UNIVERSITY OF CALIFORNIA,
IRVINE

Learning Dynamics and Convergence Barriers in Structured Games and Optimization

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Computer Science

by

Nikolas Patris

Dissertation Committee:
Assistant Professor Ioannis Panageas, Chair
Distinguished Professor Vijay V. Vazirani
Distinguished Professor David Eppstein

2026

DEDICATION

To Nefeli

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would first like to express my sincere gratitude to my advisors, Ioannis Panageas and Vijay V. Vazirani, for their guidance, support, and patience throughout my PhD. Their insight and encouragement have shaped not only this thesis, but also the way I think about research. Ioannis was always generous with his time, constantly available for discussion, and endlessly motivated by new problems and ideas. His excitement for research, and for discovering something new, has been a constant source of inspiration. I will also always remember Vijay's remarkable energy, his dedication to his work, and his attention to every detail, from the simplest to the most important. Above all, I am grateful to both of them for their support during more difficult periods, when their encouragement helped me continue my work with confidence.

I would also like to thank all of my coauthors and collaborators for the many discussions, ideas, and joint efforts that led to the results presented in this thesis. I have learned a great deal from working with them, and this thesis would not have been possible without their contributions. In particular, I would like to thank my dear friend Stratis, who was also my first collaborator. He had a great influence on me as a beginning PhD student. I was always impressed by his welcoming personality, his kindness, and the respect he showed to everyone around him. I am also grateful to my more recent collaborator, Ioannis A. I greatly admire the depth of his understanding, the clarity of his thinking, and, most importantly, how grounded he remains despite his remarkable ability. I feel fortunate to have worked with both of them.

I would also like to thank Dimitris Fotakis, who was a mentor and teacher throughout my undergraduate years. I have always admired his genuine dedication to his students, as well as the support and inspiration he offered to many of us.

I would like to thank my friends and colleagues from my time at UCI. I will always remember Stelios, who was not only a collaborator but also a good friend, and Renata, for the evenings at their home and the warmth they brought to my time in Irvine. I wish them and their family the very best, especially with the arrival of the newest member of their family. I would also like to thank Jingming, a fellow student and a friendly presence, with whom I shared several enjoyable lunches and conversations. I am also grateful to Parnian for her motivation and support; I will always be impressed by her perseverance and determination. Of course, I could not forget Vicky, for her kindness and willingness to help, despite the fact that we knew each other only for a short time during a turbulent period. She is always excited to create something new, and a person of great talent.

Lastly, I would like to honor the memory of Milena, who is no longer with us. From the first time we met, I was struck by her warm personality, the almost maternal care with which she spoke to me, and her willingness to share her thoughts and advice. She had a sincere understanding of other people's feelings, and her support was never limited to purely academic matters; it was deeply human. I am grateful to have known her.

I am thankful to my friends, old and new, for making these years more enjoyable, and for the moments away from work that made the difficult periods easier. I am especially grateful to our team from the Archimedes Research Center. Thank you to Vasilis, the youngest member of our group, whose maturity and clarity of thought often went well beyond his years, and beyond the narrow scope of research. To Andreas, for his relentless energy, humor, and constant teasing; to Vassilis, for his willingness to share his experience and perspective; to Alexandros, the quiet and hardcore mathematician of the group; to Giannis, the calmest and most relaxed among us, and perhaps a future stand-up comedian; to George, whose first conversation with me after I returned from Irvine I will never forget; and to Thodoris, the mastermind physicist and remarkably versatile researcher. I would also like to thank my undergraduate friends, whom fate brought back into my life once again: Chrysoula, Dimitra, Thodoris K., and Thomas. I am very glad that we reconnected and had the opportunity to spend more time together. Of course, I could not forget my youngest friend, Charalampos; from the first time we met, I was impressed by his friendliness and generosity. Finally, I would like to thank my close group of friends in Argyroupoli: Anastasis, Dimitris, Elina, Grigoris, and Tsampikos. We have shared so many moments over the years, and I look forward to many more in the years to come. I would also like to thank Melina, a person whose genuine love for others has always stood out to me.

Finally, I would like to thank my family: my twin sister Evita and my brother Anastasis, who were always keeping an eye on me and remained a constant source of support despite the distance. We have always been a support system for one another, and this is something I value deeply. I am also deeply grateful to my parents, Katerina and Ioannis, for their love, encouragement, and constant belief in me.

Last but not least, I would like to thank Nefeli, the person with whom I have shared this journey and so much of my life. Her support, patience, and presence have been invaluable throughout these years.

VITA

Nikolas Patris

EDUCATION

Doctor of Philosophy in Computer Science **2022–2026**
University of California, Irvine *Irvine, California*

Master of Science in Computer Science **2022–2024**
University of California, Irvine *Irvine, California*

Diploma in Electrical and Computer Engineering **2014–2021**
National Technical University of Athens *Athens, Greece*

RESEARCH EXPERIENCE

Graduate Research Assistant **2022–2026**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2022–2026**
University of California, Irvine *Irvine, California*

REFEREED CONFERENCE PUBLICATIONS

(Doubly) Exponential Lower Bounds for Follow the Regularized Leader in Potential Games — **ICML 2026, Spotlight**
Ioannis Anagnostides*, Ioannis Panageas, Nikolas Patris*, Tuomas Sandholm

Improved Bounds for Online Facility Location with Predictions — **AAAI 2025, Oral**
Dimitris Fotakis, Evangelia Gergatsouli, Themis Gouleakis, Nikolas Patris, Thanos Tolias †

Learning Nash Equilibria in Rank-1 Games — **ICLR 2024**
Nikolas Patris, Ioannis Panageas

Computing Nash Equilibria in Potential Games with Private Uncoupled Constraints — **AAAI 2024**
Nikolas Patris*, Ioannis Panageas, Stelios Andrew Stavroulakis*, Fivos Kalogiannis, Rose Zhang

Exponential Lower Bounds for Fictitious Play in Potential Games — **NeurIPS 2023**
Ioannis Panageas, Nikolas Patris*, Stratis Skoulakis*, Volkan Cevher

* Equal contribution, † Authors listed alphabetically.

ABSTRACT OF THE DISSERTATION

Learning Dynamics and Convergence Barriers in Structured Games and Optimization

By

Nikolas Patris

Doctor of Philosophy in Computer Science

University of California, Irvine, 2026

Assistant Professor Ioannis Panageas, Chair

This dissertation studies the computational and dynamical complexity of learning Nash equilibria in structured classes of games and related constrained optimization problems. A central goal in algorithmic game theory and multi-agent learning is to understand when natural learning dynamics converge efficiently to equilibrium, and when favorable structure still leaves fundamental quantitative barriers. The dissertation contributes to both sides of this question: it gives an efficient learning algorithm for rank-1 bimatrix games, proves exponential lower bounds for classical learning dynamics in potential games, and shows that follow-the-regularized-leader can fail to converge even asymptotically in constrained optimization.

The first part focuses on rank-1 bimatrix games, a natural class that strictly extends zero-sum games while retaining exploitable algebraic structure. Although rank-1 games fall outside the standard monotone frameworks used to analyze first-order methods, I show that one can still learn approximate Nash equilibria efficiently. The algorithm combines optimistic updates with a parameterized zero-sum reduction and a scalar consistency condition, yielding polynomial-time convergence to approximate equilibrium.

The second part turns to lower bounds for potential games. I show that fictitious play can require exponentially many rounds to reach an approximate Nash equilibrium even in two-

player identical-interest games. The proof constructs a payoff matrix that forces the dynamic to traverse a long sequence of strategy switches before reaching the equilibrium region.

The third part studies follow-the-regularized-leader and related no-regret dynamics in potential games. I establish positive structural properties, including potential monotonicity under suitable conditions, but also prove that FTRL can be exponentially slow in two-player potential games. In multiplayer potential games, I show that fictitious play can be doubly exponentially slow by connecting the construction to long paths in high-dimensional combinatorial structures.

The final part goes beyond finite games and studies FTRL as a first-order method for constrained nonconvex optimization. I construct an infinitely differentiable objective on a compact convex domain for which continuous-time FTRL has a Karush–Kuhn–Tucker gap bounded away from zero for all time, despite monotone improvement of the objective value.

Taken together, these results show that structural assumptions alone do not guarantee efficient learning. In some settings, structure can be exploited algorithmically; in others, asymptotic convergence and potential monotonicity hide severe computational barriers.

Chapter 1

Introduction

1.1 Motivation

Learning in games lies at the intersection of optimization, online learning, and economic behavior. In a repeated strategic interaction, each player updates its strategy from past observations while interacting with other players who are learning at the same time. The central hope is that such local adaptation eventually leads the collective system to a meaningful solution concept. The most prominent such concept is Nash equilibrium: a strategy profile at which no player can improve its payoff by a unilateral deviation.

From an algorithmic standpoint, this hope is delicate. Nash equilibria always exist in finite games, but computing them is hard in general, and the behavior of natural learning dynamics can be far more subtle than existence alone suggests. Even when a game class has favorable structure and even when a dynamic is known to converge asymptotically, the number of iterations needed to reach an approximate equilibrium may be prohibitively large. Thus, the relevant question is not only whether learning converges, but whether it converges at a rate that is meaningful from an algorithmic perspective.

This dissertation studies this quantitative question for structured games and for closely related constrained optimization problems. Its guiding theme is that structure is useful only when it can be converted into explicit convergence guarantees. In some settings, such as rank-1 bimatrix games, the underlying structure can be exploited to design efficient decentralized learning algorithms. In others, such as finite potential games, classical convergence and monotonicity properties may conceal exponential slowdowns. The final part of the dissertation moves beyond finite games and shows that, in constrained continuous optimization, even potential-like structure may fail to ensure the asymptotic convergence of continuous-time FTRL.

The thesis develops this theme through four main contributions. First, it gives a positive learning result for rank-1 bimatrix games, a natural class that strictly generalizes zero-sum games. Second, it proves that fictitious play can be exponentially slow in two-player potential games. Third, it shows that the same inefficiency extends to follow-the-regularized-leader (`FTRL`) and related no-regret dynamics in potential games, and that fictitious play can be doubly exponentially slow in multiplayer potential games. Fourth, it goes beyond finite games and shows that continuous-time `FTRL` can fail to converge even asymptotically to first-order stationary points in constrained optimization, despite monotone improvement of the objective value.

Taken together, these results separate three notions that are often conflated: structural tractability of the game, asymptotic convergence of a dynamic, and efficient last-iterate convergence to equilibrium. The dissertation shows that the first two do not automatically imply the third.

1.2 Background and scope

Most of the dissertation concerns finite normal-form games. In such a game, each player chooses an action from a finite set, possibly according to a mixed strategy, and receives a payoff determined by the joint action profile. A Nash equilibrium is a mixed strategy profile in which no player can gain by deviating unilaterally. Since exact equilibrium computation is often too demanding, the natural algorithmic target is an approximate Nash equilibrium, where every unilateral deviation improves utility by at most a prescribed accuracy parameter.

The first structural class studied in the thesis is the class of rank-1 bimatrix games. A two-player bimatrix game $(\mathbf{A}, \mathbf{B})$ is rank-1 when the matrix $\mathbf{A}+\mathbf{B}$ has rank one. Zero-sum games correspond to rank zero, so rank-1 games form the simplest nonzero extension of the zero-sum setting. They retain enough algebraic structure to admit specialized equilibrium algorithms, but they already fall outside the standard convex-concave and monotone frameworks that support many first-order convergence analyses. This makes them a useful testing ground for the possibility of efficient equilibrium learning beyond zero-sum games.

The second structural class is the class of potential games. In a potential game, all players' incentives are aligned with a single scalar potential function. This class is often considered favorable for learning because many natural dynamics have classical convergence guarantees in potential games, and Nash equilibria correspond to local optimality conditions for the potential. The results of this dissertation show that this favorable qualitative picture can be quantitatively misleading: the presence of a potential function does not by itself guarantee efficient convergence.

The final chapter broadens the perspective from finite potential games to constrained optimization. Potential games can be viewed as a special case of potential systems, where the potential function is multilinear and the feasible set is a product of simplices. The constrained-optimization viewpoint makes it possible to ask whether `FTRL` should be under-

stood as a genuine first-order optimization method, in the same way that mirror descent and gradient descent are. The answer obtained here is negative: `FTRL` may preserve monotonic improvement of the objective while failing to drive the Karush–Kuhn–Tucker gap to zero.

The learning dynamics studied in the thesis are among the most classical in game theory and online learning. Fictitious play is a canonical best-response process based on empirical frequencies. Follow-the-regularized-leader is a foundational no-regret framework that includes multiplicative weights update and is closely related to mirror descent. Optimistic first-order methods appear in the positive result for rank-1 games. The common question across these dynamics is whether their natural local update rules lead to efficient equilibrium computation.

1.3 Main questions

The dissertation is organized around the following questions.

1. Beyond zero-sum and monotone settings, are there natural classes of games in which Nash equilibria can still be learned efficiently by decentralized dynamics?

2. When a learning dynamic is known to converge asymptotically in a structured class of games, does this imply a polynomial convergence rate to approximate equilibrium?

3. What is the precise gap between no-regret guarantees, convergence of time averages, and last-iterate convergence to Nash equilibrium?

4. Can `FTRL` be interpreted as a reliable first-order optimization method in constrained nonconvex problems, or can its history dependence prevent convergence even in potential systems?

The four main chapters answer these questions from complementary directions. Chapter 3 gives a positive answer for rank-1 games. Chapters 4 and 5 give negative answers for fictitious play and `FTRL` in potential games. Chapter 6 strengthens the negative picture by showing that, outside the finite-game setting, `FTRL` can fail to converge even asymptotically in constrained optimization.

1.4 Overview of contributions

Efficient learning in rank-1 games. Chapter 3 studies the problem of learning Nash equilibria in rank-1 bimatrix games. These games strictly extend zero-sum games but do not satisfy the monotonicity conditions that typically underlie convergence analyses for first-order methods. The chapter first illustrates this obstruction by showing that rank-1 games can evade the Minty-type criteria used to analyze optimistic mirror descent and extragradient methods. It then develops an efficient decentralized learning algorithm based on a parameterized reduction to zero-sum games. The key idea is to solve a sequence of zero-sum games while simultaneously enforcing a scalar consistency condition that identifies the correct rank-1 equilibrium. The resulting optimistic procedure converges in polynomial time to an approximate Nash equilibrium.

Exponential lower bounds for fictitious play. Chapter 4 turns to fictitious play in two-player potential games. Fictitious play is known to converge asymptotically in potential games, but the chapter shows that this guarantee does not imply efficient convergence. The proof constructs a two-player identical-interest game whose payoff matrix induces a long spiral trajectory. Fictitious play is forced to traverse this trajectory through a sequence of strategy switches before it can place mass on the equilibrium region. As a result, the dynamic requires exponentially many rounds to reach an approximate Nash equilibrium. The obstruction is not cycling or nonexistence of equilibrium, but the length of the transient phase hidden behind an asymptotic convergence theorem.

Lower and upper bounds for FTRL in potential games. Chapter 5 studies follow-the-regularized-leader and related no-regret dynamics in potential games. On the positive side, the chapter establishes a potential-monotonicity property for FTRL under suitable learning-rate conditions and proves an exponential upper bound for lazy alternating no-regret dynamics. On the negative side, it proves that FTRL can require exponentially many iterations to reach an approximate Nash equilibrium in two-player potential games, for any permutation-invariant regularizer and for learning rates that may vanish over time. The chapter also extends the lower-bound perspective to multiplayer potential games, proving that fictitious play can take doubly exponentially many iterations via a connection to long paths in high-dimensional combinatorial structures.

Nonconvergence of FTRL in constrained optimization. Chapter 6 investigates FTRL as a first-order method for constrained nonconvex optimization. In contrast to mirror descent, whose convergence to first-order stationary points is well understood in many settings, FTRL is history dependent and can behave very differently near the boundary of the feasible region. The chapter constructs an infinitely differentiable objective on a compact convex domain for which continuous-time FTRL has a KKT gap bounded away from zero for all time. This holds even though the objective value is monotone along the trajectory. The construction proceeds by embedding a decelerating oscillatory gradient-flow trajectory into higher-dimensional FTRL dynamics, so that FTRL repeatedly observes profitable inactive directions but never acts on them.

1.5 Organization of the dissertation

Chapter 3 presents the positive result for rank-1 games. Chapter 4 proves the lower bound for fictitious play in two-player potential games. Chapter 5 studies FTRL and no-regret dynamics in potential games, including the two-player exponential lower bound, the lazy-dynamics upper bound, and the multiplayer doubly exponential lower bound for fictitious

play. Chapter 6 studies FTRL in constrained optimization and proves asymptotic nonconvergence in terms of the KKT gap. The dissertation concludes by synthesizing these results and outlining directions for future work.

Chapter 2

Preliminaries

This chapter collects the notation and background used throughout the thesis. The goal is to fix a common language for finite games, equilibrium notions, learning dynamics, regret, and constrained optimization. Each subsequent chapter remains largely self-contained and introduces the additional definitions, assumptions, and constructions required for its own analysis.

2.1 Notation

For a positive integer n, we write

$$[n] \coloneqq \{1, \dots, n\}.$$

Vectors are denoted in bold, for example $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, and matrices by capital letters such as $\mathbf{A}, \mathbf{B}$. Unless stated otherwise, vectors are column vectors. We write $\mathbb{1}$ for the all-ones vector, with dimension clear from context, and $\boldsymbol{e}_i$ for the i-th standard basis vector.

For vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, we write

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y}$$

for the Euclidean inner product. We write $\|\boldsymbol{x}\|$ for the Euclidean norm, and more generally $\|\boldsymbol{x}\|_p$ for the ℓ_p-norm. For a set S, we denote its cardinality by $|S|$.

The probability simplex in $\mathbb{R}^m$ is denoted by

$$\Delta_m \coloneqq \left\{ \boldsymbol{x} \in \mathbb{R}^m_{\geq 0} : \sum_{j=1}^{m} x_j = 1 \right\}.$$

When the dimension is clear from context, we simply write Δ. The support of a mixed strategy $\boldsymbol{x} \in \Delta_m$ is

$$\operatorname{supp}(\boldsymbol{x}) \coloneqq \{ j \in [m] : x_j > 0 \}.$$

For a compact convex set $\mathcal{X} \subseteq \mathbb{R}^d$, we denote its relative interior by $\operatorname{relint}(\mathcal{X})$. When needed, the tangent and normal cones of $\mathcal{X}$ at $\boldsymbol{x} \in \mathcal{X}$ are denoted by $T_{\mathcal{X}}(\boldsymbol{x})$ and $N_{\mathcal{X}}(\boldsymbol{x})$, respectively.

Throughout the thesis, asymptotic notation is used in the standard sense. We write $f(n) = O(g(n))$ if there exists a universal constant $C > 0$ such that $f(n) \leq Cg(n)$ for all sufficiently large n. The notations $f(n) = \Omega(g(n))$ and $f(n) = \Theta(g(n))$ are defined analogously. When the hidden constants depend on a parameter ϵ, we may write $O_\epsilon(\cdot)$, $\Omega_\epsilon(\cdot)$, or $\Theta_\epsilon(\cdot)$.

2.2 Finite normal-form games

A finite normal-form game consists of a finite set of players $\mathcal{N} \coloneqq \{1, \ldots, n\}$, where each player $i \in \mathcal{N}$ has a finite action set $\mathcal{A}_i \coloneqq \{1, \ldots, m_i\}$. A mixed strategy of player i is a probability distribution over $\mathcal{A}_i$, and is denoted by $\boldsymbol{x}_i \in \mathcal{X}_i \coloneqq \mathcal{X}_i$. The mixed-strategy

space of the game is $\mathcal{X} \coloneqq \prod_{i=1}^{n} \mathcal{X}_i = \prod_{i=1}^{n} \mathcal{X}_i$. A mixed-strategy profile is denoted by $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \mathcal{X}$. For a player i, we write $\boldsymbol{x}_{-i} \coloneqq (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \ldots, \boldsymbol{x}_n)$ for the profile of all players except i. Thus, $(\boldsymbol{x}_i, \boldsymbol{x}_{-i})$ denotes the full strategy profile in which player i uses $\boldsymbol{x}_i$, while the other players use $\boldsymbol{x}_{-i}$.

Each player i has a payoff function $u_i : \mathcal{X} \to \mathbb{R}$. When the players use the mixed profile $\boldsymbol{x}$, the payoff of player i is written as $u_i(\boldsymbol{x}) = u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i})$.

In a two-player matrix game, the payoff functions are induced by payoff matrices. If player 1 has payoff matrix $\mathbf{A}$ and player 2 has payoff matrix $\mathbf{B}$, then

$$\boldsymbol{u}_1(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1^\top \mathbf{A} \boldsymbol{x}_2, \qquad \text{and} \qquad u_2(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1^\top \mathbf{B} \boldsymbol{x}_2.$$

This matrix notation is used when the bilinear representation is needed explicitly.

Given a profile $\boldsymbol{x}_{-i}$ of the other players, a best response of player i is any strategy in

$$\arg\max_{\boldsymbol{x}_i' \in \mathcal{X}_i} u_i(\boldsymbol{x}_i', \boldsymbol{x}_{-i}).$$

The set of best responses of player i to $\boldsymbol{x}_{-i}$ is denoted by

$$\mathrm{BR}_i(\boldsymbol{x}_{-i}) \coloneqq \arg\max_{\boldsymbol{x}_i' \in \mathcal{X}_i} u_i(\boldsymbol{x}_i', \boldsymbol{x}_{-i}).$$

2.3 Equilibria and approximate equilibria

A mixed-strategy profile $\boldsymbol{x}^\star \in \mathcal{X}$ is a Nash equilibrium if no player can improve their payoff by changing only their own strategy. That is, for every player $i \in \mathcal{N}$,

$$u_i(\boldsymbol{x}_i^\star, \boldsymbol{x}_{-i}^\star) \geq u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}^\star) \quad \text{for all } \boldsymbol{x}_i \in \mathcal{X}_i.$$

Equivalently,

$$\boldsymbol{x}_i^\star \in \mathrm{BR}_i(\boldsymbol{x}_{-i}^\star) \quad \text{for every } i \in \mathcal{N}.$$

For $\epsilon \geq 0$, a mixed-strategy profile $\boldsymbol{x} \in \mathcal{X}$ is an ϵ-approximate Nash equilibrium if every player can gain at most ϵ by deviating unilaterally. That is, for every player $i \in \mathcal{N}$,

$$\max_{\boldsymbol{x}_i' \in \mathcal{X}_i} u_i(\boldsymbol{x}_i', \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}) \leq \epsilon.$$

The individual deviation gap of player i at $\boldsymbol{x}$ is

$$\mathtt{Gap}_i(\boldsymbol{x}) \coloneqq \max_{\boldsymbol{x}_i' \in \mathcal{X}_i} u_i(\boldsymbol{x}_i', \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}).$$

The Nash gap of the profile is

$$\mathtt{Gap}(\boldsymbol{x}) \coloneqq \max_{i \in \mathcal{N}} \mathtt{Gap}_i(\boldsymbol{x}).$$

Thus, $\boldsymbol{x}$ is an ϵ-approximate Nash equilibrium if and only if

$$\mathtt{Gap}(\boldsymbol{x}) \leq \epsilon.$$

A profile is a Nash equilibrium if and only if $\mathtt{Gap}(\boldsymbol{x}) = 0$.

In some chapters, it is more convenient to use the sum of the individual deviation gaps rather than their maximum. Whenever a different normalization is used, it is stated explicitly.

2.4 Structured classes of games

Several structured classes of games appear throughout the thesis.

Zero-sum games. A two-player game is zero-sum if

$$u_1(\boldsymbol{x}_1, \boldsymbol{x}_2) + u_2(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0 \quad \text{for all } (\boldsymbol{x}_1, \boldsymbol{x}_2).$$

For matrix games, this corresponds to $\mathbf{B} = -\mathbf{A}$. In this case, player 1 seeks to maximize $\boldsymbol{x}_1^\top \mathbf{A} \boldsymbol{x}_2$, while player 2 seeks to minimize the same quantity. A Nash equilibrium of a zero-sum game is therefore a saddle point of the bilinear payoff. In particular, in the matrix case, an equilibrium profile $(\boldsymbol{x}_1^\star, \boldsymbol{x}_2^\star)$ satisfies

$$\boldsymbol{x}_1^\top \mathbf{A} \boldsymbol{x}_2^\star \leq \boldsymbol{x}_1^{\star\top} \mathbf{A} \boldsymbol{x}_2^\star \leq \boldsymbol{x}_1^{\star\top} \mathbf{A} \boldsymbol{x}_2$$

for all $\boldsymbol{x}_1 \in \mathcal{X}_1$ and $\boldsymbol{x}_2 \in \mathcal{X}_2$.

Potential games. A game is an exact potential game if there exists a function $\Phi : \mathcal{X} \to \mathbb{R}$ such that every unilateral payoff difference is captured by the corresponding change in Φ. That is, for every player $i \in \mathcal{N}$, every $\boldsymbol{x}_i, \boldsymbol{x}_i' \in \mathcal{X}_i$, and every fixed $\boldsymbol{x}_{-i}$,

$$u_i(\boldsymbol{x}_i', \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}) = \Phi(\boldsymbol{x}_i', \boldsymbol{x}_{-i}) - \Phi(\boldsymbol{x}_i, \boldsymbol{x}_{-i}).$$

The function Φ is called a potential function.

Potential games play a central role in the later chapters because the potential provides a natural Lyapunov-type object for many learning dynamics. However, the existence of a potential does not by itself imply fast convergence, or even convergence of every natural dynamic to a desired equilibrium notion.

Remark 2.4.1. What is special about potential games is that the set of global maximizers of the potential function captures the Nash equilibria. In particular, any global maximizer of Φ is a Nash equilibrium. Indeed, suppose that $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*) \in \mathcal{X}_1 \times \mathcal{X}_2$ is a global maximizer of

Φ. For any unilateral deviation $\boldsymbol{x}_1'$ of the row player, $\Phi(\boldsymbol{x}_1', \boldsymbol{x}_2^*) \leq \Phi(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$, and therefore $(\boldsymbol{x}_1')^\top \mathbf{A} \boldsymbol{x}_2^* - (\boldsymbol{x}_1^*)^\top \mathbf{A} \boldsymbol{x}_2^* = \Phi(\boldsymbol{x}_1', \boldsymbol{x}_2^*) - \Phi(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*) \leq 0$. Thus, the row player cannot improve by deviating. The same argument applies to the column player.

Identical-payoff games. A game is identical-payoff if all players have the same payoff function. Equivalently, there exists a function $\Phi : \mathcal{X} \to \mathbb{R}$ such that $u_i(\boldsymbol{x}) = \Phi(\boldsymbol{x})$ for every player $i \in \mathcal{N}$ and every $\boldsymbol{x} \in \mathcal{X}$.

Every identical-payoff game is an exact potential game, with the common payoff function serving as a potential function. Thus, identical-payoff games form a particularly important subclass of potential games. In such games, every player seeks to optimize the same objective.

Rank-structured games. Some chapters study games whose payoff matrices have additional algebraic structure, such as low-rank perturbations of zero-sum games. The detailed rank-1 constructions and the associated parameterizations are specific to the corresponding chapter and are introduced there.

2.5 Learning dynamics

A learning dynamic is an iterative or continuous-time rule by which players update their strategies using payoff information. In discrete time, such a dynamic produces a sequence of mixed-strategy profiles $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots \in \mathcal{X}$, where $\boldsymbol{x}^{(t)} = (\boldsymbol{x}_1^{(t)}, \ldots, \boldsymbol{x}_n^{(t)})$. Here $\boldsymbol{x}_i^{(t)}$ denotes the mixed strategy of player i at time t.

In continuous time, one studies trajectories

$$t \mapsto \boldsymbol{x}(t) = (\boldsymbol{x}_1(t), \ldots, \boldsymbol{x}_n(t)).$$

The central questions studied in this thesis are whether such dynamics converge, how quickly they converge, and how their behavior depends on the structure of the underlying game.

Two learning paradigms are especially important in the sequel: fictitious play and regularized first-order learning dynamics.

Fictitious play. In fictitious play, players best respond to the empirical distribution of their opponents' past play. In a two-player game, if $\overline{\boldsymbol{x}}_1^{(t)}$ and $\overline{\boldsymbol{x}}_2^{(t)}$ denote the empirical distributions of the two players up to time t, then each player chooses a best response to the opponent's empirical distribution. The precise update rule, empirical distribution notation, and tie-breaking conventions used in the fictitious-play chapter are specified there.

Regularized first-order dynamics. The other family of dynamics used throughout the thesis consists of regularized first-order methods, including mirror descent, follow-the-regularized-leader, and their optimistic variants. These methods convert payoff or gradient information into feasible mixed strategies through a regularization term. Their general forms are recalled in Section 2.7, while the precise variants used in the proofs are stated in the relevant chapters.

2.6 Regret and no-regret learning

Regret measures the performance of an online learning algorithm relative to a fixed comparator chosen in hindsight. For player i, suppose that the algorithm chooses strategies $\boldsymbol{x}_i^{(1)}, \dots, \boldsymbol{x}_i^{(T)} \in \mathcal{X}_i$, and receives payoff vectors $\boldsymbol{u}_i^{(1)}, \dots, \boldsymbol{u}_i^{(T)} \in \mathbb{R}^{m_i}$. The cumulative payoff of player i is $\sum_{t=1}^{T} \left\langle \boldsymbol{x}_i^{(t)}, \boldsymbol{u}_i^{(t)} \right\rangle$, whereas, the regret of player i after T rounds is

$$\mathtt{Reg}_i(T) \coloneqq \max_{\boldsymbol{x}_i \in \mathcal{X}_i} \sum_{t=1}^{T} \left\langle \boldsymbol{x}_i - \boldsymbol{x}_i^{(t)}, \boldsymbol{u}_i^{(t)} \right\rangle. \tag{2.1}$$

The player has no regret if

$$\frac{\mathtt{Reg}_i(T)}{T} \to 0 \quad \text{as } T \to \infty.$$

No-regret learning is closely connected to equilibrium computation. In zero-sum games, if both players use no-regret algorithms, then their time-averaged play approaches the set of Nash equilibria. In more general classes of games, however, the relationship between no-regret learning and convergence to equilibrium is subtler. Understanding this distinction is one of the recurring themes of the thesis.

2.7 Mirror descent, FTRL, and optimistic variants

We next recall two standard first-order learning schemes that appear throughout the thesis: mirror descent and follow-the-regularized-leader. Both methods can be viewed as regularized ways of converting payoff or gradient information into a feasible decision.

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a compact convex decision set, and let $\psi : \mathcal{X} \to \mathbb{R}$ be a differentiable convex function. The Bregman divergence induced by ψ is

$$D_\psi(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \psi(\boldsymbol{x}) - \psi(\boldsymbol{y}) - \langle \nabla\psi(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle .$$

When ψ is strongly convex, D_ψ provides a geometry-adapted notion of distance on $\mathcal{X}$.

Mirror descent. Mirror descent updates the current decision by balancing the current payoff or gradient vector against the Bregman divergence from the previous iterate. In a maximization setting, the update takes the form

$$\boldsymbol{x}^{(t+1)} \in \arg\max_{\boldsymbol{x} \in \mathcal{X}} \left\{ \left\langle \boldsymbol{u}^{(t)}, \boldsymbol{x} \right\rangle - \frac{1}{\eta^{(t)}} D_\psi(\boldsymbol{x}, \boldsymbol{x}^{(t)}) \right\},$$

where $\boldsymbol{u}^{(t)}$ is the payoff or gradient vector observed at time t, and $\eta^{(t)} > 0$ is the step size. Equivalently, under a minimization setting, the sign of the linear term is reversed.

In the context of games, each player applies this update to their own strategy set. Thus, for player i, the update is

$$\boldsymbol{x}_i^{(t+1)} \in \arg\max_{\boldsymbol{x}_i \in \mathcal{X}_i} \left\{ \left\langle \boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i \right\rangle - \frac{1}{\eta^{(t)}} D_{\psi_i}(\boldsymbol{x}_i, \boldsymbol{x}_i^{(t)}) \right\}.$$

Follow-the-regularized-leader. Follow-the-regularized-leader, abbreviated `FTRL`, chooses a decision by optimizing cumulative payoff information together with a regularization term. For player i, a generic `FTRL` update is

$$\boldsymbol{x}_i^{(t+1)} \in \arg\max_{\boldsymbol{x}_i \in \mathcal{X}_i} \left\{ \left\langle \boldsymbol{x}_i, \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} \right\rangle - \frac{1}{\eta^{(t)}} R_i(\boldsymbol{x}_i) \right\},$$

where R_i is a regularizer, and $\eta^{(t)} > 0$ is a learning rate. Different choices of the regularizer recover different dynamics, including multiplicative weights update and other mirror-descent-type schemes.

Optimistic FTRL. Optimistic variants modify the update by incorporating a prediction of the next payoff vector. In optimistic `FTRL`, player i uses an additional prediction vector $\boldsymbol{m}_i^{(t+1)}$ and updates according to

$$\boldsymbol{x}_i^{(t+1)} \in \arg\max_{\boldsymbol{x}_i \in \mathcal{X}_i} \left\{ \left\langle \boldsymbol{x}_i, \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} + \boldsymbol{m}_i^{(t+1)} \right\rangle - \frac{1}{\eta^{(t)}} R_i(\boldsymbol{x}_i) \right\}.$$

Here $\boldsymbol{m}_i^{(t+1)}$ is meant to approximate the next payoff vector $\boldsymbol{u}_i^{(t+1)}$. A common choice is the one-step recency prediction

$$\boldsymbol{m}_i^{(t+1)} = \boldsymbol{u}_i^{(t)}.$$

Optimistic mirror descent. The corresponding optimistic mirror descent update uses an extrapolated payoff vector. With the one-step recency prediction, this gives

$$\boldsymbol{x}_i^{(t+1)} \in \arg\max_{\boldsymbol{x}_i \in \mathcal{X}_i} \left\{ \left\langle 2\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}, \boldsymbol{x}_i \right\rangle - \frac{1}{\eta^{(t)}} D_{\psi_i}(\boldsymbol{x}_i, \boldsymbol{x}_i^{(t)}) \right\}.$$

Thus, compared with ordinary mirror descent, optimistic mirror descent uses the extrapolated vector $2\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}$ in place of the current payoff vector $\boldsymbol{u}_i^{(t)}$.

Optimistic variants are useful when the payoff sequence is predictable, for instance in structured games where consecutive payoff vectors are strongly correlated. The rank-1 chapter uses an optimistic mirror-descent-type procedure in a parameterized zero-sum game. The precise form of the update, including the payoff vectors and parameter update, is defined in that chapter.

2.8 Continuous-time dynamics and invariance

Several arguments in the thesis use continuous-time dynamical systems. A trajectory is a differentiable map $t \mapsto \boldsymbol{x}(t)$ satisfying an ordinary differential equation of the form

$$\dot{\boldsymbol{x}}(t) = F(\boldsymbol{x}(t)),$$

where F is a vector field on the relevant state space. In game dynamics, the state variable is usually a mixed-strategy profile $\boldsymbol{x}(t) = (x_1(t), \ldots, x_d(t))$, where d the dimension of the system. A set $\mathcal{S}$ is invariant under the flow if every trajectory initialized in $\mathcal{S}$ remains in $\mathcal{S}$ for all forward time. That is, if $\boldsymbol{x}(0) \in \mathcal{S}$, then $\boldsymbol{x}(t) \in \mathcal{S}$ for all $t \geq 0$. Invariant sets are useful for proving both positive and negative results: they may identify regions in which the dynamics remain controlled, but they may also reveal obstructions to convergence.

A common way to prove convergence is to construct a Lyapunov function, namely a scalar function that is monotone along trajectories. In potential games, the potential function is often a natural candidate. Nevertheless, monotonicity of a potential-like quantity need not imply rapid convergence, nor does it preclude the existence of long transients.

2.9 Optimization background

Some parts of the thesis relate learning dynamics to constrained optimization. We recall here the minimal terminology needed for those results.

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a compact convex set, and let $f : \mathcal{X} \to \mathbb{R}$ be a differentiable function. A point $\boldsymbol{x}^\star \in \mathcal{X}$ is a first-order stationary point for the maximization problem

$$\max_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) \quad \text{if} \quad \langle \nabla f(\boldsymbol{x}^\star), \boldsymbol{x} - \boldsymbol{x}^\star \rangle \leq 0 \quad \text{for all } \boldsymbol{x} \in \mathcal{X}.$$

Equivalently, $\nabla f(\boldsymbol{x}^\star) \in N_{\mathcal{X}}(\boldsymbol{x}^\star)$, up to the sign convention associated with maximization.

Quantitative versions of this condition are often expressed through a stationarity gap or KKT gap. Such a quantity measures the extent to which the first-order optimality condition fails. The exact KKTGap used in Chapter 6 is specific to that construction and is introduced there.

2.10 Min-max optimization and optimistic mirror descent

In this section, we recall how online learning, and in particular regret minimization, can be used to solve convex-concave saddle-point problems. This material is standard; see, for instance, Orabona (2019).

Saddle-point problems. Let $\mathcal{X}_1 \subseteq \mathbb{R}^{d_1}$ and $\mathcal{X}_2 \subseteq \mathbb{R}^{d_2}$ be nonempty compact convex sets, and let $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$. We consider the min-max problem $\min_{\boldsymbol{x}_1 \in \mathcal{X}_1} \max_{\boldsymbol{x}_2 \in \mathcal{X}_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2)$.

Definition 2.10.1 (Saddle point)**.** A point $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*) \in \mathcal{X}_1 \times \mathcal{X}_2$ is a saddle point of f if

$$f(\boldsymbol{x}_1^*, \boldsymbol{x}_2) \le f(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*) \le f(\boldsymbol{x}_1, \boldsymbol{x}_2^*) \quad \text{for all } \boldsymbol{x}_1 \in \mathcal{X}_1,\ \boldsymbol{x}_2 \in \mathcal{X}_2.$$

Thus, $\boldsymbol{x}_1^*$ is optimal for the minimizing player and $\boldsymbol{x}_2^*$ is optimal for the maximizing player. We focus on the convex-concave case: for every fixed $\boldsymbol{x}_2 \in \mathcal{X}_2$, the function $f(\cdot, \boldsymbol{x}_2)$ is convex, and for every fixed $\boldsymbol{x}_1 \in \mathcal{X}_1$, the function $f(\boldsymbol{x}_1, \cdot)$ is concave.

Theorem 2.10.1 (Minimax theorem v. Neumann (1928); Sion (1958))**.** *Let $\mathcal{X}_1 \subseteq \mathbb{R}^{d_1}$ and $\mathcal{X}_2 \subseteq \mathbb{R}^{d_2}$ be nonempty compact convex sets. Let $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ be continuous and convex-concave. Then*

$$\min_{\boldsymbol{x}_1 \in \mathcal{X}_1} \max_{\boldsymbol{x}_2 \in \mathcal{X}_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2) = \max_{\boldsymbol{x}_2 \in \mathcal{X}_2} \min_{\boldsymbol{x}_1 \in \mathcal{X}_1} f(\boldsymbol{x}_1, \boldsymbol{x}_2).$$

The value in Theorem 2.10.1 is denoted by

$$\mathrm{Val}(f) \coloneqq \min_{\boldsymbol{x}_1 \in \mathcal{X}_1} \max_{\boldsymbol{x}_2 \in \mathcal{X}_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2) = \max_{\boldsymbol{x}_2 \in \mathcal{X}_2} \min_{\boldsymbol{x}_1 \in \mathcal{X}_1} f(\boldsymbol{x}_1, \boldsymbol{x}_2).$$

For a candidate pair $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, the standard saddle-point gap is

$$\texttt{SPGap}(\boldsymbol{x}_1, \boldsymbol{x}_2) \coloneqq \max_{\boldsymbol{x}_2' \in \mathcal{X}_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2') - \min_{\boldsymbol{x}_1' \in \mathcal{X}_1} f(\boldsymbol{x}_1', \boldsymbol{x}_2).$$

This quantity is always nonnegative in the convex-concave setting. Moreover, $\texttt{SPGap}(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0$ if and only if $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a saddle point.

The saddle-point gap can be decomposed into the suboptimality of the minimizing player and the suboptimality of the maximizing player. Equivalently, if $\texttt{Gap}_1$ and $\texttt{Gap}_2$ denote the individual deviation gaps of the minimizing and maximizing players in the associated saddle-point problem, then

$$\texttt{Gap}_1(\boldsymbol{x}_1, \boldsymbol{x}_2) + \texttt{Gap}_2(\boldsymbol{x}_1, \boldsymbol{x}_2) = \texttt{SPGap}(\boldsymbol{x}_1, \boldsymbol{x}_2).$$

Thus, $\texttt{SPGap}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ measures the simultaneous suboptimality of the two players.

Online learning viewpoint. The saddle-point problem can be approached through online learning. At each round $t \in [T]$, the minimizing player chooses $\boldsymbol{x}_1^{(t)} \in \mathcal{X}_1$, and the maximizing player chooses $\boldsymbol{x}_2^{(t)} \in \mathcal{X}_2$. The minimizing player receives a loss vector $\boldsymbol{u}_1^{(t)} \in \mathbb{R}^{d_1}$, while the maximizing player may equivalently be viewed as minimizing the negative payoff and therefore receives a loss vector $\boldsymbol{u}_2^{(t)} \in \mathbb{R}^{d_2}$.

For a sequence of linear losses $\ell_1^{(t)}(\boldsymbol{x}_1) \coloneqq \left\langle \boldsymbol{u}_1^{(t)}, \boldsymbol{x}_1 \right\rangle$, the external regret of the minimizing player after T rounds is

$$\texttt{Reg}_1^{(T)} \coloneqq \sum_{t=1}^{T} \ell_1^{(t)}(\boldsymbol{x}_1^{(t)}) - \min_{\boldsymbol{x}_1 \in \mathcal{X}_1} \sum_{t=1}^{T} \ell_1^{(t)}(\boldsymbol{x}_1) = \max_{\boldsymbol{x}_1 \in \mathcal{X}_1} \sum_{t=1}^{T} \left\langle \boldsymbol{u}_1^{(t)}, \boldsymbol{x}_1^{(t)} - \boldsymbol{x}_1 \right\rangle.$$

The regret $\texttt{Reg}_2^{(T)}$ of the second player is defined analogously with respect to the loss vectors $\boldsymbol{u}_2^{(t)}$. Now, if both players have sublinear regret, then the average iterates

$$\overline{\boldsymbol{x}}_1^{(T)} \coloneqq \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_1^{(t)}, \qquad \overline{\boldsymbol{x}}_2^{(T)} \coloneqq \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_2^{(t)}$$

approach the saddle-point set in terms of the saddle-point gap. This is the standard online-learning interpretation of convex-concave optimization.

2.10.1 Fast convergence under smoothness.

Optimistic mirror descent improves over the worst-case $O(1/\sqrt{T})$ ergodic convergence rate in smooth predictable settings. We state a standard form of this guarantee, following Rakhlin and Sridharan (2013); Orabona (2019).

Assumption 2.10.1. Assume that $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ is differentiable on an open set containing $\mathcal{X}_1 \times \mathcal{X}_2$, and that there exist constants $L_{11}, L_{12}, L_{22} \geq 0$ such that, for all $\boldsymbol{x}_1, \boldsymbol{x}_1' \in \mathcal{X}_1$ and all $\boldsymbol{x}_2, \boldsymbol{x}_2' \in \mathcal{X}_2$,

$$\|\nabla_{\boldsymbol{x}_1} f(\boldsymbol{x}_1, \boldsymbol{x}_2) - \nabla_{\boldsymbol{x}_1} f(\boldsymbol{x}_1', \boldsymbol{x}_2)\|_{1,*} \leq L_{11} \|\boldsymbol{x}_1 - \boldsymbol{x}_1'\|_1 ,$$

$$\|\nabla_{\boldsymbol{x}_1} f(\boldsymbol{x}_1, \boldsymbol{x}_2) - \nabla_{\boldsymbol{x}_1} f(\boldsymbol{x}_1, \boldsymbol{x}_2')\|_{1,*} \leq L_{12} \|\boldsymbol{x}_2 - \boldsymbol{x}_2'\|_2 ,$$

$$\|\nabla_{\boldsymbol{x}_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2) - \nabla_{\boldsymbol{x}_2} f(\boldsymbol{x}_1', \boldsymbol{x}_2)\|_{2,*} \leq L_{12} \|\boldsymbol{x}_1 - \boldsymbol{x}_1'\|_1 ,$$

and

$$\|\nabla_{\boldsymbol{x}_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2) - \nabla_{\boldsymbol{x}_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2')\|_{2,*} \leq L_{22} \|\boldsymbol{x}_2 - \boldsymbol{x}_2'\|_2 .$$

Here $\|\cdot\|_{1,*}$ and $\|\cdot\|_{2,*}$ are the dual norms of $\|\cdot\|_1$ and $\|\cdot\|_2$, respectively.

Theorem 2.10.2 (Optimistic mirror descent for saddle-point problems)**.** *Let $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ be continuous and convex-concave, and suppose that Assumption 2.10.1 holds. Let $\psi_1 : \mathcal{X}_1 \to \mathbb{R}$ be 1-strongly convex with respect to $\|\cdot\|_1$, and let $\psi_2 : \mathcal{X}_2 \to \mathbb{R}$ be 1-strongly convex with respect to $\|\cdot\|_2$.*

Fix $\alpha > 0$, and choose

$$\lambda_1 \geq 2\sqrt{2}\left(L_{11} + \frac{L_{12}}{\alpha}\right), \qquad \lambda_2 \geq 2\sqrt{2}\,(L_{22} + \alpha L_{12}) .$$

Then the average iterates generated by optimistic mirror descent satisfy

$$\mathtt{SPGap}(\overline{\boldsymbol{x}}_1^{(T)}, \overline{\boldsymbol{x}}_2^{(T)}) = \max_{\boldsymbol{x}_2 \in \mathcal{X}_2} f(\overline{\boldsymbol{x}}_1^{(T)}, \boldsymbol{x}_2) - \min_{\boldsymbol{x}_1 \in \mathcal{X}_1} f(\boldsymbol{x}_1, \overline{\boldsymbol{x}}_2^{(T)}) \leq \frac{C_{\mathrm{OMD}}}{T},$$

where

$$C_{\mathrm{OMD}} \coloneqq B_{\psi_1}(\boldsymbol{x}_1^\star, \boldsymbol{x}_1^{(1)}) + B_{\psi_2}(\boldsymbol{x}_2^\star, \boldsymbol{x}_2^{(1)}) + \frac{\left\|\boldsymbol{u}_1^{(1)}\right\|_{1,*}^2}{\lambda_1} + \frac{\left\|\boldsymbol{u}_2^{(1)}\right\|_{2,*}^2}{\lambda_2}.$$

This theorem shows that, in smooth convex-concave problems, optimistic mirror descent achieves an $O(1/T)$ convergence rate in the ergodic saddle-point gap. The rank-1 chapter uses this principle in the parameterized zero-sum setting developed there.

2.11 Organization of the remaining preliminaries

The material in this chapter is intentionally limited to concepts that recur across the thesis. Each subsequent chapter introduces its own specialized preliminaries.

In particular, Chapter 3 develops the rank-1 structure and the associated parameterized zero-sum games used in the analysis of rank-1 games. The fictitious-play chapter introduces the precise update rules, tie-breaking conventions, and empirical-distribution notation needed for that setting. The first `FTRL` chapter specifies the version of follow-the-regularized-leader used for potential games and develops the corresponding potential-based analysis. Finally, Chapter 6 introduces the constrained optimization formulation, the relevant KKT-gap notion, and the auxiliary constructions used to prove the lower bound.

Thus, this chapter should be read as a common foundation, while the local preliminaries in each chapter provide the exact technical framework for the results proved there.

Chapter 3

Learning Nash Equilibria in Rank-1 Games

3.1 Preface

This chapter is based on joint work Patris and Panageas (2024). The chapter has been adapted from the paper version for inclusion in this dissertation. In particular, some exposition has been streamlined and notation has been made consistent with the rest of the thesis. The purpose of this chapter is to present the positive result of the dissertation: although rank-1 games lie beyond the standard zero-sum setting and do not satisfy the usual Minty-type assumptions underlying many convergence analyses, they still admit an efficient decentralized learning procedure when their structure is exploited appropriately. We show that a modification of optimistic mirror descent converges to an ϵ-approximate Nash equilibrium (NE) after $O\left(\frac{1}{\epsilon^2}\log\left(\frac{1}{\epsilon}\right)\right)$ iterates in rank-1 games. We achieve this by leveraging structural results about the NE landscape of rank-1 games (Adsul et al., 2021).

At a high level, the chapter shows that rank-1 equilibrium learning can be reduced to the problem of solving a sequence of parameterized zero-sum games while simultaneously controlling an additional consistency condition. The main technical contribution is the design and analysis of an optimistic algorithm that carries out this program efficiently.

3.2 Introduction

Learning in games has its origins in Blackwell's influential work on the approachability theorem (Blackwell, 1956; Abernethy et al., 2011), which paved the way for the development of various learning algorithms with convergence guarantees to different game theoretic solution concepts, including coarse correlated equilibrium (CCE) and Nash equilibrium (NE). For instance, a well-known result in this context is that if both players in a zero-sum game employ a no-regret learning algorithm, the empirical averages of their strategies will converge to a NE. However, the typical guarantees provided by online learning frameworks do not shed light on how the system stabilizes toward a Nash equilibrium when the underlying two-player game is neither zero-sum nor a potential game (Anagnostides et al., 2022b). In such cases, one can only argue that the empirical averages converge to a CCE if both players employ a no-regret learning algorithm. This raises the following fundamental question:

For which classes of two-player games, beyond zero-sum and potential games, can we design online learning algorithms that guarantee convergence to a Nash equilibrium?

The work in Kannan and Theobald (2010) introduced a hierarchy of bimatrix games denoted as $(\mathbf{A}, \mathbf{B})$, where the rank of $\mathbf{A} + \mathbf{B}$ is k. These are referred to as rank-k bimatrix games. Notably, this class encompasses zero-sum games when k is zero. Intriguingly, it has been demonstrated that even for the case of $k = 1$, the set of Nash equilibria can comprise a potentially large number of disconnected components and lacks convexity. Furthermore, it has been proven that for $k \geq 3$, computing an *exact* Nash equilibria in rank-k bimatrix

games is a PPAD-hard problem (Adsul et al., 2021). In other words, it is highly unlikely that a polynomial-time algorithm exists for computing Nash equilibria in such games. On the contrary, rank-1 games have been shown to have polynomial-time algorithms based on linear programming formulations (Adsul et al., 2021, 2011), while the complexity of rank-2 games remains an open question.

Our contributions and technical overview In this chapter, we focus on rank-1 games. First, we establish that there exist rank-1 games that do not satisfy the so-called Minty criterion and, as a result, *optimistic mirror descent* (Rakhlin et al., 2011) and *extra gradient* algorithms (Korpelevich, 1976) fail to converge to a Nash equilibrium.

On a positive note, we introduce a decentralized learning algorithm that converges to an ϵ-approximate Nash equilibrium in $O\left(\frac{1}{\epsilon^2}\log\left(\frac{1}{\epsilon}\right)\right)$ iterations, as proven in Theorem 3.4.1. Both agents in the game utilize an optimistic mirror descent algorithm in a parametrized zero-sum game that evolves over time. The key to our technical analysis lies in how we adapt the underlying parameter, ensuring that when the algorithm terminates, an approximate Nash equilibrium in the parametrized zero-sum game coincides with an approximate Nash equilibrium in the rank-1 game. To provide more context, consider a rank-1 bimatrix game represented as $(\mathbf{A}, \mathbf{B})$. We can assume that $\mathbf{B} = -\mathbf{A} + \boldsymbol{a}\boldsymbol{b}^\top$ for some vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. It turns out that a Nash equilibrium $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ of the rank-1 game $(\mathbf{A}, \mathbf{B})$ is also a Nash equilibrium of the parametrized zero-sum game $(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda\boldsymbol{b}^\top)$ for an appropriate choice of the parameter λ, where $\mathbb{1}$ is the all-ones vector. The primary technical challenge is formulating a suitable optimization problem to ensure that a Nash equilibrium in the parameterized zero-sum game translates to a Nash equilibrium in the rank-1 game, as shown in Lemma 3.4.2.

Our proposed algorithm is based on *optimistic mirror descent* applied to the parameterized bilinear function $\boldsymbol{x}_1^\top(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top)\boldsymbol{x}_2$ with an additional regularization term $(\boldsymbol{x}^\top\boldsymbol{a} - \lambda)^2$. De-

pending on the sign of the difference $(\boldsymbol{x}^\top \boldsymbol{a} - \lambda)$, λ is updated appropriately. To prove our claims, we leverage the structural insights of the Nash equilibrium landscape (Adsul et al., 2021) and exploit the stationarity properties of the λ-parameterized (strongly) concave-convex function $f_\lambda(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1^\top(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top)\boldsymbol{x}_2 - \frac{1}{2}(\boldsymbol{x}_1^\top \boldsymbol{a} - \lambda)^2$. Furthermore, in Section 3.5 we provide experimental evaluation supporting our theoretical findings.

3.3 Preliminaries

3.3.1 Rank-1 games

In this chapter, we focus on bimatrix games of fixed rank, and particularly of rank 1. Thus, there are two players with finite action sets $\mathcal{A}_1$ and $\mathcal{A}_2$, of cardinalities $|\mathcal{A}_1| = m_1$ and $|\mathcal{A}_2| = m_2$, respectively. Accordingly, let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m_1 \times m_2}$ denote the payoff matrices of the two players. This hierarchy within two-player games was initially introduced by Kannan and Theobald (2010) in an attempt to offer a more detailed characterization of games that (may) admit simple polynomial-time algorithms based on the rank of the matrix $(\mathbf{A} + \mathbf{B})$. In a related model, Lipton et al. (2003) investigated games where *both* payoff matrices are of fixed rank k, and they proposed a simple (quasi) polynomial-time algorithm for constructing an ϵ-approximate NE. This seemingly similar model, however, has a significant drawback; not even zero-sum games, which have polynomial-time algorithms, entirely belong to this class. Motivated by this observation, Kannan and Theobald (2010) proposed using the rank of $(\mathbf{A} + \mathbf{B})$ as the relevant complexity measure.

Definition 3.3.1 (Games of fixed rank)**.** The rank of a bimatrix game $(\mathbf{A}, \mathbf{B})$ is the rank of matrix $\mathbf{A} + \mathbf{B}$.

Remark 3.3.1. It is apparent that zero-sum games $(\mathbf{A}, -\mathbf{A})$ are rank-0 games as $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$. Additionally, rank-1 games satisfy $\mathbf{A} + \mathbf{B} = \boldsymbol{a}\boldsymbol{b}^\top$.

Assumption 3.3.1. As Nash equilibria remain invariant under the transformations of shifting and scaling of the payoff matrices, we assume that the elements of $\mathbf{A}, \boldsymbol{a}$ and $\boldsymbol{b}$ are confined within the interval $[-1, 1]$. This assumption simplifies our calculations without impacting any of our results.

The following lemma presents an alternative method for describing a bimatrix game of fixed rank r.

Lemma 3.3.1 (Lemma 5 (Adsul et al., 2021))**.** *An* $(m_1 \times m_2)$ *bimatrix game* $(\mathbf{A}, \mathbf{B})$ *of positive rank* r *can be written as* $(\mathbf{A}, \mathbf{C} + \boldsymbol{a}\boldsymbol{b}^\top)$ *for suitable* $\boldsymbol{a} \in \mathbb{R}^{m_1}, \boldsymbol{b} \in \mathbb{R}^{m_2}$*, and a game* $(\mathbf{A}, \mathbf{C})$ *of rank* $r - 1$*.*

The following lemma, though simple, holds significant importance as it establishes that for a bimatrix game of fixed rank r, the set of its Nash equilibria is equivalent to the intersection of the set of equilibria of a parameterized game of rank $r - 1$ and a hyperplane.

Lemma 3.3.2 (Lemma 6 (Adsul et al., 2021))**.** *Let* $\mathbf{A}, \mathbf{C} \in \mathbb{R}^{m_1 \times m_2}$, $\boldsymbol{x}_1 \in \Delta_{m_1}$, $\boldsymbol{x}_2 \in \Delta_{m_2}$, $\boldsymbol{a} \in \mathbb{R}^{m_1}$, $\boldsymbol{b} \in \mathbb{R}^{m_2}$, $\lambda \in \mathbb{R}$*. The following are equivalent:*

(a) $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ *is an equilibrium of* $(\mathbf{A}, \mathbf{C} + \boldsymbol{a}\boldsymbol{b}^\top)$*.*

(b) $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ *is an equilibrium of* $(\mathbf{A}, \mathbf{C} + \mathbb{1}\lambda\boldsymbol{b}^\top)$ *and* $\boldsymbol{x}_1^\top \boldsymbol{a} = \lambda$*.*

(c) $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ *is an equilibrium of* $(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, \mathbf{C} + \mathbb{1}\lambda\boldsymbol{b}^\top)$ *and* $\boldsymbol{x}_1^\top \boldsymbol{a} = \lambda$*.*

Sketch Proof. The equivalence of (a) and (b) holds because the players get in both games the same expected payoffs for their pure strategies. The games in (b) and (c) are strategically equivalent. □

While it may not be immediately evident how this rank reduction can be practically applied, the true significance of Lemma 3.3.2 becomes fully apparent in the context of rank-1 games.

Corollary 3.3.1. *Let* $(\mathbf{A}, \mathbf{B})$ *be a bimatrix game of rank 1, i.e.* $\mathbf{A} + \mathbf{B} = \boldsymbol{a}\boldsymbol{b}^\top$. *Then, the following are equivalent.*

(a) $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ *is an equilibrium of* $(\mathbf{A}, -\mathbf{A} + \boldsymbol{a}\boldsymbol{b}^\top)$.

(b) $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ *is an equilibrium of* $(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda\boldsymbol{b}^\top)$ *and* $\boldsymbol{x}_1^\top \boldsymbol{a} = \lambda$.

The proof follows trivially from the equivalence of (a) and (c) in Lemma 3.3.2.

Parameterized zero-sum games Corollary 3.3.1 establishes a connection between rank-1 games and parameterized zero-sum games. While this link initially appears compelling, it remains uncertain whether an algorithm can *compute*, or more importantly *learn*, a Nash equilibrium $(\boldsymbol{x}, \boldsymbol{y})$ satisfying $\boldsymbol{x}^\top \boldsymbol{a} = \lambda$ at the same time. The primary challenge arises from the fact that optimal strategies vary as the parameter λ changes. It is well-known, however, that zero-sum games can be efficiently solved using various methods, including LP-based techniques (Adler, 2013; von Stengel, 2023) and no-regret algorithms (Freund and Schapire, 1999; Daskalakis et al., 2011; Syrgkanis et al., 2015; Chen and Peng, 2020). Therefore, as long as there is an easy way of updating λ, the parameterized zero-sum game serves as an *intermediary* to solve the rank-1 game.

An initial attempt was made in (Adsul et al., 2011), where they demonstrated that the set of fully-labeled points, which captures all Nash equilibria of the game, of specific best-response polytopes contains only a path that exhibits monotonicity with respect to λ. Additionally, they introduced an algorithm that employs binary search on the parameter to compute an *exact* Nash equilibrium for the rank-1 game. However, a limitation of this method is its inability to handle non-degenerate games. In their subsequent work (Adsul et al., 2021), while retaining their structural insights and building on (Adler and Monteiro, 1992), they refined their approach. They proposed a polynomial-time algorithm, similar in spirit to the one in (Adsul et al., 2011), to compute a Nash equilibrium for *any* rank-1 game.

Definition 3.3.2. Let $\mathcal{N}$ be the set of the Nash equilibria of the parameterized zero-sum game defined as follows.

$$\mathcal{N} \coloneqq \{(\lambda, \boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R} \times \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \,|\, (\boldsymbol{x}, \boldsymbol{y}) \text{ is a NE of } \left(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda \boldsymbol{b}^\top\right)\} \quad (3.1)$$

The following key lemma, essential for the algorithm in (Adsul et al., 2021), shows that the set $\mathcal{N}$ exhibits monotonicity with respect to λ.

Lemma 3.3.3 (Lemma 17 Adsul et al. (2021))**.** *Let $\underline{\lambda} \leq \overline{\lambda}$, $\underline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_1 \in \Delta_{m_1}$ and $\underline{\boldsymbol{x}}_2, \overline{\boldsymbol{x}}_2 \in \Delta_{m_2}$ so that for $\mathcal{N}$ in* (3.1)

$$(\underline{\lambda}, \underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2) \in \mathcal{N}, \quad \underline{\lambda} \leq \underline{\boldsymbol{x}}_1^\top \boldsymbol{a}, \quad \left(\overline{\lambda}, \overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2\right) \in \mathcal{N}, \quad \overline{\lambda} \geq \overline{\boldsymbol{x}}_1^\top \boldsymbol{a}. \quad (3.2)$$

Then $\boldsymbol{x}_1^\top \boldsymbol{a} = \lambda$ for some $(\lambda, \boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{N}$ with $\lambda \in [\underline{\lambda}, \overline{\lambda}]$.

This property serves as the invariant maintained by the algorithm outlined in (Adsul et al., 2021). The algorithm essentially employs a binary search with respect to the parameter λ. In brief, at each iteration, it considers the midpoint, $\lambda = (\underline{\lambda} + \overline{\lambda})/2$, and, based on its value, it solves a series of linear programs. These either return a Nash equilibrium of the zero-sum game $\left(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda \boldsymbol{b}^\top\right)$ for which $\boldsymbol{x}_1^\top \boldsymbol{a} = \lambda$ or indicate that none of the Nash equilibrium strategies satisfy this equation. Importantly, in the latter case, and due to the monotonic nature with respect to the parameter, the sign of the difference $(\boldsymbol{x}^\top \boldsymbol{a} - \lambda)$ should be maintained. Therefore, depending on it, either $\underline{\lambda}$ or $\overline{\lambda}$ is updated accordingly. For instance, in case $\boldsymbol{x}_1^\top \boldsymbol{a} - \lambda < 0$ then $\underline{\lambda} = \lambda$; otherwise, it sets $\overline{\lambda} = \lambda$ [1]. According to Lemma 3.3.3, this step is safe as there exists a Nash equilibrium $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ such that $\boldsymbol{x}_1^\top \boldsymbol{a} = \lambda$

[1] We have intentionally omitted certain technical details of this step that do not impact the core aspects of the algorithm.

and $\lambda \in [\underline{\lambda}, \overline{\lambda}]$. In Lemma 3.4.3, we extend this property to the case of an *approximate* Nash equilibrium by leveraging the convexity of Nash equilibria in zero-sum games.

3.3.2 Min-max formulation and variational inequalities

We use the min-max and optimistic mirror-descent background introduced in Section 2.10. In this chapter, the relevant saddle-point problem arises from the parameterized zero-sum game associated with a rank-1 bimatrix game, as demonstrated in Corollary 3.3.1. We recall the additional variational-inequality notation needed for the analysis of rank-1 games.

A Nash equilibrium in zero-sum games represents a solution to a saddle point problem which in its general form (not only for bilinear functions) is

$$\min_{\boldsymbol{x}_1 \in \mathcal{X}_1} \max_{\boldsymbol{y}_2 \in \mathcal{X}_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2) \tag{3.3}$$

where $\mathcal{X}_1, \mathcal{X}_2$ are convex and compact subsets of an Euclidean n-dimensional (m-dimensional resp.) space, and f is a continuously differentiable function, i.e. a smooth function. However, in practical scenarios, such as those in Generative Adversarial Networks (GANs) (Daskalakis et al., 2018), the function f does not meet the necessary convex-concave property, and so most standard optimization techniques for (3.3) do not to apply. As a result, a significant amount of research has focused instead on different notions of local min-max solutions (Daskalakis and Panageas, 2018a; Jin et al., 2020). The results in (Daskalakis et al., 2021; Bernasconi and Castiglioni, 2026), however, establish that for general smooth objectives, the computation of even approximate first-order locally optimal min-max solutions is intractable. In light of these intractability results, research has shifted its focus to investigate structural properties that circumvent this barrier, enabling the efficient computation of a solution. This will usually involve imposing some (mild or not) assumptions on the objective function. An example of such an assumption is the existence of a solution to (MVI), as presented below.

Nash Gap The goal is to find a point $(\boldsymbol{x}_1^\star, \boldsymbol{x}_1^\star) \in \mathcal{X}_1 \times \mathcal{X}_2$ that is *close* to the set of Nash equilibria. The notion of closeness is defined in terms of the best response gap, not the distance to the set of NE[2], and is defined with respect to the (Nash) gap as

$$\texttt{Gap}(\boldsymbol{x}_1^\star, \boldsymbol{x}_2^\star) \coloneqq \max_{\boldsymbol{x}_2 \in \mathcal{X}_2} f(\boldsymbol{x}_1^\star, \boldsymbol{x}_2) - \min_{\boldsymbol{x}_1 \in \mathcal{X}_1} f(\boldsymbol{x}_1, \boldsymbol{x}_2^\star), \tag{Gap}$$

which is always non-negative and zero if and only if $(\boldsymbol{x}_1^\star, \boldsymbol{x}_2^\star)$ is a NE.

Variational inequalities We introduce the set $\mathcal{Z} = \mathcal{X}_1 \times \mathcal{X}_2$ and the operator F associated with the function f, defined as $F(\boldsymbol{z}) = F(\left[\begin{smallmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{smallmatrix}\right]) = \begin{bmatrix} \nabla_{\boldsymbol{x}_1} f(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ -\nabla_{\boldsymbol{x}_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2) \end{bmatrix}$. It is readily verified that a saddle point problem (3.3) can be framed as any variational inequality problem; this unifying treatment enables the application of various techniques and results drawn from the vast literature of variational inequalities (Facchinei and Pang, 2003). Given the compact convex set $\mathcal{Z}$ and an operator F, the *Stampacchia Variational Inequality Equation* (SVI) problem consists in finding $\boldsymbol{z}^\star \in \mathcal{Z}$ such that:

$$\langle F(\boldsymbol{z}^\star), \boldsymbol{z} - \boldsymbol{z}^\star \rangle \geq 0 \quad \text{for all } \boldsymbol{z} \in \mathcal{Z} \tag{SVI}$$

In this case, $\boldsymbol{z}^\star$ is referred to as the *strong* solution to variational inequality problem corresponding to F. From a game-theoretic standpoint, $\boldsymbol{z}^\star$ corresponds to a Nash equilibrium of the underlying game. As the existence of a Nash equilibrium is always guaranteed to exist (Nash, 1950), we can safely assume that there exists a solution to (SVI).

Most of the existing literature on variational inequality problems (Facchinei and Pang, 2003) has focused on the monotone case, i.e. F is a monotone operator. In this context, the following equivalence can be shown: a solution to the (SVI) is also a solution to the *Minty*

[2]This notion of closeness is commonly referred to in the literature as weak, while the closeness in terms of the distance to a Nash Equilibrium is known as the strong notion Etessami and Yannakakis (2010).

Variational Inequality problem (MVI).

$$\langle F(\boldsymbol{z}), \boldsymbol{z} - \boldsymbol{z}^\star \rangle \geq 0 \quad \text{for all } \boldsymbol{z} \in \mathcal{Z} \tag{MVI}$$

In settings beyond the monotone case, all that can be said is that the set of solutions for (MVI) is a subset of the set of solutions for (SVI). An important (Minty) criterion that ensures tractability is the existence of a solution to (MVI). Mertikopoulos et al. (2018a) introduced the concept of coherence (see Definition 2.1 therein), which holds true if every saddle point of f (3.3), and in turn a solution to (SVI), is a solution to (MVI) and vice versa. In (Mertikopoulos et al., 2018a; Song et al., 2020; Zhou et al., 2020) (and references therein) it has been established that if the Minty criterion is satisfied (e.g. bilinear problems, quasi-convex-concave objectives), then the extra gradient algorithm (Korpelevich, 1976) does converge to a solution of saddle point problem. In recent works, such as (Diakonikolas et al., 2021; Pethick et al., 2022), a less stringent criterion, namely the weak Minty criterion, has been explored, and convergence to a solution of (3.3) was also ensured. Nevertheless, as we will demonstrate shortly in the Section 3.4.1, even relatively simple rank-1 games fail to satisfy the (weak/strong) Minty criterion.

3.4 Efficiently learning in rank-1 games

In this section, we sketch our main findings concerning the learning of Nash equilibria in rank-1 games. We start by presenting two illustrative examples that highlight the inherent challenges of these games. Despite their apparent simplicity, conventional first-order methods fail to converge to a Nash equilibrium. In Section 3.4.2, we introduce an efficient algorithm for learning a NE, that uses and extends the structural properties detailed in Section 3.3.1.

3.4.1 Technical challenges

The rank-reduction method of Corollary 3.3.1 transforms a rank-1 game $(\mathbf{A}, \mathbf{B})$ into the parameterized zero-sum game $(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda\boldsymbol{b}^\top)$. A natural question is whether the set of Nash equilibria of this zero-sum game is contained in the set of Nash equilibria of the original rank-1 game. In the case of an *exact* NE $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$, Corollary 3.3.1 implies that the answer is negative whenever the additional condition $\boldsymbol{x}_1^{*\top}\boldsymbol{a} = \lambda$ is not satisfied. For *approximate* NE, however, the situation remains unclear. More specifically, can one obtain an approximate equilibrium of the original rank-1 game solely by solving the zero-sum game and computing an approximate NE there? The following example answers this question in the negative, even when the correct value λ is known exactly.

Example 3.4.1. In this example, we demonstrate that an approximate Nash equilibrium of the zero-sum game $(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A} + \mathbf{1}\lambda\boldsymbol{b}^\top)$ does not constitute an approximate Nash Equilibrium of the rank-1 game, even when we know a *correct* value of λ.

Let

$$\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top = \begin{bmatrix} 1 & 1 \\ 1-\epsilon & 1-\epsilon \end{bmatrix}.$$

We first examine the set of Nash equilibria of the zero-sum game $\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda\boldsymbol{b}^\top\right)$. This game admits a family of exact Nash equilibria of the form

$$(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*) = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \mu \\ 1-\mu \end{bmatrix} \right) \quad \text{for any } \mu \in [0, 1].$$

As shown in (3.4), the strategy profile $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ attains the maximum payoff for player 1 and the minimum payoff for player 2; therefore, neither player has a profitable deviation.

Additionally,

$$(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2) = \left(\begin{bmatrix} \frac{1}{2} \\ \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{1}{2} \\ \\ \frac{1}{2} \end{bmatrix} \right)$$

is an $\frac{\epsilon}{2}$-approximate Nash equilibrium of the zero-sum game. Indeed, as shown in (3.6), its utility is at distance $\frac{\epsilon}{2}$ from the maximum payoff achievable by player 1 and from the minimum payoff achievable by player 2. Hence, by definition, it is an $\frac{\epsilon}{2}$-approximate Nash equilibrium.

$$\boldsymbol{x}_1^{*\top} \left(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top \right) \boldsymbol{x}_2^* = 1, \tag{3.4}$$

$$\tilde{\boldsymbol{x}}_1^\top \left(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top \right) \tilde{\boldsymbol{x}}_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \\ 1-\epsilon & 1-\epsilon \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \\ \frac{1}{2} \end{bmatrix} \tag{3.5}$$

$$= 1 - \frac{\epsilon}{2}. \tag{3.6}$$

We now choose $\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Thus, setting $\lambda = 1 \left(\coloneqq \boldsymbol{x}_1^{*\top} \boldsymbol{a} \right)$ the rank-1 game is defined as follows.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ \\ 1-\epsilon & 1-\epsilon \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ \\ 3-\epsilon & 2-\epsilon \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} -1 & -1 \\ \\ 1+\epsilon & \epsilon \end{bmatrix} \tag{3.7}$$

We now verify that $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ is no longer an $\frac{\epsilon}{2}$-approximate Nash equilibrium of the original rank-1 game, since the approximate best-response condition fails for the column player. In particular, against player 1's strategy $\tilde{\boldsymbol{x}}_1$, playing column 1 yields a strictly better response.

$$\tilde{\boldsymbol{x}}_1^\top \mathbf{B} \tilde{\boldsymbol{x}}_2 - \tilde{\boldsymbol{x}}_1^\top \mathbf{B} \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix} = \left(-\frac{1}{4} + \frac{\epsilon}{2} \right) - \left(\frac{\epsilon}{2} \right) = -\frac{1}{4} \tag{3.8}$$

Specifically, this example shows that, regardless of whether the value of λ is known, an approximate NE $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ of the zero-sum game $(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda\boldsymbol{b}^\top)$ cannot be a NE of the original rank-1 game unless it also (approximately) satisfies $(\tilde{\boldsymbol{x}}_1)^\top \boldsymbol{a} = \lambda$. In other words, it is necessary to seek approximate Nash equilibria that also satisfy this additional condition, at least approximately, as we formally prove in Lemma 3.4.1.

As described in Section 3.3.2, a recent line of work investigates structural properties of games that enable learning procedures to converge to a Nash equilibrium, beyond the case where the operator F is monotone. One compelling criterion that guarantees convergence is the existence of a solution to the Minty variational inequality (MVI) problem (Mertikopoulos et al., 2018a), or to the weak MVI problem (Diakonikolas et al., 2021). Unfortunately, but importantly for the scope of this work, it fails to hold. In this example, we show that there exists a game for which neither of these criteria holds.

Example 3.4.2. Let $\boldsymbol{z}^\star = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ be a Nash equilibrium of the rank-1 bimatrix game

$$(\mathbf{A}, \mathbf{B}) = \left(\mathbf{A}, -\mathbf{A} + \boldsymbol{a}\boldsymbol{b}^\top\right).$$

We introduce the operator F by

$$F(\boldsymbol{z}) = F\left(\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}\right) = \begin{bmatrix}-\mathbf{A}\boldsymbol{x}_2\\ -\mathbf{B}^\top\boldsymbol{x}_1\end{bmatrix},$$

that is, the gradient operator associated with the underlying rank-1 game. Since both players aim to maximize their payoffs, the gradients appear with a negative sign. Accordingly, an (ϵ-approximate) minimizer $\boldsymbol{z}^* = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ of F should satisfy the corresponding first-order condition, namely $\langle F(\boldsymbol{z}^*), \boldsymbol{z} - \boldsymbol{z}^*\rangle \geq 0$ (resp. $\geq -\epsilon$) $\forall\, \boldsymbol{z} \in \mathcal{Z}$. In this example, however, we examine whether the two important variational inequalities (Minty conditions) admit points that satisfy them. In what follows, we show that these conditions fail to hold.

$$\text{(MVI)} \quad \langle F(\boldsymbol{z}), \boldsymbol{z} - \boldsymbol{z}^\star \rangle \geq 0 \; \forall \, \boldsymbol{z} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \tag{3.9}$$

$$\text{(wMVI)} \quad \langle F(\boldsymbol{z}), \boldsymbol{z} - \boldsymbol{z}^* \rangle \geq -\frac{\rho}{2} \left\| F(\boldsymbol{z}) \right\|_2^2, \quad \forall \, \boldsymbol{z} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \text{ where } \rho \in \left[0, \frac{1}{4L} \right). \tag{3.10}$$

The parameter L is the Lipschitz constant of the operator F with respect to the ℓ_2 norm:

$$\|F(\boldsymbol{z}) - F(\boldsymbol{z}')\|_2 \leq L \|\boldsymbol{z} - \boldsymbol{z}'\|_2 \quad \forall \boldsymbol{z}, \boldsymbol{z}' \in \mathcal{Z}.$$

Now, since the solution set of (MVI) is contained in the solution set of (SVI), it suffices to examine $\boldsymbol{z}^\star = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ corresponding to the Nash equilibria of the following game:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & -2 \\ -1 & 0 \end{bmatrix}.$$

For any $\boldsymbol{z} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, we have

$$\langle F(\boldsymbol{z}), \boldsymbol{z} - \boldsymbol{z}^\star \rangle = \begin{bmatrix} -\mathbf{A}\boldsymbol{x}_2 \\ -\mathbf{B}^\top \boldsymbol{x}_1 \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{x}_1 - \boldsymbol{x}^\star{}_1 \\ \boldsymbol{x}_2 - \boldsymbol{x}^\star{}_2 \end{bmatrix} \tag{3.11}$$

$$= \langle \boldsymbol{x}_1 - \boldsymbol{x}^\star{}_1, -\mathbf{A}\boldsymbol{x}_2 \rangle + \left\langle \boldsymbol{x}_2 - \boldsymbol{x}^\star{}_2, -\mathbf{B}^\top \boldsymbol{x}_1 \right\rangle. \tag{3.12}$$

The Lipschitz constant is given by the spectral norm of the block matrix $\mathbf{C} = \left[\begin{smallmatrix} \mathbf{0} & -\mathbf{A} \\ -\mathbf{B}^\top & \mathbf{0} \end{smallmatrix} \right]$. Therefore, $L = \|\mathbf{C}\|_2 = 2.288$. Moreover, by exhaustively examining the strategy profiles,

we see that this game has two pure Nash equilibria.

$$\left(\boldsymbol{x}_1^1, \boldsymbol{x}_2^1\right) = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \quad \text{and} \quad \left(\boldsymbol{x}_1^2, \boldsymbol{x}_2^2\right) = \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right), \tag{3.13}$$

and one mixed equilibrium

$$\left(\boldsymbol{x}_1^3, \boldsymbol{x}_2^3\right) = \left(\begin{bmatrix} \frac{1}{4} \\ \frac{3}{4} \end{bmatrix}, \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \right). \tag{3.14}$$

It is important to note that, when $\boldsymbol{z} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a Nash equilibrium, the Minty condition induced by F reduces exactly to the Nash equilibrium condition for $\boldsymbol{z}$. Indeed, the two constituent parts of this inequality correspond precisely to the players' best-response conditions.

1. $\boldsymbol{z}^* = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*) = (\boldsymbol{x}_1^1, \boldsymbol{x}_2^1)$ and $\boldsymbol{z} = (\boldsymbol{x}_1, \boldsymbol{x}_2) = (\boldsymbol{x}_1^2, \boldsymbol{x}_2^2)$.

$$\begin{aligned} \langle \boldsymbol{x}_1 - \boldsymbol{x}^\star{}_1, -\mathbf{A}\boldsymbol{x}_2 \rangle &= -\left\langle \boldsymbol{x}_1^2 - \boldsymbol{x}_1^1, \mathbf{A}\boldsymbol{x}_2^2 \right\rangle && (3.15) \\ &= -\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -1 < 0, && (3.16) \\ \left\langle \boldsymbol{x}_2 - \boldsymbol{x}_2^*, -\mathbf{B}^\top \boldsymbol{x}_1 \right\rangle &= -\left\langle \boldsymbol{x}_2^2 - \boldsymbol{x}_2^1, \mathbf{B}^\top \boldsymbol{x}_1^2 \right\rangle && (3.17) \\ &= -\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} = -1 < 0. && (3.18) \end{aligned}$$

2. $\boldsymbol{z}^* = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*) = (\boldsymbol{x}_1^2, \boldsymbol{x}_2^2)$ and $\boldsymbol{z} = (\boldsymbol{x}_1, \boldsymbol{x}_2) = (\boldsymbol{x}_1^1, \boldsymbol{x}_2^1)$.

$$\langle \boldsymbol{x}_1 - \boldsymbol{x}_1^*, -\mathbf{A}\boldsymbol{x}_2 \rangle = -\left\langle \boldsymbol{x}_1^1 - \boldsymbol{x}_1^2, \mathbf{A}\boldsymbol{x}_2^1 \right\rangle \tag{3.19}$$

$$= -\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = -1 < 0, \tag{3.20}$$

$$\left\langle \boldsymbol{x}_2 - \boldsymbol{x}_2^*, -\mathbf{B}^\top \boldsymbol{x}_1 \right\rangle = -\left\langle \boldsymbol{x}_2^1 - \boldsymbol{x}_2^2, \mathbf{B}^\top \boldsymbol{x}_1^1 \right\rangle \tag{3.21}$$

$$= -\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = -3 < 0. \tag{3.22}$$

3. $\boldsymbol{z}^* = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*) = (\boldsymbol{x}_1^3, \boldsymbol{x}_2^3)$ and $\boldsymbol{z} = (\boldsymbol{x}_1, \boldsymbol{x}_2) = (\boldsymbol{x}_1^1, \boldsymbol{x}_2^1)$.

$$\langle \boldsymbol{x}_1 - \boldsymbol{x}_1^*, -\mathbf{A}\boldsymbol{x}_2 \rangle = -\left\langle \boldsymbol{x}_1^1 - \boldsymbol{x}_1^3, \mathbf{A}\boldsymbol{x}_2^1 \right\rangle \tag{3.23}$$

$$= -\begin{bmatrix} \frac{3}{4} & -\frac{3}{4} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = -\frac{3}{4} < 0, \tag{3.24}$$

$$\left\langle \boldsymbol{x}_2 - \boldsymbol{x}_2^*, -\mathbf{B}^\top \boldsymbol{x}_1 \right\rangle = -\left\langle \boldsymbol{x}_2^1 - \boldsymbol{x}_2^3, \mathbf{B}^\top \boldsymbol{x}_1^1 \right\rangle \tag{3.25}$$

$$= -\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = -\frac{6}{4} < 0. \tag{3.26}$$

Summing the two terms in each of the cases considered above, we obtain

$$\langle F(\boldsymbol{z}_2), \boldsymbol{z}_2 - \boldsymbol{z}_1 \rangle = -2, \quad \langle F(\boldsymbol{z}_1), \boldsymbol{z}_1 - \boldsymbol{z}_2 \rangle = -4, \quad \langle F(\boldsymbol{z}_1), \boldsymbol{z}_1 - \boldsymbol{z}_3 \rangle = -\frac{9}{4}.$$

Furthermore, it is straightforward to verify that $\|F(\boldsymbol{z}_1)\|_2^2 = 6,\ \|F(\boldsymbol{z}_2)\|_2^2 = 2,\ \|F(\boldsymbol{z}_3)\|_2^2 = 1$. We conclude that neither the MVI nor the wMVI condition is satisfied.

Algorithm 1. Optimistic Multiplicative Weights Update (OMWU)

Input: Payoff matrices $\mathbf{A} \in \mathbb{R}^{m_1 \times m_2}$, $\mathbf{B} \in \mathbb{R}^{m_1 \times m_2}$, initial strategies $\boldsymbol{x}_1^{(1)} \in \Delta_{m_1}$, $\boldsymbol{x}_2^{(1)} \in \Delta_{m_2}$, step size $\eta > 0$

Output: Strategy sequence $\{(\boldsymbol{x}_1^{(t)}, \boldsymbol{x}_2^{(t)})\}_{t \geq 1}$

Initialize $\boldsymbol{u}_1^{(0)} \leftarrow \mathbf{A}\boldsymbol{x}_2^{(1)},\ \boldsymbol{u}_2^{(0)} \leftarrow \mathbf{B}^\top \boldsymbol{x}_1^{(1)}$.

for $t = 1, 2, \ldots, T-1$ **do**

  Compute current payoff vectors $\boldsymbol{u}_1^{(t)} \leftarrow \mathbf{A}\boldsymbol{x}_2^{(t)},\ \boldsymbol{u}_2^{(t)} \leftarrow \mathbf{B}^\top \boldsymbol{x}_1^{(t)}$.

  for $i = 1, 2, \ldots, m_1$ **do**

$$\boldsymbol{x}_1^{(t+1)}[i] \leftarrow \frac{\boldsymbol{x}_1^{(t)}[i] \exp\Big(\eta\big(2\boldsymbol{u}_1^{(t)}[i] - \boldsymbol{u}_1^{(t-1)}[i]\big)\Big)}{\sum_{k=1}^{m_1} \boldsymbol{x}_1^{(t)}[k] \exp\Big(\eta\big(2\boldsymbol{u}_1^{(t)}[k] - \boldsymbol{u}_1^{(t-1)}[k]\big)\Big)}.$$

  end

  for $j = 1, 2, \ldots, m_2$ **do**

$$\boldsymbol{x}_2^{(t+1)}[j] \leftarrow \frac{\boldsymbol{x}_2^{(t)}[j] \exp\Big(\eta\big(2\boldsymbol{u}_2^{(t)}[j] - \boldsymbol{u}_2^{(t-1)}[j]\big)\Big)}{\sum_{\ell=1}^{m_2} \boldsymbol{x}_2^{(t)}[\ell] \exp\Big(\eta\big(2\boldsymbol{u}_2^{(t)}[\ell] - \boldsymbol{u}_2^{(t-1)}[\ell]\big)\Big)}.$$

  end

end

3.4.2 Main Results

In this section, we introduce our method and sketch the main ingredients required for the proof of Theorem 3.4.1. To facilitate the presentation, we break it down into three main steps. The first step consists of showing that only specific approximate Nash equilibria of the parameterized zero-sum correspond to approximate Nash equilibria of the underlying rank-1 game. Then, we formulate a suitable min-max optimization problem whose stationary points are extendable –under conditions– to NE. Finally, we present an efficient algorithm for solving the optimization problem.

The following lemma relies on the correspondence established in Corollary 3.3.1. Although this appears to be applicable solely to *exact* Nash equilibria, we demonstrate that by relaxing both conditions, it also extends to *approximate* Nash equilibria.

Lemma 3.4.1. *Let $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ be an ϵ-approximate Nash equilibrium of $\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda\boldsymbol{b}^\top\right)$ and assume that $\left|\boldsymbol{x}_1^\top\boldsymbol{a} - \lambda\right| \le \epsilon$. Then $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a (3ϵ)-approximate Nash equilibrium of the rank-1 game $(\mathbf{A},\, -\mathbf{A} + \boldsymbol{a}\boldsymbol{b}^\top)$.*

Proof. Since $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is an ϵ-approximate Nash equilibrium of the zero-sum game with payoff matrix $\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top$, the row player satisfies

$$\boldsymbol{x}_1^\top(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top)\boldsymbol{x}_2 \ge \boldsymbol{x}_1'^\top(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top)\boldsymbol{x}_2 - \epsilon \qquad \forall\, \boldsymbol{x}_1' \in \Delta_{m_1}. \tag{3.27}$$

Now observe that for every $\boldsymbol{x}_1' \in \Delta_{m_1}$, $\boldsymbol{x}_1'^\top\mathbb{1} = 1$, and hence $\boldsymbol{x}_1'^\top(\mathbb{1}\lambda\boldsymbol{b}^\top)\boldsymbol{x}_2 = (\boldsymbol{x}_1'^\top\mathbb{1})\lambda\boldsymbol{b}^\top\boldsymbol{x}_2 = \lambda\boldsymbol{b}^\top\boldsymbol{x}_2$. Therefore, the term $\lambda\boldsymbol{b}^\top\boldsymbol{x}_2$ is independent of $\boldsymbol{x}_1'$, and cancels from (3.27). We obtain

$$\boldsymbol{x}_1^\top\mathbf{A}\boldsymbol{x}_2 \ge \boldsymbol{x}_1'^\top\mathbf{A}\boldsymbol{x}_2 - \epsilon \qquad \forall\, \boldsymbol{x}_1' \in \Delta_{m_1}. \tag{3.28}$$

This is exactly the ϵ-best-response condition for the row player in the game $(\mathbf{A}, -\mathbf{A} + \boldsymbol{a}\boldsymbol{b}^\top)$. Next, the ϵ-best-response condition for the column player in the zero-sum game implies that

$$\boldsymbol{x}_1^\top(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top)\boldsymbol{x}_2 \le \boldsymbol{x}_1^\top(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top)\boldsymbol{x}_2' + \epsilon \qquad \forall\, \boldsymbol{x}_2' \in \Delta_{m_2}. \tag{3.29}$$

Expanding the left-hand side and rearranging terms, we obtain

$$\boldsymbol{x}_1^\top\mathbf{A}\boldsymbol{x}_2 \le \boldsymbol{x}_1^\top\mathbf{A}\boldsymbol{x}_2' - \lambda\boldsymbol{b}^\top(\boldsymbol{x}_2 - \boldsymbol{x}_2') + \epsilon \qquad \forall\, \boldsymbol{x}_2' \in \Delta_{m_2}. \tag{3.30}$$

Using the assumption $\left|\boldsymbol{x}_1^\top\boldsymbol{a} - \lambda\right| \le \epsilon$, we may write $\lambda = \boldsymbol{x}_1^\top\boldsymbol{a} + \delta$ for some $\delta \in \mathbb{R}$ with $|\delta| \le \epsilon$. Substituting this into (3.30) yields

$$\boldsymbol{x}_1^\top\mathbf{A}\boldsymbol{x}_2 \le \boldsymbol{x}_1^\top\mathbf{A}\boldsymbol{x}_2' - (\boldsymbol{x}_1^\top\boldsymbol{a} + \delta)\,\boldsymbol{b}^\top(\boldsymbol{x}_2 - \boldsymbol{x}_2') + \epsilon,$$

which after arranging the terms appropriately,

$$\boldsymbol{x}_1^\top(\mathbf{A} - \boldsymbol{a}\boldsymbol{b}^\top)\boldsymbol{x}_2 \leq \boldsymbol{x}_1^\top(\mathbf{A} - \boldsymbol{a}\boldsymbol{b}^\top)\boldsymbol{x}_2' - \delta\, \boldsymbol{b}^\top(\boldsymbol{x}_2 - \boldsymbol{x}_2') + \epsilon.$$

To bound the error term, we combine the assumption $|\delta| \leq \epsilon$ with Hölder's inequality to obtain

$$\left|\delta\, \boldsymbol{b}^\top(\boldsymbol{x}_2 - \boldsymbol{x}_2')\right| \leq \epsilon\, \|\boldsymbol{b}\|_\infty \|\boldsymbol{x}_2 - \boldsymbol{x}_2'\|_1 .$$

Since $|\delta| \leq \epsilon$, $\|\boldsymbol{b}\|_\infty \leq 1$, and $\boldsymbol{x}_2, \boldsymbol{x}_2' \in \Delta_{m_2}$, we have

$$\left|\delta\, \boldsymbol{b}^\top(\boldsymbol{x}_2 - \boldsymbol{x}_2')\right| \leq \epsilon\, \|\boldsymbol{b}\|_\infty \|\boldsymbol{x}_2 - \boldsymbol{x}_2'\|_1 \leq \epsilon(\|\boldsymbol{x}_2\|_1 + \|\boldsymbol{x}_2'\|_1) = 2\epsilon.$$

Substituting this bound into the previous inequality yields

$$\boldsymbol{x}_1^\top(\mathbf{A} - \boldsymbol{a}\boldsymbol{b}^\top)\boldsymbol{x}_2 \leq \boldsymbol{x}_1^\top(\mathbf{A} - \boldsymbol{a}\boldsymbol{b}^\top)\boldsymbol{x}_2' + 3\epsilon \qquad \forall\, \boldsymbol{x}_2' \in \Delta_{m_2}. \tag{3.31}$$

This is exactly the (3ϵ)-best-response condition for the column player in the game $(\mathbf{A}, -\mathbf{A} + \boldsymbol{a}\boldsymbol{b}^\top)$. Combining (3.28) and (3.31), we conclude that $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a (3ϵ)-approximate Nash equilibrium of the game

$$(\mathbf{A},\, -\mathbf{A} + \boldsymbol{a}\boldsymbol{b}^\top). \qquad \square$$

A study of Lemma 3.4.1 leads us to the following observation: to compute an approximate Nash equilibrium of the rank-1 game, we should seek a Nash equilibrium of the zero-sum game which simultaneously minimizes the distance $(\boldsymbol{x}_1^\top \boldsymbol{a} - \lambda)$. By virtue of this observation, we formulate the following saddle-point (min-max) optimization problem.

Definition 3.4.1 (Parameterized saddle-point function)**.** For each $\lambda \in \mathbb{R}$, define

$$f_\lambda(\boldsymbol{x}_1, \boldsymbol{x}_2) \coloneqq \boldsymbol{x}_1^\top \left(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top\right) \boldsymbol{x}_2 - \frac{1}{2}\left(\boldsymbol{x}_1^\top \boldsymbol{a} - \lambda\right)^2. \tag{PSP}$$

The function f_λ is concave in $\boldsymbol{x}_1$ (maximizer) and convex in $\boldsymbol{x}_2$; indeed, it is affine in $\boldsymbol{x}_2$ (minimizer).

Lemma 3.4.2. *Let $(\boldsymbol{x}_1^\star, \boldsymbol{x}_2^\star)$ be an ϵ-approximate Nash equilibrium of $(\mathbf{A}-\mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A}+\mathbb{1}\lambda\boldsymbol{b}^\top)$ satisfying $\left|\boldsymbol{x}_1^{\star\top}\boldsymbol{a} - \lambda\right| \le \epsilon$. Then any ϵ-approximate stationary point $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ of f_λ satisfies the following:*

1. $\left|\tilde{\boldsymbol{x}}_1^\top \boldsymbol{a} - \lambda\right| \le \sqrt{10\epsilon}$.
2. *$(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ is a $(8\sqrt{\epsilon})$ -approximate Nash equilibrium of $(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda\boldsymbol{b}^\top)$.*

Proof. Since $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ is an ϵ-approximate stationary point of f_λ, it satisfies the variational inequality associated with the saddle-point problem. We recall that $\boldsymbol{x}_1$ is the maximizing variable, whereas $\boldsymbol{x}_2$ is the minimizing variable. Therefore, the corresponding conditions are

$$\langle \nabla_{\boldsymbol{x}_1} f_\lambda(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2), \boldsymbol{x}_1' - \tilde{\boldsymbol{x}}_1\rangle \le \epsilon, \quad \forall\, \boldsymbol{x}_2' \in \Delta_{m_2}, \quad \langle \nabla_{\boldsymbol{x}_2} f_\lambda(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2), \boldsymbol{x}_2' - \tilde{\boldsymbol{x}}_2\rangle \ge -\epsilon, \quad \forall\, \boldsymbol{x}_2' \in \Delta_{m_2}.$$

By expanding the gradients, we obtain the following two variational inequalities:

VI (1) $\left\langle \left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 - \left(\tilde{\boldsymbol{x}}_1^\top\boldsymbol{a} - \lambda\right)\boldsymbol{a}, \boldsymbol{x}_1' - \tilde{\boldsymbol{x}}_1\right\rangle \le \epsilon, \quad \forall\, \boldsymbol{x}_1' \in \Delta_{m_1}.$

VI (2) $\left\langle \left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)^\top \tilde{\boldsymbol{x}}_1, \boldsymbol{x}_2' - \tilde{\boldsymbol{x}}_2\right\rangle \ge -\epsilon, \quad \forall\, \boldsymbol{x}_2' \in \Delta_{m_2}.$

We first exploit the fact that $f_\lambda(\cdot, \boldsymbol{x}_2)$ is concave for every $\boldsymbol{x}_2$, and therefore

$$f_\lambda(\boldsymbol{x}_1', \tilde{\boldsymbol{x}}_2) \le f_\lambda(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2) + \langle \nabla_{\boldsymbol{x}_1} f_\lambda(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2), \boldsymbol{x}_1' - \tilde{\boldsymbol{x}}_1\rangle \qquad \forall\, \boldsymbol{x}_1' \in \Delta_{m_1}.$$

Setting $\boldsymbol{x}_1' = \boldsymbol{x}_1^\star$, and using VI (1) together with $\left|\boldsymbol{x}_1^{\star\top}\boldsymbol{a} - \lambda\right| \le \epsilon$, we obtain

$$\begin{aligned}
\boldsymbol{x}_1^{\star\top}\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 &\le \tilde{\boldsymbol{x}}_1^\top\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 - \frac{1}{2}\left(\tilde{\boldsymbol{x}}_1^\top\boldsymbol{a} - \lambda\right)^2 + \frac{1}{2}\left(\boldsymbol{x}_1^{\star\top}\boldsymbol{a} - \lambda\right)^2 + \epsilon && (3.32)\\
&\le \tilde{\boldsymbol{x}}_1^\top\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 - \frac{1}{2}\left(\tilde{\boldsymbol{x}}_1^\top\boldsymbol{a} - \lambda\right)^2 + \frac{\epsilon^2}{2} + \epsilon && (3.33)\\
&\le \tilde{\boldsymbol{x}}_1^\top\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 - \frac{1}{2}\left(\tilde{\boldsymbol{x}}_1^\top\boldsymbol{a} - \lambda\right)^2 + 2\epsilon. && (3.34)
\end{aligned}$$

Moreover, $f_\lambda(\boldsymbol{x}_1, \cdot)$ is linear for every $\boldsymbol{x}_1 \in \Delta_{m_1}$, and hence

$$f_\lambda(\tilde{\boldsymbol{x}}_1, \boldsymbol{x}_2') = f_\lambda(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2) + \langle \nabla_{\boldsymbol{x}_2} f_\lambda(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2), \boldsymbol{x}_2' - \tilde{\boldsymbol{x}}_2\rangle \qquad \forall\, \boldsymbol{x}_2' \in \Delta_{m_2}.$$

Setting $\boldsymbol{x}_2' = \boldsymbol{x}_2^\star$, and using VI (2), we obtain

$$\begin{aligned}
\tilde{\boldsymbol{x}}_1^\top\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\boldsymbol{x}_2^\star &= \tilde{\boldsymbol{x}}_1^\top\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 - \frac{1}{2}\left(\tilde{\boldsymbol{x}}_1^\top\boldsymbol{a} - \lambda\right)^2 + \frac{1}{2}\left(\tilde{\boldsymbol{x}}_1^\top\boldsymbol{a} - \lambda\right)^2 - \epsilon && (3.35)\\
&\ge \tilde{\boldsymbol{x}}_1^\top\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 - \epsilon. && (3.36)
\end{aligned}$$

Since $(\boldsymbol{x}_1^\star, \boldsymbol{x}_2^\star)$ is an ϵ-approximate Nash equilibrium of the zero-sum game, we also have

$$\begin{aligned}
\boldsymbol{x}_1^{\star\top}\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\boldsymbol{x}_2^\star &\ge \tilde{\boldsymbol{x}}_1^\top\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\boldsymbol{x}_2^\star - \epsilon, && (3.37)\\
\boldsymbol{x}_1^{\star\top}\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\boldsymbol{x}_2^\star &\le \boldsymbol{x}_1^{\star\top}\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 + \epsilon. && (3.38)
\end{aligned}$$

We now combine the above inequalities:

$$\begin{aligned}
\tilde{\boldsymbol{x}}_1^\top\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 &\le \tilde{\boldsymbol{x}}_1^\top\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\boldsymbol{x}_2^\star + \epsilon && \text{by (3.36)}\\
&\le \boldsymbol{x}_1^{\star\top}\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\boldsymbol{x}_2^\star + 2\epsilon && \text{by (3.37)}\\
&\le \boldsymbol{x}_1^{\star\top}\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 + 3\epsilon && \text{by (3.38)}\\
&\le \tilde{\boldsymbol{x}}_1^\top\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)\tilde{\boldsymbol{x}}_2 - \frac{1}{2}\left(\tilde{\boldsymbol{x}}_1^\top\boldsymbol{a} - \lambda\right)^2 + 5\epsilon. && \text{by (3.34)}
\end{aligned}$$

Thus,

$$\frac{1}{2}\left(\tilde{\boldsymbol{x}}_1^\top \boldsymbol{a} - \lambda\right)^2 \leq 5\epsilon.$$

Multiplying both sides by 2 and taking square roots gives

$$\left|\tilde{\boldsymbol{x}}_1^\top \boldsymbol{a} - \lambda\right| \leq \sqrt{10\epsilon}. \tag{3.39}$$

Next, we show that $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ is an approximate Nash equilibrium of the (parameterized) zero-sum game $\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda\boldsymbol{b}^\top\right)$. From VI (2), we have

$$\left\langle \left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right)^\top \tilde{\boldsymbol{x}}_1, \boldsymbol{x}_2' - \tilde{\boldsymbol{x}}_2 \right\rangle \geq -\epsilon \qquad \forall\, \boldsymbol{x}_2' \in \Delta_{m_2},$$

which is equivalent to

$$\tilde{\boldsymbol{x}}_1^\top \left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right) \tilde{\boldsymbol{x}}_2 \leq \tilde{\boldsymbol{x}}_1^\top \left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right) \boldsymbol{x}_2' + \epsilon \qquad \forall\, \boldsymbol{x}_2' \in \Delta_{m_2}.$$

This is exactly the ϵ-best-response condition for the column player. Similarly, VI (1) implies that

$$\left\langle \left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right) \tilde{\boldsymbol{x}}_2 - \left(\tilde{\boldsymbol{x}}_1^\top \boldsymbol{a} - \lambda\right) \boldsymbol{a}, \boldsymbol{x}_1' - \tilde{\boldsymbol{x}}_1 \right\rangle \leq \epsilon \qquad \forall\, \boldsymbol{x}_1' \in \Delta_{m_1},$$

and therefore

$$\tilde{\boldsymbol{x}}_1^\top \left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right) \tilde{\boldsymbol{x}}_2 \geq \boldsymbol{x}_1'^\top \left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\right) \tilde{\boldsymbol{x}}_2 - \left(\tilde{\boldsymbol{x}}_1^\top \boldsymbol{a} - \lambda\right) \langle \boldsymbol{a}, \boldsymbol{x}_1' - \tilde{\boldsymbol{x}}_1 \rangle - \epsilon \qquad \forall\, \boldsymbol{x}_1' \in \Delta_{m_1}. \tag{3.40}$$

Using (3.39), Hölder's inequality, and the fact that $\boldsymbol{x}_1', \tilde{\boldsymbol{x}}_1 \in \Delta_{m_1}$, we obtain

$$\left|(\tilde{\boldsymbol{x}}_1^\top \boldsymbol{a} - \lambda)\,\langle \boldsymbol{a}, \boldsymbol{x}_1' - \tilde{\boldsymbol{x}}_1\rangle\right| \leq \left|\tilde{\boldsymbol{x}}_1^\top \boldsymbol{a} - \lambda\right|\, \|\boldsymbol{a}\|_\infty\, \|\boldsymbol{x}_1' - \tilde{\boldsymbol{x}}_1\|_1 \leq 2\sqrt{10\epsilon},$$

assuming $\|\boldsymbol{a}\|_\infty \leq 1$. Therefore,

$$\tilde{\boldsymbol{x}}_1^\top \left(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top\right) \tilde{\boldsymbol{x}}_2 \geq \boldsymbol{x}_1'^\top \left(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top\right) \tilde{\boldsymbol{x}}_2 - 2\sqrt{10\epsilon} - \epsilon \qquad \forall\, \boldsymbol{x}_1' \in \Delta_{m_1}. \tag{3.41}$$

Thus, $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ is an $(2\sqrt{10\epsilon} + \epsilon)$-approximate Nash equilibrium of the zero-sum game

$$(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda \boldsymbol{b}^\top).$$

In particular, for $\epsilon \leq 1$, this is at most an $(8\sqrt{\epsilon})$-approximate Nash equilibrium. □

So far, for a fixed value of λ, we have established that any stationary point $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ of f_λ is an approximate Nash equilibrium of the zero-sum game $\left(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top,\, -\mathbf{A} + \mathbb{1}\lambda \boldsymbol{b}^\top\right)$ that also approximately satisfies the consistency condition $|\tilde{\boldsymbol{x}}|_1^\top\, \boldsymbol{a} - \lambda \leq \epsilon$, provided that, for the chosen value of λ, there exists an exact or approximate Nash equilibrium $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ of the same zero-sum game such that $\left|\boldsymbol{x}_1^{*\top} \boldsymbol{a} - \lambda\right| \leq \epsilon$. In other words, finding a stationary point of f_λ is equivalent to finding a Nash equilibrium of the parameterized zero-sum game that simultaneously satisfies the additional condition that $\tilde{\boldsymbol{x}}_1^\top \boldsymbol{a}$ be close to λ. This is particularly useful because, by Lemma 3.4.1, such points also constitute an approximate Nash equilibria of the original rank-1 game.

There is, however, an important caveat: Lemma 3.4.2 applies only when λ is chosen correctly. Specifically, the lemma is valid only when the parameter λ is sufficiently close to a correct value $\lambda^\star$, namely one associated with a Nash equilibrium $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ of the rank-1 game such that $\boldsymbol{x}_1^{*\top} \boldsymbol{a} = \lambda^\star$. This naturally leads to the question of how to identify the appropriate value of λ. The following simple but important lemma provides the answer.

Lemma 3.4.3 (ϵ-Nash equilibria maintain sign)**.** *Let* $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ *and* $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ *be* ϵ*-approximate Nash equilibria of the parameterized zero-sum game*

$$\left(\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda \boldsymbol{b}^\top\right),$$

such that $\tilde{\boldsymbol{x}}_1^\top \boldsymbol{a} > \lambda$ *and* $(\boldsymbol{x}_1^*)^\top \boldsymbol{a} < \lambda$*. Then, there exists an* ϵ*-approximate Nash equilibrium* $(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2)$ *such that* $\overline{\boldsymbol{x}}_1^\top \boldsymbol{a} = \lambda$*.*

Proof. The set of ϵ-approximate Nash equilibria of a zero-sum game is convex. Hence, for every $\mu \in [0, 1]$, the convex combination

$$(\overline{\boldsymbol{x}}_1(\mu), \overline{\boldsymbol{x}}_2(\mu)) = \mu(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2) + (1 - \mu)(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$$

is also an ϵ-approximate Nash equilibrium.

Moreover, the map $\mu \mapsto \overline{\boldsymbol{x}}_1(\mu)^\top \boldsymbol{a}$ is continuous and affine in μ. Since $\tilde{\boldsymbol{x}}_1^\top \boldsymbol{a} > \lambda$ and $(\boldsymbol{x}_1^*)^\top \boldsymbol{a} < \lambda$, there exists $\mu^* \in [0, 1]$ such that $\overline{\boldsymbol{x}}_1(\mu^*)^\top \boldsymbol{a} = \lambda$. Setting $(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2) = (\overline{\boldsymbol{x}}_1(\mu^*), \overline{\boldsymbol{x}}_2(\mu^*))$ proves the claim. □

The lemma above answers our question as follows. If a stationary point $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ of f_λ fails to be an ϵ-approximate Nash equilibrium of the zero-sum game satisfying $\left|\boldsymbol{x}_1^{*\top} \boldsymbol{a} - \lambda\right| \leq \epsilon$, then no Nash equilibrium satisfying this condition exists for that particular value of λ. Otherwise, any stationary point of f_λ would suffice according to Lemma 3.4.2.

More importantly, this implies that $(\boldsymbol{x}_1^\top \boldsymbol{a} - \lambda)$ must have the same sign for every ϵ-approximate Nash equilibrium of the zero-sum game associated with that value of λ. Indeed, if there were two Nash equilibria for which $(\boldsymbol{x}_1^\top \boldsymbol{a} - \lambda)$ had opposite signs, then Lemma 3.4.3 would guarantee the existence of a third solution for which this quantity is equal to zero.

In a similar spirit to Lemma 3.3.3, Lemma 3.4.3 serves as an invariant. RANK1SOLVER performs binary search to parameter λ. At each iteration, it employs OMWU to obtain an ϵ-approximate stationary point of f_λ. This solution either constitutes a Nash equilibrium of the rank-1 game $(\mathbf{A}, \mathbf{B})$ or indicates that no Nash equilibrium exists for that *particular* λ and the specified precision ϵ. In the latter case, the sign of the difference $(\boldsymbol{x}_1^\top \boldsymbol{a} - \lambda)$ determines the next value of λ. For instance, if $\boldsymbol{x}_1^\top \boldsymbol{a} - \lambda < -\epsilon$, then λ is updated as $\lambda = \lambda - \frac{1}{2^k}$; otherwise, if $\boldsymbol{x}_1^\top \boldsymbol{a} - \lambda > +\epsilon$, then λ is set as $\lambda + \frac{1}{2^k}$.

In the final part, we introduce an efficient method for obtaining an ϵ-approximate stationary point of f_λ for a specific λ. Leveraging the structure of f_λ, we utilize an Optimistic Multiplicative Weights Update (OMWU), as detailed in (Daskalakis and Panageas, 2018b). This method offers two key advantages: it does not require a projection step, which can be computationally expensive, and it achieves faster convergence rates by incorporating optimism. A more comprehensive presentation of optimistic methods is deferred to Definition 2.10.1. Lemma 3.4.4 provides us with the specific number of iterations required.

Lemma 3.4.4. *Let $(\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)})$ be initialized as uniform vectors $(\mathbb{1}/n, \mathbb{1}/m)$. Suppose that both players employ OMWU for $T = \Omega\big(\frac{\ln m_1 + \ln m_2}{\epsilon}\big)$ with learning rates $\eta = \frac{1}{16\sqrt{2}\max\left\{\|\mathbf{A}\|_2 + \sqrt{m_1}\|\boldsymbol{b}\|_2, \|\boldsymbol{a}\|_2^2\right\}}$ and fixed λ. Then OMWU returns a pair of strategies $\big(\overline{\boldsymbol{x}}_1^{(T)}, \overline{\boldsymbol{x}}_2^{(T)}\big)$ that is an ϵ-approximate stationary point of f_λ.*

Proof. We consider the function $f_\lambda(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1^\top(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top)\boldsymbol{x}_2 - \frac{1}{2}(\boldsymbol{x}_1^\top \boldsymbol{a} - \lambda)^2$, where $\boldsymbol{x}_1 \in \mathcal{X}_1 \coloneqq \Delta_{m_1}$ is the maximization variable and $\boldsymbol{x}_2 \in \mathcal{X}_2 \coloneqq \Delta_{m_2}$ is the minimization variable.

Lemma 3.4.5 (Lemma 16 (Shalev-Shwartz, 2007))**.** *$\psi(\boldsymbol{x}) = \sum_{i=1}^{n} \boldsymbol{x}_i \ln \boldsymbol{x}_i$ is 1-strongly convex with respect to the L_1 norm over the simplex $\Delta_n \coloneqq \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x}_i \geq 0, \|\boldsymbol{x}\|_1 = 1\}$.*

We use the negative entropy regularizer on both domains: $\psi_{\mathcal{X}}(\boldsymbol{x}) = \sum_{i=1}^{m} \boldsymbol{x}_i \ln \boldsymbol{x}_i$. By Lemma 3.4.5, these regularizers are 1-strongly convex with respect to the ℓ_1 norm, and therefore satisfy the corresponding assumption of Theorem 2.10.2.

Next, we verify the smoothness assumptions, and in particular bound the Lipschitz constants appearing in Assumption 2.10.1. Since f_λ is twice differentiable, it suffices to bound the spectral norm of its Hessian. We have

$$\nabla^2 f_\lambda(\boldsymbol{x}_1, \boldsymbol{x}_2) = \begin{bmatrix} -\boldsymbol{a}\boldsymbol{a}^\top & \mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top \\ (\mathbf{A} - \mathbb{1}\lambda \boldsymbol{b}^\top)^\top & \mathbf{0} \end{bmatrix}.$$

We first recall a standard result from linear algebra that enables a simple computation of the spectral norm of a block matrix.

Proposition 3.4.1 (Claim C.2 in (Leonardos et al., 2021))**.** *Consider a symmetric block matrix* $\mathbf{C}$ *consisting of* $m \times m$ *blocks* $\mathbf{C}_{ij}$ *such that* $\|\mathbf{C}_{ij}\|_2 \leq L$ *for all* i, j*. Then,*

$$\|\mathbf{C}\|_2 \leq mL.$$

In other words, if all block matrices have spectral norm at most L*, then the spectral norm of* $\mathbf{C}$ *is at most* mL*.*

By Proposition 3.4.1 above, it suffices to bound the spectral norm of each block separately. For the diagonal block, since it is of rank one, we have

$$\| - \boldsymbol{a}\boldsymbol{a}^\top\|_2 = \|\boldsymbol{a}\boldsymbol{a}^\top\|_2 = \|\boldsymbol{a}\|_2^2.$$

For the off-diagonal block, using $\lambda \in [0,1]$, the triangle inequality, and the rank-1 identity $\|\mathbb{1}\boldsymbol{b}^\top\|_2 = \|\mathbb{1}\|_2\|\boldsymbol{b}\|_2$, we obtain

$$\|\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top\|_2 \le \|\mathbf{A}\|_2 + \lambda\|\mathbb{1}\boldsymbol{b}^\top\|_2 \tag{3.42}$$

$$\le \|\mathbf{A}\|_2 + \|\mathbb{1}\|_2\|\boldsymbol{b}\|_2 \tag{3.43}$$

$$= \|\mathbf{A}\|_2 + \sqrt{m_1}\,\|\boldsymbol{b}\|_2. \tag{3.44}$$

The same bound applies to the transpose block, while the bottom-right block is zero. Hence, by Proposition 3.4.1, the smoothness parameters are uniformly bounded as

$$L_{\mathcal{X}_1\mathcal{X}_1}, L_{\mathcal{X}_1\mathcal{X}_2}, L_{\mathcal{X}_2\mathcal{X}_1}, L_{\mathcal{X}_2\mathcal{X}_2} \le L, \quad \text{where} \quad L \coloneqq \max\left\{\|\mathbf{A}\|_2 + \sqrt{m_1}\|\boldsymbol{b}\|_2,\ \|\boldsymbol{a}\|_2^2\right\}.$$

Therefore, it suffices to choose the learning rate $\eta = \frac{1}{16\sqrt{2}L}$.

We now bound the initial Bregman divergences. Since $\boldsymbol{x}_1^{(1)} = \mathbb{1}/m_1$ and $\boldsymbol{x}_2^{(1)} = \mathbb{1}/m_2$, for any $\boldsymbol{x}_1 \in \Delta_{m_1}$, $\boldsymbol{x}_2 \in \Delta_{m_2}$,

$$B_{\psi_{\mathcal{X}_1}}(\boldsymbol{x}_1; \boldsymbol{x}_1^{(1)}) \le \ln m_1, \qquad B_{\psi_{\mathcal{X}_2}}(\boldsymbol{x}_2; \boldsymbol{x}_2^{(1)}) \le \ln m_2.$$

Finally, we bound the dual norms of the initial gradients. Since the primal norm is ℓ_1, the dual norm is ℓ_∞. For the first player,

$$\left\|\boldsymbol{g}_{\mathcal{X}_1}^{(1)}\right\|_\infty = \left\|(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top)\boldsymbol{x}_2^{(1)} - (\boldsymbol{x}_1^{(1)\top}\boldsymbol{a} - \lambda)\boldsymbol{a}\right\|_\infty \tag{3.45}$$

$$\le \left\|(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top)\boldsymbol{x}_2^{(1)}\right\|_\infty + \left|\boldsymbol{x}_1^{(1)\top}\boldsymbol{a} - \lambda\right| \|\boldsymbol{a}\|_\infty \tag{3.46}$$

$$\le (\mathbf{A}_{\max} + 1) + 3\,\|\boldsymbol{a}\|_\infty. \tag{3.47}$$

Similarly,

$$\left\|\boldsymbol{g}_{\mathcal{X}_2}^{(1)}\right\|_\infty \leq \left\|(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top)^\top \boldsymbol{x}_1^{(1)}\right\|_\infty \leq \mathbf{A}_{\max} + 1.$$

Thus, the initial gradients are uniformly bounded.

Consequently, all assumptions of Theorem 2.10.2 are satisfied with $\alpha = 1$. To summarize, we have shown the following:

1. The regularizers $\psi_{\mathcal{X}_1}$ and $\psi_{\mathcal{X}_2}$ are 1-strongly convex with respect to $\|\cdot\|_1$.

2. The smoothness assumptions in Assumption 2.10.1 are satisfied for the function f_λ.

3. The learning rate $\eta = \frac{1}{16\sqrt{2}L}$ satisfies

$$\frac{1}{\eta} \geq 2\sqrt{2}\big(L_{\mathcal{X}_1\mathcal{X}_1} + L_{\mathcal{X}_1\mathcal{X}_2}\big) \quad \text{and} \quad \frac{1}{\eta} \geq 2\sqrt{2}\big(L_{\mathcal{X}_2\mathcal{X}_2} + L_{\mathcal{X}_2\mathcal{X}_1}\big).$$

4. The initial gradient norms $\left\|\boldsymbol{g}_{\mathcal{X}_1}^{(1)}\right\|_\infty$ and $\left\|\boldsymbol{g}_{\mathcal{X}_2}^{(1)}\right\|_\infty$ are bounded by $(\mathbf{A}_{\max} + 1) + 3\,\|\boldsymbol{a}\|_\infty$.

5. The initial Bregman divergences satisfy

$$B_{\psi_{\mathcal{X}_1}}(\boldsymbol{x}_1; \boldsymbol{x}_1^{(1)}) \leq \ln m_1, \quad \text{and} \quad B_{\psi_{\mathcal{X}_2}}(\boldsymbol{x}_2; \boldsymbol{x}_2^{(1)}) \leq \ln m_2.$$

Applying Theorem 2.10.2, we obtain a strategy profile $(\overline{\boldsymbol{x}}_1^{(T)}, \overline{\boldsymbol{x}}_2^{(T)})$ such that

$$\texttt{Gap}(\overline{\boldsymbol{x}}_1^{(T)}, \overline{\boldsymbol{x}}_2^{(T)}) \leq \epsilon, \quad \text{for} \quad T \geq \Theta\left(\frac{\ln m_1 + \ln m_2}{\epsilon}\right).$$

It remains to show that a small gap implies approximate stationarity of f_λ. Since $f_\lambda(\cdot, \boldsymbol{x}_2)$ is concave for every fixed $\boldsymbol{x}_2$, and $f_\lambda(\boldsymbol{x}_1, \cdot)$ is linear, hence convex, for every fixed $\boldsymbol{x}_1$, the first-

Algorithm 2. Rank-1 Game Solver

Input: $\boldsymbol{x}_1^{(1)} = \mathbb{1}/m_1$, $\boldsymbol{x}_2^{(1)} = \mathbb{1}/m_2$, $\lambda^{(1)} = \frac{1}{2}$
Output: $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ NE of rank-1 $(\mathbf{A}, \mathbf{B})$

for $k = 1, \ldots, \log(\frac{1}{\epsilon})$ **do**
  $\big(\boldsymbol{x}_1^{(k+1)}, \boldsymbol{x}_2^{(k+1)}\big) = \texttt{OMWU}\big(\boldsymbol{x}_1^{(k)}, \boldsymbol{x}_2^{(k)}, \lambda^{(k)}\big)$
  if $\big(\boldsymbol{x}_1^{(k+1)}, \boldsymbol{x}_2^{(k+1)}\big)$ *is a NE of* $(\mathbf{A}, \mathbf{B})$ **then**
    return $\big(\boldsymbol{x}_1^{(k+1)}, \boldsymbol{x}_2^{(k+1)}\big)$
  end
  else
    $\lambda^{(k+1)} = \lambda^{(k)} - \frac{1}{2^{k+1}} \mathbf{sgn}\big({\boldsymbol{x}_1^{(k+1)}}^\top \boldsymbol{a} - \lambda^{(k)}\big)$ *** update λ according to sign ***
  end
end

order stationarity condition is equivalent to the absence of profitable unilateral deviations in the maximization and minimization variables.

Indeed, suppose that $\texttt{Gap}(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2) \le \epsilon$. By definition of the gap, for every $\boldsymbol{x}_1 \in \Delta_{m_1}$ and every $\boldsymbol{x}_2 \in \Delta_{m_2}$,

$$f_\lambda(\boldsymbol{x}_1, \overline{\boldsymbol{x}}_2) - f_\lambda(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2) \le \epsilon \quad \text{and} \quad f_\lambda(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2) - f_\lambda(\overline{\boldsymbol{x}}_1, \boldsymbol{x}_2) \le \epsilon.$$

The first inequality means that $\overline{\boldsymbol{x}}_1$ is an ϵ-approximate maximizer of the concave function $f_\lambda(\cdot, \overline{\boldsymbol{x}}_2)$, while the second means that $\overline{\boldsymbol{x}}_2$ is an ϵ-approximate minimizer of the convex function $f_\lambda(\overline{\boldsymbol{x}}_1, \cdot)$. Therefore, $(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2)$ satisfies the ϵ-approximate first-order optimality conditions for the saddle function f_λ, and hence is an ϵ-approximate stationary point. □

We are now prepared to present the main technical theorem that integrates all the components established thus far into a concise statement.

Theorem 3.4.1. *Let* $(\mathbf{A}, \mathbf{B})$ *be a rank-1 game. Suppose that players employ* Rank1Solver *for* $T = \Omega\big(\frac{\log m_1 + \log m_2}{\epsilon^2} \log\big(\frac{1}{\epsilon}\big)\big)$ *iterations. Then* Rank1Solver *returns a pair of strategies* $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ *that constitutes an* 18ϵ*-approximate Nash equilibrium of the rank-1 game.*

Proof. Fix a value of λ, and consider two cases.

First, suppose that λ is correct; that is, there exists a Nash equilibrium $(\boldsymbol{x}_1^\lambda, \boldsymbol{x}_2^\lambda)$ of the zero-sum game $\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top, -\mathbf{A} + \mathbb{1}\lambda\boldsymbol{b}^\top\right)$ such that $\left|\boldsymbol{x}_1^{\lambda\top}\boldsymbol{a} - \lambda\right| \leq \epsilon$. If we run `OMWU` for $T \geq \Theta\left(\frac{\ln m_1 + \ln m_2}{\epsilon^2}\right)$, then, by Lemma 3.4.4, we obtain a pair $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ that is an ϵ^2-approximate stationary point of f_λ.

By Lemma 3.4.2, this implies that $\left|\boldsymbol{x}_1^{*\top}\boldsymbol{a} - \lambda\right| \leq \sqrt{10\,\epsilon^2} \leq 6\epsilon$, and that $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ is a 6ϵ-approximate Nash equilibrium of the zero-sum game $\left(\mathbf{A} - \mathbb{1}\lambda\boldsymbol{b}^\top,\ -\mathbf{A} + \mathbb{1}\lambda\boldsymbol{b}^\top\right)$. Applying Lemma 3.4.1, we conclude that this pair is an 18ϵ-approximate Nash equilibrium of the original rank-1 game $(\mathbf{A}, \mathbf{B})$

Conversely, if λ is incorrect, it must be updated. In this case, the sign of $(\boldsymbol{x}_1^\top\boldsymbol{a} - \lambda)$ determines the direction of the update, as specified in Rank1Solver. Since the update rule performs a binary search over λ, it requires at most $K = O\left(\log\frac{1}{\epsilon}\right)$ iterations.

Combining the two cases yields the claimed complexity bound. □

3.5 Experiments

In this section, we provide experimental evaluation supporting our theoretical finding. Here we restrict to Example 3.5.1, that we constructed by randomly choosing 5×5 games defined by the matrix $\mathbf{A}$ and the vectors $\boldsymbol{a}, \boldsymbol{b}$.

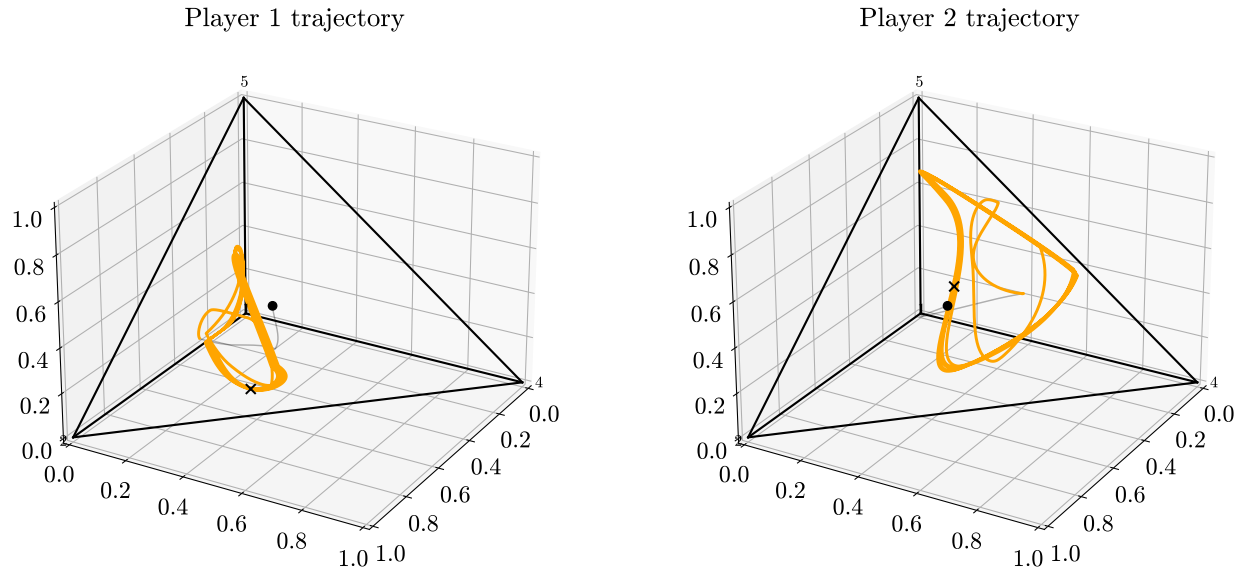


Figure 3.1: Trajectories of the players' strategies on a 4-dimensional simplex. Strategy 3 is omitted because it remains inactive throughout the iterations. The trajectories clearly enter a limit cycle from which they do not escape.

Example 3.5.1.

$$\mathbf{A} = \begin{bmatrix} -0.10 & 0.50 & 0.50 & -0.00 & -0.30 \\ 0.00 & -0.10 & -0.30 & 0.10 & -0.20 \\ -0.10 & -0.20 & -0.30 & 0.00 & -0.30 \\ -0.30 & 0.20 & -0.10 & -0.20 & 0.00 \\ -0.40 & -0.20 & -0.30 & 0.40 & -0.30 \end{bmatrix} \quad \boldsymbol{a} = \begin{bmatrix} 0.60 \\ 0.80 \\ 0.80 \\ 0.00 \\ 0.10 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 0.30 \\ 1.00 \\ 0.50 \\ 0.60 \\ 0.10 \end{bmatrix} \tag{3.48}$$

Then, the payoff matrix of the second player is derived as usual: $\mathbf{B} = -\mathbf{A} + \boldsymbol{a}\boldsymbol{b}^\top$. Our solution delineated in Section 3.4.2 is compared against a (vanilla) optimistic variant of mirror descent (`OMD`). More specifically, our tests were performed with (optimistic) gradient ascent (`OGA`) instead of multiplicative weights to avoid any numerical imprecision that might arise due to exponentiation.

In Figures 3.1 and 3.2a, we (experimentally) verify that `OGA` enters a limit cycle. This is also evident in Figure 3.2b as it oscillates, without converging to zero. In contrast, our

Algorithm 3. Optimistic Mirror Descent (OMD)

Input: Payoff matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m_1 \times m_2}$, initial strategies $\boldsymbol{x}_1^{(1)} \in \Delta_{m_1}$, $\boldsymbol{x}_2^{(1)} \in \Delta_{m_2}$, step size $\eta > 0$, mirror maps $h_1 : \Delta_{m_1} \to \mathbb{R}$, $h_2 : \Delta_{m_2} \to \mathbb{R}$

Output: Strategy sequence $\{(\boldsymbol{x}_1^{(t)}, \boldsymbol{x}_2^{(t)})\}_{t \geq 1}$

Initialize $\boldsymbol{g}_1^{(0)} \leftarrow \mathbf{A}\boldsymbol{x}_2^{(1)}$, $\boldsymbol{g}_2^{(0)} \leftarrow \mathbf{B}^\top \boldsymbol{x}_1^{(1)}$.

for $t = 1, 2, \ldots, T-1$ **do**

Compute current payoff vectors $\boldsymbol{g}_1^{(t)} \leftarrow \mathbf{A}\boldsymbol{x}_2^{(t)}$, $\boldsymbol{g}_2^{(t)} \leftarrow \mathbf{B}^\top \boldsymbol{x}_1^{(t)}$.

Update player 1 by

$$\boldsymbol{x}_1^{(t+1)} \leftarrow \arg\max_{\boldsymbol{x}_1 \in \Delta_{m_1}} \left\{ \eta \left\langle 2\boldsymbol{g}_1^{(t)} - \boldsymbol{g}_1^{(t-1)}, \boldsymbol{x}_1 \right\rangle - D_{h_1}\left(\boldsymbol{x}_1, \boldsymbol{x}_1^{(t)}\right) \right\}.$$

Update player 2 by

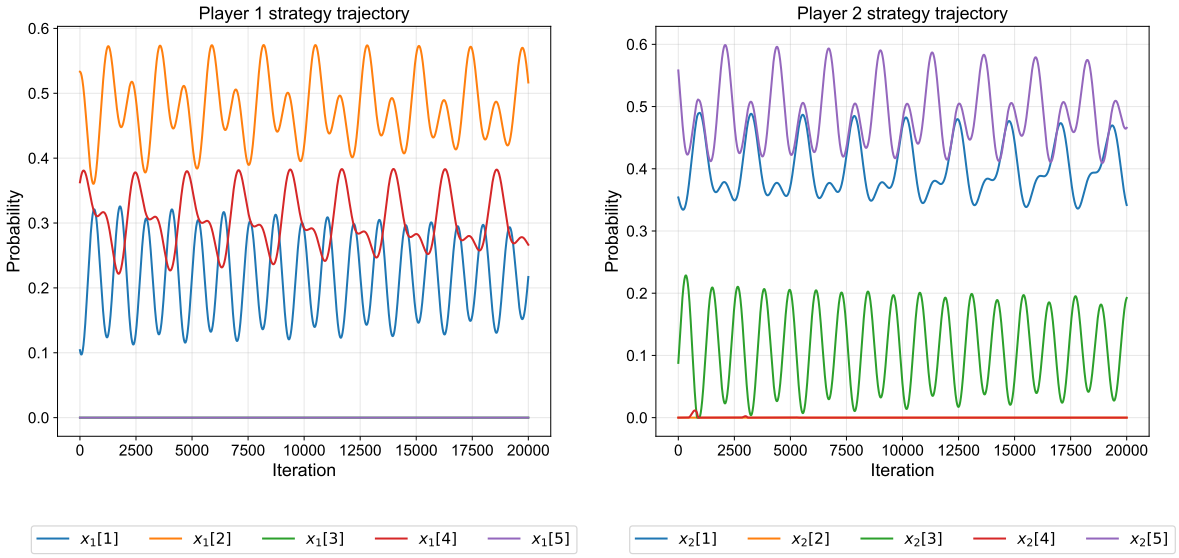
$$\boldsymbol{x}_2^{(t+1)} \leftarrow \arg\max_{\boldsymbol{x}_2 \in \Delta_{m_2}} \left\{ \eta \left\langle 2\boldsymbol{g}_2^{(t)} - \boldsymbol{g}_2^{(t-1)}, \boldsymbol{x}_2 \right\rangle - D_{h_2}\left(\boldsymbol{x}_2, \boldsymbol{x}_2^{(t)}\right) \right\}.$$

end

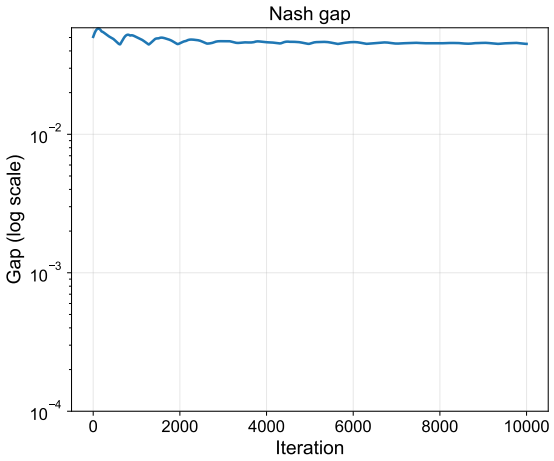
solution in Figure 3.3 reaches a Nash equilibrium. While this game may not have a natural interpretation initially, given their random construction, it is worth noting that constructing random games where OGA provably enters a limit cycle is rather challenging. This difficulty is further explained by the recent work Anagnostides et al. (2022a).

3.6 Concluding remarks

The results of this chapter show that efficient equilibrium learning is possible beyond zero-sum games, provided that the underlying structure is exploited carefully. Rank-1 games do not satisfy the standard monotonicity-based conditions that support many classical first-order analyses, and naive use of zero-sum reductions is insufficient. Nevertheless, by combining the geometry of the rank-1 equilibrium set with a parameterized optimistic procedure, one can still obtain polynomial-time convergence to approximate Nash equilibrium. This positive result serves as a useful contrast to the next chapter, where we turn to potential



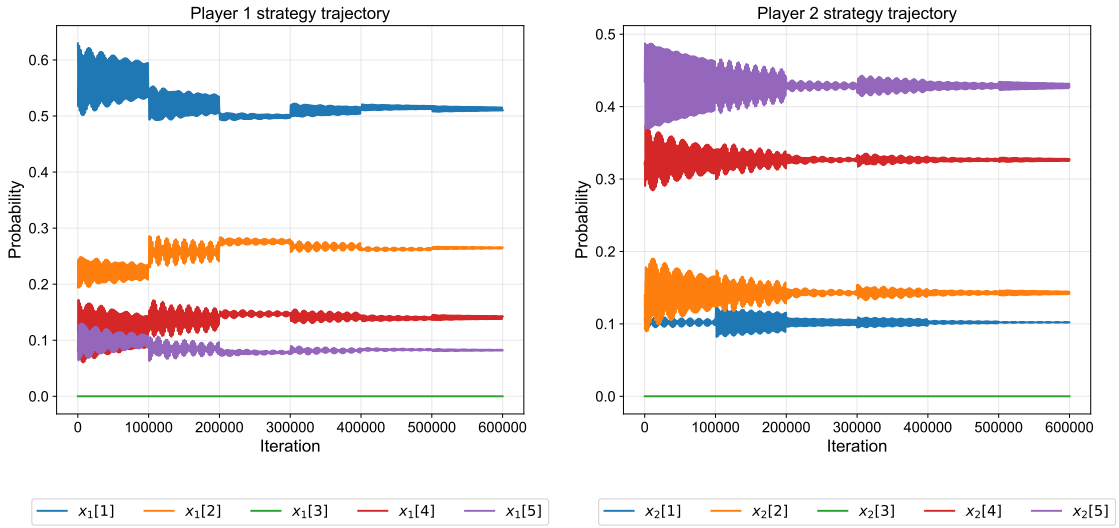
(a) Trajectories of the players' strategies over time when they employ vanilla OGA on the parameterized zero-sum game. As noted above, the method enters a limit cycle due to the zero-sum nature of the game, as reflected in the trajectories. To make the cyclic behavior evident, we plot the last iterate.



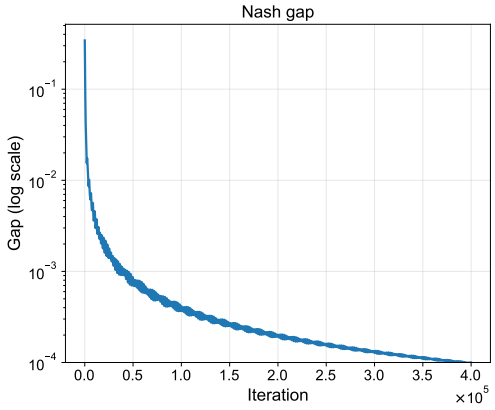
(b) The Nash gap along the players' trajectory, computed on the average trajectory rather than on the last iterate. The gap is evaluated with respect to the original rank-1 game $(\mathbf{A}, \mathbf{B})$.

Figure 3.2: Behavior of OGDA on the rank-1 game defined in Example 3.5.1. The left panel shows that, after approximately $T = 1000$ iterations, the trajectories of both players enter a limit cycle from which they do not escape. Similarly, the right panel shows that the Nash gap oscillates without converging to zero, indicating that the trajectory does not converge to a Nash equilibrium of the rank-1 game.

games and show that even extremely natural dynamics with classical asymptotic guarantees can be exponentially slow.



(a) Trajectories of the players' strategies over time under Algorithm 2. The figure shows the entire execution of the algorithm, so the irregularities in the strategy trajectories are due to changes in the value of λ.



(b) The Nash gap along the players' trajectory, computed on the average trajectory for the last computed value of λ.

Figure 3.3: Behavior of Algorithm 1 on the rank-1 game defined in Example 3.5.1. The left panel shows that the trajectories of both players eventually converge to a single strategy over the course of the execution of Algorithm 1. The right panel shows the Nash gap corresponding to the final value of λ, indicating convergence up to the approximation error $\epsilon = 10^{-6}$.

Chapter 4

Exponential Lower Bounds for Fictitious Play in Potential Games

4.1 Preface

This chapter is based on the joint work (Panageas et al., 2023), with the exposition adapted from the corresponding paper for inclusion in this dissertation. In particular, the notation and several theorem statements have been significantly reformulated, and the proofs have been reworked, in order to make them compatible with the framework and notation adopted later in Chapter 5.

Within the broader narrative of the thesis, this chapter provides the first major lower-bound result. In contrast to the previous chapter, which demonstrated that structural properties can sometimes be leveraged algorithmically, the present chapter shows that structural favorability alone does not suffice to ensure efficient convergence of canonical dynamics.

The chapter focuses on fictitious play in two-player identical-interest games, a special case of the broad class of potential games. Fictitious play (FP) is one of the most classical

and natural dynamics for repeated play, with important applications in both game theory and multi-agent reinforcement learning. Although it is known to converge asymptotically in potential games, the main theorem of this chapter shows that its convergence rate can nevertheless be exponentially slow. The proof is based on the construction of a game in which fictitious play is forced to follow a long spiral trajectory before reaching the unique equilibrium region.

4.2 Introduction

Robinson (1951a) proposed an *uncoupled* dynamic, known as fictitious play, to model the behavior of selfish agents engaged in a repeated game. Under fictitious play, at round $t = 1$, each agent selects an arbitrary action. At every subsequent round $t \geq 2$, each player chooses a pure best response to the opponents' empirical (average) strategy, namely, the empirical distribution of the actions they have played in the previous rounds.

Due to its simplicity and the naturality of its behavioral assumptions, fictitious play is one of the most influential and extensively studied dynamics in game theory (Cesa-Bianchi and Lugosi, 2006). While fictitious play does not converge to a Nash equilibrium (NE) in arbitrary normal-form games, there exist several important classes of games for which the *empirical strategies* are guaranteed to converge to an NE. In particular, Robinson (1951b) proved that, in the class of two-player zero-sum games, the empirical strategy profiles converge to a min-max equilibrium of the game; her proof is based on a clever inductive argument on the number of strategies. Subsequently, Monderer and Shapley (1996b) showed that, in n-player potential games, the empirical strategies likewise converge to a Nash equilibrium. We summarize these classical convergence guarantees in the following theorem.

Theorem 4.2.1 (Informal convergence guarantee for fictitious play (Robinson, 1951b; Monderer and Shapley, 1996b))**.** *In two-player zero-sum games and in n-player potential games,*

the empirical strategy profiles generated by fictitious play converge to a Nash equilibrium for every initialization and every tie-breaking rule.[1]

The convergence results stated above are purely asymptotic, in the sense that they do not provide quantitative bounds on the number of rounds required for fictitious play to reach an approximate Nash equilibrium. In the special case of two-player zero-sum games, Karlin (1959) conjectured that fictitious play reaches an ϵ-approximate Nash equilibrium within $O(1/\epsilon^2)$ rounds. This strong version of Karlin's conjecture was subsequently disproved by Daskalakis and Pan (2014), who constructed an adversarial tie-breaking rule for which fictitious play requires an exponential number of rounds, with respect to the number of strategies, to reach an ϵ-approximate Nash equilibrium. More recently, Wang (2025) partially resolved the weak form of the conjecture, under the assumption that ties may occur only at the first step. In contrast, no analogous lower bound is known for potential games. The purpose of this chapter is to address the following question.

Does fictitious play in potential games admit convergence to an approximate NE with rates that depend polynomially on the number of actions and the desired accuracy?

This chapter provides a negative answer on the above question. Specifically, we present a *two-player potential* game for which fictitious play requires super-exponential time with respect to the number of actions to reach an approximate NE.

4.2.1 Related work

As discussed above, much of the literature on convergence rates for fictitious play focuses on the case of zero-sum games, since quantitative convergence in this setting has long remained a central open question following the classical work of Brown (1951). In particular, (Daskalakis and Pan, 2014) made a first step toward resolving this question by proving a lower bound on the convergence rate of fictitious play in two-player zero-sum games, even for games of

[1] A tie-breaking rule is used when a player has two or more distinct best responses available.

the form $(\mathbb{1}_{m\times m}, -\mathbb{1}_{m\times m})$, under the strong assumption of an adversarial tie-breaking rule. This result is among the closest in spirit to the present work. Subsequently, Abernethy et al. (2021) showed that this negative result disappears when the tie-breaking rule is fixed in advance, for example lexicographically: for diagonal games and any such predetermined tie-breaking rule, fictitious play achieves the convergence rate conjectured by Karlin (1959), namely $O(1/\epsilon^2)$. More recently, Wang (2025) made further progress toward resolving Karlin's conjecture, again under a restriction on the handling of ties at the first step. After the initial round, no further ties arise, so their result is only partially agnostic to tie-breaking. On the other hand, another important class of games known to enjoy the fictitious-play property is that of potential games, and in particular identical-payoff games; see (Monderer and Shapley, 1996b). Other related works include the convergence of fictitious play in near-potential games (Candogan et al., 2013), as well as sufficient conditions under which fictitious play converges to a Nash equilibrium via decomposition techniques (Chen et al., 2022). However, these results do not provide quantitative convergence rates for fictitious play. Fast convergence rates for *continuous-time* fictitious play in *regular* potential games were established by Swenson and Kar (2017); here regularity is understood in the sense of Harsanyi. Additional works on continuous-time fictitious play include (Ostrovski and van Strien, 2013) and the references therein. Fictitious play has also found a range of applications in multi-agent reinforcement learning; see, for example, (Muller et al., 2020; Sayin et al., 2022; Baudin and Laraki, 2022), as well as (Yang and Wang, 2020) for a survey and further references. It has also played an important role in extensive-form games; see, for instance, (Heinrich et al., 2015). [2]

4.2.2 Technical overview

We now provide an overview of the main ingredients of the proof, together with a roadmap for the remainder of this chapter. In Section 4.4, we describe the steps leading to the proof

[2]Given the vast literature on fictitious play, it is not possible to include all works that either use this method or are inspired by it.

of Theorem 4.4.1 by introducing a recursively defined payoff matrix with carefully designed structural properties. The key feature of this construction is that, when one follows the sequence of successive increments starting from the lower-left entry, these increments trace a spiraling trajectory that eventually approaches the entry of maximum value. We then show that fictitious play is forced to follow the same trajectory: in order to reach the unique pure Nash equilibrium, it must traverse all nonzero entries of the matrix, as illustrated in Figure 4.2. A crucial part of the argument is an inductive construction that effectively emulates the evolution of fictitious play and yields a super-exponential lower bound on the number of rounds required for convergence.

4.3 Preliminaries

4.3.1 Definitions

In this chapter, we focus on bimatrix identical payoff games, as they were introduced in Section 2.4. Thus, there are two players with finite action sets $\mathcal{A}_1$ and $\mathcal{A}_2$, where $|\mathcal{A}_1| = m_1$ and $|\mathcal{A}_2| = m_2$. Accordingly, let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m_1 \times m_2}$ denote the payoff matrices of the two players.

We conclude by recalling the definition of the class of games that will be central to our lower-bound construction, namely *identical-payoff games*.

Definition 4.3.1 (Identical-payoff game)**.** A two-player normal-form game $(\mathbf{A}, \mathbf{B})$ is called an *identical-payoff game* if and only if $\mathbf{A} = \mathbf{B}$.

4.3.2 Fictitious play

Fictitious play is a natural uncoupled game dynamic in which each agent chooses a *best response* to the empirical mixed strategy of the opponent. Since there may be several best-

Algorithm 4. Fictitious Play

Input: Payoff matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m_1 \times m_2}$, initial strategies $\boldsymbol{x}_1^{(1)} \in \mathcal{X}_1 \coloneqq \Delta_{m_1}$, $\boldsymbol{x}_2^{(1)} \in \mathcal{X}_2 \coloneqq \Delta_{m_2}$, horizon T

Output: Sequence $\{(\boldsymbol{x}_1^{(t)}, \boldsymbol{x}_2^{(t)})\}_{t=0}^{T}$

Initialize empirical averages $\overline{\boldsymbol{x}}_1^{(1)} \leftarrow \boldsymbol{x}_1^{(1)}$, $\overline{\boldsymbol{x}}_2^{(1)} \leftarrow \boldsymbol{x}_2^{(1)}$

for $t = 0, 1, \ldots, T-1$ **do**

  $\boldsymbol{x}_1^{(t+1)} \leftarrow \arg\max_{\boldsymbol{x} \in \Delta_{m_1}} \boldsymbol{x}_1^\top \mathbf{A} \overline{\boldsymbol{x}}_2^{(t)}$ ** $\boldsymbol{x}_1^{(t+1)} \in \{0,1\}^{m_1}$ **

  $\boldsymbol{x}_2^{(t+1)} \leftarrow \arg\max_{\boldsymbol{x}_2 \in \Delta_{m_2}} \left(\overline{\boldsymbol{x}}_1^{(t)}\right)^\top \mathbf{B} \boldsymbol{x}_2$ ** $\boldsymbol{x}_2^{(t+1)} \in \{0,1\}^{m_2}$ **

  $\overline{\boldsymbol{x}}_1^{(t+1)} \leftarrow \dfrac{t\overline{\boldsymbol{x}}_1^{(t)} + \boldsymbol{x}_1^{(t+1)}}{t+1}$

  $\overline{\boldsymbol{x}}_2^{(t+1)} \leftarrow \dfrac{t\overline{\boldsymbol{x}}_2^{(t)} + \boldsymbol{x}_2^{(t+1)}}{t+1}$

end

response actions at a given round, fictitious play can generate different sequences of play; see Definition 4.3.2.

Definition 4.3.2 (Fictitious play)**.** An sequence of pure strategy profiles $\left\{\left(i^{(t)}, j^{(t)}\right)\right\}_{t \geq 1}$ is called a fictitious play sequence if and only if at each round $t \geq 1$,

$$i^{(t+1)} \in \arg\max_{i \in [m_1]} \sum_{\tau=1}^{t} \mathbf{A}\left[i, j^{(\tau)}\right] \quad \text{and} \quad j^{(t+1)} \in \arg\max_{j \in [m_2]} \sum_{\tau=1}^{t} \mathbf{B}\left[i^{(\tau)}, j\right] \tag{4.1}$$

The empirical strategy profiles of the row and column players at time T are defined by

$$\overline{\boldsymbol{x}}_1^{(T)} = \frac{1}{T} \sum_{\tau=1}^{T} \boldsymbol{e}_{i^{(\tau)}} \quad \text{and} \quad \overline{\boldsymbol{x}}_2^{(T)} = \frac{1}{T} \sum_{\tau=1}^{T} \boldsymbol{e}_{j^{(\tau)}}, \tag{4.2}$$

where $\boldsymbol{e}_{i^{(\tau)}}$ and $\boldsymbol{e}_{j^{(\tau)}}$ denote the corresponding standard basis vectors.

Definition 4.3.3 (Cumulative utility vector)**.** For an infinite sequence of pure strategy profiles $\left\{\left(i^{(t)}, j^{(t)}\right)\right\}_{t \geq 1}$, the *cumulative utility* vectors of the row and column player at round

$t \geq 1$ are defined as,

$$\mathtt{R}^{(t)} = \sum_{\tau=1}^{t-1} \mathbf{A}\boldsymbol{e}_{j^{(\tau)}} = \sum_{\tau=1}^{t-1} \mathbf{A}\left[\cdot, j^{(\tau)}\right] \quad \text{and} \quad \mathtt{C}^{(t)} = \sum_{\tau=1}^{t-1} \boldsymbol{e}_{i^{(\tau)}}^{\top}\mathbf{B} = \sum_{\tau=1}^{t-1} \mathbf{B}\left[i^{(\tau)}, \cdot\right] \tag{4.3}$$

Remark 4.3.1. Fictitious play assumes that, at each round $t \in [T]$, each agent selects a strategy with *maximum cumulative utility*. In the presence of ties between multiple strategies, a tie-breaking rule must be specified. Since our objective is to establish a lower bound, we impose no further assumptions on the tie-breaking rule. In addition, fictitious play is also known as *Follow the Leader* (`FTL`). We remark that this alternative interpretation yields a direct generalization of fictitious play to n-player games.

In their seminal work, Monderer and Shapley (1996b) established that, in identical-payoff games, the empirical strategies generated by any fictitious play sequence converge asymptotically to a Nash equilibrium. We restate this result informally below for later reference.

Theorem 4.3.1 (Monderer and Shapley (1996b))**.** *Suppose both players employ* `FP` *in the identical-payoff game* $(\mathbf{A}, \mathbf{A})$*. Then there exists* $T^\star \geq 1$ *such that, for every* $t \geq T^\star$*, the empirical strategy profile* $(\overline{\boldsymbol{x}}_1^{(t)}, \overline{\boldsymbol{x}}_2^{(t)})$ *is a* $1/t$*-approximate Nash equilibrium.*

On the positive side, Theorem 4.3.1 establishes that every fictitious play sequence converges to a Nash equilibrium in the case of potential games.[3] On the negative side, Theorem 4.3.1 does not provide a quantitative convergence rate, since the round $T^\star$ depends on the particular fictitious play sequence, and its dependence on the number of strategies remains unclear.

[3]For ease of exposition, we state Theorem 4.3.1 only for identical-payoff games. However, the same result holds more generally for N-player potential games.

4.4 Main result

In this section, we outline the proof of Theorem 4.4.1, stated below. We begin by introducing a carefully constructed payoff matrix $\mathbf{A} \coloneqq \mathbf{B}_{m,0}$ of size $m \times m$ for $m \coloneqq m_1 = m_2$ even, obtained as an instance of the recursively defined matrix $\mathbf{B}_{m,r}$ with $r = 0$; this construction is presented in Section 4.4.1. Next, in Section 4.4.3, we study the behavior of fictitious play in the two-player identical-payoff game with payoff matrix $\mathbf{A}$, and we state several key intermediate results needed for the proof of the main theorem.

Theorem 4.4.1. *FP requires $m^{\Omega(m)}$ rounds to converge to a $\frac{1}{\mathrm{poly}(m)}$-Nash equilibrium in two-player $m \times m$ identical-interest games.*

Remark 4.4.1. We note that Theorem 4.4.1 applies both to last-iterate convergence, that is, to the profile $(\boldsymbol{x}_1^{(t)}, \boldsymbol{x}_2^{(t)})$, and to ergodic convergence, namely to the empirical strategy profile $(\overline{\boldsymbol{x}}_1^{(t)}, \overline{\boldsymbol{x}}_2^{(t)})$.

4.4.1 Construction and analysis of $\mathbf{B}_{m,r}$

We begin by introducing our recursive construction for the payoff matrix $\mathbf{B}_{m,r}$ in Definition 4.4.1, which we use to establish the formal statement of Theorem 4.4.1. We also examine the structural properties, which are used to establish the lower bounds for `FP` and `FTRL` in Chapter 4 and Chapter 5, respectively.

$$\mathbf{B}_{4,0} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 5 & 0 & 4 \\ 0 & 6 & 7 & 0 \\ 2 & 0 & 0 & 3 \end{bmatrix}$$

(a) An example for $m = 4$ and $r = 0$.

$$\mathbf{B}_{6,0} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 4 \\ 0 & 0 & 9 & 0 & 8 & 0 \\ 0 & 0 & 10 & 11 & 0 & 0 \\ 0 & 6 & 0 & 0 & 7 & 0 \\ 2 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}$$

(b) An example for $m = 6$ and $r = 0$.

Figure 4.1: Two examples of the recursively defined matrix $\mathbf{B}_{m,r}$: one with $m = 4$ and another with $m = 6$, both for $r = 0$.

Definition 4.4.1 (Recursive payoff matrix $\mathbf{B}_{m,r}$)**.** Let $m \geq 2$ be even and $r \in \mathbb{N}$. Define $\mathbf{B}_{m,r} \in \mathbb{R}^{m \times m}$ recursively by

$$\mathbf{B}_{m,r} \coloneqq \begin{bmatrix} (r+1) & 0 & \cdots & 0 & 0 \\ 0 & & & & (r+4) \\ \vdots & & \mathbf{B}_{m-2,r+4} & & \vdots \\ 0 & & & & 0 \\ (r+2) & 0 & \cdots & 0 & (r+3) \end{bmatrix}, \text{ with base case } \mathbf{B}_{2,r} \coloneqq \begin{bmatrix} r+1 & 0 \\ r+2 & r+3 \end{bmatrix}. \quad (4.4)$$

In Figures 4.1a and 4.1b, we provide a schematic representation of Definition 4.4.1 and two illustrative examples for $m \in \{4, 6\}$ and $r = 0$. For the sake of simplicity, we have intentionally omitted the remaining zeros in the outer rows, and columns of the matrix.

The construction of the payoff matrix $\mathbf{B}_{m,r}$ exhibits an interesting circular pattern, which begins at the upper left corner and extends along the outer layer of the matrix. More specifically, the first increment occurs in the same column as the starting point, i.e., at position $(m, 1)$ on the last row. The pattern then proceeds to the next greater element on the same row but a different column, i.e., at position (m, m), and the last increment before

entering the inner sub-matrix is located on the same column but on the 2-th row.

$$\text{Sequence of increments: } (1,1) \to (m,1) \to (m,m) \to (2,m) \to \underbrace{(2,2) \to \cdots}_{\mathbf{B}_{m-2,r+4}}$$

The increments are chosen so that, starting from the lower-left corner and following the successive increments, one encounters alternating changes between rows and columns until reaching the central submatrix. Once inside this submatrix, the same pattern continues recursively. As we shall see later, the structure of the payoff matrix determines the behavior of fictitious play. We now record some structural properties of $\mathbf{B}_{m,r}$.

Lemma 4.4.1 (Support entries of $\mathbf{B}_{m,r}$)**.** *Let $m \geq 2$ be even and $r \in \mathbb{N}$. Then the set of nonzero entries of $\mathbf{B}_{m,r}$ is exactly $\{r+1, r+2, \ldots, r+(2m-1)\}$, and each value in this set appears exactly once.*

Proof. We proceed by induction on m. For the base case $m = 2$, the matrix $\mathbf{B}_{2,r}$ has exactly three nonzero entries, namely $\{r+1,\ r+2,\ r+(2\cdot 2-1)\}$, each appearing once. Assume the claim holds for $m-2$, and consider $\mathbf{B}_{m,r}$ with $m \geq 4$. By construction, the only nonzero entries on the outer frame are

$$r+1 \text{ at } (1,1), \qquad r+2 \text{ at } (m,1), \qquad r+3 \text{ at } (m,m), \qquad r+4 \text{ at } (2,m),$$

and all other outer entries are zero. The inner block is $\mathbf{B}_{m-2,r+4}$, supported on rows and columns in $\{2,\ldots,m-1\}$. By the induction hypothesis, its nonzero entries are exactly

$$\{(r+4)+1, \ldots, (r+4)+(2(m-2)-1)\} = \{r+5, \ldots, r+(2m-1)\},$$

each appearing exactly once. Since $\{r+1, r+2, r+3, r+4\}$ is disjoint from $\{r+5, \ldots, r+(2m-1)\}$, all nonzero entries of $\mathbf{B}_{m,r}$ are distinct, and their union is precisely $\{r+1, \ldots, r+(2m-1)\}$. □

For each payoff value $k \in \{1, \ldots, 2m-1\}$, we introduce the locator functions $a_1(k), a_2(k) \in [m]$.

Definition 4.4.2 (Locator functions a_1, a_2)**.** Let $m \geq 2$ be even and let $r \in \mathbb{N}$. For each value $s \in \{r+1, \ldots, r+(2m-1)\}$ that appears as an entry of $\mathbf{B}_{m,r}$, define $(a_1(s), a_2(s)) \in [m] \times [m]$ as the unique pair of indices (guaranteed by Lemma 4.4.1) such that $\mathbf{B}_{m,r}\big[a_1(s), a_2(s)\big] = s$.

We denote the payoff matrix under consideration by $\mathbf{A}$, defined as $\mathbf{A} \coloneqq \mathbf{B}_{m,0}$ for m even. The following statements establish the key properties of $\mathbf{A}$. In particular, to avoid reproving the same result with minor modifications, we restate Proposition A.3.1, originally used in Chapter 5, in a form adapted to the notation of the present chapter.

Proposition 4.4.1 (Structural properties of $\mathbf{A}$)**.** *The following properties hold for* $\mathbf{A}$*:*

(i) $\max_{i,j} \mathbf{A}[i,j] = 2m-1$.

(ii) For even $k \in \{1, \ldots, 2m-2\}$*,* $a_1(k) = a_1(k+1)$.

(iii) For odd $k \in \{1, \ldots, 2m-2\}$*,* $a_2(k) = a_2(k+1)$.

Proposition 4.4.2. *Let* $\mathbf{A}[i,j]$ *be a nonzero entry of the matrix* $\mathbf{A}$*. Then there exists at most one nonzero entry in the same row or column that is strictly greater than* $\mathbf{A}[i,j]$*, and at most one nonzero entry in the same row or column that is strictly smaller than* $\mathbf{A}[i,j]$.

Proof. By Lemma 4.4.1, since $\mathbf{A}$ is an instance of $\mathbf{B}_{m,r}$, each value $k \in \{1, \ldots, 2m-1\}$ appears exactly once in $\mathbf{A}$. Moreover, Items (ii) and (iii) in Item (i) imply that, for every such k, the corresponding entry of value k in $\mathbf{A} \coloneqq \mathbf{B}_{n,0}$ shares its row or column with at

most two other nonzero entries. Consequently, at most one of them can be larger and at most one can be smaller. □

4.4.2 Properties of ϵ-Nash equilibria

One of the central ingredients of the main theorem is provided by the two results stated below. First, Lemma 4.4.4 shows that there exists a constant δ_m, depending on the dimension m, such that any ϵ-approximate Nash equilibrium must assign positive probability mass to the entry $\left(\frac{m}{2}+1, \frac{m}{2}+1\right)$ whenever $\epsilon < \delta_m$. However, this statement alone does not clarify how δ_m depends on m. Ideally, one would like this dependence to be inverse-polynomial, that is,

$$\delta_m \geq \frac{1}{\mathrm{poly}(m)}.$$

Fortunately, Lemma 4.4.4 establishes precisely such a bound.

We include both lemmas because they illustrate two different techniques for proving closely related claims, while also highlighting the advantages and limitations of each approach.

Transformation of $\mathbf{B}_{m,0}$ Before turning to the proofs of Lemmas 4.4.3 and 4.4.4, we describe a transformation of the matrix $\mathbf{A} \coloneqq \mathbf{B}_{m,0}$ that simplifies the proofs of both lemmas.

By Definition 4.4.1, the matrix $\mathbf{A}$ has size $m \times m$, where m is even. We define a permutation $\pi \in \mathfrak{S}_m$ by $\pi(2k-1) = k$ and $\pi(2k) = m-k+1$, $k = 1, 2, \ldots, \frac{m}{2}$. For instance, the permutation is given by

$$\pi(1) = 1, \quad \pi(2) = m, \quad \pi(3) = 2, \quad \pi(4) = m-1, \quad \ldots, \quad \pi(m-1) = \frac{m}{2}, \quad \pi(m) = \frac{m}{2}+1.$$

Let $\mathbf{P}$ be the permutation matrix determined by $\mathbf{P}\boldsymbol{e}_i = \boldsymbol{e}_{\pi(i)}$, and set $\mathbf{G}_m \coloneqq \mathbf{P}^\top \mathbf{B}_{m,0}\mathbf{P}$. A direct inspection of the recursive definition of $\mathbf{B}_{m,0}$ shows that

$$\mathbf{G}_m = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 2 & 3 & 0 & \cdots & 0 \\ 0 & 4 & 5 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 2m-2 & 2m-1 \end{bmatrix}. \tag{4.5}$$

Thus, the original strategy $\frac{m}{2} + 1$ corresponds to the last strategy m in the permuted game. Since a permutation of strategies preserves the Nash equilibrium property, as shown in Lemma 4.4.2, it suffices to prove the claim for the game with common payoff matrix $\mathbf{G}_m$.

Lemma 4.4.2. *Let $\mathbf{B}_{m,0} \in \mathbb{R}^{m\times m}$, and consider the identical-payoff game $(\mathbf{B}_{m,0}, \mathbf{B}_{m,0})$. Let $\mathbf{P}$ be the permutation matrix defined above, and let*

$$\mathbf{G}_m \coloneqq \mathbf{P}^\top \mathbf{B}_{m,0}\mathbf{P}.$$

Then the Nash equilibria are preserved under the corresponding permutation of strategies. More precisely,

$$(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathrm{NE}(\mathbf{B}_{m,0}, \mathbf{B}_{m,0}) \quad \Longleftrightarrow \quad (\mathbf{P}^\top \boldsymbol{x}_1, \mathbf{P}^\top \boldsymbol{x}_2) \in \mathrm{NE}(\mathbf{G}_m, \mathbf{G}_m).$$

Proof. Since every permutation matrix is orthogonal, we have $\mathbf{P}^{-1} = \mathbf{P}^\top$. Moreover, both maps

$$\boldsymbol{x} \mapsto \mathbf{P}\boldsymbol{x} \qquad \text{and} \qquad \boldsymbol{x} \mapsto \mathbf{P}^\top \boldsymbol{x}$$

define bijections from the simplex Δ_m onto itself.

We first prove the forward implication. Suppose that

$$(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathrm{NE}(\mathbf{B}_{m,0}, \mathbf{B}_{m,0}).$$

Then, by definition of Nash equilibrium, for every $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \Delta_m$,

$$\boldsymbol{x}_1^\top \mathbf{B}_{m,0} \boldsymbol{x}_2 \geq \boldsymbol{z}_1^\top \mathbf{B}_{m,0} \boldsymbol{x}_2 \quad \text{and} \quad \boldsymbol{x}_1^\top \mathbf{B}_{m,0} \boldsymbol{x}_2 \geq \boldsymbol{x}_1^\top \mathbf{B}_{m,0} \boldsymbol{z}_2.$$

We claim that

$$(\mathbf{P}^\top \boldsymbol{x}_1, \mathbf{P}^\top \boldsymbol{x}_2) \in \mathrm{NE}(\mathbf{G}_m, \mathbf{G}_m).$$

Let $\widetilde{\boldsymbol{z}}_1 \in \Delta_m$ be an arbitrary deviation of the first player in the permuted game. Since $\mathbf{P}^\top$ is a bijection on Δ_m, there exists $\boldsymbol{z}_1 \in \Delta_m$ such that

$$\widetilde{\boldsymbol{z}}_1 = \mathbf{P}^\top \boldsymbol{z}_1.$$

Equivalently, $\boldsymbol{z}_1 = \mathbf{P}\widetilde{\boldsymbol{z}}_1$. Using $\mathbf{G}_m = \mathbf{P}^\top \mathbf{B}_{m,0} \mathbf{P}$, we obtain

$$\begin{aligned} \widetilde{\boldsymbol{z}}_1^\top \mathbf{G}_m \mathbf{P}^\top \boldsymbol{x}_2 &= (\mathbf{P}^\top \boldsymbol{z}_1)^\top \mathbf{P}^\top \mathbf{B}_{m,0} \mathbf{P} \mathbf{P}^\top \boldsymbol{x}_2 && (4.6) \\ &= \boldsymbol{z}_1^\top \mathbf{P} \mathbf{P}^\top \mathbf{B}_{m,0} \boldsymbol{x}_2 && (4.7) \\ &= \boldsymbol{z}_1^\top \mathbf{B}_{m,0} \boldsymbol{x}_2. && (4.8) \end{aligned}$$

Similarly,

$$\begin{aligned} (\mathbf{P}^\top \boldsymbol{x}_1)^\top \mathbf{G}_m \mathbf{P}^\top \boldsymbol{x}_2 &= (\mathbf{P}^\top \boldsymbol{x}_1)^\top \mathbf{P}^\top \mathbf{B}_{m,0} \mathbf{P} \mathbf{P}^\top \boldsymbol{x}_2 && (4.9) \\ &= \boldsymbol{x}_1^\top \mathbf{B}_{m,0} \boldsymbol{x}_2. && (4.10) \end{aligned}$$

Therefore, by the Nash equilibrium condition in the original game,

$$(\mathbf{P}^\top \boldsymbol{x}_1)^\top \mathbf{G}_m \mathbf{P}^\top \boldsymbol{x}_2 = \boldsymbol{x}_1^\top \mathbf{B}_{m,0} \boldsymbol{x}_2 \geq \boldsymbol{z}_1^\top \mathbf{B}_{m,0} \boldsymbol{x}_2 = \widetilde{\boldsymbol{z}}_1^\top \mathbf{G}_m \mathbf{P}^\top \boldsymbol{x}_2.$$

Thus, $\mathbf{P}^\top \boldsymbol{x}_1$ is a best response to $\mathbf{P}^\top \boldsymbol{x}_2$ in the permuted game.

The argument for the second player is analogous. Let $\widetilde{\boldsymbol{z}}_2 \in \Delta_m$ be an arbitrary deviation of the second player. Since $\mathbf{P}^\top$ is a bijection on Δ_m, there exists $\boldsymbol{z}_2 \in \Delta_m$ such that

$$\widetilde{\boldsymbol{z}}_2 = \mathbf{P}^\top \boldsymbol{z}_2.$$

Then

$$(\mathbf{P}^\top \boldsymbol{x}_1)^\top \mathbf{G}_m \widetilde{\boldsymbol{z}}_2 = (\mathbf{P}^\top \boldsymbol{x}_1)^\top \mathbf{P}^\top \mathbf{B}_{m,0} \mathbf{P} \mathbf{P}^\top \boldsymbol{z}_2 \tag{4.11}$$

$$= \boldsymbol{x}_1^\top \mathbf{B}_{m,0} \boldsymbol{z}_2, \tag{4.12}$$

and

$$(\mathbf{P}^\top \boldsymbol{x}_1)^\top \mathbf{G}_m \mathbf{P}^\top \boldsymbol{x}_2 = \boldsymbol{x}_1^\top \mathbf{B}_{m,0} \boldsymbol{x}_2.$$

Since $\boldsymbol{x}_2$ is a best response to $\boldsymbol{x}_1$ in the original game,

$$\boldsymbol{x}_1^\top \mathbf{B}_{m,0} \boldsymbol{x}_2 \geq \boldsymbol{x}_1^\top \mathbf{B}_{m,0} \boldsymbol{z}_2.$$

Hence,

$$(\mathbf{P}^\top \boldsymbol{x}_1)^\top \mathbf{G}_m \mathbf{P}^\top \boldsymbol{x}_2 \geq (\mathbf{P}^\top \boldsymbol{x}_1)^\top \mathbf{G}_m \widetilde{\boldsymbol{z}}_2.$$

Thus, $\mathbf{P}^\top \boldsymbol{x}_2$ is a best response to $\mathbf{P}^\top \boldsymbol{x}_1$ in the permuted game. Therefore,

$$(\mathbf{P}^\top \boldsymbol{x}_1, \mathbf{P}^\top \boldsymbol{x}_2) \in \mathrm{NE}(\mathbf{G}_m, \mathbf{G}_m).$$

For the converse implication, suppose that

$$(\widetilde{\boldsymbol{x}}_1, \widetilde{\boldsymbol{x}}_2) \in \mathrm{NE}(\mathbf{G}_m, \mathbf{G}_m).$$

Since

$$\mathbf{G}_m = \mathbf{P}^\top \mathbf{B}_{m,0} \mathbf{P},$$

we equivalently have

$$\mathbf{B}_{m,0} = \mathbf{P} \mathbf{G}_m \mathbf{P}^\top.$$

Applying the forward implication to the matrix $\mathbf{G}_m$ with the inverse permutation matrix $\mathbf{P}^\top$, we obtain

$$(\mathbf{P}\widetilde{\boldsymbol{x}}_1, \mathbf{P}\widetilde{\boldsymbol{x}}_2) \in \mathrm{NE}(\mathbf{B}_{m,0}, \mathbf{B}_{m,0}).$$

Equivalently,

$$(\widetilde{\boldsymbol{x}}_1, \widetilde{\boldsymbol{x}}_2) = (\mathbf{P}^\top \boldsymbol{x}_1, \mathbf{P}^\top \boldsymbol{x}_2)$$

for some

$$(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathrm{NE}(\mathbf{B}_{m,0}, \mathbf{B}_{m,0}).$$

Consequently, the Nash equilibria of the original game and of the permuted game are in one-to-one correspondence through the map

$$(\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto (\mathbf{P}^\top \boldsymbol{x}_1, \mathbf{P}^\top \boldsymbol{x}_2).$$

This proves the claim. □

Lemma 4.4.3. *There exists a constant δ_m, such that any ϵ-Nash equilibrium $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ for $\epsilon < \delta_m$ of the identical-payoff game $(\mathbf{G}_m, \mathbf{G}_m)$ satisfies*

$$\boldsymbol{x}_1[m] > 0 \quad \textit{or} \quad \boldsymbol{x}_2[m] > 0. \tag{4.13}$$

Proof. Let

$$K_m \coloneqq \{(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \Delta_m \times \Delta_m \,:\, \boldsymbol{x}_1[m] = 0,\ \boldsymbol{x}_2[m] = 0\}.$$

We first show that K_m contains no exact Nash equilibrium. We argue by induction on m. For $m = 2$,

$$\mathbf{G}_2 = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}.$$

If $\boldsymbol{x}_1[2] = \boldsymbol{x}_2[2] = 0$, then $(\boldsymbol{x}_1, \boldsymbol{x}_2) = (\boldsymbol{e}_1, \boldsymbol{e}_1)$, but against column 1 player 1 strictly prefers row 2 to row 1. Hence K_2 contains no exact Nash equilibrium.

Assume now that $m \geq 3$ and that K_m contains no exact Nash equilibrium for the game with payoff matrix $\mathbf{G}_{m-1}$. Suppose, toward a contradiction, that $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in K_n$ is an exact Nash equilibrium of the game with payoff matrix $\mathbf{G}_m$. Since $\boldsymbol{x}_1[m] = \boldsymbol{x}_2[m] = 0$, both players use only the first $m-1$ strategies. Let $\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2 \in \Delta_{m-1}$ be obtained by deleting the last coordinate from $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively.

Because the upper-left $(m-1)\times(m-1)$ principal submatrix of $\mathbf{G}_m$ is exactly $\mathbf{G}_{m-1}$, for every $i,j \le m-1$ we have $\boldsymbol{e}_i^\top \mathbf{G}_m \boldsymbol{x}_2 = \boldsymbol{e}_i^\top \mathbf{G}_{m-1}\overline{\boldsymbol{x}}_1,$ and $\boldsymbol{x}_1^\top \mathbf{G}_m \boldsymbol{e}_j = \overline{\boldsymbol{x}}_1^\top \mathbf{G}_{m-1}\boldsymbol{e}_j$. Therefore $(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2)$ is an exact Nash equilibrium of the game with payoff matrix $\mathbf{G}_{m-1}$. Moreover, since $\boldsymbol{x}_1[n] = \boldsymbol{x}_2[n] = 0,$ we have $\overline{\boldsymbol{x}}_1[m-1] = 0,$ and $\overline{\boldsymbol{x}}_2[m-1] = 0,$ so $(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2) \in K_{m-1}$, contradicting the induction hypothesis. Hence K_m contains no exact Nash equilibrium.

Now define the regret (Nash gap) function

$$R_m(\boldsymbol{x}_1, \boldsymbol{x}_2) \coloneqq \max\Big\{\max_{i\in[m]} \boldsymbol{e}_i^\top \mathbf{G}_m \boldsymbol{x}_2 - \boldsymbol{x}_1^\top \mathbf{G}_m \boldsymbol{x}_2,\ \max_{j\in[m]} \boldsymbol{x}_1^\top \mathbf{G}_m \boldsymbol{e}_j - \boldsymbol{x}_1^\top \mathbf{G}_m \boldsymbol{x}_2\Big\}.$$

Then $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is an ϵ-Nash equilibrium if and only if $R_m(\boldsymbol{x}_1, \boldsymbol{x}_2) \le \epsilon$. The function R_m is continuous, and K_m is compact. Since K_m contains no exact Nash equilibrium, $R_m(\boldsymbol{x}_1, \boldsymbol{x}_2) > 0$ for all $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in K_m$. Thus, the smallest approximation level, δ_m, for which there exists an ϵ-Nash equilibrium inside K_m is defined as

$$\delta_m \coloneqq \min_{(\boldsymbol{x}_1,\boldsymbol{x}_2)\in K_m} R_m(\boldsymbol{x}_1, \boldsymbol{x}_2) > 0.$$

It follows that any ϵ-Nash equilibrium with $\epsilon < \delta_m$ cannot belong to K_m. Hence such an equilibrium must satisfy $\boldsymbol{x}_1[m] > 0$ or $\boldsymbol{x}_2[m] > 0$. This proves the claim. □

Lemma 4.4.3 shows not only that, for some constant $\delta_m > 0$, any ϵ-approximate Nash equilibrium of the identical-payoff game with payoff matrix $\mathbf{G}_m$ must assign positive probability mass to the strategy profile (m, m) whenever $\epsilon < \delta_m$, but also that this profile is the unique exact Nash equilibrium, corresponding to the case $\epsilon = 0$. The key point is to determine the order of δ_m with respect to m. More specifically, we seek a bound that is inverse-polynomial in the size of the game, that is, in the dimension of the payoff matrix $m \times m$. As we shall see, Lemma 4.4.4 provides precisely such a bound, and in fact proves a stronger statement.

Lemma 4.4.4. *Let $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ be an ϵ^2-approximate Nash equilibrium of the identical-payoff game $(\mathbf{G}_m, \mathbf{G}_m)$, and assume that $\epsilon \leq \frac{1}{16m^3}$. Then*

$$\boldsymbol{x}_1[k] < \epsilon \quad \textit{and} \quad \boldsymbol{x}_2[k] < \epsilon \quad \textit{for every } k \in [m-1]. \tag{4.14}$$

Proof. To simplify notation, we write $\boldsymbol{x}_{1,i}$ (respectively, $\boldsymbol{x}_{2,j}$) instead of $\boldsymbol{x}_1[i]$ (respectively, $\boldsymbol{x}_2[j]$) for the probability assigned to strategy $i \in [m]$ (respectively, $j \in [m]$). Moreover, for a mixed strategy profile $(\boldsymbol{x}_1, \boldsymbol{x}_2)$, we denote by $\boldsymbol{u}_{1,i}$ and $\boldsymbol{u}_{2,j}$ the expected utilities of the row and column players, respectively, from playing the pure strategies i and j.

Let $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \Delta_m \times \Delta_m$ be an ϵ^2-approximate Nash equilibrium of the identical-payoff game $(\mathbf{G}_m, \mathbf{G}_m)$, where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are the mixed strategies of the row and column players, respectively. For a pure strategy $i \in [m]$, the row player's expected utility from playing i against $\boldsymbol{x}_2$ is $\boldsymbol{u}_{1,i} \coloneqq \boldsymbol{u}_1(i, \boldsymbol{x}_2) = \boldsymbol{e}_i^\top \mathbf{G}_m \boldsymbol{x}_2$. By the structure of $\mathbf{G}_m$, we have

$$\boldsymbol{u}_{1,i} = (2i-2)\boldsymbol{x}_{2,i-1} + (2i-1)\boldsymbol{x}_{2,i}, \quad \text{and} \quad \boldsymbol{u}_{1,1} = \boldsymbol{x}_{2,1}. \tag{4.15}$$

Similarly, for a pure strategy $j \in [m]$, the column player's expected utility from playing j against $\boldsymbol{x}_1$ is $\boldsymbol{u}_{2,j} \coloneqq \boldsymbol{u}_2(\boldsymbol{x}_1, j) = \boldsymbol{x}_1^\top \mathbf{G}_m \boldsymbol{e}_j$, and therefore

$$\boldsymbol{u}_{2,j} = (2j-1)\boldsymbol{x}_{1,j} + 2j\,\boldsymbol{x}_{1,j+1}, \quad \text{and} \quad \boldsymbol{u}_{2,m} = (2m-1)\,\boldsymbol{x}_{1,m} \tag{4.16}$$

We prove the claim by induction on $k \in [m]$. We begin with the base case $k = 1$, where we show that

$$\boldsymbol{x}_{1,1} < \epsilon \qquad \text{and} \qquad \boldsymbol{x}_{2,1} < \epsilon.$$

Suppose, toward a contradiction, that

$$\boldsymbol{x}_{1,1} \geq \epsilon \qquad \text{or} \qquad \boldsymbol{x}_{2,1} \geq \epsilon.$$

We distinguish three cases and show that in each case the condition of being an ϵ^2-approximate Nash equilibrium is violated.

1. $\boldsymbol{x}_{1,1} \geq \epsilon$ and $\boldsymbol{x}_{2,1} \geq \epsilon$.

   In this case, the row player has a profitable deviation from strategy 1 to strategy 2. Define the deviation $\tilde{\boldsymbol{x}}_1$ by

   $$\tilde{\boldsymbol{x}}_{1,i} = \begin{cases} 0, & i = 1, \\ \boldsymbol{x}_{1,1} + \boldsymbol{x}_{1,2}, & i = 2, \\ \boldsymbol{x}_{1,i}, & \text{otherwise.} \end{cases}$$

   Then the gain of the row player is

   $$\langle \tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1, \mathbf{G}_m \boldsymbol{x}_2 \rangle = \boldsymbol{x}_{1,1}\big(2\boldsymbol{x}_{2,1} + 3\boldsymbol{x}_{2,2}\big) - \boldsymbol{x}_{1,1}\boldsymbol{x}_{2,1} > \boldsymbol{x}_{1,1}\boldsymbol{x}_{2,1} \geq \epsilon^2.$$

   This contradicts the assumption that $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is an ϵ^2-approximate Nash equilibrium.

2. $\boldsymbol{x}_{1,1} \geq \epsilon$ and $\boldsymbol{x}_{2,1} < \epsilon$.

   As before, the row player has a profitable deviation away from row 1. Since $\epsilon < \frac{1}{m}$, there exists some column $j > 1$ such that $\boldsymbol{x}_{2,j} > \epsilon$. The row player deviates by moving

the probability mass of strategy 1 to strategy j. Define $\tilde{\boldsymbol{x}}_1$ by

$$\tilde{\boldsymbol{x}}_{1,i} = \begin{cases} 0, & i = 1, \\ \boldsymbol{x}_{1,j} + \boldsymbol{x}_{1,1}, & i = j, \\ \boldsymbol{x}_{1,i}, & \text{otherwise.} \end{cases}$$

Then the gain is lower bounded as follows:

$$\begin{aligned} \langle \tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1, \mathbf{G}_m \boldsymbol{x}_2 \rangle &> \boldsymbol{x}_{1,1}(2j-1)\boldsymbol{x}_{2,j} - \boldsymbol{x}_{1,1}\boldsymbol{x}_{2,1} && (4.17) \\ &= \boldsymbol{x}_{1,1}\big((2j-1)\boldsymbol{x}_{2,j} - \boldsymbol{x}_{2,1}\big) && (4.18) \\ &\geq \boldsymbol{x}_{1,1}\boldsymbol{x}_{2,j} \geq \epsilon^2. && (4.19) \end{aligned}$$

This is again a contradiction.

3. $\boldsymbol{x}_{1,1} < \epsilon$ and $\boldsymbol{x}_{2,1} \geq \epsilon$.

This case is symmetric. There exists some $i > 1$ such that $\boldsymbol{x}_{1,i} > \epsilon$, and the column player can profitably deviate from column 1 to column i. Define $\tilde{\boldsymbol{x}}_2$ by

$$\tilde{\boldsymbol{x}}_{2,j} = \begin{cases} 0, & j = 1, \\ \boldsymbol{x}_{2,i} + \boldsymbol{x}_{2,1}, & j = i, \\ \boldsymbol{x}_{2,j}, & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \left\langle \boldsymbol{x}_1^\top \mathbf{G}_m, \tilde{\boldsymbol{x}}_2 - \boldsymbol{x}_2 \right\rangle &> \boldsymbol{x}_{2,1}(2i-1)\boldsymbol{x}_{1,i} - \boldsymbol{x}_{2,1}\boldsymbol{x}_{1,1} && (4.20) \\ &> \boldsymbol{x}_{2,1}\boldsymbol{x}_{1,i} > \epsilon^2. && (4.21) \end{aligned}$$

This again contradicts the ϵ^2-approximate Nash equilibrium condition.

We have therefore established the base case, and now we proceed with the inductive step. Assume as induction hypothesis that for some $k \in [m-1]$,

$$\boldsymbol{x}_{1,i} < \epsilon \quad \text{and} \quad \boldsymbol{x}_{2,j} < \epsilon \qquad \text{for all } i, j \in [k-1].$$

We must show that

$$\boldsymbol{x}_{1,k} < \epsilon \qquad \text{and} \qquad \boldsymbol{x}_{2,k} < \epsilon.$$

We again argue by contradiction, considering each of the two inequalities separately. The proof follows the same general strategy as in the base case of the induction.

1. $\boldsymbol{x}_{1,k} < \epsilon$.

   Suppose, for the sake of contradiction, that $\boldsymbol{x}_{1,k} \geq \epsilon$. Since $k \in [2, m-1]$, the strategy $k+1$ is also available to the row player. We first examine the natural local deviation in which the row player transfers the probability mass assigned to row k to row $k+1$. To this end, define the mixed strategy $\tilde{\boldsymbol{x}}_1$ by

$$\tilde{\boldsymbol{x}}_{1,i} = \begin{cases} 0, & i = k, \\ \boldsymbol{x}_{1,k+1} + \boldsymbol{x}_{1,k}, & i = k+1, \\ \boldsymbol{x}_{1,i}, & \text{otherwise.} \end{cases}$$

   By (4.15), the expected utilities of rows k and $k+1$ are

$$\boldsymbol{u}_{1,k} = (2k-2)\boldsymbol{x}_{2,k-1} + (2k-1)\boldsymbol{x}_{2,k}, \qquad \boldsymbol{u}_{1,k+1} = 2k\,\boldsymbol{x}_{2,k} + (2k+1)\boldsymbol{x}_{2,k+1}.$$

Hence, the gain of the row player from deviating from $\boldsymbol{x}_1$ to $\tilde{\boldsymbol{x}}_1$ is

$$\begin{aligned}\langle \tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1, \mathbf{G}_m \boldsymbol{x}_2 \rangle &= \boldsymbol{x}_{1,k}\big(\boldsymbol{u}_{1,k+1} - \boldsymbol{u}_{1,k}\big) && (4.22)\\ &= \boldsymbol{x}_{1,k}\Big(\boldsymbol{x}_{2,k} + (2k+1)\boldsymbol{x}_{2,k+1} - (2k-2)\boldsymbol{x}_{2,k-1}\Big) && (4.23)\\ &> \boldsymbol{x}_{1,k}\Big(\boldsymbol{x}_{2,k} + \boldsymbol{x}_{2,k+1} - (2k-2)\epsilon\Big), && (4.24)\end{aligned}$$

where the last inequality follows from the induction hypothesis, since $\boldsymbol{x}_{2,k-1} < \epsilon$.

Now, if

$$\boldsymbol{x}_{2,k} + \boldsymbol{x}_{2,k+1} - (2k-2)\epsilon > \epsilon,$$

then (4.24) yields

$$\langle \tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1, \mathbf{G}_m \boldsymbol{x}_2 \rangle > \boldsymbol{x}_{1,k}\epsilon \geq \epsilon^2,$$

which contradicts the assumption that $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is an ϵ^2-approximate Nash equilibrium.

Therefore, we may assume that

$$\boldsymbol{x}_{2,k} + \boldsymbol{x}_{2,k+1} \leq (2k-1)\epsilon. \tag{4.25}$$

Under this assumption, we can bound the utility of row k. Indeed,

$$\begin{aligned}\boldsymbol{u}_{1,k} &= (2k-2)\boldsymbol{x}_{2,k-1} + (2k-1)\boldsymbol{x}_{2,k} && (4.26)\\ &< (2k-2)\epsilon + (2k-1)^2\epsilon < 8m^2\epsilon, && (4.27)\end{aligned}$$

where we used the induction hypothesis $\boldsymbol{x}_{2,k-1} < \epsilon$, together with (4.25). Next, observe from (4.15) that for every i,

$$\boldsymbol{u}_{1,i} = (2i-2)\boldsymbol{x}_{2,i-1} + (2i-1)\boldsymbol{x}_{2,i} \geq \boldsymbol{x}_{2,i}.$$

Hence,

$$\sum_{i=k+2}^{m} \boldsymbol{u}_{1,i} \geq \sum_{i=k+2}^{m} \boldsymbol{x}_{2,i}.$$

Now,

$$\begin{aligned}
\sum_{i=k+2}^{m} \boldsymbol{x}_{2,i} &= 1 - \sum_{i=1}^{k+1} \boldsymbol{x}_{2,i} && (4.28)\\
&= 1 - (\boldsymbol{x}_{2,k} + \boldsymbol{x}_{2,k+1}) - \sum_{i=1}^{k-1} \boldsymbol{x}_{2,i} && (4.29)\\
&> 1 - (2k-1)\epsilon - (k-1)\epsilon && (4.30)\\
&> 1 - 3m\epsilon, && (4.31)
\end{aligned}$$

where we used (4.25) and the induction hypothesis $\boldsymbol{x}_{2,i} < \epsilon$ for all $i \in [k-1]$.

Therefore, there exists some $i^* \in \{k+2, \ldots, m\}$ such that

$$\boldsymbol{u}_{1,i^*} > \frac{1-3m\epsilon}{m}.$$

We now consider the deviation in which the row player transfers the probability mass of row k to row i^*, namely

$$\tilde{\boldsymbol{x}}_{1,i} = \begin{cases} 0, & i = k, \\ \boldsymbol{x}_{1,i^*} + \boldsymbol{x}_{1,k}, & i = i^*, \\ \boldsymbol{x}_{1,i}, & \text{otherwise.} \end{cases}$$

The gain from this deviation is

$$\begin{aligned} \langle \tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1, \mathbf{G}_m \boldsymbol{x}_2 \rangle &= \boldsymbol{x}_{1,k} \big(\boldsymbol{u}_{1,i^*} - \boldsymbol{u}_{1,k} \big) && (4.32) \\ &> \boldsymbol{x}_{1,k} \left(\frac{1 - 3m\epsilon}{m} - 8m^2\epsilon \right). && (4.33) \end{aligned}$$

Thus, it is enough that the term inside the parenthesis is greater than ϵ. Equivalently, it should hold $\frac{1}{m} > (8m^2 + 4)\epsilon$. Hence the above inequality holds whenever

$$\epsilon < \frac{1}{8m^3 + 4m} \leq \frac{1}{16m^3}.$$

Under this condition, we obtain

$$\langle \tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1, \mathbf{G}_m \boldsymbol{x}_2 \rangle > \boldsymbol{x}_{1,k}\epsilon \geq \epsilon^2,$$

again contradicting the ϵ^2-approximate Nash equilibrium condition. Therefore, the assumption $\boldsymbol{x}_{1,k} \geq \epsilon$ is impossible, and so $\boldsymbol{x}_{1,k} < \epsilon$.

2. $\boldsymbol{x}_{2,k} < \epsilon$.

Suppose, again for the sake of contradiction, that

$$\boldsymbol{x}_{2,k} \geq \epsilon.$$

At this point we may use the first part of the inductive step, which has already established that

$$\boldsymbol{x}_{1,k} < \epsilon.$$

Combined with the inductive hypothesis, this implies that all rows in $[k]$ carry only negligible probability mass. More precisely,

$$\boldsymbol{x}_{1,i} < \epsilon \qquad \text{for every } i \in [k].$$

Hence the total probability mass assigned by the row player to strategies $k+1, \ldots, m$ satisfies

$$\sum_{i=k+1}^{m} \boldsymbol{x}_{1,i} = 1 - \sum_{i=1}^{k} \boldsymbol{x}_{1,i} > 1 - k\epsilon > 1 - m\epsilon.$$

Therefore, there exists some row $i^* \in \{k+1, \ldots, m\}$ whose probability mass is at least the average over these remaining coordinates. In particular,

$$\boldsymbol{x}_{1,i^*} \geq \frac{1}{m-k} \sum_{i=k+1}^{m} \boldsymbol{x}_{1,i} > \frac{1-m\epsilon}{m-k} \geq \frac{1-m\epsilon}{m} = \frac{1}{m} - \epsilon.$$

Let us choose $i^* \in \arg\max_{i \in \{k+1,\ldots,m\}} \boldsymbol{x}_{1,i}$. Then $\boldsymbol{x}_{1,i^*} \geq \boldsymbol{x}_{1,i}$ for all $i \in \{k+1, \ldots, m\}$, and in particular $\boldsymbol{x}_{1,i^*} \geq \frac{1}{m} - \epsilon$.

We now consider the deviation in which the column player transfers the probability mass assigned to column k to column i^*. Define $\tilde{\boldsymbol{x}}_2$ by

$$\tilde{\boldsymbol{x}}_{2,j} = \begin{cases} 0, & j = k, \\ \boldsymbol{x}_{2,i^*} + \boldsymbol{x}_{2,k}, & j = i^*, \\ \boldsymbol{x}_{2,j}, & \text{otherwise.} \end{cases}$$

By (4.16), the expected utilities of columns k and i^* are

$$\boldsymbol{u}_{2,k} = (2k-1)\boldsymbol{x}_{1,k} + 2k\,\boldsymbol{x}_{1,k+1}, \qquad \boldsymbol{u}_{2,i^*} = (2i^*-1)\boldsymbol{x}_{1,i^*} + 2i^*\,\boldsymbol{x}_{1,i^*+1}.$$

Hence, the gain of the column player from deviating from $\boldsymbol{x}_2$ to $\tilde{\boldsymbol{x}}_2$ is

$$\begin{aligned} \left\langle \boldsymbol{x}_1^\top \mathbf{G}_m, \tilde{\boldsymbol{x}}_2 - \boldsymbol{x}_2 \right\rangle &= \boldsymbol{x}_{2,k}\big(\boldsymbol{u}_{2,i^*} - \boldsymbol{u}_{2,k}\big) && (4.34) \\ &> \boldsymbol{x}_{2,k}\Big((2i^*-1)\boldsymbol{x}_{1,i^*} - 2k\,\boldsymbol{x}_{1,k+1} - (2k-1)\epsilon\Big), && (4.35) \end{aligned}$$

where we dropped the nonnegative term $2i^*\,\boldsymbol{x}_{1,i^*+1}$, and used the fact, established earlier in the inductive step, that $\boldsymbol{x}_{1,k} < \epsilon$.

Since $i^* \in \{k+1, \ldots, m\}$, we have $2i^* - 1 \geq 2k+1$. Moreover, because i^* maximizes $\boldsymbol{x}_{1,i}$ over $\{k+1, \ldots, m\}$, it follows that $\boldsymbol{x}_{1,i^*} \geq \boldsymbol{x}_{1,k+1}$. Hence,

$$\begin{aligned} (2i^*-1)\boldsymbol{x}_{1,i^*} - 2k\,\boldsymbol{x}_{1,k+1} &\geq (2k+1)\boldsymbol{x}_{1,i^*} - 2k\,\boldsymbol{x}_{1,k+1} && (4.36) \\ &\geq (2k+1)\boldsymbol{x}_{1,i^*} - 2k\,\boldsymbol{x}_{1,i^*} && (4.37) \\ &= \boldsymbol{x}_{1,i^*}. && (4.38) \end{aligned}$$

Using this bound in (4.35), we obtain

$$\left\langle \boldsymbol{x}_1^\top \mathbf{G}_m, \tilde{\boldsymbol{x}}_2 - \boldsymbol{x}_2 \right\rangle > \boldsymbol{x}_{2,k}\big(\boldsymbol{x}_{1,i^*} - (2k-1)\epsilon\big) \tag{4.39}$$

$$> \boldsymbol{x}_{2,k}\left(\frac{1}{m} - \epsilon - (2k-1)\epsilon\right) \tag{4.40}$$

$$> \boldsymbol{x}_{2,k}\left(\frac{1}{m} - 2m\epsilon\right), \tag{4.41}$$

where the second inequality follows from the previously established bound $\boldsymbol{x}_{1,i^*} \geq \frac{1}{m} - \epsilon$.

It therefore remains to verify that $\frac{1}{m} - 2m\epsilon > \epsilon$. This is ensured by our choice $\epsilon \leq \frac{1}{16m^3}$, since for every $m \geq 1$,

$$(2m+1)\epsilon \leq \frac{2m+1}{16m^3} < \frac{1}{m}.$$

Consequently,

$$\left\langle \boldsymbol{x}_1^\top \mathbf{G}_m, \tilde{\boldsymbol{x}}_2 - \boldsymbol{x}_2 \right\rangle > \boldsymbol{x}_{2,k}\epsilon \geq \epsilon^2,$$

which contradicts the assumption that $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is an ϵ^2-approximate Nash equilibrium.

Therefore, the assumption $\boldsymbol{x}_{2,k} \geq \epsilon$ is impossible, and we conclude that

$$\boldsymbol{x}_{2,k} < \epsilon.$$

Therefore, we conclude that $\boldsymbol{x}_{1,k} < \epsilon$ and $\boldsymbol{x}_{2,k} < \epsilon$. This establishes the inductive step, and hence the lemma follows by induction.

□

Remark 4.4.2. Our original goal was to determine the dependence of δ_m on m, so as to identify the approximation threshold for which strategy (m, m) must receive positive probability mass in Lemma 4.4.3. In fact, Lemma 4.4.4 proves a strictly stronger statement: not only must (m, m) receive positive probability mass, but it must in fact carry the majority of the probability mass.

We now translate the above lemmas back to the original matrix $\mathbf{A} = \mathbf{B}_{m,0}$. Applying the inverse permutation, we obtain

$$\pi^{-1}(m) = \frac{m}{2} + 1.$$

Thus, the fact that an equilibrium of the permuted game assigns positive probability to strategy m implies that, in the original game, the corresponding equilibrium assigns positive probability to strategy $\frac{m}{2} + 1$. We record this observation formally below.

Corollary 4.4.1. *Let $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ be an ϵ-approximate Nash equilibrium of the identical-payoff game with payoff matrix $\mathbf{A} = \mathbf{B}_{m,0}$, with $\epsilon = O\big(\frac{1}{m^6}\big)$. Then*

$$\boldsymbol{x}_1\Big[\frac{m}{2} + 1\Big] > 0 \quad \textit{or} \quad \boldsymbol{x}_2\Big[\frac{m}{2} + 1\Big] > 0.$$

4.4.3 Exponential lower bound for fictitious play

In this subsection, we prove Theorem 4.4.1. To this end, we first show that fictitious play requires (super-)exponential time before assigning positive probability mass to the entry $(a_1(2m-1, a_2(2m-1)) = \big(\frac{m}{2} + 1, \frac{m}{2} + 1\big)$. This is established through the main technical result of the subsection, namely Lemma 4.4.5. For ease of exposition, we introduce the notion of a *period*.

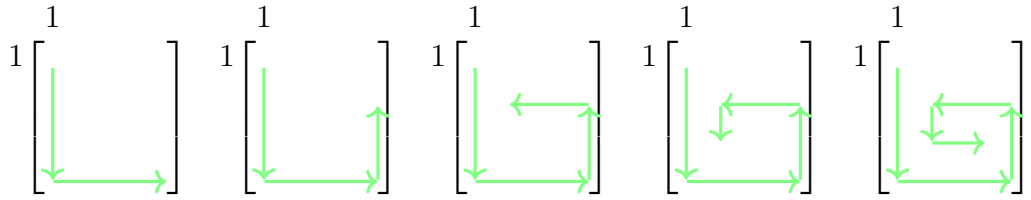


Figure 4.2: Spiral trajectory of fictitious play in the game with payoff matrix $\mathbf{A}$. Starting from the upper-left entry, fictitious play visits all nonzero entries in increasing order of their values.

Definition 4.4.3 (Transition period)**.** For each $k \in \{1, \ldots, 2m-1\}$, the *transition period* k is the set of rounds t for which the following conditions hold $(i^{(t)}, j^{(t)}) = (a_1(k), a_2(k))$.

For $k \geq 1$, we let $\underline{t_k}$ and $\overline{t_k}$ denote the first and last rounds of period k, respectively. $T_k \coloneqq \overline{t_k} - \underline{t_k} + 1$ denote its length, so that period k corresponds to the interval $t \in [\underline{t_k}, \overline{t_k}]$.

Period consistency We are now ready to state an important lemma, which we refer to as the *period consistency* lemma. As already suggested throughout the chapter, the structure of the identical-payoff game $(\mathbf{A}, \mathbf{A})$ dictates the trajectory of fictitious play, namely the spiral path illustrated in Figure 4.2. We now formalize this claim and prove it inductively.

Property 4.4.1 (Period consistency)**.** Suppose both players employ `FP` in the identical-payoff game $(\mathbf{A}, \mathbf{A})$, starting from the pure strategy profile $(i^{(1)}, j^{(1)}) = (1, 1)$, and under an arbitrary tie-breaking rule. If a round t belongs to period k for some $3 \leq k \leq 2m-1$, then the subsequent round $t+1$ belongs either to the same period k or to the next period $k+1$.

Proof. We prove the claim by induction on the period index k.

Assume that round t belongs to period k, where $3 \leq k \leq 2m-2$. By definition of the periods, the strategy profile played at round t is $(a_1(k), a_2(k))$. We show that round $t+1$ belongs either to period k or to period $k+1$.

The key observation is given by Proposition 4.4.2, which states that every nonzero entry has at most one larger nonzero entry in its row or column. Assume first that k is even. Then, by Item (ii) in Proposition 4.4.1, we have $a_1(k) = a_1(k+1)$, whereas by Item (iii) we have $a_2(k-2) = a_2(k-1) \neq a_2(k) = a_2(k+1)$, since each nonzero entry appears exactly once by Lemma 4.4.1. Thus, it remains to inspect the cumulative utility vectors of the two players.

$$\mathbf{A}\boldsymbol{e}_{a_1(k)} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ k-1 \\ 0 \\ 0 \\ k \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \succ a_1(k-1) \\ \\ \\ \succ a_1(k) \\ \\ \\ \end{matrix} \qquad \boldsymbol{e}_{a_2(k)}^{\top}\mathbf{A} = \begin{bmatrix} 0 & \cdots & 0 & \underbrace{k}_{a_2(k)} & 0 & 0 & \underbrace{k+1}_{a_2(k+1)} & \cdots & 0 \end{bmatrix}. \tag{4.42}$$

Figure 4.3: Utility vectors received by the players at a round $t \in [\underline{t_k}, \overline{t_k}]$, for k even. In this case, only the column player has a better response available, while the row player has no incentive to change strategy.

Since $t \in [\underline{t_k}, \overline{t_k}]$, it holds that $a_1(k) \in \arg\max_{a_1 \in \mathcal{A}_1} \mathtt{R}^{(t)}[a_1]$ and $a_2(k) \in \arg\max_{a_2 \in \mathcal{A}_2} \mathtt{C}^{(t)}[a_2]$. From (4.42), it follows that after the update at round $t+1$, the row player still has no incentive to deviate, so $a_1(k) \in \arg\max_{a_1 \in \mathcal{A}_1} \mathtt{R}^{(t+1)}[a_1]$. On the other hand, for the column player, either $a_2(k)$ remains a best response or $a_2(k+1)$ becomes a best response.

Therefore, when the players update at round $t+1$, any deviation from the current profile can only move the dynamics from the entry of value k to the adjacent entry of value $k+1$, since $(a_1(k), a_2(k+1)) = (a_1(k+1), a_2(k+1))$. No jump to a period $k' > k+1$ is possible, because such an entry lies in neither the same row nor the same column as the current one. Likewise, the dynamics cannot move backward to a period $k' < k$, since by construction the update rule only allows the current period either to persist or to advance to the next one.

Hence, after a round in period k, the next round $t+1$ belongs either to period k, if the current strategy profile is repeated, or to period $k+1$. This proves the induction step for even k. The case where k is odd is analogous.

Since the argument applies to every $k \in \{3, \ldots, 2m-2\}$, the claim follows. □

Lemma 4.4.5 (Recursive structure of the fictitious play trajectory)**.** *Suppose both players employ* FP *in the identical-payoff game* $(\mathbf{A}, \mathbf{A})$*, starting from the pure strategy profile* $(i^{(1)}, j^{(1)}) = (1,1)$*, and under an arbitrary tie-breaking rule. Then, for every* $k \in \{2, \ldots, 2m-1\}$*, the following hold:*

(1) $(i^{(t)}, j^{(t)}) = (a_1(k), a_2(k))$ *for all* $t \in [\underline{t_k}, \overline{t_k}]$.

(2) If k *is even, then for every* $t \in [\underline{t_k}, \overline{t_k}]$ *and every* $\ell \geq k+2$, $R^{(t)}[a_1(\ell)] = 0$.

(3) If k *is odd, then for every* $t \in [\underline{t_k}, \overline{t_k}]$ *and every* $\ell \geq k+2$, $C^{(t)}[a_2(\ell)] = 0$

Sketch Proof. Lemma 4.4.5 now follows directly from Property 4.4.1, since all the necessary components have already been established formally in the proof of the latter. □

Item (1) states that there exists a consecutive block of rounds during which a fixed pure strategy profile is played throughout. Moreover, Items (2) and (3) state that the cumulative utilities of all strategies with payoff value larger than $(k+1)$ remain zero. Thus, Lemma 4.4.5 provides precisely the inductive hypothesis needed for the proof of Theorem 4.4.1.

Lemma 4.4.6 (Recursive formula for the period lengths)**.** *Suppose both players employ* FP *in the identical-payoff game* $(\mathbf{A}, \mathbf{A})$*, starting from the pure strategy profile* $(i^{(1)}, j^{(1)}) = (1,1)$*, and under an arbitrary tie-breaking rule. Then, for every* $k \in \{3, \ldots, 2m-2\}$*, it holds* $T_k \geq (k-1)\big(T_{k-2} + T_{k-1}\big)$.

Proof. Assume that k is even. Then, by Item (1) in Lemma 4.4.5, there exists a round $\underline{t_k}$ at which the players choose the strategy profile $(a_1(k), a_2(k))$ for the first time. Since k is even, Item (ii) in Proposition 4.4.1 yields $a_1(k) = a_1(k+1)$. Hence, the larger entry lies in the same row, and therefore, by Proposition 4.4.2, we must have $a_2(k) \neq a_2(k+1)$.

Thus, using once again Item (3) in Lemma 4.4.5, we obtain that at round $\underline{t_k} - 1 \coloneqq \overline{t_{k-1}}$, which belongs to period $k-1$,

$$\mathtt{C}^{(\underline{t_k}-1)}[a_2(k+1)] = 0 \quad \text{and} \quad \mathtt{R}^{(\underline{t_k}-1)}[a_1(k+1)] = \mathtt{R}^{(\underline{t_k}-1)}[a_1(k)] > 0. \tag{4.43}$$

On the other hand, $\mathtt{C}^{(\underline{t_k})}[a_2(k+1)] > 0$. In particular, the profile $(a_1(k), a_2(k))$ cannot have been played before round $\underline{t_k}$; otherwise, the cumulative utility of column $a_2(k+1)$ would already be positive at round $\overline{t_{k-1}}$, contradicting (4.43).

We now estimate the number of rounds required before a switch to the strategy profile $(a_1(k+1), a_2(k+1))$ can occur. Since after round $\underline{t_k}$ only the column player has a profitable deviation, it suffices to estimate the first round $\overline{t_k} \coloneqq \underline{t_{k+1}} - 1$ at which strategy $a_2(k+1)$ becomes a best response to the empirical strategy $\overline{\boldsymbol{x}}_2^{(\overline{t_k})}$. Equivalently, in terms of the cumulative utility vectors introduced in Definition 4.3.3, $\overline{t_k}$ is the first round such that $\mathtt{C}^{(\overline{t_k})}[a_2(k+1)] \geq \mathtt{C}^{(\overline{t_k})}[a_2(k)]$.

Now, the column player updates their cumulative utility vector according to

$$\boldsymbol{e}_{a_2(k)}^{\top}\mathbf{A} = \Big[0 \quad \cdots \quad 0 \quad \underbrace{k}_{a_2(k)} \quad 0 \quad 0 \quad \underbrace{k+1}_{a_2(k+1)} \quad \cdots \quad 0\Big]. \tag{4.44}$$

Thus, the two strategies differ by exactly one unit of utility. Moreover, by (4.43), $\mathtt{C}^{(\underline{t_k}-1)}[a_2(k+1)] = 0$, and $\mathtt{C}^{(\underline{t_k}-1)}[a_2(k)] > 0$. Therefore, at least $\mathtt{C}^{(\underline{t_k}-1)}[a_2(k)]$ additional rounds are re-

quired before the two cumulative utilities can become equal, that is, before

$$\mathtt{C}^{(\overline{t_k})}[a_2(k)] = \mathtt{C}^{(\overline{t_k})}[a_2(k+1)].$$

It follows that

$$T_k = \overline{t_k} - \underline{t_k} + 1 \geq \mathtt{C}^{(\underline{t_k}-1)}[a_2(k)]. \tag{4.45}$$

Hence, this lower bound on the length of period k is independent of the tie-breaking rule, since the argument does not rely on how ties are resolved once equality is reached.

It remains to estimate the quantity $\mathtt{C}^{(\overline{t_{k-1}})}[a_2(k)]$. To this end, we identify the periods during which the cumulative utility of strategy $a_2(k)$ increases. By Item (ii) in Proposition 4.4.1, we have $a_2(k-2) = a_2(k-1)$. Therefore, for every $t \in [\underline{t_{k-2}}, \overline{t_{k-2}}] \cup [\underline{t_{k-1}}, \overline{t_{k-1}}]$, the quantity $\mathtt{C}^{(t)}[a_2(k)]$ increases exactly by $k-1$. Consequently,

$$\mathtt{C}^{(\overline{t_{k-1}})}[a_2(k)] \geq (k-1)\big(T_{k-2} + T_{k-1}\big). \tag{4.46}$$

Combining (4.45) and (4.46), we conclude that

$$T_k \geq (k-1)\big(T_{k-2} + T_{k-1}\big).. \tag{4.47}$$

Of course, the case where k is odd is analogous. □

From Lemma 4.4.6, it follows that, for a very long time, the Nash equilibrium profile $(a_1(2m-1), a_2(2m-1))$ has not yet been played. However, our goal is to identify a round up to which neither of its two components, $a_1(2m-1)$ and $a_2(2m-1)$, has been selected by fictitious play. For this purpose, it suffices to consider period $2m-3$, since Proposition 4.4.1 shows that neither of these two components played during that period coincides with the components of

the Nash equilibrium. Moreover, due to the super-exponential growth of the period lengths, it is not necessary to account for all preceding periods: the contribution of T_{2m-3} alone is already sufficient to establish the claim.

The final step is therefore to unfold the recursive relation for the period lengths given in Lemma 4.4.6 and derive a lower bound on T_{2m-3}. We will provide a crude lower bound, as this is even enough to prove Theorem 4.4.1.

Lemma 4.4.7. *Let $T_1 = 1$, $T_2 = 2$, and suppose that $T_k \geq (k-1)\big(T_{k-2} + T_{k-1}\big)$ for all $k \in \{3, 4, \ldots, 2m-2\}$. Then $T_{2m-3} = m^{\Omega(m)}$.*

Proof. Since $T_k \geq (k-1)T_{k-1}$ for every $k \geq 3$, iterating this inequality gives

$$T_{2m-3} \geq \prod_{k=3}^{2m-3} (k-1)\, T_2 = 2 \prod_{j=2}^{2m-4} j = 2(2m-4)!,$$

and Stirling's formula implies that $(2m-4)! = m^{\Omega(m)}$. Hence, $T_{2m-3} = m^{\Omega(m)}$. □

Having established all the technical ingredients above, we now turn to the proof of Theorem 4.4.1. By Property 4.4.1, fictitious play progresses through successive periods $k \in \{2, 3, \ldots, 2m-1\}$, starting from period 1. It therefore remains to show that, throughout this trajectory and before entering period $2m-2$, the empirical strategy profile $(\overline{\boldsymbol{x}}_1^{(t)}, \overline{\boldsymbol{x}}_2^{(t)})$ is not a Nash equilibrium.

Proof of Theorem 4.4.1. By Property 4.4.1, fictitious play visits the periods successively. Hence, up to round $\overline{t_{2m-3}} = \sum_{k=1}^{2m-3} T_k$, the dynamics has not yet entered period $2m-2$. In particular, by the definition of the periods and the structure of the trajectory, neither of the two components of the strategy profile

$$(a_1(2m-1), a_2(2m-1)) = \Big(\frac{m}{2} + 1, \frac{m}{2} + 1\Big)$$

has been played by fictitious play up to round $t = \overline{t_{2m-3}}$. On the one hand, since fictitious play selects only pure strategies and Lemma 4.4.3 shows that

$$\left(\frac{m}{2}+1, \frac{m}{2}+1\right)$$

is the unique pure Nash equilibrium, we conclude that FP has not converged in the last-iterate sense. On the other hand, the empirical strategy profile $(\overline{\boldsymbol{x}}_1^{(t)}, \overline{\boldsymbol{x}}_2^{(t)})$ assigns zero probability to strategy $\frac{m}{2}+1$ for both players. It then follows from Lemma 4.4.4 that, for every $t \leq \overline{t_{2m-3}}$, the empirical strategy profile $(\overline{\boldsymbol{x}}_1^{(t)}, \overline{\boldsymbol{x}}_2^{(t)})$ cannot be an ϵ-approximate Nash equilibrium. Thus fictitious play requires strictly more than $\overline{t_{2m-3}}$ rounds before reaching such an equilibrium.

Finally, by Lemma 4.4.6, we have

$$T_{2m-3} = m^{\Omega(m)}.$$

Since $\overline{t_{2m-3}} \geq T_{2m-3}$, we conclude that fictitious play requires at least $m^{\Omega(m)}$ rounds to reach a Nash equilibrium. This proves the theorem. □

4.5 Experiments

In this section, we aim to experimentally validate Theorem 4.4.1 in the 4×4 example of Figure 4.1a.

Our analysis focuses on three key quantities: the rounds at which strategy switches occur, the Nash gap throughout the trajectory, and the empirical strategy profiles $\left(\overline{\boldsymbol{x}}_1^{(t)}, \overline{\boldsymbol{x}}_2^{(t)}\right)$.

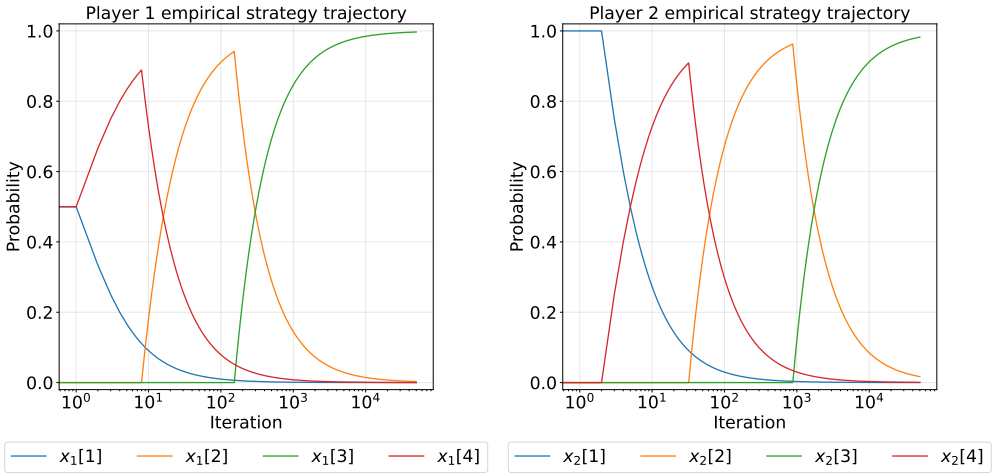


Figure 4.4: The figure displays the empirical strategy profiles of both players. To preserve the crucial qualitative features in both diagrams, the x-axis was set to a logarithmic scale.

t	(i, j)
1	$(1, 1)$
2	$(4, 1)$
3	$(4, 4)$
7	$(2, 4)$
22	$(2, 2)$
98	$(3, 2)$
553	$(3, 3)$

(4.48)

In (4.48), we report the rounds at which strategy switches occur. As a tie-breaking rule, we choose the fastest possible one: whenever a tie occurs, fictitious play selects the strategy that was not played in the previous round. As predicted by the analysis, fictitious play *visits* all strategies, and does so in increasing order of their utility values, before eventually reaching the pure Nash equilibrium.

In Figure 4.4, we present the empirical strategy profiles of both players. To preserve the relevant qualitative features in both plots, the horizontal axis is displayed on a logarithmic

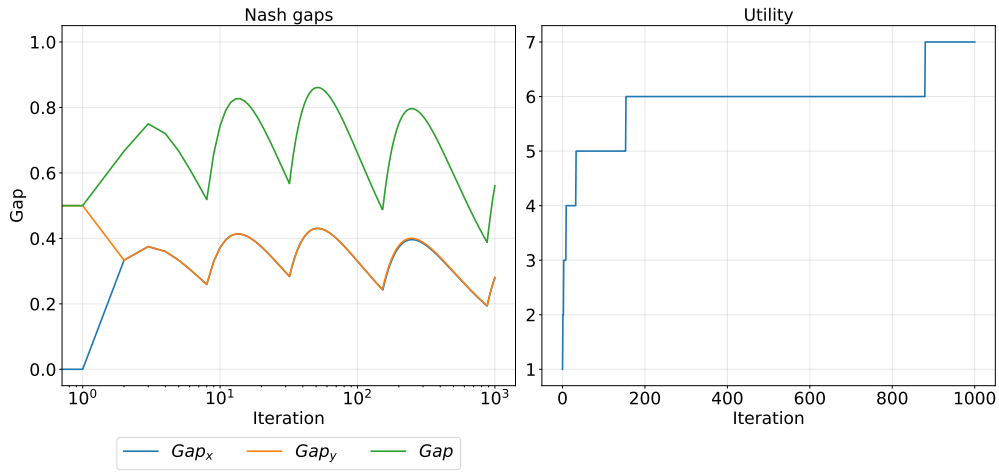


Figure 4.5: On the left, we depict the Nash gap of the empirical strategy profile, while on the right we plot the utility of the last iterate $(\boldsymbol{x}_1^{(t)}, \boldsymbol{x}_2^{(t)})$.

scale. It is clear that, after each strategy switch, the effect of the newly selected strategy takes increasingly longer to become visible.

More importantly, Figure 4.5 reveals a recurring pattern in the Nash gap: the gap increases whenever a new, higher-utility strategy is selected, and then decreases until the next strategy switch. However, it remains bounded away from zero, indicating that a profitable deviation is always available and hence that a Nash equilibrium has not yet been reached. This pattern disappears once the pure Nash equilibrium is reached, which, in accordance with Theorem 4.4.1, is the unique approximate Nash equilibrium.

4.6 Concluding remarks

The results of this chapter show that asymptotic convergence guarantees for fictitious play in potential games can conceal severe quantitative inefficiencies. Although potential games possess a global potential function and fictitious play is known to converge in this class, the construction above demonstrates that the number of rounds required to reach an approximate Nash equilibrium can be exponential in the number of actions.

The lower bound is driven by a recursive payoff construction that forces fictitious play to follow a long sequence of strategy switches before it can place probability mass on the equilibrium region. Thus, the obstruction is not nonconvergence, but rather the length of the transient phase before convergence becomes visible at the level of approximate equilibrium.

This result complements the positive result of Chapter 3. Whereas rank-1 structure can be exploited to obtain an efficient learning algorithm, the potential-game structure studied here does not by itself guarantee efficient convergence of canonical dynamics. The next chapter continues this lower-bound perspective by studying follow-the-regularized-leader and related no-regret dynamics in potential games.

Chapter 5

(Doubly) Exponential Lower Bounds for Follow the Regularized Leader in Potential Games

5.1 Preface

This chapter is based on joint work Anagnostides et al. (2026a), with the exposition adapted from the corresponding paper for inclusion in this dissertation. In particular, the notation has been aligned with the rest of the thesis, several theorem statements have been reformulated, and parts of the proofs have been reworked to emphasize the connection with the lower-bound framework developed in Chapter 4.

The chapter continues the lower-bound direction initiated in Chapter 4. There, we showed that fictitious play can be exponentially slow in two-player potential games, despite its classical asymptotic convergence guarantees. The present chapter asks whether this inefficiency is specific to fictitious play, or whether it persists for broader families of learning dynam-

ics arising from online optimization. The focus is on follow-the-regularized-leader (FTRL), a foundational no-regret algorithmic framework that includes multiplicative weights update and is closely related to mirror descent.

We show that FTRL can take exponential time to converge to a Nash equilibrium in two-player potential games, for any permutation-invariant regularizer (see Section A.1) and for learning rates that may vanish over time. By standard equivalences, this lower bound also applies to certain mirror-descent counterparts, most notably multiplicative weights update. Thus, even in potential games, the combination of no-regret guarantees and a global potential function does not imply efficient last-iterate convergence to Nash equilibrium.

The chapter also establishes complementary positive results. We prove a potential-monotonicity property for FTRL under suitable learning-rate conditions, and we obtain an exponential upper bound, $\exp(O_\epsilon(1/\epsilon^2))$, for no-regret dynamics executed in a lazy, alternating fashion. This upper bound matches the lower-bound behavior up to factors in the exponent, thereby clarifying the quantitative scale of the obstruction.

Finally, we extend the lower-bound perspective to multiplayer potential games. In this setting, we show that fictitious play, which can be viewed as the extreme version of FTRL, can require *doubly* exponentially many iterations to reach a Nash equilibrium. This yields an exponentially stronger lower bound for fictitious play than the two-player construction of Chapter 4, and further reinforces the central message of the thesis: asymptotic convergence and structural favorability do not by themselves guarantee efficient equilibrium learning.

5.2 Introduction

Multiplicative weights update (MWU) is the quintessential online algorithm (Littlestone and Warmuth, 1994), attaining the optimal *regret* bound in many fundamental online learning problems (Fan et al., 2025). It has found applications in diverse areas ranging from approx-

imation algorithms to boosting in machine learning (Arora et al., 2012). MWU is an instantiation of the seminal online learning paradigm *follow the regularized leader* (*FTRL*) (Kalai and Vempala, 2005). Viewed differently, MWU can also be cast as an instance of online *mirror descent* (*MD*) (Nemirovsky et al., 1983).

A celebrated connection inextricably links the no-regret property, attained by algorithms such as FTRL and MD, to *coarse correlated equilibria (CCEs)*, a seminal game-theoretic solution concept (Moulin and Vial, 1978; Aumann, 1974). Specifically, if each player in a multi-player game employs a no-regret algorithm to update its strategy, the *average* correlated distribution of play converges to the set of CCEs. This connection, however, provides no information about the *last-iterate* convergence of the dynamics.

Characterizing the convergence of MWU and other online algorithms has received extensive attention. Kleinberg et al. (2009) first established that MWU converges to Nash equilibria for almost all *potential* games. In this setting, players are effectively ascending a global potential function; thus, the dynamics can be viewed through the lens of constrained optimization, with Nash equilibria corresponding to first-order stationary points. Non-asymptotic rates soon emerged for some versions of MD, most notably gradient descent, but despite decades of research on this topic, a major question remained hitherto wide open:

How many iterations are needed for FTRL algorithms to converge in potential games?

Viewed differently, can MWU or FTRL efficiently find first-order stationary points in constrained optimization?

Although MWU has been shown to asymptotically converge in potential games—at least when the set of Nash equilibria comprises isolated points, a condition that holds for almost all potential games (Kleinberg et al., 2009; Palaiopanos et al., 2017)—this result has not been extended to the general class of FTRL. FTRL presents distinct challenges that complicate

the analysis compared to algorithms such as gradient descent. For example, it can exhibit *spurious fixed points*: due to its history dependence, the dynamics may stagnate even when the current strategy is highly suboptimal relative to the observed utility (Example 5.4.1).

5.2.1 Our results

Algorithm 5. Follow-the-Regularized-Leader (FTRL)

Input: Payoff matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m_1 \times m_2}$, initial strategies $\boldsymbol{x}_1^{(1)} \in \Delta_{m_1}$, $\boldsymbol{x}_2^{(1)} \in \Delta_{m_2}$, regularizers $\mathcal{R}_1 : \Delta_{m_1} \to \mathbb{R}$, $\mathcal{R}_2 : \Delta_{m_2} \to \mathbb{R}$, learning-rate sequence $\{\eta^{(t)}\}_{t \geq 1}$

Output: Strategy sequence $\{(\boldsymbol{x}_1^{(t)}, \boldsymbol{x}_2^{(t)})\}_{t \geq 1}$

for $t = 1, 2, \ldots, T-1$ **do**

  $\boldsymbol{x}_1^{(t+1)} \leftarrow \arg\max_{\boldsymbol{x}_1 \in \Delta_{m_1}} \left\{ \left\langle \boldsymbol{x}_1, \sum_{\tau=1}^{t} \boldsymbol{u}_1^{(\tau)} \right\rangle - \frac{1}{\eta^{(t)}} \mathcal{R}_1(\boldsymbol{x}_1) \right\}$.

  $\boldsymbol{x}_2^{(t+1)} \leftarrow \arg\max_{\boldsymbol{x}_2 \in \Delta_{m_2}} \left\{ \left\langle \boldsymbol{x}_2, \sum_{\tau=1}^{t} \boldsymbol{u}_2^{(\tau)} \right\rangle - \frac{1}{\eta^{(t)}} \mathcal{R}_2(\boldsymbol{x}_2) \right\}$.

end

We provide new upper and lower bounds for the convergence of FTRL in potential games and constrained optimization.

Our first positive result goes beyond FTRL and covers the entire class of no-regret dynamics. We establish a non-asymptotic rate of convergence to Nash equilibria, provided the updates are alternating and *lazy*—the player performs an update only when the proposed strategy guarantees a sufficient improvement.

Theorem 5.2.1. *In any potential game, alternating ϵ-lazy no-regret dynamics converge to an ϵ-Nash equilibrium after at most* $\exp(O_\epsilon(1/\epsilon^2))$ *iterations.*

The proof proceeds by bounding the number of updates that can occur and then deriving a recursion for the time required between successive updates, paving the way to Theorem 5.2.1. To put this into context, it is well-known that no-regret dynamics can fail to converge to Nash equilibria in potential games (Kleinberg et al., 2009); Theorem 5.2.1 offers a natural way to circumvent those impossibility results.

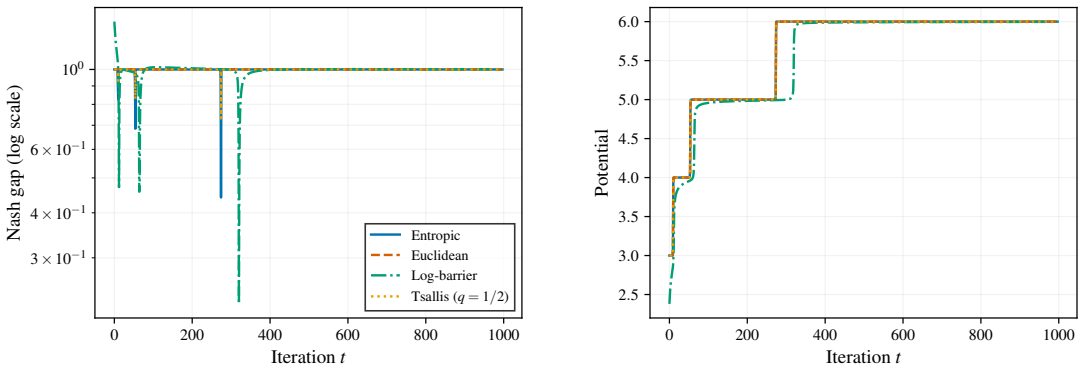


Figure 5.1: Illustration of our results: FTRL algorithms have the potential property (right), but can take exponential time to converge (left).

Turning to the special case of FTRL, we show that, for sufficiently small learning rate, the potential value is nondecreasing (Figure 5.1, right), and this does not require alternation nor laziness (Proposition 5.4.1). As a result, FTRL is bound to converge to a set in which the value of the potential is constant.

While Theorem 5.2.1 only establishes an exponential upper bound in $1/\epsilon^2$, our next result justifies this by establishing a fundamental barrier for FTRL dynamics, which manifests itself with or without alternation or laziness.

Theorem 5.2.2. *In two-player $m \times m$ potential games, FTRL can take $2^{\Omega(m \log m)}$ iterations to converge to an ϵ-Nash equilibrium for $\epsilon = 1/\operatorname{poly}(m)$. This holds for any permutation-invariant regularizer and learning rate $\eta^{(t)} = 1/t^\alpha$ for $\alpha \in [0, 1)$.*

Crucially, this theorem applies to *any* member of FTRL with a permutation invariant regularizer, which encompasses most common incarnations of FTRL (Figure 5.1, left). By virtue of known equivalences, Theorem 5.2.2 thereby circumscribes specific MD algorithms as well, most notably MWU. Furthermore, Theorem 5.2.2 holds even under a vanishing learning rate $\eta^{(t)} = 1/t^\alpha$ for $\alpha \in [0, 1)$. The analytical challenge in this regime stems from the propensity of FTRL to mix actions, especially when the learning rate is close to zero. Another notable

aspect of our lower bound construction is that it tightly mirrors the mechanism driving our upper bound.

Finally, we turn our attention to n-player potential games. We analyze *fictitious play (FP)* (Robinson, 1951a; Brown, 1951), the foundational learning algorithm in games, which can be viewed as the extreme version of FTRL when $\eta \to \infty$. We show that FP can take *doubly* exponentially many iterations to converge to a Nash equilibrium.

Theorem 5.2.3. *There is an n-player binary-action potential game in which FP takes $2^{\Omega(2^n)}$ iterations to reach an ϵ-Nash equilibrium for $\epsilon = 1/2^n$.*

This constitutes an exponentially stronger lower bound relative to prior work Brandt et al. (2013) or the result presented in Chapter 4. From a technical standpoint, our multi-player construction distills the essential property behind the two-player lower bound by making a connection to a classic graph-theoretic problem connected to coding theory (Section 5.6).

5.2.2 Related work

Learning in potential games There is a rich literature on the convergence of learning dynamics in potential games. Unlike general games, potential games always admit a *pure* Nash equilibrium, and better-response dynamics converge to one (Rosenthal, 1973; Monderer and Shapley, 1996a). Fictitious play is also known to converge in this setting, though it was recently shown that it may require exponential time to do so Panageas et al. (2023). Our lower bound in two-player game adapts their construction; unlike fictitious play, FTRL has, by definition, a propensity to mix, which significantly complicates the analysis. The construction of Panageas et al. (2023) was also recently used by Anagnostides et al. (2025) to show that *regret matching* (Hart and Mas-Colell, 2000) can also require exponential time to converge.

Beyond these classic dynamics, significant effort has been devoted to analyzing other learning algorithms in potential games (Héliou et al., 2017; Palaiopanos et al., 2017; Anagnostides et al., 2022b; Hart and Mas-Colell, 2003; Marden et al., 2007; Candogan et al., 2013; Maheshwari et al., 2025). From a complexity standpoint, computing Nash equilibria in potential games is likely intractable when the precision is exponentially small (Babichenko and Rubinstein, 2021, 2020), but amenable to algorithms such as gradient descent when the precision is inverse polynomial. Our results reveal that `FTRL` algorithms are poor first-order optimizers even in the inverse polynomial regime.

From a broader vantage point, the convergence of `FTRL` in games is a topic of active research (Lotidis et al., 2025a,b).

Fictitious play Following the pioneering work of Robinson (1951a), the convergence of `FP` has received extensive attention in two-player zero-sum games. Daskalakis and Pan (2014) proved an exponential lower bound, albeit under adversarial tie breaking. Characterizing its convergence beyond adversarial tie breaks remains open, although Wang (2025) recently made progress. Our construction holds regardless of how ties are broken.

Forgetfulness and convergence speed Our finding that `FTRL` is exponentially slower than `MD` in potential games mirrors a recent result by Cai et al. (2024), albeit in a fundamentally different problem and class of algorithms. Specifically, their result pertains to (two-player) zero-sum games and *optimistic* algorithms.

5.3 Preliminaries

Our analysis centers on the `FTRL` algorithm, which was reviewed in Section 2.7. For completeness, we briefly recall the parts most relevant to our setting and introduce several properties that will be useful in the sequel. We do so in Section 5.3.1. We also provide some background on potential games in Section 5.3.2.

5.3.1 Online learning and FTRL

We examine the convergence of *follow the regularized leader (FTRL)*. This is a classic online algorithm introduced by Kalai and Vempala (2005), which became a mainstay in online optimization (Shalev-Shwartz, 2012). Taking a step back, in the usual online learning framework, a learner interacts with an environment over a sequence of T rounds. In each round $t \in [T]$, the learner selects a strategy $\boldsymbol{x}^{(t)} \in \mathcal{X}$, where $\mathcal{X}$ is a convex and compact set such as the probability simplex. The environment then specifies a utility vector $\boldsymbol{u}^{(t)}$, so that the utility of the learner in round t reads $\langle \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \rangle$. FTRL is a specific update rule that maps the history of observed utilities to the next strategy of the learner. Specifically, upon observing a sequence of utilities $(\boldsymbol{u}^{(\tau)})_{\tau=1}^{t}$, it computes

$$\boldsymbol{x}^{(t+1)} := \arg\max_{\boldsymbol{x} \in \mathcal{X}} \left\{ \left\langle \boldsymbol{x}, \sum_{\tau=1}^{t} \boldsymbol{u}^{(\tau)} \right\rangle - \frac{1}{\eta^{(t)}} \mathcal{R}(\boldsymbol{x}) \right\}. \tag{5.1}$$

Here, $\mathcal{R}$ is a 1-strongly convex and continuously differentiable[1] regularizer with respect to some norm $\|\cdot\|$: $\mathcal{R}(\boldsymbol{x}) \geq \mathcal{R}(\boldsymbol{x}') + \langle \nabla \mathcal{R}(\boldsymbol{x}'), \boldsymbol{x} - \boldsymbol{x}' \rangle + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}'\|^2$ for any $\boldsymbol{x}, \boldsymbol{x}'$; and $\eta^{(t)} > 0$ is the (nonincreasing) *learning rate* sequence. When $\eta^{(t)} = +\infty$, FTRL approaches *fictitious play* (Robinson, 1951a)—also known as *follow the leader* in the online learning literature—formally defined in the sequel. We will denote by R the *range* of the regularizer $\mathcal{R}$, that is, $|\mathcal{R}(\boldsymbol{x}) - \mathcal{R}(\boldsymbol{x}')| \leq R$ for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$. Also, $\|\cdot\|_*$ denotes the dual norm of $\|\cdot\|$; that is, $\|\boldsymbol{u}\|_* = \sup_{\|\boldsymbol{x}\| \leq 1} \langle \boldsymbol{x}, \boldsymbol{u} \rangle$.

A classic result due to Kalai and Vempala (2005) shows that, unlike fictitious play, FTRL has the *no-regret* property even against an adversarially produced sequence of utilities. Regret (see (2.1) in Section 2.6) is the most common way of measuring performance in an online

[1] Differentiability is assumed over some open set $\tilde{\mathcal{X}} \supseteq \mathcal{X}$.

learning environment. It is defined as

$$\texttt{Reg}^{(T)} = \max_{\boldsymbol{x}' \in \mathcal{X}} \sum_{t=1}^{T} \langle \boldsymbol{x}' - \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \rangle. \tag{5.2}$$

We are now ready to formally state the regret guarantee of FTRL, as defined in (5.1).

Proposition 5.3.1 (Kalai and Vempala, 2005; Shalev-Shwartz, 2012)**.** *For any sequence of utilities* $(\boldsymbol{u}^{(t)})_{t=1}^{T}$*, the regret of* FTRL *can be bounded as*

$$\textit{Reg}^{(T)} \leq \frac{R}{\eta^{(T)}} + \sum_{t=1}^{T} \eta^{(t)} \|\boldsymbol{u}^{(t)}\|_*^2$$

for any nonincreasing learning rate sequence $\eta^{(t)}$*. In particular, if* $\|\boldsymbol{u}^{(t)}\|_* \leq B$ *for all* t *and* $\eta^{(t)} = {}^{C(m,B)}/_{t^\alpha}$ *for some* $\alpha \in (0,1)$*,*

$$\textit{Reg}^{(T)} \leq \frac{RT^\alpha}{C(m,B)} + B^2 C(m,B) \left(1 + \frac{T^{1-\alpha}}{1-\alpha}\right).$$

Above, we used the fact that $\sum_{t=1}^{T} {}^{1}/_{t^\alpha} \leq 1 + {}^{T^{1-\alpha}}/_{(1-\alpha)}$. The key takeaway of Proposition 5.3.1 is that $\texttt{Reg}^{(T)} = o(T)$ for any $\alpha \in (0,1)$. We proceed to review some common instantiations of FTRL, most notably MWU.

Example 5.3.1 (MWU)**.** A canonical member of FTRL is MWU. It is the instantiation of FTRL on the probability simplex $\mathcal{X} = \Delta(\mathcal{A})$ when $\mathcal{R}$ is the (negative) entropy regularizer, $\mathcal{R}(\boldsymbol{x}) = \sum_{a \in \mathcal{A}} \boldsymbol{x}[a] \log \boldsymbol{x}[a]$, which can be shown to be 1-strongly convex with respect to $\|\cdot\|_1$. In that case, FTRL admits the closed-form solution

$$\boldsymbol{x}^{(t+1)}[a] = \frac{e^{\eta^{(t)} \sum_{\tau=1}^{t} \boldsymbol{u}^{(\tau)}[a]}}{\sum_{a' \in \mathcal{A}} e^{\eta^{(t)} \sum_{\tau=1}^{t} \boldsymbol{u}^{(\tau)}[a']}} \quad \forall a \in \mathcal{A}. \tag{MWU}$$

Algorithm 6. Multiplicative Weights Update (MWU)

Input: Initial strategies $\boldsymbol{x}_1^{(1)} \in \Delta(\mathcal{A}_1)$, $\boldsymbol{x}_2^{(1)} \in \Delta(\mathcal{A}_2)$, step size $\eta > 0$
Output: Strategy sequence $\{(\boldsymbol{x}_1^{(t)}, \boldsymbol{x}_2^{(t)})\}_{t \geq 1}$

for $t = 1, 2, \ldots, T-1$ **do**
  Compute current payoff vectors $\boldsymbol{u}_1^{(t)}$, $\boldsymbol{u}_2^{(t)}$.
  for $a_1 \in \mathcal{A}_1$ **do**
    $$\boldsymbol{x}_1^{(t+1)}[a_1] \leftarrow \frac{\boldsymbol{x}_1^{(t)}[a_1] \exp\left(\eta \boldsymbol{u}_1^{(t)}[a_1]\right)}{\sum_{a_1' \in \mathcal{A}_1} \boldsymbol{x}_1^{(t)}[a_1'] \exp\left(\eta \boldsymbol{u}_1^{(t)}[a_1']\right)}.$$
  end
  for $a_2 \in \mathcal{A}_2$ **do**
    $$\boldsymbol{x}_2^{(t+1)}[a_2] \leftarrow \frac{\boldsymbol{x}_2^{(t)}[a_2] \exp\left(\eta \boldsymbol{u}_2^{(t)}[a_2]\right)}{\sum_{a_2' \in \mathcal{A}_2} \boldsymbol{x}_2^{(t)}[a_2'] \exp\left(\eta \boldsymbol{u}_2^{(t)}[a_2']\right)}.$$
  end
end

When the learning rate is a constant, this can be equivalently written as

$$\boldsymbol{x}^{(t+1)}[a] = \boldsymbol{x}^{(t)}[a] \frac{e^{\eta \boldsymbol{u}^{(t)}[a]}}{\sum_{a' \in \mathcal{A}} \boldsymbol{x}^{(t)}[a'] e^{\eta \boldsymbol{u}^{(t)}[a']}} \quad \forall a \in \mathcal{A}.$$

Since the range of the entropy regularizer is $\log |\mathcal{A}|$, Proposition 5.3.1 implies that the regret of MWU is bounded by $\sqrt{T \log |\mathcal{A}|}$ when $\|\boldsymbol{u}^{(t)}\|_\infty \leq 1$ for all t.

Example 5.3.2 (Euclidean regularization)**.** Another canonical member of the FTRL family arises when $\mathcal{R} : \boldsymbol{x} \mapsto \frac{1}{2}\|\boldsymbol{x}\|_2^2$. In that case, the update rule can be cast as

$$\boldsymbol{x}^{(t+1)} = \Pi_{\mathcal{X}} \left(\eta \sum_{\tau=1}^{t} \boldsymbol{u}^{(\tau)} \right),$$

where $\Pi_{\mathcal{X}}(\cdot)$ denotes the Euclidean projection onto $\mathcal{X}$.

Two other notable regularizers are i) the *logarithmic regularizer* $\mathcal{R} : \boldsymbol{x} \mapsto -\sum_{a \in \mathcal{A}} \log \boldsymbol{x}[a]$, and ii) *Tsallis entropy* $\mathcal{R} : \boldsymbol{x} \mapsto -\frac{1}{q(1-q)} \sum_{a \in \mathcal{A}} (\boldsymbol{x}[a])^q$ for $q \in (0, 1)$, both of which have found

many applications in, among others, learning with bandit feedback (Zimmert and Seldin, 2021; Abernethy et al., 2008; Wei and Luo, 2018). All of these regularizers are permutation invariant (Definition A.1.1).

A regularizer $\mathcal{R}$ is called a *Legendre regularizer* if, in addition to the previous properties, for any sequence of points $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots$ converging to a boundary point of $\mathcal{X}$, $\nabla\mathcal{R}(\boldsymbol{x}^{(T)}) \to \infty$ as $T \to \infty$ (Cesa-Bianchi and Lugosi, 2006). A well-known fact is that `FTRL` and `MD` are equivalent under any Legendre regularizer; for example, we refer to McMahan (2011, 2017). This means that all our results concerning `FTRL`—both upper and lower bounds—immediately translate to certain `MD` variants as well.

Fictitious play Fictitious play (Robinson, 1951a; Brown, 1951), also known as *follow the leader* in the online learning literature, can be viewed as the extreme case of `FTRL` with $\eta \to \infty$. Specifically,

$$\boldsymbol{x}^{(t+1)} := \arg\max_{\boldsymbol{x}\in\mathcal{X}} \left\{ \left\langle \boldsymbol{x}, \sum_{\tau=1}^{t} \boldsymbol{u}^{(\tau)} \right\rangle \right\}.$$

When $\mathcal{X} = \Delta(\mathcal{A})$, it is assumed that fictitious play selects a pure strategy.

5.3.2 Games and solution concepts

To be clear with respect to the notation used in the previous cchapters we restate the notation for a game of n-players In an n-player normal-form game, each player $i \in [n]$ selects as strategy a probability distribution $\boldsymbol{x}_i \in \Delta(\mathcal{A}_i) =: \mathcal{X}_i$ from a finite set of actions $\mathcal{A}_i$. We denote by $u_i(\boldsymbol{x})$ the expected utility of player $i \in [n]$ under the joint strategy $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$. We also use the notation $\boldsymbol{x}_{-i} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \dots, \boldsymbol{x}_n)$. The most standard solution concept in games is the *Nash equilibrium*, recalled below.

Definition 5.3.1 (Nash, 1950)**.** A joint strategy $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is an ϵ-*Nash equilibrium* if for any player $i \in [n]$ and deviation $\boldsymbol{x}_i' \in \mathcal{X}_i$,

$$u_i(\boldsymbol{x}_i', \boldsymbol{x}_{-i}) \leq u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}) + \epsilon.$$

As we alluded to in our introduction, there is an intimate connection between the no-regret property and a game-theoretic solution concept known as *coarse correlated equilibrium (CCE)* (Aumann, 1974; Moulin and Vial, 1978). CCEs relax Definition 5.3.1 by allowing for a *correlated* distribution.

Definition 5.3.2. A distribution $\mu \in \Delta(\mathcal{A}_1 \times \cdots \times \mathcal{A}_n)$ is an ϵ-CCE if for any player $i \in [n]$ and deviation $\boldsymbol{x}_i' \in \mathcal{X}_i$,

$$\mathbb{E}_{\boldsymbol{a} \sim \mu}[u_i(\boldsymbol{x}_i', \boldsymbol{a}_{-i})] \leq \mathbb{E}_{\boldsymbol{a} \sim \mu}[u_i(\boldsymbol{a})] + \epsilon.$$

By virtue of Proposition 5.3.1, it follows that `FTRL` dynamics converge to the set of CCEs. Specifically, if the cumulative regret of each player is bounded as $O_T(\sqrt{T})$, convergence to the set of CCEs occurs at a rate of $O_\epsilon(1/\epsilon^2)$.

Potential games A game is called a *potential game* if there exists a function $\Phi : \mathcal{X} \to \mathbb{R}$ such that for any player i, $\boldsymbol{x}_{-i}$, and $\boldsymbol{x}_i, \boldsymbol{x}_i' \in \Delta(\mathcal{A}_i)$,

$$u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}_i', \boldsymbol{x}_{-i}) = \Phi(\boldsymbol{x}_i, \boldsymbol{x}_{-i}) - \Phi(\boldsymbol{x}_i', \boldsymbol{x}_{-i}).$$

This implies that the utility gradient of each player can be expressed as $\boldsymbol{u}_i(\boldsymbol{x}) = \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x})$. In a normal-form potential game, the potential function is multilinear, so it is smooth, but not necessarily concave. Our upper bounds apply more broadly beyond normal-form games to constrained optimization whenever the potential is *L-smooth*: $\|\nabla \Phi(\boldsymbol{x}) - \nabla \Phi(\boldsymbol{x}')\|_* \leq$

$L\|\boldsymbol{x} - \boldsymbol{x}'\|$ for any $\boldsymbol{x}, \boldsymbol{x}'$, where $\|\cdot\|_*$ is the dual norm. For convenience, we define L-smoothness with respect to the global norm $\|\boldsymbol{x}\| \coloneqq \sqrt{\sum_{i=1}^n \|\boldsymbol{x}_i\|_{(i)}^2}$, where $\|\cdot\|_{(i)}$ is the norm induced by the regularizer $\mathcal{R}_i$ employed by player i. To ease the notation, we will assume that all players employ the same regularizer, although our results apply more broadly.

Further notation We denote by D_i the diameter of $\mathcal{X}_i$ with respect to some norm $\|\cdot\|$; the choice of norm will be clear from the context. We use B for an upper bound on the norm of utilities in terms of the dual norm; that is, $\|\boldsymbol{u}_i(\boldsymbol{x})\|_* \leq B$ for any $i \in [n]$ and joint strategy $\boldsymbol{x}$.

5.4 Exponential upper bound and the potential property for FTRL

We begin by establishing exponential upper bounds for `FTRL` in potential games. Our first goal is to show that when multiple players perform an `FTRL` update simultaneously, the value of the potential function cannot decrease. The following lemma takes the perspective of a single player, and shows that `FTRL` always yields an improvement relative to its current utility.

Lemma 5.4.1. *If the sequence $(\boldsymbol{x}_i^{(t)})_{t\geq 1}$ is updated using `FTRL` under a 1-strongly convex regularizer $\mathcal{R}$ with respect to some norm $\|\cdot\|$,*

$$\langle \boldsymbol{x}_i^{(t+1)}, \boldsymbol{u}_i^{(t)} \rangle - \langle \boldsymbol{x}_i^{(t)}, \boldsymbol{u}_i^{(t)} \rangle \geq \frac{1}{\eta} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|^2.$$

In particular, the improvement is proportional to the squared movement $\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|$. The proof of Lemma 5.4.1 relies on the first-order optimality conditions of the `FTRL` update. Together with all the other proofs of this section, it is deferred to Section A.2.

Lemma 5.4.1 already implies that FTRL can never decrease the value of the potential function under *alternating updates*, which means that players perform the update one by one; in fact, this holds no matter the choice of the learning rate.

We will now argue about the simultaneous case. Since Φ is L-smooth, we can lower bound the difference $\Phi(\boldsymbol{x}^{(t+1)}) - \Phi(\boldsymbol{x}^{(t)})$ by

$$\langle \nabla \Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)} \rangle - \frac{L}{2} \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|^2.$$

By Lemma 5.4.1, the first term can in turn be lower bounded by

$$\sum_{i=1}^{n} \langle \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}, \boldsymbol{u}_i^{(t)} \rangle \geq \frac{1}{\eta} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|^2.$$

As a result, when $\eta \leq 1/L$, we have shown that

$$\Phi(\boldsymbol{x}^{(t+1)}) - \Phi(\boldsymbol{x}^{(t)}) \geq \frac{1}{2\eta} \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|^2.$$

The telescopic summation yields the following implication.

Proposition 5.4.1. *Suppose that each player in a potential game employs FTRL with learning rate $\eta \leq 1/L$, where L is the smoothness parameter of the potential. For any $t \geq 1$,*

$$\Phi(\boldsymbol{x}^{(t+1)}) - \Phi(\boldsymbol{x}^{(t)}) \geq \frac{1}{2\eta} \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|^2. \tag{5.3}$$

In particular,

$$\sum_{t=1}^{T} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|^2 \leq 2\eta \Phi_{\mathsf{range}}.$$

The second implication—the fact that FTRL has a bounded second-order path length—follows from (5.3) from a telescopic summation, noting that Φ_{range} denotes the range of the potential function, which is bounded.

An interesting implication of Proposition 5.4.1 is that simultaneous FTRL asymptotically converges to a set in which the value of the potential is the same (Proposition A.2.1). Proposition 5.4.1 also implies that $\lim_{t\to\infty} \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\| \to 0$. Specifically, for any $\epsilon > 0$, $O_\epsilon(1/\epsilon^2)$ iterations suffice so that $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\| \leq \epsilon$. For MD with a smooth regularizer, this in fact implies that $\boldsymbol{x}^{(t+1)}$ is an $O_\epsilon(\epsilon)$-Nash equilibrium (Anagnostides et al., 2022b); however, this is not the case for FTRL, no matter the choice of the regularizer. In other words, FTRL can have *spurious fixed points*.

Example 5.4.1 (Spurious fixed points of FTRL)**.** Consider a sequence of utilities $\boldsymbol{u}_i^{(\tau)} = (1, 0)$ for all $\tau \in [t]$. Under this sequence, FTRL is producing a strategy that plays the first action deterministically, up to an error proportional to $1/t$. Now, if $\boldsymbol{u}_i^{(t+1)} = (0, 1)$, it holds that $\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\| \leq O_t(1/t)$, and consequently $\boldsymbol{x}_i^{(t+1)}$ is highly suboptimal relative to $\boldsymbol{u}_i^{(t+1)}$.

This inertia of FTRL is indeed what drives our lower bounds presented later in Section 5.5.

5.4.1 Non-asymptotic convergence of no-regret dynamics

Despite this pathological behavior, we will now establish an upper bound for FTRL. In fact, our result applies to any no-regret dynamics, provided that the updates are alternating and ϵ-lazy. To put this into context, we highlight that no-regret dynamics can fail to converge in potential games (Kleinberg et al., 2009), so our positive results offer a natural way of bypassing that impossibility.

More precisely, let $\mathfrak{R}(\boldsymbol{u}_i^{(1)}, \dots, \boldsymbol{u}_i^{(t)}) \in \mathcal{X}$ denote the output of the regret minimization algorithm upon observing the sequence $(\boldsymbol{u}_i^{(\tau)})_{\tau=1}^t$, which we assume guarantees $\mathtt{Reg}_i^{(t)} \leq O_t(t^{1-\alpha})$ for some $\alpha \in (0, 1]$. If $\tilde{\boldsymbol{x}}_i^{(t+1)} = \mathfrak{R}(\boldsymbol{u}_i^{(1)}, \dots, \boldsymbol{u}_i^{(t)})$ at each time t, the strategy $\boldsymbol{x}_i^{(t+1)}$ in the

ϵ-*lazy* version is defined as

$$\boldsymbol{x}_i^{(t+1)} = \begin{cases} \tilde{\boldsymbol{x}}_i^{(t+1)} & \text{if } \langle \tilde{\boldsymbol{x}}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}, \boldsymbol{u}_i^{(t)} \rangle \geq \epsilon, \\ \boldsymbol{x}_i^{(t)} & \text{otherwise.} \end{cases} \tag{5.4}$$

That is, the update occurs only if it delivers at least an ϵ improvement relative to the current utility. In this context, we prove the following result.

Theorem 5.4.1. *Alternating ϵ-lazy no-regret dynamics converge to a 2ϵ-Nash equilibrium after at most $\exp(O_\epsilon(1/\epsilon^2))$ rounds.*

Our analysis proceeds by bounding the total number of updates that can occur (Lemma A.2.1) and the duration of the intervals between consecutive updates (Lemma A.2.2). Specifically, we derive the recursion $T_{k+1} \leq T_k \left(1 + O_\epsilon\left(\frac{1}{\epsilon}\right)\right)$. What is intriguing is that the mechanism driving our upper bound mirrors the construction of our lower bound.

5.5 Exponential lower bound for FTRL

In this section, we show that `FTRL` can take exponentially many rounds to reach an approximate Nash equilibrium even in a two-player identical-interest games. Our lower bound applies to either alternating or simultaneous updates; in what follows, we consider simultaneous updates for concreteness. The construction is also robust in terms of using laziness per (5.4), establishing the tightness of Theorem 5.4.1 up to factors in the exponent.

In what follows, we provide a high-level overview of the main steps in our argument; the formal proofs are deferred to Section A.3. The class of games we study is based on the construction of Panageas et al. (2023), originally introduced to analyze fictitious play in. We employ a variant of their class of games, formally defined in Section A.3.2. For an odd dimension m, we define a payoff matrix $\mathbf{A}$ whose maximum entry equals $2m - 1$, and we

take the action sets to be $\mathcal{A}_1 = \mathcal{A}_2 = [m]$. For each payoff value $k \in \{3, \dots, 2m-1\}$, we denote by $a_1(k), a_2(k) \in [m]$ the unique row and column indices, respectively, such that $\mathbf{A}[a_1(k), a_2(k)] = k$. An illustration of $\mathbf{A}$ is given in Figure 5.2b. The role of the extra row and column compared to $\mathbf{B}$ concerns the initialization; as we shall see, FTRL initialized uniformly at random on $\mathbf{A}$ will end up at the beginning of the spiral in the $\mathbf{B}$ submatrix.[2]

To examine the behavior of FTRL in this class of games, it is essential to keep track of the *cumulative utility gap* of each action, defined as

$$\mathtt{Gap}_i^{(t)}[a_i] \coloneqq \mathtt{Gap}_i^{(t)}(\boldsymbol{e}_{a_i}) = \max_{a_i' \in \mathcal{A}_i} \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i'] - \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i] \geq 0$$

for each player $i \in \{1, 2\}$. An important observation is that for FTRL algorithms, an action with large gap will always be played with small probability, as we establish in the lemma below.

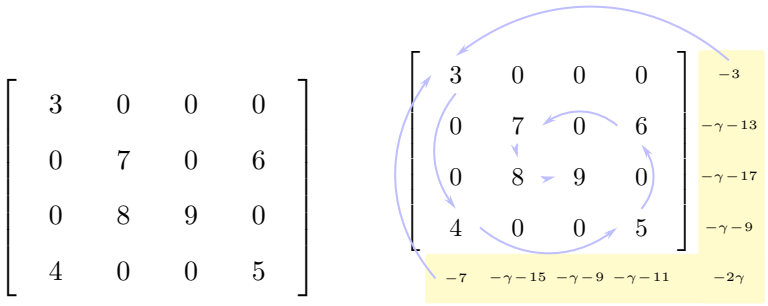
Lemma 5.5.1. *Suppose that $(\boldsymbol{x}_i^{(t)})_{t \geq 1}$ is updated using FTRL. If $\mathtt{Gap}_i^{(t)}[a_i] \geq \gamma > 0$ for some action $a_i \in \mathcal{A}_i$, then*

$$\boldsymbol{x}_i^{(t+1)}[a_i] \leq \frac{R}{\eta^{(t)}\gamma}, \tag{5.5}$$

where R denotes the range of the regularizer.

An important point about Lemma 5.5.1 is that the bound relating the gap and the probability of playing the corresponding action in (5.5) becomes *weaker* as the learning rate gets closer to 0, which happens at a rate of $1/t^{\alpha}$. In the beginning, $\eta^{(t)}$ is relatively large, so the regularizer has a limited impact. As the learning rate decreases, the propensity of FTRL to mix over actions intensifies, introducing a considerable obstacle in the lower bound construction; it is harder to control the trajectory of FTRL under significant mixing. Our key observation is that

[2]For simplicity in the exposition, we allow the payoffs to be larger than 1. As we explain in Remark A.3.1, it is straightforward to adjust our construction even when all payoffs are in $[-1, 1]$.



(a) Original matrix **B**. (b) New matrix **A**.

the gap for the relevant actions will grow *linearly* in t (Lemma 5.5.2), thereby subsuming the vanishing effect of the learning rate in (5.5).

Periods The key invariance we establish is that the trajectory of `FTRL` can be divided into k *periods*, so that in the kth period `FTRL` places high probability on the profile $(a_1(k), a_2(k))$. Because `FTRL` has a tendency to mix, especially under Legendre regularizers, our analysis needs to account for the residue probability. To do so, we fix the threshold $\delta \coloneqq \frac{1}{4m}$ and define the periods as follows.

Definition 5.5.1 (Transition period)**.** For each integer $k \in \{3, 4, \dots, 2m-1\}$, the *transition period* k is the set of rounds t for which the following conditions hold:

(i) If $k \geq 4$ is even, then for all $t \in [\underline{t_k}, \overline{t_k}]$, $\boldsymbol{x}_2^{(t)}[a_2(k)] \geq 1 - \delta$ and $\boldsymbol{x}_1^{(t)}[a_1(k-1)] + \boldsymbol{x}_1^{(t)}[a_1(k)] \geq 1 - \delta$.

(ii) If $k \geq 5$ is odd, then for all $t \in [\underline{t_k}, \overline{t_k}]$, $\boldsymbol{x}_1^{(t)}[a_1(k)] \geq 1 - \delta$ and $\boldsymbol{x}_2^{(t)}[a_2(k-1)] + \boldsymbol{x}_2^{(t)}[a_2(k)] \geq 1 - \delta$.

For $k \geq 3$, we let $\underline{t_k}$ and $\overline{t_k}$ denote the first and last rounds of period k, respectively. $T_k \coloneqq \overline{t_k} - \underline{t_k} + 1$ denote its length, so that period k corresponds to the interval $t \in [\underline{t_k}, \overline{t_k}]$.

A key property that we establish is that `FTRL` transitions from each period k to its successor period $k+1$.

Property 5.5.1 (Period consistency)**.** Suppose both players employ FTRL. If a round t belongs to period k for some $3 \leq k \leq 2m-1$, then the subsequent round $t+1$ belongs either to the same period k or to the next period $k+1$.

In other words, there are no shortcuts: FTRL needs to traverse the entire path to reach the maximum payoff.

To argue about Property 5.5.1, let us say that k is even. We need to analyze the transition of the row player from $a_1(k-1)$ to $a_1(k)$. When the row player starts playing $a_1(k)$ with higher and higher probability, it triggers a change in the utility observed by the column player. What we need to show is that the transition from $a_1(k-1)$ to $a_1(k)$ will be faster than the time it takes for the column player to react.

This is indeed the case for the following reason. The transition from $a_1(k-1)$ to $a_1(k)$ takes time proportional to $1/\eta^{(t)}$, which is of the order of t^{α}. On the other hand, the time it takes for the column player to react is $\Omega(t)$. This is a high-level simplification; the precise argument is given in Section A.3.6.

Growth of the cumulative utility gaps Having established this key invariance, we show that the gap of actions corresponding to future periods grows considerably.

Lemma 5.5.2. *Let $k \geq 6$ be even. Then*

$$\mathit{Gap}_1^{(\overline{t_{k-2}})}[a_1(k)] \;\geq\; \sum_{\ell=4}^{k-2} \Big[(1-\delta)(\ell-1) - \delta(2m-1)\Big]\, T_\ell.$$

Similarly, if $k \geq 5$ is odd, then

$$\mathit{Gap}_2^{(\overline{t_{k-2}})}[a_2(k)] \;\geq\; \sum_{\ell=4}^{k-2} \Big[(1-\delta)(\ell-1) - \delta(2m-1)\Big]\, T_\ell.$$

When δ is small enough, we use Lemma 5.5.2 to derive the following recursion.

Lemma 5.5.3 (Recurrence relation of $T_k + T_{k-1}$)**.** *For a small enough $\delta > 0$,*

$$T_k + T_{k-1} \;\geq\; \frac{1}{2}\sum_{\ell=4}^{k-2}(\ell - 2)\, T_\ell$$

for all $k \geq 6$.

The only way for FTRL to have a small Nash equilibrium gap under the invariance established in Property 5.5.1 is for it to reach the last period (Lemma A.3.10). However, on account of Lemma 5.5.3, this takes $2^{\Omega(m \log m)}$ rounds, establishing our lower bound, as formalized in Theorem A.3.1.

Basis of the induction The validity of our previous argument rests on t being large enough, so that the linear gap $\Omega(t)$ subsumes factors that grow as t^α. To complete the proof, we establish the basis of the induction. Specifically, we construct a suitable gadget that forces FTRL to i) immediately transition at the beginning of the spiral and ii) spend a considerable amount of time in the initial period. This can be achieved by appending an additional row and column to the matrix with sufficiently negative entries (Figure 5.2b). When FTRL is initialized uniformly at random, which is the optimal choice for permutation-invariant regularizers (Lemma A.1.1), all actions except the first will experience large regret. This means that escaping from the initial period will require many rounds so as to overcome the negative regret. Setting the negative entries in the extra row and column appropriately guarantees the desired bound on the length of the initial period, as we formalize in Section A.3.

5.6 Doubly exponential lower bound for fictitious play

We now turn to multi-player potential games. We will prove an exponentially stronger lower bound for fictitious play (FP).

Our construction is based on a connection we make to a graph-theoretic problem known as *snake in the box*. This was first introduced by Kautz (1958) in the context of coding theory, and has many interesting applications (Klee, 1970, 1967). The basic version of the problem is to identify a path along the edges of a high-dimensional hypercube with the following property. The path begins at some vertex of the hypercube and traverses the edges—two vertices are connected if they differ by a single bit—to as many vertices as it can reach, subject to the constraint that every time it arrives at a new vertex, the previous one and *all of its neighbors can never be used* going forward. Such a path is called a *snake*. Figure 5.3 (left) portrays an example for the 4-dimensional hypercube.

This problem can be defined for any graph. In our context, we associate each joint action with a vertex in the graph, and two vertices are connected with an edge if there is a unilateral deviation that goes from one to the other. Under this mapping, Figure 5.3 (right) shows a snake in a 3-player game where each player has 4 actions. Our key observation is that what essentially drove our lower bound in two-player games is the defining property of a snake. Indeed, the payoff matrix used in our two-player lower bound (Figure 5.2a) induces a snake that spirals toward the center.

For our purposes, it suffices to restrict our attention to an n-dimensional hypercube. A well-known fact is that there exists a snake with exponential length (Abbott and Katchalski, 1988; Evdokimov, 1969), although characterizing the precise length is a major open problem in graph theory.

Lemma 5.6.1 (Evdokimov, 1969)**.** *For any $n \in \mathbb{N}$, there exists a snake on the n-dimensional hypercube with length $C \cdot 2^n$, for some absolute constant $C > 0$.*

Let P be such a snake in the n-dimensional hypercube. We proceed by embedding this path into a potential game. Specifically, we construct a binary-action n-player game with identical interests as follows. A joint action $(a_1, \ldots, a_n)$ is in one-to-one correspondence with vertices

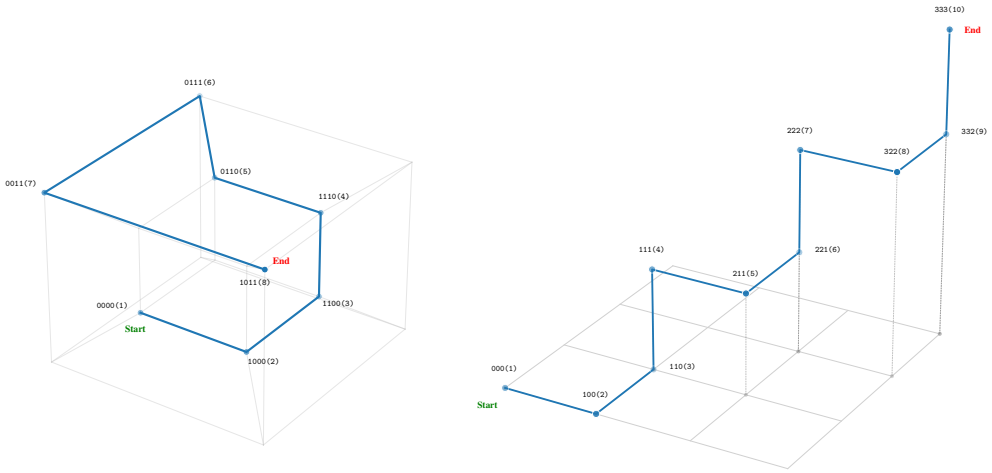


Figure 5.3: A snake on a 4-dimensional hypercube (left) and a snake corresponding to a 3-player 4-action game (right).

of the hypercube. For a joint action $\boldsymbol{a} = (a_1, \ldots, a_n) \in P$, we let $p(\boldsymbol{a})$ be the position of $\boldsymbol{a}$ on the path P, starting from 1. We define the utility, for each player $i \in [n]$, as

$$u_i(\boldsymbol{a}) = \begin{cases} 0 & \text{if } \boldsymbol{a} \notin P, \\ p(\boldsymbol{a}) & \text{otherwise.} \end{cases} \tag{5.6}$$

We assume that `FP` is initialized at the start of the path P. By definition, it will always transition to pure strategies. Furthermore, because of the structure of the game, a basic invariance is that it will always remain on the path. Now, by construction, in every joint action on the path, there is a player who can improve the utility by at least 1 through a unilateral deviation, unless the joint action corresponds to the last vertex in the path, which is the global maximum of the utility function.

The crucial role of the snake property is this: at all iterations before reaching the end of the path, *exactly* one player will have positive best-response gap. This is so because, by definition, all subsequent vertices except the immediate successor cannot be reached through

a unilateral deviation. In other words, there are no shortcuts, and each edge traversal is caused by the movement of a single player.

If T_k is the number of iterations that FP spent on the action profile corresponding to payoff k, we establish the following recursion (the proof is in Section A.4).

Lemma 5.6.2. *For any $k \geq 2$, $T_k \geq (k-1)T_{k-1} + T_{k-2} + T_{k-3} - \sum_{\kappa=1}^{k-4} \kappa T_\kappa$. ($T_k$ for $k \leq 0$ is to be interpreted as 0.)*

We draw attention to the fact that except the first three leading terms, the rest are negative. This is unlike the basic recursion that can be set up in two-player games, and happens because a player continually switches from one action to the other; in two-player games, a player always switches to a new action, which has accumulated negative regret throughout.

We next argue that, despite the negative terms, $T_k \geq (k-1)!$ for $k \geq 1$.

Lemma 5.6.3. *Any sequence satisfying the recurrence relation of Lemma 5.6.2 grows at least as $(k-1)!$ for $k \geq 1$.*

Combining, we arrive at the following *doubly* exponential lower bound for FP in potential games.

Theorem 5.6.1. *FP requires at least $(C2^n)!$ to converge to a Nash equilibrium of an n-player potential game, where $C > 0$ is some absolute constant.*

This holds even for converging to a ϵ-Nash equilibrium with $\epsilon < 1$, although the payoffs in our construction can be exponentially large. By rescaling—FP is scale invariant—it follows that doubly exponentially many iterations are needed to converge to an ϵ-Nash equilibrium with $\epsilon \approx 1/2^n$. This closely matches our upper bound in Theorem 5.4.1.

Theorem 5.6.1 improves exponentially over existing lower bounds for fictitious play in potential games (Panageas et al., 2023; Brandt et al., 2013).

5.7 Concluding remarks

This chapter shows that the inefficiency phenomenon identified for fictitious play is not limited to that particular dynamic. Follow-the-regularized-leader, despite its central role in online learning and its no-regret guarantees, can also require exponentially many iterations to reach an approximate Nash equilibrium in two-player potential games.

The results also clarify the gap between no-regret learning and last-iterate equilibrium convergence. No-regret guarantees imply convergence of empirical play to coarse correlated equilibrium, but they do not ensure that the actual strategy profile produced by the dynamics approaches Nash equilibrium efficiently. In potential games, the existence of a global potential function provides useful monotonicity structure, yet this structure is insufficient to rule out exponentially long transient behavior.

Together with Chapter 4, these lower bounds indicate that classical learning dynamics may be computationally inefficient even in highly structured games. This motivates the broader conclusion of the dissertation: efficient equilibrium learning requires more than asymptotic convergence, and the relevant structural assumptions must support explicit quantitative rates.

Chapter 6

Follow the Regularized Leader Does Not Converge in Constrained Optimization

6.1 Preface

This chapter is based on a joint work with Anagnostides et al. (2026b), with the exposition adapted from the corresponding paper for inclusion in this dissertation. In particular, the notation has been aligned with the rest of the thesis, and some of the arguments have been reorganized to emphasize the connection with the preceding chapters.

Follow-the-regularized-leader (`FTRL`) is a foundational algorithm in online learning, whose regret guarantees have been studied extensively. However, in contrast to mirror descent—and gradient descent in particular—its convergence to first-order stationary points in constrained nonconvex optimization has remained poorly understood. In this chapter, we show that

continuous-time FTRL can fail to converge even asymptotically in constrained optimization: the Karush–Kuhn–Tucker gap can remain bounded away from zero for all time.

This phenomenon is surprising because continuous-time FTRL still guarantees monotonic improvement of the objective value. Thus, monotonicity of the objective alone is insufficient to ensure convergence to stationarity. In particular, our result shows that the nonconvergent behavior of no-regret dynamics, familiar from general variational inequality problems, can also arise in potential systems.

The construction proceeds in two steps. First, we build an infinitely differentiable objective whose gradient-flow dynamics exhibit decelerating oscillations and do not converge pointwise. We then embed this behavior into a higher-dimensional constrained optimization problem whose FTRL dynamics inherit the oscillatory behavior while maintaining a uniformly positive KKT gap. Consequently, the dynamics continue to improve the objective, but never approach the set of first-order stationary points.

6.2 Introduction

Follow the regularized leader (FTRL) is a premier online optimization algorithm that has shaped modern machine learning (Kalai and Vempala, 2005; Shalev-Shwartz, 2012; Hazan, 2016). The regret properties of FTRL—and variations thereof such as follow the perturbed leader (Kalai and Vempala, 2005)—have been a vibrant area of research in the last two decades, and are by now well understood. On top of establishing minimax optimal regret bounds, FTRL provides a broad, unifying framework, encompassing celebrated algorithms such as multiplicative weights update (Littlestone and Warmuth, 1994; Arora et al., 2012). As such, FTRL has driven empirical breakthroughs in complex domains, serving as the engine for large-scale click-through rate (CTR) prediction (McMahan et al., 2013), and being at

the heart of AI milestones in solving zero-sum games (Moravčík et al., 2017; Brown and Sandholm, 2017, 2019; Farina et al., 2021; Zhang and Sandholm, 2026).

However, despite the tremendous amount of interest FTRL has received, its convergence properties as a first-order optimization algorithm have remained poorly understood. This is in sharp contrast to online mirror descent (MD) (Nemirovskij and Yudin, 1983), which is another seminal class of algorithms in online learning. In particular, different incarnations of MD algorithms are known to converge to first-order stationary points in time polynomial in $1/\epsilon$, where ϵ is the suboptimality gap. For example, the convergence of gradient descent—which can be viewed as MD endowed with Euclidean regularization—even in nonconvex problems is a cornerstone in optimization theory (Nesterov, 2004). This raises the following fundamental question:

Does FTRL converge to first-order stationary points in constrained optimization?

For unconstrained problems, there is a well-known equivalence between FTRL and MD, which holds more generally even in the constrained setting under suitable assumptions on the regularizer (*e.g.* McMahan, 2011). This implies that FTRL inherits the convergence of MD in such settings. However, the general question beyond such special cases has remained wide open despite decades of research on the behavior of FTRL. Recent work by Anagnostides et al. (2026a) established an exponential lower bound for FTRL in *potential games*, which can be viewed as a special case of constrained (nonconvex) optimization problems. Potential games have played a central role in the development of game theory since the pathbreaking work of Rosenthal (1973), and the convergence of FTRL to Nash equilibria—the natural counterpart of first-order stationary points in the case of games—has received considerable attention (we elaborate on related work in Section 6.2.2). While that result exposes severe inefficiency of FTRL, it leaves open the possibility of eventual convergence.

In this chapter, we take a significant step further: we show that FTRL can fail to converge *even asymptotically* in constrained optimization problems, revealing a stark separation between FTRL and MD.

6.2.1 Our results

In more detail, we examine the convergence of continuous-time FTRL in constrained optimization problems. We focus on the continuous-time version of FTRL because it always has the potential property (Lemma 6.3.1), in that it can never decrease the value of the potential function (which is to be maximized). As a result, one would expect that the suboptimality gap of the dynamics would asymptotically be driven to zero; this would indeed be in line with prior results showing that MWU—an incarnation of FTRL—converges to Nash equilibria in (generic) potential games (Kleinberg et al., 2009; Palaiopanos et al., 2017). On the other hand, the discrete-time version of FTRL only has the potential property when the learning rate is sufficiently small (Anagnostides et al., 2026a); indeed, FTRL with large learning rate is known to exhibit a wide spectrum of non-convergent behavior, such as Li-Yorke chaos (Palaiopanos et al., 2017; Bielawski et al., 2021); those results do not apply in our setting because we examine the limiting case where the learning rate tends to zero.

As alluded to, the notion of convergence we consider is measured in terms of the *Karush-Kuhn-Tucker (KKT)* gap (Definition 6.3.1), which is the standard notion of first-order suboptimality in optimization. This is weaker than pointwise convergence; as we shall see, dynamics such as gradient flow always converge in terms of the KKT gap but can still admit non-trivial limit sets.

In this context, our main result is summarized below.

Theorem 6.2.1. *There exists an infinitely differentiable function $\Phi \in C^\infty$ on a convex and compact domain such that for any time t, the KKT gap of the* `FTRL` *dynamics on Φ is at least some constant bounded away from 0. This holds even in constant dimensions.*

This result contributes to the fundamental understanding of the convergence of learning algorithms—and `FTRL` in particular—in nonconvex constrained optimization problems, establishing that `FTRL` is inherently incapable of serving as a first-order optimizer.

Proof overview The starting point of our construction concerns (continuous-time) gradient flow. Specifically, we rely on an instance in which gradient flow does not converge pointwise—but does converge in terms of the KKT gap. The fact that this phenomenon can occur is a well-known fact in the theory of dynamical systems (Palis and De Melo, 2012). In fact, by virtue of the Łojasiewicz inequality (Polyak, 1963), this can only happen under a non-analytic (but infinitely differentiable) function. In more detail, we first show that gradient flow follows the trajectory shown in Figure 6.1, governed by $x_2 = \sin\left(\frac{\pi x_1}{1-x_1}\right)$ (Lemma 6.4.2), thereby approaching the limit set $\{1\} \times [-1, 1]$. As time goes by, the dynamics in this instance move ever so slowly as the KKT gap approaches zero. The stalling of the dynamics notwithstanding, a critical property for the purpose of our construction is that the dynamics do not decelerate too quickly (Lemma 6.4.3).

Our next step is to embed this instance into `FTRL` dynamics executed on a higher-dimensional space. The upshot is that we will force `FTRL` to always move on the (x_1, x_2) plane, where the potential function is essentially flat when x_1 approaches 1. At the same time, there will always be a direction perpendicular to the (x_1, x_2) plane in which `FTRL` will observe a large KKT gap, but will still never act upon it. To realize this counterintuitive behavior, we maintain the invariance that the x_1 and x_2 coordinates of `FTRL` remain in the interior throughout, whereas the rest of the coordinates always remain inactive, in that they are stuck at zero. For the active coordinates, namely (x_1, x_2), it is easy to show that `FTRL` is

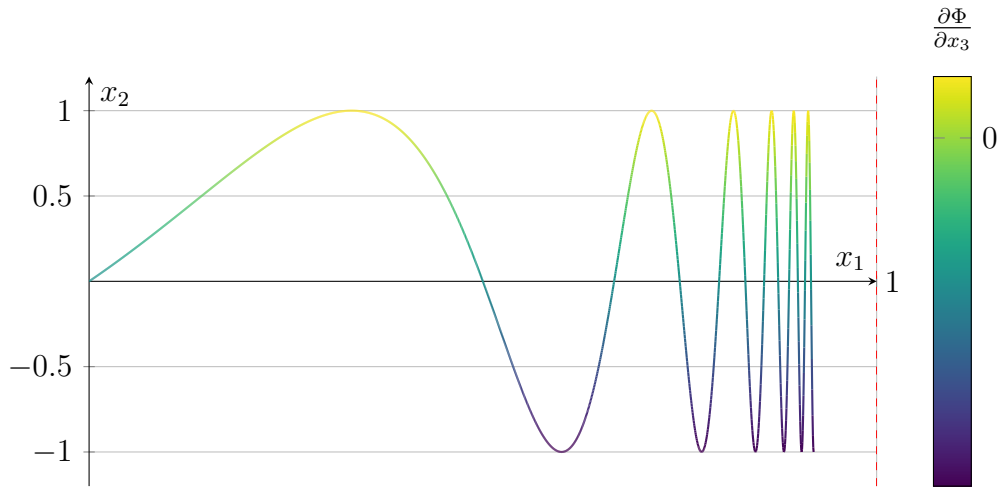


Figure 6.1: The trajectory of FTRL remains on the curve $x_2 = \sin\left(\frac{\pi x_1}{1-x_1}\right)$. The color indicates the gradient observed on the x_3 axis, which is perpendicular to the page. The crux in the argument lies in showing that FTRL never acts upon the positive gradient on the x_3 axis.

equivalent to gradient flow (Claim 6.5.2), matching the trajectory shown in Figure 6.1 . The more interesting claim is that at least one of the rest of the coordinates will observe a large KKT gap. The reason why this happens is because the *cumulative* gradient observed by each inactive coordinate will be negative throughout, even though for a constant fraction of the time the instantaneous gradient is positive. This is where the oscillating pattern in the (x_1, x_2) plane becomes critical: the key idea is to create a large negative gradient in some region of the space that neutralizes the positive gradient in aggregate (Figure 6.1).

Finally, to arrive at Theorem 6.2.1, we need to tie two loose ends. First, we need to guarantee that *for any time* the KKT gap of FTRL is at least some constant bounded away from zero. To do so, we lift not just to three dimensions, as shown in Figure 6.1, but to higher dimensions so that there is always a direction of improvement. Second, to argue about the basis of the induction—that movement only occurs on the (x_1, x_2) plane—we incorporate an additional term in the objective value that induces large negative gradients in the beginning of the trajectory, but vanishes as $x_1 \to 1$.

These steps are formalized and described in more detail in Sections 6.4 and 6.5.

6.2.2 Further related work

We conclude this section by highlighting some key lines of research closely related to result of the chapter.

Learning in potential games The convergence of learning dynamics in potential games has been a fruitful area of research. Potential games can be viewed as the special case of constrained optimization where the potential function is multilinear (Section 6.3.1). In that class of games, as discussed, first-order stationary points amount to Nash equilibria (Nash, 1950). Many natural learning algorithms have been analyzed in potential games (Monderer and Shapley, 1996a; Candogan et al., 2013; Héliou et al., 2017; Hart and Mas-Colell, 2003; Panageas et al., 2023; Anagnostides et al., 2026c; Palaiopanos et al., 2017; Kleinberg et al., 2009; Maheshwari et al., 2025; Marden et al., 2007). It is worth clarifying that these results concern specifically (last-)iterate convergence to Nash equilibria; a well-known fact is that the *average* correlated distribution induced by any no-regret dynamics converges to the set of coarse correlated equilibria (Moulin and Vial, 1978), which is a weaker solution concept—not least in terms of worst-case welfare (Blum et al., 2008). Our main result does not apply to finite potential games because there the potential function is a polynomial, and thus analytic—so gradient flow always converges pointwise (Polyak, 1963). Whether `FTRL` converges when the potential function is analytic is left as an interesting open problem. A key reference point is the result of Kleinberg et al. (2009), which showed that multiplicative weights update—the special case of `FTRL` under entropic regularization—converges in generic potential games, meaning that the payoffs are subjected to some arbitrarily small perturbation. Also, Anagnostides et al. (2026a) showed that a lazy version of `FTRL`—in which a step is only taken if it is bound to improve the potential by a sufficient amount—always converges when the potential function is smooth. Lazy updates preclude our counterexample as the dynamics would eventually stop moving in the direction where the potential function is essentially flat.

Complexity of constrained optimization From a complexity standpoint, computing a first-order stationary point of a smooth function is complete for the class `CLS`, which stands for continuous local search (Daskalakis and Papadimitriou, 2011; Fearnley et al., 2022), thus likely requiring exponential time or number of queries (Babichenko and Rubinstein, 2020, 2021). These hardness results apply in high dimensions when the desired suboptimality gap is exponential small, otherwise algorithms such as gradient descent converge in polynomial time. In contrast, our result concerning `FTRL` applies even when both the number of dimensions and the suboptimality gap are absolute constants.

Cycling behavior in games From a broader standpoint, our main result shows that, surprisingly, natural no-regret dynamics can cycle even when there is an underlying potential. It is well known that learning dynamics can fail to converge in general-sum games; for example, Daskalakis et al. (2010) gave an example where multiplicative weights update can fail to converge to Nash equilibria, and there is a rich literature on this subject (*e.g.*, Mertikopoulos et al., 2018b; Vlatakis-Gkaragkounis et al., 2020). Our work significantly strengthens those prior results by reconciling such counterexamples with the potential property, albeit applying beyond finite games.

Forgetfulness and speed of convergence Finally, Cai et al. (2024) established a separation between *optimistic* variants of `FTRL` and `MD` in zero-sum games. Their main finding, that fast convergence requires forgetfulness, mirrors our main result. We take a step further by showing that `FTRL` does not merely suffer from slow convergence, but can fail to converge altogether. Also, our result concerns a fundamentally different class of problems and algorithms compared to Cai et al. (2024).

6.3 Preliminaries

Before we dive into our analysis, in this section, we provide the necessary background on constrained optimization problems (Section 6.3.1) and follow the regularized leader (Section 6.3.2).

6.3.1 Constrained optimization and KKT gap

We begin by introducing the underlying class of problems in which we are interested. We let $\mathcal{X}$ be a convex and compact subset of a Euclidean space, and let $\Phi : \mathcal{X} \to \mathbb{R}$ be a function to be maximized. Φ will generally be nonconcave, making global optimization intractable. Instead, the main goal will be to compute a first-order optimum of Φ, in a sense formally specified below (Definition 6.3.1). Φ is assumed to be *smooth*, meaning that there exists a parameter $L > 0$ such that $\|\nabla\Phi(\boldsymbol{x}) - \nabla\Phi(\boldsymbol{x}')\|_2 \leq L\|\boldsymbol{x} - \boldsymbol{x}'\|_2$ for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$. In fact, as we shall see, our lower bound applies even when $\Phi \in C^\infty$, which is to say that the kth derivatives of Φ are continuous for any $k \in \mathbb{N}$.

We measure local suboptimality in terms of the *Karush-Kuhn-Tucker (KKT)* gap (Boyd and Vandenberghe, 2004). There are multiple equivalent ways (up to polynomial factors) of defining the KKT gap; in this chapter, we make use of the following definition.

Definition 6.3.1 (KKT gap)**.** For a function $\Phi \in C^1$ defined on a convex and compact domain $\mathcal{X}$, the *KKT gap* of a point $\boldsymbol{x} \in \mathcal{X}$ is defined as

$$\mathsf{KKTGap}(\boldsymbol{x}) \coloneqq \max_{\boldsymbol{x}' \in \mathcal{X}} \langle \boldsymbol{x}' - \boldsymbol{x}, \nabla\Phi(\boldsymbol{x}) \rangle. \tag{6.1}$$

An equivalent definition is in terms of the fixed-point gap of the gradient descent operator $\boldsymbol{x} \mapsto \Pi_{\mathcal{X}}(\boldsymbol{x} + \eta\nabla\Phi(\boldsymbol{x}))$, where $\Pi_{\mathcal{X}}(\cdot)$ denotes the Euclidean projection (Facchinei and Pang, 2003).

We will be interested in the case where the constraint set $\mathcal{X}$ can be decomposed as the Cartesian product $\mathcal{X} \coloneqq \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$. In that case, the KKT gap per (6.1) of a point $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ can be equivalently expressed as $\sum_{i=1}^{n} \max_{\boldsymbol{x}_i' \in \mathcal{X}_i} \langle \boldsymbol{x}_i' - \boldsymbol{x}_i, \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}) \rangle$. In particular, the KKT gap will be large if there is a considerable first-order deviation benefit *for at least* a single component $i \in [n]$; our lower bound relies on this simple fact.

As discussed, a special case of computing KKT points in constrained optimization is identifying Nash equilibria in potential games (Rosenthal, 1973; Monderer and Shapley, 1996a). Specifically, in n-player potential games, $\mathcal{X}$ can be decomposed as a Cartesian product of n probability simplices, and Φ is a multilinear function with respect to the strategies of each player.

6.3.2 Follow the regularized leader and the potential property

We proceed by introducing follow the regularized leader (`FTRL`), which was introduced by Kalai and Vempala (2005). We focus on the continuous-time version of `FTRL`, which can be thought of as the limit case of discrete-time `FTRL` when the learning rate approaches zero. To put this into context, it is worth reiterating that `FTRL` with a large learning rate is known to exhibit non-convergent behavior (Palaiopanos et al., 2017; Bielawski et al., 2021); our result is fundamentally different as it is not an artifact of a poorly tuned learning rate. `FTRL` was historically developed as a regularized version of *fictitious play* (Robinson, 1951a), the foundational algorithm in learning in games. Fictitious play can be viewed as the extreme version of `FTRL` when the learning rate is arbitrarily large; unlike `FTRL`, fictitious play does not guarantee the no-regret property.

In this context, continuous-time `FTRL` with Euclidean regularization can be defined as

$$\boldsymbol{x}(t) = \arg\max_{\boldsymbol{x}' \in \mathcal{X}} \left\{ \left\langle \boldsymbol{x}', \int_0^t \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}(\tau)) d\tau \right\rangle - \frac{1}{2} \|\boldsymbol{x}'\|_2^2 \right\}, \tag{6.2}$$

where Φ is the differentiable objective function; in particular, if $\mathbf{0} \in \mathcal{X}$, (6.2) prescribes initializing at $\mathbf{0}$. Equivalently, (6.2) can be expressed as

$$\boldsymbol{x}(t) = \arg\min_{\boldsymbol{x}' \in \mathcal{X}} \left\| \boldsymbol{x}' - \int_0^t \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}(\tau)) d\tau \right\|_2 = \Pi_{\mathcal{X}} \left(\int_0^t \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}(\tau)) d\tau \right). \tag{6.3}$$

A key property of continuous-time `FTRL` is that it can never decrease the value of Φ. In other words, as one would expect, Φ serves as a potential under the `FTRL` dynamics; for the discrete-time version of this property when the learning rate is small enough, we refer to Anagnostides et al. (2026a).

Lemma 6.3.1 (Potential property for `FTRL`)**.** *For any t, `FTRL` on $\Phi \in C^1$ guarantees*

$$\frac{d}{dt} \Phi(\boldsymbol{x}(t)) \geq 0.$$

Proof. For convenience, we define the accumulated gradient at time t as

$$\boldsymbol{y}(t) = \int_0^t \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}(\tau)) d\tau.$$

Taking the time derivative yields

$$\dot{\boldsymbol{y}}(t) = \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}(t)). \tag{6.4}$$

We can rewrite the update rule in terms of $\boldsymbol{y}(t)$ as

$$\boldsymbol{x}(t) = \arg\max_{\boldsymbol{x}' \in \mathcal{X}} \left\{ \langle \boldsymbol{x}', \boldsymbol{y}(t) \rangle - \frac{1}{2} \|\boldsymbol{x}'\|_2^2 \right\}.$$

We define the regularizer function as $h(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|_2^2 + I_{\mathcal{X}}(\boldsymbol{x})$, where $I_{\mathcal{X}}$ is the indicator function for the convex and compact set $\mathcal{X}$. The convex (Fenchel) conjugate of $h(\boldsymbol{x})$ is given

by

$$h^*(\boldsymbol{y}) = \max_{\boldsymbol{x}' \in \mathcal{X}} \left\{ \langle \boldsymbol{x}', \boldsymbol{y} \rangle - \frac{1}{2} \|\boldsymbol{x}'\|_2^2 \right\}.$$

By a basic property of the convex conjugate, we know that the gradient of $h^*(\boldsymbol{y})$ evaluates exactly to

$$\nabla h^*(\boldsymbol{y}) = \arg\max_{\boldsymbol{x}' \in \mathcal{X}} \left\{ \langle \boldsymbol{x}', \boldsymbol{y} \rangle - \frac{1}{2} \|\boldsymbol{x}'\|_2^2 \right\}.$$

Therefore, `FTRL` can be equivalently expressed as

$$\boldsymbol{x}(t) = \nabla h^*(\boldsymbol{y}(t)). \tag{6.5}$$

Now, using the multivariate chain rule, the time derivative of the potential function can be expressed as

$$\frac{d}{dt}\Phi(\boldsymbol{x}(t)) = \langle \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}(t)), \dot{\boldsymbol{x}}(t) \rangle.$$

Combining with (6.4), we have

$$\frac{d}{dt}\Phi(\boldsymbol{x}(t)) = \langle \dot{\boldsymbol{y}}(t), \dot{\boldsymbol{x}}(t) \rangle \tag{6.6}$$

Next, taking the time derivative in (6.5) gives

$$\dot{\boldsymbol{x}}(t) = \nabla^2 h^*(\boldsymbol{y}(t)) \dot{\boldsymbol{y}}(t)$$

Combining with (6.6),

$$\frac{d}{dt}\Phi(\boldsymbol{x}(t)) = \langle \dot{\boldsymbol{y}}(t), \nabla^2 h^*(\boldsymbol{y}(t)) \dot{\boldsymbol{y}}(t) \rangle. \tag{6.7}$$

Since $h(\boldsymbol{x})$ is a strongly convex function, its conjugate $h^*(\boldsymbol{y})$ is convex. The Hessian of any convex function, $\nabla^2 h^*(\boldsymbol{y})$, is positive semi-definite. As a result, the associated quadratic form must be nonnegative:

$$\langle \dot{\boldsymbol{y}}(t), \nabla^2 h^*(\boldsymbol{y}(t)) \dot{\boldsymbol{y}}(t) \rangle \geq 0,$$

in turn implying that

$$\frac{d}{dt} \Phi(\boldsymbol{x}(t)) \geq 0,$$

as claimed. □

As a result, Lemma 6.3.1 reassures us that the potential value is monotonically non-decreasing over time. We next point out a simple decomposition property of FTRL when the constraint set is a Cartesian product; the proof is immediate.

Corollary 6.3.1. *If $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$, then FTRL can be equivalently expressed as*

$$\boldsymbol{x}_i(t) = \operatorname*{arg\,max}_{\boldsymbol{x}_i' \in \mathcal{X}_i} \left\{ \left\langle \boldsymbol{x}_i', \int_0^t \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}(\tau)) d\tau \right\rangle - \frac{1}{2} \|\boldsymbol{x}_i'\|_2^2 \right\} \quad \forall i \in [n].$$

Gradient flow Before we proceed, it is worth comparing FTRL with the (projected) gradient flow dynamics, whereby

$$\dot{\boldsymbol{x}}(t) = \Pi_{T_{\mathcal{X}}(\boldsymbol{x}(t))}(\nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}(t))),$$

where $T_{\mathcal{X}}(\boldsymbol{x})$ denotes the tangent cone of $\mathcal{X}$ at $\boldsymbol{x}$. When the maximum in (6.2) is consistently attained in the interior of $\mathcal{X}$, the first-order optimality conditions imply that FTRL coincides with (unconstrained) gradient flow; we will use this simple observation in our counterexample (Claim 6.5.2). Moreover, gradient flow also enjoys the potential property, as it

guarantees $\frac{d}{dt}\Phi(\boldsymbol{x}(t)) = \langle \dot{\boldsymbol{x}}, \nabla_{\boldsymbol{x}}\Phi(\boldsymbol{x}(t))\rangle \geq \|\dot{\boldsymbol{x}}(t)\|_2^2$ (by the chain rule and standard properties of the tangent cone projection). Integrating yields $\int_0^\infty \|\dot{\boldsymbol{x}}(t)\|_2^2 dt \leq O(1)$ since Φ is bounded. Importantly, the norm of the velocity $\|\dot{\boldsymbol{x}}(t)\|_2$ under (constrained) gradient flow bounds the KKT gap at $\boldsymbol{x}(t)$, implying first-order stationarity as the dynamics go along. As we shall see, this is not so for `FTRL`.

6.4 Gradient flow without pointwise convergence

The starting point of our construction is a two-dimensional instance in which gradient flow does not converge pointwise. The fact that this phenomenon can occur is well known in dynamical systems (Palis and De Melo, 2012), but the point of this section is to establish some key properties that we rely on when we turn to `FTRL` on a higher dimensional space (Section 6.5).

We begin by defining the potential. It is designed so that the trajectory approaches (from the left) the boundary $x_1 = 1$ while perpetually oscillating in the x_2-coordinate. For $x_1 < 1$, we set

$$u = \frac{1}{1 - x_1}, \qquad v = \pi(u - 1), \qquad \delta = x_2 - \sin v.$$

Here, u measures the proximity to the boundary $x_1 = 1$, v is the total phase of the oscillation, and δ is the signed deviation from the oscillatory curve $x_2 = \sin v$. The goal is to construct a potential whose gradient flow maintains the invariance $\delta(t) = 0$. To this end, we let

$$a(u) \coloneqq \frac{u^4 e^{-u}}{1 + \pi^2 u^4 \cos^2 v}, \qquad L(u) \coloneqq -\pi a(u) \cos v,$$

and we define $\Phi : [0, 1] \times [-1, 1] \to \mathbb{R}$ by

$$\Phi(x_1, x_2) \coloneqq -e^{-u} - L(u)\delta - \frac{1}{2}e^{-u}\delta^2, \qquad x_1 < 1. \tag{6.8}$$

At the boundary $x_1 = 1$, we define $\Phi(1, x_2)$ by the limit $x_1 \to 1^-$ of the right-hand side of (6.8), which is 0. The resulting potential is smooth up to the boundary, as shown below.

Lemma 6.4.1 (Smoothness of Φ)**.** Φ *defined in* (6.8) *is* C^∞ *up to the boundary* $x_1 = 1$*.*

Proof. For $x_1 < 1$, the function Φ is smooth. Indeed, it is obtained by composition, addition, multiplication, and division of smooth functions, and the denominator in $L(u)$,

$$1 + \pi^2 u^4 \cos^2 v,$$

is strictly positive. Hence the only issue is smoothness at the boundary $x_1 = 1$, which corresponds to $u \to +\infty$.

We shall use the standard fact that exponential decay dominates polynomial growth: for every $M > 0$,

$$u^M e^{-u} \to 0 \qquad \text{as } u \to +\infty.$$

Consequently, any expression of the form

$$e^{-u} R(u, \sin v, \cos v),$$

where R grows at most polynomially in u, tends to zero as $u \to +\infty$, since $\sin v$ and $\cos v$ are bounded.

This property is preserved under differentiation with respect to x_1. Indeed,

$$\frac{\partial u}{\partial x_1} = u^2,$$

and therefore

$$\partial_{x_1} = u^2 \partial_u.$$

Applying $u^2\partial_u$ to an expression of the form

$$e^{-u} R(u, \sin v, \cos v)$$

again gives an expression of the same type: an exponentially decaying factor e^{-u}, multiplied by a function with at most polynomial growth in u and bounded trigonometric factors. Repeating this argument shows that every x_1-derivative of such an expression tends to zero as $u \to +\infty$.

We now apply this observation to the three terms in

$$\Phi = -e^{-u} - L(u)\delta - \frac{1}{2} e^{-u} \delta^2.$$

First, $-e^{-u}$ is flat at $x_1 = 1$. That is, all of its x_1-derivatives are of the form

$$e^{-u} P(u),$$

where P is a polynomial. Hence all of them tend to zero as $u \to +\infty$.

Next, consider the linear term $-L(u)\delta$. Since

$$L(u) = -\frac{\pi u^4 e^{-u} \cos v}{1 + \pi^2 u^4 \cos^2 v},$$

and the denominator is bounded below by 1, the factor $L(u)$ contains the exponentially decaying factor e^{-u}. Therefore $L(u)$, together with all of its x_1-derivatives, tends to zero as $u \to +\infty$.

Moreover,

$$\delta = x_2 - \sin v.$$

Differentiating δ with respect to x_1 only produces polynomial powers of u and bounded trigonometric factors. Hence, by the product rule, every x_1-derivative of $L(u)\delta$ is a finite sum of terms of the form

$$e^{-u} R(u, \sin v, \cos v, x_2),$$

where R grows at most polynomially in u, locally uniformly in x_2. Thus all such derivatives tend to zero as $u \to +\infty$.

Finally, consider the quadratic term

$$-\frac{1}{2} e^{-u} \delta^2.$$

The factor e^{-u} is flat at the boundary. Differentiating δ^2 with respect to x_1 produces only polynomial powers of u, powers of δ, and bounded trigonometric factors. Since smoothness is a local property in the x_2-variable, we may restrict x_2 to a compact set; on such a set, $\delta = x_2 - \sin v$ is bounded. Therefore every x_1-derivative of $e^{-u}\delta^2$ also tends to zero as $u \to +\infty$.

It remains to consider derivatives with respect to x_2. These cause no difficulty, since Φ is at most quadratic in x_2. More explicitly,

$$\partial_{x_2}\Phi = -L(u) - e^{-u}\delta,$$

$$\partial_{x_2}^2\Phi = -e^{-u},$$

and

$$\partial_{x_2}^k\Phi = 0 \qquad \text{for all } k \geq 3.$$

Each of these expressions, and each of their x_1-derivatives, tends to zero as $u \to +\infty$ by the preceding argument.

Therefore every mixed partial derivative

$$\partial_{x_1}^\alpha \partial_{x_2}^\beta \Phi$$

has a continuous extension to $x_1 = 1$, and its boundary value is zero. Hence the extension defined by

$$\Phi(1, x_2) = 0$$

is C^∞ up to the boundary $x_1 = 1$. □

The role of the linear term $L(u)\delta$ in (6.8) is to make the oscillatory curve $\mathcal{M} = \{\delta = 0\}$ forward invariant under the gradient flow. As we shall show, along this curve, the dynamics reduce to the scalar differential equation $\dot{u} = a(u) > 0$, so $u(t)$ increases monotonically and the trajectory moves through successive oscillations while approaching the boundary $x_1 = 1$.

In the remainder of this section, we first formalize that the oscillatory curve $\{\delta = 0\}$ is forward invariant (Lemma 6.4.2). The second and more involved step involves obtaining dwell-time estimates for the motion through each oscillation (Lemmas 6.4.3 and 6.4.4).

6.4.1 Invariant oscillatory curve

We first show that the potential function defined in (6.8) guarantees the following key invariance property.

Lemma 6.4.2 (Forward invariance)**.** *Let*

$$\mathcal{M} \coloneqq \{(x_1, x_2) : \delta = 0\} = \left\{(x_1, x_2) : x_2 = \sin\left(\frac{\pi x_1}{1 - x_1}\right)\right\}.$$

Then $\mathcal{M}$ is forward invariant under the gradient flow on Φ.

Proof. We use the gradient-ascent dynamics $\dot{\boldsymbol{x}} = \nabla\Phi(\boldsymbol{x})$, where the maximization potential is

$$\Phi = -e^{-u} - L(u)\delta - \frac{1}{2}e^{-u}\delta^2, \qquad \delta = x_2 - \sin v, \qquad v = \pi(u - 1).$$

We prove forward invariance by showing that the tracking error cannot move away from zero once it is zero; namely, we show that $\dot{\delta} = 0$ whenever $\delta = 0$.

We first compute the x_2-velocity on the curve $\mathcal{M} = \{\delta = 0\}$. Since $\partial_{x_2}\delta = 1$, we have $\partial_{x_2}\Phi = -L(u) - e^{-u}\delta$. Therefore, on $\mathcal{M}$,

$$\dot{x}_2 = \partial_{x_2}\Phi = -L(u). \tag{6.9}$$

Next we compute the induced u-velocity. Since $\frac{\partial u}{\partial x_1} = u^2$ and $\frac{\partial v}{\partial x_1} = \pi u^2$, we get

$$\frac{\partial \delta}{\partial x_1} = -\pi u^2 \cos v. \tag{6.10}$$

When $\delta = 0$, every term in $\partial_{x_1}\Phi$ containing an explicit factor of δ vanishes. Hence, using (6.10), we obtain

$$\begin{aligned} \partial_{x_1}\Phi &= u^2 e^{-u} - L(u)\frac{\partial \delta}{\partial x_1} \\ &= u^2 e^{-u} + \pi u^2 L(u) \cos v. \end{aligned} \tag{6.11}$$

Since we are using gradient ascent, $\dot{x}_1 = \partial_{x_1}\Phi$. Therefore

$$\dot{x}_1 = u^2 e^{-u} + \pi u^2 L(u) \cos v. \tag{6.12}$$

Using $\dot{u} = u^2 \dot{x}_1$, this gives

$$\dot{u} = u^4 e^{-u} + \pi u^4 L(u) \cos v. \tag{6.13}$$

We now use the specific choice of the linear coefficient, $L(u) = -\pi a(u) \cos v$. Substituting this into (6.13) yields

$$\dot{u} = u^4 e^{-u} - \pi^2 u^4 a(u) \cos^2 v. \tag{6.14}$$

By definition of $a(u)$, it follows that

$$u^4 e^{-u} = a(u)\left(1 + \pi^2 u^4 \cos^2 v\right). \tag{6.15}$$

Combining (6.14) and (6.15), we conclude that, on $\mathcal{M}$,

$$\dot{u} = a(u) > 0. \tag{6.16}$$

It remains to verify that the tracking error is preserved. Differentiating $\delta = x_2 - \sin v$ along the flow gives

$$\dot{\delta} = \dot{x}_2 - \cos v\, \dot{v} = \dot{x}_2 - \pi \cos v\, \dot{u}.$$

Using (6.9) and (6.16), we get, on $\mathcal{M}$,

$$\dot{\delta} = -L(u) - \pi a(u) \cos v.$$

Finally, since $L(u) = -\pi a(u) \cos v$, the right-hand side vanishes:

$$\dot{\delta} = 0.$$

Therefore $\mathcal{M} = \{\delta = 0\}$ is forward invariant under the gradient-ascent dynamics. Moreover, (6.16) gives $\dot{u} = a(u) > 0$ on $\mathcal{M}$. □

$\mathcal{M}$ is forward invariant, meaning that if the gradient flow is initialized in $\mathcal{M}$, then the trajectory remains in $\mathcal{M}$ for all future times. For completeness, we also provide an experimental verification of this invariance in Figure 6.2.

We state two consequences of Lemma 6.4.2. The first, Corollary 6.4.1, was already established in the proof above, while the second, Corollary 6.4.2, follows naturally from the fact that $\delta(t) = 0$ is preserved for all $t \geq 0$.

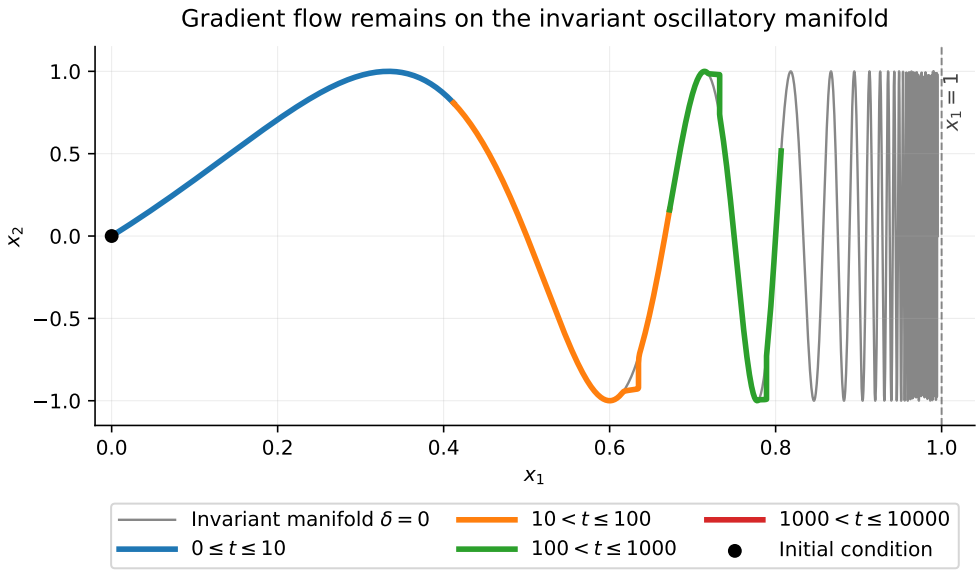


Figure 6.2: Coordinate x_2 closely exactly the curve $\sin v$ throughout the trajectory. The small deviations are due to numerical approximation error and occur near the critical points, where the velocity of $\sin v$ changes sign. The figure also illustrates that the velocity of u decreases over time: in later oscillations, the movement of u, and consequently the movement of x_1 toward 1, becomes progressively slower.

Corollary 6.4.1. *If gradient flow is initialized at $(x_1(0), x_2(0)) = (0, 0)$, then $\delta(t) = 0$ for all $t \geq 0$. Moreover, $\dot{u}(t) = a(u) > 0$, so $u(t)$ is strictly increasing, and it decreases over time.*

Corollary 6.4.2 (Limit sets)**.** *If gradient flow is initialized on $\mathcal{M}$, then $u(t) \to +\infty$ and $x_1(t) \to 1$. Moreover, the ω-limit set reads $\omega(\boldsymbol{x}(0)) = \{1\} \times [-1, 1]$.*

6.4.2 Dwell-time bounds

Next, we estimate how much time the gradient flow spends in different parts of each oscillation. These dwell-time bounds will be crucial when we embed the two-dimensional gradient flow instance into the higher-dimensional `FTRL` dynamics, as they allow us to control the cumulative gradient observed by each auxiliary coordinate. Referring back to Figure 6.1, the key requirement is that, within each oscillation, the positive cumulative gradient is subsumed by the negative cumulative gradient. This requires the flow to spend sufficiently little

time in the positive-gradient region; this subsection sets the groundwork for establishing that property.

Phase windows We first introduce the phase windows. We assume that the gradient flow is initialized on the invariant curve $\mathcal{M}$. During the kth oscillation, with $k \geq 2$, we write

$$v = 2\pi(k-1) + \phi, \qquad \text{for} \quad \phi \in [0, 2\pi], \quad \text{and} \quad U_k \coloneqq 2k - 1.$$

In words, ϕ represents the local phase within a single oscillation, while v is the absolute phase. For $m \geq 1$, we let

$$\ell^{(m)} \coloneqq \frac{\pi}{2^m}, \qquad h^{(m)} \coloneqq \frac{\ell^{(m)}}{2} = \frac{\pi}{2^{m+1}}, \quad \text{and} \quad N^{(m)} = \frac{2\pi}{h^{(m)}} = 2^{m+2}.$$

Here, $\ell^{(m)}$ is the window length and $h^{(m)}$ is the shift between consecutive windows, so adjacent windows overlap by $h^{(m)}$. Furthermore, for $i = 0, \ldots, N^{(m)} - 2$, we define the non-wrapping windows

$$I_i^{(m)} \coloneqq [ih^{(m)}, ih^{(m)} + \ell^{(m)}] \subset [0, 2\pi].$$

The final window crosses 2π, and is written as

$$I_{N^{(m)}-1}^{(m)} \coloneqq [2\pi - h^{(m)}, 2\pi] \cup [0, h^{(m)}],$$

where the first piece is counted in the k-th oscillation and the second one in the next oscillation. Figure 6.3 contains a visualization of the phase windows.

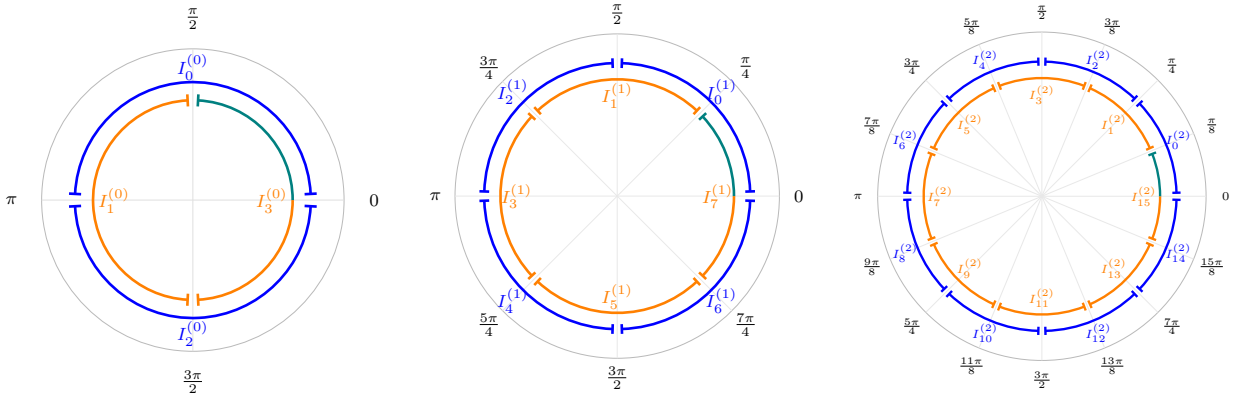


Figure 6.3: Overlapping intervals of decreasing size. The part shown in teal is the final wrapped part of the last interval, which extends into the next oscillation.

Dwell times The motion in $\mathcal{M}$ can be parametrized by the phase. Since $v = \pi(u-1)$, we have

$$\dot{v} = \pi\dot{u} = \pi a(u) = \frac{\pi u^4 e^{-u}}{1 + \pi^2 u^4 \cos^2 v} > 0,$$

by Corollary 6.4.1. Thus, the phase is strictly increasing, and we can compute the dwell time by changing variables from t to the phase. Since $\dot{\phi} = \dot{v}$, the time density with respect to the local phase ϕ is $g_k(\phi) = \frac{dt}{d\phi} = 1/\dot{v}$. Therefore, since $\cos v = \cos(2\pi k + \phi) = \cos\phi$, we can write

$$g_k(\phi) = \frac{1 + \pi^2 u^4 \cos^2\phi}{\pi u^4 e^{-u}} = \frac{e^u}{\pi u^4} + \pi e^u \cos^2\phi.$$

Now, $g_k(\phi)\, d\phi$ is precisely the time spent while the local phase moves through the infinitesimal interval $d\phi$. For $i = 0, \ldots, N^{(m)} - 2$, we define

$$T_{k,i}^{(m)} \coloneqq \int_{I_i^{(m)}} g_k(\phi)\, d\phi, \qquad Q_{k,i}^{(m)} \coloneqq \int_{[0,2\pi]\setminus I_i^{(m)}} g_k(\phi)\, d\phi.$$

Finally, for the wrapping window, we define

$$T^{(m)}_{k,N^{(m)}-1} \coloneqq \int_{2\pi-h^{(m)}}^{2\pi} g_k(\phi)\,d\phi + \int_0^{h^{(m)}} g_{k+1}(\phi)\,d\phi, \qquad Q^{(m)}_{k,N^{(m)}-1} \coloneqq \int_0^{2\pi-h^{(m)}} g_k(\phi)\,d\phi.$$

Lemma 6.4.3 (Dwell-time bounds for overlapping phase windows)**.** *Suppose that the gradient flow is initialized on the invariant curve $\mathcal{M}$. Then, for every $k \geq 2$, $m \geq 1$, and $i = 0, \ldots, N^{(m)} - 1$,*

$$T^{(m)}_{k,i} \leq \frac{15}{2^m} Q^{(m)}_{k,i}.$$

For the construction of the potential in Section 6.5, we want to express the dwell time in terms of level sets of the x_2-coordinate rather than phase windows. To this end, for each phase window $I^{(m)}_i$, $i = 0, \ldots, N^{(m)} - 1$, we define the associated x_2-level set by

$$A^{(m)}_i \coloneqq \sin\big(I^{(m)}_i\big) \coloneqq \left\{ y \in [-1,1] : y = \sin\phi \text{ for some } \phi \in I^{(m)}_i \right\},$$

which should be viewed as the corresponding interval in the x_2-coordinate. Then, during the kth oscillation, we let $\widetilde{T}^{(m)}_{k,i}$ denote the amount of time for which $x_2(t) \in A^{(m)}_i$, and let $\widetilde{Q}^{(m)}_{k,i}$ denote the amount of time for which $x_2(t) \notin A^{(m)}_i$.

Lemma 6.4.4 (Dwell-time bounds in terms of x_2-level sets)**.** *Suppose that the gradient flow is initialized on the invariant curve $\mathcal{M}$. Then, for every $k \geq 2$, $m \geq 5$, and $i = 0, \ldots, N^{(m)} - 1$,*

$$\widetilde{T}^{(m)}_{k,i} \leq \frac{60}{2^m} \widetilde{Q}^{(m)}_{k,i}. \tag{6.17}$$

Roughly speaking, for the purpose of Section 6.5, one can think of $\widetilde{T}_{k,i}^{(m)}$ as the total time during the kth oscillation in which the corresponding gradient is positive, and $\widetilde{Q}_{k,i}^{(m)}$ as the complement of $\widetilde{T}_{k,i}^{(m)}$. When m is large, (6.17) shows that $\widetilde{T}_{k,i}^{(m)}$ is a small fraction of the oscillation's duration.

6.5 Main result: from gradient flow to FTRL

To obtain our main result, we now embed the gradient flow dynamics of Section 6.4 into a higher-dimensional space. Specifically, the first two coordinates (x_1, x_2) are the original variables of gradient flow, while the remaining coordinates $(x_3, \ldots, x_{N^{(m)}+2})$ serve an auxiliary role. The key invariance we will establish is that $x_i(t) = 0$ for any $i = 3, \ldots, N^{(m)} + 2$ and $t \geq 0$, so `FTRL` in this higher-dimensional space will remain trapped in the (x_1, x_2) plane, mimicking the trajectory of gradient flow. In accordance with the dwell-time bounds of Lemma 6.4.4, we associate x_{3+i} with the ith x_2-level interval. The constraint set is defined as the product of intervals $\mathcal{X}_1 \times \mathcal{X}_2 \times [0,1] \times \cdots \times [0,1]$, where $\mathcal{X}_1$ and $\mathcal{X}_2$ are strict supersets of $[0,1]$ and $[-1,1]$, respectively; this guarantees that constrained gradient flow with respect to (x_1, x_2) coincides with the unconstrained gradient flow by virtue of Lemma 6.4.2.

Higher-dimensional potential We now introduce the augmented potential function. We recall from Section 6.4.2 that each phase window $I_i^{(m)}$ is in one-to-one correspondence with the interval

$$A_i^{(m)} = [\alpha_i^{(m)}, \beta_i^{(m)}], \quad \text{where } \alpha_i^{(m)} := \min_{\phi \in I_i^{(m)}} \sin\phi, \ \beta_i^{(m)} := \max_{\phi \in I_i^{(m)}} \sin\phi.$$

In particular, for the wrapping window, we have

$$A_{N^{(m)}-1}^{(m)} = \sin\left([2\pi - h^{(m)}, 2\pi] \cup [0, h^{(m)}]\right) = [-\sin h^{(m)}, \sin h^{(m)}].$$

Inside each interval $A_i^{(m)} = [\alpha_i^{(m)}, \beta_i^{(m)}]$, we define a trimmed interval $\widetilde{A}_i^{(m)} \coloneqq [\widetilde{\alpha}_i^{(m)}, \widetilde{\beta}_i^{(m)}]$; the role of these trimmed intervals will become clear shortly. Now, if $A_i^{(m)}$ lies strictly inside $[-1, 1]$, we remove one quarter of its length from each endpoint, and define

$$\widetilde{A}_i^{(m)} \coloneqq \left[\frac{3\alpha_i^{(m)} + \beta_i^{(m)}}{4}, \frac{\alpha_i^{(m)} + 3\beta_i^{(m)}}{4}\right], \quad \text{and hence} \quad \left|\widetilde{A}_i^{(m)}\right| = \frac{\left|A_i^{(m)}\right|}{2}.$$

The only intervals that need separate treatment are the two endpoint intervals: one reaches 1 and the other reaches -1. For $m \geq 1$, these are distinct because each phase window has length at most $\pi/2$. For the endpoint intervals, we modify the construction slightly. At the endpoint -1 or 1, we extend the interval by one quarter of its original length, while trimming the opposite side by one quarter. This choice is made only to keep the calculations uniform with the non-endpoint case. That is, if $\beta_i^{(m)} = 1$, we define

$$\widetilde{A}_i^{(m)} \coloneqq \left[\frac{3\alpha_i^{(m)} + \beta_i^{(m)}}{4}, \beta_i^{(m)} + \frac{\beta_i^{(m)} - \alpha_i^{(m)}}{4}\right], \quad \text{and hence} \quad \left|\widetilde{A}_i^{(m)}\right| = \left|A_i^{(m)}\right|.$$

Similarly, if $\alpha_i^{(m)} = -1$, we define

$$\widetilde{A}_i^{(m)} \coloneqq \left[\alpha_i^{(m)} - \frac{\beta_i^{(m)} - \alpha_i^{(m)}}{4}, \frac{\alpha_i^{(m)} + 3\beta_i^{(m)}}{4}\right], \quad \text{and hence} \quad \left|\widetilde{A}_i^{(m)}\right| = \left|A_i^{(m)}\right|.$$

We associate with the interval $\widetilde{A}_i^{(m)}$ the quadratic term

$$q_i^{(m)}(x_2) \coloneqq (x_2 - \widetilde{\alpha}_i^{(m)})(\widetilde{\beta}_i^{(m)} - x_2),$$

so that the initial potential function Φ per (6.8) in Section 6.4 is now recast as

$$\widetilde{\Phi}(x_1, x_2, x_3, \ldots, x_{N^{(m)}+2}) \coloneqq \Phi(x_1, x_2) + \sum_{i=0}^{N^{(m)}-1} x_{3+i}\, q_i^{(m)}(x_2). \tag{6.18}$$

As a result, the gradient of (6.18) with respect to x_{3+i} is precisely $q_i^{(m)}(x_2)$, which is positive when x_2 lies in $\widetilde{A}_i^{(m)}$. At the same time, as long as $x_{3+i} = 0$, the added terms do not change the (x_1, x_2)-trajectory since the added contribution to $\partial\widetilde{\Phi}/\partial x_2$ is always multiplied by one of the auxiliary variables. Furthermore, the intervals constructed above are such that at any point in time at least one coordinate will observe a positive gradient (Lemma 6.5.1), so `FTRL` will experience a large KKT gap.

Now, under the constraint set $\mathcal{X}_1 \times \mathcal{X}_2 \times [0,1] \times \cdots \times [0,1]$, `FTRL` can be decomposed as $x_i(t) = \Pi_{\mathcal{X}_i}(\int_0^t \nabla_{x_i}\widetilde{\Phi}(\boldsymbol{x}(\tau))d\tau)$ for each coordinate i (Corollary 6.3.1). The first direct consequence of this decomposition is that maintaining the invariance $x_{3+i} = 0$ reduces to showing that the observed *cumulative* gradient is nonpositive—the projection of any nonpositive number into $[0,1]$ is 0.

Claim 6.5.1. For $i \geq 3$, if $\int_0^t \nabla_{x_i}\widetilde{\Phi}(\boldsymbol{x}(\tau))d\tau \leq 0$ for any $t \geq 0$, then $x_i(t) = 0$.

Bounding the cumulative gradients To bound the cumulative utility over a full oscillation, we will rely on the dwell-time arguments developed in Section 6.4. There is, however, a caveat. The original phase windows $I_i^{(m)}$ were defined so that consecutive windows overlap by $h^{(m)}$. By construction, these windows cover the entire trigonometric circle. Consequently, their images under sin, denoted by $A_i^{(m)}$, also form an overlapping cover of the relevant x_2-values.

We then introduced the trimmed intervals $\widetilde{A}_i^{(m)}$, obtained by removing a fixed portion from the endpoints of each $A_i^{(m)}$. This trimming was needed in order to obtain uniform upper bounds on the gradient throughout an oscillation, as used in the proof of Lemma 6.5.2. However, because the map $v \mapsto \sin v$ is nonlinear and non-monotone over the full circle, it is not immediate that the trimmed intervals $\widetilde{A}_i^{(m)}$ still cover all possible x_2-values. In what follows, we show that this is not the case: the intervals $\widetilde{A}_i^{(m)}$ still cover the entire range of x_2-values; the proof is deferred to Section A.6.

Lemma 6.5.1 (Coverage and overlap of the trimmed intervals)**.** *For every $m \geq 1$, the trimmed intervals $\widetilde{A}_i^{(m)}$, $i = 0, \ldots, N^{(m)} - 1$, cover the feasible x_2-axis:*

$$[-1, 1] \subseteq \bigcup_{i=0}^{N^{(m)}-1} \widetilde{A}_i^{(m)}.$$

Moreover, consecutive intervals overlap by a quantitatively positive amount. If

$$\omega^{(m)} := 2 \sin^3\left(\frac{h^{(m)}}{2}\right) \sin\left(\frac{3h^{(m)}}{2}\right),$$

then, with indices interpreted modulo $N^{(m)}$,

$$\left|\widetilde{A}_i^{(m)} \cap \widetilde{A}_{i+1}^{(m)}\right| \geq \omega^{(m)} \qquad \text{for every } i = 0, \ldots, N^{(m)} - 1.$$

In particular, $\omega^{(m)} > 0$.

Thus, the coverage and overlap properties from Lemma 6.5.1 ensure that these positive first-order gradients, $q_i^{(m)}$, are available throughout the x_2-axis, except possibly at the global endpoints. We now explain why both cases are handled properly.

Indeed, the intervals $\widetilde{A}_i^{(m)}$ cover $[-1, 1]$, and consecutive intervals overlap nontrivially. Hence, if $x_2 \in (-1, 1)$, then x_2 lies in the interior of at least one trimmed interval. This is precisely where the nontrivial overlap of Lemma 6.5.1 is used: if x_2 is at the boundary of one trimmed interval, then, away from the global endpoints $\{-1, 1\}$, it lies in the interior of a neighboring trimmed interval. The only remaining cases are $x_2 = 1$ and $x_2 = -1$. By construction, the quadratic function $q_i^{(m)}(x_2)$ is defined so that, whenever $A_i^{(m)}$ contains one of the endpoints -1 or 1, the corresponding root is shifted accordingly. Therefore, even if x_2 reaches one of these endpoints, the resulting first-order gradient is nonpositive by definition.

Now, we may employ the dwell-time estimate from Lemma 6.4.4 to control the cumulative gradient observed by each auxiliary coordinate. We fix an oscillation k and an index $i = 0, \dots, N^{(m)} - 1$. Since $\widetilde{A}_i^{(m)} \cap [-1, 1] \subset A_i^{(m)}$, the time spent inside $\widetilde{A}_i^{(m)}$ is at most $\widetilde{T}_{k,i}^{(m)}$, while the time spent outside $\widetilde{A}_i^{(m)}$ is at least $\widetilde{Q}_{k,i}^{(m)}$. If $\Delta_i^{(m)} = |A_i^{(m)}| = \beta_i^{(m)} - \alpha_i^{(m)} > 0$, the quadratic function $q_i^{(m)}(x_2)$ satisfies the following.

$$\max_{x_2 \in \widetilde{A}_i^{(m)}} q_i^{(m)}(x_2) = \frac{(\Delta_i^{(m)})^2}{16}, \quad \text{and} \quad q_i^{(m)}(\alpha_i^{(m)}) = q_i^{(m)}(\beta_i^{(m)}) = -\frac{3(\Delta_i^{(m)})^2}{16}.$$

Hence, by concavity, $q_i^{(m)}(x_2) \leq \frac{-3(\Delta_i^{(m)})^2}{16}$, whenever $x_2 \notin A_i^{(m)}$. For the endpoint intervals, we consider two cases. First, if $\beta_i^{(m)} = 1$, then

$$\max_{x_2 \in \widetilde{A}_i^{(m)}} q_i^{(m)}(x_2) = \frac{(\Delta_i^{(m)})^2}{4}, \quad \text{and} \quad q_i^{(m)}(\alpha_i^{(m)}) = -\frac{5(\Delta_i^{(m)})^2}{16}.$$

Similarly, if $\alpha_i^{(m)} = -1$, then

$$\max_{x_2 \in \widetilde{A}_i^{(m)}} q_i^{(m)}(x_2) = \frac{(\Delta_i^{(m)})^2}{4}, \quad \text{and} \quad q_i^{(m)}(\beta_i^{(m)}) = -\frac{5(\Delta_i^{(m)})^2}{16}.$$

In either case, concavity gives $q_i^{(m)}(x_2) \leq -\frac{5(\Delta_i^{(m)})^2}{16}$ whenever $x_2 \notin A_i^{(m)}$ on the side of $A_i^{(m)}$ that was trimmed away.

Thus, whenever x_2 lies in an interval $\widetilde{A}_i^{(m)}$, the coordinate x_{3+i} receives a positive first-order signal: increasing x_{3+i} from the boundary value 0 is a beneficial feasible deviation. Outside $\widetilde{A}_i^{(m)}$, the quadratic is always nonpositive; more importantly, outside the larger interval $A_i^{(m)}$, it is strictly negative.

We are now ready to prove our claim. By Lemma 6.4.4, it follows that over the k-th oscillation the cumulative gradient of variable x_i can be bounded as

$$\int \nabla_{x_i} \Phi(\boldsymbol{x}(t))\, dt = \int q_i^{(m)}(x_2(t))\, dt \le \int_{\widetilde{T}_{k,i}^{(m)}} \frac{(\Delta_i^{(m)})^2}{4}\, dt + \int_{\widetilde{Q}_{k,i}^{(m)}} -\frac{5(\Delta_i^{(m)})^2}{16}\, dt \tag{6.19}$$

$$\le \frac{(\Delta_i^{(m)})^2}{16}\left(4\widetilde{T}_{k,i}^{(m)} - 5\,\widetilde{Q}_{k,i}^{(m)}\right) \tag{6.20}$$

$$\le \frac{(\Delta_i^{(m)})^2 \widetilde{Q}_{k,i}^{(m)}}{16}\left(4\frac{60}{2^m} - 5\right), \tag{6.21}$$

which is negative for $m \ge 6$. Thus, even though the auxiliary coordinate observes positive gradient when $x_2 \in \widetilde{A}_i^{(m)}$, this gets subsumed by the negative contribution accumulated while $x_2 \notin \widetilde{A}_i^{(m)}$. We summarize this key property in the lemma below.

Lemma 6.5.2. *Suppose that $(x_1(t), x_2(t))$ evolves as the (unconstrained) gradient flow executed on (6.8). If $m \ge 6$ in (6.18), then $\int_0^t q_i^{(m)}(x_2(\tau))d\tau \le 0$ for any $i \ge 0$ and $t \ge 0$.*

Initialization To complete the construction, we need to handle the initialization. To do so, we incorporate an additional term in (6.18), namely

$$\widetilde{\Phi}'(x_1, x_2, x_3, \ldots, x_{N^{(m)}+2}) \coloneqq \widetilde{\Phi}(x_1, x_2, x_3, \ldots, x_{N^{(m)}+2}) - C \sum_{i=0}^{N^{(m)}-1} x_{3+i} e^{-1/(1-x_1)}, \tag{6.22}$$

where $C > 0$ is a sufficiently large constant. The role of the last term in (6.22) is this: when C is large enough, each coordinate x_{3+i} will observe a large negative gradient in the beginning (so it will remain at 0), but as x_1 approaches 1 that term will quickly vanish and become negligible. As before, this additional term will not affect the gradient observed by x_1 or x_2.

For any k, we can pick $C = C(k)$ in (6.22) large enough so that, for each $i \ge 3$, the cumulative gradient is nonpositive up to the kth oscillation. To argue after that point, we use Lemma 6.5.2 to readily conclude that the positive gradient accumulated inside $\widetilde{A}_i^{(m)}$ is

outweighed by the cumulative negative gradient that comes before it, leading to the following consequence.

Corollary 6.5.1. *Suppose that* $(x_1(t), x_2(t))$ *evolves as the (unconstrained) gradient flow executed on* (6.8). *If* $C > 0$ *is large enough,* $\int_0^t \nabla_{x_i} \widetilde{\Phi}'(\boldsymbol{x}(\tau))d\tau \leq 0$ *for any* $i \geq 3$ *and* $t \geq 0$.

Combining with Claim 6.5.1, this would show that $x_i(t) = 0$ for any $t \geq 0$ and $i \geq 3$. What remains to argue is that on the (x_1, x_2) plane, `FTRL` executed on (6.22) adheres to the motion described in Lemma 6.4.2. This follows from the equivalence of `FTRL` and gradient flow when the latter remains on the interior of the constraint set, as we point out below; the proof follows by simply integrating.

Claim 6.5.2 (Equivalence of gradient flow to `FTRL` on the interior)**.** Consider gradient flow on a function Φ and constraint set $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ such that for each $i \in S \subseteq [n]$, $x_i(t)$ is always in the interior of $\mathcal{X}_i$. For each coordinate $i \in S$, $x_i(t) = \int_0^t \nabla_{x_i} \Phi(\boldsymbol{x}(\tau))d\tau = \Pi_{\mathcal{X}_i}(\int_0^t \nabla_{x_i} \Phi(\boldsymbol{x}(\tau))d\tau)$.

It remains to show that, throughout the execution of `FTRL`, there are always variables x_{3+i} with a constant KKTGap. The point of Lemma 6.5.1 is that such an index i always exists. Indeed, the trimmed intervals cover the whole feasible x_2-axis, and consecutive intervals overlap by at least $\omega^{(m)} > 0$. Hence x_2 cannot pass from one interval to the next only through points where all adjacent quadratic gates vanish. Moreover, at the global endpoints ± 1, the shifted-root convention ensures that the corresponding endpoint gate remains strictly positive. We split the proof into two parts.

Lemma 6.5.3 (Positive coordinate-wise KKTGap)**.** *Fix* $m \geq 1$ *and* $i \in \{0, \ldots, N^{(m)} - 1\}$. *Suppose that the auxiliary coordinate* x_{3+i} *is kept at the boundary value* 0. *If* $q_i^{(m)}(x_2(t)) > 0$,

then the coordinate-wise first-order gap of x_{3+i} is strictly positive:

$$\max_{x'_{3+i}\in\mathcal{X}_{3+i}} \left\langle x'_{3+i} - x_{3+i}, \nabla_{x_{3+i}}\Phi(\boldsymbol{x}(t))\right\rangle > 0.$$

Proof. By construction,

$$\nabla_{x_{3+i}}\Phi(\boldsymbol{x}(t)) = q_i^{(m)}(x_2(t)).$$

Since $x_{3+i} = 0$, choosing any feasible $x'_{3+i} > 0$ gives

$$\left\langle x'_{3+i} - x_{3+i}, \nabla_{x_{3+i}}\Phi(\boldsymbol{x}(t))\right\rangle = x'_{3+i}q_i^{(m)}(x_2(t)) > 0.$$

Hence the coordinate-wise first-order gap of x_{3+i} is strictly positive. □

Lemma 6.5.4 (Pointwise positive KKTGap)**.** *Fix $m \geq 1$, and suppose that the auxiliary coordinates are kept at $x_{3+i} = 0$ for all $i = 0, \ldots, N^{(m)} - 1$. Then, at every time t,*

$$\mathsf{KKTGap}(\boldsymbol{x}(t)) > 0.$$

Proof. Since the trajectory always satisfies $x_2(t) \in [-1, 1]$, it is enough to find one auxiliary coordinate whose quadratic gate is strictly positive. By Lemma 6.5.1, the intervals $\widetilde{A}_i^{(m)}$ cover $[-1, 1]$, and consecutive intervals overlap by at least $\omega^{(m)} > 0$. Therefore, for every $x_2(t) \in [-1, 1]$, there exists an index $i \in \{0, \ldots, N^{(m)} - 1\}$ such that $q_i^{(m)}(x_2(t)) > 0$. For points in the interior of $[-1, 1]$, this follows from the nontrivial overlap of consecutive trimmed intervals. At the endpoint values $x_2(t) = \pm 1$, it follows from the shifted-root convention for the endpoint intervals.

Applying Lemma 6.5.3 to this index gives a strictly positive coordinate-wise first-order gap. Since the full KKTGap dominates every coordinate-wise first-order gap, we conclude that

$$\mathsf{KKTGap}(\boldsymbol{x}(t)) > 0.$$

□

Combining all these ingredients, we arrive at Theorem 6.2.1.

6.6 Concluding remarks

In this chapter, we showed that continuous-time FTRL can fail to converge to first-order stationary points in constrained nonconvex optimization. More precisely, we constructed a smooth objective on a compact convex domain for which the KKTGap of the FTRL trajectory remains bounded away from zero for all time. This happens despite the fact that the objective value improves monotonically along the trajectory.

The construction highlights a fundamental distinction between objective monotonicity and stationarity. For gradient-based dynamics, monotone ascent is often tightly connected to convergence toward first-order stationary points. For FTRL, however, the dynamics are governed by cumulative payoff information rather than by the instantaneous gradient alone. This history dependence can prevent the trajectory from reacting to currently profitable feasible directions, especially near the boundary of the domain. As a result, FTRL can continue to make progress in objective value while systematically failing to satisfy the first-order optimality conditions.

Technically, the proof proceeds by first constructing a smooth objective whose gradient-flow dynamics exhibit decelerating oscillations without pointwise convergence. We then embed this oscillatory behavior into a higher-dimensional constrained problem in which FTRL

inherits the same nonconvergent behavior, while additional inactive directions maintain a uniformly positive KKTGap. The resulting example shows that the obstruction is not caused by discontinuities, lack of compactness, or absence of objective improvement, but rather by the interaction between constraint geometry and the cumulative nature of the FTRL update.

Within the broader narrative of the dissertation, this result strengthens the negative message developed in the preceding chapters. Chapters 4 and 5 showed that classical learning dynamics can be exponentially slow in potential games, even when they enjoy qualitative convergence or monotonicity properties. The present chapter goes further: in the continuous constrained-optimization setting, a potential-like objective may not merely hide a long transient, but may fail to drive FTRL toward stationarity at all. Thus, potential monotonicity and no-regret structure are insufficient, by themselves, to guarantee asymptotic convergence to a meaningful first-order solution concept.

Several questions remain open. One natural direction is to identify additional regularity, curvature, or constraint qualifications under which FTRL does converge to stationary points. Another is to understand whether discrete-time FTRL, under natural learning-rate schedules, can exhibit analogous nonconvergent behavior in smooth constrained optimization. Finally, it would be interesting to characterize more precisely which features distinguish mirror descent, whose convergence behavior is much better understood, from FTRL in constrained nonconvex environments. These questions point to a broader theme: understanding learning dynamics as optimization methods requires not only regret guarantees or monotonicity identities, but also a careful analysis of how the update rule interacts with the geometry of the feasible set.

Chapter 7

Conclusion and future directions

7.1 Summary of the dissertation

This dissertation studied the computational and dynamical complexity of learning in structured games and related constrained optimization problems. The central theme was that structure alone does not guarantee efficient learning. Rather, the relevant question is whether the structure of a game or optimization problem can be converted into quantitative convergence guarantees for natural learning dynamics.

The results developed in the thesis show that the answer depends sharply on the interaction between the structure of the problem and the update rule used by the players. In rank-1 games, the algebraic structure of the payoff matrices can be exploited algorithmically: even though these games lie beyond the zero-sum setting and need not satisfy the monotonicity assumptions that underlie many first-order convergence analyses, one can still design a decentralized learning procedure that converges efficiently to an approximate Nash equilibrium. In this sense, rank-1 games provide a positive example of a structured class where nontrivial equilibrium geometry can be used to guide learning.

The situation is markedly different for potential games. Potential games are often viewed as favorable for learning because the incentives of all players are aligned with a common potential function. Nevertheless, the lower bounds proved in this dissertation show that this qualitative structure can conceal severe quantitative barriers. Fictitious play, despite its classical asymptotic convergence guarantees in potential games, can require exponentially many rounds to reach an approximate Nash equilibrium. Likewise, follow-the-regularized-leader can exhibit exponentially slow convergence in two-player potential games, especially for all commonly used regularizers. In multiplayer potential games, the lower-bound phenomenon becomes even more pronounced, with fictitious play requiring doubly exponentially many rounds in the worst case.

The final chapter moved beyond finite games and examined FTRL as a first-order method for constrained nonconvex optimization. The result is negative in a stronger sense: continuous-time FTRL may fail to converge even asymptotically to first-order stationary points. More precisely, there exist smooth constrained optimization problems for which the objective value increases monotonically along the FTRL trajectory, while the KKT gap remains bounded away from zero for all time. This shows that potential monotonicity and objective improvement, although useful, are not sufficient to guarantee stationarity for history-dependent dynamics such as FTRL.

Taken together, these results separate several notions that are often closely associated but are fundamentally distinct. A game may possess favorable structure without admitting efficient convergence under a given dynamic. A learning rule may have no-regret guarantees without yielding fast last-iterate convergence to Nash equilibrium. A potential function may increase monotonically along a trajectory without forcing the dynamics toward a stationary point. Thus, the thesis identifies a gap between qualitative convergence theory and algorithmically meaningful convergence guarantees.

7.2 Future directions

The results of this dissertation suggest several directions for future work.

A first direction is to further understand the boundary between efficiently learnable and hard structured games. Rank-1 games admit efficient learning through a parameterized zero-sum reduction, while potential games admit exponential lower bounds for classical dynamics. It would be valuable to identify additional structural assumptions that make equilibrium learning efficient beyond zero-sum and rank-1 settings. In particular, the complexity of learning in rank-k games for small $k \geq 2$, and the extent to which rank structure can be combined with other assumptions such as sparsity or regularity, remain natural questions.

A second direction concerns the design of learning dynamics for potential games. The lower bounds in this thesis apply to several canonical dynamics, including fictitious play and FTRL, but they do not rule out the possibility that other uncoupled or partially coupled procedures may converge efficiently. A central challenge is therefore to determine which additional algorithmic features are necessary to overcome the long transients constructed here. Such features might include carefully chosen perturbations, adaptive exploration, second-order information, or update rules that respond more directly to the geometry of the potential landscape.

A third direction is to clarify the relationship between regret, time-average convergence, and last-iterate convergence. No-regret learning is a powerful and robust framework, but the results of this dissertation reinforce that small regret is not the same as fast convergence to Nash equilibrium in the last iterate. Developing sharper conditions under which no-regret dynamics do imply last-iterate convergence, and understanding the quantitative price of such implications, remains an important problem at the interface of online learning and game theory.

A fourth direction is to investigate the optimization-theoretic behavior of FTRL more broadly. The nonconvergence result of the final chapter shows that FTRL is not merely a mirror-descent variant in constrained nonconvex optimization: its cumulative, history-dependent nature can lead to qualitatively different behavior. This raises several questions. Which regularizers or feasible-set geometries rule out the nonconvergent behavior constructed here? Can one characterize the constrained optimization problems for which continuous-time FTRL converges to KKT points? Are there modified FTRL dynamics that preserve the favorable regret properties of the method while recovering the stationarity guarantees associated with mirror descent or projected gradient methods?

Finally, the lower-bound constructions in this dissertation are deliberately worst-case. An important complementary direction is to understand their robustness and practical relevance. This includes studying whether the constructed slowdowns persist under perturbations, whether analogous phenomena occur in stochastic or noisy learning environments, and whether empirical benchmarks can be designed to expose the difference between asymptotic convergence and efficient convergence in structured games.

Overall, the thesis shows that the convergence behavior of learning dynamics is governed not only by the structure of the underlying game or objective, but also by the precise way in which the dynamic interacts with that structure. Understanding this interaction is essential for developing learning algorithms whose guarantees are not merely asymptotic, but genuinely algorithmic.

Bibliography

H.L Abbott and M Katchalski. On the snake in the box problem. *Journal of Combinatorial Theory, Series B*, 45(1):13–24, 1988.

Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Conference on Learning Theory (COLT)*, 2008.

Jacob Abernethy, Peter L Bartlett, and Elad Hazan. Blackwell approachability and no-regret learning are equivalent. In *COLT*, pages 27–46, 2011.

Jacob D. Abernethy, Kevin A. Lai, and Andre Wibisono. Fast convergence of fictitious play for diagonal payoff matrices. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 1387–1404. SIAM, 2021. doi: 10.1137/1.9781611976465.84. URL `https://doi.org/10.1137/1.9781611976465.84`.

Ilan Adler. The equivalence of linear programs and zero-sum games. *International Journal of Game Theory*, 42:165–177, 2013.

Ilan Adler and Renato DC Monteiro. A geometric view of parametric linear programming. *Algorithmica*, 8:161–176, 1992.

Bharat Adsul, Jugal Garg, Ruta Mehta, and Milind Sohoni. Rank-1 bimatrix games: a homeomorphism and a polynomial time algorithm. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 195–204, 2011.

Bharat Adsul, Jugal Garg, Ruta Mehta, Milind Sohoni, and Bernhard Von Stengel. Fast algorithms for rank-1 bimatrix games. *Operations Research*, 69(2):613–631, 2021.

Ioannis Anagnostides, Gabriele Farina, Ioannis Panageas, and Tuomas Sandholm. Optimistic mirror descent either converges to nash or to strong coarse correlated equilibria in bimatrix games. *Advances in Neural Information Processing Systems*, 35:16439–16454, 2022a.

Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, and Tuomas Sandholm. On last-iterate convergence beyond zero-sum games. In *International Conference on Machine Learning (ICML)*, 2022b.

Ioannis Anagnostides, Emanuel Tewolde, Brian Hu Zhang, Ioannis Panageas, Vincent Conitzer, and Tuomas Sandholm. Convergence of regret matching in potential games and constrained optimization. *arXiv:2510.17067*, 2025.

Ioannis Anagnostides, Ioannis Panageas, Nikolas Patris, and Tuomas Sandholm. (Doubly) exponential lower bounds for follow the regularized leader in potential games. In *International Conference on Machine Learning (ICML)*, 2026a.

Ioannis Anagnostides, Ioannis Panageas, Nikolas Patris, and Tuomas Sandholm. Follow the regularized leader does not converge in constrained optimization. Manuscript, 2026b.

Ioannis Anagnostides, Emanuel Tewolde, Brian Hu Zhang, Ioannis Panageas, Vincent Conitzer, and Tuomas Sandholm. Convergence of regret matching in potential games and constrained optimization. In *International Conference on Learning Representations (ICLR)*, 2026c.

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of computing*, 8(1):121–164, 2012.

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

Yakov Babichenko and Aviad Rubinstein. Communication complexity of Nash equilibrium in potential games. In *Proceedings of the Annual Symposium on Foundations of Computer Science (FOCS)*, 2020.

Yakov Babichenko and Aviad Rubinstein. Settling the complexity of Nash equilibrium in congestion games. In *Symposium on Theory of Computing (STOC)*, 2021.

Lucas Baudin and Rida Laraki. Fictitious play and best-response dynamics in identical interest and zero-sum stochastic games. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 1664–1690. PMLR, 2022. URL `https://proceedings.mlr.press/v162/baudin22a.html`.

Martino Bernasconi and Matteo Castiglioni. The complexity of min-max optimization with product constraints. *arXiv preprint arXiv:2602.04665*, 2026.

Jakub Bielawski, Thiparat Chotibut, Fryderyk Falniowski, Grzegorz Kosiorowski, Michal Misiurewicz, and Georgios Piliouras. Follow-the-regularized-leader routes to chaos in routing games. In *International Conference on Machine Learning (ICML)*, 2021.

David Blackwell. An analog of the minmax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

Avrim Blum, MohammadTaghi Hajiaghayi, Katrina Ligett, and Aaron Roth. Regret minimization and the price of total anarchy. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 2008.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Felix Brandt, Felix Fischer, and Paul Harrenstein. On the rate of convergence of fictitious play. *Theory of Computing Systems*, 53(1):41–52, 2013.

George W. Brown. Iterative solutions of games by fictitious play. In Tjalling C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 374–376. John Wiley & Sons, 1951.

Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, page eaao1733, Dec. 2017.

Noam Brown and Tuomas Sandholm. Superhuman AI for multiplayer poker. *Science*, 365 (6456):885–890, 2019.

Yang Cai, Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Weiqiang Zheng. Fast last-iterate convergence of learning in games requires forgetful algorithms. In *Neural Information Processing Systems (NeurIPS)*, 2024.

Ozan Candogan, Asuman E. Ozdaglar, and Pablo A. Parrilo. Dynamics in near-potential games. *Games and Economic Behavior*, 82:66–90, 2013.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006. ISBN 978-0-521-84108-5. doi: 10.1017/CBO9780511546921. URL `https://doi.org/10.1017/CBO9780511546921`.

Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. In *Neural Information Processing Systems (NeurIPS)*, 2020.

Yurong Chen, Xiaotie Deng, Chenchen Li, David Mguni, Jun Wang, Xiang Yan, and Yaodong Yang. On the convergence of fictitious play: A decomposition approach. In Luc De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 179–185. ijcai.org, 2022. doi: 10.24963/ijcai.2022/26. URL `https://doi.org/10.24963/ijcai.2022/26`.

Constantinos Daskalakis and Qinxuan Pan. A counter-example to karlin's strong conjecture for fictitious play. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 11–20. IEEE Computer Society, 2014. doi: 10.1109/FOCS.2014.10. URL `https://doi.org/10.1109/FOCS.2014.10`.

Constantinos Daskalakis and Ioannis Panageas. The limit points of (optimistic) gradient descent in min-max optimization. In *NeurIPS 2018*, pages 9256–9266, 2018a.

Constantinos Daskalakis and Ioannis Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. *arXiv preprint arXiv:1807.04252*, 2018b.

Constantinos Daskalakis and Christos Papadimitriou. Continuous local search. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2011.

Constantinos Daskalakis, Rafael M. Frongillo, Christos H. Papadimitriou, George Pierrakos, and Gregory Valiant. On learning algorithms for Nash equilibria. In *International Symposium on Algorithmic Game Theory (SAGT)*, 2010.

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2011.

Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training gans with optimism. In *6th International Conference on Learning Representations, ICLR 2018*. OpenReview.net, 2018.

Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. The complexity of constrained min-max optimization. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1466–1478, 2021.

Jelena Diakonikolas, Constantinos Daskalakis, and Michael I. Jordan. Efficient methods for structured nonconvex-nonconcave min-max optimization. In Arindam Banerjee and Kenji Fukumizu, editors, *The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021, April 13-15, 2021, Virtual Event*, volume 130 of *Proceedings of Machine Learning Research*, pages 2746–2754. PMLR, 2021. URL `http://proceedings.mlr.press/v130/diakonikolas21a.html`.

Kousha Etessami and Mihalis Yannakakis. On the complexity of nash equilibria and other fixed points. *SIAM Journal on Computing*, 39(6):2531–2597, 2010.

A. A. Evdokimov. Maximal length of a chain in a unit n-dimensional cube. *Matematicheskie Zametki*, 6:309–319, 1969.

Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

Zhiyuan Fan, Arnab Maiti, Kevin Jamieson, Lillian J Ratliff, and Gabriele Farina. On the universal near optimality of hedge in combinatorial settings. In *Neural Information Processing Systems (NeurIPS)*, 2025.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Faster game solving via predictive Blackwell approachability: Connecting regret matching and mirror descent. In *Conference on Artificial Intelligence (AAAI)*, 2021.

John Fearnley, Paul Goldberg, Alexandros Hollender, and Rahul Savani. The complexity of gradient descent: CLS= PPAD $\cap$ PLS. *Journal of the ACM*, 70(1):1–74, 2022.

Yoav Freund and Robert E Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103, 1999.

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

Sergiu Hart and Andreu Mas-Colell. Regret-based continuous-time dynamics. *Games and Economic Behavior*, 45(2):375–394, 2003.

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

Johannes Heinrich, Marc Lanctot, and David Silver. Fictitious self-play in extensive-form games. In Francis R. Bach and David M. Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 805–813. JMLR.org, 2015. URL `http://proceedings.mlr.press/v37/heinrich15.html`.

Amélie Héliou, Johanne Cohen, and Panayotis Mertikopoulos. Learning with bandit feedback in potential games. In *Neural Information Processing Systems (NeurIPS)*, 2017.

Chi Jin, Praneeth Netrapalli, and Michael I. Jordan. What is local optimality in nonconvex-nonconcave minimax optimization? In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 4880–4889. PMLR, 2020. URL `http://proceedings.mlr.press/v119/jin20e.html`.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71:291–307, 2005.

Ravi Kannan and Thorsten Theobald. Games of fixed rank: A hierarchy of bimatrix games. *Economic Theory*, 42:157–173, 2010.

Samuel Karlin. *Mathematical Methods and Theory in Games, Programming & Economics*. Addison-Wesley Professional, 1959.

W. H. Kautz. Unit-distance error-checking codes. *IRE Transactions on Electronic Computers*, EC-7(2):179–180, 1958. doi: 10.1109/TEC.1958.5222529.

Victor Klee. A method for constructing circuit codes. *Journal of the ACM (JACM)*, 14(3): 520–528, 1967.

Victor Klee. What is the maximum length of ad-dimensional snake? *The American Mathematical Monthly*, 77(1):63–65, 1970.

Robert Kleinberg, Georgios Piliouras, and Éva Tardos. Multiplicative updates outperform generic no-regret learning in congestion games: extended abstract. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 2009.

Galina M Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

Stefanos Leonardos, Will Overman, Ioannis Panageas, and Georgios Piliouras. Global convergence of multi-agent policy gradient in markov potential games. *arXiv preprint arXiv:2106.01969*, 2021.

Richard J Lipton, Evangelos Markakis, and Aranyak Mehta. Playing large games using simple strategies. In *Proceedings of the 4th ACM Conference on Electronic Commerce*, pages 36–41, 2003.

Nick Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

Viktor Losert and Ethen Akin. Dynamics of games and genes: Discrete versus continuous time. *Journal of Mathematical Biology*, 17(2):241–251, 1983.

Kyriakos Lotidis, Panayotis Mertikopoulos, Nicholas Bambos, and Jose Blanchet. Multi-agent learning under uncertainty: Recurrence vs. concentration. *arXiv:2512.08132*, 2025a.

Kyriakos Lotidis, Panayotis Mertikopoulos, Nicholas Bambos, and Jose Blanchet. Robust equilibria in continuous games: From strategic to dynamic robustness. *arXiv:2512.08138*, 2025b.

Chinmay Maheshwari, Manxi Wu, Druv Pai, and Shankar Sastry. Independent and decentralized learning in Markov potential games. *IEEE Transactions on Automatic Control*, 2025.

Jason R. Marden, Gürdal Arslan, and Jeff S. Shamma. Regret based dynamics: convergence in weakly acyclic games. In *Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2007.

H. Brendan McMahan. Follow-the-regularized-leader and mirror descent: Equivalence theorems and L1 regularization. In *Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

H Brendan McMahan. A survey of algorithms and analysis for adaptive online learning. *Journal of Machine Learning Research*, 18(90):1–50, 2017.

H. Brendan McMahan, Gary Holt, David Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg, Arnar Mar Hrafnkelsson, Tom Boulos, and Jeremy Kubica. Ad click prediction: a view from the trenches. In *Conference on Knowledge Discovery and Data Mining (KDD)*, 2013.

Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. *arXiv preprint arXiv:1807.02629*, 2018a.

Panayotis Mertikopoulos, Christos H. Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018*, pages 2703–2717. SIAM, 2018b.

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

Dov Monderer and Lloyd S Shapley. Fictitious play property for games with identical interests. *Journal of economic theory*, 68(1):258–265, 1996b.

Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, May 2017.

H. Moulin and J.-P. Vial. Strategically zero-sum games: The class of games whose completely mixed equilibria cannot be improved upon. *International Journal of Game Theory*, 7(3-4): 201–221, 1978.

Paul Muller, Shayegan Omidshafiei, Mark Rowland, Karl Tuyls, Julien Pérolat, Siqi Liu, Daniel Hennes, Luke Marris, Marc Lanctot, Edward Hughes, Zhe Wang, Guy Lever, Nicolas Heess, Thore Graepel, and Rémi Munos. A generalized training approach for multiagent learning. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL `https://openreview.net/forum?id=Bkl5kxrKDr`.

John F Nash. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

Arkadij Semenovič Nemirovskij and David Borisovich Yudin. Problem complexity and method efficiency in optimization. In *Wiley-Interscience*, 1983.

AS Nemirovsky, DB Yudin, and E Dawson. Wiley-interscience series in discrete mathematics, 1983.

Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2004.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

Georg Ostrovski and Sebastian van Strien. Payoff performance of fictitious play. *CoRR*, abs/1308.4049, 2013. URL `http://arxiv.org/abs/1308.4049`.

Gerasimos Palaiopanos, Ioannis Panageas, and Georgios Piliouras. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. In *Neural Information Processing Systems (NeurIPS)*, 2017.

J Jr Palis and Welington De Melo. *Geometric theory of dynamical systems: an introduction*. Springer Science & Business Media, 2012.

Ioannis Panageas, Nikolas Patris, Stratis Skoulakis, and Volkan Cevher. Exponential lower bounds for fictitious play in potential games. In *Neural Information Processing Systems (NeurIPS)*, 2023.

Nikolas Patris and Ioannis Panageas. Learning nash equilibria in rank-1 games. In *The Twelfth International Conference on Learning Representations*, 2024.

Thomas Pethick, Puya Latafat, Panos Patrinos, Olivier Fercoq, and Volkan Cevher. Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=2_vhkAMARk.

Boris T Polyak. Gradient methods for the minimisation of functionals. *USSR Computational Mathematics and Mathematical Physics*, 3(4):864–878, 1963.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Beyond regret. In *Conference on Learning Theory (COLT)*, 2011.

Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems*, 26, 2013.

Julia Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54:296–301, 1951a.

Julia Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54(2): 296–301, 1951b. ISSN 0003486X. URL http://www.jstor.org/stable/1969530.

Robert W Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.

Muhammed O. Sayin, Francesca Parise, and Asuman E. Ozdaglar. Fictitious play in zero-sum stochastic games. *SIAM J. Control. Optim.*, 60(4):2095–2114, 2022. doi: 10.1137/21m1426675. URL https://doi.org/10.1137/21m1426675.

Shai Shalev-Shwartz. *Thesis submitted for the degree of "Doctor of Philosophy"*. PhD thesis, Thesis submitted for the degree of doctor of philosophy, 2007.

Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4, 2012.

Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171 – 176, 1958.

Chaobing Song, Zhengyuan Zhou, Yichao Zhou, Yong Jiang, and Yi Ma. Optimistic dual extrapolation for coherent non-monotone variational inequalities. *Advances in Neural Information Processing Systems*, 33:14303–14314, 2020.

Brian Swenson and Soummya Kar. On the exponential rate of convergence of fictitious play in potential games. In *55th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2017, Monticello, IL, USA, October 3-6, 2017*, pages 275–279. IEEE, 2017. doi: 10.1109/ALLERTON.2017.8262748. URL https://doi.org/10.1109/ALLERTON.2017.8262748.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, pages 2989–2997, 2015.

J v. Neumann. Zur theorie der gesellschaftsspiele. *Mathematische annalen*, 100(1):295–320, 1928.

Emmanouil-Vasileios Vlatakis-Gkaragkounis, Lampros Flokas, Thanasis Lianeas, Panayotis Mertikopoulos, and Georgios Piliouras. No-regret learning and mixed nash equilibria: They do not mix. In *Neural Information Processing Systems (NeurIPS)*, 2020.

Bernhard von Stengel. Zero-sum games and linear programming duality. *Mathematics of Operations Research*, 2023.

Yuanhao Wang. Tie-breaking agnostic lower bound for fictitious play. *arXiv:2507.09902*, 2025.

Chen-Yu Wei and Haipeng Luo. More adaptive algorithms for adversarial bandits. In *Conference on Learning Theory (COLT)*, 2018.

Yaodong Yang and Jun Wang. An overview of multi-agent reinforcement learning from game theoretical perspective. *CoRR*, abs/2011.00583, 2020. URL `https://arxiv.org/abs/2011.00583`.

Brian Hu Zhang and Tuomas Sandholm. General search techniques without common knowledge for imperfect-information games, and application to superhuman fog of war chess. In *ICLR*, 2026.

Zhengyuan Zhou, Panayotis Mertikopoulos, Nicholas Bambos, Stephen P Boyd, and Peter W Glynn. On the convergence of mirror descent beyond stochastic convex programming. *SIAM Journal on Optimization*, 30(1):687–716, 2020.

Julian Zimmert and Yevgeny Seldin. Tsallis-inf: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(28):1–49, 2021.

Appendix A

Supplementary material

A.1 Properties of permutation-invariant regularizers

As we alluded to in Section 5.1, our super-exponential lower bound for `FTRL` applies to the class of permutation-invariant regularizers. As we will show shortly, this class includes essentially all of the regularizers most commonly used in practice.

Definition A.1.1 (Permutation invariance)**.** A function $\mathcal{R} : \Delta_m \to \mathbb{R}$ is called *permutation invariant* if for every permutation $\pi : [m] \to [m]$ and every $\boldsymbol{x} \in \Delta_m$,

$$\mathcal{R}(\boldsymbol{x}) = \mathcal{R}(\pi(\boldsymbol{x})),$$

where $\pi(\boldsymbol{x})[i] := \boldsymbol{x}[\pi(i)]$.

Lemma A.1.1 (Permutation-invariant function has uniform minimizer on the simplex)**.** *Let $\mathcal{R} : \Delta_m \to \mathbb{R}$ be a convex, permutation-invariant function. Then the uniform distribution $\frac{1}{m}\mathbf{1}$ is a minimizer of $\mathcal{R}$ over Δ_m. If $\mathcal{R}$ is strictly convex, $\frac{1}{m}\mathbf{1}$ is the unique minimizer.*

Proof. Let $\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x}\in\Delta_m} \mathcal{R}(\boldsymbol{x})$. By permutation invariance (Definition A.1.1), for any permutation π,

$$\mathcal{R}(\pi(\boldsymbol{x}^\star)) = \mathcal{R}(\boldsymbol{x}^\star),$$

so $\pi(\boldsymbol{x}^\star)$ is also a minimizer.

Let $\mathfrak{S}_m$ denote the set of all permutations of $\{1, \dots, m\}$. By convexity of $\mathcal{R}$ and Jensen's inequality,

$$\mathcal{R}\left(\frac{1}{|\mathfrak{S}_m|}\sum_{\pi\in\mathfrak{S}_m} \pi(\boldsymbol{x}^\star)\right) \le \frac{1}{|\mathfrak{S}_m|}\sum_{\pi\in\mathfrak{S}_m} \mathcal{R}(\pi(\boldsymbol{x}^\star)) = \mathcal{R}(\boldsymbol{x}^\star).$$

The average over all permutations of $\boldsymbol{x}^\star$—and more generally of any vector in Δ_m—is the uniform vector $\frac{1}{m}\mathbf{1}$. To see this, define

$$\overline{\boldsymbol{x}} \coloneqq \frac{1}{|\mathfrak{S}_m|}\sum_{\pi\in\mathfrak{S}_m} \pi(\boldsymbol{x}^\star),$$

and denote its jth coordinate by $\overline{\boldsymbol{x}}[j]$. Decompose $\mathfrak{S}_m$ as $\mathfrak{S}_m = \bigcup_{i=1}^m \mathfrak{S}_m^{(i)}$, where $\mathfrak{S}_m^{(i)}$ is the set of permutations mapping coordinate j to i. Clearly, $|\mathfrak{S}_m^{(i)}| = (m-1)!$ for each $i \in [m]$. Hence,

$$\begin{aligned}\overline{\boldsymbol{x}}[j] &= \frac{1}{|\mathfrak{S}_m|}\sum_{\pi\in\mathfrak{S}_m} \pi(\boldsymbol{x}^\star)[j] = \frac{1}{|\mathfrak{S}_m|}\sum_{i=1}^m |\mathfrak{S}_m^{(i)}|\, \boldsymbol{x}^\star[i] \\ &= \frac{(m-1)!}{m!}\sum_{i=1}^m \boldsymbol{x}^\star[i] = \frac{1}{m}.\end{aligned}$$

Consequently, $\overline{\boldsymbol{x}} = \frac{1}{m}\mathbf{1}$ and $\mathcal{R}(\overline{\boldsymbol{x}}) \le \mathcal{R}(\boldsymbol{x}^\star)$, so $\overline{\boldsymbol{x}}$ is also a minimizer. If $\mathcal{R}$ is strictly convex, equality holds only if $\boldsymbol{x}^\star = \overline{\boldsymbol{x}} = \frac{1}{m}\mathbf{1}$. □

Lemma A.1.2 (FTRL with a permutation-invariant regularizer preserves order)**.** *Let* $\mathcal{R} : \Delta(\mathcal{A}_i) \to \mathbb{R}$ *be a permutation-invariant, strictly convex, and differentiable regularizer, and let* $\eta > 0$*. Then, for any actions* $a_i, a_i' \in \mathcal{A}_i$ *and any round* t*, it holds that*

$$\sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i] \geq \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i'] \quad \iff \quad \boldsymbol{x}_i^{(t+1)}[a_i] \geq \boldsymbol{x}_i^{(t+1)}[a_i'].$$

Proof. Since $\mathcal{R}$ is strictly convex and differentiable, the FTRL optimizer $\boldsymbol{x}_1^{(t+1)}$ is unique and satisfies the first-order condition

$$\nabla \mathcal{R}(\boldsymbol{x}_i^{(t+1)}) = \eta \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} + \lambda \mathbf{1}$$

for some scalar λ. Hence, for any actions $a_i, a_i' \in \mathcal{A}_i$,

$$\frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i]} - \frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i']} = \eta \left(\sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i] - \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i'] \right). \tag{A.1}$$

Suppose $\sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i] \geq \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i']$. Then (A.1) gives

$$\frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i]} \geq \frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i']}. \tag{A.2}$$

We claim this implies

$$\boldsymbol{x}_i^{(t+1)}[a_i] \geq \boldsymbol{x}_i^{(t+1)}[a_i'].$$

For the sake of contradiction, assume $\boldsymbol{x}_i^{(t+1)}[a_i] < \boldsymbol{x}_i^{(t+1)}[a_i']$. Let π be the permutation that swaps a_i and a_i', and set $\boldsymbol{y} \coloneqq \pi(\boldsymbol{x}_i^{(t+1)})$. By permutation invariance, $\mathcal{R}(\boldsymbol{y}) = \mathcal{R}(\pi(\boldsymbol{x}_i^{(t+1)})) = \mathcal{R}(\boldsymbol{x}_i^{(t+1)})$; moreover, the gradient operator is linear, which implies that the corresponding

partial derivatives are swapped:

$$\frac{\partial \mathcal{R}(\boldsymbol{y})}{\partial \boldsymbol{x}_i[a_i]} = \frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i']}, \qquad \frac{\partial \mathcal{R}(\boldsymbol{y})}{\partial \boldsymbol{x}_i[a_i']} = \frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i]}. \tag{A.3}$$

Since $\mathcal{R}$ is strictly convex, we have

$$\langle \nabla \mathcal{R}(\boldsymbol{x}_i^{(t+1)}) - \nabla \mathcal{R}(\boldsymbol{y}),\, \boldsymbol{x}_i^{(t+1)} - \boldsymbol{y} \rangle > 0 \quad \text{for } \boldsymbol{x}_i^{(t+1)} \neq \boldsymbol{y}. \tag{A.4}$$

By definition of $\boldsymbol{y}$, it differs from $\boldsymbol{x}_i^{(t+1)}$ only in the coordinates a_i and a_i'. Therefore, using (A.3), the inner product in (A.4) reduces to

$$2\left(\frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i]} - \frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i']}\right)\left(\boldsymbol{x}_i^{(t+1)}[a_i] - \boldsymbol{x}_i^{(t+1)}[a_i']\right) > 0. \tag{A.5}$$

We note that the case $\boldsymbol{x} = \boldsymbol{y}$ is trivial, since the claim follows immediately from (A.5). Under our assumption $\boldsymbol{x}_i^{(t+1)}[a_i] - \boldsymbol{x}_i^{(t+1)}[a_i'] < 0$, the above inequality forces

$$\frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i]} - \frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i']} < 0,$$

contradicting (A.2). Therefore, $\boldsymbol{x}_i^{(t+1)}[a_i] \geq \boldsymbol{x}_i^{(t+1)}[a_i']$.

Conversely, suppose that $\boldsymbol{x}_i^{(t+1)}[a_i] \geq \boldsymbol{x}_i^{(t+1)}[a_i']$ but $\sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i] < \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i']$. Then (A.1) implies

$$\frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i]} < \frac{\partial \mathcal{R}(\boldsymbol{x}_i^{(t+1)})}{\partial \boldsymbol{x}_i[a_i']},$$

which, by the same argument as above, forces $\boldsymbol{x}_i^{(t+1)}[a_i] < \boldsymbol{x}_i^{(t+1)}[a_i']$, contradicting the assumption. Combining both directions establishes the equivalence. □

This class contains most standard regularizers used in online learning. For example, the negative entropy regularizer

$$\mathcal{R}(\boldsymbol{x}) = \sum_{i=1}^{m} x_i \log x_i,$$

which gives rise to multiplicative-weights-type updates, is permutation-invariant. The Euclidean regularizer

$$\mathcal{R}(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{x}\|_2^2$$

is also permutation-invariant, as are more general ℓ_p-type regularizers such as $\mathcal{R}(\boldsymbol{x}) = \frac{1}{p} \|\boldsymbol{x}\|_p^p$. Other examples include separable regularizers of the form

$$\mathcal{R}(\boldsymbol{x}) = \sum_{i=1}^{m} r(x_i),$$

for a fixed scalar function r, such as Tsallis-entropy-type regularizers or logarithmic-barrier regularizers on the relative interior of the simplex. Thus, the permutation-invariance assumption covers the canonical regularizers used in `FTRL`, while excluding only regularizers that encode an artificial preference for particular action labels.

A.2 Omitted proofs from Section 5.4

This section contains the proofs omitted from Section 5.4. We begin with the one-step improvement property of `FTRL`.

Lemma 5.4.1. *If the sequence $(\boldsymbol{x}_i^{(t)})_{t\geq 1}$ is updated using `FTRL` under a 1-strongly convex regularizer $\mathcal{R}$ with respect to some norm $\|\cdot\|$,*

$$\langle \boldsymbol{x}_i^{(t+1)}, \boldsymbol{u}_i^{(t)} \rangle - \langle \boldsymbol{x}_i^{(t)}, \boldsymbol{u}_i^{(t)} \rangle \geq \frac{1}{\eta} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|^2.$$

Proof. Let $\psi^{(t)}(\boldsymbol{x}_i) = \left\langle \boldsymbol{x}_i, \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} \right\rangle - \frac{1}{\eta}\mathcal{R}(\boldsymbol{x}_i)$. By definition, $\boldsymbol{x}_i^{(t+1)}$ maximizes $\psi^{(t)}(\boldsymbol{x}_i)$ with respect to $\mathcal{X}$. The first-order optimality condition implies

$$\langle \nabla\psi^{(t)}(\boldsymbol{x}_i^{(t+1)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i' \rangle \geq 0 \quad \forall \boldsymbol{x}_i' \in \mathcal{X}_i.$$

Substituting the gradient $\nabla\psi^{(t)}(\boldsymbol{x}_i) = \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} - \frac{1}{\eta}\nabla\mathcal{R}(\boldsymbol{x}_i)$, we have

$$\left\langle \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} - \frac{1}{\eta}\nabla\mathcal{R}(\boldsymbol{x}_i^{(t+1)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i' \right\rangle \geq 0 \quad \forall \boldsymbol{x}_i' \in \mathcal{X}_i. \tag{A.6}$$

Similarly, the optimality condition for $\boldsymbol{x}_i^{(t)}$ (which maximizes $\psi^{(t-1)}$) implies

$$\left\langle \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)} - \frac{1}{\eta}\nabla\mathcal{R}(\boldsymbol{x}_i^{(t)}), \boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i' \right\rangle \geq 0 \quad \forall \boldsymbol{x}_i' \in \mathcal{X}_i. \tag{A.7}$$

Substituting $\boldsymbol{x}_i' = \boldsymbol{x}_i^{(t)}$ in (A.6) and $\boldsymbol{x}_i' = \boldsymbol{x}_i^{(t+1)}$ in (A.7),

$$\left\langle \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} - \frac{1}{\eta}\nabla\mathcal{R}(\boldsymbol{x}_i^{(t+1)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle \geq 0, \tag{A.8}$$

$$\left\langle \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)} - \frac{1}{\eta}\nabla\mathcal{R}(\boldsymbol{x}_i^{(t)}), \boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t+1)} \right\rangle \geq 0. \tag{A.9}$$

Summing (A.8) and (A.9) yields

$$\left\langle \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}, \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle + \left\langle \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}, \boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t+1)} \right\rangle \tag{A.10}$$

$$- \frac{1}{\eta}\left\langle \nabla\mathcal{R}(\boldsymbol{x}_i^{(t+1)}) - \nabla\mathcal{R}(\boldsymbol{x}_i^{(t)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle \geq 0. \tag{A.11}$$

Since

$$\left\langle \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}, \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle + \left\langle \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}, \boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t+1)} \right\rangle = \langle \boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \rangle,$$

we have

$$\langle \boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \rangle - \frac{1}{\eta} \left\langle \nabla \mathcal{R}(\boldsymbol{x}_i^{(t+1)}) - \nabla \mathcal{R}(\boldsymbol{x}_i^{(t)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle \geq 0.$$

Since $\mathcal{R}$ is a 1-strongly convex function,

$$\left\langle \nabla \mathcal{R}(\boldsymbol{x}_i^{(t+1)}) - \nabla \mathcal{R}(\boldsymbol{x}_i^{(t)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle \geq \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|^2,$$

and the claim follows. □

We next leverage Proposition 5.4.1 to establish that simultaneous FTRL converges to a limit set in which the potential has the same value. We refer to Losert and Akin (1983) for a related argument.

Proposition A.2.1. *In an n-player potential game, the ω-limit set of simultaneous FTRL has the same potential value.*

Proof. By Proposition 5.4.1, we know that

$$\Phi(\boldsymbol{x}^{(t+1)}) - \Phi(\boldsymbol{x}^{(t)}) \geq \frac{1}{2\eta} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|^2 \geq 0.$$

Since Φ is bounded and nondecreasing along the trajectory, the sequence $\{\Phi(\boldsymbol{x}^{(t)})\}$ converges to a limit Φ^*. Let Ω be the ω-limit set—the set of accumulation points—of the sequence $\{\boldsymbol{x}^{(t)}\}$. Since $\mathcal{X}$ is compact, Ω is nonempty. By continuity, $\Phi(\boldsymbol{x}) = \Phi^*$ for all $\boldsymbol{x} \in \Omega$. Because the potential is constant on Ω, the improvement at every step must be zero. By Lemma 5.4.1, this implies that any $\boldsymbol{x} \in \Omega$ is a *fixed point* of FTRL. □

Furthermore, it is worth pointing out that if the limit point exists, it is a Nash equilibrium. Indeed, let us assume that $\boldsymbol{x} = \lim_{t \to \infty} \boldsymbol{x}^{(t)}$, so FTRL converges pointwise. For the sake of

contradiction, suppose that $\boldsymbol{x}$ is not a Nash equilibrium. This means that there exists player i, deviation $\boldsymbol{x}_i^*$, and $\epsilon > 0$ such that

$$\langle \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}), \boldsymbol{x}_i^* \rangle > \langle \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}), \boldsymbol{x}_i \rangle + \epsilon. \tag{A.12}$$

Since Φ is continuously differentiable, $\lim_{t\to\infty} \boldsymbol{u}_i^{(t)} = \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x})$. This also implies

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} = \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x})$$

. Now, the `FTRL` update for player i at time $t+1$ can be expressed as

$$\boldsymbol{x}_i^{(t+1)} = \arg\max_{\boldsymbol{x}_i' \in \mathcal{X}_i} \left(\left\langle \boldsymbol{x}_i', \frac{1}{t} \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} \right\rangle - \frac{1}{\eta t} \mathcal{R}_i(\boldsymbol{x}_i') \right).$$

By Berge's maximum theorem, this implies

$$\boldsymbol{x}_i = \lim_{t\to\infty} \boldsymbol{x}_i^{(t)} \in \arg\max_{\boldsymbol{x}' \in \mathcal{X}_i} \langle \boldsymbol{x}_i', \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}) \rangle,$$

which is a contradiction to (A.12).

In fact, this argument connecting pointwise convergence to Nash equilibria is significantly more general, and holds under any no-regret algorithm and general-sum game.

We continue with the proof of Theorem 5.4.1, which is recalled below.

Theorem 5.4.1. *Alternating ϵ-lazy no-regret dynamics converge to a 2ϵ-Nash equilibrium after at most $\exp(O_\epsilon(1/\epsilon^2))$ rounds.*

Proof. We first point out that the total number of updates is bounded by $O_\epsilon(1/\epsilon)$. This follows directly from the potential property, the fact that the updates are alternating, and the definition of ϵ-lazy updates.

Lemma A.2.1 (Number of updates)**.** *Let K denote the total number of times an update occurs under ϵ-lazy alternating no-regret dynamics, that is, $\boldsymbol{x}_i^{(t+1)} \neq \boldsymbol{x}_i^{(t)}$ for some player i. Then*

$$K \leq \frac{\Phi_{\mathsf{range}}}{\epsilon}.$$

What remains is to bound the number of rounds in between two consecutive updates.

Lemma A.2.2 (Time between updates)**.** *Let T_k be the time index of the kth update, which is assumed to be large enough so that $\mathtt{Reg}_i^{(t)} \leq \frac{1}{2}\epsilon t$ for all $t \geq T_k$. Suppose that during $[T_k, T_{k+1} - 1]$, the strategy is fixed at $\boldsymbol{x}^{(T_k)}$ and is not an 2ϵ-Nash equilibrium. Then there exists a game-dependent parameter $P > 0$ such that*

$$T_{k+1} \leq T_k \left(1 + \frac{P}{\epsilon}\right).$$

Proof. Let $t < T_{k+1}$. By definition of the no-regret property,

$$\sum_{\tau=1}^{t} \langle \boldsymbol{u}_i^{(\tau)}, \boldsymbol{x}_i' - \tilde{\boldsymbol{x}}_i^{(\tau)} \rangle \leq \mathtt{Reg}_i^{(t)}.$$

Bounding the difference in the first T_k rounds as $\sum_{\tau=1}^{T_k} \langle \boldsymbol{u}_i^{(\tau)}, \boldsymbol{x}_i' - \tilde{\boldsymbol{x}}_i^{(\tau)} \rangle \leq \sum_{\tau=1}^{T_k} \|\boldsymbol{x}_i' - \tilde{\boldsymbol{x}}_i^{(\tau)}\| \|\boldsymbol{u}_i^{(t)}\|_* \leq T_k B D_i$, we have

$$\sum_{\tau=T_k+1}^{t} \langle \boldsymbol{x}_i' - \tilde{\boldsymbol{x}}_i^{(\tau)}, \boldsymbol{u}_i^{(T_k)} \rangle \leq \mathtt{Reg}_i^{(t)} + T_k B D_i.$$

Here we used the fact that for $T_k < t < T_{k+1}$ no updates occur, so $\boldsymbol{u}_i^{(t)} = \boldsymbol{u}_i^{(T_k)}$. We now write

$$(t - T_k)\langle \boldsymbol{x}_i' - \boldsymbol{x}_i^{(T_k)}, \boldsymbol{u}_i^{(T_k)} \rangle + \sum_{\tau=T_k+1}^{t} \langle \boldsymbol{x}_i^{(T_k)} - \tilde{\boldsymbol{x}}_i^{(\tau)}, \boldsymbol{u}_i^{(T_k)} \rangle \leq T_k B D_i + \mathtt{Reg}_i^{(t)}.$$

Since $\boldsymbol{x}_i^{(\tau)}$ is not getting updated before T_{k+1}, it follows that $\langle \boldsymbol{x}_i^{(\tau)}, \boldsymbol{u}_i^{(T_k)} \rangle \geq \langle \tilde{\boldsymbol{x}}_i^{(\tau)}, \boldsymbol{u}_i^{(T_k)} \rangle - \epsilon$. Furthermore, since $\boldsymbol{x}_i^{(T_k)}$ is not a 2ϵ-Nash equilibrium, we have

$$\epsilon(t - T_k) \leq T_k B D_i + \mathtt{Reg}_i^{(t)}.$$

Using the fact that $\mathtt{Reg}_i^{(t)} \leq \frac{1}{2}\epsilon t$ and rearranging concludes the proof. □

In other words, we have shown that

$$T_{k+1} \leq T_k \left(1 + O_\epsilon\left(\frac{1}{\epsilon}\right)\right).$$

So,

$$T_k \leq \left(1 + O_\epsilon\left(\frac{1}{\epsilon}\right)\right)^k \leq \exp\left(k O_\epsilon\left(\frac{1}{\epsilon}\right)\right).$$

Combining with Lemma A.2.1, the total number of iterations needed is $\exp\left(O_\epsilon\left(\frac{1}{\epsilon^2}\right)\right)$. □

A.3 Omitted proofs from Section 5.5

We begin with the following simple but crucial observation. It shows that any pure strategy whose *cumulative utility* up to round t is smaller than the maximal *cumulative utility* by at least a constant γ is assigned, in round $t + 1$, a probability that is inversely proportional

to $\eta^{(t)}\gamma$. To simplify the exposition, we introduce the *cumulative utility gap*, which we henceforth refer to as the *gap*.

Definition A.3.1 (Cumulative utility gap)**.** For a player i, the gap of an action $a_i \in \mathcal{A}_i$ at time t is

$$\texttt{Gap}_i^{(t)}[a_i] \coloneqq \max_{a_i' \in \mathcal{A}_i} \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i'] - \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i].$$

Lemma 5.5.1. *Suppose that $(\boldsymbol{x}_i^{(t)})_{t\geq 1}$ is updated using* FTRL*. If $\texttt{Gap}_i^{(t)}[a_i] \geq \gamma > 0$ for some action $a_i \in \mathcal{A}_i$, then*

$$\boldsymbol{x}_i^{(t+1)}[a_i] \leq \frac{R}{\eta^{(t)}\gamma}, \tag{5.5}$$

where R denotes the range of the regularizer.

Proof. Define

$$\boldsymbol{x}_i' \coloneqq \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t+1)}[a_i]\boldsymbol{e}_{a_i} + \boldsymbol{x}_i^{(t+1)}[a_i]\boldsymbol{e}_{a_i'},$$

where

$$a_i' \in \arg\max_{a \in \mathcal{A}_i} \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a],$$

and $\boldsymbol{e}_a$ denotes the ath standard basis vector. Equivalently, $\boldsymbol{x}_i'$ differs from $\boldsymbol{x}_i^{(t+1)}$ only in the coordinates a_i and a_i', with $\boldsymbol{x}_i'[a_i] = 0$ and $\boldsymbol{x}_i'[a_i'] = \boldsymbol{x}_i^{(t+1)}[a_i] + \boldsymbol{x}_i^{(t+1)}[a_i']$. In particular, $\boldsymbol{x}_i' \in \Delta(\mathcal{A}_i)$.

Since $\texttt{Gap}_i^{(t)}[a_i] \geq \gamma$, we obtain

$$\begin{aligned}
\left\langle \boldsymbol{x}_i', \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} \right\rangle - \left\langle \boldsymbol{x}_i^{(t+1)}, \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} \right\rangle &= \langle \boldsymbol{x}_i' - \boldsymbol{x}_i^{(t+1)}, \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} \rangle \\
&= \boldsymbol{x}_i^{(t+1)}[a_i] \left(\sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i'] - \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i] \right) \\
&= \boldsymbol{x}_i^{(t+1)}[a_i] \; \texttt{Gap}_i^{(t)}[a_i] \\
&\geq \boldsymbol{x}_i^{(t+1)}[a_i] \; \gamma. \qquad \text{(A.13)}
\end{aligned}$$

Moreover, the regularizer has bounded range, so $\mathcal{R}(\boldsymbol{x}_i') \leq \mathcal{R}(\boldsymbol{x}_i^{(t+1)}) + R$. On the other hand, $\boldsymbol{x}_i^{(t+1)}$ is the FTRL maximizer of the objective $\langle \boldsymbol{x}_i, \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} \rangle - \frac{1}{\eta^{(t)}} \mathcal{R}(\boldsymbol{x}_i)$, and therefore its value is at least that of feasible $\boldsymbol{x}_i'$:

$$\left\langle \boldsymbol{x}_i^{(t+1)}, \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} \right\rangle - \frac{1}{\eta^{(t)}} \mathcal{R}(\boldsymbol{x}_i^{(t+1)}) \;\geq\; \left\langle \boldsymbol{x}_i', \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)} \right\rangle - \frac{1}{\eta^{(t)}} \mathcal{R}(\boldsymbol{x}_i'). \qquad \text{(A.14)}$$

Combining (A.13) and (A.14) with the bound on $\mathcal{R}(\boldsymbol{x}_i')$ yields

$$\boldsymbol{x}_i^{(t+1)}[a_i] \leq \frac{R}{\eta^{(t)} \gamma},$$

which implies the claimed bound. □

A.3.1 Definition of transition period

As explained in the main body, we partition the FTRL dynamics into *periods*. A period k is a block of consecutive rounds during which FTRL concentrates its probability mass on a small set of strategies—the *competing actions*—while all remaining actions receive negligible mass. We fix the threshold $\delta := \frac{1}{4m}$ and define periods accordingly. We now recall the basic setup.

For $k \geq 3$, let $\underline{t_k}$ and $\overline{t_k}$ denote the first and last rounds of period k, respectively, and let $T_k \coloneqq \overline{t_k} - \underline{t_k} + 1$ denote its length, so that period k corresponds to the interval $t \in [\underline{t_k}, \overline{t_k}]$.

Definition 5.5.1 (Transition period)**.** For each integer $k \in \{3, 4, \ldots, 2m-1\}$, the *transition period* k is the set of rounds t for which the following conditions hold:

(i) If $k \geq 4$ is even, then for all $t \in [\underline{t_k}, \overline{t_k}]$, $\boldsymbol{x}_2^{(t)}[a_2(k)] \geq 1 - \delta$ and $\boldsymbol{x}_1^{(t)}[a_1(k-1)] + \boldsymbol{x}_1^{(t)}[a_1(k)] \geq 1 - \delta$.

(ii) If $k \geq 5$ is odd, then for all $t \in [\underline{t_k}, \overline{t_k}]$, $\boldsymbol{x}_1^{(t)}[a_1(k)] \geq 1 - \delta$ and $\boldsymbol{x}_2^{(t)}[a_2(k-1)] + \boldsymbol{x}_2^{(t)}[a_2(k)] \geq 1 - \delta$.

Since the actions $a_1(2)$ and $a_2(2)$ are undefined, the above definition does not apply to $k = 3$. We therefore let period 3 consist only of the first round, *i.e.*, $\underline{t_3} = \overline{t_3} = 1$, and let period 4 start immediately thereafter. The remainder of the analysis relies on Property 5.5.1, which we prove by induction in Section A.3.6.

Property 5.5.1 (Period consistency)**.** Suppose both players employ `FTRL`. If a round t belongs to period k for some $3 \leq k \leq 2m-1$, then the subsequent round $t+1$ belongs either to the same period k or to the next period $k+1$.

A.3.2 Uniform initialization

We begin with the initialization, which serves as the base case for the induction in the proof of Property 5.5.1. A standard initialization for `FTRL` is the uniform strategy. In particular, for permutation-invariant regularizers, the first update ($t = 1$) selects a minimizer of $\mathcal{R}$, which is the uniform distribution. Indeed, at initialization the cumulative utility gaps satisfy

$$\mathtt{Gap}_i^{(0)}[a_i] = 0 \quad \text{for all } i \in \{1, 2\} \text{ and } a_i \in \mathcal{A}_i.$$

Thus, at $t = 1$ the FTRL update reduces to minimizing $\mathcal{R}$; by Lemma A.1.1, we obtain

$$\boldsymbol{x}_1^{(1)} = \text{Uniform}(\mathcal{A}_1), \qquad \boldsymbol{x}_2^{(1)} = \text{Uniform}(\mathcal{A}_2).$$

As we mentioned in Section 5.5, our construction builds on the class of games introduced by Panageas et al. (2023). For completeness, Section 4.4.1 reviews their recursive construction of the matrix $\mathbf{B}_{m,r}$ (*cf.* (4.4) in Definition 4.4.1) and establishes the structural properties we require. Although this class is well suited for proving exponential lower bounds for fictitious play, it is not directly applicable to FTRL: under uniform initialization, running FTRL on $(\mathbf{B}_{m,r}, \mathbf{B}_{m,r})$ for any r quickly identifies the actions attaining the maximum payoff and converges to that profile within a constant number of rounds. To preclude this behavior, we modify the payoff matrix as described below.

The underlying idea is straightforward. By Lemma 5.5.1, the gap $\texttt{Gap}_1^{(1)}[a_1]$ (resp. $\texttt{Gap}_2^{(1)}[a_2]$) at round 1 controls the probability assigned to action a_1 (resp. a_2) at round 2. Hence, by ensuring that a sufficiently large gap is incurred for all actions except $(1, 1)$ after round 1, we can guarantee that at round 2 the only action profile played with probability nearly 1 (up to δ) is $(1, 1)$.

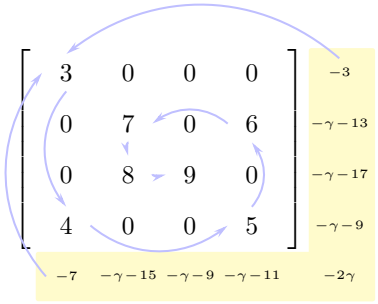
For an odd dimension m, we consider the matrix $\mathbf{B} = \mathbf{B}_{m-1,2}$ of size $(m-1) \times (m-1)$ as defined in Definition 4.4.1. Starting from the matrix $\mathbf{B}$, we embed it into a larger matrix

$$\mathbf{A} \in [-\gamma - (4m-1),\, 2m-1]^{m \times m},$$

which is obtained by adding one extra row and one extra column to $\mathbf{A}$. The transformation is illustrated in Figures A.1a and A.1b.

$$\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 7 & 0 & 6 \\ 0 & 8 & 9 & 0 \\ 4 & 0 & 0 & 5 \end{bmatrix}$$

(a) Original matrix $\mathbf{B} = \mathbf{B}_{4,2}$ of size 4×4.



(b) New matrix $\mathbf{A}$ of size 5×5.

$$\mathbf{A}[a_1, a_2] = \begin{cases} \mathbf{B}[a_1, a_2], & 1 \le a_1 \le m-1 \text{ and } 1 \le a_2 \le m-1, \\ -\sum_{a_1'=1}^{m-1} \mathbf{B}[a_1', 1], & a_1 = m \text{ and } a_2 = 1, \\ -\gamma - \sum_{a_1'=1}^{m-1} \mathbf{B}[a_1', a_2], & a_1 = m \text{ and } 2 \le a_2 \le m-1, \\ -\sum_{a_2'=1}^{m-1} \mathbf{B}[1, a_2'], & a_2 = m \text{ and } a_1 = 1, \\ -\gamma - \sum_{a_2'=1}^{m-1} \mathbf{B}[a_1, a_2'], & a_2 = m \text{ and } 2 \le a_1 \le m-1, \\ -2\,\gamma, & a_1 = m \text{ and } a_2 = m, \\ 0, & \text{otherwise.} \end{cases} \tag{A.15}$$

Lemma A.3.1 (Structural properties of $\mathbf{A}$)**.** *The following properties hold for* $\mathbf{A}$*:*

(i) *The positive entries* $\{3, 4, \ldots, 2m-1\}$ *each appear exactly once.*

(ii) $\max_{i,j} \mathbf{A}[i, j] = (2m-1)$.

(iii) *For even* $k \in \{3, 4, \ldots, 2m-2\}$, $a_1(k) = a_1(k+1)$.

(iv) *For odd* $k \in \{3, 4, \ldots, 2m-2\}$, $a_2(k) = a_2(k+1)$.

We first state a lemma on the structural properties of the generic class of payoff matrices $\mathbf{B}_{m,r}$ that will be exclusively used throughout the proof of the immediately following the next lemma.

Proposition A.3.1 (Structural properties of $\mathbf{B}_{m,r}$)**.** *Let $m \geq 2$ and $r \in \mathbb{N}$ be even. The following properties hold for $\mathbf{B}_{m,r}$:*

(i) $\max_{i,j} \mathbf{B}_{m,r}[i,j] = r + (2m-1)$.

(ii) For even $k \in \{r+1, \ldots, r+(2m-2)\}$, $a_1(k) = a_1(k+1)$.

(iii) For odd $k \in \{r+1, \ldots, r+(2m-2)\}$, $a_2(k) = a_2(k+1)$.

Proof. Item (i) is immediate. By Lemma 4.4.1, the nonzero entries of $\mathbf{B}_{m,r}$ are exactly $\{r+1, r+2, \ldots, r+(2m-1)\}$, and hence the maximum is $r+(2m-1)$.

We prove Items (ii) and (iii) simultaneously by induction on m. For $m = 2$, and since r is even, we have $a_2(r+1) = a_2(r+2) = 1$ and $a_1(r+2) = a_1(r+3) = 2$, establishing the claim.

Assume the claim holds for $m-2$ (with $m \geq 4$), and consider $\mathbf{B}_{m,r}$. By construction,

$$r+1 \text{ is at } (1,1), \quad r+2 \text{ is at } (m,1), \quad r+3 \text{ is at } (m,m), \quad r+4 \text{ is at } (2,m),$$

and the submatrix on rows/cols $\{2, \ldots, m-1\}$ is $\mathbf{B}_{m-2,r+4}$.

Fix $k \in \{r+1, \ldots, r+(2m-2)\}$.

- If $k \in \{r+1, r+2, r+3\}$, the conclusions follow by inspection:

$$a_2(r+1) = a_2(r+2) = 1, \qquad a_1(r+2) = a_1(r+3) = m, \qquad a_2(r+3) = a_2(r+4) = m.$$

Thus, for odd $k \in \{r+1, r+3\}$ we have $a_2(k) = a_2(k+1)$, and for even $k = r+2$ we have $a_1(k) = a_1(k+1)$.

- If $k = r+4$, then $r+4$ is at $(2, m)$ and $r+5$ is the top-left entry of the inner block, hence at $(2, 2)$. Therefore $a_1(r+4) = a_1(r+5) = 2$, which is exactly Item (ii) (since $r+4$ is even).

- If $k \geq r+5$, then k and $k+1$ are inside the inner block $\mathbf{B}_{m-2,r+4}$. Let $(\tilde{a}_1(\cdot), \tilde{a}_2(\cdot))$ denote the location functions within that $(m-2) \times (m-2)$ block. Because the inner block is embedded with a $(+1, +1)$ index shift,

$$a_1(t) = \tilde{a}_1(t) + 1, \qquad a_2(t) = \tilde{a}_2(t) + 1 \qquad \text{for all } \tilde{r} \in \{r+5, \ldots, r+(2m-1)\}.$$

  Applying the induction hypothesis to $\mathbf{B}_{m-2,r+4}$ yields $\tilde{a}_1(k) = \tilde{a}_1(k+1)$ when k is even, and $\tilde{a}_2(k) = \tilde{a}_2(k+1)$ when k is odd. Shifting by $+1$ gives the desired equalities for a_1, a_2 in $\mathbf{B}_{m,r}$.

This completes the induction and proves Items (ii) and (iii). □

Proof. All claims follow from the properties established in Section 4.4.1 for the embedded submatrix $\mathbf{B}_{m-1,2}$ of $\mathbf{A}$. In particular, Item (i) is an immediate consequence of Lemma 4.4.1. Moreover, by Item (i) in Proposition A.3.1, the maximum entry of $\mathbf{B}_{m-1,2}$ equals $2 + \big(2(m-1) - 1\big) = 2m - 1$. proving Item (ii). Since the embedding preserves the values of $\mathbf{B}_{m-1,2}$, and all additional entries in $\mathbf{A}$ (in particular those in the extra row and column) are chosen to be negative, this entry is also the maximum of $\mathbf{A}$. The remaining properties follow directly from Items (ii) and (iii) in Proposition A.3.1, applied to the same embedded block. □

As illustrated below, this construction ensures that, under uniform initialization, all actions other than $(1, 1)$ start with a utility gap of order $\Omega(\gamma/m)$, thereby enforcing that the action

profile $(1,1)$ is played with probability nearly 1. The remainder of the analysis relies on Property 5.5.1, which we prove by induction in Section A.3.6. In what follows, we establish Property A.3.1, which will be used throughout the analysis.

Property A.3.1. At round 1, the utility gaps satisfy

$$\texttt{Gap}_1^{(1)}[a_1] \geq \gamma^{(1)} \quad \text{and} \quad \texttt{Gap}_2^{(1)}[a_2] \geq \gamma^{(1)},$$

for all actions $a_1 \in \mathcal{A}_1$ and $a_2 \in \mathcal{A}_2$ except for action 1, which has zero gap, and $\gamma^{(1)}$ satisfies

$$\gamma^{(1)} \coloneqq \max\Big\{ \frac{4m^3\, R}{1}, 2\left(\frac{64Rm}{\delta}\right)^{\frac{1}{1-\alpha}}, \left(\frac{Rm}{\delta}\right)^{\frac{1}{1-\alpha}} \alpha^{\frac{\alpha}{1-\alpha}}(1-\alpha) \Big\} \tag{A.16}$$

Then, it also holds

$$\boldsymbol{x}_2^{(2)}[a_2(4)] \geq 1-\delta, \quad \boldsymbol{x}_1^{(2)}[a_1(3)] + \boldsymbol{x}_1^{(2)}[a_1(4)] \geq 1-\delta, \quad \text{where } \delta = \frac{1}{4m}.$$

In other words, period $k = 4$ starts at round 2, satisfying the Definition 5.5.1.

Lemma A.3.2. *Let $\boldsymbol{x}_1^{(1)} = \boldsymbol{x}_2^{(1)} = \text{Uniform}(m)$. If both players employ* `FTRL` *on the game $(\mathbf{A}, \mathbf{A})$, then Property A.3.1 is satisfied.*

Proof. From Lemma A.1.1, the `FTRL` update at round 1 dictates that both players choose the uniform strategy over their action sets, *i.e.*,

$$\boldsymbol{x}_1^{(1)} = \boldsymbol{x}_2^{(1)} = \text{Uniform}(m).$$

Although the effect of (A.15) on the utility vectors $\boldsymbol{u}_1^{(1)}, \boldsymbol{u}_2^{(1)}$ is apparent, we present a detailed derivation for completeness. We describe Player 1's utility vector; Player 2's is analogous.

1. **Row** $a_1 = 1$

$$\begin{aligned}
\boldsymbol{u}_1^{(1)}[1] &\coloneqq \sum_{a_2=1}^{m} \boldsymbol{x}_2^{(1)}[a_2]\, \mathbf{A}[1, a_2] \\
&= \frac{1}{m} \left(\sum_{a_2=1}^{m-1} \mathbf{A}[1, a_2] + \mathbf{A}[1, m] \right) \\
&= \frac{1}{m} \left(\sum_{a_2=1}^{m} \mathbf{B}[1, a_2] + \left(- \sum_{a_2'=1}^{m} \mathbf{B}[1, a_2'] \right) \right) \\
&= 0.
\end{aligned}$$

2. **Rows** $2 \leq a_1 \leq m - 1$

$$\begin{aligned}
\boldsymbol{u}_1^{(1)}[a_1] &\coloneqq \sum_{a_2=1}^{m} \boldsymbol{x}_2^{(1)}[a_2]\, \mathbf{A}[a_1, a_2] \\
&= \frac{1}{m} \left(\sum_{a_2=1}^{m-1} \mathbf{A}[a_1, a_2] + \mathbf{A}[a_1, m] \right) \\
&= \frac{1}{m} \left(\sum_{a_2=1}^{m-1} \mathbf{B}[a_1, a_2] + \left(-\gamma - \sum_{a_2'=1}^{m-1} \mathbf{B}[a_1, a_2'] \right) \right) \\
&= -\frac{\gamma}{m}.
\end{aligned}$$

3. **Row** $a_1 = m$

$$
\begin{aligned}
\boldsymbol{u}_1^{(1)}[m] &\coloneqq \sum_{a_2=1}^{m} \boldsymbol{x}_2^{(1)}[a_2]\, \mathbf{A}[m, a_2] \\
&= \frac{1}{m}\left(\mathbf{A}[m,1] + \sum_{a_2=2}^{m-1} \mathbf{A}[m, a_2] + \mathbf{A}[m,m] \right) \\
&= \frac{1}{m}\left(-\sum_{a_1'=1}^{m-1} \mathbf{B}[a_1', 1] + \sum_{a_2=2}^{m-1} \big(-\gamma - \sum_{a_1'=1}^{m-1} \mathbf{B}[a_1', a_2] \big) - 2\gamma \right) \\
&= \frac{1}{m}\left(-(m-2)\gamma - 2\gamma - \sum_{a_1'=1}^{m-1} \mathbf{B}[a_1', 1] - \sum_{a_2=2}^{m-1} \sum_{a_1'=1}^{m-1} \mathbf{B}[a_1', a_2] \right) \\
&= \frac{1}{m}\left(-m\gamma - \sum_{a_1=1}^{m-1} \sum_{a_2=1}^{m-1} \mathbf{B}[a_1, a_2] \right) \\
&= -\gamma - \frac{1}{m} \sum_{a_1=1}^{m-1} \sum_{a_2=1}^{m-1} \mathbf{B}[a_1, a_2].
\end{aligned}
$$

Only action 1 has non-negative utility at round 1, while all other actions have utility gaps of order $\Omega(\gamma/m)$, as shown below. Analogously, the same holds for Player 2's utility vector $\boldsymbol{u}_2^{(1)}$. Since γ is a free parameter that can be chosen arbitrarily large, Lemma 5.5.1 and this construction ensure that under uniform initialization, the action profile $(1, 1)$ will be played

at round 2 with probability nearly 1.

$$\mathtt{Gap}_1^{(1)}[a_1] = \begin{cases} 0 & a_1 = 1, \\ \frac{\gamma}{m} & 2 \le a_1 \le m-1, \\ \gamma + \frac{\sum_{a_1,a_2} \mathbf{B}[a_1,a_2]}{m} & a_1 = m. \end{cases}$$

$$\mathtt{Gap}_2^{(1)}[a_2] = \begin{cases} 0 & a_2 = 1, \\ \frac{\gamma}{m} & 2 \le a_2 \le m-1, \\ \gamma + \frac{\sum_{a_1,a_2} \mathbf{B}[a_1,a_2]}{m} & a_2 = m. \end{cases}$$

Setting $\gamma^{(1)} \coloneqq \frac{\gamma}{m}$, it follows that $\mathtt{Gap}_i^{(1)}[a] \ge \gamma^{(1)}$ for every $a \ne 1$. Now, by Lemma 5.5.1 and setting $\gamma^{(1)} = \frac{4m^3R}{1}$ it follows that for all $a_1 \ne 1$,

$$\boldsymbol{x}_1^{(2)}[a_1] \;\le\; \frac{R}{\eta^{(1)}\mathtt{Gap}_1^{(1)}[a_1]} \;\le\; \frac{R}{\left(\frac{1}{1^\alpha}\right)\left(\frac{4m^3R}{1}\cdot\frac{1}{m}\right)} \;=\; \frac{1}{4m^2} \;\le\; \frac{1}{4m}\frac{1}{m} = \frac{\delta}{m}.$$

and therefore

$$\boldsymbol{x}_1^{(2)}[1] \;=\; 1 - \sum_{a_1=2}^{m} \boldsymbol{x}_1^{(2)}[a_1] \;\ge\; 1 - \frac{m}{4m^2} \;\ge\; 1 - \frac{1}{4m} \;=\; 1-\delta.$$

An analogous argument applies to Player 2's strategy $\boldsymbol{x}_2^{(1)}$. Therefore, in accordance with Definition 5.5.1, period $k = 3$ starts at round 1 and terminates immediately, so period 4 begins at round 2. Indeed, at round 2 the defining conditions hold:

$$\boldsymbol{x}_1^{(2)}[a_1(3)] + \boldsymbol{x}_1^{(2)}[a_1(4)] \ge 1-\delta, \qquad \boldsymbol{x}_2^{(2)}[a_2(4)] \ge 1-\delta.$$

Finally, setting

$$\gamma^{(1)} \;=\; \max\Big\{\frac{4m^3\,R}{1}, 2\left(\frac{64Rm}{\delta}\right)^{\frac{1}{1-\alpha}}, \left(\frac{Rm}{\delta}\right)^{\frac{1}{1-\alpha}} \alpha^{\frac{\alpha}{1-\alpha}}(1-\alpha)\Big\}, \quad \text{where} \quad \delta = \frac{1}{4m}.$$

ensures that Property A.3.1 is satisfied. □

Remark A.3.1 (Normalized payoff matrix $\mathbf{A}$)**.** Our lower bound construction can be implemented with payoffs in $[-1, 1]$. Although `FTRL` is not scale invariant—rescaling payoffs changes the updates—we can modify the construction so that it does not rely on large-magnitude entries.

Specifically, assume henceforth that payoffs must lie in $[-1, 1]$. We start from the recursive matrix $\mathbf{B}_{m-1,2}$ in Definition 4.4.1. By Item (i) in Proposition A.3.1, its maximum entry equals $2m - 1$. Define the rescaled matrix $\mathbf{B} \coloneqq \frac{1}{2m-1}\mathbf{B}_{m-1,2}$, so that $\mathbf{B} \in [0, 1]^{(m-1)\times(m-1)}$. We then construct our payoff matrix $\mathbf{A}$ by embedding $\mathbf{B}$, as in the original construction, and padding it with $\lceil\gamma^{(1)}\rceil$ additional *rows* and *columns*, where $\gamma^{(1)}$ is as in (A.16). All newly introduced entries are set to -1, except that in each added row and each added column we set the first entry to 0. This should be contrasted with the initial construction in (A.15), which adds only one row and one column but uses entries of magnitude $\Theta(\gamma^{(1)})$. Since $\lceil\gamma^{(1)}\rceil = \text{poly}(m)$, the resulting game has polynomial dimension and all payoffs lie in $[-1, 1]$. It is easy to verify that this modification does not affect the subsequent arguments; consequently, the same lower bound holds: `FTRL` still requires $2^{\Omega(\text{poly}(m))}$ rounds.

A.3.3 Growth of the cumulative utility gaps

As shown in Lemma 5.5.1, to control the probability assigned to an action it suffices to track its *utility gap* over time, rather than the full cumulative utility vector.

Lemma A.3.3. *Let $k \geq 4$ be even. For any action $a_1 \in \mathcal{A}_1 \setminus \{a_1(k), a_1(k-2)\}$, the utility gap $\textit{Gap}_1^{(t)}[a_1]$ increases by at least $(1-\delta)(k-1) - \delta(2m-1)$ at every round $t \in [\underline{t_k}, \overline{t_k}]$. Similarly, if $k \geq 5$ is odd, then for any action $a_2 \in \mathcal{A}_2 \setminus \{a_2(k), a_2(k-2)\}$, the utility gap $\textit{Gap}_2^{(t)}[a_2]$ increases by at least $(1-\delta)(k-1) - \delta(2m-1)$ at every round $t \in [\underline{t_k}, \overline{t_k}]$.*

Proof. Let $a_1 \in \mathcal{A}_1 \setminus \{a_1(k), a_1(k-1)\}$ for even k. By the definition of period k, we have

$$\boldsymbol{x}_1^{(t)}[a_1(k-1)] + \boldsymbol{x}_1^{(t)}[a_1(k)] \geq 1-\delta, \qquad \boldsymbol{x}_2^{(t)}[a_2(k)] \geq 1-\delta.$$

Since during period k Player 1 primarily mixes between the actions $a_1(k-1)$ and $a_1(k)$, it follows from Lemma A.3.11 that at either $\boldsymbol{x}_1^{(t)}[a_1(k-1)]$ or $\boldsymbol{x}_1^{(t)}[a_1(k)]$ is maximal among the components of $\boldsymbol{x}_1^{(t)}$. By Lemma A.1.2, the same action also maximizes the cumulative utility at rounds $t-1$ and t. Consequently,

$$\{a_1(k-1), a_1(k)\} \supseteq \arg\max_{a_1' \in \mathcal{A}_1} \left\{ \sum_{\tau=1}^{t-1} \boldsymbol{u}_1^{(\tau)}[a_1'] \right\} \neq \emptyset.$$

Since k is even, Item (iv) in Lemma A.3.1 implies $a_2(k-1) = a_2(k)$, and so the utilities of actions $a_1(k-1)$ and $a_1(k)$ at round t satisfy

$$\boldsymbol{u}_1^{(t)}[a_1(k-1)] \geq \boldsymbol{x}_2^{(t)}[a_2(k-1)]\,(k-1) \geq (1-\delta)\,(k-1), \qquad \boldsymbol{u}_1^{(t)}[a_1(k)] \geq \boldsymbol{x}_2^{(t)}[a_2(k)]\,k \geq (1-\delta)\,k. \tag{A.17}$$

Next, consider the one-step change in the gap of action a_1 after round t:

$$\begin{aligned}\Delta\mathtt{Gap}_1^{(t)}[a_1] \;&\coloneqq\; \mathtt{Gap}_1^{(t)}[a_1] - \mathtt{Gap}_1^{(t-1)}[a_1] \\ &= \left(\max_{a_1'\in\mathcal{A}_1}\sum_{\tau=1}^{t}\boldsymbol{u}_1^{(\tau)}[a_1'] - \sum_{\tau=1}^{t}\boldsymbol{u}_1^{(\tau)}[a_1]\right) - \left(\max_{a_1'\in\mathcal{A}_1}\sum_{\tau=1}^{t-1}\boldsymbol{u}_1^{(\tau)}[a_1'] - \sum_{\tau=1}^{t-1}\boldsymbol{u}_1^{(\tau)}[a_1]\right) \\ &= \left(\max_{a_1'\in\mathcal{A}_1}\sum_{\tau=1}^{t}\boldsymbol{u}_1^{(\tau)}[a_1'] - \max_{a_1'\in\mathcal{A}_1}\sum_{\tau=1}^{t-1}\boldsymbol{u}_1^{(\tau)}[a_1']\right) - \boldsymbol{u}_1^{(t)}[a_1]. \qquad \text{(A.18)}\end{aligned}$$

Even though (during period k) one of $a_1(k-1)$ and $a_1(k)$ attains the maximal cumulative utility at rounds $t-1$ and t, the maximizer could in principle change from one round to the next. Fortunately, Lemma A.3.12 avoids a case distinction.

$$\max_{a_1'\in\mathcal{A}_1}\sum_{\tau=1}^{t}\boldsymbol{u}_1^{(\tau)}[a_1'] - \max_{a_1'\in\mathcal{A}_1}\sum_{\tau=1}^{t-1}\boldsymbol{u}_1^{(\tau)}[a_1'] \;\geq\; \min\Big\{\boldsymbol{u}_1^{(t)}[a_1(k-1)], \boldsymbol{u}_1^{(t)}[a_1(k)]\Big\}. \qquad \text{(A.19)}$$

Combining (A.18), (A.19), and (A.17), we obtain

$$\Delta\mathtt{Gap}_1^{(t)}[a_1] \;\geq\; (1-\delta)(k-1) - \boldsymbol{u}_1^{(t)}[a_1].$$

For any $a_1 \in \mathcal{A}_1 \setminus \{a_1(k), a_1(k-1)\}$, the identity $a_2(k-1) = a_2(k)$ (by Item (iv) in Lemma A.3.1) implies $\mathbf{A}[a_1, a_2(k)] = 0$; equivalently, row a_1 has no nonzero entry in column $a_2(k)$. Moreover, by the definition of period k, Player 2 assigns total probability at most δ to columns other than $a_2(k)$. Finally, Item (ii) in Lemma A.3.1 implies that every payoff outside column $a_2(k)$ is at most $2m-1$. Hence,

$$\boldsymbol{u}_1^{(t)}[a_1] \;\leq\; \delta(2m-1).$$

Therefore, due to $a_1(k-2) = a_1(k-1)$ from Item (iii) in Lemma A.3.1, it follows that for any $a_1 \in \mathcal{A}_1 \setminus \{a_1(k), a_1(k-1)\} = \mathcal{A}_1 \setminus \{a_1(k), a_1(k-2)\}$,

$$\Delta \mathtt{Gap}_1^{(t)}[a_1] \;\geq\; (1-\delta)(k-1) - \delta(2m-1).$$

The proof for odd k and for Player 2 is analogous.

□

Lemma A.3.4. *Let $k \geq 4$ be even. For any action $a_2 \in \mathcal{A}_2 \setminus \{a_2(k), a_2(k+2)\}$, the utility gap $\mathtt{Gap}_2^{(t)}[a_2]$ increases by at least $(1-\delta)(k-1) - \delta(2m-1)$ at every round $t \in [\underline{t_k}, \overline{t_k}]$. Similarly, if $k \geq 5$ is odd, then for any action $a_1 \in \mathcal{A}_1 \setminus \{a_1(k), a_1(k+2)\}$, the utility gap $\mathtt{Gap}_1^{(t)}[a_1]$ increases by at least $(1-\delta)(k-1) - \delta(2m-1)$ at every round $t \in [\underline{t_k}, \overline{t_k}]$.*

Proof. Let $a_2 \in \mathcal{A}_2 \setminus \{a_2(k), a_2(k+2)\}$ and suppose that k is even. By the definition of transition period k, for every $t \in [\underline{t_k}, \overline{t_k}]$,

$$\boldsymbol{x}_1^{(t)}[a_1(k-1)] + \boldsymbol{x}_1^{(t)}[a_1(k)] \geq 1-\delta, \qquad \boldsymbol{x}_2^{(t)}[a_2(k)] \geq 1-\delta.$$

Since $\boldsymbol{x}_2^{(t)}[a_2(k)] \geq 1-\delta$, action $a_2(k)$ has maximal probability among actions in $\mathcal{A}_2$, and thus Lemma A.1.2 implies that it also attains the maximal cumulative utility up to time $t-1$, *i.e.*,

$$a_2(k) \in \arg\max_{a_2' \in \mathcal{A}_2} \sum_{\tau=1}^{t-1} \boldsymbol{u}_2^{(\tau)}[a_2'].$$

Consequently, the one-step change in the gap of a_2 at round t satisfies

$$\begin{aligned}\Delta\mathtt{Gap}_2^{(t)}[a_2] &\coloneqq \mathtt{Gap}_2^{(t)}[a_2] - \mathtt{Gap}_2^{(t-1)}[a_2] \\ &= \left(\max_{a_2' \in \mathcal{A}_2} \sum_{\tau=1}^{t} \boldsymbol{u}_2^{(\tau)}[a_2'] - \max_{a_2' \in \mathcal{A}_2} \sum_{\tau=1}^{t-1} \boldsymbol{u}_2^{(\tau)}[a_2']\right) - \boldsymbol{u}_2^{(t)}[a_2] \\ &\geq \left(\sum_{\tau=1}^{t} \boldsymbol{u}_2^{(\tau)}[a_2(k)] - \sum_{\tau=1}^{t-1} \boldsymbol{u}_2^{(\tau)}[a_2(k)]\right) - \boldsymbol{u}_2^{(t)}[a_2] \\ &= \boldsymbol{u}_2^{(t)}[a_2(k)] - \boldsymbol{u}_2^{(t)}[a_2]. \end{aligned} \tag{A.20}$$

For the first term, the defining condition of period k yields $\boldsymbol{u}_2^{(t)}[a_2(k)] = \boldsymbol{x}_1^{(t)}[a_1(k)]\, k + \boldsymbol{x}_1^{(t)}[a_1(k-1)]\,(k-1) \geq (1-\delta)(k-1)$. For the second term, the identities $a_2(k-1) = a_2(k)$ and $a_2(k+1) = a_2(k+2)$ (by Item (iv) in Lemma A.3.1) imply that, for any $a_2 \in \mathcal{A}_2 \setminus \{a_2(k), a_2(k+2)\}$, $\mathbf{A}[a_1(k-1), a_2] = 0$ and $\mathbf{A}[a_1(k), a_2] = 0$. Moreover, by the definition of period k, Player 1 assigns total probability at most δ to actions other than $a_1(k-1)$ and $a_1(k)$, and Item (ii) in Lemma A.3.1 implies that every payoff outside these two rows is at most $2m-1$. Therefore,

$$\boldsymbol{u}_2^{(t)}[a_2] \leq \delta(2m-1). \tag{A.21}$$

Combining (A.20) and (A.21), we conclude that for any $a_2 \in \mathcal{A}_2 \setminus \{a_2(k), a_2(k+2)\}$,

$$\Delta\mathtt{Gap}_2^{(t)}[a_2] = \boldsymbol{u}_2^{(t)}[a_2(k)] - \boldsymbol{u}_2^{(t)}[a_2] \;\geq\; (1-\delta)(k-1) - \delta(2m-1).$$

This establishes the claim. The proof for odd k (and the corresponding case for Player 1) is analogous. □

A.3.4 Lower bounds on period k duration

So far, we have characterized how the gap of an action $a_i \in \mathcal{A}_i$ evolves over the rounds within period k. We now use this characterization to bound the duration of period k.

Lemma A.3.5. *Let $k \geq 4$ be even. Then*

$$T_k \;\geq\; \frac{1}{1+\delta(2m-1)}\, \mathit{Gap}_1^{(\overline{t_{k-1}})}[a_1(k)].$$

Similarly, if $k \geq 5$ is odd, then

$$T_k \;\geq\; \frac{1}{1+\delta(2m-1)}\, \mathit{Gap}_2^{(\overline{t_{k-1}})}[a_2(k)].$$

Proof. Let k be even (the odd-k case is symmetric). By definition, period k ends at the round t when $\boldsymbol{x}_1^{(t)}[a_1(k)]$ reaches the threshold $1-\delta$ (and, for odd k, once the probability mass on $\boldsymbol{x}_2^{(t)}[a_2(k)]$ reaches $1-\delta$). Before this happen, the action $a_1(k)$ needs first to become a maximizer of the cumulative gap over $\mathcal{A}_1$ (as shown in Property 5.5.1) ; this occurs at round $\hat{t}_k \in [\underline{t_k}, \overline{t_k}]$.

For every $t \in [\underline{t_k}, \hat{t}_k - 1]$, the one-step change in the utility gap of $a_1(k)$ is lower bounded by

$$\begin{aligned}
\Delta\mathtt{Gap}_1^{(t)}[a_1(k)] \;&=\; \left(\max_{a_1' \in \mathcal{A}_1} \sum_{\tau=1}^{t} \boldsymbol{u}_1^{(\tau)}[a_1'] - \max_{a_1' \in \mathcal{A}_1} \sum_{\tau=1}^{t-1} \boldsymbol{u}_1^{(\tau)}[a_1']\right) - \boldsymbol{u}_1^{(t)}[a_1(k)] \\
&=\; \left(\sum_{\tau=1}^{t} \boldsymbol{u}_1^{(\tau)}[a_2(k-1)] - \sum_{\tau=1}^{t-1} \boldsymbol{u}_1^{(\tau)}[a_2(k-1)]\right) - \boldsymbol{u}_1^{(t)}[a_1(k)] \\
&=\; \boldsymbol{u}_1^{(t)}[a_1(k-1)] - \boldsymbol{u}_1^{(t)}[a_1(k)] \\
&\geq\; \boldsymbol{x}_2^{(t)}[a_2(k-1)](k-1) - \boldsymbol{x}_2^{(t)}[a_2(k)]\, k - \boldsymbol{x}_2^{(t)}[a_2(k+1)](k+1).
\end{aligned}$$

Item (iv) in Lemma A.3.1 implies that $a_2(k-1) = a_2(k)$ for k even, and so

$$\begin{aligned}\Delta\mathtt{Gap}_1^{(t)}[a_1(k)] &\geq -\boldsymbol{x}_2^{(t)}[a_2(k)] - \boldsymbol{x}_2^{(t)}[a_2(k+1)](k+1) \\ &= -\boldsymbol{x}_2^{(t)}[a_2(k)] - \boldsymbol{x}_2^{(t)}[a_2(k+1)](2m-1) \\ &\geq -1 - \delta(2m-1),\end{aligned}$$

where the last inequality uses $\boldsymbol{x}_2^{(t)}[a_2(k+1)] \leq \delta$ and $\boldsymbol{x}_2^{(t)}[a_2(k)] \leq 1$. Equivalently, over period k the gap $\mathtt{Gap}_1^{(t)}[a_1(k)]$ can decrease by at most $1 + \delta(2m+1)$ per round. Moreover, since period k lasts at least until $\hat{t}_k$, we have $T_k \geq (\hat{t}_k - 1) - \underline{t_k} + 1$. Because the gap at the beginning of the period is finite, it follows that

$$T_k \;\geq\; \frac{1}{1+\delta(2m-1)}\,\mathtt{Gap}_1^{(\overline{t_{k-1}})}[a_1(k)],$$

which establishes the claim. The odd-k case follows analogously.

□

Lemma 5.5.2. *Let $k \geq 6$ be even. Then*

$$\mathit{Gap}_1^{(\overline{t_{k-2}})}[a_1(k)] \;\geq\; \sum_{\ell=4}^{k-2}\Big[(1-\delta)(\ell-1) - \delta(2m-1)\Big]\,T_\ell.$$

Similarly, if $k \geq 5$ is odd, then

$$\mathit{Gap}_2^{(\overline{t_{k-2}})}[a_2(k)] \;\geq\; \sum_{\ell=4}^{k-2}\Big[(1-\delta)(\ell-1) - \delta(2m-1)\Big]\,T_\ell.$$

Proof. The claim follows by repeatedly applying Lemma A.3.3. We present the argument for even $k \geq 6$; the odd-k case is analogous.

For any period index $\ell \in \{4, \dots, k-2\}$, applying Lemma A.3.3 within period ℓ yields that the regret gap of action $a_1(k)$ increases by at least

$$\Big[(1-\delta)(\ell-1) - \delta(2m-1)\Big]$$

per round throughout that period. Therefore, over the entire duration T_ℓ of period ℓ, we obtain the increment bound

$$\texttt{Gap}_1^{(\overline{t_\ell})}[a_1(k)] \;\geq\; \texttt{Gap}_1^{(\overline{t_{\ell-1}})}[a_1(k)] \;+\; \Big[(1-\delta)(\ell-1) - \delta(2m-1)\Big]\, T_\ell.$$

Summing the above inequality over $\ell = 4, \dots, k-2$ and telescoping gives

$$\texttt{Gap}_1^{(\overline{t_{k-2}})}[a_1(k)] \;\geq\; \texttt{Gap}_1^{(\overline{t_3})}[a_1(k)] \;+\; \sum_{\ell=4}^{k-2} \Big[(1-\delta)(\ell-1) - \delta(2m-1)\Big]\, T_\ell.$$

Dropping the nonnegative term $\texttt{Gap}_1^{(\overline{t_3})}[a_1(k)] = \texttt{Gap}_1^{(1)}[a_1(k)] \geq 0$ results to the stated lower bound:

$$\texttt{Gap}_1^{(\overline{t_{k-2}})}[a_1(k)] \;\geq\; \sum_{\ell=4}^{k-2} \Big[(1-\delta)(\ell-1) - \delta(2m-1)\Big]\, T_\ell.$$

The odd-k case follows analogously. □

Although Lemmas A.3.5 and 5.5.2 appear to admit a recursive application, a mismatch in their reference points prevents this directly. Specifically, Lemma 5.5.2 provides a lower bound on the regret gap only up to the end of period $k-2$, whereas Lemma A.3.5 concerns the subsequent period $k-1$. To bridge this gap, we require an auxiliary lemma that tracks the evolution of the regret gap of $a_1(k)$ from the end of $k-2$ through the end of $k-1$.

Lemma A.3.6. *Let $k \geq 4$ be even. Then,*

$$\mathit{Gap}_1^{(\overline{t_{k-1}})}[a_1(k)] \;\geq\; \mathit{Gap}_1^{(\overline{t_{k-2}})}[a_1(k)] - (1+\delta(2m-1))\,T_{k-1}.$$

Similarly, if $k \geq 5$ is odd, then

$$\mathit{Gap}_2^{(\overline{t_{k-1}})}[a_2(k)] \;\geq\; \mathit{Gap}_2^{(\overline{t_{k-2}})}[a_2(k)] - (1+\delta(2m-1))\,T_{k-1}.$$

Proof. We track the evolution of $\mathtt{Gap}_1^{(t)}[a_1(k)]$ during period $k-1$ for even k. Fix any $t \in [\underline{t_{k-1}}, \overline{t_{k-1}}]$. By the definition of transition period $k-1$, it holds that

$$\boldsymbol{x}_2^{(t)}[a_2(k-2)] + \boldsymbol{x}_2^{(t)}[a_2(k-1)] \geq 1-\delta \quad \text{and} \quad \boldsymbol{x}_1^{(t)}[a_1(k-1)] \geq 1-\delta.$$

Since Player 1 assigns probability at least $1-\delta$ to $a_1(k-1)$, Lemma A.1.2 implies that $a_1(k-1)$ attains the maximal cumulative utility in $\mathcal{A}_1$ throughout period $k-1$. Therefore, the maximizer in the definition of the gap is $a_1(k-1)$ and

$$\begin{aligned}
\Delta\mathtt{Gap}_1^{(t)}[a_1(k)] \;&\coloneqq\; \mathtt{Gap}_1^{(t)}[a_1(k)] - \mathtt{Gap}_1^{(t-1)}[a_1(k)] \\
&=\; \left(\max_{a_1' \in \mathcal{A}_1} \sum_{\tau=1}^{t} \boldsymbol{u}_1^{(\tau)}[a_1'] - \max_{a_1' \in \mathcal{A}_1} \sum_{\tau=1}^{t-1} \boldsymbol{u}_1^{(\tau)}[a_1']\right) - \boldsymbol{u}_1^{(t)}[a_1(k)] \\
&=\; \boldsymbol{u}_1^{(t)}[a_1(k-1)] - \boldsymbol{u}_1^{(t)}[a_1(k)].
\end{aligned}$$

Expanding the utilities gives

$$\Delta\mathtt{Gap}_1^{(t)}[a_1(k)] = \boldsymbol{x}_2^{(t)}[a_2(k-1)](k-1) + \boldsymbol{x}_2^{(t)}[a_2(k-2)](k-2) - \boldsymbol{x}_2^{(t)}[a_2(k)]\,k - \boldsymbol{x}_2^{(t)}[a_2(k+1)](k+1).$$

Since k is even, Item (iv) in Lemma A.3.1 implies $a_2(k-1) = a_2(k)$. We now derive a crude but sufficient bound. The terms satisfy $\boldsymbol{x}_2^{(t)}[a_2(k-2)] \geq 0$ and $\boldsymbol{x}_2^{(t)}[a_2(k-1)] \leq 1$. Moreover,

by the definition of the period, Player 2 assigns probability at most δ to any action other than $a_2(k-2), a_2(k-1)$, hence $\boldsymbol{x}_2^{(t)}[a_2(k+1)] \le \delta$. Substituting these bounds yields

$$\Delta\mathtt{Gap}_1^{(t)}[a_1(k)] \ge (k-1) - k - \delta(k+1) = -1 - \delta(k+1) \ge -1 - \delta(2m-1),$$

where the last inequality uses $k+1 \le 2m-1$. Thus, throughout period $k-1$, the gap $\mathtt{Gap}_1^{(t)}[a_1(k)]$ can decrease by at most $1+\delta(2m-1)$ per round, and therefore

$$\mathtt{Gap}_1^{(\overline{t_{k-1}})}[a_1(k)] \ge \mathtt{Gap}_1^{(\overline{t_{k-2}})}[a_1(k)] - \big(1+\delta(2m-1)\big)\, T_{k-1}.$$

The odd-k case follows by an analogous argument. □

A.3.5 Exponential lower bound

Here we combine the lemmas from the preceding subsection to obtain an exponential lower bound. In particular, we show that the combined duration of periods $2m-4$ and $2m-3$ is already exponential in m.

Lemma A.3.7 (Lower bound on T_4)**.** *Under Property A.3.1, it holds that*

$$T_4 \;\ge\; \frac{\gamma_1}{2}.$$

Proof. By Lemma A.3.5, for $k=4$ (even) we have $T_4 \ge \frac{1}{1+\delta(2m+1)}\,\mathtt{Gap}_1^{(\overline{t_3})}[a_1(4)]$. By Property A.3.1, period 3 is the single round 1, so $\mathtt{Gap}_1^{(\overline{t_3})}[a_1(4)] = \mathtt{Gap}_1^{(1)}[a_1(4)] \ge \gamma^{(1)}$. Substituting yields

$$T_4 \ge \frac{\gamma_1}{1+\delta(2m+1)}.$$

Finally, since $\delta = \frac{1}{4m}$, it follows that $1+\delta(2m+1) \le 1+\frac{2m+1}{4m} \le 2$, and therefore $T_4 \ge \gamma_1/2$, as claimed. □

Lemma 5.5.3 (Recurrence relation of $T_k + T_{k-1}$)**.** *For a small enough $\delta > 0$,*

$$T_k + T_{k-1} \;\geq\; \frac{1}{2}\sum_{\ell=4}^{k-2}(\ell-2)\,T_\ell$$

for all $k \geq 6$.

Proof. We present the argument for even k; the case of odd k is analogous. By Lemma A.3.5, the length of period k satisfies

$$T_k \;\geq\; \frac{1}{1+\delta(2m-1)}\,\mathtt{Gap}_1^{(\overline{t_{k-1}})}[a_1(k)]. \tag{A.22}$$

To lower bound the gap at time $\overline{t_{k-1}}$, we first invoke Lemma 5.5.2, which yields

$$\mathtt{Gap}_1^{(\overline{t_{k-2}})}[a_1(k)] \;\geq\; \sum_{\ell=4}^{k-2}\Big[(1-\delta)(\ell-1)-\delta(2m-1)\Big]\,T_\ell. \tag{A.23}$$

Next, to relate the gap at times $\overline{t_{k-2}}$ and $\overline{t_{k-1}}$, we apply Lemma A.3.6, obtaining

$$\mathtt{Gap}_1^{(\overline{t_{k-1}})}[a_1(k)] \;\geq\; \mathtt{Gap}_1^{(\overline{t_{k-2}})}[a_1(k)] - \big(1+\delta(2m-1)\big)\,T_{k-1}. \tag{A.24}$$

Substituting (A.23) into (A.24) and then into (A.22) gives

$$T_k \geq \frac{\sum_{\ell=4}^{k-2}\Big[(1-\delta)(\ell-1)-\delta(2m-1)\Big]\,T_\ell}{1+\delta(2m-1)} - T_{k-1}. \tag{A.25}$$

Rearranging (A.25) yields

$$T_k + T_{k-1} \;\geq\; \frac{\sum_{\ell=4}^{k-2}\Big[(1-\delta)(\ell-1)-\delta(2m-1)\Big]\,T_\ell}{1+\delta(2m-1)}. \tag{A.26}$$

By Lemma A.3.13, for $\ell \geq 4$ we have

$$(1-\delta)(\ell-1) - \delta(2m-1) \;\geq\; \ell - 2.$$

Moreover, since $\delta = \frac{1}{4m}$ (from Property A.3.1), we have

$$1 + \delta(2m-1) \;\leq\; 1 + \frac{2m-1}{4m} \;\leq\; 2.$$

Plugging these bounds into (A.26) proves that claim

$$T_k + T_{k-1} \;\geq\; \frac{1}{2}\sum_{\ell=4}^{k-2}(\ell-2)\,T_\ell.$$

This completes the proof for even k; the odd-k case follows analogously. □

Lemma A.3.8 (Exponential lower bound)**.** *Under Property A.3.1, it holds that*

$$T_{2m-3} + T_{2m-4} \;\geq\; \frac{\gamma^{(1)}}{4}\left(\frac{m-3}{e}\right)^{m-3}.$$

Proof. Define, for $3 \leq k \leq m$,

$$S_k \;\coloneqq\; T_{2k-1} + T_{2k-2}, \qquad P_k \;\coloneqq\; \sum_{\ell'=3}^{k}(\ell'-2)\,S_{\ell'}.$$

Applying Lemma 5.5.3 with index $2k-1$ gives

$$S_k = T_{2k-1} + T_{2k-2} \;\geq\; \frac{1}{2}\sum_{\ell=4}^{2k-3}(\ell-2)\,T_\ell.$$

Grouping the terms $(T_{2\ell-2}, T_{2\ell-1})$ for $\ell = 4, \ldots, k-1$ and using that all coefficients are nonnegative yields a lower bound

$$\frac{1}{2}\sum_{\ell=4}^{2k-3}(\ell-2)\,T_\ell \;\geq\; \frac{1}{2}\sum_{\ell=3}^{k-1}((2\ell-2)-2)\,(T_{2\ell-1}+T_{2\ell-2}) = \sum_{\ell'=3}^{k-1}(\ell'-2)\,S_{\ell'} = P_{k-1}.$$

Therefore, for all $k \geq 4$,

$$S_k \;\geq\; P_{k-1}. \tag{A.27}$$

It follows that for every $k \geq 4$,

$$P_k = P_{k-1} + (k-2)S_k \;\geq\; P_{k-1} + (k-2)P_{k-1} = (k-1)P_{k-1},$$

where we used (A.27). Iterating from $k = 4$ up to $k = m-2$ yields

$$P_{m-1} \;\geq\; \left(\prod_{j=4}^{m-2}(j-1)\right) P_3 = \frac{(m-3)!}{2}\,P_3 = \frac{(m-3)!}{2}\,S_3,$$

since $P_3 = (3-2)S_3 = S_3$. Finally, (A.27) with $k = m-1$ gives $S_{m-1} \geq P_{m-2}$, and $S_{m-1} = T_{2m-3} + T_{2m-4}$ by definition, so

$$T_{2m-3} + T_{2m-4} = S_{m-1} \;\geq\; P_{m-2} \;\geq\; \frac{(m-3)!}{2}\,S_3.$$

Moreover, $S_3 = T_5 + T_4 \geq T_4$, and Lemma A.3.7 gives $T_4 \geq \gamma^{(1)}/2$, proving the factorial bound.

Applying Stirling's lower bound $(m-3)! \geq \left(\frac{m-3}{e}\right)^{m-3}$ yields

$$T_{2m-3} + T_{2m-4} \;\geq\; \frac{\gamma^{(1)}}{4}\left(\frac{m-3}{e}\right)^{m-3}.$$

□

A.3.6 Period consistency

Here, we justify the definition of a *period* (Definition 5.5.1) and establish its consistency. To this end, we prove two lemmas. Throughout, we focus on the case where k is even; the case of odd k follows by an entirely analogous argument.

We show that the FTRL dynamics are consistent with the definition of a period. In particular, if round t belongs to period k, then round $t+1$ should belong only to period k, unless $\boldsymbol{x}_1^{(t+1)}[a_1(k)] \geq 1-\delta$, in which case period k terminates at round t. Hence, we have to rule out *period mixing*: since FTRL induces mixed strategies, it is not a priori clear that the update cannot move directly from period k to some period other than $k+1$. To exclude this, we proceed in two steps. First, Lemma A.3.9 shows that any action whose gap grows linearly in t receives at most δ/m probability mass under the next FTRL update. We then leverage this fact in Property 5.5.1 to preclude transitions to periods other than $k+1$.

Lemma A.3.9 (Uniformly small probability under linear gap growth)**.** *Let $a_i \in \mathcal{A}_i$ and suppose that action a increases its gap by at least 1 in each round, i.e., $\boldsymbol{Gap}_i^{(t+1)}[a_i] \geq \boldsymbol{Gap}_i^{(t)}[a_i]+1$ for $t \geq 1$. Then, it holds*

$$\boldsymbol{x}_i^{(t+1)}[a_i] \leq \frac{R}{\eta^{(t)}\, \boldsymbol{Gap}_i^{(t)}[a_i]} \leq \frac{\delta}{m}, \quad \text{for every } t \geq 1\ . \tag{A.28}$$

Proof. The proof relies only on Lemma 5.5.1, which implies that for every $t \geq 1$,

$$\boldsymbol{x}_i^{(t+1)}[a_i] \ \leq\ \frac{R}{\eta^{(t)}\, \mathtt{Gap}_i^{(t)}[a_i]}.$$

By (A.16) in Property A.3.1, the initial gap $\gamma^{(1)}$ satisfies $\mathtt{Gap}_i^{(1)}[a_i] = \gamma^{(1)} \geq \frac{Rm}{\delta}$, which ensures the desired bound on $\boldsymbol{x}_i^{(2)}[a_i]$ at round 2. To control all subsequent rounds under

$\eta^{(t)} = t^{-\alpha}$, it is enough to guarantee that

$$\mathtt{Gap}_i^{(t)}[a_i] \;\geq\; \frac{Rm}{\delta}\, t^{\alpha} \qquad \forall\, t \geq 1, \tag{A.29}$$

since this immediately implies $\boldsymbol{x}^{(t+1)}[a_i] \leq \delta/m$ for all $t \geq 1$. To establish (A.29), we use the assumed linear growth of the gap: $\mathtt{Gap}_i^{(t)}[a_i] \geq \mathtt{Gap}_i^{(1)}[a_i] + (t-1) = \gamma^{(1)} + (t-1)$for all $t \geq 1$. Thus, a sufficient condition for (A.29) is $\gamma^{(1)} + (t-1) \geq \frac{Rm}{\delta}\, t^{\alpha}$ for all $t \geq 1$. Although the linear term on the left eventually dominates the sublinear term on the right, we need the bound to hold uniformly from the start; this is ensured by $\gamma^{(1)}$ from (A.16) in Property A.3.1. Since $t - 1 \leq t$ for all $t \geq 1$, it suffices to require

$$\gamma^{(1)} \;\geq\; \frac{Rm}{\delta}\, t^{\alpha} - t \qquad \forall\, t \geq 1. \tag{A.30}$$

The tightest sufficient condition in (A.30) is obtained by maximizing the right-hand side over $t \geq 1$. Let $c \coloneqq Rm/\delta$ and define $f(t) \coloneqq ct^{\alpha} - t$ for $t \geq 1$. We view t as a continuous variable and upper bound $\sup_{t \geq 1} f(t)$ by maximizing f over $t \in [1, \infty)$. Since $\alpha \in [0, 1)$, f is concave on $(0, \infty)$ and hence attains its maximum at a unique critical point. Differentiating yields $f'(t) = c\alpha t^{\alpha - 1} - 1$, so the maximizer satisfies $t^{\star} = (c\alpha)^{\frac{1}{1-\alpha}}$. Evaluating f at $t^{\star}$, we obtain

$$f(t^{\star}) = c(t^{\star})^{\alpha} - t^{\star} = c(c\alpha)^{\frac{\alpha}{1-\alpha}} - (c\alpha)^{\frac{1}{1-\alpha}} = \left(c^{\frac{1}{1-\alpha}}\right) \alpha^{\frac{\alpha}{1-\alpha}}(1 - \alpha)$$

Define $\gamma^{\star} \coloneqq f(t^{\star})$. Then, by (A.16) in Property A.3.1, we have $\gamma^{(1)} \geq \gamma^{\star}$, and therefore (A.30) holds. This implies (A.29), and consequently

$$\boldsymbol{x}_i^{(t+1)}[a_i] \leq \frac{\delta}{m} \qquad \text{for all } t \geq 1.$$

□

Property 5.5.1 (Period consistency)**.** Suppose both players employ `FTRL`. If a round t belongs to period k for some $3 \le k \le 2m-1$, then the subsequent round $t+1$ belongs either to the same period k or to the next period $k+1$.

Proof. Assume that k is even. We prove by induction on the period index that the `FTRL` dynamics satisfy the requirements of Definition 5.5.1.

By Definition 5.5.1, period 3 consists of the single round 1, *i.e.*, $\underline{t_3} = \overline{t_3} = 1$. This case is already verified and is included purely by convention. At round 2, Property A.3.1 shows us

$$\boldsymbol{x}_1^{(2)}[a_1(3)] + \boldsymbol{x}_1^{(2)}[a_1(4)] \;\ge\; 1-\delta, \qquad \boldsymbol{x}_2^{(2)}[a_2(4)] \;\ge\; 1-\delta. \tag{A.31}$$

Hence, period 4 begins at round $\underline{t_4} = 2$, and the defining conditions are satisfied at its start.

Assume now that periods $3, 4, \ldots, k-1$ satisfy Definition 5.5.1, and consider the initial round $\underline{t_k}$ of period k. By Definition 5.5.1, we have

$$\boldsymbol{x}_2^{(\underline{t_k})}[a_2(k)] \;\ge\; 1-\delta, \tag{A.32}$$

$$\boldsymbol{x}_1^{(\underline{t_k})}[a_1(k-1)] + \boldsymbol{x}_1^{(\underline{t_k})}[a_1(k)] \;\ge\; 1-\delta. \tag{A.33}$$

It therefore suffices to show that the period-k conditions continue to hold at every round $t \in \{\underline{t_k}, \underline{t_k}+1, \ldots, \overline{t_k}\}$. Our main tools are Lemma A.3.9 and the gap-to-probability bound Lemma 5.5.1.

Player 1. By Lemma A.3.3 and Lemma A.3.13, every action $a_1 \in \mathcal{A}_1(k+2)$ increases its gap by at least $(k-2)$ at each round $t \in [1, \underline{t_k}]$. Consequently, Lemma A.3.9 implies that every $a_1 \in \mathcal{A}_1(k+2)$ receives negligible probability mass at the next update:

$$\boldsymbol{x}_1^{(\underline{t_k}+1)}[a_1] \;\le\; \frac{\delta}{m}.$$

Next, consider actions $a_1 \in [m] \setminus \mathcal{A}_1(k-2)$. Although these actions may have been active in earlier periods, they have remained *inactive* (*i.e.*, assigned probability at most δ/m) throughout the last two periods, $(k-2)$ and $(k-1)$. By Lemma A.3.3 and Lemma A.3.13, each such a_1 increases its gap by at least $(k-2)$ per round over the interval $[\underline{t_{k-2}}, \overline{t_{k-1}}]$. Moreover, Lemma A.3.5 ensures that the length of this interval is linear in the elapsed horizon:

$$T_{k-1} + T_{k-2} \;\geq\; \frac{1}{4}\sum_{\ell=3}^{k-1} T_\ell = \frac{1}{4}\,\overline{t_{k-1}}.$$

Therefore, the cumulative gap of any $a_1 \in [m] \setminus \mathcal{A}_1(k-2)$ is still linear in time, and another application of Lemma A.3.9 yields the uniform bound

$$\boldsymbol{x}_1^{(\underline{t_k}+1)}[a_1] \;\leq\; \frac{\delta}{m}.$$

It remains to control the evolution of the competing actions $a_1(k-1)$ and $a_1(k)$. By the definition of the gap, we have

$$\begin{aligned}
\Delta\mathtt{Gap}_1^{(\underline{t_k})}[a_1(k)] &\coloneqq \boldsymbol{u}_1^{(\underline{t_k})}[a_1(k-1)] - \boldsymbol{u}_1^{(\underline{t_k})}[a_1(k)] \\
&= \left(\boldsymbol{x}_2^{(\underline{t_k})}[a_2(k-1)](k-1) + \boldsymbol{x}_2^{(\underline{t_k})}[a_2(k-2)](k-2)\right) \\
&\quad - \left(\boldsymbol{x}_2^{(\underline{t_k})}[a_2(k)]k + \boldsymbol{x}_2^{(\underline{t_k})}[a_2(k+1)](k+1)\right). \qquad \text{(A.34)}
\end{aligned}$$

Since k is even, Item (iv) in Lemma A.3.1 implies $a_2(k-1) = a_2(k)$, and therefore (A.34) simplifies to

$$\Delta\mathtt{Gap}_1^{(\underline{t_k})}[a_1(k)] = -\boldsymbol{x}_2^{(\underline{t_k})}[a_2(k)] - \boldsymbol{x}_2^{(\underline{t_k})}[a_2(k+1)](k+1) + \boldsymbol{x}_2^{(\underline{t_k})}[a_2(k-2)](k-2). \quad \text{(A.35)}$$

By (A.32), $\boldsymbol{x}_2^{(t_k)}[a_2(k)] \geq 1-\delta$, so the total probability mass on columns other than $a_2(k)$ is at most δ. Using (A.35) and dropping the negative term $-\boldsymbol{x}_2^{(t_k)}[a_2(k+1)](k+1)$, we obtain

$$\begin{aligned}\Delta\mathtt{Gap}_1^{(t_k)}[a_1(k)] &\leq -\boldsymbol{x}_2^{(t_k)}[a_2(k)] + \delta(k-2) \\ &\leq -(1-\delta) + \delta(k-2) = -1 + \delta(k-1).\end{aligned}$$

Under the bounds $k \leq 2m+1$ and $\delta = \frac{1}{4m}$, we have $\delta(k-1) \leq \frac{1}{2}$, and therefore

$$\Delta\mathtt{Gap}_1^{(t_k)}[a_1(k)] \;\leq\; -\frac{1}{2}. \tag{A.36}$$

In particular, as long as $\boldsymbol{x}_2^{(t_k)}[a_2(k)] \geq 1-\delta$, the gap of $a_1(k)$ decreases by at least $1/2$ per round at time $\underline{t_k}$.

Player 2. The same reasoning applies symmetrically to Player 2. In particular, every action in $\mathcal{A}_2 \setminus \{a_2(k), a_2(k+1)\}$ continues to have probability at most δ/m by another application of Lemma A.3.9. The argument is identical to the one for Player 1 after splitting the actions into (i) the *future* actions $\mathcal{A}_2(k+2)$ and (ii) the actions *active* in *earlier* periods, namely $[m] \setminus \mathcal{A}_2(k-2)$.

The remaining nontrivial case concerns the competing actions $a_2(k)$ and $a_2(k+1)$. Since $a_2(k)$ has the largest probability mass, Lemma A.1.2 implies that it also attains the maximal cumulative utility among actions in $\mathcal{A}_2$. Thus, the one-step update of $\mathtt{Gap}^{(t_k)}[a_2(k+1)]$ is

$$\begin{aligned}\Delta\mathtt{Gap}^{(t_k)}[a_2(k+1)] \;&\coloneqq\; \boldsymbol{u}_2^{(t_k)}[a_2(k)] - \boldsymbol{u}_2^{(t_k)}[a_2(k+1)] \\ &=\; \boldsymbol{x}_1^{(t_k)}[a_1(k)]k + \boldsymbol{x}_1^{(t_k)}[a_1(k-1)](k-1) \\ &\quad - \boldsymbol{x}_1^{(t_k)}[a_1(k+1)](k+1) - \boldsymbol{x}_1^{(t_k)}[a_1(k+2)](k+2).\end{aligned}$$

Since k is even, Item (iii) in Lemma A.3.1 implies $a_1(k) = a_1(k+1)$, and therefore

$$\Delta\mathtt{Gap}^{(\underline{t_k})}[a_2(k+1)] = -\boldsymbol{x}_1^{(\underline{t_k})}[a_1(k)] + \boldsymbol{x}_1^{(\underline{t_k})}[a_1(k-1)](k-1) - \boldsymbol{x}_1^{(\underline{t_k})}[a_1(k+2)](k+2). \quad \text{(A.37)}$$

The sign of $\Delta\mathtt{Gap}^{(\underline{t_k})}[a_2(k+1)]$ is not immediate, as it depends on the relative magnitudes of $\boldsymbol{x}_1^{(\underline{t_k})}[a_1(k)]$ and $\boldsymbol{x}_1^{(\underline{t_k})}[a_1(k-1)]$, while $\boldsymbol{x}_1^{(\underline{t_k})}[a_1(k+2)]$ is uniformly small. Nevertheless, we can extract a meaningful lower bound. As long as $\boldsymbol{x}_1^{(\underline{t_k})}[a_1(k-1)] \geq \boldsymbol{x}_1^{(\underline{t_k})}[a_1(k)]$ (which holds at round $\underline{t_k}$ by the definition of the period), (A.33) implies

$$\boldsymbol{x}_1^{(\underline{t_k})}[a_1(k-1)] \;\geq\; \frac{1}{2}(1-\delta).$$

Using this in (A.37) and the bound $\boldsymbol{x}_1^{(\underline{t_k})}[a_1(k+2)] \leq \delta$ yields

$$\Delta\mathtt{Gap}^{(\underline{t_k})}[a_2(k+1)] \;\geq\; \frac{1}{2}(1-\delta)\,(k-2) - \delta(k+2).$$

Under our standing choice of parameters and Lemma A.3.13, the right-hand side is at least $(k-3)/2$. Hence, since $\boldsymbol{x}_1^{(\underline{t_k})}[a_1(k-1)] \geq \boldsymbol{x}_1^{(\underline{t_k})}[a_1(k)]$, the gap increases by at least the constant $(k-3)/2$, *i.e.*,

$$\Delta\mathtt{Gap}^{(\underline{t_k})}[a_2(k+1)] \;\geq\; \frac{(k-3)}{2}. \quad \text{(A.38)}$$

Probability mass transfer from $a_1(k-1)$ to $a_1(k)$. As in (A.38), we can show that as long as $\boldsymbol{x}_1^{(t)}[a_1(k-1)] \geq \boldsymbol{x}_1^{(t)}[a_1(k)]$, the gap of $a_2(k+1)$ increases by at least $\frac{k-3}{2}$. This inequality, however, cannot hold for the entire period. Indeed, by (A.36), the gap of $a_1(k)$ decreases by at least $\frac{1}{2}$ in every round in which $\boldsymbol{x}_2^{(t)}[a_2(k)] \geq 1-\delta$, and therefore the probability mass on $a_1(k)$ increases over time. As a result, there exists a first round $\hat{t}_k$ at

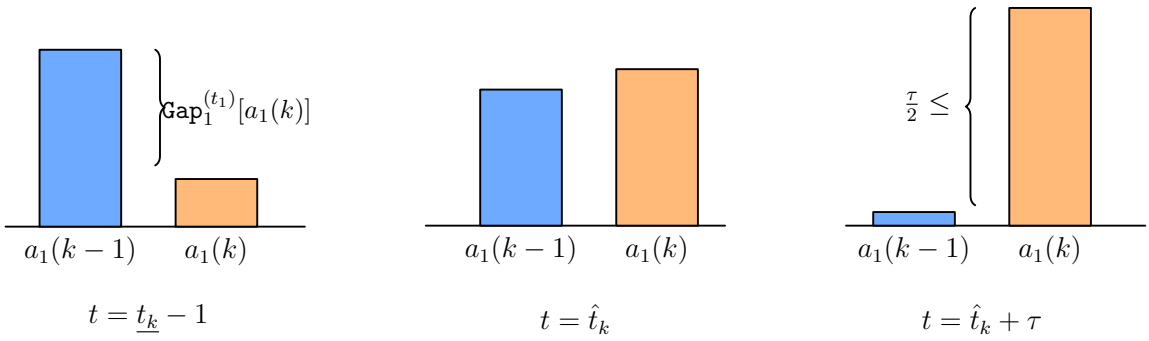


Figure A.2: Probability mass shift between the two competing actions. Transitioning from $a_1(k-1)$ to $a_1(k)$ takes a long time when the initial gap is large. Eventually $a_2(k)$ will surpass $a_1(k)$, and this will trigger a shift in the other player's cumulative utility. The consistency of the argument requires bounding the time it takes for $a_2(k)$ to be played with high probability.

which the cumulative utility of $a_1(k)$ overtakes that of $a_1(k-1)$, *i.e.*,

$$\sum_{t=1}^{\hat{t}_k} \boldsymbol{u}_1^{(t)}[a_1(k)] \;\geq\; \sum_{t=1}^{\hat{t}_k} \boldsymbol{u}_1^{(t)}[a_1(k-1)].$$

After $\hat{t}_k$, an identical argument to (A.36) implies that the gap $\mathtt{Gap}_1^{(t)}[a_1(k-1)]$ increases by at least $1/2$ per round. Our goal is to show that when $\boldsymbol{x}_1^{(t_{k+1})}[a_1(k)] \geq 1-\delta$, we still have $\boldsymbol{x}_2^{(t_{k+1})}[a_2(k)] \geq 1-\delta$. Starting from round $\hat{t}_k$, consider τ additional rounds until round $\hat{t}_k + \tau$. To track the evolution, we outline the timeline of the gaps for $a_1(k-1), a_1(k)$; see Figure A.2.

1. **Time $\underline{t_k} - 1$.** The cumulative utility of $a_1(k)$ is smaller than that of $a_1(k-1)$.

2. **Time $\hat{t}_k$.** The cumulative utility of $a_1(k)$ first exceeds that of $a_1(k-1)$, so $a_1(k)$ becomes the action of maximal cumulative utility.

3. **Time $\hat{t}_k + \tau$.** The gap of $a_1(k-1)$ has increased to at least $\frac{\tau}{2}$.

Then,

$$\mathtt{Gap}_1^{(\hat{t}_k+\tau)}[a_1(k-1)] \;\geq\; \mathtt{Gap}_1^{(\hat{t}_k)}[a_1(k-1)] + \frac{\tau}{2} \;\geq\; \frac{\tau}{2}.$$

We next bound the smallest τ such that, after τ additional rounds, $a_1(k-1)$ becomes inactive, *i.e.*, $\boldsymbol{x}_1^{(t_2+\tau+1)}[a_1(k-1)] \leq \frac{\delta}{m}$. By Lemma 5.5.1, it suffices to require

$$\boldsymbol{x}_1^{(\hat{t}_k+\tau+1)}[a_1(k-1)] \;\leq\; \frac{R}{\eta^{(\hat{t}_k+\tau)}\,\mathtt{Gap}_1^{(\hat{t}_k)}[a_1(k-1)]} \;\leq\; \frac{R}{\eta^{(\hat{t}_k+\tau)}\,(\tau/2)} \;\leq\; \frac{\delta}{m}. \tag{A.39}$$

Setting $c \coloneqq \frac{Rm}{\delta}$ and using $\eta^{(t)} = t^{-\alpha}$, (A.39) is equivalent to

$$\frac{(\hat{t}_k+\tau)^\alpha}{\tau} \;\leq\; \frac{1}{2c}. \tag{A.40}$$

A sufficient condition is obtained by bounding $(\hat{t}_k+\tau)^\alpha \leq 2^\alpha\left(\hat{t}_k^\alpha+\tau^\alpha\right) \leq 2\left(\hat{t}_k^\alpha+\tau^\alpha\right)$, for $\alpha \in [0,1)$. Substituting into (A.40) concludes the sufficient requirement $(\hat{t}_k^\alpha/\tau) + \tau^{\alpha-1} \leq (1/4c)$. To simplify the algebra, we impose the stronger pair of inequalities $\hat{t}_k^\alpha/\tau \leq 1/(8c)$ and $\tau^{\alpha-1} \leq 1/(8c)$. These are equivalent to

$$\tau \geq 8c\,\hat{t}_k^\alpha \qquad \text{and} \qquad \tau \geq (8c)^{\frac{1}{1-\alpha}},$$

since $\alpha - 1 < 0$ implies $\tau^{1-\alpha} \geq 8c$. Therefore,

$$\tau \;=\; \max\Big\{(8c)\,\hat{t}_k^\alpha,\; (8c)^{\frac{1}{1-\alpha}}\Big\}.$$

Moreover, Lemma A.3.7 together with (A.16) implies $\hat{t}_k \geq \underline{t_k} \geq T_4 \geq \gamma^{(1)}/2 \geq (64c)^{\frac{1}{1-\alpha}}$. Hence, the maximum is attained by the first term. In particular, this shows that it takes at most τ rounds for $\boldsymbol{x}_1^{(t)}[a_1(k-1)]$ to become inactive, where

$$\tau \coloneqq (8c)\,\hat{t}_k^\alpha = \left(\frac{8\,Rm}{\delta}\right)\hat{t}_k^\alpha. \tag{A.41}$$

It remains to show that, even after these τ rounds, Player 2 still assigns probability at least $1-\delta$ to action $a_2(k)$. As argued previously for Player 2, every action other than $a_2(k+1)$ accumulates a gap that is linear in $\hat{t}_k + \tau$; hence, by Lemma A.3.9, each such action remains assigned probability at most δ/m. Therefore, if $\boldsymbol{x}_2^{(\hat{t}_k+\tau)}[a_2(k)]$ were to drop below $1-\delta$, then only $a_2(k+1)$ could have increased its probability.

To this end, we lower bound the gap of the competing action $a_2(k+1)$. As noted above, the implication in (A.38) does not hold after round $\hat{t}_k$. We therefore bound the worst-case one-step change in the gap of $a_2(k+1)$. From (A.37), for any $t \geq \hat{t}_k$,

$$\begin{aligned}
\Delta\texttt{Gap}^{(t)}[a_2(k+1)] &= -\boldsymbol{x}_1^{(t)}[a_1(k)] + \boldsymbol{x}_1^{(t)}[a_1(k-1)](k-1) - \boldsymbol{x}_1^{(t)}[a_1(k+2)](k+2) \\
&\geq -\boldsymbol{x}_1^{(t)}[a_1(k)] - \boldsymbol{x}_1^{(t)}[a_1(k+2)](k+2) \\
&\geq -1 - (\delta/m)(k+2) \geq -2. \qquad \text{(A.42)}
\end{aligned}$$

By Lemma 5.5.2, (A.38) and (A.42), we obtain

$$
\begin{aligned}
\mathtt{Gap}_2^{(\hat{t}_k+\tau)}[a_2(k+1)] &= \mathtt{Gap}_2^{(\hat{t}_k)}[a_2(k+1)] + \sum_{t=\hat{t}_k+1}^{\hat{t}_k+\tau} \Delta\mathtt{Gap}^{(t)}[a_2(k+1)] \\
&\geq \mathtt{Gap}_2^{(\underline{t_k}-1)}[a_2(k+1)] \\
&\quad + \sum_{t=\underline{t_k}}^{\hat{t}_k} \Delta\mathtt{Gap}^{(t)}[a_2(k+1)] + \sum_{t=\hat{t}_k+1}^{\hat{t}_k+\tau} (-2) && \text{(From (A.42))} \\
&\geq \mathtt{Gap}_2^{(\underline{t_k}-1)}[a_2(k+1)] + \sum_{t=\underline{t_k}}^{\hat{t}_k} \frac{(k-3)}{2} + \tau(-2) && \text{(From (A.38))} \\
&\geq \sum_{\ell=4}^{k-1} (\ell-2)T_\ell + \frac{(\hat{t}_k - \underline{t_k}+1)}{2} - 2\tau && \text{(From (5.5.2))} \\
&\geq \sum_{\ell=4}^{k-1} T_\ell + \frac{(\hat{t}_k - \underline{t_k}+1)}{2} - 2\tau \\
&\geq \frac{\underline{t_k}-1}{2} + \frac{(\hat{t}_k - \underline{t_k}+1)}{2} - 2\tau \\
&= \frac{\hat{t}_k}{2} - (16c)\,\hat{t}_k^{\alpha}, && \text{(A.43)}
\end{aligned}
$$

where in the last step we used $\tau = (8c)\,\hat{t}_k^{\alpha}$ from (A.41).

The correction term $(8c)\,\hat{t}_k^{\alpha}$ is sublinear in $\hat{t}_k$ for $\alpha \in [0,1)$, so after a particular round the linear term should dominates the second term. It suffices to find t such as the following is true

$$
\frac{t}{2} - (16c)\,t^{\alpha} \;\geq\; \frac{t}{4}, \tag{A.44}
$$

which is equivalent to $\frac{t}{4} \geq (16c)\,t^{\alpha} \Leftrightarrow t \geq (64c)^{\frac{1}{1-\alpha}}$. Therefore, (A.44) holds whenever $\hat{t}_k \geq (64c)^{\frac{1}{1-\alpha}}$. Moreover, due to Lemma A.3.7 and (A.16)

$$
\hat{t}_k \geq \sum_{\ell=3}^{k-1} T_\ell \geq T_4 \geq \frac{\gamma^{(1)}}{2} \geq (64c)^{\frac{1}{1-\alpha}}, \tag{A.45}
$$

so from (A.43) we conclude

$$\texttt{Gap}_2^{(\hat{t}_k+\tau)}[a_2(k+1)] \;\geq\; \frac{\hat{t}_k}{4}. \tag{A.46}$$

The final step is to verify that $\boldsymbol{x}_2^{(\hat{t}_k+\tau)}[a_2(k+1)] \leq \delta/m$. This ensures that the conditions of Definition 5.5.1 are satisfied, and hence periods do not blend despite the propensity of FTRL to mix actions; in particular, period k can only be followed by period $k+1$. To this end, we invoke Lemma 5.5.1 together with (A.46).

$$\begin{aligned}
\boldsymbol{x}_2^{(\hat{t}_k+\tau)}[a_2(k+1)] \;&\leq\; \frac{R}{\eta^{(\hat{t}_k+\tau)}\texttt{Gap}_2^{(\hat{t}_k+\tau)}[a_2(k+1)]} \\
&\leq\; \frac{R}{(\hat{t}_k+\tau)^{-\alpha}(\hat{t}_k/4)} \\
&= \frac{4R\,(\hat{t}_k+\tau)^{\alpha}}{\hat{t}_k} \leq \frac{\delta}{m}
\end{aligned} \tag{A.47}$$

Hence, it suffices to show that $\frac{(\hat{t}_k+\tau)^\alpha}{\hat{t}_k} \leq \frac{1}{4c}$, where $c = \frac{Rm}{\delta}$. From (A.40), we already have $\frac{(\hat{t}_k+\tau)^\alpha}{\tau} \leq \frac{1}{2c}$. Therefore,

$$\frac{(\hat{t}_k+\tau)^\alpha}{\hat{t}_k} = \frac{(\hat{t}_k+\tau)^\alpha}{\tau}\frac{\tau}{\hat{t}_k} \leq \frac{1}{2c}\frac{\tau}{\hat{t}_k}.$$

Thus, it remains to ensure $\tau/\hat{t}_k \leq 1/2$, which, in turn, implies (A.47) and concludes the proof. By (A.41), $\tau = (8c)\,\hat{t}_k^{\alpha}$, and hence $\tau/\hat{t}_k = 8c\,\hat{t}_k^{-(1-\alpha)} = 8c/\hat{t}_k^{1-\alpha}$. Therefore, the condition $\tau/\hat{t}_k \leq 1/2$ is equivalent to

$$\frac{\tau}{\hat{t}_k} \leq \frac{1}{2} \iff \frac{8c}{\hat{t}_k^{1-\alpha}} \leq \frac{1}{2} \iff \hat{t}_k^{1-\alpha} \geq 16c \iff \hat{t}_k \geq (16c)^{\frac{1}{1-\alpha}},$$

where the last step uses $1-\alpha > 0$. Since (A.45) guarantees $\hat{t}_k \geq (64c)^{\frac{1}{1-\alpha}}$, the above condition is satisfied; hence (A.47) holds as well, completing the proof. □

A.3.7 Proof of Theorem 5.2.2

Having established in Property 5.5.1 that the FTRL iterates traverse the periods sequentially (from $k = 3$ up to $k = 2m - 1$), we can now invoke the preceding results to prove the main theorem of Section 5.5, namely Theorem 5.2.2. We begin with a lemma showing that, in every period except the last, the FTRL iterates do not reach an ϵ-Nash equilibrium for any $\epsilon \leq \frac{1}{8m}$.

Throughout, we work with the identical-payoff game $(\mathbf{A}, \mathbf{A})$ defined in Section A.3.2, and we fix δ and $\gamma^{(1)}$ as in Property A.3.1.

Lemma A.3.10. *For every $k \in \{3, \ldots, 2m-2\}$, throughout period k the FTRL iterates $(\boldsymbol{x}_1^{(t)}, \boldsymbol{x}_2^{(t)})$ are not an ϵ-Nash equilibrium for $\epsilon = \frac{1}{8m}$.*

Proof. Fix a round t in period k, where $3 \leq k \leq 2m-2$ and k is even. By Definition 5.5.1,

$$\boldsymbol{x}_1^{(t)}[a_1(k-1)] + \boldsymbol{x}_1^{(t)}[a_1(k)] \geq 1 - \delta, \qquad \boldsymbol{x}_2^{(t)}[a_2(k)] \geq 1 - \delta.$$

Then,

$$\sum_{a_1 \notin \{a_1(k-1), a_1(k)\}} \boldsymbol{x}_1^{(t)}[a_1] \leq \delta, \qquad \sum_{a_2 \neq a_2(k)} \boldsymbol{x}_2^{(t)}[a_2] \leq \delta.$$

Using the uniform bound $\mathbf{A}[a_1, a_2] \leq 2m-1$ from Item (ii) in Lemma A.3.1, we upper bound the realized payoff as

$$\begin{aligned}\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{x}_2^{(t)} \rangle &= \sum_{a_1, a_2} \boldsymbol{x}_1^{(t)}[a_1] \boldsymbol{x}_2^{(t)}[a_2] \mathbf{A}[a_1, a_2] \\ &\leq \delta^2(2m-1) + \boldsymbol{x}_2^{(t)}[a_2(k)] \Big(\boldsymbol{x}_1^{(t)}[a_1(k-1)]\, \mathbf{A}[a_1(k-1), a_2(k)] \\ &\qquad + \boldsymbol{x}_1^{(t)}[a_1(k)]\, \mathbf{A}[a_1(k), a_2(k)] \Big). \end{aligned} \tag{A.48}$$

Since k is even, Item (iv) in Lemma A.3.1 implies $a_2(k-1) = a_2(k)$, and hence

$$\mathbf{A}[a_1(k-1), a_2(k)] = \mathbf{A}[a_1(k-1), a_2(k-1)] = k-1.$$

Substituting this identity and $\mathbf{A}[a_1(k), a_2(k)] = k$ into (A.48) yields

$$\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{x}_2^{(t)} \rangle \leq \delta^2(2m-1) + \boldsymbol{x}_2^{(t)}[a_2(k)] \left(\boldsymbol{x}_1^{(t)}[a_1(k-1)]\,(k-1) + \boldsymbol{x}_1^{(t)}[a_1(k)]\,k \right). \tag{A.49}$$

Consider Player 1 deviating to the pure action $a_1(k)$. By definition, $\mathbf{A}[a_1(k), a_2(k)] = k$ and all other entries, but the mth entry, are nonnegative, hence

$$\left\langle \boldsymbol{e}_{a_1(k)}, \mathbf{A}\boldsymbol{x}_2^{(t)} \right\rangle \geq \boldsymbol{x}_2^{(t)}[a_2(k)]k + \boldsymbol{x}_2^{(t)}[m]\ \left(-2\gamma^{(1)}\right). \tag{A.50}$$

The probability assigned to the mth action can easily be bounded. The key observation is that the gap of action m grows $\gamma^{(1)}$ times faster than the gap of any action whose gap increases linearly with time. In particular,

$$\mathtt{Gap}_2^{(t-1)}[m]\ \geq\ \gamma^{(1)}\,\mathtt{Gap}_2^{(t-1)}[a_2(2m-1)].$$

By Lemma 5.5.1, we have

$$\boldsymbol{x}_2^{(t)}[m] \leq \frac{R}{\eta^{(t-1)}\,\mathtt{Gap}_2^{(t-1)}[m]} \leq \frac{R}{\eta^{(t-1)}\,\gamma^{(1)}\,\mathtt{Gap}_2^{(t-1)}[a_2(2m-1)]} \leq \frac{1}{\gamma^{(1)}} \cdot \frac{\delta}{m}, \tag{A.51}$$

where the last inequality uses Lemma A.3.9 applied to $a_2(2m-1)$, whose gap is linear in $t-1$.

Combining this bound with (A.50) yields

$$\langle \boldsymbol{e}_{a_1(k)}, \mathbf{A}\boldsymbol{x}_2^{(t)} \rangle \;\geq\; \boldsymbol{x}_2^{(t)}[a_2(k)]\, k - \frac{2\delta}{m}. \tag{A.52}$$

Since $\boldsymbol{x}_1^{(t)}[a_1(k-1)] + \boldsymbol{x}_1^{(t)}[a_1(k)] \leq 1$, (A.49) can be rewritten as

$$\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{x}_2^{(t)} \rangle \leq \delta^2(2m-1) + \boldsymbol{x}_2^{(t)}[a_2(k)]\Big((k-1) + \boldsymbol{x}_1^{(t)}[a_1(k)]\Big).$$

Subtracting it from (A.52) yields

$$\begin{aligned}\Big\langle \boldsymbol{e}_{a_1(k)}, \mathbf{A}\boldsymbol{x}_2^{(t)} \Big\rangle - \Big\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{x}_2^{(t)} \Big\rangle \geq -\delta^2(2m-1) - \frac{2\delta}{m} \\ + \boldsymbol{x}_2^{(t)}[a_2(k)] \left(1 - \boldsymbol{x}_1^{(t)}[a_1(k)]\right).\end{aligned} \tag{A.53}$$

If $\boldsymbol{x}_1^{(t)}[a_1(k)] \leq 1 - \frac{1}{k+1}$, then $1 - \boldsymbol{x}_1^{(t)}[a_1(k)] \geq \frac{1}{k+1}$ and $\boldsymbol{x}_2^{(t)}[a_2(k)] \geq 1 - \delta$, so

$$\Big\langle \boldsymbol{e}_{a_1(k)}, \mathbf{A}\boldsymbol{x}_2^{(t)} \Big\rangle - \Big\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{x}_2^{(t)} \Big\rangle \geq -\delta^2(2m-1) - \frac{2\delta}{m} + (1-\delta)\frac{1}{k+1}. \tag{A.54}$$

Substituting $\delta = \frac{1}{4m}$ and using $k+1 \leq 2m$ yields

$$(1-\delta)\frac{1}{k+1} \;\geq\; (1-\delta)\frac{1}{2m}.$$

Hence

$$\begin{aligned}
-\delta^2(2m-1)-\frac{2\delta}{m}+(1-\delta)\frac{1}{k+1} &\geq -\delta^2(2m-1)-\frac{2\delta}{m}+(1-\delta)\frac{1}{2m} \\
&= -\frac{2m-1}{16m^2}-\frac{1}{2m^2}+\Big(1-\frac{1}{4m}\Big)\frac{1}{2m} \\
&= -\frac{2m-1}{16m^2}-\frac{1}{2m^2}+\frac{1}{2m}-\frac{1}{8m^2} \\
&= \frac{3}{8m}-\frac{9}{16m^2}=\frac{6m-9}{16m^2}\geq\frac{1}{8m}, \qquad (\text{for } m>2)
\end{aligned}$$

$$\Big\langle \boldsymbol{e}_{a_1(k)}, \mathbf{A}\boldsymbol{x}_2^{(t)}\Big\rangle - \Big\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{x}_2^{(t)}\Big\rangle \geq \frac{1}{8m}. \tag{A.55}$$

Otherwise, $\boldsymbol{x}_1^{(t)}[a_1(k)] > 1-\frac{1}{k+1}$, and therefore

$$\boldsymbol{x}_1^{(t)}[a_1(k-1)] \;\leq\; 1-\boldsymbol{x}_1^{(t)}[a_1(k)] \;<\; \frac{1}{k+1}.$$

Consider player 2 deviating to $a_2(k+1)$. Since k is even, Item (iii) (in Lemma A.3.1) implies $a_1(k)=a_1(k+1)$. Hence

$$\mathbf{A}[a_1(k),a_2(k+1)] = \mathbf{A}[a_1(k+1),a_2(k+1)] = k+1.$$

Moreover, all payoff entries appearing in the deviation comparison are nonnegative, except the mth entry, so

$$\Big\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{e}_{a_2(k+1)}\Big\rangle \geq \boldsymbol{x}_1^{(t)}[a_1(k)]\,(k+1) + \boldsymbol{x}_1^{(t)}[m]\,\left(-2\gamma^{(1)}\right). \tag{A.56}$$

As before, we can bound $\boldsymbol{x}_1^{(t)}[m] \le \delta/(\gamma^{(1)}m)$ (analogously to the bound on $\boldsymbol{x}_2^{(t)}[m]$). Using this fact, subtracting (A.49) from $\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{e}_{a_2(k+1)}\rangle$, and using $\boldsymbol{x}_2^{(t)}[a_2(k)] \le 1$, we obtain

$$\begin{aligned}\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{e}_{a_2(k+1)}\rangle - \langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{x}_2^{(t)}\rangle &\ge \boldsymbol{x}_1^{(t)}[a_1(k)](k+1) - \frac{2\delta}{m} - \delta^2(2m-1) \\ &\quad - \left(\boldsymbol{x}_1^{(t)}[a_1(k-1)](k-1) + \boldsymbol{x}_1^{(t)}[a_1(k)]k\right) \\ &= -\delta^2(2m-1) - \frac{2\delta}{m} + \boldsymbol{x}_1^{(t)}[a_1(k)] \\ &\quad - (k-1)\boldsymbol{x}_1^{(t)}[a_1(k-1)]. \end{aligned} \tag{A.57}$$

Using $\boldsymbol{x}_1^{(t)}[a_1(k)] > 1 - \frac{1}{k+1}$ and $\boldsymbol{x}_1^{(t)}[a_1(k-1)] \le \frac{1}{k+1}$ yields $\boldsymbol{x}_1^{(t)}[a_1(k)] - (k-1)\boldsymbol{x}_1^{(t)}[a_1(k-1)] \ge \frac{1}{k+1}$, and hence from (A.57)

$$\begin{aligned}\left\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{e}_{a_2(k+1)}\right\rangle - \left\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{x}_2^{(t)}\right\rangle &\ge -\delta^2(2m-1) - \frac{2\delta}{m} + \frac{1}{k+1} \\ &\ge -\delta^2(2m-1) - \frac{2\delta}{m} + \frac{(1-\delta)}{k+1}. \end{aligned} \tag{A.58}$$

In (A.54) and (A.55) we lower bounded the right-hand side by $1/(8m)$. The same bound applies to (A.58), yielding

$$\left\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{e}_{a_2(k+1)}\right\rangle - \left\langle \boldsymbol{x}_1^{(t)}, \mathbf{A}\boldsymbol{x}_2^{(t)}\right\rangle \ \ge \ \frac{1}{8m}. \tag{A.59}$$

Combining (A.55) and (A.59), every round t of every even period $k \le 2m-2$ admits a unilateral deviation that improves the payoff by at least $1/(8m)$. The odd-k case is analogous (with the roles of the players swapped), so no iterate that lies in a period $k \le 2m-2$ can be an ϵ-Nash equilibrium for any $\epsilon \le \frac{1}{8m}$. □

We now prove the formal version of Theorem 5.2.2 stated in Section 5.2.

Theorem A.3.1. *Consider the identical-payoff game* $(\mathbf{A}, \mathbf{A})$*, and suppose both players run* `FTRL` *with a permutation-invariant regularizer and learning rate* $\eta^{(t)} = t^{-\alpha}$ *for* $\alpha \in [0, 1)$*. Then, for any* $\epsilon \leq \frac{1}{8m}$*, the* `FTRL` *iterates do not constitute an* ϵ*-Nash equilibrium for at least* $2^{\Omega(m \log m)}$ *rounds.*

Proof. The claim follows by combining Property 5.5.1, Lemma A.3.10, and Lemma A.3.8. By Property 5.5.1, the `FTRL` iterates sequentially pass through periods $k = 3, \ldots, 2m - 3$. For every round t occurring in these periods, Lemma A.3.10 guarantees the existence of a unilateral deviation that increases payoff by at least $1/(8m)$. Hence, at every such round the iterate fails to be an ϵ-Nash equilibrium for any $\epsilon \leq 1/(8m)$.

It remains to lower bound the time elapsed by the end of period $2m - 3$. By Lemma A.3.8, the combined duration of periods $2m - 4$ and $2m - 3$ satisfies

$$T_{2m-4} + T_{2m-3} \;\geq\; \frac{\gamma^{(1)}}{4} \left(\frac{m-3}{e} \right)^{m-3}.$$

To express the dominant term in base 2, note that

$$\left(\frac{m-3}{e} \right)^{m-3} = 2^{(m-3)\log_2\left(\frac{m-3}{e}\right)} = 2^{(m-3)(\log_2(m-3) - \log_2 e)}.$$

For $m \geq 10$, we have $m - 3 \geq m/2$ and

$$\log_2(m-3) \;\geq\; \log_2(m/2) \;=\; \log_2 m - 1.$$

Therefore,

$$\begin{aligned}(m-3)\left(\log_2(m-3) - \log_2 e\right) &\geq \frac{m}{2}\left((\log_2 m - 1) - \log_2 e\right) \\ &= \frac{m}{2}\left(\log_2 m - (1 + \log_2 e)\right).\end{aligned}$$

Let $C \coloneqq 1 + \log_2 e$. For all $m \geq 2^{2C} (< 10)$ we have $\log_2 m - C \geq \frac{1}{2} \log_2 m$, and hence

$$(m-3)\left(\log_2(m-3) - \log_2 e\right) \;\geq\; \frac{m}{4} \log_2 m.$$

Plugging this back yields

$$\left(\frac{m-3}{e}\right)^{m-3} \;\geq\; 2^{\frac{m}{4}\log_2 m} = 2^{\Omega(m \log m)}.$$

□

A.3.8 Auxiliary lemmas

Lemma A.3.11. *Let $\boldsymbol{x}_i \in \Delta(\mathcal{A}_i)$. Suppose there exist $a_i, a_i' \in \mathcal{A}_i$ such that $\boldsymbol{x}_i[a_i] + \boldsymbol{x}_i[a_i'] \geq 1 - \delta$ for $\delta = \frac{1}{4m}$. Then*

$$\max\{\boldsymbol{x}_i[a_i], \boldsymbol{x}_i[a_i']\} \geq \boldsymbol{x}_i[\tilde{a}_i] \quad \textit{for all } \tilde{a}_i \in \mathcal{A}_i.$$

In particular, either $\boldsymbol{x}_i[a_i]$ or $\boldsymbol{x}_i[a_i']$ is a maximal coordinate of $\boldsymbol{x}_i$.

Proof. Since $\boldsymbol{x}_i$ is a probability vector,

$$\sum_{\tilde{a}_i \in \mathcal{A}_i \setminus \{a_i, a_i'\}} \boldsymbol{x}_i[\tilde{a}_i] = 1 - (\boldsymbol{x}_i[a_i] + \boldsymbol{x}_i[a_i']) \leq \delta.$$

Hence, for every $\tilde{a}_i \in \mathcal{A}_i \setminus \{a_i, a_i'\}$,

$$\boldsymbol{x}_i[\tilde{a}_i] \leq \sum_{\hat{a}_i \in \mathcal{A}_i \setminus \{a_i, a_i'\}} \boldsymbol{x}_i[\hat{a}_i] \leq \delta. \tag{A.60}$$

On the other hand,

$$\max\{\boldsymbol{x}_i[a_i], \boldsymbol{x}_i[a_i']\} \geq \frac{\boldsymbol{x}_i[a_i] + \boldsymbol{x}_i[a_i']}{2} \geq \frac{1-\delta}{2}. \tag{A.61}$$

Because $\delta = \frac{1}{4m}$, we have $\frac{1-\delta}{2} \geq \delta$, and therefore from (A.60) and (A.61)

$$\max\{\boldsymbol{x}_i[a_i], \boldsymbol{x}_i[a_i']\} \geq \boldsymbol{x}_i[\tilde{a}_i] \quad \text{for all } \tilde{a}_i \in \mathcal{A}_i \setminus \{a_i, a_i'\}.$$

The conclusion is trivial for $\tilde{a}_i \in \{a_i, a_i'\}$, so

$$\max\{\boldsymbol{x}_i[a_i], \boldsymbol{x}_i[a_i']\} \geq \boldsymbol{x}_i[\tilde{a}_i] \quad \text{for all } \tilde{a}_i \in \mathcal{A}_i,$$

which implies that either $\boldsymbol{x}_i[a_i]$ or $\boldsymbol{x}_i[a_i']$ is a maximal coordinate of $\boldsymbol{x}_i$. □

Lemma A.3.12. *For any player i and any $t \geq 1$, it holds that*

$$\max_{a_i \in \mathcal{A}_i} \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i] - \max_{a_i \in \mathcal{A}_i} \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}[a_i] \;\geq\; \min\Big\{\boldsymbol{u}_i^{(t)}[a_i^{(t-1)}],\; \boldsymbol{u}_i^{(t)}[a_i^{(t)}]\Big\},$$

where

$$a_i^{(t)} \in \arg\max_{a_i' \in \mathcal{A}_i} \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i'], \quad a_i^{(t-1)} \in \arg\max_{a_i' \in \mathcal{A}_i} \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}[a_i'].$$

Proof. Define

$$M_t \coloneqq \max_{a_i' \in \mathcal{A}_i} \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i'], \qquad M_{t-1} \coloneqq \max_{a_i' \in \mathcal{A}_i} \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}[a_i'].$$

By definition of $a_i^{(t)}$ and $a_i^{(t-1)}$, we have $M_t = \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i^{(t)}]$, $M_{t-1} = \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}[a_i^{(t-1)}]$. Hence,

$$\begin{aligned} M_t - M_{t-1} &= \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i^{(t)}] - \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}[a_i^{(t-1)}] \\ &= \left(\sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}[a_i^{(t)}] - \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}[a_i^{(t-1)}] \right) + \boldsymbol{u}_i^{(t)}[a_i^{(t)}]. \end{aligned} \tag{A.62}$$

By optimality of $a_i^{(t-1)}$, we have $\sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}[a_i^{(t-1)}] \geq \sum_{\tau=1}^{t-1} \boldsymbol{u}_i^{(\tau)}[a_i^{(t)}]$, so this term in (A.62) is nonpositive and therefore

$$M_t - M_{t-1} \;\geq\; \boldsymbol{u}_i^{(t)}[a_i^{(t)}]. \tag{A.63}$$

Finally,

$$\boldsymbol{u}_i^{(t)}[a_i^{(t)}] \;\geq\; \min\Big\{ \boldsymbol{u}_i^{(t)}[a_i^{(t-1)}],\, \boldsymbol{u}_i^{(t)}[a_i^{(t)}] \Big\}. \tag{A.64}$$

Combining (A.63) and (A.64) proves the claim. □

Lemma A.3.13. *Let $m \geq 2$ and $\delta = \frac{1}{4m}$. Then, for every integer k with $3 \leq k \leq 2m-1$, it holds that*

$$(1-\delta)(k-1) - \delta(2m-1) \;\geq\; k-2.$$

Proof. Starting from the left-hand side and subtracting $k-2$, we obtain

$$(1-\delta)(k-1) - \delta(2m-1) - (k-2) = 1 - \delta\big((k-1) + (2m-1)\big) = 1 - \delta(k+2m-2).$$

It therefore suffices to show that $1 - \delta(k + 2m - 2) \geq 0$, equivalently $\delta(k + 2m - 2) \leq 1$. Using $\delta = \frac{1}{4m}$ and $k \leq 2m - 1$, we have

$$\delta(k + 2m - 2) = \frac{k + 2m - 2}{4m} \leq \frac{(2m - 1) + 2m - 2}{4m} = \frac{4m - 3}{4m} < 1.$$

Hence $1 - \delta(k + 2m - 2) > 0$, and thus

$$(1 - \delta)(k - 1) - \delta(2m - 1) \geq k - 2, \quad \text{for } 3 \leq k \leq (2m - 1),$$

as claimed. □

A.4 Ommited proofs from Section 5.6

We conclude with the proofs from Section 5.6.

Lemma 5.6.2. *For any $k \geq 2$, $T_k \geq (k-1)T_{k-1} + T_{k-2} + T_{k-3} - \sum_{\kappa=1}^{k-4} \kappa T_\kappa$. ($T_k$ for $k \leq 0$ is to be interpreted as 0.)*

Proof. We let $\mathcal{T}_k$ be the set of time indices corresponding to T_k. Let i be the unique player who has a positive best-response gap during $\mathcal{T}_k$; uniqueness follows from the snake property of P and the construction of the game. We denote by $a_i(k) \in \{0, 1\}$ the action of player i corresponding to payoff k, and similarly for $a_i(k+1)$. $\mathcal{T}_k$ lasts as long as it takes so that $\sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i(k+1)] > \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i(k)]$. This is so because `FP` picks with probability 1 the action with the highest cumulative utility; without any loss, we assume for convenience that ties are broken adversarially, so as to maximize the number of steps.

Now, for all times t during the iterations in $\mathcal{T}_{k-1}$, we have $\boldsymbol{u}_i^{(t)}[a_i(k+1)] = \boldsymbol{u}_i^{(t)}[a_i(k)] - (k-1)$. This holds because during those iterations i was obtaining $(k-1)$ by playing $a_i(k)$ while switching to $a_i(k+1)$ would result in 0 utility since the corresponding action is off the path.

Furthermore, we claim that during all times in $\mathcal{T}_{k-2}$ and $\mathcal{T}_{k-3}$, it holds that $\boldsymbol{u}_i^{(t)}[a_i(k+1)] \leq \boldsymbol{u}_i^{(t)}[a_i(k)] - 1$. Specifically, during $\mathcal{T}_{k-2}$, there exists some player $i' \neq i$ with positive best-response gap, for otherwise the snake property of P would be violated—there would exist an edge between the vertex corresponding to $k-2$ and the vertex corresponding to k. This means that during those iterations switching to $a_i(k+1)$ would lead to a utility smaller by at least 1. During $\mathcal{T}_{k-3}$, it is possible that i has a positive best-response gap, in which case player i eventually transitions from $a_i(k+1) = a_i(k-3)$ (actions are binary and $a_i(k+1) \neq a_i(k)$) to $a_i(k)$, which means that again $a_i(k)$ is the preferred action and satisfies $\boldsymbol{u}_i^{(t)}[a_i(k)] \geq \boldsymbol{u}_i^{(t)}[a_i(k+1)] + 1$. If it is not player i who transitions, playing action $a_i(k+1)$ leads off the path, so the same inequality applies.

As a result, if t is the last round in $\mathcal{T}_k$,

$$\sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i(k+1)] - \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i(k)] \leq T_k - (k-1)T_{k-1} - T_{k-2} - T_{k-3} + \sum_{\kappa=1}^{k-4} \kappa T_\kappa.$$

Here we used the fact that $\boldsymbol{u}_i^{(t)}[a_i(k+1)] = \boldsymbol{u}_i^{(t)}[a_i(k)] + 1$ for all iterations t in $\mathcal{T}_k$, and during $\mathcal{T}_\kappa$ we have $\boldsymbol{u}_i^{(t)}[a_i(k+1)] \leq \boldsymbol{u}_i^{(t)}[a_i(k)] + \kappa$. Since t is the last round in T_k, it follows that $\sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i(k+1)] - \sum_{\tau=1}^{t} \boldsymbol{u}_i^{(\tau)}[a_i(k)] > 0$. This implies

$$T_k \geq (k-1)T_{k-1} + T_{k-2} + T_{k-3} - (k-4)T_{k-4} - \cdots - 1T_1,$$

as claimed. □

Lemma 5.6.3. *Any sequence satisfying the recurrence relation of Lemma 5.6.2 grows at least as* $(k-1)!$ *for* $k \geq 1$.

Proof. We define, for $k \geq 2$,

$$S_k \coloneqq T_k - (k-1)T_{k-1}.$$

It suffices to prove that $S_k \geq 0$ for all $k \geq 2$, as it would imply $T_k \geq (k-1)T_{k-1}$, which in turn yields $T_k \geq (k-1)!$ since $T_3 = 2!$ Using the recurrence relation for T_k,

$$\begin{aligned} S_k &= T_k - (k-1)T_{k-1} \\ &\geq T_{k-2} + T_{k-3} - \sum_{j=1}^{k-4} jT_j. \end{aligned} \tag{A.65}$$

Let R_k denote the right-hand side of (A.65). We proceed by strong induction. We assume $R_\kappa \geq 0$ for all $2 \leq \kappa < k$, which in turn implies $T_\kappa \geq (\kappa - 1)T_{\kappa-1}$ for $2 \leq \kappa < k$. We will show that $R_k \geq 0$. The basis of the induction $k \in \{2, 3, 4\}$ follows trivially, so we take $k \geq 5$.

We consider the difference

$$\begin{aligned} R_k - R_{k-1} &= \left(T_{k-2} + T_{k-3} - \sum_{j=1}^{k-4} jT_j \right) - \left(T_{k-3} + T_{k-4} - \sum_{j=1}^{k-5} jT_j \right) \\ &= T_{k-2} - T_{k-4} - (k-4)T_{k-4} \\ &= T_{k-2} - (k-3)T_{k-4}. \end{aligned}$$

By the inductive hypothesis,

$$T_{k-2} \geq (k-3)T_{k-3} \geq (k-3)(k-4)T_{k-4}.$$

So, $R_k - R_{k-1} \geq (k-3)(k-5)T_{k-4} \geq 0$. Given that $R_{k-1} \geq 0$, it follows that $R_k \geq 0$, and the claim follows. □

A.5 Omitted proofs from Section 6.4

Corollary 6.4.2 (Limit sets)**.** *If gradient flow is initialized on $\mathcal{M}$, then $u(t) \to +\infty$ and $x_1(t) \to 1$. Moreover, the ω-limit set reads $\omega(\boldsymbol{x}(0)) = \{1\} \times [-1, 1]$.*

Proof. By Corollary 6.4.1, if the trajectory is initialized on $\mathcal{M}$, then it remains on $\mathcal{M}$ for all future times, and $\dot{u} > 0$ holds.

We now show that $u(t)$ cannot converge to a finite value. Suppose, toward a contradiction, that

$$u(t) \to u_* < +\infty.$$

Since $u(t)$ is increasing, this would mean that the trajectory approaches a finite point on the invariant curve. But $\dot{u}$ is continuous and strictly positive at u_* by definition. Therefore there exists $\eta > 0$ such that, for all sufficiently large t,

$$\dot{u}(t) \geq \eta > 0.$$

This is incompatible with convergence of $u(t)$ to the finite value u_*. Hence

$$u(t) \to +\infty.$$

Since $x_1(t) = 1 - 1/u(t)$, we immediately obtain

$$x_1(t) \to 1.$$

It remains to identify the accumulation set in the x_2-coordinate. Since $u(t) \to +\infty$, the phase

$$v(t) = \pi(u(t) - 1)$$

also tends to $+\infty$. Moreover $v(t)$ is continuous and strictly increasing because $\dot{v}(t) = \pi\dot{u}(t) > 0$. Hence the phase passes through infinitely many full periods. Along the invariant curve,

$$x_2(t) = \sin(v(t)),$$

so $x_2(t)$ has every value in $[-1, 1]$ as an accumulation value.

More explicitly, fix any $y \in [-1, 1]$, and choose $\alpha \in [-\pi/2, \pi/2]$ such that $\sin\alpha = y$. Since $v(t)$ is continuous, strictly increasing, and tends to $+\infty$, for every sufficiently large integer n there is a time t_n such that

$$v(t_n) = \alpha + 2\pi n.$$

Then

$$x_2(t_n) = \sin(v(t_n)) = y,$$

while $x_1(t_n) \to 1$. Therefore

$$(x_1(t_n), x_2(t_n)) \to (1, y).$$

Since $y \in [-1, 1]$ was arbitrary, every point of $\{1\} \times [-1, 1]$ is an accumulation point.

Conversely, every accumulation point must lie in $\{1\} \times [-1, 1]$, because $x_1(t) \to 1$ and $x_2(t) \in [-1, 1]$. Therefore

$$\omega(\boldsymbol{x}(0)) = \{1\} \times [-1, 1].$$

□

Lemma 6.4.3 (Dwell-time bounds for overlapping phase windows)**.** *Suppose that the gradient flow is initialized on the invariant curve $\mathcal{M}$. Then, for every $k \geq 2$, $m \geq 1$, and $i = 0, \ldots, N^{(m)} - 1$,*

$$T_{k,i}^{(m)} \leq \frac{15}{2^m} Q_{k,i}^{(m)}.$$

Proof. Since the trajectory is initialized on the invariant curve $\mathcal{M}$, the tracking error satisfies

$$\delta(t) = 0 \qquad \text{for all } t \geq 0.$$

As shown above, the time spent in a phase interval is computed by integrating the phase density

$$g_k(\phi) = e^{U_k} \left[\frac{e^{\phi/\pi}}{\pi(U_k + \phi/\pi)^4} + \pi e^{\phi/\pi} \cos^2 \phi \right].$$

Since $k \geq 2$, we have $U_k = 2k - 1 \geq 3$. Therefore, for every $\phi \in [0, 2\pi]$,

$$\frac{1}{(U_k + \phi/\pi)^4} \leq \frac{1}{81}, \qquad 1 \leq e^{\phi/\pi} \leq e^2. \tag{A.66}$$

We first consider the non-wrapping windows $i = 0, \ldots, N_m - 2$. Since $|I_i^{(m)}| = \ell_m$, it follows from (A.66) that

$$\int_{I_i^{(m)}} e^{\phi/\pi} \, d\phi \leq e^2 \ell_m. \tag{A.67}$$

Moreover, because $\cos^2 \phi \leq 1$, we also have

$$\int_{I_i^{(m)}} e^{\phi/\pi} \cos^2 \phi \, d\phi \leq e^2 \int_{I_i^{(m)}} \cos^2 \phi \, d\phi \leq e^2 \ell_m. \tag{A.68}$$

Thus, combining (A.66), (A.67), and (A.68), we obtain

$$T_{k,i}^{(m)} = e^{U_k} \int_{I_i^{(m)}} \left[\frac{e^{\phi/\pi}}{\pi(U_k + \phi/\pi)^4} + \pi e^{\phi/\pi} \cos^2 \phi \right] d\phi$$
$$\leq e^{U_k} e^2 \ell_m \left(\frac{1}{81\pi} + \pi \right). \tag{A.69}$$

We now derive a lower bound on the complementary time $Q_{k,i}^{(m)}$. Recall that $Q_{k,i}^{(m)}$ is obtained by integrating g_k over the complement of the interval $I_i^{(m)}$. To obtain a convenient lower bound, we use two elementary observations. First, the first term in g_k is nonnegative, so removing it can only decrease the integral. Second, since $e^{\phi/\pi} \geq 1$ for every $\phi \in [0, 2\pi]$, we may replace this factor by 1 from below. Therefore,

$$Q_{k,i}^{(m)} \geq e^{U_k} \pi \int_{[0,2\pi] \setminus I_i^{(m)}} \cos^2 \phi \, d\phi.$$

It remains to estimate the integral of $\cos^2 \phi$ over the complementary set. Over one full period, we have

$$\int_0^{2\pi} \cos^2 \phi \, d\phi = \pi.$$

Therefore,

$$\int_{[0,2\pi] \setminus I_i^{(m)}} \cos^2 \phi \, d\phi = \int_0^{2\pi} \cos^2 \phi \, d\phi - \int_{I_i^{(m)}} \cos^2 \phi \, d\phi.$$

As before, since $\cos^2 \phi \leq 1$, the integral over $I_i^{(m)}$ is at most the length of that interval:

$$\int_{I_i^{(m)}} \cos^2 \phi \, d\phi \leq |I_i^{(m)}| = \ell^{(m)}.$$

Substituting this into the previous identity yields

$$\int_{[0,2\pi]\setminus I_i^{(m)}} \cos^2\phi \, d\phi \geq \pi - \ell^{(m)}.$$

Consequently,

$$Q_{k,i}^{(m)} \geq e^{U_k}\pi(\pi - \ell^{(m)}). \tag{A.70}$$

Finally, combining the upper bound from (A.69) with the lower bound in (A.70), we conclude that

$$\frac{T_{k,i}^{(m)}}{Q_{k,i}^{(m)}} \leq \frac{e^2\ell^{(m)}\left(\frac{1}{81\pi}+\pi\right)}{\pi(\pi-\ell^{(m)})}.$$

Since $m \geq 1$, we have $\ell^{(m)} = \pi 2^{-m} \leq \pi/2$, and therefore

$$\pi - \ell^{(m)} \geq \frac{\pi}{2}.$$

Hence,

$$\frac{T_{k,i}^{(m)}}{Q_{k,i}^{(m)}} \leq \frac{2e^2\ell^{(m)}\left(\frac{1}{81\pi}+\pi\right)}{\pi^2}.$$

Using $\ell^{(m)} = \pi 2^{-m}$, this becomes

$$\frac{T_{k,i}^{(m)}}{Q_{k,i}^{(m)}} \leq C_{\mathrm{nw}}\, 2^{-m}, \qquad C_{\mathrm{nw}} := 2e^2\left(1+\frac{1}{81\pi^2}\right). \tag{A.71}$$

The constant C_{nw} is universal. Moreover,

$$C_{\mathrm{nw}} = 2e^2\left(1+\frac{1}{81\pi^2}\right) < 15.$$

Therefore,

$$T_{k,i}^{(m)} \le 15\, 2^{-m} Q_{k,i}^{(m)} \qquad \text{for } i = 0, \dots, N^{(m)} - 2. \tag{A.72}$$

It remains to handle the final wrapping window. By definition,

$$T_{k,N_m-1}^{(m)} = \int_{2\pi - h^{(m)}}^{2\pi} g_k(\phi)\, d\phi + \int_0^{h^{(m)}} g_{k+1}(\phi)\, d\phi.$$

We estimate the two pieces separately.

For the first piece, the interval has length h_m, so applying the same bound as in (A.69) gives

$$\int_{2\pi - h_m}^{2\pi} g_k(\phi)\, d\phi \le e^{U_k} e^2 h^{(m)} \left(\frac{1}{81\pi} + \pi \right). \tag{A.73}$$

For the second piece, note that $U_{k+1} = U_k + 2$. Also, on $[0, h^{(m)}]$, we have

$$e^{\phi/\pi} \le e^{h^{(m)}/\pi} \le e^{1/4},$$

because $h^{(m)} \le \pi/4$ for $m \ge 1$. Since $U_{k+1} \ge 3$, the same polynomial bound applies, and hence

$$\int_0^{h^{(m)}} g_{k+1}(\phi)\, d\phi \le e^{U_{k+1}} e^{1/4} h^{(m)} \left(\frac{1}{81\pi} + \pi \right) = e^{U_k} e^{9/4} h^{(m)} \left(\frac{1}{81\pi} + \pi \right). \tag{A.74}$$

Summing (A.73) and (A.74), we obtain

$$T_{k,N^{(m)}1}^{(m)} \le e^{U_k} \left(e^2 + e^{9/4} \right) h^{(m)} \left(\frac{1}{81\pi} + \pi \right). \tag{A.75}$$

We now derive a lower bound for the comparison interval

$$Q^{(m)}_{k,N^{(m)}-1} = \int_0^{2\pi - h^{(m)}} g_k(\phi)\, d\phi,$$

which lies entirely within the kth oscillation. Therefore, we may apply (A.70) on this interval. Since $h^{(m)} \le \pi/4$, we obtain

$$Q^{(m)}_{k,N_m-1} \ge e^{U_k}\pi(\pi - h^{(m)}) \ge e^{U_k}\pi\left(\frac{3\pi}{4}\right). \tag{A.76}$$

Combining (A.75) and (A.76), we obtain

$$\frac{T^{(m)}_{k,N_m-1}}{Q^{(m)}_{k,N_m-1}} \le \frac{\left(e^2 + e^{9/4}\right) h^{(m)} \left(\frac{1}{81\pi} + \pi\right)}{\frac{3\pi^2}{4}}.$$

Using $h^{(m)} = \pi 2^{-(m+1)}$, we can rewrite the right-hand side as

$$\frac{T^{(m)}_{k,N_m-1}}{Q^{(m)}_{k,N_m-1}} \le C_{\text{wrap}}\, 2^{-m}, \tag{A.77}$$

where

$$C_{\text{wrap}} := \frac{2}{3}\left(e^2 + e^{9/4}\right)\left(1 + \frac{1}{81\pi^2}\right) < 15.$$

Therefore,

$$T^{(m)}_{k,N_m-1} \le 15\, 2^{-m} Q^{(m)}_{k,N_m-1}. \tag{A.78}$$

Finally, combining (A.72) with (A.78), we conclude that

$$T^{(m)}_{k,i} \le 15\, 2^{-m} Q^{(m)}_{k,i}, \qquad \text{for every } i = 0, \dots, N^{(m)} - 1.$$

□

Lemma 6.4.4 (Dwell-time bounds in terms of x_2-level sets)**.** *Suppose that the gradient flow is initialized on the invariant curve $\mathcal{M}$. Then, for every $k \geq 2$, $m \geq 5$, and $i = 0, \ldots, N^{(m)} - 1$,*

$$\widetilde{T}_{k,i}^{(m)} \leq \frac{60}{2^m} \widetilde{Q}_{k,i}^{(m)}. \tag{6.17}$$

Proof. On the invariant curve, $x_2(t) = \sin \phi(t)$ during each oscillation. Hence the condition $x_2(t) \in A_i^{(m)}$ corresponds to the phase lying in the window $I_i^{(m)}$, together with its sine-reflection

$$(I_i^{(m)})^\sharp := \pi - I_i^{(m)} \mod 2\pi,$$

because $\sin \phi = \sin(\pi - \phi)$. Therefore, if $(T_{k,i}^{(m)})^\sharp$ denotes the time spent in the reflected window, then the time spent in the x_2-level set $A_i^{(m)}$ is

$$\widetilde{T}_{k,i}^{(m)} = T_{k,i}^{(m)} + (T_{k,i}^{(m)})^\sharp.$$

Let P denote the full comparison-period time associated with the window $I_i^{(m)}$. Thus, for a non-wrapping window, P is the full kth oscillation time, while for a wrapping window the interval is split across the kth and $(k+1)$st oscillations exactly as in the definition of $T_{k,N^{(m)}-1}^{(m)}$. Since $\widetilde{Q}_{k,i}^{(m)}$ is the time spent outside $A_i^{(m)}$ during the same comparison period, we have

$$\widetilde{Q}_{k,i}^{(m)} = P - \widetilde{T}_{k,i}^{(m)} = P - T_{k,i}^{(m)} - (T_{k,i}^{(m)})^\sharp.$$

By Lemma 6.4.3, applied to $I_i^{(m)}$, we have

$$T_{k,i}^{(m)} \le 15\,2^{-m}\left(P - T_{k,i}^{(m)}\right).$$

Applying the same estimate to the reflected window $(I_i^{(m)})^\sharp$ gives

$$(T_{k,i}^{(m)})^\sharp \le 15\,2^{-m}\left(P - (T_{k,i}^{(m)})^\sharp\right).$$

Using

$$P - T_{k,i}^{(m)} = \widetilde{Q}_{k,i}^{(m)} + (T_{k,i}^{(m)})^\sharp$$

and

$$P - (T_{k,i}^{(m)})^\sharp = \widetilde{Q}_{k,i}^{(m)} + T_{k,i}^{(m)},$$

we obtain

$$\begin{aligned}\widetilde{T}_{k,i}^{(m)} &= T_{k,i}^{(m)} + (T_{k,i}^{(m)})^\sharp \\ &\le 15\,2^{-m}\left[2\widetilde{Q}_{k,i}^{(m)} + T_{k,i}^{(m)} + (T_{k,i}^{(m)})^\sharp\right] \\ &= 15\,2^{-m}\left(2\widetilde{Q}_{k,i}^{(m)} + \widetilde{T}_{k,i}^{(m)}\right). \end{aligned} \tag{A.79}$$

Rearranging gives

$$\left(1 - 15\,2^{-m}\right)\widetilde{T}_{k,i}^{(m)} \le 30\,2^{-m}\widetilde{Q}_{k,i}^{(m)}.$$

Since $m \ge 5$, we have $15\,2^{-m} \le 1/2$. Therefore

$$\widetilde{T}_{k,i}^{(m)} \le 60\,2^{-m}\widetilde{Q}_{k,i}^{(m)}.$$

The argument applies uniformly for every $i = 0, \dots, N^{(m)} - 1$, including the final phase window. □

A.6 Omitted proofs from Section 6.5

Lemma 6.5.1 (Coverage and overlap of the trimmed intervals)**.** *For every $m \geq 1$, the trimmed intervals $\widetilde{A}_i^{(m)}$, $i = 0, \dots, N^{(m)} - 1$, cover the feasible x_2-axis:*

$$[-1, 1] \subseteq \bigcup_{i=0}^{N^{(m)}-1} \widetilde{A}_i^{(m)}.$$

Moreover, consecutive intervals overlap by a quantitatively positive amount. If

$$\omega^{(m)} := 2 \sin^3\left(\frac{h^{(m)}}{2}\right) \sin\left(\frac{3h^{(m)}}{2}\right),$$

then, with indices interpreted modulo $N^{(m)}$,

$$\left|\widetilde{A}_i^{(m)} \cap \widetilde{A}_{i+1}^{(m)}\right| \geq \omega^{(m)} \qquad \text{for every } i = 0, \dots, N^{(m)} - 1.$$

In particular, $\omega^{(m)} > 0$.

Proof. Since the phase windows $I_i^{(m)}$ cover one full period and $x_2 = \sin\phi$ along the invariant curve, the intervals $A_i^{(m)} = \sin(I_i^{(m)})$ cover $[-1, 1]$. It remains to check that passing from $A_i^{(m)}$ to the trimmed intervals $\widetilde{A}_i^{(m)}$ does not create gaps, and in fact leaves a uniformly positive overlap between consecutive trimmed intervals.

We first consider two consecutive non-wrapping phase windows whose union lies on a monotone branch of sin:

$$I_i^{(m)} = [ih^{(m)}, (i+2)h^{(m)}], \qquad I_{i+1}^{(m)} = [(i+1)h^{(m)}, (i+3)h^{(m)}].$$

Suppose first that sin is increasing on $[ih^{(m)}, (i+3)h^{(m)}]$. Then

$$A_i^{(m)} = [\sin(ih^{(m)}), \sin((i+2)h^{(m)})], \qquad A_{i+1}^{(m)} = [\sin((i+1)h^{(m)}), \sin((i+3)h^{(m)})].$$

Using

$$\widetilde{A}_j^{(m)} = \left[\frac{3\alpha_j^{(m)} + \beta_j^{(m)}}{4}, \frac{\alpha_j^{(m)} + 3\beta_j^{(m)}}{4}\right], \qquad j \in \{i, i+1\},$$

the overlap length is

$$\begin{aligned}\left|\widetilde{A}_i^{(m)} \cap \widetilde{A}_{i+1}^{(m)}\right| &= \frac{\sin(ih^{(m)}) + 3\sin((i+2)h^{(m)}) - 3\sin((i+1)h^{(m)}) - \sin((i+3)h^{(m)})}{4} \\ &= 2\sin^3\left(\frac{h^{(m)}}{2}\right)\cos\left(\left(i+\frac{3}{2}\right)h^{(m)}\right).\end{aligned}$$

On a monotone increasing branch, the cosine factor is nonnegative. Moreover, for the monotone pairs under consideration, its smallest possible value occurs for the pair closest to a critical point of sin, and is at least $\sin(3h^{(m)}/2)$. Therefore

$$\left|\widetilde{A}_i^{(m)} \cap \widetilde{A}_{i+1}^{(m)}\right| \geq 2\sin^3\left(\frac{h^{(m)}}{2}\right)\sin\left(\frac{3h^{(m)}}{2}\right) = \omega^{(m)}.$$

The case where sin is decreasing on $[ih^{(m)}, (i+3)h^{(m)}]$ is identical, with the order of the endpoints reversed.

It remains to check the pairs near the extrema of sin. We treat the maximum 1; the minimum -1 is symmetric. Let

$$r := 2^m, \qquad rh^{(m)} = \frac{\pi}{2}.$$

The phase windows whose A-intervals contain 1 are those with indices $r-2$, $r-1$, and r. By the endpoint rule, these intervals are trimmed on the side away from 1 and extended on

the side beyond 1. Hence 1 is strictly inside each of the corresponding $\widetilde{A}$-intervals, so these endpoint intervals overlap with one another.

We only need to check the overlap with the adjacent ordinary intervals. Consider the left adjacent pair $\widetilde{A}^{(m)}_{r-3}$ and $\widetilde{A}^{(m)}_{r-2}$. Since $rh^{(m)} = \pi/2$, we have

$$A^{(m)}_{r-3} = [\cos(3h^{(m)}), \cos h^{(m)}], \qquad A^{(m)}_{r-2} = [\cos(2h^{(m)}), 1].$$

Thus the overlap length is at least

$$\frac{\cos(3h^{(m)}) + 3\cos h^{(m)} - 3\cos(2h^{(m)}) - 1}{4}.$$

Using the identity

$$\cos(3h^{(m)}) + 3\cos h^{(m)} - 3\cos(2h^{(m)}) - 1 = 2(1 - \cos h^{(m)})^2(2\cos h^{(m)} + 1),$$

this becomes

$$\frac{1}{2}(1 - \cos h^{(m)})^2(2\cos h^{(m)} + 1) = 2\sin^3\left(\frac{h^{(m)}}{2}\right)\sin\left(\frac{3h^{(m)}}{2}\right) = \omega^{(m)}.$$

The right adjacent pair around the maximum is handled by symmetry around $\pi/2$. The minimum -1 is handled in the same way by symmetry around $3\pi/2$.

Finally, the wrapping pair is covered by the monotone-branch argument after identifying

$$I^{(m)}_{N^{(m)}-1} = [2\pi - h^{(m)}, 2\pi] \cup [0, h^{(m)}]$$

with the interval $[-h^{(m)}, h^{(m)}]$. Its union with $I^{(m)}_0 = [0, 2h^{(m)}]$ lies on a monotone branch of sin for $m \geq 2$.

Thus every consecutive pair of trimmed intervals overlaps by at least $\omega^{(m)} > 0$. In particular, the intervals form an overlapping chain whose union contains both endpoints -1 and 1. Therefore they cover the whole feasible x_2-axis:

$$[-1, 1] \subseteq \bigcup_{i=0}^{N^{(m)}-1} \widetilde{A}_i^{(m)}.$$

□