N$ which maps a pair $(\vec{p}, \Gamma)$ to $(\vec{p}, h(\Gamma))$. Let $X^\sharp = H^{-1}(X) \subseteq \mathbb{R}^M \times U_*$. We aim to show that the projection of $X^\sharp$ to $U_*$ has zero measure. As before, we can accomplish this by covering $X^\sharp$ with finitely many smooth manifolds $X^\sharp_j$ and using Sard's Theorem on each of these pieces. The necessary ingredient in this argument, as before, is a smooth real-valued function $F^\sharp$ which vanishes on $X^\sharp_j$, such that at every point $x \in X^\sharp_j$ the gradient $\nabla F^\sharp(x)$ is a nonzero vector whose first $M$ components vanish. We can obtain such a function by composing $F$ and $H$: $F^\sharp = F \circ H$. Using the fact that $H$ is the identity map on the first $M$ coordinates, and on the last $N$ coordinates is given by the map $\Gamma \mapsto (c_e(k))_{e,k}$, we obtain the following formulas:

$$\frac{\partial F^\sharp}{\partial p_{iR}} = \frac{\partial F}{\partial p_{iR}} = 0 \tag{3.7}$$

$$\frac{\partial F^\sharp}{\partial \gamma_\ell^e} = \sum_{k=1}^{n} \left( \frac{\partial F}{\partial c_e(k)} \right) \left( \frac{\partial c_e(k)}{\partial \gamma_\ell^e} \right). \tag{3.8}$$

This proves that the first $M$ components of $\nabla F^\sharp$ vanish. Now recall that $\partial c_e(k)/\partial \gamma_0^e > 0$ by the definition of a smooth, monotonically parameterized class of cost functions. Fix any edge $e \in R \setminus R'$ and let $v = (0, w)$ be the vector whose first $M$ components are 0 and whose last $N$ components are given by

$$v_{e,\ell} = \begin{cases} \frac{1}{\partial c_e(k)/\partial \gamma_0^e} & \text{if } \ell = 0 \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$\nabla F^\sharp \cdot v = \sum_{k=1}^{n} \frac{\partial F^\sharp/\partial \gamma_0^e}{\partial c_e(k)/\partial \gamma_0^e} = \sum_{k=1}^{n} \frac{\partial F}{\partial c_e(k)} = 1,$$

by part (2) of Lemma 3.4.3, and this clearly shows that $\nabla F^\sharp \neq 0$. Applying Sard's Theorem as in the proof of Theorem 3.4.1, it follows that the image of $X_j^\sharp$ in $\mathbb{R}^N$ has measure zero. □

Turning now from measure-theoretic statements to algebraic ones, we will present the proof of Theorem 3.4.2.

**Proof of Theorem 3.4.2** We work over the field $\mathbb{C}$ of complex numbers in order to apply theorems about varieties over an algebraically closed field; in the last sentence of the proof we will translate the result back to the field $\mathbb{R}$. Let $I$ be the ideal defined in Section 3.4, and let $X_{\mathbb{C}} = V(I)$, the zero-set of $I$ over $\mathbb{C}$. By Lemma 3.4.4, $X_{\mathbb{C}}$ can be covered by finitely many nonsingular varieties $X_1, \ldots, X_m$.

We will prove that each of them projects to a proper closed subset of the affine space $\mathbb{C}^N$. If $F$ is the polynomial defined in Lemma 3.4.3, then as in the proof of Theorem 3.4.1, the equation $\nabla F(x) \cdot v = 0$ implies[8] that $\text{rank}(d_x f) < N$, where $f$ denotes the projection map $X_j \to \mathbb{C}^N$. However, if the set $f(X_j)$ were dense in $\mathbb{C}^N$ then by ([93], Lemma II.6.2.2) we would have $\text{rank}(d_x f) = N$ for all $x$ in a nonempty open subset of $X_j$. It follows that $f(X_j)$ is contained in a proper closed subset of $\mathbb{C}^N$ for all $j$, hence $f(X_{\mathbb{C}}) = \cup_j f(X_j)$ is also contained in a proper closed subset of $\mathbb{C}^N$. This means there is a nonzero polynomial $P \in \mathbb{C}[c_1, c_2, \ldots, c_N]$ that vanishes on $f(X_{\mathbb{C}})$. If $\bar{P}$ denotes the complex conjugate of $P$, then $W = P\bar{P}$ is a polynomial in $\mathbb{R}[c_1, c_2, \ldots, c_N]$ that vanishes on $f(X)$. $\square$

## 3.5 The Discrete-Time Process

In this section we give an approximate version of our main result using the discrete-time learning process with a small amount of added noise. The added noise is a useful artifact for proving general convergence guarantees in the discrete time setting.

Let $p'(t)$ denote the mixed strategy profile at time $t$ before we introduce the noise and let $p(t)$ express the resulting distribution. The original strategy distribution $p'(t)$ is scaled down by a factor of $1-\varepsilon^2$ and then shifted by mixing it with the uniform distribution. The updated vector has the form $(1-\varepsilon^2)p'(t) + \varepsilon^2\vec{1}$. Next, a player is chosen at random (i.e. player $i$) who in turns chooses uniformly and independently two of her strategies $R$ and $\bar{R}$. If both $p_{iR}, p_{i\bar{R}}$ are greater than $\varepsilon$ then the player proceeds to move $\varepsilon/2$ from the first strategy to the second. If that

[8] See Lemma 3.6.2 in section 3.6 for a rigorous derivation of this step using the standard algebraic-geometry definitions of tangent spaces and differentials.

is not the case, then she proceeds to move merely $\varepsilon^2/2$ . This update is always possible because of our first mixing step.

**Approximately Stable Equilibria** We need to extend the definitions of stable and unstable fixed points to approximate stability. For the dynamical systems definition, we define a *$\nu$-stable fixed point*, for any real number $\nu \geq 0$, to be a fixed point at which the Jacobian has no eigenvalues whose real part is greater than $\nu$.

For simplicity, we assume in this section that $\beta(i, p) = 1$ for all players $i$ and all probability distributions $p$. This section outlines a proof of the following result.

**Theorem 3.5.1** *The discrete-time learning process with a small amount of added noise satisfies the following guarantee for all congestion games with and non-decreasing cost functions.*

> *For all $\nu > 0$ there exists an $\epsilon > 0$, and for all $\epsilon < \epsilon_0$ there exists $T_0$, so that when all players are using $Hedge(\epsilon)$ to optimize their strategies, then for all times $T > T_0$ with probability at least $1 - \nu$, the mixed strategy profile is a $\nu$-stable equilibrium point in all but $\nu T$ of the first $T$ steps of the history of play.*

To use this theorem in the context of games, we need to relate our game-theoretic notion of weak stability to this notion of $\nu$-stable fixed points. We say that a Nash equilibrium $p$ is $\bar{\nu}$-weakly stable, if the following holds for each pair of players $i$ and $j$. Suppose we randomly sample a strategy of $j$ with probability distribution $p_j$, and assume that $j$ plays this sampled strategy $Q$ and all other players $k \neq i, j$ play with the given probability distribution $p_k$. Now sampling two strategies $R$ and $\bar{R}$ of $i$ with probability distribution $p_i$, the expected difference of payoffs of player $i$ between strategies $R$ and $\bar{R}$ is at most $\bar{\nu}$. Note that 0-weakly stable is exactly our definition of weakly stable. We will show that the learning process spends almost all the time near a $\nu$-stable equilibrium point. To conclude the same for $\bar{\nu}$-weakly stable Nash equilibrium, we need to show that $\nu$-stable implies $\bar{\nu}$-weakly stable.

**Theorem 3.5.2** *For every $\bar{\nu} > 0$ there is a $\nu > 0$ so that if a Nash equilibrium $p$ is $\nu$-stable for the dynamical system, then it is a $\bar{\nu}$-weakly stable Nash equilibrium.*

**Proof:** Assume $p$ is $\nu$-stable, that is, for all eigenvalues $\lambda$ of the Jacobian $J$ we have $\Re(\lambda) \leq \nu$. If there are $N$ players and a total of $M$ strategies, this implies $\mathrm{Tr}(\mathfrak{J}^2) \leq \nu^2 M$. We know from Theorem 3.3.6 that this trace is equal to

$$\sum_{i,j} \sum_{R<\bar{R},Q<\bar{Q}} p_{iR}p_{i\bar{R}}p_{jQ}p_{j\bar{Q}}(M_{i,j}^{R,\bar{R},Q,\bar{Q}})^2 \tag{3.9}$$

where $M_{i,j}^{R,\bar{R},Q,\bar{Q}} = mc_{ij}(R \cap Q) - mc_{ij}(R \cap \bar{Q}) - mc_{ij}(\bar{R} \cap Q) + mc_{ij}(\bar{R} \cap \bar{Q})$, and recall that we are assuming here that $\beta(i,p) = 1$ for all $i$ and $p$ for simplicity.

Now consider the change in the cost of strategy $R$ for player $i$ when player $j$ selects a strategy $Q$. This change is $mc_{ij}(R \cap Q) - \sum_{\bar{Q}} p_{j\bar{Q}} mc_{ij}(R \cap \bar{Q})$. The difference in cost between strategies $R$ and $\bar{R}$ is then

$$\begin{aligned} &\sum_{\bar{Q}} p_{j\bar{Q}}(mc_{ij}(R \cap Q) - mc_{ij}(R \cap \bar{Q}) - \\ &- mc_{ij}(\bar{R} \cap Q) + mc_{ij}(\bar{R} \cap \bar{Q})) = \sum_{\bar{Q}} p_{j\bar{Q}} M_{i,j}^{R,\bar{R},Q,\bar{Q}}. \end{aligned}$$

Taking the expectation defining the $\bar{\nu}$-weakly stable Nash equilibrium, the sum that we need to bound is

$$\sum_{R,\bar{R},Q,\bar{Q}} p_{iR}p_{i\bar{R}}p_{jQ}p_{j\bar{Q}} M_{i,j}^{R,\bar{R},Q,\bar{Q}}. \tag{3.10}$$

Restricting the sum of squares in (3.9) to only the $(i,j)$ pair and using the Cauchy-Schwarz inequality we can bound the sum in (3.10) by $2\nu\sqrt{M}$, which establishes the theorem with $\bar{\nu} = 2\nu\sqrt{M}$. $\square$

### 3.5.1 Overview of the Discrete Analysis

In deriving Theorem 3.5.1 from the foregoing continuous-time analysis, we must address several sources of error: the noise introduced by the players' random sampling, the evolution of mixed strategies in discrete jumps rather than along continuous flow lines, and the approximation error resulting from treating $\varepsilon$ as infinitesimally small when estimating the coefficients of the vector field $\xi$. Resolving these issues requires careful manipulations of Taylor series, but it is also possible to distinguish a few main ideas which constitute a road map for this stage of the proof.

**Amortizing unstable-fixed-point steps against potential-diminishing steps.** Our analysis of the stochastic process $p(t)$ distinguishes three types of time steps: *potential-shrinking* steps in which the expected decrease in $\Psi$ is at least $\Omega(\varepsilon^2)$, *stable* steps in which $p(t)$ is near a $\nu$-stable equilibrium point, and *unstable* steps in which $p(t)$ is near an equilibrium point which is not $\nu$-stable. To deal with unstable steps, we show that a sufficiently long time window starting at an unstable step $t$ will contain (with high probability) many more potential-shrinking steps than unstable steps. Amortizing the change in $\Psi$ over the entire time window, we can show that the expected potential decrease is $\Omega(\varepsilon^2)$. Thus, the entire time history $t = 1, 2, \ldots$ can be broken up into *good stages* consisting of a single stable step, and *bad stages* of bounded length such that the expected potential decrease during a bad stage is $\Omega(\varepsilon^2)$. Since $\Psi$ can only decrease by a bounded amount over the entire history of play, we may easily conclude that with high probability, the good stages vastly outnumber the bad stages in any sufficiently long time history. We omit most details of the proof from this extended abstract, and for the remainder of this section we focus on one of the most crucial steps.

**Balancing error terms by completing the square.** The most involved step in the preceding outline is the proof that every sufficiently long time window which begins with an unstable step is likely to contain many more potential-shrinking steps than unstable steps. The difficulty is as follows: the rate at which an unstable fixed point $p^0$ repels points at distance $\rho$ from $p^0$ (along an unstable direction) is $O(\rho^2)$. This second-order effect is offset by a second-order correction term arising from our approximation of the multiplicative-update rule by the vector field $\xi$. To compare these two effects we use an analogue of "completing the square": instead of basing a Taylor expansion at the fixed point $p^0$ we choose a nearby basepoint $p^1$, resulting in a Taylor expansion whose leading-order term has an unambiguous sign reflecting the system's tendency to move away from $p^1$ in the repelling direction.

### 3.5.2 Details of the Discrete-time analysis

Throughout this section we normalize the edge costs (multiplying each cost function by a scalar if necessary) so that the cost of each path is bounded above by 1 even if every edge of the path experiences a load of $n$.

We present a noisy variant of the multiplicative updates algorithm. The added noise is a useful artifact for proving general convergence guarantees in the discrete time setting. Going back to the replicator dynamics analogy, the added noise introduces mutations which allows the system to fully explore the strategy space and recover even from very unlikely configurations.

Let $p'(t)$ denote the mixed strategy profile at time $t$ before we introduce the noise and let $p(t)$ express the resulting distribution. The original strategy distribution $p'(t)$ is scaled down by a factor of $1-\varepsilon^2$ and then shifted by mixing it with the uniform distribution. The updated vector has the form $(1-\varepsilon^2)p'(t)+\varepsilon^2\vec{1}$. Next, a player is chosen at random [9] (i.e. player $i$) who in turns chooses uniformly and independently two of her strategies $R$ and $\bar{R}$. If both $p_{iR}, p_{i\bar{R}}$ are greater than $\varepsilon$ then the player proceeds to move $\varepsilon/2$ from the first strategy to the second. If that is not the case, then she proceeds to move merely $\varepsilon^2/2$ . This update is always possible because of our first mixing step. We refer to the process of adding this noise as "wiggling" henceforth. Let's denote by $N_{iR}(t)$ the random variable $p_{iR}(t)-p'_{iR}(t)$. It is easy to verify that for each $i$ and $R$ $\mathbb{E}(N_{iR}(t)) = O(\varepsilon^2)$ for all $t$, since the expected drift of the second step is zero.

The random variables $p(t)$ constitute a Markov chain on the state space $\Sigma$, the set of all mixed profiles. We will represent the state change that takes place at time $t$ by a random vector $\hat{q}(t) = p(t+1)-p(t)$ whose distribution depends only on the value of $p(t)$. Hence for every mixed strategy profile $p$ there is a well-defined "expected state change" vector $q(p)$. For any player $i$ and path $R$, if $p(t)=p$ then

[9] Our results hold regardless of the numbers players that perform this update. In other words, there is no need for communication between the players to decide who performs the update.

$$
\begin{aligned}
p(t+1)_{iR} &= \frac{(1-\varepsilon)^{\mathrm{cost}(R,t)} p_{iR}}{\sum_Q (1-\varepsilon)^{\mathrm{cost}(Q,t)} p_{iQ}} + N_{iR}(t) \\
&= p_{iR} \frac{1 - \mathrm{cost}(R,t)\varepsilon + \frac{1}{2}\mathrm{cost}(R,t)(\mathrm{cost}(R,t)-1)\varepsilon^2 + O(\varepsilon^3)}{1 - \left(\sum_Q \mathrm{cost}(Q,t) p_{iQ}\right)\varepsilon + \frac{1}{2}\left(\sum_Q \mathrm{cost}(Q,t)(\mathrm{cost}(Q,t)-1) p_{iQ}\right)\varepsilon^2 + O(\varepsilon^3)} \\
&+ N_{iR}(t) \\
&= p_{iR} + \left(\sum_Q \left(\mathrm{cost}(Q,t) - \mathrm{cost}(R,t)\right) p_{iR} p_{iQ}\right)\varepsilon + O(p_{iR})\varepsilon^2 + N_{iR}(t)
\end{aligned}
$$

Therefore $\|\hat{q}(t)\|_1 = O(n\varepsilon)$ for all $t$, and for all $p$ there is a vector $\eta(p)$ (depending on $\varepsilon$) such that $\|\eta(p)\|_1 = O(n)$ and

$$
q(p) = \mathbb{E}(p(t+1) - p(t) \,|\, p(t) = p) = \xi(p)\varepsilon + \eta(p)\varepsilon^2. \tag{3.11}
$$

We now proceed to estimate the expected change in potential when $p(t) = p$, using $\nabla\Psi(p)$ and $H_\Psi(p)$ to denote the gradient and the Hessian matrix of $\Psi$ at $p$, respectively. These are given by the formulas

$$
\nabla\Psi(p)_{iR} = c^i(R) \tag{3.12}
$$

$$
H_\Psi(p)_{iR,jQ} = mc^{ij}(R \cap Q), \tag{3.13}
$$

from which it follows that $\|\nabla\Psi(p)\|_\infty \le 1$ and $|H_\Psi(p)|_{\max} \le 1$. Using the fact that $|u^\mathsf{T} A v| \le \|u\|_1 |A|_{\max} \|v\|_1$ for any vectors $u, v$ and matrix $A$ such that the product $u^\mathsf{T} A v$ is well-defined, we obtain the bound

$$
|\hat{q}(t)^\mathsf{T} H_\Psi(p(t)) \hat{q}(t)| = O(n^2\varepsilon^2). \tag{3.14}
$$

Now applying Taylor's Theorem to the function $\Psi$, we obtain

$$\begin{aligned}\mathbb{E}(\Psi(p(t+1)) - \Psi(p(t)) \mid p(t) = p) &= \nabla\Psi(p)^{\mathsf{T}} q(p) + \mathbb{E}(\hat{q}(t)^{\mathsf{T}} H_{\Psi}(p)\hat{q}(t) \mid p(t) = p) + \\ &\quad + O(n^3\varepsilon^3) \qquad (3.15)\\ &= \nabla\Psi(p)^{\mathsf{T}}\xi(p)\varepsilon + O(n\varepsilon^2) + O(n^2\varepsilon^2)\\ &= O(n\varepsilon^2) - \Omega\left(\frac{1}{n}\|\xi(p)\|_1^2\varepsilon\right), \qquad (3.16)\end{aligned}$$

where the last line was derived using Theorem 3.3.2. When $p(t)$ is far from the fixed point set of $\xi$, i.e. $\|\xi(p(t))\|_1^2 > Cn^3\varepsilon$ for a sufficiently large constant $C$, equation (3.16) ensures that we are in a potential-shrinking state.

Furthermore we will justify that condition (3.16) coupled with the fact that we keep decreasing the $\varepsilon$ in the multiplicative update algorithms[10] implies convergence to fixed points of $\xi$. Indeed, for any (congestion) game, there exists a function $f : \mathbb{R} \to \mathbb{R}$ such that $\lim_{x\to 0} f(x) = 0$ such that all $\kappa$-approximate fixed points (in the sense $\|\xi\|_1 \leq k$) are at most $f(k)$ distance away from a Nash. Assume that this is not the case. Let's define for any $\kappa$ the set of such $\kappa$-approximate Nash as $N_\kappa$. We denote the max min distance between a set of points $S$ in our strategy space and the set of Nash equilibrium points as $d_N(S)$. Since $N_\kappa$ is compact and $d_N$ is continuous, $d_N$ attains its supremum over $N_\kappa$ for every $\kappa > 0$ and by hypothesis it is greater than 0. However the sequence of $d_N(N_\kappa)$ must have a converging subsequence, and since it is decreasing it has a limit. By hypothesis the limit would have a non-zero value and hence we reach a contradiction.

Next, we will proceed to show that our process will not converge to an unstable fixed point. To analyze the behavior of the discrete-time process near an unstable fixed point $p^0$, we will need to use Taylor's Theorem applied to the vector field $\xi$. A fact which aids in the analysis is the following lemma, which bounds the

[10] which we must in order to ensure no-regret in the discrete time setting

magnitude of each term of the Taylor series for $\xi(p^0+r)$ in terms of $\|r\|_1$ rather than the customary bounds in terms of $\|r\|_2$.

**Lemma 3.5.3** *For any vectors $p^0, r$ such that both $p^0$ and $p^0+r$ belong to $\Sigma$, we have*

$$\xi(p^0+r) = \xi(p^0) + J(p^0)r + O(\|r\|_1^2)\tilde{r}, \tag{3.17}$$

*where $\|\tilde{r}\|_1 \le 1$. Moreover, $\|J(p^0)r\|_1 = O(n\|r\|)$.*

**Proof:** Write $r = ar_0$, where $a = \|r\|_1$ and $\|r_0\|_1 = 1$. Now apply Taylor's Theorem with remainder to the vector-valued univariate function $\zeta(x) = \xi(p^0 + xr_0)$.

$$\zeta(a) = \zeta(0) + a\zeta'(0) + \frac{1}{2}a^2\zeta''(b), \quad \text{for some } b \in [0,a]. \tag{3.18}$$

Now

$$\zeta'(0)_{iR} = \sum_{j,Q} \left(\frac{\partial \xi_{iR}}{\partial p_{jQ}}\right)_{p=p^0} (r_0)_{jQ} = \left(J(p^0)r_0\right)_{iR}, \tag{3.19}$$

and

$$\zeta''(b)_{iR} = \sum_{j,Q}\sum_{j',Q'} \left(\frac{\partial^2 \xi_{iR}}{\partial p_{jQ}\partial p_{j'Q'}}\right)_{p=p^0+br_0} (r_0)_{jQ}(r_0)_{j'Q'}. \tag{3.20}$$

In (3.20), the partial derivative in each term of the sum is $O(p_{iR})$, except when $(j,Q) = (i,R)$ or $(j',Q') = (i,R)$, and in those cases it is $O(1)$.

If there is a $\nu$-unstable fixed point $p^0$ near $p(t)$, then the Jacobian matrix $J = J(p^0)$ has an eigenvalue $\lambda \in \mathbb{C}$ and left eigenvector $w^\mathsf{T}$ such that $w^\mathsf{T} J = \lambda w^\mathsf{T}$ and $\Re(\lambda) \ge \nu$. We rescale $w$ if necessary so that $\|w\|_\infty = 1$. To show that $p^0$ has a tendency to repel $p(t)$, let $r(t) = p(t) - p^0$ and we estimate the expected change in the quantity $|w^\mathsf{T} r(t)|^2 = r(t)^\mathsf{T}\bar{w}w^\mathsf{T} r(t)$. For notational convenience, we abbreviate $q(t), \hat{q}(t), r(t), \|r(t)\|_1$ by $q, \hat{q}, r, \rho$, and we use $\mathbb{E}(\cdots)$ to denote conditional expectation given $p(t) = p$. Note that $r(t+1) = r + \hat{q}$.

$$
\begin{aligned}
\mathbb{E}\left(|w^\intercal(r+\hat{q})|^2\right) &= \mathbb{E}\left(w^\intercal(r+\hat{q})\overline{w^\intercal(r+\hat{q})}\right) && (3.21)\\
&= \mathbb{E}\left(w^\intercal(r+\hat{q})\overline{w^\intercal}(\bar{r}+\bar{\hat{q}})\right) && (3.22)\\
&= \mathbb{E}\left(w^\intercal r\overline{w^\intercal}\bar{r} + w^\intercal\hat{q}\overline{w^\intercal}\bar{r} + w^\intercal r\overline{w^\intercal}\bar{\hat{q}} + w^\intercal\hat{q}\overline{w^\intercal}\bar{\hat{q}}\right) && (3.23)\\
&= \mathbb{E}\left(w^\intercal r\overline{w^\intercal r} + w^\intercal\hat{q}\overline{w^\intercal r} + w^\intercal r\overline{w^\intercal\hat{q}} + w^\intercal\hat{q}\overline{w^\intercal\hat{q}}\right) && (3.24)\\
&= |w^\intercal r|^2 + 2\Re(\bar{w^\intercal}\bar{r}w^\intercal q) + \mathbb{E}\left(|w^\intercal\hat{q}|^2\right) && (3.25)\\
&= |w^\intercal r|^2 + 2\Re(\bar{w^\intercal}\bar{r}w^\intercal q) + \mathbb{E}\left(|w^\intercal\hat{q}|^2\right) && (3.26)\\
&= |w^\intercal r|^2 + 2\Re\left[\bar{w^\intercal}\bar{r}w^\intercal\left(\xi(p)\varepsilon+\eta(p)\varepsilon^2\right)\right] + \mathbb{E}\left(|w^\intercal\hat{q}|^2\right) && (3.27)\\
&= |w^\intercal r|^2 + 2\Re[\bar{w^\intercal}\bar{r}w^\intercal((\xi(p^0)+J(p^0)r+O(\|r\|_1^2)\tilde{r})\varepsilon+ \\
&+ \eta(p)\varepsilon^2)] + \mathbb{E}\left(|w^\intercal\hat{q}|^2\right) && (3.28)\\
&= |w^\intercal r|^2 + 2\Re\left[\bar{w^\intercal}\bar{r}w^\intercal\left(\left(J(p^0)r+O(\|r\|_1^2)\tilde{r}\right)\varepsilon+\eta(p)\varepsilon^2\right)\right] + \\
&+ \mathbb{E}\left(|w^\intercal\hat{q}|^2\right) && (3.29)\\
&= |w^\intercal r|^2 + 2\Re\left[\bar{w^\intercal}\bar{r}w^\intercal\left(J(p^0)r\varepsilon+\eta(p)\varepsilon^2\right)\right] + \mathbb{E}\left(|w^\intercal\hat{q}|^2\right) \pm \\
&\pm O(\varepsilon\rho^3) && (3.30)\\
&= |w^\intercal r|^2 + 2\Re\left[\bar{w^\intercal}\bar{r}w^\intercal\left(Jr\varepsilon+\eta(p^0)\varepsilon^2+(\eta(p)-\eta(p^0))\varepsilon^2\right)\right] + \\
&+ \mathbb{E}\left(|w^\intercal\hat{q}|^2\right) \pm O(\varepsilon\rho^3) && (3.31)\\
&= |w^\intercal r|^2 + 2\Re\left[\bar{w^\intercal}\bar{r}w^\intercal\left(Jr\varepsilon+\eta(p^0)\varepsilon^2\right)\right] + \mathbb{E}\left(|w^\intercal\hat{q}|^2\right) \pm \\
&\pm O(\varepsilon^2\rho^2) \pm O(\varepsilon\rho^3) && (3.32)\\
&= |w^\intercal r|^2 + 2\Re\left[\bar{w^\intercal}\bar{r}w^\intercal\left(\lambda r\varepsilon+\eta(p^0)\varepsilon^2\right)\right] + \mathbb{E}\left(|w^\intercal\hat{q}|^2\right) \pm \\
&\pm O(\varepsilon\rho^2(\rho+\varepsilon)). && (3.33)
\end{aligned}
$$

The expression $\lambda r\varepsilon+\eta(p^0)\varepsilon^2$ is problematic because it is unclear which term is of leading order (it depends on the relative magnitudes of $\rho,\varepsilon$) and the two terms could even have opposite signs and cancel each other. To deal with this, we let

$p^1 = p^0 - (\lambda)^{-1}\eta(p^0)\varepsilon$ and $s(t) = p(t) - p^1$. By proving that we diverge from this nearby point $p^1$, we will have that we diverge from $p^0$ as well. Let $\sigma$ be $\|s(t)\|_1$ :

$$
\begin{aligned}
\mathbb{E}\left(|w^\intercal(s+\hat{q})|^2\right) &= \mathbb{E}\left(w^\intercal(s+\hat{q})\overline{w^\intercal(s+\hat{q})}\right) && (3.34)\\
&= \dots && (3.35)\\
&= |w^\intercal s|^2 + 2\Re(\bar{w^\intercal}\bar{s}w^\intercal q) + \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) && (3.36)\\
&= |w^\intercal s|^2 + 2\Re\left[\bar{w^\intercal}\bar{s}w^\intercal\left(\xi(p)\varepsilon + \eta(p)\varepsilon^2\right)\right] + \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) && (3.37)\\
&= |w^\intercal s|^2 + 2\Re[\bar{w^\intercal}\bar{s}w^\intercal((\xi(p^0) + J(p^0)r + O(\|r\|_1^2)\tilde{r})\varepsilon + \\
&+ \eta(p)\varepsilon^2)] + \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) && (3.38)\\
&= |w^\intercal s|^2 + 2\Re\left[\bar{w^\intercal}\bar{s}w^\intercal\left(\left(J(p^0)r + O(\|r\|_1^2)\tilde{r}\right)\varepsilon + \eta(p)\varepsilon^2\right)\right] + \\
&+ \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) && (3.39)\\
&= |w^\intercal s|^2 + 2\Re\left[\bar{w^\intercal}\bar{s}w^\intercal\left(J(p^0)r\varepsilon + \eta(p)\varepsilon^2\right)\right] + \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) \pm \\
&\pm O(\varepsilon\sigma\rho^2) && (3.40)\\
&= |w^\intercal s|^2 + 2\Re\left[\bar{w^\intercal}\bar{s}w^\intercal\left(Jr\varepsilon + \eta(p^0)\varepsilon^2 + (\eta(p) - \eta(p^0))\varepsilon^2\right)\right] + \\
&+ \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) \pm O(\varepsilon\sigma\rho^2) && (3.41)\\
&= |w^\intercal s|^2 + 2\Re\left[\bar{w^\intercal}\bar{s}w^\intercal\left(Jr\varepsilon + \eta(p^0)\varepsilon^2\right)\right] + \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) \pm \\
&\pm O(\varepsilon^2\sigma\rho) \pm O(\varepsilon\sigma\rho^2) && (3.42)\\
&= |w^\intercal s|^2 + 2\Re\left[\bar{w^\intercal}\bar{s}w^\intercal\left(\lambda r\varepsilon + \eta(p^0)\varepsilon^2\right)\right] + \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) \pm \\
&\pm O(\varepsilon\sigma\rho(\rho+\varepsilon)) && (3.43)\\
&= |w^\intercal s|^2 + 2\Re\left[\lambda\varepsilon\bar{w^\intercal}\bar{s}w^\intercal\left(r + (\lambda)^{-1}\eta(p^0)\varepsilon\right)\right] + \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) \pm \\
&\pm O(\varepsilon\sigma\rho(\rho+\varepsilon)) && (3.44)\\
&= |w^\intercal s|^2 + 2\Re(\lambda\varepsilon\bar{w^\intercal}\bar{s}w^\intercal s) + \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) \pm O(\varepsilon\sigma\rho(\rho+\varepsilon)) && (3.45)\\
&= (1 + 2\Re(\lambda)\varepsilon)|w^\intercal s|^2 + \mathbb{E}\left(|w^\intercal \hat{q}|^2\right) \pm O(\varepsilon\sigma\rho(\rho+\varepsilon)). && (3.46)
\end{aligned}
$$

However, we have by the definition of $\sigma$ that: $\sigma = \|s(t)\|_1$.

$$\begin{aligned} s(t) &= p(t) - p^1 \\ &= p(t) - p^0 + p^0 - p^1 \\ &= r(t) + (\lambda)^{-1}\eta(p^0)\varepsilon \end{aligned}$$

From the above equation, we derive that $\sigma = \rho + O(\frac{\varepsilon}{\nu})$. Our analysis depends on the effective ratio between $\varepsilon, \nu$. Namely, we need to define a relation $\varepsilon = f(\nu)$ such that for every $\nu > 0$, the gameplay diverges from $\nu$-stable equilibria, if $\varepsilon \leq f(\nu)$. Any such relation must fit the requirement that $\frac{\varepsilon}{\nu}$ converges to zero as $\varepsilon$ goes to zero, so that $p^1$ is a true approximation to $p^0$. Here, we set $\nu = \Theta(\varepsilon^{0.05})$. By combining this with (3.46), we derive that:

$$\mathbb{E}\left(|w^\mathsf{T}(s+\hat{q})|^2 \;\middle|\; p(t) = p\right) \geq (1+2\Re(\lambda)\varepsilon)|w^\mathsf{T}s|^2 + \mathbb{E}\left(|w^\mathsf{T}\hat{q}|^2\right) \pm O(\varepsilon\rho(\rho+\varepsilon)(\rho+\varepsilon^{0.95})) \tag{3.47}$$

In the range $\rho = O(\varepsilon^{0.4})$, the error term is of size $O(\varepsilon^{2.2})$. On the other hand, as we will show $\mathbb{E}\left(|w^\mathsf{T}\hat{q}|^2\right) = \Omega(\varepsilon^{2.1})$ when we are near an unstable fixed point.

**Lemma 3.5.4** *Let $J_p$ be the Jacobian of our game at a mixed Nash equilibrium $p$ and let $w^\mathsf{T}$ be a left eigenvector corresponding to an eigenvalue with a positive real part, then there exists player $i$ and strategies $R, \bar{R}$ with $p_{iR}, p_{i\bar{R}} > 0$ such that $w_{iR} \neq w_{i\bar{R}}$.*

First, we will prove that any vector $w^\mathsf{T}$ such that for each player $i$ and strategies $R, \bar{R}$ of that player $w_{iR} = w_{i\bar{R}}$ holds, is a left eigenvector of $J_p$ with eigenvalue of zero. We will prove this by showing that for each column $jQ$ of the jacobian the sum of the terms corresponding to strategies played by any player $i$ is equal to 0.The sum of all the terms of column $jQ$ corresponding to player $i$ is the partial derivative of $\sum_R \xi_{iR}$ with respective to $p_{jQ}$ . However $\sum_R \xi_{iR} = 0$ since $\xi$ is a tangent vector to the mixed strategy polytope. Hence, each partial derivative of it will be zero as well, implying the desired $\sum_{iR} \frac{\partial \xi_{iR}}{\partial p_{jQ}} = 0$

Starting from such a vector let's examine what happens to the vector $w^\mathsf{T} J_p$ when we alter some $w_{iR}$ entry of $w$ such that $p_{iR} = 0$. However, on the $iR$ row of $J_p$ there is at most one non zero element. That is the element of the diagonal whose entry is $\sum_{\bar{R}} p_{i\bar{R}}(c^i(\bar{R}) - c^i(R)) \leq 0$, since $p$ is a Nash equilibrium. Any such change on $w$ only affects the $iR$ entry of $w^\mathsf{T} J_p$ whereas all other remain zero. Therefore for any vector $w$ with the property for each player $i$ and strategies $R, \bar{R}$ with $p_{iR}, p_{i\bar{R}} > 0$ $w_{iR} = w_{i\bar{R}}$, we have that every $jQ$ coordinate of $w^\mathsf{T} J_p$ with $p_{jQ} > 0$ remains equal to 0. Hence, for such a vector $w$ to be an eigenvector with an nonzero eigenvalue it has to be the case that $w_{iR} = 0$ for each $p_{iR} > 0$. In such a case though, the only terms of $J_p$ that come into play are the nonpositive terms on the diagonal. The corresponding eigenvalue cannot have a positive real part. Hence if $J_p$ has an eigenvalue with a positive real then there must exist player $i$ and strategies $R, \bar{R}$ with $p_{iR}, p_{i\bar{R}} > 0$ such that $w_{iR} \neq w_{i\bar{R}}$. □

**Lemma 3.5.5** *Let $J_p$ be the Jacobian of our game at a mixed Nash equilibrium $p$ and let $w^\mathsf{T}$ be a left eigenvector corresponding to an eigenvalue $\lambda$ with positive real part $\Re(\lambda) \geq \nu > 0$. Assume $\|w\|_\infty = 1$. Then for some constant $c$ depending only on the number of players and strategies, there exists a player $i$ with strategies $R, \bar{R}$ satisfying $p_{iR}, p_{i\bar{R}} > 0$ and $|w_{iR} - w_{i\bar{R}}| > c\nu$.*

**Proof:** Let us partition the pairs of strategies $(i, R)$ such that $i$ is a player and $R$ is a strategy in $\mathcal{S}_i$ into two sets $A$ and $B$: $A$ consists of the pairs such that $p_{iR} > 0$ and $B$ consists of the pairs such that $p_{iR} = 0$. Consider the matrix $J_p$ as composed of blocks $J_{AA}, J_{AB}, J_{BA}, J_{BB}$, where, for instance, an entry $\partial\xi_{iR}/\partial p_{jQ}$ belongs to the submatrix $J_{AB}$ if $(i, R) \in A$ and $(j, Q) \in B$.

$$J_p = \begin{pmatrix} J_{AA} & J_{AB} \\ J_{BA} & J_{BB} \end{pmatrix}$$

By inspection of the formulas for $\partial\xi_{iR}/\partial p_{jQ}$, we see that $J_{BA} = 0$ and $J_{BB}$ is a diagonal matrix. Moreover, the diagonal entries of $J_{BB}$ are non-positive because of our assumption that $p$ is a mixed Nash equilibrium.

Let us partition the components of our eigenvector $w$ according to the partition $A, B$, obtaining vectors $w_A, w_B$. We first claim that

$$\|w_A\|_\infty \geq 1/M \tag{3.48}$$

where $M$ denotes the combined number of (player,strategy) pairs, i.e. the number of rows and columns in the matrix $J_p$. If the largest component of $w$ belongs to $w_A$, then (3.48) follows trivially from our assumption that $\|w\|_\infty \geq 1$. Let $(j, Q) \in B$ be the index such that $|w_{jQ}| = 1$, and let $J_{*,jQ}$ denote the $(j, Q)$ column of $J$. The equation $w^\mathsf{T} J = \lambda w^\mathsf{T}$ implies

$$\begin{aligned} w_A^\mathsf{T} J_{A,jQ} + w_B^\mathsf{T} J_{B,jQ} = \lambda w_{jQ} \\ w_A^\mathsf{T} J_{A,jQ} = (\lambda - J_{jQ,jQ}) w_B^\mathsf{T} \\ \|w_A^\mathsf{T}\|_\infty \|J_{A,jQ}\|_1 \geq \nu - J_{jQ,jQ} \geq \nu, \end{aligned}$$

and the claim follows immediately because each entry of $J$ lies in $[-1, 1]$, hence $\|J_{A,jQ}\|_1 \leq M$.

Now for each player $i$ let $A_i$ be the subset of $A$ consisting of pairs $(i, R)$. Note that $A_i$ is nonempty for every $i$ because not every probability $p_{iR}$ can be equal to zero. Let $z_i$ denote the average of the numbers $w_{iR}$, $(i, R) \in A_i$, i.e.

$$z_i = \frac{1}{|A_i|} \sum_{(i,R)\in A_i} w_{iR}.$$

Define a vector $v$ as follows:

$$v_{iR} = \begin{cases} z_i & \text{if } (i, R) \in A_i \\ 0 & \text{otherwise.} \end{cases}$$

Observe that $v_A J_{AA} = 0$ because for any $(j, Q) \in A$,

$$\sum_{(i,R)\in A} v_{iR} J_{iR,jQ} = z_j \left( J_{jQ,jQ} + \sum_{\bar{Q}\neq Q,(j,\bar{Q}\in A)} J_{j\bar{Q},jQ} \right) + \sum_{i\neq j} z_i \left( \sum_{(i,R)\in A} J_{iR,jQ} \right)$$

and each of the terms in parentheses on the right side vanishes, using the formula for the entries of $J$.

Now consider the equation

$$(w_A - v_A)^\mathsf{T} J_{AA} = w_A^\mathsf{T} J_{AA} = \lambda w_A^\mathsf{T}. \tag{3.49}$$

If $(j, Q) \in A$ is such that $|w_{jQ}| \geq 1/M$ then (3.49) implies

$$\|w_A - v_A\|_\infty \|J_{A,jQ}\|_1 \geq \nu/M,$$

hence $\|w_A - v_A\|_\infty \geq \nu/M^2$. If $(i, R)$ is such that $|w_{iR} - z_i| \geq \nu/M^2$, it means that $w_{iR}$ differs from the average of the numbers $\{w_{i\bar{R}} \mid (i, \bar{R}) \in A\}$ by at least $\nu/M^2$. Consequently there exists $(i, \bar{R}) \in A$ such that $w_{iR} - w_{i\bar{R}} \geq \nu/M^2$.

We proceed to bound the term $\mathbb{E}_{i,R,\bar{R}} \left( |w^\mathsf{T} \hat{q}|^2 \right)$. We have assumed that the state of the game is at most distance $\rho$ away from a unstable Nash equilibrium. However, all such unstable Nash are mixed, since otherwise they would be stable. If the unstable point is at a distance $\Omega(\varepsilon^{1/3})$ from its nearest pure strategy profile, then any strategy profile within distance $\rho = O(\varepsilon^{0.4})$ has players that choose at least 2 strategies with probability $\Omega(\varepsilon)$. As a result, during the wiggling step we wiggle by $\varepsilon/2$ with probability at least $1/(n \max_i |S_i|^2)$. Let's $\Gamma$ denote $|(1-\varepsilon^2)|w^\intercal(p'(t+1)-p(t)) + \varepsilon^2 w^\intercal(\vec{1}-p(t))$ and let $W$ express the wiggling size.

$$\begin{aligned}
\mathbb{E}_{i,R,\bar{R}}\left(|w^\intercal \hat{q}|^2\right) &= \mathbb{E}\big[\big(|(1-\varepsilon^2)|w^\intercal(p'(t+1)-p(t)) + \varepsilon^2(\vec{1}-p(t)) + \\
&+ W(w_{iR} - w_{i\bar{R}})\big)^2\big] \\
&\geq \frac{\left(\Gamma + \frac{\varepsilon}{2}(w_{iR} - w_{i\bar{R}})\right)^2 + \left(\Gamma + \frac{\varepsilon}{2}(w_{i\bar{R}} - w_{iR})\right)^2}{n \max_i |S_i|^2} \\
&= \Omega((\varepsilon\nu)^2) \\
&= \Omega(\varepsilon^{2.1})
\end{aligned}$$

therefore the $\mathbb{E}\left(|w^\intercal \hat{q}|^2\right)$ term dominates the error term in (3.47) and since it is positive (3.47) turns into:

$$\mathbb{E}\left(|w^\intercal(s+\hat{q})|^2 \;\mid\; p(t) = p\right) \geq (1 + 2\Re(\lambda)\varepsilon)|w^\intercal s|^2 \tag{3.50}$$

Equation (3.50) allows to conclude that with probability at least 1/2, starting from an unstable time $t_0$ we reach a state in which $\rho = \Theta(\varepsilon^{0.4})$ after no more than $u = O((\nu\varepsilon)^{-1}\log(\rho/\varepsilon^2)) = O(\varepsilon^{-1.05}\log(1/\varepsilon))$ time steps. We have that $c\varepsilon^{0.4} \leq \sum_{t=t_0}^{t_0+u-1} \|\hat{q}(t)\|_1$ (where $c$ suitable constant). We will argue that $\xi(p(t))\varepsilon$ approximates the expectation of $\hat{q}(t)$ close enough so as obtain $\bar{c}\varepsilon^{0.4} \leq \sum_{t=t_0}^{t_0+u-1} \|\xi(p(t))\varepsilon\|_1$ for some other constant $\bar{c}$. The terms that we are not taking in consideration in this summation are $O(\varepsilon^2)$ order terms plus the possible $O(\varepsilon)$ shifting of probabilites in the wiggling. The $O(\varepsilon^2)$ terms are negligible since in the allowed time period can only add up to $0(\varepsilon^{0.95})$. Furthermore, the shifting probability steps on each coordinate constitute a random walk with zero drift and step size at most $\varepsilon/2$. The expectation of the translation of such a random walk is equal to $O(\varepsilon u^{1/2}) = 0(\varepsilon^{0.475})$, which is again dominated by $\rho$. Finally, we can use Cauchy-Schwartz to lower bound $\sum_{t=t_0}^{t_0+u-1} \|\xi(p(t))\|_1^2$. We have that

$$(\sum_{t=t_0}^{t_0+u-1} \varepsilon^2)(\sum_{t=t_0}^{t_0+u-1} \|\xi(p(t))\|_1^2) \geq (\sum_{t=t_0}^{t_0+u-1} \|\xi(p(t))\varepsilon\|_1)^2 \geq \bar{c}^2\varepsilon^{0.8} \tag{3.51}$$

$$u\varepsilon^2(\sum_{t=t_0}^{t_0+u-1} \|\xi(p(t))\|_1^2) \geq \bar{c}^2\varepsilon^{0.8} \tag{3.52}$$

$$\sum_{t=t_0}^{t_0+u-1} \|\xi(p(t))\|_1^2/2n \geq c'\frac{\varepsilon^{-0.15}}{\log(1/\varepsilon)} \tag{3.53}$$

Finally, we can prove that we are escape from the attraction of unstable point $p_0$ by moving to a point with significantly lower potential.

$$\begin{aligned}
\mathbb{E}(\Psi(p(t_0+u-1)) - \Psi(p(t_0))) &= \mathbb{E}(\sum_{t=t_0}^{t_0+u-1} (\Psi(p(t+1)) - \Psi(p(t)))) && (3.54)\\
&= \sum_{t=t_0}^{t_0+u-1} \mathbb{E}(\Psi(p(t+1)) - \Psi(p(t))) && (3.55)\\
&= \sum_{t=t_0}^{t_0+u-1} (O(n\varepsilon^2) - \Omega\left(\frac{1}{n}\|\xi(p)\|_1^2\varepsilon\right) && (3.56)\\
&= O\left(\frac{\varepsilon}{\nu}\log(1/\varepsilon)\right) - \Omega\left(\sum_{t=t_0}^{t_0+u-1}\frac{1}{n}\|\xi(p)\|_1^2\varepsilon\right) && (3.57)\\
&= O(\varepsilon^{0.95}\log(1/\varepsilon)) - \Omega(\frac{\varepsilon^{0.85}}{\log(1/\varepsilon)}) && (3.58)
\end{aligned}$$

The foregoing discussion proves that if the state of the discrete-time process wanders sufficiently close to a mixed Nash equilibrium that is not $\nu$-weakly stable, then the potential function $\Psi$ will decrease by at least $\Omega(\varepsilon^{0.85}/\log(1/\varepsilon))$ in expectation during a period lasting no longer than $O(\varepsilon^{-1.05}\log(1/\varepsilon))$ steps. It remains for us to prove a similar statement in case the discrete-time process wanders sufficiently close to a fixed point $p^0$ of $\xi$ that is not even a $\nu$-Nash equilibrium. In that case there is some player $i$ and strategy $R \in \mathcal{S}_i$ such that $c^i(\bar{R}) - c^i(R) \geq \nu$ for all $\bar{R}$ such that $p^0_{i\bar{R}} > 0$. We use a logarithmic potential function to prove that the discrete-time process is repelled from $p^0$. Specifically, let $Y_t = \ln(p(t)_{iR})$. We wish to estimate $\mathbb{E}[Y_{t+1} - Y_t | p(t)]$. We have

$$p(t+1)_{iR} = \frac{(1-\varepsilon)^{\mathrm{cost}(R,t)} p_{iR}}{\sum_Q (1-\varepsilon)^{\mathrm{cost}(Q,t)} p_{iQ}} + N_{iR}(t).$$

Let $Z_t$ denote the denominator of the first term on the right side of this equation. Taking the logarithm of both sides and using the identity $\ln(x+y) > \ln(x) + \frac{y}{x+y}$, we obtain

$$\begin{aligned} Y_{t+1} &= \ln\left( \frac{(1-\varepsilon)^{\mathrm{cost}(R,t)} p_{iR}}{\sum_Q (1-\varepsilon)^{\mathrm{cost}(Q,t)} p_{iQ}} + N_{iR}(t) \right) \\ &> \ln(1-\varepsilon)\mathrm{cost}(R,t) + Y_t - \ln(Z_t) + \frac{N_{iR}(t)}{p(t+1)_{iR}}. \end{aligned}$$

We move $Y_t$ over to the left side and take the conditional expectation of both sides with respect to $p(t)$. For notational convenience, we omit the conditioning on $p(t)$ from all expectation operators.

$$\begin{aligned} \mathbb{E}[Y_{t+1} - Y_t] &> \ln(1-\varepsilon)\mathbb{E}[\mathrm{cost}(R,t)] - \mathbb{E}[\ln Z_t] + \mathbb{E}\left[\frac{N_{iR}(t)}{p(t+1)_{iR}}\right] \\ &\geq \ln(1-\varepsilon)\mathbb{E}[\mathrm{cost}(R,t)] - \mathbb{E}[\ln Z_t], \end{aligned}$$

where the second line follows because $\mathbb{E}[N_{iR}] \geq 0$ whenever $p_{iR}$ is sufficiently small. Now, to bound $\mathbb{E}[\ln Z_t]$ we use the following derivation that is familiar from the analysis of the Hedge algorithm.

$$\begin{aligned}\mathbb{E}[Z_t] &= \mathbb{E}\left[\sum_Q (1-\varepsilon)^{\mathrm{cost}(Q,t)} p_{iQ}\right] \\ &\leq \mathbb{E}\left[1 - \varepsilon \sum_Q \mathrm{cost}(Q,t) p_{iQ}\right] \\ &= 1 - \varepsilon \mathbb{E}\left[\sum_Q \mathrm{cost}(Q,t) p_{iQ}\right] \\ &= 1 - \varepsilon c^i(t),\end{aligned}$$

where $c^i(t)$ denotes the expected cost experienced by player $i$ on the path it randomly samples at time $t$. Taking the logarithm of both sides of the inequality above, and using Jensen's inequality,

$$\mathbb{E}[\ln Z_t] \leq \ln \mathbb{E}[Z_t] \leq \ln(1 - \varepsilon c^i(t)) \leq -\varepsilon c^i(t).$$

Plugging this bound on $\mathbb{E}[\ln Z_t]$ into the lower bound for $\mathbb{E}[Y_{t+1} - Y_t]$ from above, we obtain

$$\begin{aligned}\mathbb{E}[Y_{t+1} - Y_t] &\geq \ln(1-\varepsilon) c^i(R,t) + \varepsilon c^i(t) \\ &\geq \varepsilon(c^i(t) - c^i(R,t)) - O(\varepsilon^2) \\ &\geq \varepsilon\nu - O(\varepsilon^2),\end{aligned}$$

using the fact that $c^i(\bar{R},t) - c^i(R,t) \geq \nu$ for all $\bar{R}$ such that $p(t)_{i\bar{R}} > 0$.

Starting from any state in which $p_{iR}(t) < \varepsilon^2$, it must be the case that $p_{iR}(t+1) \geq \varepsilon^2$ because we mix in $\varepsilon^2$ times the uniform distribution after performing the multiplicative update, and when $p_{iR} = O(\varepsilon^2)$ it is not large enough to be eligible for the $\Omega(\varepsilon)$ wiggling. So we may assume without loss of generality that we start at time $t_0$ such that $Y_{t_0} \geq 2\ln(\varepsilon)$. The stochastic process $Y_t$ for $t = t_0, t_0+1, \ldots$ experiences a constant positive drift bounded below by $\frac{1}{3}\varepsilon\nu$, which persists as long as $c^i(t) - c^i(R,t) \geq \nu/2$. Since $c^i(t) - c^i(R,t)$ is a Lipschitz function of the mixed strategy profile $p$, we see that $p$ must traverse a distance of $\Omega(\nu)$ before the positive drift bounded below by $\Omega(\varepsilon\nu)$ ceases. It must cease after at most $O(\log(1/\varepsilon)/\varepsilon\nu)$ steps in expectation, because and $Y_{t_0} \geq 2\ln(\varepsilon)$ and $Y_t$ is never a positive number. Thus, as above, for some $u = O(\log(1/\varepsilon)/\varepsilon\nu)$ we have

$$\sum_{t=t_0}^{t_0+u-1} \|\hat{q}(t)\|_1 = \Omega(\nu).$$

We may argue, as above, that replacing $\sum \|\hat{q}(t)\|_1$ with $\sum \|\xi(p(t))\|_1$ introduces error terms of order $O(\varepsilon^{0.475})$, which is of lower order than the $\Omega(\nu)$ term on the right side. Now, using Cauchy-Schwartz,

$$(\sum_{t=t_0}^{t_0+u-1} \varepsilon^2)(\sum_{t=t_0}^{t_0+u-1} \|\xi(p(t))\|_1^2) \geq (\sum_{t=t_0}^{t_0+u-1} \|\xi(p(t))\varepsilon\|_1)^2 \geq c^2\varepsilon^{0.1} \tag{3.59}$$

$$u\varepsilon^2(\sum_{t=t_0}^{t_0+u-1} \|\xi(p(t))\|_1^2) \geq c^2\varepsilon^{0.1} \tag{3.60}$$

$$\sum_{t=t_0}^{t_0+u-1} \|\xi(p(t))\|_1^2/2n \geq c'\frac{\varepsilon^{-0.85}}{\log(1/\varepsilon)} \tag{3.61}$$

Finally, we can prove that we are escape from the attraction of unstable point $p_0$ by moving to a point with significantly lower potential.

$$\begin{aligned}
\mathbb{E}(\Psi(p(t_0+u-1)) - \Psi(p(t_0))) &= \mathbb{E}(\sum_{t=t_0}^{t_0+u-1} (\Psi(p(t+1)) - \Psi(p(t)))) && (3.62)\\
&= \sum_{t=t_0}^{t_0+u-1} \mathbb{E}(\Psi(p(t+1)) - \Psi(p(t))) && (3.63)\\
&= \sum_{t=t_0}^{t_0+u-1} (O(n\varepsilon^2) - \Omega\left(\frac{1}{n}\|\xi(p)\|_1^2\varepsilon\right) && (3.64)\\
&= O\left(\frac{\varepsilon}{\nu}\log(1/\varepsilon)\right) - \Omega\left(\sum_{t=t_0}^{t_0+u-1}\frac{1}{n}\|\xi(p)\|_1^2\varepsilon\right) && (3.65)\\
&= O(\varepsilon^{0.95}\log(1/\varepsilon)) - \Omega(\frac{\varepsilon^{0.15}}{\log(1/\varepsilon)}) && (3.66)
\end{aligned}$$

## 3.6 Standard Facts from Algebraic Geometry

In this section, we prove some standard facts from algebraic geometry that are needed in Section 3.4.

**Lemma 3.6.1** *If $I \subseteq \mathbb{R}[x_1, x_2, \ldots, x_n]$ then the variety $Z = V(I)$ is the union of finitely many subsets $Z_1, Z_2, \ldots, Z_m$, each of which is a nonsingular quasi-affine algebraic variety*[11]*, and therefore also a smooth manifold. For any polynomial $P \in I$ and any point $z \in Z_j$ $(1 \leq j \leq m)$, the gradient vector $\nabla P$ at $z$ is orthogonal to the entire tangent space $T_z Z_j$ of the manifold $Z_j$.*

**Proof:** To prove the existence of the decomposition $Z_1, Z_2, \ldots, Z_m$, we induct on the dimension of $Z$. Every algebraic variety is the union of finitely many irreducible varieties ([93], Theorem I.3.1.1), so it suffices to prove the statement when $Z$ is irreducible. A zero-dimensional irreducible variety is a point, so the base case is trivial. Assuming now that the theorem holds for all varieties of dimension less than $d$, let $Z$ be an arbitrary irreducible variety of dimension $d$. The nonsingular points of $Z$ form an open, hence quasi-affine, subset $Z_1$ ([93], Section II.1.4) and the singular points of $Z$ form a closed proper subvariety ([93], Section II.1.4), whose dimension is strictly smaller than $d$ because $Z$ is irreducible ([93], Theorem I.6.1.1). By the induction hypothesis, the set of singular points of $Z$ is the union of finitely many nonsingular quasi-affine algebraic varieties $Z_2, \ldots, Z_m$. This completes the induction.

A nonsingular quasi-affine algebraic variety over $\mathbb{R}$ is a smooth manifold ([93], Section II.2.3). For any $z \in Z_j$ $(1 \leq j \leq m)$ and any tangent vector $v \in T_z Z_j$,

[11] A quasi-affine algebraic variety is any variety isomorphic to an open subset of an affine algebraic variety

let $\gamma : [-1, 1] \to Z_j$ be any smooth parameterized curve in $Z_j$ such that $\gamma(0) = z, \gamma'(0) = v$. Letting $h(t) = P(\gamma(t))$, we have

$$h'(0) = \nabla P(\gamma(0)) \cdot \gamma'(0) = \nabla P(z) \cdot v.$$

But $h(t) = 0$ for all $t$ because $P$ vanishes on $Z_j$. Thus $h'(0) = 0$, which establishes that $\nabla P(z) \cdot v = 0$ as claimed.

We now present some definitions about tangent spaces and differentials, leading up to Lemma 3.6.2 below, which presents a rigorous proof of one of the steps in the proof of Theorem 3.4.2.

If $k$ is a field and $I \subset k[x_1, \ldots, x_n]$ is an ideal, then the affine algebraic variety $X = V(I)$ is the set of points $x \in k^n$ such that $P(x) = 0$ for all $P \in I$. (In algebraic geometry it is customary to denote $k^n$ by $\mathbb{A}^n_k$ and to call if *affine $n$-space over $k$.*) The *coordinate ring* $k[X]$ is the quotient ring $k[x_1, \ldots, x_n]/I$. Note that every $P \in k[X]$ determines a well-defined function on $X$, because if $P \equiv Q \pmod{I}$ then $P(x) = Q(x)$ for all $x \in X$. Functions from $X$ to $k$ defined in this way are called *regular functions* on $X$.

An $m$-tuple of $n$-variate polynomials $f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)$ collectively determine a *regular map* $f : \mathbb{A}^n_k \to \mathbb{A}^m_k$ that sends every point $(x_1, \ldots, x_n)$ to the point $(f_1(\vec{x}), \ldots, f_m(\vec{x}))$. Note that a regular map uniquely determines a homomorphism from $k[y_1, \ldots, y_m]$ to $k[x_1, \ldots, x_n]$ by mapping the polynomial $y_i$ to $f_i$ for each $i$. If $X = V(I), Y = V(J)$ are two affine algebraic varieties in $\mathbb{A}^n_k, \mathbb{A}^m_k$, respectively, a *regular map* from $X$ to $Y$ is a mapping from the points of $X$ to the points of $Y$ obtained by restricting a regular map $f : \mathbb{A}^n_k \to \mathbb{A}^m_k$ to $X$. A regular map from $X$ to $Y$ uniquely determines a homomorphism from $k[Y]$ to $k[X]$.

If $p \in X = V(I) \subseteq \mathbb{A}_k^n$, then the degree-1 polynomials $x_1 - p_1, \ldots, x_n - p_n$ generate a maximal ideal of $k[X]$ denoted by $\mathfrak{m}_p$. Every function in $k[X]$ that vanishes at $p$ belongs to $\mathfrak{m}_p$. If $\mathfrak{m}_p^2$ denotes the ideal generated by all products of pairs of elements of $\mathfrak{m}_p$ (i.e., all the regular functions on $X$ that vanish to second order at $p$) then the quotient $\mathfrak{m}_p/\mathfrak{m}_p^2$ is a $k$-vector-space called the *cotangent space of $X$ at $p$*. The dual vector space $T_pX = (\mathfrak{m}_p/\mathfrak{m}_p^2)^*$ is called the *tangent space of $X$ at $p$*.

If $f : X \to Y$ is a regular map and $f(p) = q$, then the induced homomorphism $h : k[Y] \to k[X]$ sends $\mathfrak{m}_q$ to a subset of $\mathfrak{m}_p$, hence it induces a well-defined linear transformation from $\mathfrak{m}_q/\mathfrak{m}_q^2$ to $\mathfrak{m}_p/\mathfrak{m}_p^2$. The dual of this linear transformation is denoted by $d_pf : T_pX \to T_qY$.

It's useful to consider how these ideas play out in the case of the projection map that sends $\mathbb{A}_k^{r+s}$ to $\mathbb{A}_k^s$ by mapping an $(r+s)$-tuple to its last $s$ coordinates. Denoting the coordinate rings of $\mathbb{A}_k^{r+s}$ and $\mathbb{A}_k^s$ by $k[x_1, \ldots, x_r, y_1, \ldots, y_s] = k[\vec{x}, \vec{y}]$ and $k[y_1, \ldots, y_s] = k[\vec{y}]$, respectively, then the projection map induces the homomorphism $k[\vec{y}] \to k[\vec{x}, \vec{y}]$ that simply includes $k[\vec{y}]$ as a subring of $k[\vec{x}, \vec{y}]$. If $X = V(I) \subseteq \mathbb{A}_k^{r+s}$ is an affine algebraic variety, then the regular map $f$ defined by the function composition $X \hookrightarrow \mathbb{A}_k^{r+s} \to \mathbb{A}_k^s$ induces the ring homomorphism $k[\vec{y}] \to k[\vec{x}, \vec{y}]/I$ that maps a polynomial in the variables $y_1, \ldots, y_s$ to its equivalence class modulo $I$.

If $a = (b, c)$ is a point of $\mathbb{A}_k^{r+s}$ then a polynomial $P \in k[\vec{x}, \vec{y}]$ belongs to $\mathfrak{m}_a$ if and only if it vanishes at $a$, and $P$ belongs to $\mathfrak{m}_a^2$ if and only if its Taylor expansion at $a$ has vanishing degree-0 and degree-1 terms, i.e. $P(a) = 0$ and $\nabla P(a) = 0$. If $X = V(I) \subseteq \mathbb{A}_k^{r+s}$ and $a = (b, c)$ is a point of $X$, then we can also consider $\mathfrak{m}_a$ and $\mathfrak{m}_a^2$ as ideals in $k[X]$ — unfortunately, the notation doesn't distinguish between these two meanings. In the case of $\mathfrak{m}_a$ the distinction is insignificant: an element of $\mathfrak{m}_a$ in $k[X]$ is an equivalence class (modulo $I$) of polynomials that vanish at $a$. But with $\mathfrak{m}_a^2$ we need to be more careful: a polynomial $P$ represents an element of $\mathfrak{m}_a^2$ in $k[X]$ if and only if it is equivalent (modulo $I$) to another polynomial $Q$ that satisfies $Q(a) = 0$ and $\nabla Q(a) = 0$.

So, for example, suppose that the gradient $\nabla F(a)$ of a polynomial $F \in I$ at a point $a = (b, c) \in X$ is a vector of the form $(0, 0, \ldots, 0, w_1, \ldots, w_s)$ whose first $r$ components are 0 and whose last $s$ components constitute a nonzero vector $w$. Let $P$ denote the degree-1 polynomial $\sum_{j=1}^{s} w_j(y_j - c_j) \in k[y_1, \ldots, y_s]$. Then $P \in \mathfrak{m}_a^2$ because $P$ is congruent, modulo $I$, to the polynomial $Q = P - F$, and $Q$ satisfies $Q(a) = 0$ and

$$\nabla Q(a) = \nabla P(a) - \nabla F(a) = (0, \vec{w}) - (0, \vec{w}) = 0.$$

Thus, $P$ is a nonzero element of $\mathfrak{m}_c/\mathfrak{m}_c^2$ that maps to zero in $\mathfrak{m}_a/\mathfrak{m}_a^2$; i.e. the induced map of cotangent spaces $\mathfrak{m}_c/\mathfrak{m}_c^2 \to \mathfrak{m}_a/\mathfrak{m}_a^2$ is not one-to-one. Dualizing this statement, we conclude that the differential $d_a f$ is not surjective, i.e. its rank is strictly less than $s$. Thus we have established the following lemma.

**Lemma 3.6.2** *Suppose $X = V(I) \subseteq \mathbb{A}_k^{r+s}$ is an affine algebraic variety and $f : X \to \mathbb{A}_k^s$ is the composition of the inclusion and projection maps $X \hookrightarrow \mathbb{A}_k^{r+s} \to \mathbb{A}_k^s$. If $F \in k[x_1, \ldots, x_r, y_1, \ldots, y_s]$ is a polynomial in $I$ whose gradient $\nabla F(a)$ at $a$ is a vector of the form $(0, \vec{w})$ for some nonzero $w \in k^s$, then the differential $d_a f : T_a X \to T_c \mathbb{A}_k^s$ is a linear transformation of rank strictly less than $s$.*

CHAPTER 4

# LOAD BALANCING WITHOUT REGRET IN THE BILLBOARD MODEL

## 4.1 Introduction

In the previous chapter, we studied the performance of the multiplicative weights learning algorithm in atomic congestion games. We showed that in almost all such games, the multiplicative-weights learning algorithm results in convergence to pure equilibria. Our setting was the standard full information one, where all players have access to an accurate model of the underlying game. In this chapter, we will be shifting away from the full information model and moving closer to the reality of distributed systems.

Traditional learning theory assumes that after playing a round of a game, each player can discover the cost of each possible strategy they could have used given the actions of their opponents. This is a reasonable assumption in games with infinitesimally small players, when actions of a single player have (essentially) no effect on the system. It is also reasonable when the underlying game is well-defined and common knowledge amongst all players. In distributed systems, however, this assumption is rather unnatural. Indeed, different subsystems only need to share some common functionality, whereas their inner workings can vary widely and even be updated seamlessly, and every single process can have significant impact on the behavior of the system.

Here, we take a significant additional step towards modeling distributed systems, by moving away from the standard full information setting. We consider load balancing in the so-called "bulletin board model" (similar to the ones in [10, 72]). In this model players can find out the delay on all machines, but do not have information on what their experienced delay would have been if they had selected another machine. Namely, players can query the current state of the system but cannot reliably predict the effect of their actions on it.

Although this change in the players' information might appear benign at first glance, it significantly alters the system behavior. Most importantly, the system becomes symmetric because all players observe the same feedback signal and respond to it using identical algorithms. Thus, at any point in time the players will sample their strategies from identical distributions, and our analysis only needs to focus on how this distribution evolves over time. This is quite different from our analysis of the full information setting in [58], which focused on the *symmetry-breaking* that inevitably occurs when atomic players use the Hedge algorithm in that setting. The symmetric setting that we study here allows for a significantly simpler analysis, incorporating techniques that are standard in the analysis of multiplicative-weight algorithms in learning theory (such as the use of KL-divergence as a potential function) as well as some new techniques specific to our setting (such as the martingale argument used to analyze the random walk in Lemma 4.2.7). Another benefit of this analysis, in addition to its simplicity, is the considerably better dependence of the convergence time on the number of players and congestible resources.

**Our Results and Techniques** We show that a natural and simple multiplicative weights algorithm achieves exponential improvement over the worst correlated equilibrium, for a class of load-balancing games. Our main result is that using the Hedge algorithm [43] in the bulletin board model, the expected makespan of the outcome is bounded by $O(\log n)$, exponentially better than the known lower bounds for generic no-regret algorithms. We also show that Hedge continues to satisfy the no-regret property even in the bulletin board model.

We utilize KL-divergence to express the distance between the mixed strategy employed by a player at time $t$ and her projected strategy at the symmetric Nash equilibrium of the non-atomic version of the game. We show that when this distance is large enough, then it has a tendency to shrink. As a result we can predict the evolution of the system by analyzing a random walk, that has negative drift only when we are far away from the origin, an analysis that is of independent interest.

**Prior work** The theory of learning in games has a long history; see [44] for an extensive exposition of the literature in this field, which has primarily focused on analyzing the convergence behavior of various classes of learning processes and relating this behavior to Nash equilibrium, correlated equilibrium, and their refinements. See [18] for a more recent survey. The relationship between regret minimization, calibrated forecasting, and correlated equilibrium has been studied by [39, 38], and the connection between these topics and the price of anarchy was first made in [15, 16]. Whereas these papers use regret bounds to establish *static* equilibrium properties of the limiting distribution of play, our work requires directly analyzing the *dynamics* of the stochastic process induced by these algorithms.

There has been considerable research in algorithmic game theory on understanding the behavior of adaptive procedures in load-balancing games and other congestion games, including best-response dynamics [46] and replication protocols [33]. These simple distributed protocols are well motivated, but they lack desirable learning-theoretic properties such as the no-regret property. An exception is [35], which analyzes a continuous-time process in non-atomic congestion games that can be regarded as the continuum limit of the multiplicative-weights learning process studied here. The shift from atomic to non-atomic congestion games eliminates the distinction between the solution quality of correlated, mixed Nash, and pure Nash equilibria, thus eliminating the motivating question in our work while also evading most of the technical difficulties we address in analyzing the discrete-time process in atomic congestion games. In the context of atomic congestion games, Roughgarden [87] has recently shown that for a wide class of games, including congestion games that satisfy a natural smoothness condition, bounds on the price of anarchy automatically extend to the total price of anarchy, when the global quality is defined to be the average cost.

In the previous chapter we introduced the study of the multiplicative weights learning algorithm in atomic congestion games. Our setting was the standard full information one, where all players have access to an accurate model of the underlying game. We show that in almost all such games, the multiplicative-weights learning algorithm results in convergence to pure equilibria. As discussed earlier, shifting from the full information setting to the more realistic bulletin board model invalidates the results of [58]; in particular this shift falsifies the prediction of convergence to pure equilibria and necessitates an analysis of the dynamics using completely different tools.

## 4.2 System Analysis

In this section we study the performance of learning algorithms in load-balancing games, i.e. congestion games on parallel links using the "bulletin board model" in which players assess edge costs according to the actual cost incurred on that edge, and not the hypothetical cost if the player had used it. We demonstrate that using the Hedge algorithm in the "bulletin board model" the process remains close to the symmetric fully mixed equilibrium of the non-atomic version of the game. As a result, its performance is exponentially better than the worst correlated equilibrium of the game.

In this section we first present the definition of the games we will be focusing on (section 4.2.1). Next, we introduce the multiplicative updates algorithm and the bulletin board model in section 4.2.2, where we prove that the no-regret property persists in the bulletin board model. The main part of the analysis is in section 4.2.3, while we defer a few technical lemmas to section 4.2.4.

### 4.2.1 Defining the Game and the Social Cost

The congestion game we consider in this section is an atomic congestion game with a set of $n$ players, each having weight $w_i = 1/n$, and $n$ edges with cost functions $c_e(x)$. In each period $t = 1, 2, \ldots$, each player chooses one edge $e$. We define $f_t(e)$ to be the total amount of flow on edge $e$ in period $t$, i.e. $f_t(e) = j/n$ where $j$ is the number of players choosing $e$ in period $t$. We make the following standing assumptions: for the edge $e$, the function $c_e(x)$ is twice continuously differentiable, satisfies $c_e(0) = 0$ and $c_e(1) \leq 1$, and for some positive constants $A, B$ it satisfies $c_e'(x) \geq A$ and $0 \leq c_e''(x) \leq B$ for all $x \in [0, 1]$. In section 4.2.4, lemma 4.2.8 proves that these hypotheses imply the following inequalities for all $x \in [0, 1]$:

$$Ax \leq c_e(x) \leq (B + 1)x \tag{4.1}$$

As a measure of social cost, we adopt the maximum edge cost, $\max_e c_e[f_e(t)]$. Interpreting players as jobs and edges as machines, this interpretation of the social cost is equivalent to the makespan. The inequality $Ax \leq c_e(x) \leq (B+1)x$ implies that for any flow vector $f$ the social cost $\max_e c_e(f_e)$ lies between $A\|f\|_\infty$ and $(B + 1)\|f\|_\infty$. In particular, the social optimum is $\Theta(1/n)$. As we have mentioned in the introduction, even for the extremely simple case in which $c_e(x) = x$ for all $e, x$ — i.e., a load-balancing game in which players schedule $n$ jobs on $n$ machines, and the cost experienced by player $i$ is proportional to the number of jobs on its machine — the correlated equilibria of the game can be exponentially worse than any Nash equilibrium.

### 4.2.2 The Learning Algorithm and the Bulletin Board Model

To define the learning algorithm used by each player, we let $\varepsilon$ be a small positive number (we'll need to have $\varepsilon \leq 1/n^3$ for the analysis) and we introduce the following notations.

$$c_e[t] = c_e(f_t(e)), \quad c_e[1:t] = \sum_{r=1}^{t} c_e[r]$$

$$Z(t) = \sum_{e \in E} \exp(-\varepsilon c_e[1:t-1]).$$

In period $t$, each player samples a random edge $e$ with probability

$$P(e, t) = \frac{\exp(-\varepsilon c_e[1:t-1])}{Z(t)}, \tag{4.2}$$

i.e., to obtain $P(e,t)$ from $P(e,t-1)$ we multiply it by $\exp(-\varepsilon c_e[t-1])$ and then renormalize all probabilities so that they sum to 1. At the first time step, the algorithm samples an edge uniformly at random. This algorithm for specifying a mixed strategy in period $t$ is a version of the Hedge algorithm [43], modified so that players assess edge costs according to the actual cost $c_e[t-1]$ incurred on that edge, and not the hypothetical cost $c_e(f_{t-1}(e)+1)$ if the player had used it for players that do not use the edge in this iteration. This model is usually referred to as the bulletin board model. Using the well-known fact that Hedge itself is a no-regret learning algorithm[1] first we prove that the bulletin board variant of Hedge is also a no-regret learning algorithm.

**Proposition 4.2.1** *The bulletin board variant of Hedge in any load-balancing game with non-decreasing cost functions retains the $\epsilon$-regret property.*

**Proof:** Hedge is known to have the $\epsilon$-regret property even in settings when the cost functions of the edges can vary with time[43]. For the proof, let us consider such a setting, where the actual cost/latency of each edge at period $t$ as $c_e^t(x_e^t)$, where $x_e^t$ is the load of the edge in question at period $t$. Naturally, all cost functions $c_e^t$ are non-decreasing functions of $x_e$. Now, we will define a new cost function $C_e^t$ as follows:

$$C_e^t(x) = \begin{cases} c_e^t(x) & \text{if } x \leq x_e^t \\ c_e^t(x_e^t) & \text{otherwise.} \end{cases}$$

Let us examine what this new cost function expresses. Under these cost functions, the latency of any edge observed at time $t$ is actually the worst possible and any further increase on the load of any edge would have no effect on its latency. If this

[1]provided that $\varepsilon$ converges to zero at an appropriate rate depending on $t$, e.g. $\varepsilon(t) = O(1/n^3\sqrt{t})$

optimistic view of the cost of the edges were actually true, then the algorithm we have proposed would perform exactly as the Hedge algorithm. Hedge is known to have the no-regret property, hence, the expected performance of the algorithm as $t$ goes to infinity is roughly as good as that of the best edge/strategy in hindsight under this modified costs $C$. However, the actual cost of any strategy under the real cost functions $c$, when taking into account the effect of the deviating player, would be at least as bad as that under the optimistic costs $C$. As a result the performance of our algorithm is also of $\epsilon$-regret in regards to the best strategy in hindsight under the true cost evaluations. □

Although the proposition above in its current form will suffice for our purposes, it can be straightforwardly extended to any no-regret algorithm and all congestion games with non-decreasing cost functions.

### 4.2.3 Main Theorems

The main result of this section is the following bound on the distribution $P(t)$ determined by the Hedge algorithm (4.2).

**Theorem 4.2.2** *If all players sample their strategies at time $t$ using the distribution $P(t)$ determined by the Hedge algorithm (4.2), then there exist positive constants $\alpha, \beta_0$ such that for all times $t$ and all $\beta > \beta_0$ it holds with probability at least* $1 - \exp(-\alpha\beta)$ *that* $\max_e |P(e,t)| < 2\beta/n$.

Combining this theorem with Chernoff bounds leads to a price-of-anarchy type result — the long-run average social cost exceeds the social optimum by a factor of at most $O(\log n)$. More precisely:

**Corollary 4.2.3** *In the setting of Theorem 4.2.2, there exist constants $c_1, c_2$ such that for all $t$, with probability at least $1 - 1/n^{c_1}$, the flow $f_t$ sampled by the players satisfies*

$$\max_e c_e(f_t(e)) \le \frac{c_2 \log n}{n}.$$

The proof of Theorem 4.2.2 rests on analyzing a stochastic process $KL(t)$ defined as the KL-divergence between the Nash equilibrium and $P(t)$. Let $Q$ be the symmetric Nash equilibrium of the non-atomic congestion game (where all players play the same strategy) with edge set $E$ and cost functions $(c_e)_{e\in E}$. KL-divergence between $P$ and $Q$ is defined as

$$KL(t) = \sum_{j \in E} Q(j) \log \left(Q(j)/P(j,t)\right).$$

KL-divergence measures the distance[2] between the distributions $Q(j)$ and $P(j,t)$. It is zero if they are equal and positive otherwise. We will show that when this distance is large enough, then it has a tendency to shrink (Lemma 4.2.6). This reduces the analysis of $KL(t)$ to exploring the behavior of a kind of random walks, which face negative drift only when they are far away from the origin. Lemma 4.2.7 provides this analysis.

Theorem 4.2.2 will follow from proving an exponential tail bound for $KL(t)$.

**Theorem 4.2.4** *There exist positive constants $\alpha, \beta_0$ such that* $\Pr(KL(t) > \beta/n) < e^{-\alpha\beta}$ *for all $\beta > \beta_0$.*

We next sketch the proof of this tail bound. In all of the following arguments, "log" denotes the natural logarithm function. We'll need the following technical lemma.

[2]although it is not a true distance metric since it is not symmetric

**Lemma 4.2.5**

$$\log Z(t+1) - \log Z(t) \le (\exp(-\varepsilon) - 1) \sum_e P(e,t)c(e,t).$$

**Proof:** We will use the fact that if $0 \le y \le 1$, then $\exp(-\varepsilon y) \le 1 + y(\exp(-\varepsilon) - 1)$; this can be verified by checking that the left side is a convex function, the right side is a linear function, and the left and right sides are equal when $y$ is an endpoint of the interval $[0,1]$.

$$\begin{aligned}
\frac{Z(t+1)}{Z(t)} &= \frac{\sum_e e^{-\varepsilon c_e[1:t-1]} e^{-\varepsilon c_e[t]}}{Z(t)} \\
&\le \frac{\sum_e e^{-\varepsilon c_e[1:t-1]} \left[1 + c_e[t](e^{-\varepsilon} - 1)\right]}{Z(t)} \\
&= 1 + \left(e^{-\varepsilon} - 1\right) \sum_e P(e,t) c_e[t].
\end{aligned}$$

The lemma follows by taking the logarithm of both sides and using the identity $\log(1+y) \le y$. □

We denote the difference $KL(t+1) - KL(t)$ as $\Delta_t$.

**Lemma 4.2.6** *The stochastic process $KL(t)$ satisfies*

$$\mathbf{E}[\Delta_t \mid P(t)] \le -(AC\varepsilon/n)KL(t) + C\varepsilon/n^2. \tag{4.3}$$

*In particular, $KL(t)$ drifts to the left at a rate of $\Omega(\varepsilon/n^2)$ whenever it is greater than $2/(An)$.*

**Proof:** A simple calculation using equation (4.2) using Lemma 4.2.5 justifies the bound

$$\log\left(\frac{P(e,t)}{P(e,t+1)}\right) \le \varepsilon c_e[t] - \left(1 - e^{-\varepsilon}\right) \sum_{e' \in E} P(e',t) c_{e'}[t].$$

Taking a weighted average of the above inequalities, weighted by $Q(e)$, we obtain

$$\begin{aligned}\Delta_t &= \sum_e Q(e)\left(\log \frac{P(e,t)}{P(e,t+1)}\right)\\ &\leq \varepsilon \sum_e Q(e) c_e[t] - (1-\exp(-\varepsilon)) \sum_e P(e,t) c_e[t].\end{aligned}$$

Now, using $\bar{c}_e[t]$ to denote $\mathbf{E}[c_e[t] \mid P(t)]$ and using $c_e[\bar{f}(t)]$ denote $c_e(P(e,t)/n)$, we may take the conditional expectation of both sides and apply the identity $1-\exp(-\varepsilon) > \varepsilon - \frac{1}{2}\varepsilon^2$ to obtain:

$$\begin{aligned}\mathbf{E}\left[\Delta_t \mid P(t)\right] &\leq \varepsilon \sum_e [Q(e) - P(e,t)]\bar{c}_e[t] + \frac{\varepsilon^2}{2} \sum_e P(e,t)\bar{c}_e[t]\\ &\leq \varepsilon(Q-P(t)) \cdot c[\bar{f}(t)] + \varepsilon(Q-P(t)) \cdot (\bar{c}[t] - c[\bar{f}(t)]) + \frac{\varepsilon^2}{2}.\end{aligned}$$

We denote the usual convex potential function $\sum_e \int_0^{x_e} c_e(y)\,dy$ as $\Phi(\vec{x})$. As a result, we have for the first term above that

$$\varepsilon(Q-P(t)) \cdot \nabla\Phi(P(t)) \leq \varepsilon\left[\Phi(Q) - \Phi(P(t))\right] \leq -A\varepsilon\|P(t)-Q\|_2^2,$$

where the last inequality uses the fact the $Q$ minimizes $\Phi$, combined with our assumption that $c_e'(y) \geq A$ for all $y$. It is not hard to prove that for some constant $C$, the additional inequalities

$$\|P(t)-Q\|_2^2 \geq \frac{C}{n} KL(t), \tag{4.4}$$

$$\varepsilon(Q-P(t)) \cdot (\bar{c}[t] - c[\bar{f}(t)]) + \frac{1}{2}\varepsilon^2 \leq C\varepsilon/n^2 \tag{4.5}$$

hold (Lemmas 4.2.11 and 4.2.12 in section 4.2.4), implying that the stochastic process $KL(t)$ satisfies

$$\mathbf{E}[\Delta_t \,|\, P(t)] \leq -(AC\varepsilon/n)KL(t) + C\varepsilon/n^2, \tag{4.6}$$

as claimed. $\square$

Next we give the submartingale argument to show that the fact that $KL(t)$ has negative drift when its large implies that the probability of $KL(t) > \beta/n$ is exponentially small in $\beta$ as claimed by Theorem 4.2.4.

**Lemma 4.2.7** *Let $(Y_t)_{t\geq 0}$ be a random walk satisfying the following for some constant $M \geq 1$:*

**bounded differences:** $|Y_{t+1} - Y_t| \leq 1$;

**negative drift:** $\mathbf{E}(Y_{t+1} - Y_t \mid Y_t) \leq -1/M$ *whenever* $Y_t \geq M$.

*Then there exist constants $\alpha, \lambda_0$ such that for all $\lambda > \lambda_0$ and $t \geq 0$, we have* $\Pr(Y_t > \lambda M) < e^{-\alpha\lambda}$.

**Proof:** For $t \in \mathbb{N}, r \in \mathbb{R}_+$, let $\mathcal{E}(r,t)$ denote the event that $Y_t > M + r + 1$. For $0 \leq s \leq t$ let $\mathcal{E}(r,t,s)$ denote the event

$$\begin{aligned}\mathcal{E}(r,t,s) = \{Y_{s-1} \leq M\} \cap \{Y_s, Y_{s+1}, \ldots, Y_{t-1} > M\} \\ \cap \{Y_t > M + r + 1\}.\end{aligned}$$

Note that the events $\mathcal{E}(r,t,s)$ $(s = 0, 1, \ldots, t)$ are disjoint and their union is $\mathcal{E}(r,t)$. Our upper bound on $\Pr(\mathcal{E}(r,t))$ will be established by proving separate upper bounds on each probability $\Pr(\mathcal{E}(r,t,s))$.

To this end, for a specified value of $s$, define a random variable $q = \min\{i \geq s \mid Y_i \leq M\}$ and a stochastic process

$$Z_i = \begin{cases} MY_i + i & \text{if } s \leq i \leq q \\ MY_q + q & \text{if } i > q. \end{cases}$$

Our negative drift assumption for the stochastic process $(Y_i)_{i\geq 0}$ implies that the process $(Z_i)_{i\geq 0}$ is a supermartingale:

$$\mathbf{E}\left[Z_i \mid Z_s, \ldots, Z_{i-1}\right] \leq Z_{i-1}.$$

Also, the bound $|Z_{i+1} - Z_i| \leq M + 1$ holds with probability 1. Applying Azuma's supermartingale inequality, for every $\gamma > 0$ we have

$$\Pr\left(Z_t - Z_s > \gamma\right) < \exp\left(-\frac{\gamma^2}{2(M+1)^2(t-s)}\right).$$

If event $\mathcal{E}(r, t, s)$ occurs, then we have

$$\begin{aligned}
Z_t = MY_t + t &> M(M + r + 1) + t \\
&= M^2 + Mr + M + t \\
Z_s = MY_s + s &\leq M(Y_{s-1} + 1) + s \\
&\leq M^2 + M + s \\
Z_t - Z_s &> Mr + (t - s).
\end{aligned}$$

Therefore,

$$\Pr(\mathcal{E}(r, t, s)) < \exp\left(-\frac{[Mr + (t-s)]^2}{2(M+1)^2(t-s)}\right).$$

Summing over $s$, we obtain

$$\Pr(\mathcal{E}(r, t)) < \sum_{s=0}^{t} \exp\left(-\frac{[Mr + (t-s)]^2}{2(M+1)^2(t-s)}\right).$$

Let $k = \lfloor r(M+1)/2 \rfloor$, and break up the sum into terms in which $t - s \leq k$ and those in which $t - s > k$.

$$\begin{aligned}
\Pr(\mathcal{E}(r,t)) &< \sum_{u=0}^{k} \exp\left(-\frac{[Mr+u]^2}{2(M+1)^2 u}\right) \\
&\quad + \sum_{u=k+1}^{\infty} \exp\left(-\frac{[Mr+u]^2}{2(M+1)^2 u}\right) \\
&< \sum_{u=0}^{k} \exp\left(-\frac{M^2 r^2}{2(M+1)^2 k}\right) \\
&\quad + \sum_{u=k+1}^{\infty} \exp\left(-\frac{u^2}{2(M+1)^2 u}\right) \\
&< (k+1)\exp\left(-\frac{r^2}{8k}\right) \\
&\quad + \int_k^{\infty} \exp\left(-\frac{x}{2(M+1)^2}\right) dx \\
&\leq (k+1)\exp\left(-\frac{r^2}{4r(M+1)}\right) \\
&\quad + 2(M+1)^2 \exp\left(-\frac{\frac{1}{2}r(M+1)-1}{2(M+1)}\right) \\
&\leq \left[1+\frac{r(M+1)}{2}+2(M+1)^2 e^{1/4}\right] e^{-\frac{r}{4(M+1)}}
\end{aligned}$$

For $r > e^{1/4}$ the last line implies that $\Pr(\mathcal{E}(r,t)) < [1 + 3r(M+1)^2]\, e^{-r/8M}$. By setting $r = \lambda M - M - 1 > e^{1/4}$ and $c = \frac{1}{8} + \frac{1}{8M}$, we obtain

$$\Pr(Y_t > \lambda M) < \left[1 + 3(\lambda M - M - 1)(M+1)^2\right] e^{-\frac{\lambda}{8}+c},$$

which shows that the lemma holds whenever $\alpha < 1/8$ and $\lambda_0$ is a sufficiently large constant depending on $\alpha$ and $M$. $\square$

**Proof of Theorems 4.2.2 and 4.2.4:**

Let $Y_t = KL(t)/\varepsilon$. We can show that for all $t \geq 0$, $|Y_{t+1} - Y_t| \leq 1$. Indeed, since $KL(t+1) - KL(t) = \sum_{j \in E} Q(j) \log\left(P(j,t)/P(j,t+1)\right)$ and each of the terms of the form $P(j,t)/P(j,t+1)$ lies in the $[e^{-\varepsilon}, e^{\varepsilon}]$ interval, $Y_t$ satisfies the property of bounded difference of Lemma 4.2.7.

We apply Lemma 4.2.7 with $M = (A+C)n^2/AC$. Moreover, the inequality $\mathbf{E}[\Delta_t \mid P(t)] \leq -(AC\varepsilon/n)KL(t) + C\varepsilon/n^2$ implies that there exist positive constants $\alpha, \beta_0$ such that $\Pr(KL(t) > \beta/n) < e^{-\alpha\beta}$ for all $\beta > \beta_0$. This proves Theorem 4.2.4. The bound on $\max_e |P(e,t)|$ in Theorem 4.2.2 now follows by combining the KL-divergence bound in Theorem 4.2.4 with Lemma 4.2.9 below, which bounds the infinity-norms of two distributions $P, Q$ in terms of their corresponding KL-divergence. $\square$

### 4.2.4 Technical Lemmas

The following technical lemmas complete the analysis the performance of Hedge:

**Lemma 4.2.8** *Let $c_e(x)$ be a function in $C^2([0,1])$ satisfying*

- $c_e(0) = 0, c_e(1) \leq 1$;
- *for all* $x \in [0,1]$, $c_e'(x) \geq A$;
- *for all* $x \in [0,1]$, $0 \leq c_e''(x) \leq B$.

*Then* $Ax \leq c_e(x) \leq A(B+1)x$ *for all* $x \in [0,1]$.

**Proof:** For all $x$ we have $c_e(x) = \int_0^x c_e'(y)\,dy \geq \int_0^x A\,dy$, which establishes that $Ax \leq c_e(x)$. To establish the upper bound on $c_e(x)$, we first use the mean value theorem to deduce that there exists some $x \in [0,1]$ such that

$$c_e'(x) = \frac{c_e(1) - c_e(0)}{1-0} \leq 1.$$

If there exists $y \in [0,1]$ such that $c'_e(y) > B+1$, then a second application of the mean value theorem would imply the existence of $z \in [0,1]$ satisfying

$$|c''_e(z)| = \left| \frac{c'_e(y) - c'_e(x)}{y - x} \right| > B,$$

contradicting our hypothesis about $c_e$. Hence $c'_e(y) \leq B+1$ for all $y \in [0,1]$. Now, for all $x \in [0,1]$, $c_e(x) = \int_0^x c'_e(y)\,dy \leq \int_0^x B+1\,dy$, which establishes that $c_e(x) \leq (B+1)x$. $\square$

**Lemma 4.2.9** *If $P, Q$ are two probability distributions on a finite set $S$, satisfying $\|P\|_\infty \geq 2\|Q\|_\infty$, then $KL(Q;P) \geq \frac{\|P\|_\infty}{16}$.*

**Proof:** Let $s_0$ be a point at which $P(s_0) = \|P\|_\infty$. Let $a = Q(s_0), b = P(s_0)$. Then

$$\begin{aligned} KL(Q;P) &= Q(s_0)\log\left(\frac{Q(s_0)}{P(s_0)}\right) + \sum_{s \neq s_0} Q(s)\log\left(\frac{Q(s)}{P(s)}\right) \\ &= a\log\left(\frac{a}{b}\right) + (1-a)\sum_{s\neq s_0} \frac{Q(s)}{1-a}\left[-\log\frac{P(s)}{Q(s)}\right] \end{aligned}$$

Since $\sum_{s\neq s_0} Q(s)/(1-a) = 1$, the sum on the right side can be interpreted as a weighted average of value of the convex function $-\log(x)$ at the points $P(s)/Q(s)$. Using Jensen's inequality, we see that this is greater than or equal to $-\log(x)$ evaluated at the point

$$x = \sum_{s\neq s_0}\left(\frac{Q(s)}{1-a}\right)\frac{P(s)}{Q(s)} = \sum_{s\neq s_0}\frac{P(s)}{1-a} = \frac{1-b}{1-a}.$$

Hence we have derived the first line of the following series of bounds.

$$\begin{aligned} KL(Q;P) &\geq a\log\left(\frac{a}{b}\right) + (1-a)\log\left(\frac{1-a}{1-b}\right) \\ &= \int_a^b \frac{x-a}{x(1-x)}\,dx \end{aligned}$$

The integrand is a strictly increasing function of $x$ for $0 < x < 1$, so letting $c = (a+b)/2$ we have

$$\begin{aligned}
\int_a^b \frac{x-a}{x(1-x)}\,dx &\geq \int_c^b \frac{x-a}{x(1-x)}\,dx \\
&\geq (b-c)\frac{c-a}{c(1-c)} = \frac{1}{4}\frac{(b-a)^2}{c(1-c)} \\
&\geq \frac{(b-a)^2}{4b}.
\end{aligned}$$

The assumption that $\|P\|_\infty \geq 2\|Q\|_\infty$ implies $a \leq b/2$, and the lemma follows immediately. $\square$

**Lemma 4.2.10** *In a non-atomic load balancing game*[3] *with $n$ edges whose cost functions satisfy the conditions of Lemma 4.2.8, the Nash equilibrium $Q$ satisfies for every edge $e$:*

$$\frac{A}{(B+1)n} \leq Q(e) \leq \frac{B+1}{An}$$

**Proof:** Since $Q$ is a Nash equilibrium, there exists[4] a $z > 0$ such that $c_e(Q(e)) = z$ for all $e$. Since there is some $e_0$ such that $Q(e_0) \leq 1/n$, we have $z \leq (B+1)/n$. Now for any edge $e$, the relations $AQ(e) \leq c_e(Q(e))$ and $c_e(Q(e)) = z \leq (B+1)/n$ together imply that $Q(e) \leq (B+1)/(An)$. Similarly, the existence of an edge $e_1$ such that $Q(e_1) \geq 1/n$ implies that $z \geq A/n$ from which it follows that $Q(e) \geq A/((B+1)n)$ for all $e$. $\square$

[3] i.e. a non-atomic parallel-links congestion game

[4] Since $c_e(0) = 0$ for all $e$, in the symmetric Nash $Q$ of the non-atomic congestion game we have that $Q(e), c_e(Q(e) > 0$ for all $e$. Otherwise, players could decrease their expected latency by utilizing the empty resource.

**Lemma 4.2.11** *For any distributions $P, Q$ on an $n$-element set $S$, if $C/n \leq Q(s) \leq 1/2$ for all $s \in S$, then*

$$\|P - Q\|_2^2 \geq \frac{C}{n} KL(Q; P).$$

**Proof:** Let $x(s) = P(s) - Q(s)$. We have

$$\begin{aligned} KL(Q;P) &= -\sum_s Q(s) \log \frac{P(s)}{Q(s)} \\ &= -\sum_s Q(s) \log \left(1 + \frac{x(s)}{Q(s)}\right). \end{aligned}$$

Using the identity $\log(1 + x) \geq x - x^2$, valid for $-1/2 \leq x \leq 1$, we obtain

$$\begin{aligned} KL(Q;P) &\leq -\sum_s Q(s) \left[\frac{x(s)}{Q(s)} - \frac{x(s)^2}{Q(s)^2}\right] \\ &\leq \sum_s \frac{x(s)^2}{Q(s)} \leq \frac{n}{C}\|x\|_2^2, \end{aligned}$$

from which the lemma follows immediately. □

**Lemma 4.2.12** *Let $P$ be any probability distribution on edges and let $f = (f_e)_{e \in E}$ be the random flow vector obtained by letting $n$ players each sample an edge in $E$ according to $P$ and send $1/n$ units of flow on that edge. Let $\bar{\mathbf{c}}, \mathbf{c}$ denote the vectors*

$$\bar{\mathbf{c}}_e = \mathbf{E}(c_e(f_e)), \qquad \mathbf{c}_e = c_e(\mathbf{E}(f_e)) = c_e(P(e)),$$

*respectively. There is a constant $C$ such that*

$$\varepsilon(Q - P) \cdot (\bar{\mathbf{c}} - \mathbf{c}) + \frac{1}{2}\varepsilon^2 \leq \frac{C\varepsilon}{n^2}.$$

**Proof:** Let us fix our attention on one edge $e$ and let $x_0 = P(e)$. Taylor's theorem with remainder ensures that for all $x \in [0, 1]$,

$$c_e'(x_0)(x - x_0) \leq c_e(x) - c_e(x_0) \leq c_e'(x_0)(x - x_0) + \frac{B}{2}(x - x_0)^2$$

This holds, since $0 \leq c_e''(y) \leq B$ for all $y$. Plugging the random variable $f_e$ into this bound, we find that

$$\mathbf{c}_e \leq \bar{\mathbf{c}}_e \leq \mathbf{c}_e + c_e'(x_0)\mathbf{E}(f_e - x_0) + \frac{B}{2}\mathbf{E}((f_e - x_0)^2)$$

$$0 \leq \bar{\mathbf{c}}_e - \mathbf{c}_e \leq \frac{B}{2}\,\mathrm{Var}(f_e).$$

If $z_i$ $(i = 1, 2, \ldots, n)$ denotes a collection of independent Bernoulli random variables with $\Pr(z_i = 1) = P(e)$, then the random variable $f_e$ has the same distribution as $\frac{1}{n}\sum_{i=1}^{n} z_i$, so its variance is

$$\mathrm{Var}(f_e) = \frac{1}{n^2} \cdot n \cdot \mathrm{Var}(z_i) = \frac{P(e)(1 - P(e))}{n}.$$

To bound the dot product $(Q - P) \cdot (\bar{\mathbf{c}} - \mathbf{c})$ from above, we first note that when $Q(e) < P(e)$ we have $(Q(e) - P(e))(\bar{\mathbf{c}}_e - \mathbf{c}_e) \leq 0$. The remaining terms of the dot product according to lemma 4.2.10 satisfy $Q(e) - P(e) \leq Q(e) \leq (B + 1)/(An)$, and $P(e)(1 - P(e))/n \leq P(e)/n \leq Q(e)/n \leq (B + 1)/(An^2)$. Hence the dot product is bounded above by $\frac{B}{2}\sum_e \frac{B+1}{An} \cdot \frac{B+1}{An^2} = \frac{B}{2}\left(\frac{B+1}{An}\right)^2$. Recalling that $\varepsilon \leq 1/n^2$, we see that the inequality in the statement of the lemma is satisfied by setting $C = \frac{1}{2} + \frac{B}{2}\left(\frac{B+1}{A}\right)^2$. □

## 4.3 Summary

Given that online learning is quite thoroughly understood in the setting of a single learner [19], it is rather natural to hope for a thorough understanding of systems consisting of multiple learners, but the characterization of such systems in existing work is far from thorough. Several recent papers have pursued this direction in the context of no-regret learning [15, 16, 87], but their findings have been limited to games in which the no-regret property by itself suffices to establish bounds on the overall system performance. Our work establishes that in many cases of interest — and specifically in settings closely related to the reality of distributed systems — this optimistic view does not materialize. Two systems consisting of no-regret learners can exhibit huge performance differences. Nevertheless, our result is in essence a positive result. It shows that "natural" candidates (e.g. Hedge) of no-regret algorithms perform well. An interesting direction for future research is the question of how much we can extend the family of allowable no-regret algorithms while still allowing for strong provable performance bounds on the overall system behavior.

CHAPTER 5

# LEARNING INSPIRED EQUILIBRIA AND COMPUTATION

## 5.1 Introduction

One of the main goals of game theory is to predict the behavior of self-interested agents that interact with each other. The proposed solution concepts usually come in the form of equilibrium notions. Nash equilibrium constitutes the most predominant such concept. Defined as (possibly mixed) strategy profiles against which no profitable deviation exists, Nash equilibria are both natural and convincing as a solution concept. Furthermore, in his much celebrated paper [76], John Nash showed that Nash equilibria are universal, in the sense that every finite game exhibits at least one.

Within the CS community, however, it is widely accepted that a solution concept should also be efficiently computable[77]. The CS sentiment is accurately captured in Kamal Jains quote: *"If your laptop cant find it then neither can the market"*. In recent years there has been a lot of work on characterizing the computational complexity of finding Nash equilibria. Unfortunately, most results in this area tend to be negative [27],[24] and currently the attention of the research community is mostly focused on approximation algorithms and identifying tractable special cases. Although the importance of these current approaches is beyond doubt, if we take a step back we see that incorporating learning theory into this framework can lead us to new insights.

Looking back at the desirable properties of a solution concept, we see that universality and computational efficiency are well-defined, whereas what it means for a solution concept to be convincing is largely open to interpretation. Learning theory can actually provide us with an elegant solution to this predicament. The legitimacy of a solution concept can be characterized by its robustness to multiple learning dynamics. In fact, under this interpretation Nash (or even correlated) equilibria can be cast into doubt since there exist simple games in which most natural dynamics (or even their averages) tend to not to converge not to converge to any single point[88]. In such games any single-point behavioral solution is bound to fail.

A more promising approach would be to consider the set of all (coarse) correlated equilibria, since they are closely connected with the behavior of no-regret dynamics (see Chapter 2). Unfortunately, correlated equilibria offer no insights on the behavior of some rather common dynamics such as best-response dynamics. Sink equilibria[46] on the other hand are specifically tailored to capture the long-run behavior of best response dynamics, but are inconsistent with any other learning behavior (e.g. AMS-dynamics).

However, there exist set-valued fundamental solution concepts, whose predictions can be shown to be closely attuned to the actual behavior of several natural learning dynamics. CURB (Closed Under Rational Behavior) sets, introduced by Basu and Weibull [11] and CUBR (Closed Under Better Responses) sets [81], introduced by Ritzberger and Weibull, are both set-wise extensions of strict Nash equilibria. A CURB (CUBR) set is a cartesian product of pure strategy sets such that each player's component contains all best (weakly better replies) replies to itself given any product of mixed actions of his opponents with support within their CURB sets. A CURB (CUBR) set is said to be minimal if it does not contain any proper subset that is also a CURB (CUBR) set.

Several plausible learning processes eventually settle down in a minimal CURB set (see the works of Hurkens[56], Sanchirico[90] , Young[107] and Fudenberg and Levine[44]). Moreover, there exist learning procedures that eventually settle down in minimal CUBR sets but not necessarily to minimal CURB sets (see [56] and[57]). CUBR sets are asymptotically stable for the class of payoff-positive selection dynamics (which contains replicator dynamics as well as AMS-dynamics)[81]. Finally, (minimal) CURB sets can be computed in polynomial time in the case of extensive games[79] as well as two player canonical forms games [13].

These solution concepts are based on the assumption that each player's belief about the strategies of the other players are independent. However this assumption can be put into question. Indeed, the choices of all players depend on the common history of play. History, therefore may very well act as a correlation mechanism and self-interested players have an incentive to keep track of such correlations and exploit them. In the next section we will be discussing solution concepts which allow for correlated beliefs about opponents' play.

### 5.1.1 Strong CURB and CUBR Sets

In this chapter, we will look into learning inspired solution concepts such as *strong CURB (Closed Under Rational Behavior) and CUBR (Closed Under weakly Better Replies) sets*. A strong CURB (CUBR) set is a cartesian product of pure strategy sets such that each player's component contains all best (weakly better) responses to itself given any joint probability distributions of its opponents over the set. A strong CURB (CUBR) set is said to be minimal if it does not contain any proper subset that is also a strong CURB (CUBR) set. Such minimal sets exist in all finite games and are asymptotically stable for a great number of natural learning dynamics. Both strong CURB (CUBR) sets are refinements of the notion of the CURB set and every minimal CURB set is contained in a minimal strong CURB set as well as a minimal strong CUBR set, so stability properties for CURB sets (for any learning procedure) carry over to strong CURB (CUBR) sets as well. In finite games CUBR and strong CUBR sets coincide, but the definition of strong CURB sets will he helpful to provide a unified treatment of strong CURB and CUBR sets. Moreover, there exist learning procedures that eventually settle down in minimal strong CURB (CUBR) sets but not necessarily to minimal CURB sets (see [56] and[57]). Furthermore, we prove that we can also compute all minimal strong CURB (as well as CUBR) sets for any normal form game with a constant number of players in polynomial time. As a result, these concepts are shown to be *universal, robust to different learning dynamics and efficiently computable* for any normal form game with a constant number of players.

The algorithms for computing minimal strong CURB (CUBR) sets are based on the ideas presented in [13] for the computation of CURB sets in the case of two player normal form games. The strong CURB set solution concept has also appeared under the names of primitive formations (Harsanyi and Selten[50] for the case of finite extensive-form games) and in independent working papers by Klimm and Weibull[59] (as sCURB sets) and by Zapechelnyck [108](as strict curb sets). In [59], Klimm and Weibull also extend Benisch etal algorithm to one that finds strong CURB sets in finite games and runs in polynomial time if the number of players is a constant. Their algorithm involves an initialization phase that leads to better performance than the original Benisch algorithm when applied to random two player games, where the notions of CURB and strong CURB sets coincide.

## 5.2 Preliminaries

Let $G = \{N, S_1, S_2, \ldots, S_n, u_1, u_2, \ldots, u_n\}$ be a finite game with a set of players $N = \{1, 2, \ldots, n\}$. We also extend the player's utilities over probability distributions in the usual multilinear manner. Let $m = \max_i |S_i|$. Also, let $S = \Pi_{i=1}^n S_i$ and $S_{-i} = \Pi_{j \neq i} S_i$. For a given set $X$, we denote as $\Delta(X)$ the set of probability distribution over $X$. For a distribution $p \in \Delta(S)$, we denote as $p_i \in \Delta(S_i)$ the marginal distribution of player $i$ on $S_i$. Similarly, we denote as $p_{-i}$ the marginal of $\Delta(S_{-i})$. Specifically,

$$p_i(s_i) = \sum_{s_{-i} \in S_{-i}} p(s_i, s_{-i}) \qquad \text{for} \;\; s_i \in S_i$$

$$p_{-i}(s_{-i}) = \sum_{s_i \in S_i} p(s_i, s_{-i}) \qquad \text{for} \;\; s_{-i} \in S_{-i}$$

For $p \in \Delta(S)$, let $BR_i(p_{-i})$ denote the set of pure best replies to the distribution $p_{-i}$. Also, if $x \in \Delta(S_i)$, let $BTR_i(x, p_{-i})$ denote the set of pure weakly better replies to (possibly mixed) strategy $x$ of player $i$, given the distribution $p_{-i}$ of his opponents. Finally, if $F \subset \Delta(S)$, we denote as $BR_i(F) = \cup_{p_{-i} \in F_{-i}} BR_i(p_{-i})$ and $BR(F) = \Pi_{i=1}^n BR_i(F)$. Similarly, for $F \subset \Delta(S)$ and $X \subset S_i$, we denote as $BTR_i(X, F) = \cup_{x \in \Delta(X)} \cup_{p_{-i} \in F_{-i}} BTR_i(x, p_{-i})$ and $BTR(X_1, X_2, \ldots, X_n, F) = \Pi_{i=1}^n BTR_i(X_i, F)$.

**Definition:** A (non-empty) cartesian product $P = \Pi_{i=1}^n P_i \subset S$ is a CURB set iff $BR(\Pi_{i=1}^n \Delta(P_i)) \subset P$.

**Definition:** A (non-empty) cartesian product $P = \Pi_{i=1}^n P_i \subset S$ is a strong CURB set iff $BR(\Delta(\Pi_{i=1}^{n} P_i)) \subset P$.

**Definition:** A (non-empty) cartesian product $P = \Pi_{i=1}^{n} P_i \subset S$ is a CUBR set iff $BTR(P_1, P_2, \ldots, P_n, \Pi_{i=1}^{n} \Delta(P_i)) \subset P$.

**Definition:** A (non-empty) cartesian product $P = \Pi_{i=1}^{n} P_i \subset S$ is a strong CUBR set iff $BTR(P_1, P_2, \ldots, P_n, \Delta(\Pi_{i=1}^{n} P_i)) \subset P$.

## 5.3 Algorithms for Computing Strong Minimal CURB and CUBR Sets

The algorithms for computing minimal strong CURB (CUBR) sets are based on the observation that if a cartesian product $P = \Pi_{i=1}^{n} P_i \subset S$ is not a strong CURB (CUBR) set, then this can be verified efficiently by solving a linear feasibility problem (LFP). The main idea behind both these algorithms is to start from a random pure strategy profile $s_1 \times s_2 \times \cdots \times s_n$ and keep iteratively increasing the sets $P_i$, by adding the violating strategies $q_i$ until we reach a fixed point. The resulting set will be the minimal CURB (CUBR) set which contains the strategy profile $s_1 \times s_2 \times \cdots \times s_n$. From that point on, finding truly minimal (set inclusion-wise) CURB (CUBR) sets is trivial and can be accomplished even by pairwise comparison of all the $m^n$ resulting strong CURB (CUBR) sets. However, we will also discuss ways to expedite this process.

Since both algorithms for finding minimal strong CURB sets as well as minimal strong CUBR sets are rather similar to each other, we will focus on analyzing one of the two (i.e. the one for strong CUBR sets). We will also describe how this algorithm can be transformed into one that returns all minimal strong CURB sets.

We begin, by providing an alternate characterization of strong CUBR sets.

**Proposition 5.3.1** *A (non-empty) cartesian product $P = \Pi_{i=1}^{n} P_i \subset S$ is not a strong CUBR set iff there exist player $i \in N$ and strategies $p_i \in P_i$, $q_i \in S_i \setminus P_i$ and strategy distribution $r_{-i} \in \Delta(\Pi_{j \neq i} P_i)$ s.t. $u_i(q_i, r_{-i}) \geq u_i(p_i, r_{-i})$.*

**Proof:** A (non-empty) cartesian product $P = \Pi_{i=1}^{n} P_i \subset S$ is a strong CUBR set iff $BTR(P_1, P_2, \ldots, P_n, \Delta(\Pi_{i=1}^{n} P_i)) \subset P$. Equivalently, $P$ is a strong CUBR set iff for all players $i \in N$, $BTR(P_i, \Delta(\Pi_{i=1}^{n} P_i)) \subset P_i$.

Next, we will show that $BTR_i(X, F) = \cup_{x \in \Delta(X)} \cup_{p_{-i} \in F_{-i}} BTR_i(x, p_{-i})$ is equivalent to $\cup_{x \in X} \cup_{p_{-i} \in F_{-i}} BTR_i(x, p_{-i})$. Indeed, for any $p_{-i} \in F_{-i}$, there exists a pure strategy in $X$ such that it experiences minimal utility amongst all strategies $x \in \Delta(X)$. We denote that strategy as $w(X, p_{-i})$. Hence, $BTR_i(X, F) = \cup_{p_{-i} \in F_{-i}} \cup_{x \in \Delta(X)} BTR_i(x, p_{-i}) = \cup_{p_{-i} \in F_{-i}} BTR_i(w(X, p_{-i}), p_{-i}) = \cup_{p_{-i} \in F_{-i}} \cup_{x \in X} BTR_i(x, p_{-i}) = \cup_{x \in X} \cup_{p_{-i} \in F_{-i}} BTR_i(x, p_{-i})$.

As a result, $P = \Pi_{i=1}^{n} P_i \subset S$ is a strong CUBR set iff for all players $i \in N$, $\cup_{x \in P_i} \cup_{p_{-i} \in \Delta(P_{-i})} BTR_i(x, p_{-i}) \subset P_i$. However, the negation of this condition implies exactly that there exist strategy $p_i \in P_i$, and distribution $r_{-i} \in \Delta(\Pi_{j \neq i} P_i)$ and strategy $q_i \in S_i \setminus P_i$ s.t. $u_i(q_i, r_{-i}) \geq u_i(p_i, r_{-i})$.

$\square$

In fact, it suffices to check for strategy distributions $r_{-i} \in \Pi_{j \neq i} P_i$, which implies the equivalence between strong CUBR sets and CUBR sets, but for consistency between the treatments of strong CURB and CUBR sets we will utilize characterization 5.3.1.

Below follows the process for finding given a set $P = \Pi_{i=1}^{n} P_i$, the set of strategies $Q_i$ which can be cast as weakly better replies to some strategy $p_i \in P_i$, as a result of a possibly correlated probability distribution of his opponents actions over $\Pi_{j \neq i} P_i$.

**Input**: set $P_i$, set $S_i$, set $\Pi_{j \neq i} P_j$, $u_i$

**Output**: set $Q_i$ of weakly-better-replies

$Q_i = \emptyset$;

**foreach** *strategy* $q_i \in S_i \setminus P_i$ **do**

  **foreach** *strategy* $p_i \in P_i$ **do**

    **if** *exists* $r_{-i} \in \Delta(\Pi_{j \neq i} P_j)$ *such that* $u_i(q_i, r_{-i}) \geq u_i(p_i, r_{-i})$ **then**

      $Q_i = Q_i \cup q_i$;

    **end**

  **end**

**end**

**return** $Q_i$;

**Algorithm 1:** Procedure Find-Weakly-Better-Replies

We will show that this process runs in polynomial time in normal form games with a constant number of players. Indeed, the line exists $r_{-i} \in \Delta(\Pi_{j \neq i} P_j)$ such that $u_i(q_i, r_{-i}) \geq u_i(p_i, r_{-i})$, corresponds to a linear feasibility problem since both conditions are linear relations over $|\Pi_{j \neq i} P_j| \leq m^{n-1}$ many variables. Linear feasibility problems as easier than LPs (they are LPs with no objective) and can be solved in low-order polynomial time that we denote as LFP(). The two outer loops can lead at most to $m^2$ calls to the LFP problem, so the overall complexity of the algorithm is $O(m^2 LFP(m^{n-1}))$.

The main routine finds for each pure strategy profile the minimal strong CUBR that contains it. In order to do so it iteratively executes the procedure Find-Weakly-Better-Replies to examine whether its current cartesian product of pure strategies is closed under weakly better replies.

**Input**: outcome $s_1 \times s_2 \times \cdots \times s_n$, action sets $S_1, S_2, \ldots, S_n$, utilities $u_1, u_2, \ldots, u_n$

**Output**: Minimal CUBR set $\Pi_{i=1}^{n} P_i$ containing outcome $s_1 \times s_2 \times \cdots \times s_n$

**foreach** $i \in N$ **do**

  $P_i = \{s_i\}$;

**end**

converged=false;

**while** *(¬ converged)* **do**

  **foreach** $i \in N$ **do**

    $Q_i$=Find-Weakly-Better-Replies($P_i$, $S_i$, $\Pi_{j \neq i} P_j$, $u_i$);

    $P_i = P_i \cup Q_i$;

  **end**

  **if** $\cup_{i=1}^{n} Q_i = \emptyset$ **then**

    convergence=true;

  **end**

**end**

**return** $\Pi_{i=1}^{n} P_i$;

**Algorithm 2:** Finding Minimal Strong CUBR set containing outcome $s_1 \times s_2 \times \cdots \times s_n$

**Theorem 5.3.2** *Algorithm 2 terminates in time* $O(n^2 m^3 LFP(m^{n-1}))$.

**Proof:** The while loop will terminate after at most $nm$ iterations since, each time is called it either terminates or it increases the size of at least one set $P_i$. This can be repeated at most $nm$ times. Furthermore, each while loop performs $n$ calls to the Find-Weakly-Better-Replies procedures, which has running time of $O(m^2 LFP(m^{n-1}))$. Putting all these together we derive, the above running time.

$\square$

**Theorem 5.3.3** *The cartesian product returned by algorithm 2 is the minimal strong CUBR set that includes the outcome $s_1 \times s_2 \times \cdots \times s_n$.*

**Proof:** We will argue that the returned cartesian product $\Pi_{i=1}^{n} P_i$ is a strong CUBR set and does not contain any strong CUBR subset, which includes the outcome $s_1 \times s_2 \times \cdots \times s_n$.

We have already argued that the algorithm terminates. At that point, for all strategies $p_i \in P_i$, $q_i \in S_i \setminus P_i$ and strategy distribution $r_{-i} \in \Delta(\Pi_{j \neq i} P_i)$ we have that $u_i(q_i, r_{-i}) < u_i(p_i, r_{-i})$. By proposition 5.3.1, we derive that the returned cartesian product $\Pi_{i=1}^{n} P_i$ is a strong CUBR set.

We will prove that the returned product is indeed minimal by induction on the number of iterations of the while loop that it takes for the program to terminate.

**Base Case:** If the program terminates after one while loop, this implies that the starting strategy profile is a pure Nash eq and indeed the absolutely minimum strong CUBR set including the outcome $s_1 \times s_2 \times \cdots \times s_n$.

**Inductive step:** The strategies added in the $k$-th iteration of the loop cannot be removed without breaking the strong CUBR property. Indeed, all of these strategies are weakly better responses to some preexisting strategy $p_i$ in the current $P_i$, given some possibly correlated probability distribution of his opponents of the current $\Pi_{j \neq i} p_j$. Since, no strategies are removed from the set $P_i$, in future iterations of the while loop, all these strategies will remain necessary so as for our returned cartesian product be a strong CUBR. □

Using each strategy profile as a starting configuration, we find a super-set of all truly minimal strong CUBR sets, since not all pure strategy profiles are included in a minimal strong CUBR set, but every minimal CUBR set includes (at least) one pure strategy profile. By pairwise comparison of these $m^n$, in terms of set-inclusion, we can discard all returned CURB sets for which there exist subsets within our set. All remaining strong CUBR sets are indeed minimal. Indeed, if that was not the case, there would exist a strategy profile within one surviving (but not minimal) strong CUBR set $C$, whose minimal strong CUBR that included it was indeed minimal and specifically a strict subset of strong CUBR set $C$. However, this would imply that the strong CUBR set $C$ would have already been discarded.

Although this algorithm is polynomial, when the number of players is considered a constant, one can improve its performance by reducing the number of calls to algorithm 2. Instead of trying all starting configurations, it suffices to examine a starting set of cartesian products of strategies with the property that each minimal strong CUBR contains at least one of them. The set of all minimal strong CURB sets has this property and as we will discuss can be computed in polynomial time. Another such starting set is the set of all minimal weak CURB sets[59]. Weak CURB sets sets are cartesian products of strategies such that they contain all best responses to any pure strategy profile within the weak CURB sets and can be computed in polynomial time [1], which is proposed by Klimm and Weibull [59] as a method to speed up the computation of minimal strong CURB sets.

Given the presentation of the algorithm for finding minimal strong CUBR sets,

[1] if the number of players is a constant

it is pretty straightforward to transform it into one that finds minimal strong CURB sets. Specifically, we merely need to substitute the call to procedure Find-Weakly-Better-Replies($P_i, S_i, \Pi_{j \neq i} P_j, u_i$) to a call to the following Find-Best-Reply($P_i, S_i, \Pi_{j \neq i} P_j, u_i$) procedure. This procedure given a set $P = \Pi_{i=1}^{n} P_i$, finds the set of strategies $Q_i$ with the following two properties: i) $Q_i \cap P_i = \emptyset$ ii) each strategy in $Q_i$ can be cast as best reply to some possibly correlated probability distribution of his opponents actions over $\Pi_{j \neq i} P_i$.

**Input**: set $P_i$, set $S_i$, set $\Pi_{j \neq i} P_j$, $u_i$

**Output**: set $Q_i$ of best-replies

$Q_i = \emptyset$;

**foreach** *strategy* $q_i \in S_i \setminus P_i$ **do**

  **if** *exists* $r_{-i} \in \Delta(\Pi_{j \neq i} P_j)$ *such that* $\forall$ *strategy* $s_i \in S_i$

  $u_i(q_i, r_{-i}) \geq u_i(s_i, r_{-i})$ **then**

    $Q_i = Q_i \cup q_i$;

  **end**

**end**

**return** $Q_i$;

**Algorithm 3:** Procedure Find-Best-Replies

The runtime and correctness analysis of this algorithm is directly analogous to the one for finding minimal strong CUBR sets.

CHAPTER 6

# NO REGRET LEARNING IN OLIGOPOLIES: COURNOT VS BERTRAND

## 6.1 Introduction

Oligopoly theory deals with the fundamental economic problem of competition between two or more firms. In this work we study the conditions under which an oligopoly arrives at stability. We focus on the two most notable models in oligopoly theory: Cournot oligopoly[25], and Bertrand oligopoly[14]. In the Cournot model, firms control their production level, which influences the *market price*. In the Bertrand model, firms choose the price to charge for a unit of product, which affects the *market demand*.

Competition among firms in an oligopolistic market is a setting of strategic interaction, and is therefore analyzed within a game theoretic framework. Cournot and Bertrand oligopolies are modeled as strategic games, with continuous action sets (either production levels or prices). In both models the revenues of a firm are the product of the firm's part of the market times the price; In addition, a firm incurs a production cost, which depends on its production level.

In the most simple oligopoly model, the firms play a single game, where they all take actions simultaneously. All the firms produce the same good; the demand for this product is a linear in the total production; the cost of production is fixed per unit of production. In this oligopolistic market, a Nash equilibrium in pure strategies exists in both Cournot and Bertrand models. Interestingly, despite the strong similarity between these models, the Nash equilibrium points are very different: in Bertrand oligopoly, Nash equilibrium drives prices to their competitive levels, that is, the price equals the cost of production, while in Cournot oligopoly, the price in the unique Nash equilibrium is strictly above its competitive level. Liu [64] showed that the uniqueness of equilibrium in the linear demand, linear cost model, carries on to correlated equilibrium. Yi [105] have extended Liu's work to the case of Cournot oligopoly where firms produce different products, that are strategic substitutes, and to the case of weakly convex production cost functions.

Equilibrium analysis alone, however, cannot capture the dynamic nature of markets. In the real world, trading is performed over long periods of time, which gives firms the chance to adjust their actions e.g, their prices or production levels. If we assume that the essential market attributes remain unchanged, then this situation gives rise to a repeated game, obtained by repeated play of the original simultaneous, one shot game.

One approach for analyzing the repeated oligopolistic game, is through studying the Nash equilibrium of the repeated game. This models a situation where the firms "commit" to a strategy, and their joint commitment forms an equilibrium (see [66], Chapter 12.D). In practice however, an important feature of an oligopolistic market is that different firms are not perfectly informed about different aspects of the market, e.g., the attributes of the other participants, and cannot pre-compute, or agree on a Nash equilibrium of the repeated game before they begin interacting.

A more pragmatic approach for studying such repeated interactions is through the analysis of adaptive behavior dynamics (see, [44, 106]). The goal here is to investigate the evolution of the repeated game, when the agents (firms) play in accordance to some "natural" rule of behavior. In the setting of an oligopolistic market, we would want a natural behavior to comply with "rationality" and hopefully give rise to some sort of profit maximization on the side of a firm. Another natural aspiration is that our behavior rules should be "distributed", which means that firms should be able to make their choices in each period based only on their own payoffs, and independently of other firms (in most markets, a firm cannot tell with certainty what are the payoffs, and costs of other firms). The central question, in such a setting, is whether the behavior dynamics finally converge, as this would imply long term stability of the market.

Dynamic behavior in Cournot and Bertrand oligopolies have been studied before. Cournot [25] considered the simple best response dynamics, where at every step of the repeated game, firms react to what happened in the market on the previous step. Cournot showed that in the case of a duopoly, the simultaneous best response dynamics converges to the unique Nash equilibrium of the one shot game, i.e., after sufficiently many steps the two firms will play their Nash equilibrium strategies on every subsequent step. However, this result does not generalize to an arbitrary number of firms, as shown by Theocharis [100].

Milgrom and Roberts [68, 69] were the first to explore connections between Cournot competition and super-modular games as a way to show convergence results for learning dynamics. In their work (as well as in followup papers [4, 101]) Cournot duopolies as well as specific models of Cournot oligopolies are shown to exhibit strategic complementarities. This identifying property of supermodular games is shown to imply convergence to Nash equilibrium for a specific class of learning dynamics, known as adaptive choice behavior. This class of learning dynamics encompasses best-response dynamics and Bayesian learning but is generally orthogonal to the class of dynamics that we will be focusing on.

In this work we are interested in dynamic behaviors where firms minimize their long term *regret*[1]. Regret compares the firm's average utility to that of the best fixed constant action (e.g., constant production level in Cournot, and constant price in Bertrand). Having no-regret means that no deviating action would significantly improve the firm's utility (see [20]). Several learning algorithms [109], [54],are known to offer such guarantees, as their average regret bounds are $o(T)$, where $T$ is the number of time steps.

Regret minimization procedures prescribe to some rather desirable requirements in regards to modeling market behavior. Firstly, they are rational, in the sense that an agent is given guarantees on her own utility regardless of how the other agents act. Moreover, they are distributed, since an agent needs to be aware only of her own utility. Many of the no regret procedures[42] are rather intuitive, as they share the idea that agents increase the probability of choosing actions that have been performing well in the past. Several learning procedures are known to be of no-regret, but more importantly, the assumption is not tied to any specific algorithmic procedure, but merely captures successful long-term behavior. Lastly, no-regret guarantees can be achieved even in the "multi-armed bandit" setting[7, 36, 1], where the input for the algorithm consists only of the payoffs received. This feature is important in the case that firms are not fully aware of the market structure (i.e., demand function), and are maybe even uncertain regarding their own production costs.

In the most relevant result to our work, Even-Dar *et al.*[31] study no regret dynamics in a class of games that includes Cournot competition with linear inverse market demand function, and convex costs functions. They show that the *average* production level of every firm, as well as it's *average* profits, converge to the ones in the unique Nash equilibrium of the one shot game.

[1]Regret is sometimes also referred to as external regret.

Table 6.1: Overview of results

| | Bertrand | Cournot with perfect substitutes | Cournot with product differentiation |
|---|---|---|---|
| Nash equilibria | Infinite, Unique prices, Unique profits | Unique | Unique |
| Correlated equilibria | **Infinite**[2] | Unique | Unique |
| No Regret | **Infinite, Different prices, Different profits** | **Infinite, Unique prices, Different profits** | **Unique** |

### 6.1.1 Our Results

In this work we examine the behavior of no regret dynamics in Cournot and Bertrand oligopoly models.

In the classic model of Bertrand oligopoly [14], it is well known that oligopolies with more than two firms exhibit several trivial Nash equilibria but in all of them the prices are equal to the marginal costs and all players make zero profit (Bertrand paradox). This phenomenon has been verified for correlated equilibria only for the case of a duopoly[2], where correlated equilibria are unique[68]. In our work, we show that under no-regret behavior the zero-profit postulate does not hold even in the case of two players. In fact, we show that not only does the market not

[2]In [104] it is claimed that the correlated equilibria of Bertrand games are unique under some special cases.

necessarily converge to zero profit outcomes, but that the players can actually enjoy significant profits. In summary, our main results for Bertrand oligopolies under no-regret have as follows:

1. The Bertrand paradox does not hold anymore; firms enjoy non-zero profits under no-regret behavior.
2. Moreover, the identified profits can be rather significant when the number of players is small (e.g. 17% of optimal profits in the case of a duopoly). Profits however, tend to go to zero quickly as the number of firms increases.

Our observations about no-regret behavior in Bertrand oligopolies agree to a large extent both with experimental work [29], as well as with empirical observations about real world oligopolistic markets [95].

The study of correlated equilibria [64, 105] as well as of no-regret dynamics in [31] in Cournot oligopolies, has been an area of interest in both economics as well as computer science. In our work, we analyze a model of Cournot equilibria, which is a strict generalization of all the previously examined models, under no-regret dynamics. In fact, our results can be extended to all dynamics, in which each player's average payoff dominates the one they would receive if they always deviated to their respective Nash equilibrium strategy. This is a strict generalization of no-regret dynamics, since no-regret dynamics must fare well against all fixed strategies. In a novel approach in this line of work, we consider the evolution of the market not only from the perspective of the firms (individual production levels, profits), but also from the consumers' perspective (aggregate production level, prices) which leads to new insights. As a result, we can prove a single unifying message for all models examined: the daily prices converge to their level at Nash equilibrium.

In summary, our main results for Cournot oligopolies under no-regret have as follows:

1. In Cournot oligopoly with linear inverse demand function, and weakly convex costs, when every firm experiences no-regret, the empirical distribution of the daily overall production level, as well as of the daily prices, converges to a single point that corresponds to the Nash equilibrium of the one shot game.
2. When the firms produce products that are not perfect substitutes i.e., when even the tiniest of *product differentiation* is introduced, the empirical distributions of all market characteristics including the daily production levels of every firm converge to their levels in Nash equilibrium.
3. Some product differentiation is necessary in order to alleviate the nondeterminism of the day-to-day behavior on the side of the firms.

Table 6.1 summarizes what is known about equilibrium, and no-regret in Cournot and Bertrand oligopolies.

## 6.2 Preliminaries

### 6.2.1 Models of Oligopoly

We formally define Cournot oligopoly, and Bertrand oligopoly, as strategic games, with continuous action space.

**Definition:** A Cournot oligopoly is a game between $n$ firms, where the strategy space $S_i$ of firm $i$ is the span of its production level $q_i$. Typically, $S_i$ is defined to be the interval $[0, \infty)$. The utility function for firm $i$ is $u_i(q_1, \ldots, q_n) = P_i(q_1, \ldots, q_n)q_i - c_i(q_i)$, where $P_i$ is the market inverse demand function for the good of firm $i$, which maps the vector of production levels to a market clearing price in $\mathbb{R}^+$.

Our focus is on the case of linear inverse demand function. The utility of a firm $i$ as a function of the firms' production levels is $u_i(q_1, \ldots, q_n) = (a - bQ)q_i - c_i(q_i)$, where $a$ and $b$ are positive constants, and $Q = \sum_i q_i$ denotes the total product supply. In Section 6.4 we consider an extension of Cournot oligopoly with perfect substitutes, to the case of product differentiation, where the price of firm $i$ depends in an asymmetric manner on his own production level, and the production levels of the other firms. In this case the market inverse demand function $P^i(q)$ is given by $P^i(q) = a_i - b_i q_i - b_i \gamma_i Q_{-i} = a_i - b_i(1 - \gamma_i)q_i - b_i \gamma_i Q$, where $\gamma_i$ denotes the degree of product differentiation between products, $0 < \gamma_i \leq 1$, $b_i > 0$.

**Definition:** A Bertrand oligopoly is a strategic game between $n$ firms, where the strategy space $P_i$ of firm $i$ is its declared price $p_i$, which lies in the interval of all possible prices $[0, \infty)$, and its utility function is
$u_i(p_1, \ldots, p_n) = D_i(p_1, \ldots, p_n)p_i - c_i(D_i(p_1, \ldots, p_n))$, where $D_i$ is the market demand function of firm $i$, that maps from the vector of firms prices to a demand in $\mathbb{R}^+$.

We consider Bertrand oligopoly with a linear demand function, in which the market demand is equally shared among the firms with the least price:

$$D_i(p_1, \ldots, p_n) = \begin{cases} 0 & p_i > p_j, \text{ for some } j \\ \frac{a-p_i}{b(m+1)} & p_i \le p_j \text{ for all } j, \text{ and } m = |\{j \ne i | p_j = p_i\}| \end{cases}$$

Intuitively, this means that the market demand goes down linearly as the minimal announced price increases. If the minimal price has been offered by more than one firms, these firms share the market demand equally.

### 6.2.2 Regret Minimization

We will merely apply the general definitions of Chapter 2 to our setting.

**The regret of a firm in a repeated oligopoly game:** Consider the case of $n$ firms that engage in a repeated Cournot (alternatively Bertrand) oligopoly game, and suppose that $\{x^t\}_{t=1}^{\infty}$ is a sequence of vectors, where $x^t$ represents the production levels (alternatively, prices), set by the firms at time $t$. The regret of firm $i$ at time $T$ is defined as $\mathcal{R}_i(T) = \max_{y \in S_i} \sum_{t=1}^{T} u_i(y, x_{-i}^t) - \sum_{t=1}^{T} (u_i(x^t))$, where $u_i$ is the utility function of firm $i$, and $S_i$ is the strategy set of $i$.

## 6.3 Bertrand Oligopolies

We will be focusing on the case where are the all firms share the same linear cost function (i.e. $C_i(x) = cx$ for all $i$). The set of Nash equilibria of this game consists of all price vectors such that the prices of at least two firms are equal to $c$, whereas all others are greater than $c$. Although there exist multiple Nash equilibria, all of them imply the same market prices where the firms sell at marginal cost and hence no profit is being made. On the contrary, we will show that firms can achieve positive payoffs while experiencing no-regret. Moreover, we will show that infinitely many positive profit vectors are sustainable under no-regret guarantees.

We will show that by producing a probability distribution on outcomes of Bertrand oligopolies such that when the market outcomes are chosen according to this distribution, then each player's expected payoff is at least as large as the expected payoff of her best deviating strategy, given that all other players follow the distribution. More formally, we will produce a probability measure $F$ on ($P$, $\Sigma$),[4] such that for all $i, p_i'$

$$\int_P [u_i(p_i, p_{-i}) - u_i(p_i', p_{-i})] dF(p) \geq 0$$

Such probability distributions are referred to as *coarse correlated equilibria (CCE)* [106]. It is straightforward to check that, any market history whose empirical distribution of outcomes converges to a CCE imposes no regret on the involved players. Indeed, the average profits of the players, will converge to their expected values, which by definition of the CCE exhibit no-regret. Conversely, any CCE can give rise to such a history, merely by infinitely choosing outcomes according to it. Therefore it suffices to prove that we can achieve positive payoffs in a CCE. Our constructions are inspired by observations regarding the structure of Nash in Bertrand games made in [12].

**Theorem 6.3.1** *All symmetric linear Bertrand games exhibit coarse correlated equilibria (CCE) in which all players exhibit positive profits.*

**Proof:** We denote $(p-c)(a-p)/b$ by $\pi(p)$, which is equal to the utility function when the winning player is unique. This function in strictly increasing in [c,

[4] $P$ is the set of all strategy (price) profiles and $\Sigma$ is the Borel $\sigma$-algebra on it

(a+c)/2]. As a result, we can define the following distribution:

$$F_0(p) = \begin{cases} 0 & p \leq \beta \\ 1 - (\frac{\pi(\beta)}{\pi(p)})^{\frac{1}{n-1}} & \beta < p < \gamma \\ 1 & p \geq \gamma \end{cases} \tag{6.1}$$

where $\beta > c$ and $\gamma \leq (a+c)/2$. Before, we construct the CCE, we will examine some properties of the mixed strategy profile where each player chooses a strategy according to $F_0(p)$. We will show that each action in the support of the mixed strategy $F_0(p)$ is optimal [3] except from $\beta$.

The probability distribution $F_0(p)$ sets $p = \gamma$ with probability $(\frac{\pi(\beta)}{\pi(\gamma)})^{\frac{1}{n-1}}$. The rest of the probability distribution is atomless, that is $Pr(p = x|x < \gamma) = 0$. Suppose that the rest n-1 players play according to this distribution. The expected payoff for playing price $\beta \leq p < \gamma$ would be equal to:

$$E[u] = [1 - F_0(p)]^{n-1}\pi(p) = \pi(\beta)$$

Next, we will compute the expected payoff for playing $\beta$. The only way for someone to win when playing $\beta$ is for everyone else to be playing $\beta$. However, in this case they share the pot. So,

$$E[u] = [(\frac{\pi(\beta)}{\pi(\gamma)})^{\frac{1}{n-1}}]^{n-1}\frac{\pi(\gamma)}{n} = \frac{\pi(\beta)}{n}$$

Also, just to complete the picture, the expected cost for playing $p > \gamma$ is 0 and the expected profit for playing $p < \beta$ is less than $\pi(\beta)$. Lastly, let us compute

[3] given that all players play according to $F_0(p)$

the expected utility of the players when all of them play according to this strategy distribution. In this case and if we denote $(\frac{\pi(\beta)}{\pi(\gamma)})^{\frac{1}{n-1}}$ as $\rho$, we have that:

$$E[u] = (1-\rho)\pi(\beta) + \rho\frac{\pi(\beta)}{n} = (1 - \frac{n-1}{n}\rho)\pi(\beta)$$

Now, we will define a probability distribution over outcomes of the Bertrand games and we will prove that it is a CCE. we will be using three prices $\alpha, \beta, \gamma$ such that $c < \alpha < \beta < \gamma \leq (a+c)/2$. With probability 1/2 all players play $\alpha$ and with probability 1/2 all players play according to $F_0$. Regarding the expected payoff for each player, we have that with probability 1/2 they all share the profit at price $\alpha$ and with probability 1/2 they gain the precomputed payoff of the defined mixed strategy profile. Specifically,

$$E[u] = 1/2\frac{\pi(\alpha)}{n} + 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)]$$

In order for this to be a CCE it must be the case any deviating player cannot increase his payoff by deviating to a single strategy given that the rest of the players keep playing according to this distribution. Let us examine what are the best deviating strategies for a player. First a player can deviate and play $\alpha - \epsilon$ for some small $\epsilon > 0$. Her expected payoff in that case is essentially $\pi(\alpha)$ since she will always be winning the competition. It is obvious that any strategy less than that is clearly worse for him since $\pi$ is increasing in the range $[0, \alpha] \subset [0, (a+c)/2]$. Another good deviating strategy for the player is to play a strategy in $[\beta, \gamma)$ since this is a best response to the second probability distribution. Actually, given that a player deviates to a price which is greater than $\alpha$ her best choice is to deviate to any price in the $[\beta, \gamma)$ range. This is true because the only way to incur payoff at this point is to maximize her payoff when her opponents play according to $F_0(p)$. As we have seen, the player achieves a maximum expected payoff of $\pi(\beta)$ when playing within that range. So, the best deviating strategy is either $\alpha - \epsilon$ or something in the range $[\beta, \gamma)$. If our current expected payoff exceeds the payoffs at these points then our distribution is a CCE. So, we wish to have:

$$1/2\frac{\pi(\alpha)}{n} + 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)] \geq \pi(\alpha)$$

and

$$1/2\frac{\pi(\alpha)}{n} + 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)] \geq 1/2\ \pi(\beta)\ .$$

Let us try to analyze each relation separately:

$$\begin{aligned} 1/2\frac{\pi(\alpha)}{n} + 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)] \geq \pi(\alpha) \quad &\Leftrightarrow \\ 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)] \geq (1 - \frac{1}{2n})\pi(\alpha) \quad &\Leftrightarrow \\ (1 - \frac{n-1}{n}\rho)\frac{n}{2n-1} \geq \frac{\pi(\alpha)}{\pi(\beta)} & \end{aligned}$$

Similarly, from the second inequality we have:

$$\begin{aligned} 1/2\frac{\pi(\alpha)}{n} + 1/2[(1 - \frac{n-1}{n}\rho)\pi(\beta)] \geq 1/2\ \pi(\beta) \quad &\Leftrightarrow \\ \frac{\pi(\alpha)}{n} + [(1 - \frac{n-1}{n}\rho)\pi(\beta)] \geq \pi(\beta) \quad &\Leftrightarrow \\ \frac{\pi(\alpha)}{n} \geq \frac{n-1}{n}\rho\pi(\beta) \quad &\Leftrightarrow \\ \frac{\pi(\alpha)}{\pi(\beta)} \geq (n-1)\rho & \end{aligned}$$

So, in order for our probability distribution over outcomes to be a CCE, it suffices that we choose $\alpha, \beta$ so that:

$$(n-1)\rho \leq \frac{\pi(\alpha)}{\pi(\beta)} \leq (1 - \frac{n-1}{n}\rho)\frac{n}{2n-1}$$

However $\pi(\alpha), \pi(\beta)$ are positive payoffs in the range $(0, \frac{(a-c)^2}{4b}]$ with $\pi(\alpha) < \pi(\beta)$. So, by choosing proper $\alpha, \beta$ we have reproduce any number in the range $(0,1)$. Hence, all we have to do is show that we can choose $\rho$ appropriately such that:

$$(n-1)\rho \leq (1 - \frac{n-1}{n}\rho)\frac{n}{2n-1}$$

as well as $(n-1)\rho < 1$ and $0 < (1 - \frac{n-1}{n}\rho)\frac{n}{2n-1}$. Again, by manipulating the given inequality we get:

$$(n-1)\rho \leq (1 - \frac{n-1}{n}\rho)\frac{n}{2n-1} \Leftrightarrow (n - 1 + \frac{n-1}{2n-1})\rho \leq \frac{n}{2n-1}$$

It suffices to choose $\rho = \frac{1}{2n-1}$ and $\frac{\pi(\alpha)}{\pi(\beta)} = \frac{n-1}{2n-1}$ to satisfy all inequalities. However, $\rho = (\frac{\pi(\beta)}{\pi(\gamma)})^{\frac{1}{n-1}}$. So, we have that we need to choose $\beta$ and $\gamma$ such that $\frac{\pi(\beta)}{\pi(\gamma)} = (\frac{1}{2n-1})^{n-1}$. So, given any $\pi(\gamma) \in (0, \pi(\frac{a+c}{2})] = (0, \frac{(a-c)^2}{4b}]$, we can define $\beta, \alpha$ such that the distribution we have defined is a CCE. The expected payoffs of all players are positive in this CCE and can vary widely. Hence, no regret behavior can support infinitely many different positive average payoff profiles, in contrast to Bertrand's paradox. Finally, this construction establishes that increased competition is necessary for converging to marginal cost pricing. $\square$

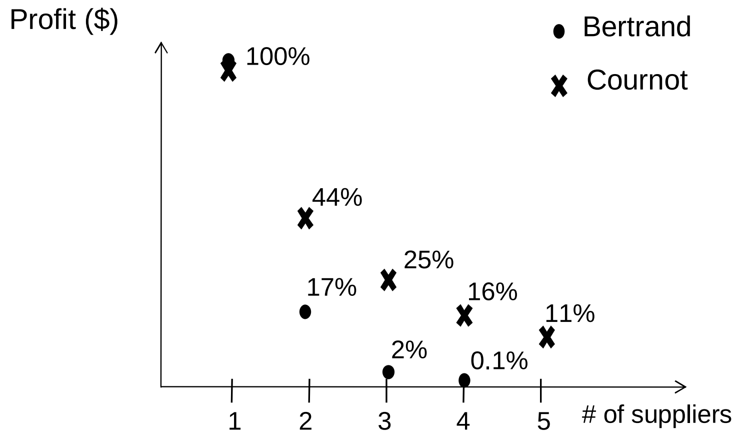


Figure 6.1: Profit decrease

As we see in the figure 6.1, the profitability of the families of Bertrand no-regret histories we have identified, decreases much faster than the profitability of the no-regret histories in the Cournot oligopolies as the number of agents (firms) increases. In fact, for $n = 4$ players we see that essentially the prices reach the level of marginal costs as profitability drops to zero. This theoretical projection is in perfect agreement both with experimental work in the case of Bertrand games [29], as well as with empirical observations about real world oligopolistic markets [95]. Specifically, "the rule of three", as is presented in [95], states that in most markets three major players will emerge (e.g. ExxonMobil, Texaco and Chevron in petroleum). In order for the smaller companies to be successful they need to specialize and address niche markets. Our works suggests a possible quantitative explanation behind this phenomenon, as a result of the steep drop in profitability in the case of Bertrand markets.

## 6.4 Cournot Oligopolies

We will be analyzing a generalization of the Cournot model with product differentiation that was introduced by Yi[105]. By exploring ideas from that work, we will show how we can generalize its results and prove tight convergence guarantees in the case of no-regret algorithms. Our model will be the Cournot competition in the case of linear demand functions with symmetric product differentiation, where the inverse demand function $P^i(q)$ is given by $P^i(q) = a_i - b_i q_i - b_i \gamma_i Q_{-i} = a_i - b_i(1-\gamma_i)q_i - b_i\gamma_i Q$, where $\gamma_i$ denotes the degree of product differentiation between products, $0 < \gamma_i \leq 1$, $b_i > 0$ and $Q = \sum_i q_i$ denotes the total product supply. We will assume that the cost functions are convex and twice continuously differentiable. We denote by $q^* = (q_1^*, \ldots, q_n^*)$ a pure Nash equilibrium of the one-shot game, which is known to exist by [82]. Finally, $Q^*$ denotes the aggregate production level at the Nash.

**Lemma 6.4.1** *Let $q_i^\tau, Q^\tau$ denote respectively the production level of company $i$ and the aggregate production level in period $\tau$ of a differentiated Cournot market with differentiation levels $\gamma_i$ for each product. If each player's regret converges to zero, then*

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t}\left(\frac{\gamma_i - 1}{\gamma_i}\sum_i (q_i^\tau - q_i^*)^2 - (Q^\tau - Q^*)^2\right) \quad = \quad 0$$

**Proof:** By assumption, we have that each player $i$ experiences vanishing regret against any deviating action $s_i \in S_i$. Specifically, we can apply this to their respective Nash equilibrium actions $q_i^*$.

$$\frac{1}{t}\sum_{\tau=1}^{t} u_i(q^\tau) = \max_{s_i\in S_i}\frac{1}{t}\sum_{\tau=1}^{t}\left(u_i(s_i, q_{-i}^\tau) - \mathcal{R}_i(t)\right) \geq \frac{1}{t}\sum_{\tau=1}^{t} u_i(q_i^*, q_{-i}^\tau) - \frac{\mathcal{R}_i(t)}{t} \quad (6.2)$$

Let us denote the difference $u_i(q^\tau) - u_i(q_i^*, q_{-i}^\tau)$ as $\Delta(u_i^\tau)$. Equation 6.2 allows us to bound $\sum_{\tau=1}^{t} \Delta(u_i^\tau)$ from below. Next, we will work on bounding this quantity from above.

$$
\begin{aligned}
\sum_{\tau=1}^{t} \Delta(u_i^\tau) &= \sum_{\tau=1}^{t} \left( P^i(q^\tau)q_i^\tau - C_i(q_i^\tau) - \left(P^i(q_i^*, q_{-i}^\tau)q_i^* - C_i(q_i^*)\right) \right) \\
&= \sum_{\tau=1}^{t} \left( P^i(q^\tau)q_i^\tau - P^i(q_i^*, q_{-i}^\tau)q_i^* - (C_i(q_i^\tau) - C_i(q_i^*)) \right) \\
&= \sum_{\tau=1}^{t} (q_i^\tau - q_i^*) \left( P^i(q^\tau) - C_i'(\bar{q}_i^\tau) - b_i q_i^* \right) \qquad (6.3)
\end{aligned}
$$

The last line is derived from the mean value theorem [3] and the fact that $P^i(q_i^*, q_{-i}^\tau) = P^i(q_i^\tau, q_{-i}^\tau) + b_i(q_i^\tau - q_i^*)$.

$$
\begin{aligned}
\sum_{\tau=1}^{t} \Delta(u_i^\tau) &= \sum_{\tau=1}^{t} (q_i^\tau - q_i^*) \left( P^i(q^\tau) - C_i'(\bar{q}_i^\tau) - b_i q_i^* \right) \\
&\leq \sum_{\tau=1}^{t} (q_i^\tau - q_i^*) \left( P^i(q^\tau) - P^i(q^*) - (C_i'(\bar{q}_i^\tau) - C_i'(q_i^*)) \right) \\
&= \sum_{\tau=1}^{t} (q_i^\tau - q_i^*) \Bigg( b_i(\gamma_i - 1)(q_i^\tau - q_i^*) - b_i\gamma_i(Q^\tau - Q^*) - \\
&- C_i''(\tilde{q}_i^\tau)(\bar{q}_i^\tau - q_i^*) \Bigg)
\end{aligned}
$$

where the inequality in the second line follows by the fact at a Nash equilibrium $q^*$, we have that

$$\frac{\partial u_i(q^*)}{\partial q_i} = P^i(q^*) - b_i q_i^* - C_i'(q_i^*) \leq 0.$$

Finally, the last line is derived by another application of the mean value theorem[4]

[3]There exists $\bar{q}_i^\tau$ between $q_i^\tau, q_i^*$ such that $C_i(q_i^\tau) - C_i(q_i^*) = C_i'(\bar{q}_i^\tau)(q_i^\tau - q_i^*)$.
[4]There exists $\tilde{q}_i^\tau$ between $\bar{q}_i^\tau, q_i^*$ such that $C_i'(\bar{q}_i^\tau) - C_i'(q_i^*) = C_i''(\tilde{q}_i^\tau)(q_i^\tau - q_i^*)$.

and the definition of the demand functions $P^i$. Now, we will take the following weighted sum of the resulting inequalities over all players $i \in N$.

$$\sum_i \frac{1}{b_i\gamma_i}\sum_{\tau=1}^{t}\Delta(u_i^\tau) \le \sum_{\tau=1}^{t}\left(\frac{\gamma_i - 1}{\gamma_i}\sum_i (q_i^\tau - q_i^*)^2 - (Q^\tau - Q^*)^2 - \right.$$
$$\left. - \sum_i \frac{C_i''(\tilde{q}_i^\tau)}{b_i\gamma_i}(q_i^\tau - q_i^*)(\bar{q}_i^\tau - q_i^*)\right)$$

By dividing the above inequality with $t$ and combining with 6.2 we conclude that:

$$-\sum_i \frac{\mathcal{R}_i(t)/b_i\gamma_i}{t} \le \frac{1}{t}\sum_{\tau=1}^{t}\left(\frac{\gamma_i - 1}{\gamma_i}\sum_i (q_i^\tau - q_i^*)^2 - (Q^\tau - Q^*)^2 - \right.$$
$$\left. - \sum_i \frac{C_i''(\tilde{q}_i^\tau)}{b_i\gamma_i}(q_i^\tau - q_i^*)(\bar{q}_i^\tau - q_i^*)\right) \tag{6.4}$$

Given the definition of regret minimizing behavior we have that for all $i$, $\limsup_{t\to\infty} \frac{\mathcal{R}_i(t)}{t} \le 0$, therefore the $\limsup_{t\to\infty}$ of the second terms will greater or equal to 0. If the cost functions are weakly convex all three terms in the summand are less or equal to 0. As a result, we have that:

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t}\left(\frac{\gamma_i - 1}{\gamma_i}\sum_i (q_i^\tau - q_i^*)^2 - (Q^\tau - Q^*)^2 - \sum_i \frac{C_i''(\tilde{q}_i^\tau)}{b_i\gamma_i}(q_i^\tau - q_i^*)(\bar{q}_i^\tau - q_i^*)\right) = 0 \tag{6.5}$$

The lemma follows immediately. □

Depending on the details of the Cournot model, we have the following cases:

## A) Perfect Substitutes

This is to the simplest case of Cournot competition and was the model analyzed by Even-Dar et. al. in [31]. We have that $\gamma_i = 1$ and $C_i'''(q_i) \geq 0$ for all $i, q_i$. Equation 6.5 implies that:

$$\limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} (Q^\tau - Q^*)^2 = 0$$

Intuitively, this equation suggests that if we exclude a statistically insignificant (sublinear) number of periods for the history of our no-regret play, then for the rest of the history the overall production levels converge to $Q^\tau$ (and therefore the prices $(P^i(q) = a_i - b_i Q, ))$ converge to their levels at the Nash equilibrium $Q^*$.

**Theorem 6.4.2** *Suppose that $n$ firms participate in a homogeneous Cournot oligopoly game of perfect substitutes with linear demand $(P^i(q) = a_i - b_i Q, )$ and convex cost functions. If all firms experience no-regret as t grows to infinity, then given any $\epsilon > 0$, for all but $o(t)$ periods $\tau$ in $[1, t]$ we have that $|Q^\tau - Q^*| < \epsilon$.*

**Proof:** We will prove this by contradiction. Indeed, suppose that this did not hold, then there would exist $\epsilon > 0$ such that it would not be the case that for all but $o(t)$ periods $\tau$ in $[1, t]$ we have that $|Q^\tau - Q^*| < \epsilon$. Namely, if we define $s_t = \{\tau | \tau \in [1, t] \text{ and } |Q^\tau - Q^*| \geq \epsilon\}$ then there exists $c > 0$ such for all $k$ there exists $t \geq k$ such that $|s_t| \geq ct$. Hence, we can define an infinite subsequence $t_0, t_1, \ldots$ such that for all $k$, $|s_{t_k}| \geq ct_k$. Hence for this subsequence, we have that for all $k$, $\frac{1}{t_k} \sum_{\tau=1}^{t_k} (Q^\tau - Q^*)^2 \geq c\epsilon^2 > 0$. Therefore, we reach a contradiction, since $\limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} (Q^\tau - Q^*)^2 = 0$. $\square$

We should stress here that this is a statement about the day-to-day behavior (i.e. aggregate production levels) instead of average behavior as in [31](Theorem 3.1.). In particular, this statement implies that the average action vector and the average utility of each player converge to their respective levels at the Nash equilibrium, a result that has been shown in [31]. Given the convergence of the day-to-day characteristics of the market prices and total supply, it is rather tempting to try to prove a similar statement about the convergence of the action vector and utilities of the firms and not merely of their averages. Here, we show that this cannot be the case by providing sufficient conditions for a market history to be of no-regret.

This is essentially a negative result, so it suffices to prove that this holds for as simple a model as possible. Therefore, we will focus on the special case of the fully symmetric Cournot oligopoly ($a_i = a$ and $b_i = b$) with linear cost functions. It is well known that these games exhibit a unique Nash $q^* = (q_1^*, q_2^*, \ldots, q_2^*)$ where $q_i^* = (a - (n+1)c_i - \sum_{j \in N} c_j)/((n+1)b)$.

**Theorem 6.4.3** *Suppose that $n$ firms participate in a homogeneous Cournot oligopoly game with linear demand ($P^i(q) = a - bQ,$) and linear cost functions and let $q*$ denote the unique Nash of this game. Any market history, where for all time periods $\tau$, $Q^\tau = Q^*$ and where the time average $\hat{q}_i$ of each player's actions converges to her Nash strategy $q_i^*$ does not induce regret to any player.*

**Proof:** As we have shown in equation 6.3 for all deviating strategies $q_i'$ we have that:

$$\sum_{\tau=1}^{t} \big(u_i(q^\tau) - u_i(q_i', q_{-i}^\tau)\big) = \sum_{\tau=1}^{t} (q_i^\tau - q_i') \left(P^i(q^\tau) - C_i'(\bar{q}_i^\tau) - bq_i'\right)$$

where $\bar{q}$ between $q_i^\tau, q_i^*$ is such that $C_i(q_i^\tau) - C_i(q_i^*) = C_i'(\bar{q}_i^\tau)(q_i^\tau - q_i^*)$. However, in this case we have that since $C_i(x) = c_i x$, $C_i'(x) = c_i$ for all $x$. We can also substitute $P^i(q^\tau)$ with $a - bQ^\tau$. As a result, we derive that

$$\begin{aligned}
& \frac{1}{t}\sum_{\tau=1}^{t}\left(u_i(q^\tau) - u_i(q_i', q_{-i}^\tau)\right) \\
= \ & \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i') \cdot (a - bQ^\tau - c_i - bq_i') \\
= \ & \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i')\,(a - bQ^* - c_i - bq_i') && (6.6) \\
= \ & \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i')\left(a - \frac{na - \sum_i c_i}{n+1} - c_i - bq_i'\right) && (6.7) \\
= \ & \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i')\left(\frac{a + \sum_i c_i - (n+1)c_i}{n+1} - bq_i'\right) \\
= \ & \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i')b(q_i^* - q_i') && (6.8) \\
= \ & \frac{b(q_i^* - q_i')}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i') \\
= \ & b(q_i^* - q_i')(\hat{q}_i - q_i')
\end{aligned}$$

where line 6.6 is derived by hypothesis. Also, lines 6.7, 6.8 follow from the fact that the unique Nash $q^* = (q_1^*, q_2^*, \ldots, q_2^*)$ of these games is of the form

$$q_i^* = \frac{a - (n+1)c_i - \sum_{j \in N} c_j}{(n+1)b}.$$

Lastly, since by hypothesis we have that $\lim_{t\to\infty}\hat{q}_i = q_i^*$, we derive that for each player $i$ and all deviating actions $q_i'$ we have that:

$$\lim_{t\to\infty}\frac{1}{t}\sum_{\tau=1}^{t}\left(u_i(q^\tau) - u_i(q_i', q_{-i}^\tau)\right) = b(q_i^* - q_i')^2 \geq 0$$

Hence in any such market history no players experience regret. $\square$

An immediate corollary of the above theorem is that one cannot hope to prove convergence of the day-to-day action profiles in any model that generalizes the basic linear Cournot model. Surprisingly, if we introduce product differentiation in the market, then we can actually prove convergence of all attributes (i.e. action profiles, profits, prices e.t.c) of the market.

## B) Symmetric Product Differentiation

In this case, we have that $0 < \gamma_i < 1$ and $C_i''(q_i) \geq 0$ for all $i, q_i$. Equation 6.5 implies that for all firms $i$,

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i^*)^2 = 0$$

**Theorem 6.4.4** *Suppose that n firms participate in a differentiated good Cournot oligopoly game with linear demand ($P^i(q) = a_i - q_i - \gamma Q, 0 < \gamma < 1$). If all firms experience no-regret, then given any $\epsilon > 0$, as $t$ grows to infinity, for all but $o(t)$ periods $\tau$ in $[1,t]$ we have that $|q_i^\tau - q_i^*| < \epsilon$, where $q^*$ is the unique Nash equilibrium.*

**Proof:** We will prove this by contradiction. Indeed, suppose that this did not hold, then there would exist $\epsilon > 0$ such that it would not be the case that for all but $o(t)$ periods $\tau$ in $[0,t]$ we have that $|q_i^\tau - q_i^*| < \epsilon$. Namely, if we define $s_t = \{\tau | \tau \in [1,t] \text{ and } |q_i^\tau - q_i^*| \geq \epsilon\}$ then there exists $c > 0$ such for all $k$ there exists $t \geq k$ such that $|s_t| \geq ct$. Hence, we can define an infinite subsequence $t_0, t_1, \ldots$ such that for all $k$, $|s_{t_k}| \geq ct_k$. Hence for this subsequence, we have that for all $k$, $\frac{1}{t_k}\sum_{\tau=1}^{t_k}(q_i^\tau - q_i^*)^2 \geq c\epsilon^2 > 0$. Therefore, we reach a contradiction, since $\limsup_{t\to\infty}\frac{1}{t}\sum_{\tau=1}^{t}(q_i^\tau - q_i^*)^2 = 0$. $\square$

From the analysis above we have that in Cournot games, the market participants can derive a reasonable estimate of their long-term average (and sometimes their daily) profits by focusing on the unique Nash equilibrium of these games. In the next chapter, we will argue how market producers can exploit such knowledge and try to improve their position by dynamically forming coalitions with the other market participants.

CHAPTER 7

# COALITION FORMATION AND THE PRICE OF ANARCHY IN COURNOT OLIGOPOLIES

## 7.1 Introduction

It is a basic tenet of algorithmic game theory that agents act not maliciously or cooperatively, but in their own self-interest. Borrowing from economics, the literature purports that these agents will take actions which collectively form a Nash equilibrium (or a related solution concept). Hence the actions could be potentially far from the social optimal. For example, in a road network, each driver, observing traffic patterns, selects the route which minimizes his own delay. The resulting total delay can be much greater than that of the optimal flow.

In a seminal paper in 1999, Koutsoupias and Papadimitriou [62] initiated the investigation of the so-called *price of anarchy* which measures the ratio of the social value in the worst-case equilibrium to the optimal social value. Recent years have seen a profusion of results exploring the price of anarchy of various non-cooperative games. The traffic example mentioned above, known as *selfish routing* in the literature, has a bounded price of anarchy of 4/3 for linear latency functions [86]. This can be viewed as a positive result. However, many settings have a drastically large price of anarchy, e.g., Cournot oligopoly games, which model competition between firms, have a linear price of anarchy for certain production functions [60].

As shown by Hayrapetyan, Tardos and Wexler [53], the price of anarchy can degrade even further in the presence of collusion. When agents collude, they effectively act as a super-player whose strategy is a vector consisting of a single strategy for each member. The utility of the super-player is then the aggregate utility of its members. In such settings, the authors demonstrate that, although the price of anarchy clearly drops to one if all agents collude, for inter mediate partitions of agents into cooperative groups, the price of anarchy can actually rise above that of the completely non-cooperative setting. In some instances (e.g., convex games), the price of anarchy is arbitrarily high in the presence of coalitions. This is however a worst case result under the assumption that any partition of players into coalitions is feasible.

The basic thesis of this chapter is that many such price of anarchy results are overly pessimistic, because worst case instances of games do not arise in practice. Intuitively, if the price of anarchy is high, agents can try to change the nature of the game by forming coalitions. These coalitions will not be arbitrary, but rather carefully selected ones that reduce the price of anarchy and at the same time are sustainable, i.e., stable against deviations.

With this in mind, we define a coalition formation game to capture this selection process and use Cournot oligopoly games to demonstrate our main point that price of anarchy can be significantly reduced via such coalition formation processes. The approach can be applied to other classes of games (which we leave for future work) as long as certain conditions are satisfied (see Section 7.5) and we believe this to be a promising path for alleviating negative price of anarchy results.

In our coalition formation game actions correspond to changes in the current coalition structure. Specifically, a new coalition can be created by a merger between two or more existing coalitions. An existing coalition can also be destroyed due to a deviation of a subset of its current players who decide either to form a coalition by themselves or join an existing coalition. For a new coalition to be formed, it must be the case that its creation benefits all its members.

Given a current coalition structure, we treat each coalition as a super-player who, as in [53], acts on behalf of its members and tries to maximize its aggregate utility. In the Cournot oligopolies that we consider, any such game between the super-players (coalitions) has Nash equilibria and in fact we show that the utilities of the super-players at Nash equilibria are unique. This is a crucial property that allows us to assign a unique value to a coalition given the current partition, which is reminiscent of the approach in [80]. Finally we divide this utility equally among the members of a coalition since we are focusing on symmetric Cournot games where all players have equal production costs.

Given the rules of the game described above, we are interested in stable coalition configurations, i.e., partitions where no profitable deviating actions exist with regard to the allowed actions we have defined. We analyze the social welfare of the worst such stable partition and compare it to the cost of the optimum and refer to this ratio as the price of anarchy of our coalition formation game. We find that the price of anarchy of our coalition formation game for Cournot oligopolies is $\Theta(n^{2/5})$, where $n$ is the number of firms that participate in the market, implying a significant improvement of the actual price of anarchy of Cournot oligopolies which is $\Theta(n)$.

The value assignment to coalitions, as described in the previous paragraphs relies on the assumption that if a coalition structure is stable and hence not permeable, then the super-players coalitions will reach a Nash equilibrium of the underlying game. We show that we can weaken this assumption considerably in the case of Cournot games. Specifically, we can show that if the coalitions participate in the Cournot oligopoly repeatedly in a fashion that minimizes their long term regret then the average utility of the super-players (coalitions) will converge to their levels at Nash equilibria. Regret compares a players average utility to that of the best fixed constant action with hindsight. Having no-regret means that no deviating action would significantly improve the firm's utility. Several learning algorithms ([20]). More importantly, the assumption is not tied to any specific algorithmic procedure, but instead captures successful long-term behavior. Finally, since the setting of oligopolies markets is in its nature repeated, this observation significantly strengthens the justification of our model.

**Chapter Structure.** Section 7.2 offers the definition of Cournot oligopolies as well as a detailed exposition of our coalition formation model. In Section 7.3 we prove tight bounds for the price of anarchy of the Cournot coalition formation game. Finally, Section 7.4 extends our analysis to the case of no-regret behavior.

### 7.1.1 Related Work

Dynamic coalition formation problems have been studied extensively in the economics and computer science literature. We refer the reader to [28], [21], and [2][Section 5.1] as well as the numerous references therein. The main goals of these works have been to provide appropriate game theoretic solution concepts (both from a cooperative and noncooperative point of view) and to design intuitive procedures that converge experimentally or theoretically to such solution concepts.

Quantifying the inefficiency of outcomes when coalitions are allowed to form has been the subject of much recent work. In [53], the authors initiate the study of the *price of collusion*, which is the worst case inefficiency over all possible partitions of the set of players. Another interesting direction has been to analyze the *price* *of strong anarchy*, see [5], i.e. the inefficiency of Nash equilibria which are resilient to deviations by coalitions[1]. The models above are static and any partition is allowed to form. In contrast, we focus on a dynamic context, where we allow only stable partitions, i.e., partitions where agents have no incentive to deviate, as we define in Section 7.2. In [22], a different measure is introduced, namely the *price of democracy*. This notion captures the inefficiency of a given coalition formation process (e.g. a bargaining process) with respect to a cooperative game. The authors study this notion in the context of weighted voting games for certain intuitive bargaining processes. Hence the inefficiency is measured with regard to the arising partitions in the subgame perfect equilibria of the corresponding bargaining game.

Regarding Cournot games, it has been long known that the loss of efficiency at Nash equilibria can be quite high. Earlier studies focused on empirical analysis [49] whereas more recently, price of anarchy bounds have been obtained in [47, 60]. Collusion and cartel enforcement in Cournot games have been studied experimentally, see e.g., [99]. There have also been mechanism design aspects of collusion, see [26]. Conceptually the closest example to our approach that we know of is the work of Ray and Vohra in [80]. The authors propose a solution concept ("binding agreement") that allows for the formation of coalition structures and examine the inefficiency of stable partitions. Unlike in our work, their deviations can only make the existing coalition structure finer- never coarser. In the case of symmetric Cournot games, it is shown that there always exists a stable partition with social welfare $O(\sqrt{n})$ worse than the optimal. However, the social welfare of the worst stable partition is always at least as bad as that of the worst Nash.

[1]Unfortunately, for several classes of games including Cournot equilibria strong Nash do not exist.

Finally, there has been some recent work on the behavior of no-regret algorithms in Cournot oligopolies. In [31], [75] several convergence results are shown for different classes of Cournot oligopolies. To our knowledge this work is the first to consider the behavior of *coalitions* which are behaving in a no-regret fashion in any kind of setting.

## 7.2 The Model

We will demonstrate the main point of our work in the context of Cournot games. The definitions presented in this Section can be easily generalized and applied to other contexts but we postpone a more general treatment for an extended version.

Cournot games describe a fundamental model of competition between firms. They were introduced by Cournot in his much celebrated work [25]. In the Cournot games, firms control their production levels and by doing so influence the market prices. In the simplest Cournot model all the firms produce the same good; the demand for this product is linear in the total production (i.e. the price decreases linearly with total production); the unit cost of production is fixed and equal across all firms. The revenues of a firm are the product of the firm's part of the market production times the price. Finally, the utility of a firm is equal to its revenue minus its total production cost. Overproducing leads to low prices, while at the same time an overly cautious production rate leads to a small market share and reduced revenue. The balancing act between these two competing tendencies is known to give rise to a unique Nash equilibrium. More formally:

**Definition:** A linear and symmetric Cournot oligopoly is a noncooperative game between a set $N = \{1, 2, ..., n\}$ of players (firms), all capable of producing the same product. The strategy space of each firm is $\mathbb{R}_+$, corresponding to the quantity of the product that the firm decides to produce. Given a profile of strategies, $q = (q_1, ..., q_n)$, the utility of firm $i$ is $u_i(q) = q_i p(q) - cq_i$, where $p(q)$ is the price of the product, determined by $p(q) = \max\{0, a - b\sum_i q_i\}$, for some parameters $a, b$, and $c$ is a production cost, with $a > c$.

**Proposition 7.2.1 ([25])** *In the unique Nash equilibrium of a Cournot oligopoly with $n$ players, the production level is the same for all players and equal to $q_i = q^* = \frac{(a-c)}{b(n+1)}$. The utility of each player is $u_i = \frac{(a-c)^2}{b(n+1)^2}$.*

For more on Cournot games and their variants, we refer the reader to [65].

### 7.2.1 Cournot Games with a Fixed Partitioning of the Players

Suppose now that the players are given the opportunity to form coalitions and sign agreements with other firms, as a means of reducing competition and improving on their welfare. Given a partition of the players into coalitions, we can think of the new situation as a super-game whose super-players are the coalitions themselves. The strategy for a coalition, or super-player, is now a vector assigning a strategy to each of its members. The payoff to the super-player is the aggregate payoff its members would achieve with their assigned strategies in the original game. This definition can be used to model coalitions in general games as in [53].

**Definition:** Let $\mathcal{G}$ be a game of $n$ players, with $A_j$ being the set of available actions and $u_j^{\mathcal{G}}(a_1, \ldots, a_n)$ the utility function for each player $j$. Given a partitioning

$\Pi = (S_1, ..., S_k)$ of the players, then the corresponding *super-game* consists of the following:

- $k$ *super-players*
- The strategy set for super-player $S_i$ is the set of vectors $\overrightarrow{a}_{S_i} \in \prod_{j \in S_i} A_j$.
- The utility of super-player $S_i$ is $u_i(\overrightarrow{a}_{S_1}, \ldots, \overrightarrow{a}_{S_k}) = \sum_{j \in S_i} u_j^{\mathcal{G}}(a_1, \ldots, a_n)$ where $a_j$ is the strategy assigned to player $j$ by his coalition $S_i$ in the coalition's strategy $\overrightarrow{a}_{S_i}$.

Henceforth, when it is clear from the context, we will use game instead of super-game and player instead of super-player.

The following Lemma demonstrates that for Cournot games, the super-game with $k$ super-players is essentially equivalent to a Cournot game with $k$ players. This allows us to use theorems regarding Cournot games to study the Nash equilibria and welfare of Cournot games with coalitions.

**Lemma 7.2.2** *Consider a Cournot oligopoly super-game for a fixed partitioning $\Pi = (S_1, ..., S_k)$ of players. The social welfare in this game under any strategy profile $\overrightarrow{q}_{S_1}, \ldots, \overrightarrow{q}_{S_k}$ (where $\overrightarrow{q}_{S_i} \in \mathcal{R}_+^{|S_i|}$) is equivalent to the social welfare of a linear and symmetric Cournot game with $k$ players where each player $i$ produces the aggregate production $\sum_{j \in S_i} (\overrightarrow{q}_{S_i})_j$ of the corresponding coalition $S_i$. Furthermore, a strategy profile for the super-game with the fixed partitioning is a Nash equilibrium if and only if the $k$-tuple of the aggregate levels of productions for each coalition is the unique Nash equilibrium for the Cournot game on $k$ players (without coalitions).*

**Proof:** Let's denote by $Q$ the total production level, i.e. $Q = \sum_i^N q_i$. Similarly, for any coalition $S_i$, we define as $Q_{S_i} = \sum_{i \in S_i} q_i$. For each (super-)player $S_i$, we have that $u_{S_i}(\overrightarrow{q}) = \sum_{i\in S_i}(a-c-bQ)q_i = (a-c-bQ)Q_{S_i}$. In terms of social welfare we have that, $SW = \sum_{i=1}^{k} u_{S_i} = \sum_{i=1}^{k}(a-c-bQ)Q_{S_i} = (a-c-bQ)Q$. Similarly, in the case where, coalitions are replaced with single players with $q_i = Q_{S_i}$, we derive the same formula for the social welfare by the definition of Cournot games (i.e. $u_i(\overrightarrow{q}) = (a-c-bQ)q_i$).

If $\overrightarrow{q}$ is a Nash equilibrium of the partition game, then for each $i$, $u_{S_i}(\overrightarrow{q}) = (a-c-bQ)Q_{S_i}$ is maximal given the choices of all the players not in $S_i$. Since, the utility of $S_i$ depends only on the aggregate production level of the other coalitions, we have that $Q_{S_i}$ maximizes $u_{S_i}(\overrightarrow{q})$, given the rest $Q_{S_j}$. Hence, $(Q_{S_1},\ldots,Q_{S_k})$ is a Nash equilibrium of a Cournot oligopoly with $k$ players. Similarly, if $Q^* = (q_1,\ldots,q_k)$ is a Nash equilibrium of a Cournot oligopoly with $k$ players, then any vector $\overrightarrow{q} \in \mathcal{R}_+^{|S_i|}$ whose aggregate production levels for each coalition agree with their respective levels in $Q^*$ is also a Nash equilibrium of the Cournot game with coalitions. Indeed, if that was not the case, there would exist a (super-)player with a deviating strategy such that her resulting utility would increase her utility given the strategies of the other (super-)players. Such a deviating strategy, must induce a differentiated aggregate production for (super-)player $S_i$, since all strategies with the same aggregate utility lead to the same payoff ($u_{S_i}(\overrightarrow{q}) = (a-c-bQ)Q_{S_i}$). Let's define this new aggregate production level as $Q'_{S_i}$, then in the original Cournot game $Q'_{S_i}$ is a strictly better response for player $i$ that her current strategy and we reach a contradiction. $\square$

Lemma 7.2.2 implies that the social welfare is the same in all Nash equilibria of the Cournot game with a fixed partitioning. Hence we can define the price of anarchy as the ratio of this social welfare over the optimal social welfare, which is realized when all agents unite into a single coalition. The following Lemma shows that the price of anarchy depends only on the number of coalitions in the partition.

**Lemma 7.2.3** *The price of anarchy of a Cournot oligopoly with a fixed partition* $\Pi = (S_1, ..., S_k)$ *is* $\frac{(k+1)^2}{4k}$.

**Proof:** From lemma 7.2.2, we have that the social welfare of all Nash equilibria of a Cournot oligopoly with a fixed coalition partition $\Pi = (S_1, ..., S_k)$, is equal to the social welfare of the unique Nash of a Cournot oligopoly with $k$ players. In the case of symmetric Cournot oligopolies, this social welfare is equal to $\frac{k}{(k+1)^2}\frac{(a-c)^2}{b}$. This is the case, because at the unique Nash the production level of all players is equal to $\frac{a-c}{(k+1)b}$. The optimal outcome arises when there is no competition (i.e. there exists a single producing player or coalition). In this case, the optimal payoff is equal to $\frac{1}{4}\frac{(a-c)^2}{b}$, since this is the global maximum of the function $\frac{(a-c-Q)Q}{b}$. By dividing the resulting social welfares we derive the price of anarchy. $\square$

As a consequence, the price of anarchy in the original noncooperative Cournot oligopoly with $n$ players is very high, namely linear in the number of players, as has been observed previously [60].

**Corollary 7.2.4** *The price of anarchy in the original Cournot game with n players, where no coalitions are allowed to form is* $\Theta(n)$.

### 7.2.2 Cournot Coalition Formation Games

Next, we move away from the fixed coalition structure assumption and instead we will allow the players to dynamically form coalitions. We will call this game the

*Cournot coalition formation game*. Given some initial partition, players or sets of players can consider deviations according to the rules that we define below. As we have seem by Lemma 7.2.2, for any resulting partition, say with $k$ coalitions, the utility of each coalition is unique in all Nash equilibria of the Cournot game with fixed coalitions, and equal to the utility of a player in the unique Nash equilibrium of a symmetric Cournot game with $k$ players. In the coalition formation game, each of the $n$ players, when evaluating a possible action of hers, estimates her resulting utility to be equal to her equiproportional share of the Nash equilibrium utility of the coalition to which she belongs, given the resulting coalition structure. More formally:

**Definition:** We define a coalition formation game on top of a symmetric Cournot game to consist of the following:

- $n$ players and a current partitioning of them into $k$ coalitions $\Pi = (S_1, ..., S_k)$,
- Given the current partition $\Pi$, the allowed moves (deviations) that players can use along with the consequences for the coalition left behind (i.e., the non-deviators) are as follows:

  *Type 1:* A strict subset $S_i'$ of a current coalition $S_i$ decides to deviate and form a new coalition. The rest of the members of the original coalition (i.e. $S_i/S_i'$) dissolve into singletons. [2]

  *Type 2:* A strict subset $S_i'$ of a current coalition $S_i$ decides to leave its current coalition $S_i$ and join another coalition of $\Pi$, say $S_j$. The rest of the members, if any, of the original coalition (i.e. $S_i/S_i'$) dissolve into singletons.

[2] This type of actions also includes the non-action option (i.e. the coalition structure remains unaltered), when $S_i' = S_i$.

*Type 3:* A set of coalitions of $\Pi$ decide to unite and form a coalition. The rest of the coalitions remain as they were.

- Given a partition $\Pi$ and a player $i$ in coalition $S_j$ of $\Pi$, denote by $u(S_j)$, the uniquely defined utility of coalition $S_j$ in the symmetric Cournot game with fixed coalition structure $\Pi$. The utility of player $i$ in this case, is defined to be equal to $u(S_j)/|S_j|$.

Some justification is in order here for our definition. First, we have assumed that members of a coalition share proportionally the utility of the coalition since we are in a symmetric Cournot game, where all firms have the same production cost. Hence it would be unfair to resort to non-proportional shares. Secondly, we have made the assumption in some of the deviating actions (Type 1 and 2) that the left over coalition from where the deviation emerged, dissolves into singleton players. This is an assumption that has also been considered in other works on Cournot games, such as [26], where cartel enforcement is studied in Cournot games from a mechanism design point of view. Furthermore, we want to note here that despite the large amount of work in coalition formation, there is no unanimously accepted model for coalition formation and especially on how the left over players react once a subset of the coalition they belong to decides to deviate. This has been a disputed issue especially in the literature on partition function games, where the value of a coalition depends on how the rest of the players partition themselves. Several approaches have been considered there. One of them is based on assuming that the non-deviators are trying to partition themselves so as to hurt the deviators ($\alpha$-core, see [8]); a more optimistic approach has also been considered, where non-deviators react in the best possible way for the deviators, leading to the so-called $\omega$-core [94]. And within these two extremes, another suggestion that has been made in [23] is to have the left over players dissolve into singletons and play individual best reply strategies. We believe this last approach can be a good fit for an oligopoly market as in commercial agreements among firms, once there is a deviation, the left over firms may not be able to sustain the agreement and have to dissolve before reconsidering the terms of continuing their cooperation. For more on alternative models for the behavior of non-deviators see [61][Section 3.1].

We will be interested in analyzing the price of anarchy for partitions in which no player or set of players has an incentive to change the current coalition structure. In order to characterize stable coalitions, we need to define when a deviation is successful. A deviation is successful if and only if the utility of all the players that induce this deviation strictly increases as a result. More formally, we have that:

**Definition:** A deviation is successful iff all the players that facilitate the deviation strictly increase their payoff by doing so. Specifically, a deviation of

- *Type 1* is successful iff all the players in $S_i'$ increase their payoffs.
- *Type 2* is successful iff all the deviating players in $S_i'$ as well as all the members of the coalition $S_j$ who accept them increase their payoffs.
- *Type 3* is successful iff all the players of all the merging coalitions increase their payoffs.

**Definition:** A partition $\Pi$ is stable if there is no successful deviation of any type.

In the usual manner of the "price of anarchy" literature, we are interested in bounding the ratio of the social welfare of the worst stable outcome (i.e. coalition partition) divided by the optimal social welfare. In our setting, the stable outcomes do not correspond exactly to Nash, since we allow bilateral moves (e.g. type 3).

Nevertheless, we will still use the term price of anarchy to refer to this ratio, since it characterizes the loss in performance due to the lack of a centralized authority that could enforce the optimal (grand) coalition.

**Definition:** Given a Cournot coalition formation game, we define the price of anarchy as the ratio of the social welfare that is achieved at the worst stable partition divided by the optimal social welfare.

## 7.3 The Main Result

The starting point of our work is the observation of Corollary 7.2.4 that without coalition formation the price of anarchy is $\Theta(n)$. Hence our goal is to understand the quality of the worst stable partition structure and compare it to the optimal. The optimal partition structure is trivially the one where all players have united in a single coalition, as there is no competition in such setting. Our main result is that the price of anarchy can be reduced when coalition formation is allowed. Formally:

**Theorem 7.3.1** *The price of anarchy of the coalition formation game is* $\Theta(n^{2/5})$.

### 7.3.1 The Proof of the Upper Bound

We begin by proving that the price of anarchy is $O(n^{2/5})$. We will first establish this upper bound on a restricted version of our model. In particular, we restrict each type of the allowed deviations of Definition 7.2.2 as follows:

*Type 1:* A member of a coalition of $\Pi$, decides to form a singleton coalition on his own. The coalition from which the player left dissolves into singleton players.

*Type 2:* A member of a coalition of $\Pi$ decides to leave its current coalition $S_i$ and join another coalition of $\Pi$, say $S_j$. The rest of coalition $S_i$ (if $|S_i| \geq 2$) dissolves into singleton players.

*Type 3:* A set of singleton players of $\Pi$ decide to unite and form a coalition.

We will refer to this game as the *restricted coalition formation game*. Once we establish the upper bound in the restricted model, it is trivial to extend it to the general model since the set of stable partitions only gets smaller in the general model. The reason for choosing this restricted version is that stability against this restricted set of deviations is sufficient to impose the desired upper bound as we exhibit below.

To analyze the price of anarchy, we first derive a characterization of stable partitions in the restricted game. This will depend on a number of lemmas. The bound of $n^{2/5}$ on the price of anarchy is finally derived as a solution to a certain mathematical program, which is determined by the constraints that are imposed by stability against deviations (see Theorem 7.3.8).

Throughout the analysis we will normally denote the cardinality of a coalition $S_i$ by $s_i = |S_i|$. The first Lemma below says that for coalitions of size at least 2, its members need only consider Type 1 deviations.

**Lemma 7.3.2** *Consider a partition* $\Pi = (S_1, ..., S_k)$*, with* $k \geq 2$*. For a player that belongs to a coalition of* $\Pi$ *of size at least* 2*, the most profitable deviation (though not necessarily a successful one) is the deviation where the player forms a singleton* *coalition on his own.*

**Proof:** Consider a coalition $S_i$ of $\Pi$ and denote its cardinality by $s_i = |S_i|$. Suppose $s_i \geq 2$ and consider a player $j \in S_i$. The available deviations for $j$ are either to form a coalition on his own or to join an existing coalition. In the former case, the coalition $S_i$ will dissolve and the total number of coalitions in the new game will be $k + s_i - 1$. Hence the payoff of $j$ will be

$$u = \frac{(a-c)^2}{b(k+s_i)^2}$$

On the other hand, if $j$ goes to an existing coalition, then $S_i$ again dissolves but the total number of coalitions is now $k + s_i - 2$. Since $j$ will be in a coalition with at least 2 members, the payoff to $j$ will be at most:

$$u' \leq \frac{(a-c)^2}{2b(k+s_i-1)^2}$$

We want to prove $u \geq u'$, which is equivalent to $(k+s_i)^2 \leq 2(k+s_i-1)^2$, which in turn is equivalent to $(k+s_i)^2 - 4(k+s_i) + 2 \geq 0$. For this it suffices to show that $(k+s_i) \geq 2 + \sqrt{2}$. But we have assumed that $k \geq 2$ and that $s_i \geq 2$, hence the proof is complete. $\square$

The next lemma is based on Lemma 7.3.2 and characterizes coalitions of size at least 2, for which there are no successful deviations for its members.

**Lemma 7.3.3** *Consider a partition $\Pi = (S_1, ..., S_k)$, with $k \geq 2$. For a coalition $S_i$ with $s_i \geq 2$, there is no successful deviation for its members iff $s_i \geq k^2$.*

**Proof:** Consider a coalition of partition $\Pi$, say $S_i$ with $s_i \geq 2$. The payoff that a player in $S_i$ now receives is $u = \frac{(a-c)^2}{s_i b(k+1)^2}$. By Lemma 7.3.2 the most profitable deviation for any player of $S_i$ is to form a singleton coalition, in which case he would receive a payoff of $u = \frac{(a-c)^2}{b(k+s_i)^2}$. Hence in order that no player has an incentive to deviate, we need that $(k+s_i)^2 \geq s_i(k+1)^2$, which is equivalent to $s_i \geq k^2$. $\square$

We now deal with deviations of players that form singleton coalitions in a partition $\Pi$. The next Lemma eliminates the potential deviations of such players.

**Lemma 7.3.4** *Let $n \geq 3$, and consider a partition $\Pi = (S_1, ..., S_k)$, with $k \geq 2$. No singleton player of $\Pi$ has an incentive to perform a Type 2 deviation.*

**Proof:** The payoff that a singleton player receives in $\Pi$ is $(a-c)^2/(b(k+1)^2)$. If he joins a coalition, say $S_i$, then the total number of players becomes $k-1$ and the payoff of the deviating player is now $\frac{(a-c)^2}{(s_i+1)bk^2}$. This is not a successful deviation if $(k+1)^2 \leq (s_i+1)k^2$, which is equivalent to $s_i k^2 \geq 2k+1$. If $k \geq 3$, this is satisfied. For $k = 2$, the only case it would not be satisfied is if $s_i = 1$. But since $n \geq 3$, it is impossible that both $k = 2$ and $s_i = 1$. $\square$

Next, we consider Type 3 deviations for singleton players.

**Lemma 7.3.5** *Consider a partition $\Pi = (S_1, ..., S_k)$, with $k \geq 2$. Suppose that $\Pi$ contains $k_1$ singleton coalitions with $k_1 \geq 2$, and $k_2$ non-singleton ones ($k_1 + k_2 = k$). The merge of the $k_1$ singletons is not a successful deviation iff $k_1 \leq (k_2+1)^2$.*

**Proof:** The $k_1$ singletons receive in $\Pi$ a payoff of $(a-c)^2/(b(k_1+k_2+1)^2)$. After the merge, their payoff will be $(a-c)^2/(k_1 b(k_2+2)^2)$. Hence, the merge will not be successful, iff $(a-c)^2/(b(k_1+k_2+1)^2) \geq (a-c)^2/(k_1 b(k_2+2)^2)$. As a result, we have that:

$$\begin{aligned}
(a-c)^2/(b(k_1+k_2+1)^2) &\geq (a-c)^2/(k_1 b(k_2+2)^2) \Leftrightarrow \\
k_1(k_2+2)^2 &\geq (k_1+k_2+1)^2 \Leftrightarrow \\
k_1(k_2+1)^2+2k_1(k_2+1)+k_1 &\geq k_1^2+(k_2+1)^2+2k_1(k_2+1) \Leftrightarrow \\
(k_1-1)(k_2+1)^2 &\geq k_1^2-k_1 \Leftrightarrow \\
(k_2+1)^2 &\geq k_1
\end{aligned}$$

□

Finally we show that for ensuring stability there is no need to consider any other Type 3 deviation of smaller coalitions.

**Lemma 7.3.6** *Consider a partition $\Pi = (S_1, ..., S_k)$, with $k \geq 2$ and suppose that it contains $k_1$ singleton coalitions with $k_1 \geq 2$, and $k_2$ non-singleton ones. There is a successful Type 3 deviation iff the merge of all $k_1$ singletons is a successful deviation.*

**Proof:** One direction is trivial, namely if the merge of all $k_1$ singletons is a successful deviation. For the reverse direction, suppose there is a successful type 3 deviation which is not the merge of all the $k_1$ singletons. Let $m$ be the number of players who merge and suppose $2 \leq m < k_1$. By arguing as in Lemma 7.3.5, we get that in order for the deviation to be successful, it should hold that $(k_1+k_2+1)^2 > m(k_1+k_2-m+2)^2$. Let $\lambda = k_2+1$ and $\theta = \lambda + k_1 - m$. Restating the condition in terms of $\lambda$ and $\theta$ we get:

$$\begin{aligned}
(k_1+\lambda)^2 = (\theta+m)^2 &> m(\theta+1)^2 \Rightarrow \\
\theta^2+m^2+2\theta m &> m\theta^2+2\theta m+m \Rightarrow \\
m &> \theta^2
\end{aligned}$$

This implies that $k_1 > m > (\lambda + k_1 - m)^2 > \lambda^2 = (k_2+1)^2$. By Lemma 7.3.5, this means that the merge of all $k_1$ singletons is also a successful deviation and the proof is complete. $\square$

All the above can be summarized as follows:

**Corollary 7.3.7** *Consider a partition* $\Pi$. *For* $n \leq 2$, $\Pi$ *is stable iff it is the grand coalition. For* $n \geq 3$, *suppose* $\Pi = (S_1, ..., S_k)$ *with* $k_1$ *singleton coalitions and* $k_2$ *non-singleton ones. Then* $\Pi$ *is stable iff it is either the grand coalition or the following hold:*

- $k_1 \leq (k_2+1)^2$.
- *For every non-singleton coalition* $S_i$, $s_i \geq k^2$.

Having acquired a characterization of the stable partitions, we are ready to analyze the (pure) price of anarchy of the restricted coalition formation game on top of a symmetric Cournot oligopoly.

**Theorem 7.3.8** *The price of anarchy of the restricted coalition formation game under symmetric Cournot oligopoly is* $O(n^{2/5})$, *where* $n$ *is the total number of players.*

**Proof:** By Lemma 7.2.3 we have that the cost of a partition with $k$ coalitions is $O(k)$ times worse than that of optimal and that this ratio is increasing with $k$. Hence, the worst stable partition for any instance of a coalition formation game is the one with the maximal number of coalitions. Therefore, in order to bound the price of anarchy it suffices to find an upper bound on the number of coalitions $k$

of a stable coalition structure. By Corollary 7.3.7, we thus need to optimize the following mathematical program:

$$\begin{aligned}
\max \quad & k \\
\text{s.t.} \quad & k = k_1 + k_2 \\
& k_1 + \textstyle\sum_{i=1}^{k_2} s_i = n \\
& k_1 \leq (k_2 + 1)^2 \\
& s_i \geq k^2 & \forall i \in \{1, ..., k_2\} \\
& k, k_1, k_2 \geq 0
\end{aligned}$$

We have that

$$n = k_1 + \sum_{i:s_i \geq 2} s_i \geq k_1 + k_2 \min_{i:s_i \geq 2} s_i \geq k_1 + k_2(k_1 + k_2)^2 \tag{7.1}$$

Trivially this implies that, $n \geq k_2^3$ or equivalently $k_2 \leq n^{\frac{1}{3}}$. Also, since by Corollary 7.3.7 we have that $k_2 \geq \sqrt{k_1} - 1$, equation (7.1) implies that $n \geq k_1 + (\sqrt{k_1} - 1)(k_1 + \sqrt{k_1} - 1)^2$. This is equivalent to $n \geq k^{\frac{5}{2}} + k^2 - 3k^{\frac{3}{2}} + 3k^{\frac{1}{2}} - 1$. As a result $k_1 \leq \max\{9, n^{\frac{2}{5}}\}$. Putting all these together we derive the desired bound on $k$ and hence on the price of anarchy as well. $\square$

Finally, we come back to the original coalition formation game of Definition 7.2.2. Since in the game we are interested in, we have only enlarged the set of possible deviations with regard to the restricted coalition formation game, the set of stable partitions can only decrease. As a result, the price of anarchy for the original game is also $O(n^{2/5})$. This completes the proof for the upper bound of Theorem 7.3.1.

### 7.3.2 The Proof of the Lower Bound

The lower bound is obtained by the construction in the following Lemma:

**Lemma 7.3.9** *For any $N$, let $n$ be the number: $n = \lceil 4N^{4/5} \rceil \lfloor N^{1/5} \rfloor + \lfloor N^{2/5} \rfloor$. Consider a game on $n$ players and a partition of the $n$ players consisting of $k_1 = \lfloor N^{2/5} \rfloor$ singletons and $k_2 = \lfloor N^{1/5} \rfloor$ coalitions of size $s = \lceil 4N^{4/5} \rceil$ each. This coalition structure is stable for the Cournot coalition formation game.*

**Proof:** Here, we provide the proof of lemma 7.3.9. For any $N$, let $n$ be the number: $n = \lceil 4N^{4/5} \rceil \lfloor N^{1/5} \rfloor + \lfloor N^{2/5} \rfloor$. For a game on $n$ players, let $\Pi$ be a partition of the $n$ players into $k_1 = \lfloor N^{2/5} \rfloor$ singletons and $k_2 = \lfloor N^{1/5} \rfloor$ coalitions of size $s = \lceil 4N^{4/5} \rceil$ each. We will prove the desired Lemma, by exhaustively showing that for this coalition structure all types of deviating moves are not successful. We commence our analysis by looking into deviations of type 1.

**Proposition 7.3.10** *There exists no successful deviation of Type 1 for* $\Pi$.

**Proof:** A deviation of type 1 is one in which a subset $S_i'$ of a current coalition $S_i$ decides to deviate and form a new coalition. The rest of the members of the original coalition (i.e. $S_i/S_i'$) dissolve into singletons. In our instance, the only such possible deviation is one where a number of $s' < s$ players deviate from a coalition of size $s$ and form a new coalition with the remaining $s - s'$ members of the coalitions breaking off into singletons. The initial payoff of the players in that coalitions were $(a-c)^2/(bs(k+1)^2)$. After the deviation the payoffs for the deviating players will be equal to $(a-c)^2/(bs'(k+(s-s')+1)^2)$. In order for this coalition to not be successful it suffices that we have:

$$
\begin{aligned}
(a-c)^2/\Big(bs(k+1)^2\Big) &\geq (a-c)^2/\Big(bs'(k+(s-s')+1)^2\Big) \Leftrightarrow \\
s'\Big(k+1+(s-s')\Big)^2 &\geq s(k+1)^2 \Leftrightarrow \\
s'\Big((k+1)^2+2(k+1)(s-s')+(s-s')^2\Big) &\geq s(k+1)^2 \Leftrightarrow \\
2(k+1)s'(s-s')+s'(s-s')^2 &\geq (s-s')(k+1)^2 \Leftrightarrow \\
2(k+1)s'+s'(s-s') &\geq (k+1)^2 \Leftrightarrow \\
-s'^2+\big(2(k+1)+s\big)s' &\geq (k+1)^2 \Leftrightarrow
\end{aligned}
$$

Given that $s' \in [1, s-1]$ the minimum of the quantity $X = -s'^2+\big(2(k+1)+s\big)s'$ is achieved at either $s' = 1$ or $s' = s-1$. If $s' = 1$, then $X = s+2k+1 \geq k^2+2k+1 = (k+1)^2$. Similarly, if $s' = s-1$, then $X = s'\big(-s'+2(k+1)+s\big) = (s-1)(2k+3) \geq (k+1)^2$. □

Similarly, we continue to argue that no deviations of type 2 are successful either.

**Proposition 7.3.11** *There exists no successful deviation of Type 2 for* $\Pi$*.*

**Proof:** A deviation of type 2 is one in which either a singleton decides to join another coalition $S_j$ or a subset $S_i'$ of a current coalition $S_i$ (with $|S_i'| \leq |S_i|-1$) decides to leave its current coalition $S_i$ and join another coalition $S_j$. By Lemma 7.3.4 the singleton players do not have an incentive to perform a Type 2 deviation. Hence in our instance, the only such possible deviation is one where a number of $s' < s$ players deviate from a coalition and join either a singleton or another coalition of size $s$. In order to prove that such deviations are not successful, we will argue that the members of the receiving coalition $S_j$ (regardless of whether this is a singleton or a coalition of size $s$) are better off in the initial coalition structure. Indeed, in the new coalition structure, we will have $k + s - s' - 1 \geq k$ coalitions. Similarly, the size of the coalition $S_j$ that receives the deviating players will strictly increase. Therefore, the resulting payoffs for the members of coalition $S_j$ will decrease in any such deviation and as a result no such deviation can be successful. □

We conclude, by proving that deviations of type 3 cannot be successful. Deviations of type 3 allow for any number of coalitions to unite and form a new coalition. We exhaustively examine all possible compositions of the deviating sets and prove that there exists no set of coalitions for which a deviation of type 3 is profitable. Specifically, we have to examine the following cases for our instance:

1) All the coalitions that participate in the move are singletons.

2) All the coalitions that participate in the move are of size $s$.

3) Some of the participating coalitions are singletons and some are of size $s$.

It is straightforward to show that for the first two cases a deviation of type 3 cannot be successful. For the third case, we show that if any such deviation is successful, then there exists a successful deviation of type 3 in which all singletons participate. The proof concludes by showing that no such deviation can be successful either.

**Proposition 7.3.12** *There exists no successful deviation of Type 3 for* $\Pi$*.*

**Proof:** We will show that no deviating moves of type 3 are successful, regardless of the composition of sets of coalitions participating in the move.

1) All the coalitions that participate in the move are singletons.

This essentially corresponds to a move of type 3 of the restricted coalition formation game of Section 7.3.1. By Lemma 7.3.5 and Lemma 7.3.6 it suffices to show that $k_1 \leq (k_2+1)^2$. This is true in our instance, since $\lfloor N^{2/5} \rfloor \leq N^{2/5} = (N^{1/5})^2 < (\lfloor N^{1/5} \rfloor + 1)^2$.

2) All the coalitions that participate in the move are of size $s$.

If we have $m$ sets of size $s$ uniting to form a single coalition, then the original payoff of the players were $(a-c)^2/s(k+1)^2$, whereas their utilities in the new coalition will be equal to $(a-c)^2/ms(k-m+2)^2$. Equivalently, if we denote $\lambda = k+1$, the respective costs will be expressed as $(a-c)^2/s\lambda^2$ and $(a-c)^2/ms(\lambda - m+1)^2$. In order to have stability it suffices that we have :

$$\begin{aligned}
(a-c)^2/s\lambda^2 &\geq (a-c)^2/ms(\lambda-m+1)^2 \Leftrightarrow \\
ms(\lambda-m+1)^2 &\geq s\lambda^2 \Leftrightarrow \\
m(\lambda^2+m^2+1-2m\lambda+2\lambda-2m) &\geq \lambda^2 \Leftrightarrow \\
m\lambda^2+m^3+m-2m^2\lambda+2m\lambda-2m^2 &\geq \lambda^2 \Leftrightarrow \\
(m-1)\lambda^2+(-2m^2+2m)\lambda+m^3-2m^2+m &\geq 0 \Leftrightarrow \\
\lambda^2-2m\lambda+m(m-1) &\geq 0 \Leftrightarrow \\
(\lambda-m-\sqrt{m})(\lambda-m+\sqrt{m}) &\geq 0 \Leftrightarrow \qquad (7.2)
\end{aligned}$$

However, $\lambda = k+1 = \lfloor N^{2/5} \rfloor + \lfloor N^{1/5} \rfloor + 1$, whereas $m \leq k_2 = \lfloor N^{1/5} \rfloor$. Hence, $\lambda \geq m + \sqrt{m}$, for all $N$.

3) Some of the participating coalitions are singletons and some are of size $s$.

We will start the analysis of this case by showing that if any such deviation is successful, then there exists a successful deviation of type 3 in which all the singletons participate.

Indeed, suppose that we have a profitable move of type 3 such that $m_1$ singletons and $m_2$ coalitions of size $s$ participate, where $m_1 < k_1$. Each player, in the resulting coalition earns $(a-c)^2/b(m_1+m_2 s)(k-m_1-m_2+2)^2$ and since this is a successful move, this payoff dominates the original payoff of the singleton players. Now, swap each coalitions of size $s$ with singletons that do not participate currently in this move. If $m_1 + m_2 \leq k_1$, we will end up swapping all the non-singleton coalitions, so finally the sets that participate in the move are all singletons. As a result of this swapping the new payoffs for the players will be equal to $(a-c)^2/b(m_1+m_2)(k-m_1-m_2+2)^2$, which is strictly better that the profits when coalitions of size $s$ were included. Hence, this move will be successful as well, but this is not possible, since this is essentially a move of type 3 of the restricted coalition formation game which has already been ruled out. The only case left, is the one where we have a move of type 3 with $m_1$ singletons and $m_2$ coalitions of size $s$ and $m_1 + m_2 > k_1$. As a result of the same swap, we define a move of type 3 with $k_1$ singletons and $m_1+m_2-k_1$ coalitions of size $s$. The payoff of the players in the resulting coalition will be $(a-c)^2/\big(b(k+(m_1+m_2-k_1)s)(k-m_1-m_2+2)^2\big)$, which dominates the payoffs of the first move of type 3 $(a-c)^2/\big(b(m_1+m_2 s)(k-m_1-m_2+2)^2\big)$. Hence, the new utilities of the singletons are higher than their utilities before the deviation and as result this deviation will be successful as well. So, if there exists a successful deviation of type 3, then there exists a successful deviation of type 3 where all the singletons participate (as well as some coalitions of size $s$).

At this point, in order to show that no moves of type 3 are successful for our family of examples, it suffices to show that there exist no successful moves of type 3 where we have all $k_1$ singletons participating as well $m \geq 1$ coalitions of size $s$. We will show that this move is not successful by showing that the singletons have no incentive to participate in this move.

Indeed, currently the singleton players enjoy a payoff of $(a-c)^2/\big(b(k_1+k_2+1)^2\big)$. After the deviation, their payoff will be equal to $(a-c)^2/\big(b(k_1+ms)(k_2-m+2)^2\big)$. In order for this move to not be successful, it suffices to have that $(k_1+ms)(k_2-m+2)^2 \geq (k_1+k_2+1)^2$.

$$
\begin{aligned}
(k_1+ms)(k_2-m+2)^2 &> 4s \\
&= 4\lceil 4N^{4/5} \rceil \\
&> (\lfloor N^{2/5} \rfloor + \lfloor N^{1/5} \rfloor + 1)^2 \\
&= (k_1+k_2+1)^2
\end{aligned}
$$

□

By combining propositions 7.3.10, 7.3.11 and 7.3.12, we have finally proven lemma 7.3.9. □

Since the total number of coalitions in the construction is $k = k_1+k_2 \geq N^{2/5} = \Omega(n^{2/5})$, by Lemma 7.2.3, we obtain the desired lower bound for any number of players $n$.

**Corollary 7.3.13** *For any number of players $n$, there exist stable partitions with cost $\Omega(n^{\frac{2}{5}})$ the cost of the optimal partition.*

## 7.4 Coalition Formation under No-regret

So far, given a partition $\Pi$, we assign to each coalition $S_i$ value equal to its uniquely defined utility at the Nash equilibria of the Cournot game with a fixed coalition partition $\Pi$. The reasoning behind this is that if the coalition partition $\Pi$ is not permeable, then the players/coalitions will hopefully reach a Nash equilibrium and hence their uniquely defined utility at it, is a good estimator of how much they value their current coalition partition.

In this paragraph, we show that we can significantly weaken the assumption that the players/coalitions will reach an equilibrium. In fact, we will show that if the coalitions participate in the Cournot oligopoly repeatedly in a fashion that minimizes their long term regret then their average utility will converge to their levels at Nash equilibria. Intuitively, having no-regret means that no deviating action would significantly improve the utility of the player. This notion captures successful long-term behavior and can be achieved in practice by several natural learning algorithms ([20] and references therein). Putting all these together, we have that the values assigned to coalitions by our model, are in excellent agreement with the average utilities they would actually receive by participating repeatedly (and successfully) in the market. Since the setting of oligopolies markets is in its nature repeated, this observation significantly strengthens the justification of our model.

**Theorem 7.4.1** *Consider a Cournot oligopoly game with a fixed coalition partitioning of the $n$ players in $k$ coalitions $\Pi = (S_1, ..., S_k)$. If all $k$ (super-)players employ no-regret strategies, then their average utilities converge to their Nash levels.*

**Proof:** As we have argued in the proof of lemma 7.2.2, given any strategy profile of the Cournot game with coalitions the superplayers' payoff is exactly equal to the payoff of regular players that play the same game with production levels equal to the aggregate production of the corresponding superplayers.

If the superplayers experience no-regret, that means that deviating to any fixed strategy only decreases their average payoff[1]. Specifically, since their payoff on any day depends only on the aggregate production, the average utility of any player only decreases when a player deviates and only plays strategies of a specific production level. Now, consider the same history of play but replace every superplayer by a simple player that produces on each day $t$ quantity equal to the aggregate production quantity of the respective coalition on the same day. We will show that this history is also of no-regret. Indeed, for each player any deviation to a constant production level will only decrease their average utility[1], as is true for the superplayers. But, in the case of simple Cournot games, as was shown in [31], if all players experience no-regret then the average utilities of the players converge to their levels at Nash equilibrium. Translating this result to the history of the Cournot game with coalitions, we have that the average utilities of the superplayers converge to the levels of the unique Nash of the same Cournot game with $k$ simple players. As was shown in lemma 7.2.2, these levels are also in agreement with the Nash equilibria of the Cournot game with coalitions. $\square$

[1]More accurately, it means that at times goes to infinity the limsup of the difference between the average utility for any fixed deviation and the experienced average utility will be at most 0.

## 7.5 Discussion and Future Work

We have introduced a model for coalition formation and have studied the inefficiency of stable coalitions. Our main finding is that the price of anarchy can be greatly reduced once players are allowed to start forming agreements with other players. We believe that this opens up a very promising avenue for future research. Our approach requires that we define, in a sensible manner, the utility of a coalition at a given coalition structure. In the case of Cournot games, this was easy to do because all Nash equilibria of the coalition-game are payoff unique. One possible extension of our work is to asymmetric Cournot games, which still have unique Nash equilibria. However, in the case of asymmetric producers we would need a new way of dividing the utility of a coalition amongst its members, since equal sharing does not reflect the true contributions of the members anymore. Other candidate games that we are planning to study include socially concave games and routing games.

## BIBLIOGRAPHY


[1] J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *COLT*, pages 263–274, 2008.

[2] S. Airiau. Tutorial notes of the 11th European Agent Systems Summer School (EASSS-2009).

[3] E. Akin. *The Geometry of Population Genetics*. Springer-Verlag, 1979.

[4] R. Amir. Cournot oligopoly and the theory of supermodular games. *Games and Economic Behavior*, 1996.

[5] Nir Andelman, Michal Feldman, and Yishay Mansour. Strong price of anarchy. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 189–198, 2007.

[6] S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: A meta-algorithm and applications. *Preprint*, 2005.

[7] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002. (A preliminary version appeared in FOCS 1995 as "Gambling in a rigged casino: The adversarial multi-armed bandit problem".).

[8] J. Aumann, R. and B. Peleg. Von neumann-morgenstern solutions to cooperative games without side payments. *Bulletin of the American Mathematical Society*, 66:173–179, 1960.

[9] R. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1:67–96, 1974.

[10] B. Awerbuch, R. Khandekar, and S. Rao. Distributed algorithms for multicommodity flow problems via approximate steepest descent framework. In *SODA*, pages 949–957, 2007.

[11] K. Basu and J.W. Weibull. Strategy subsets closed under rational behavior. *Economics Letters*, 36:141–146, 1991.

[12] M.R. Baye and J. Morgan. A folk theorem for one-shot bertrand games. *Economic Letters*, 65(1):59–65, 1999.

[13] Michael Benisch, George Davis, and Tuomas Sandholm. Algorithms for rationalizability and curb sets. In *IN PROCEEDINGS OF AAAI06*, pages 598–604, 2006.

[14] J. Bertrand. Theorie mathematique de la richesse sociale. *Journal des Savants.*, 67:499–508, 1883.

[15] A. Blum, E. Even-Dar, and K. Ligett. Routing without regret: on convergence to nash equilibria of regret-minimizing algorithms in routing games. In *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, pages 45–52. ACM Press, 2006.

[16] A. Blum, M. Hajiaghayi, K. Ligett, and A. Roth. Regret minimization and the price of total anarchy. In *STOC*, pages 373–382, 2008.

[17] A. Blum and Y. Mansour. From external to internal regret. In *COLT*, 2005.

[18] A. Blum and Y. Mansour. Learning, regret minimization and equilibria. *Algorithmic Game Theory N. Nisan, T. Roughgarden, E. Tardos and V. Vazirani (eds)*, 2007.

[19] N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

[20] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning and Games*. Cambridge University Press, 2006.

[21] G. Chalkiadakis. *A Bayesian Approach to Multiagent Reinforcement Learning and Coalition Formation under Uncertainty*. PhD thesis, Department of Computer Science, University of Toronto, Canada, 2007.

[22] Georgios Chalkiadakis, Edith Elkind, Maria Polukarov, and Nicholas R. Jennings. The price of democracy in coalition formation. In *International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 401–408, 2009.

[23] P. Chander and H. Tulkens. The core of an economy with multilateral environmental externalities. *International Journal of Game Theory*, 26:379–401, 1997.

[24] Xi Chen and Xiaotie Deng. Settling the complexity of two-player nash equilibrium. In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium on*

*Foundations of Computer Science*, pages 261–272, Washington, DC, USA, 2006. IEEE Computer Society.

[25] A. A. Cournot. *Recherches sur les Principes Mathmatiques de la Thorie des Richesses.(English translation: Researches into the Mathematical Principles of the Theory of Wealth).* New York: Macmillan, 1838.

[26] P. Cramton and T. Palfrey. Cartel enforcement with uncertainty about costs. *International Economic Review*, 31(1):17–47, 1990.

[27] C. Daskalakis, P. Goldberg, and C. Papadimitriou. The complexity of computing a nash equilibrium. In *STOC*, 2006.

[28] T. Dieckmann and U. Schwalbe. Dynamic Coalition Formation and the Core, 1998. Economics Department Working Paper Series, Department of Economics, National University of Ireland - Maynooth.

[29] M. Dufwenberg and U. Gneezy. Price competition and market concentration: An experimental study. *International Journal of Industrial Organization 18*, pages 7–22, 2000.

[30] E. Even-Dar and Y. Mansour. Fast convergence of selfish rerouting. In *SODA*, 2005.

[31] E. Even-Dar, Y. Mansour, and U. Nadav. On the convergence of regret minimization dynamics in concave games. In *41st ACM Symposium on Theory of Computing. STOC*, 2009.

[32] S. Fischer, H. Räcke, and B. Vöcking. Fast convergence to wardrop equilibria by adaptive sampling methods. In *STOC*, 2006.

[33] S. Fischer, H. Räcke, and B. Vöcking. Fast convergence to wardrop equilibria by adaptive sampling methods. In *STOC*, 2006.

[34] S. Fischer and B. Vöcking. On the evolution of selfish routing. In *12th European Symposium on Algorithms (ESA '04)*, pages 323–334. Springer, 2004.

[35] S. Fischer and B. Vöcking. On the evolution of selfish routing. In *ESA*, 2004.

[36] A. Flaxman, A. T. Kalai, and B. McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the*

*Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 385–394, 2005.

[37] D. Foster and S. M. Kakade. Deterministic calibration and nash equilibrium. In *Conference on Learning Theory (COLT)*, pages 33–48, 2004.

[38] D. Foster and R. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 24:73–96, 1997.

[39] D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 21:40–55, 1997.

[40] D. Foster and H. P. Young. Learning, hypothesis testing, and nash equilibrium. *Games and Economic Behavior*, 96:45–73, 2003.

[41] D. Foster and H. P. Young. Regret testing: Learning to play nash equilibrium without knowing you have an opponent. *Theoretical Economics*, 1(3):341–367, 2006.

[42] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119139, 1997.

[43] Y. Freund and R. E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29:79–103, 1999.

[44] D. Fudenberg and D. K. Levine. *The theory of learning in games*. MIT press, 1998.

[45] F. Germano and G. Lugosi. Global nash convergence of foster and youngs regret testing. *Games and Economic Behavior*, 60:135–154, 2007.

[46] M. Goemans, V. Mirrokni, and A. Vetta. Sink equilibria and convergence. In *FOCS*, 2005.

[47] X. Guo and H. Yang. The price of anarchy of cournot oligopoly. In *Workshop on Internet and Network Economics (WINE)*, pages 246–257, 2005.

[48] J. Hannan. Approximation to bayes risk in repeated plays. *In M. Dresher, A. Tucker, and P. Wolfe eds, Contributions to the Theory of Games*, 3:79–139, 1957.

[49] A. Harberger. Monopoly and resource allocation. *Amer. Econ. Review*, 44:77–87, 1954.

[50] J. C. Harsanyi and R. Selten. *A General Theory of Equilibrium Selection in Games*. The MIT Press, 1988.

[51] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

[52] S. Hart and A. Mas-Colell. Stochastic uncoupled dynamics and nash equilibrium. *Games and Economic Behavior*, 57:286–303, 2006.

[53] Ara Hayrapetyan, Éva Tardos, and Tom Wexler. The effect of collusion in congestion games. In *ACM Symposium on Theory of Computing (STOC)*, pages 89–98, 2006.

[54] E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

[55] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[56] S. Hurkens. Learning by forgetful players. *Games and Economic Behavior*, 11(2):304–329, November 1995.

[57] J. Josephson. Stochastic better-reply dynamics in games. Working Paper Series in Economics and Finance 400, Stockholm School of Economics, July 2000.

[58] R. Kleinberg, G. Piliouras, and E. Tardos. Multiplicative updates outperform generic no-regret learning in congestion games. In *STOC*, 2009.

[59] M. Klimm and J.W. Weibull. Finding all minimal curb sets. *Working Paper*, 2009.

[60] J. Kluberg and G. Perakis. Generalized quantity competition for multiple products and loss of efficiency. In *46th Annual Allerton Conference on Communication, Control, and Computing*, pages 930 – 936, 2008.

[61] L. Kóczy. A recursive core for partition function form games. *Theory and Decision*, 63:41–51, 2007.

[62] E. Koutsoupias and C. H. Papadimitriou. Worst-case equilibria. In *16th Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413, 1999.

[63] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–260, 1994.

[64] L. Liu. Correlated equilibrium of cournot oligopoly competition. *Journal of Economic Theory*, 68(2):544–548, 1996.

[65] A. Mas-Colell, J. Green, and M. D. Whinston. *Microeconomic Theory*. Oxford University Press, 1995.

[66] A. Mas-Colell, M.D. Winston, and J.R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

[67] J. Maynard-Smith and G. R. Price. The logic of animal conflict. *Nature*, 246(15), 1973.

[68] P. Milgrom and J. Roberts. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica*, 58(6):1255–77, November 1990.

[69] P. Milgrom and J. Roberts. Adaptive and sophisticated learning in repeated normal form games. *Games and Economic Behavior*, pages 82–100, February 1991.

[70] J. W. Milnor. *Topology from the differentiable viewpoint*. Princeton Landmarks in Mathematics. Princeton University Press, 1997.

[71] V. Mirrokni and A. Vetta. Convergence issues in competitive games. In *RANDOM-APPROX*, 2004.

[72] M. Mitzenmacher. How useful is old information? In *16th Ann. ACM SIGACT- SIGOPS Symp. on Principles of Distributed Computing (PODC)*, pages 83–91, 1997.

[73] D. Monderer and L. S. Shapley. Fictitious play property for games with identical interests. *J. Economic Theory*, 68:258–265, 1996.

[74] D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, 14:124–143, 1996.

[75] U. Nadav and G. Piliouras. No-regret learning in oligopolies: Cournot vs bertrand. In *Submitted*, 2010.

[76] J. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[77] C. H. Papadimitriou. The search for equilibrium concepts. In *SAGT*, pages 1–3, 2008.

[78] L. Perko. *Differential Equations and Dynamical Systems*. Springer, 1991.

[79] V. Pruzhansky. Strategy subsets closed under rational behavior. *International Journal of Game Theory*, 32(2):205–210, 2003.

[80] D. Ray and R Vohra. Equilibrium binding agreements. *Journal of Economic Theory*, 73:30–78, 1997.

[81] K. Ritzberger and J. Weibull. Evolutionary selection in normal-form games. *Econometrica*, 63:1371–1399, 1995.

[82] J.B. Rosen. Existense and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33:520–534, 1965.

[83] R. W. Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.

[84] T. Roughgarden. Routing games. *Algorithmc Game Theory (eds. N. Nisan, T. Roughgarden, E. Tardos, V. Vazirani)*, 2007.

[85] T. Roughgarden. Intrinsic robustness of the price of anarchy. In *STOC*, 2009.

[86] T. Roughgarden and E. Tardos. How bad is selfish routing? *J. ACM*, 49(2):236–259, 2002.

[87] T. Roughgarden and É. Tardos. Bounding the inefficiency of equilibria in nonatomic congestion games. *Games and Economic Behavior*, 2004.

[88] Jordan J. S. Three problems in learning mixed-strategy nash equilibria. *Games and Economic Behavior*, 5(3):368–386, July 1993.

[89] L. Samuelson and J. Zhang. Evolutionary stability in asymmetric games. *Journal of Economic Theory*, 57:363–391, 1992.

[90] C. Sanchirico. A probabilistic model of learning in games. *Econometrica*, 64:375–1394, 1996.

[91] W. H. Sandholm. Potential games with continuous player sets. *Journal of Economic Theory*, 97:81–108, 2001.

[92] W. H. Sandholm. *Population Games and Evolutionary Dynamics*. MIT Press, 2010.

[93] I. R. Shafarevich. *Basic Algebraic Geometry, volume 1.* Springer-Verlag, 1994.

[94] P. Shenoy, P. Coalition formation: a game theoretical approach. *International Journal of Game Theory*, 8(3):133–164, 1979.

[95] J. N. Sheth and R. S. Sisodia. *The Rule of Three: Surviving and Thriving in Competitive Markets.* Free Press, 2002.

[96] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, Cambridge, UK, 2009.

[97] J. M. Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.

[98] J. M. Steele. An efron-stein inequality for nonsymmetric statistics. *Annals of Statistics*, 14(2):753–758, 1986.

[99] S. Suetens and J. Potters. Bertrand colludes more than Cournot. *Experimental Economics*, 10:71–77, 2007.

[100] R. D. Theocharis. On the stability of the cournot solution on the oligopoly problem. *The Review of Economic Studies.*, 27(2):133–134, 1960.

[101] X. Vives. *Oligopoly Pricing: Old Ideas and New Tools*. The MIT Press, 2001.

[102] B. Vöcking. Selfish load balancing. *Algorithmc Game Theory (eds. N. Nisan, T. Roughgarden, E. Tardos, V. Vazirani)*, 2007.

[103] J. Weibull. *Evolutionary Game Theory*. MIT Press, 1997.

[104] J. Wu. Correlated equilibrium of bertrand competition. In *4th International Workshop on Internet and Network Economics, WINE*, 2008.

[105] S. Yi. On the existence of a unique correlated equilibrium in cournot oligopoly. *Economic Letters*, 54:235–239, 1997.

[106] H.P. Young. *Strategic Learning and Its Limits.* Oxford University Press, Oxford, 2004.

[107] P.H Young. *Individual Strategy and Social Structure*. Princeton University Press, Princeton, NJ, 1998.

[108] A. Zapechelnyck. Limit behavior of no-regret dynamics. *Working Paper*, 2009.

[109] M. Zinkevich. Online convex programming and generalized infitesimal gradient ascent. In *ICML*, pages 928–936, 2003.