Book of abstracts

Workshop on Learning in Games

July 1–3, 2024

Institut de Mathématiques de Toulouse

Toulouse, France

Preface

The past decade has witnessed a paradigm shift in system design from model-based to a data-driven or learning-based one fueled by success stories in computer vision and natural language processing. This has given hope for a wider applicability of this paradigm especially in large-scale systems where there is a need for autonomous and adaptive control. Through a series of workshops, the CIMI semester *Stochastic control and learning for complex networks (SOLACE)* seeks to showcase current trends in data-driven design of systems that do not necessarily make the headlines but are nevertheless fundamental to our society.

The present workshop, which is the second event of the semester, explores the exciting intersection between game theory and learning. On one hand, classical game theory assumes that all agents are aware of the primitives of the game, including payoff distributions. On the other hand, the classical theory of online learning analyses a single agent interacting with a stochastic/adversarial environment. The workshop aims to shed light on recent developments in the area of learning in games, where multiple strategic agents interact without a priori knowledge of all the primitives of the game/environment.

The program features two tutorials – one on multi-player bandit learning; and another on no-regret learning in repeated games – and fifteen invited talks by renowned researchers on topics ranging from algorithms and fundamental limits for no-regret learning, learning in (repeated) auctions, and learning equilibria of games. To encourage participation of young researchers, they were invited to present a poster and an accompanying flash-talk to 'elevator-pitch' their work to the audience. This booklet collects titles and abstracts of all presentations that took place during the workshop.

The organisers would like to express their gratitude to CIMI for providing the majority of the funding for the workshop through the SOLACE semester, as well as to IMT, CNRS, LAAS-CNRS, and ANITI for their support. A big thanks is due to the speakers and the poster presenters without whom the workshop would have not been as rich and lively, and of course to the audience for their participation. It is our hope that this workshop will trigger new lines of research in the area, and foster collaborations between participants.

Sébastien Gerchinovitz
Jayakrishnan Nair
Balakrishna Prabhu
Bary Pradelski
Jérôme Renault

July 3, 2024
Toulouse, France

Contents

I Tutorials

1 Simina Branzei

Title: Multiplayer bandit learning

Abstract: The classic theory of multi-armed bandits shows how one decision maker can take decisions sequentially; the optimal strategies highlight the tradeoff between exploration and exploitation.

I will describe the basic model for stochastic multi-armed bandits with one decision maker and explain the optimal (Gittins index) strategy. Then I will discuss the extension of the model to multiple players, and describe a set of results from the paper "Multiplayer bandit learning, from competition to cooperation" (https://arxiv.org/abs/1908.01135, joint with Peres, in COLT 21).

In the multi-player multi-armed model, there are two players and two arms; one arm has a known success probability (i.e. is safe) and the other arm does not (i.e. is risky). In each round, each player pulls an arm, gets the reward from that arm, and observes the action of the other player but not their reward.

The main results are that competing players explore less than a single player, but they are not completely myopic. Thus under optimal play, information (acquired by experimenting with the risky arm) is less valuable in the zero-sum game than in the single player setting. This leads to reduced exploration compared to the one player optimum; but information still has positive value. In contrast, cooperating players explore more than a single player, while neutral players learn from each other, getting strictly higher total rewards than they would playing alone. Finally, competing and neutral players eventually settle on the same arm in every Nash equilibrium, while this can fail for cooperating players.

2 Sylvain Sorin

Title: No-regret Algorithms and Games

Abstract: No regret-algorithms appeared in game dynamics, prediction of sequences and convex optimization. We will describe the basic framework, the main tools and the fundamental results. Part I will deal with the initial discrete (in space) and random procedures. Part II will be concern with continuous deterministic algorithms.

Part I

A. Introduction: on-line algorithms, external and internal regret; Blackwell theorem; existence of no-regret algorithms; Calibration

   Extensions: Bandit, signals; expected version; experts, wide range; from external to internal regret

B. Application to finite games: unilateral procedures; external regret and Hannan set; internal regret and correlated equilibrium distributions (dual viewpoint)

   Other approaches: smooth fictitious play, stochastic approximation, continuous time and replicator dynamics

Part II

A. Comparison:

   1. Continuous/discrete time
   2. On-line learning, vector field, convex optimization
   3. Gradient descent, mirror descent, dual averaging

B. Extensions and recent advances: optimistic procedures, adaptive stepsize.

II Invited talks

1 Venkat Anantharam

Title: Game theory for cumulative-prospect-theoretic agents

Abstract: Classical game theory is based on modeling the agents as maximizers of an expected utility. Empirical research on human behavior has demonstrated that humans are not well modeled as utility maximizers. An attractive theory that appears to be capable of capturing many of the peculiarities of human behavior, and which includes expected utility theory as a special case, is the cumulative prospect theory (CPT) of Kahneman and Tversky.

In the doctoral thesis work of Soham Phade, which we will give an exposition of in this talk, we have aimed at developing versions of some of the core results in classical game theory in the context where the agents have CPT preferences. We will discuss the relationship between Nash and correlated equilibria in this context, develop an analog of the calibrated learning theory justification of correlated equilibria for CPT agents, and build a theory of CPT mechanism design for implementing social choice functions. A potential application of some of these techniques in communication networking will also be discussed.

2 Galit Ashkenazi-Golan

Title: Policy Gradient Methods in Repeated Games

Abstract: Policy gradient methods have emerged as powerful tools for optimising complex decision-making processes. These methods utilise gradient ascent to iteratively enhance policies based on observed rewards. While they frequently achieve local convergence, global convergence guarantees remain elusive outside of specific classes of games. We illustrate this phenomenon through examples, illuminating the challenges surrounding global optimisation.

Following that, we explore the utilisation of policy gradient methods in repeated games and define the set of policies that are learnable using these methods. Lastly, we provide a preview of our follow-up presentation, wherein we will present a Folk theorem result for learning in repeated games.

3 Martin Bichler

Title: Learning in Bayesian Games

Abstract: Auctions are modeled as Bayesian games with continuous type and action spaces. Determining equilibria in auction games is computationally hard in general and no exact solution theory is known. We introduce an algorithmic framework in which we discretize type and action space and then learn distributional strategies via online optimization algorithms. We show that the equilibrium of the discretized game approximates an equilibrium in the continuous game. In a wide variety of auction games, we provide empirical evidence that the approach approximates the analytical (pure) Bayes Nash equilibrium closely. In standard models where agents are symmetric, we find equilibrium in seconds. The method allows for interdependent valuations and different types of utility functions and it provides a foundation for broadly applicable equilibrium solvers that can push the boundaries of equilibrium analysis in auction markets and beyond.

4 Tommaso Cesari

Title: The Role of Transparency in Repeated First-Price Auctions with Unknown Valuations

Abstract: I will discuss the problem of regret minimization for a single bidder in a sequence of first-price auctions where the bidder discovers the item's value only if the auction is won. In particular, I will show how minimax regret rates change with the amount of information on competing bids disclosed by the auctioneer at the end of each auction. Results will be presented under different assumptions (stochastic, adversarial, and their smoothed variants) on the environment generating the bidder's valuations and competing bids. These minimax rates reveal how the interplay between transparency and the nature of the environment affects how fast one can learn to bid optimally in first-price auctions.

5 Julien Grand-Clement

Title: Fast Last-Iterate Convergence of Learning in Games Requires Forgetful Algorithms

Abstract: Self play via online learning is one of the premier ways to solve large-scale zero-sum games, both in theory and practice. Particularly popular algorithms include optimistic multiplicative weights update (OMWU) and optimistic gradient-descent-ascent (OGDA). While both algorithms enjoy $O(1/T)$ ergodic convergence to Nash equilibrium in two-player zero-sum games, OMWU offers several advantages, including logarithmic dependence on the size of the payoff matrix and $\tilde{O}(1/T)$ convergence to coarse correlated equilibria even in general-sum games. However, in terms of last-iterate convergence in two-player zero-sum games, an increasingly popular topic in this area, OGDA guarantees that the duality gap shrinks at a rate of $(1/\sqrt{T})$, while the best existing last-iterate convergence for OMWU depends on some game-dependent constant that could be arbitrarily large. This begs the question: is this potentially slow last-iterate convergence an inherent disadvantage of OMWU, or is the current analysis too loose? Somewhat surprisingly, we show that the former is true. More generally, we prove that a broad class of algorithms that do not forget the past quickly all suffer the same issue: for any arbitrarily small $\delta > 0$, there exists a 2×2 matrix game such that the algorithm admits a constant duality gap even after $1/\delta$ rounds. This class of algorithms includes OMWU and other standard optimistic follow-the-regularized-leader algorithms.

This is joint work with Y. Cai, G. Farina, C. Kroer, H. Luo, C.-W. Lee and W. Zheng.

6 Sergiu Hart

Title: ”Calibeating”: Beating Forecasters at Their Own Game

Abstract: In order to identify expertise, forecasters should not be tested by their calibration score, which can always be made arbitrarily small, but rather by their Brier score. The Brier score is the sum of the calibration score and the refinement score; the latter measures how good the sorting into bins with the same forecast is, and thus attests to ”expertise.” This raises the question of whether one can gain calibration without losing expertise, which we refer to as ”calibeating.” We provide an easy way to calibeat any forecast, by a deterministic online procedure. We moreover show that calibeating can be achieved by a stochastic procedure that is itself calibrated, and then extend the results to simultaneously calibeating multiple procedures, and to deterministic procedures that are continuously calibrated.

Joint work with Dean Foster.

7 Chi Jin

Title: Beyond Equilibrium Learning

Abstract: While classical game theory primarily focuses on finding equilibria, modern machine learning applications introduce a series of new challenges where standard equilibrium notions are no longer sufficient, and the development of new efficient algorithmic solutions is urgently needed. In this talk, we will demonstrate two such scenarios:

1. a natural goal in multiagent learning is to learn rationalizable behavior, which avoids iteratively dominated actions. Unfortunately, such rationalizability is not guaranteed by standard equilibria, especially when approximation errors are present. Our work presents the first line of efficient algorithms for learning rationalizable equilibria with sample complexities that are polynomial in all problem parameters, including the number of players;
2. In multiplayer symmetric constant-sum games like Mahjong or Poker, a natural baseline is to achieve an equal share of the total reward. We demonstrate that the self-play meta-algorithms used by existing state-of-the-art systems can fail to achieve this simple baseline in general symmetric games.

We will then discuss the new principled solution concept required to achieve this goal.

8 Sam Jindani

Title: Learning efficient equilibria in repeated games

Abstract: The folk theorem tells us that a wide range of payoffs can be sustained as equilibria in an infinitely repeated game. Existing results about learning in repeated games suggest that players may converge to an equilibrium, but do not address selection between equilibria. I propose a stochastic learning rule that selects a subgame-perfect equilibrium of the repeated game in which payoffs are efficient. The exact payoffs selected depend on how players experiment; two natural specifications yield the Kalai–Smorodinsky and maxmin bargaining solutions, respectively.

9 Maryam Kamgarpour

Title: Learning equilibria in games with bandit feedback

Abstract: A significant challenge in managing large-scale engineering systems, such as energy and transportation networks, lies in enabling autonomous decision-making among interacting agents. Game theory offers a framework for modeling and analyzing these types of problems. In many practical applications, like power markets, each player only has partial information about the cost functions and actions of others. Therefore, a decentralized learning approach is essential to devise optimal strategies for each player.

My talk will focus on recent advances in decentralized learning algorithms in games under bandit feedback. The first part will discuss conditions for the convergence of decentralized learning to a Nash equilibrium in continuous action static games. The second part will explore Markov games, presenting our methods for decentralized learning of Nash equilibria in zero-sum Markov games and coarse-correlated equilibria in general-sum Markov games. I will demonstrate the practical applications of the developed algorithms using real-world problems.

This presentation is primarily based on the following papers:

[1] T. Tatarenko and M. Kamgarpour. Learning generalized nash equilibria in a class of convex games. *IEEE Trans. on Automatic Control*, 64(4), 2019.

[2] R. Ouhamma and M. Kamgarpour. Learning nash equilibria in zero-sum markov games: a single time-scale algorithm under weak reachability, 2024. arXiv: 2312.08008.

[3] T. Tatarenko and M. Kamgarpour. Convergence rate of learning a strongly variationally stable equilibrium, 2024. arXiv: 2304.02355.

[4] P.G. Sessa, M. Kamgarpour, and A. Krause. Efficient model-based multi-agent reinforcement learning via optimistic equilibrium computation. In *Proc. of the 39th ICML*, 2022.

10 Panayotis Mertikopoulos

Title: Potential vs. harmonic games, convergence vs. recurrence, regret vs. optimism

Abstract: The long-run behavior of multi-agent learning – and, in particular, no-regret learning – is relatively well-understood in potential games, where players have common interests. By contrast, in harmonic games – the strategic counterpart of potential games, where players have conflicting interests – very little is known outside the narrow subclass of 2-player zero-sum games with a fully-mixed equilibrium. Our paper seeks to partially fill this gap by focusing on the full class of (generalized) harmonic games and examining the convergence properties of ”follow-the-regularized-leader” (FTRL), the most widely studied class of no-regret learning schemes. As a first result, we show that the continuous-time dynamics of FTRL are Poincaré recurrent, that is, they return arbitrarily close to their starting point infinitely often, and hence fail to converge. In discrete time, the standard, ”vanilla” implementation of FTRL may lead to even worse outcomes, eventually trapping the players in a perpetual cycle of best-responses. However, if FTRL is augmented with a suitable extrapolation step – which includes as special cases the optimistic and mirror-prox variants of FTRL – we show that learning converges to a Nash equilibrium from any initial condition, and all players are guaranteed at most O(1) regret. These results provide an in-depth understanding of no-regret learning in harmonic games, nesting prior work on 2-player zero-sum games, and showing at a high level that harmonic games are the canonical complement of potential games, not only from a strategic, but also from a dynamic viewpoint.

11 Mehryar Mohri

Title: Pseudonorm Approachability and Applications to Regret Minimization

Abstract: Blackwell's approachability theory is a powerful tool for various learning problems. While traditionally studied under the Euclidean distance (ℓ_2), we argue that the ℓ_∞-metric is more suitable for many applications like regret minimization. However, existing ℓ_∞-approachability algorithms suffer from high dimensionality. We present a framework to overcome this issue by converting high-dimensional ℓ_∞ problems into low-dimensional pseudonorm problems. This enables efficient algorithms with convergence independent of the original problem's dimension. We also provide a logarithmic convergence algorithm. Finally, we demonstrate the advantages of our framework in several regret minimization problems.

Joint work with Chris Dann, Yishay Mansour, Jon Schneider, and Balubramanian Sivan.

12 Vianney Perchet

Title: Optimizing the coalition gain in Online Auctions with Greedy Structured Bandits

Abstract: Motivated by online display advertising, this work considers repeated second-price auctions, where agents sample their value from an unknown distribution with cumulative distribution function F. In each auction t, a decision-maker bound by limited observations selects n_t agents from a coalition of N to compete for a prize with p other agents, aiming to maximize the cumulative reward of the coalition across all auctions.

The problem is framed as an N-armed structured bandit, each number of player sent being an arm n, with expected reward $r(n)$ fully characterized by F and $p + n$. We present two algorithms, Local-Greedy (LG) and Greedy-Grid (GG), both achieving constant problem-dependent regret. This relies on three key ingredients:

1. an estimator of $r(n)$ from feedback collected from any arm k,
2. concentration bounds of these estimates for k within an estimation neighborhood of n, and
3. the unimodality property of r under standard assumptions on F.

Additionally, GG exhibits problem-independent guarantees on top of best problem-dependent guarantees. However, by avoiding to rely on confidence intervals, LG practically outperforms GG, as well as standard unimodal bandit algorithms such as OSUB or multi-armed bandit algorithms.

13 Marco Scarsini

Title: Best-response dynamics on a large class of random games

Abstract: Several papers study games with random payoffs with the goal of finding the distribution of the number of pure Nash equilibria and of determining the behavior of best-response dynamics. In most of the existing literature the payoffs are assumed to be independent and identically distributed (i.i.d.) with a continuous distribution function. In this paper we consider classes of random games where the payoffs are not i.i.d. but have a structure that is more suitable for game-theoretic applications. We will focus our analysis on the behavior of best-response dynamics when games have multiple equilibria.

14 Bassel Tarbush

Title: Game connectivity and adaptive dynamics

Abstract: We analyse the typical structure of games in terms of the connectivity properties of their best-response graphs. Our central result shows that almost every game that is 'generic' (without indifferences) and has a pure Nash equilibrium and a 'large' number of players is connected, meaning that every action profile that is not a pure Nash equilibrium can reach every pure Nash equilibrium via best-response paths. This has important implications for dynamics in games. In particular, we show that there are simple, uncoupled, adaptive dynamics for which period-by-period play converges almost surely to a pure Nash equilibrium in almost every large generic game that has one (which contrasts with the known fact that there is no such dynamic that leads almost surely to a pure Nash equilibrium in every generic game that has one). We build on recent results in probabilistic combinatorics for our characterisation of game connectivity.

This is joint work with T. Johnston, M. Savery, and A. Scott.

15 Long Tran-Thanh

Title: Learning against No-Regret Learners: Multi-agent Learning with Strategic Agents

Abstract: In this talk I will discuss our recent results on how to efficiently exploit the fact that we are playing against no-regret learning agents. In particular, I will show how to achieve sub-linear forward and dynamic regrets, notions that are stronger than the (external) regrets, when playing against an agent who uses no-regret learning algorithm to minimise their (external) regret. I will also show how these results can be extended to online Markov decision processes. Finally, I will discuss how to achieve last-round/last-iterate convergence against generic no-regret learners.

III Posters

1 Aurélien Bechler

Title: Bidding efficiently in Simultaneous Ascending Auctions using Monte Carlo Tree Search

Abstract: For decades, Simultaneous Ascending Auction (SAA) has been the most popular mechanism used for spectrum auctions. It has recently been employed by many countries for the allocation of 5G licences. Although SAA presents relatively simple rules, it induces a complex strategical game for which the optimal bidding strategy is unknown. Considering the fact that sometimes billions of euros are at stake in a SAA, establishing an efficient bidding strategy is crucial. In this work, we model the auction as a n-player simultaneous move game with complete information and propose the first efficient bidding algorithm that tackles simultaneously its four main strategical issues: the *exposure problem*, the *own price effect*, *budget constraints* and the *eligibility management problem*. Our solution, called SMS^{α}, is based on Monte Carlo Tree Search (MCTS) and relies on a new method for the prediction of closing prices. By introducing scalarised rewards in SMS^{α}, we give the possibility to bidders to define their own level of risk-aversion. Through extensive numerical experiments on instances of realistic size, we show that SMS^{α} largely outperforms state-of-the-art algorithms, notably by achieving higher expected utility while taking less risks.

Joint work with Alexandre Pacaud and Marceau Coupechoux.

2 Tommaso Cesari

Title: A Contextual Online Learning Theory of Brokerage

Abstract: We investigate brokerage between traders from an online learning perspective. At any round t, two traders arrive with private valuations about an asset they wish to trade. The broker proposes a trading price based on contextual information about the asset. We focus on the scenario where buyer and seller roles are not predetermined: each trader will attempt to either buy or sell depending on the current price.

We assume that the asset market value is an unknown linear function of a d-dimensional vector representing the contextual information available to the broker. Additionally, we model traders' valuations as independent bounded zero-mean perturbations of the asset market value, allowing for potentially different unknown distributions across traders and time steps. The performance is evaluated via the standard notion of *gain from trade*. If the noise distributions admit densities bounded by some constant M, then, for any time horizon T:

- If the agents' valuations are revealed after each interaction, we provide an algorithm achieving $O\big(Md\ln(T)\big)$ regret, and show a corresponding matching lower bound of $\Omega\big(Md\ln(T)\big)$.
- If only their willingness to sell or buy at the proposed price is revealed after each interaction, we provide an algorithm achieving $O\big(\sqrt{MdT\ln(T)}\big)$ regret, and show that this rate is quasi-optimal, via a lower bound of $\Omega\big(\sqrt{MdT}\big)$.

Finally, if we drop the bounded density assumption, we show that the problem becomes unlearnable, even in full feedback.

Joint work with François Bachoc and Roberto Colomboni.

3 Atulya Jain

Title: Calibrated Forecasting and Persuasion

Abstract: How should an expert send forecasts to maximize her utility subject to passing a calibration test? We consider a dynamic game where an expert sends probabilistic forecasts to a decision maker. The decision maker uses a calibration test based on past outcomes to verify the expert's forecasts. We characterize the optimal forecasting strategy by reducing the dynamic game to a static persuasion problem. A distribution of forecasts is implementable by a calibrated strategy if and only if it is a mean-preserving contraction of the distribution of conditionals (honest forecasts). We characterize the value of information by comparing what an informed and uninformed expert can attain. Moreover, we consider a decision maker who uses regret minimization, instead of the calibration test, to take actions. We show that the expert can achieve the same payoff against a regret minimizer as under the calibration test, and in some instances, she can achieve strictly more.

Joint work with Vianney Perchet.

4 Davide Legacci

Title: Convergence vs. Recurrence under No-Regret Learning

Abstract: A fundamental question in multi-agent learning theory is whether players eventually learn to emulate rational behavior through repeated interactions, while minimizing their incurred regret. This question finds a positive answer in the class of potential games, where players have common interests. By contrast, in harmonic games – the strategic counterpart of potential games, where players have conflicting interests – very little is known outside the narrow subclass of 2-player zero-sum games with a fully-mixed equilibrium.

In light of this, our objective is to examine the convergence properties of FTRL – the most widely studied class of no-regret learning schemes – in harmonic games.

As a first result, we show that FTRL dynamics in continuous time – including in particular the replicator dynamics – are Poincaré recurrent in harmonic games, nesting existing results for 2-player zero-sum games.

In discrete time, the standard implementation of FTRL may lead to even worse outcomes, spiraling towards the boundary of the game's strategy space and eventually trapping the players in a perpetual cycle of best-responses. However, if FTRL is augmented with a suitable extrapolation step – which includes as special cases the optimistic and mirror-prox variants of FTRL – we show that learning converges to a Nash equilibrium from any initial condition, and all players are guaranteed at most O(1) regret.

Joint work with Panayotis Mertikopoulos, Christos H. Papadimitriou, Georgios Piliouras, and Bary Pradelski.

5 Prasanna Maddila

Title: Learning in Competitive-Cooperative Games for Anti-Poaching

Abstract: Widespread poaching threatens many endangered species today, requiring robust strategies to coordinate ranger patrols and effectively deter poachers within protected areas. Recent research has modelled this problem as a strategic game between rangers and poachers, resulting in anti-poaching becoming a popular application domain within game theory and multi-agent research communities. Unfortunately, the lack of a standard open-source implementation of the anti-poaching game hinders the reproducibility and advancement of current research in the field. This paper aims to fill this gap by providing the first open-source standardised environment for the anti-poaching game. Our contributions are as follows: (1) we formalise anti-poaching as a Partially Observable Stochastic Game; (2) we provide the Anti-Poaching Environment (APE), an open-source Python implementation of a simulator for this game using the PettingZoo API, which is compatible with many existing multi-agent reinforcement learning (MARL) libraries; and (3) we illustrate how to apply deep reinforcement-learning algorithms from the RLlib library, in order to compute cooperative and cooperative-competitive equilibria of APE instances. Our project is published at https://forgemia.inra.fr/chip-gt/antipoaching.

Joint work with Régis Sabbadin and Meritxell Vinyals.

6 Anna Maddux

Title: Finite-time convergence to an ϵ-efficient Nash equilibrium in potential games

Abstract: This paper investigates the convergence time of log-linear learning to an ϵ-efficient Nash equilibrium (NE) in potential games. In such games, an efficient NE is defined as the maximizer of the potential function. Existing results are limited to potential games with stringent structural assumptions and entail exponential convergence times in $1/\epsilon$. Unaddressed so far, we tackle general potential games and prove the first finite-time convergence to an ϵ-efficient NE. In particular, by using a problem-dependent analysis, our bound depends polynomially on $1/\epsilon$. Furthermore, we provide two extensions of our convergence result: first, we show that a variant of log-linear learning that requires a factor A less feedback on the utility per round enjoys a similar convergence time; second, we demonstrate the robustness of our convergence guarantee if log-linear learning is subject to small perturbations such as alterations in the learning rule or noise-corrupted utilities.

Joint work with Reda Ouhamma and Maryam Kamgarpour.

7 Mathieu Molina

Title: Trading-off price for data quality to achieve fair online allocation

Abstract: We consider the problem of online allocation subject to a long-term fairness penalty. Contrary to existing works, however, we do not assume that the decision-maker observes the protected attributes—which is often unrealistic in practice. Instead they can purchase data that help estimate them from sources of different quality; and hence reduce the fairness penalty at some cost. We model this problem as a multi-armed bandit problem where each arm corresponds to the choice of a data source, coupled with the online allocation problem. We propose an algorithm that jointly solves both problems and show that it has a regret bounded by $\mathcal{O}(\sqrt{T})$. A key difficulty is that the rewards received by selecting a source are correlated by the fairness penalty, which leads to a need for randomization (despite a stochastic setting). Our algorithm takes into account contextual information available before the source selection, and can adapt to many different fairness notions. We also show that in some instances, the estimates used can be learned on the fly.

Joint work with Nicolas Gast, Patrick Loiseau, and Vianney Perchet.

8 Francesco Morri

Title: Learning in Conjectural Stackelberg Games

Abstract: In this work we tackle the class of Stackelberg games considering multiple leaders (N) and a single follower. Specifically, we introduce a new type of game, called **Conjectural Stackelberg Game**, unifying the concepts of Stackelberg and conjectural games: this novel approach allows us to combine the hierarchical structure of Stackelberg games, with the notion of conjectures, while being flexible on the number of agents and on the form of their objective functions.

Formally, this game is represented as follows:

$$\min_{x_i \in X_i, y \in Y} f_i(x_i, \gamma_i(x_i, x_{-i}, y)) \quad \forall i = 1, \ldots, N, \tag{1a}$$

$$\text{s.t.} \quad y \in \arg\min_{y \in Y} g(\boldsymbol{x}, y). \tag{1b}$$

The leaders solve their respective optimization simultaneously, given the set of conjectures $(\gamma_i)_i$, the follower acts second, optimizing its objective function with full access to all the decision variables. Changing the form of $(\gamma_i)_i$ allows us to span different settings for our game: we can remove the dependencies on the other leaders variables and also on the follower's variables leading to games which can be interpreted through different solutions concepts. In particular, we formulate a **conjecture-agnostic** game, with the possibility of allowing each leader to learn its strategy and its conjecture at the same time. We study this type of games both from a theoretical perspective and from an algorithmic one. On the theoretical side, we analyze in depth how the equilibria are influenced by the form of $(\gamma_i)_i$, and whether or not we need to apply restrictions on the conjectures. Furthermore, we compare our general conjectural game with the standard Stackelberg one, through metrics measuring differences in the strategies of the leaders and in the values obtained for their objective functions.

On the algorithmic side, we consider gradient-based learning for the leaders, allowing them to update and adapt their strategy and their conjecture. In particular, we are able to provide convergence guarantees depending on the form and complexity of the $(\gamma_i)_i$ functions. Finally, we validate our result running multiple numerical simulations of different games formulations.

Joint work with Hélène Le Cadre and Luce Brotcorne.

9 Reda Ouhamma

Title: Learning Nash Equilibria in Zero-Sum Markov Games: A Single-Timescale Algorithm Under Weak Reachability

Abstract: We consider decentralized learning for zero-sum games, where players only see their payoff information and are agnostic to the opponent's actions and payoffs. Previous works demonstrated convergence to a Nash equilibrium in this setting using double timescale algorithms under strong reachability assumptions. We address the open problem of achieving an approximate Nash equilibrium efficiently with an uncoupled and single-timescale algorithm under weaker conditions. Our contribution is a rational and convergent algorithm, utilizing Tsallis-entropy regularization in a value-iteration-based approach. The algorithm learns an approximate Nash equilibrium in polynomial time, requiring only the existence of a policy pair that induces an irreducible and aperiodic Markov chain, thus considerably weakening past assumptions. Our analysis leverages negative drift inequalities and introduces novel properties of Tsallis entropy that are of independent interest.

Joint work with Maryam Kamgarpour.

10 Giulio Salizzoni

Title: On the number of Nash equilibria in scalar discrete-time linear quadratic games

Abstract: An open problem in linear quadratic games has been determining the cardinality of Nash equilibria. This problem has renewed relevance given the surge of work on understanding the convergence of learning algorithms in dynamic games. In this paper, we investigate scalar discrete-time infinite-horizon linear quadratic (LQ) games with two agents. Even in this arguably simple setting, there are no results on the number of Nash equilibria. By analyzing the best response map, we formulate a polynomial system of equations characterizing the linear feedback Nash equilibria. This enables us to bring in tools from algebraic geometry, in particular the Gröbner basis, to study the roots of this polynomial system. We prove that the LQ games under consideration admit at most three Nash equilibria. Moreover, we provide sufficient conditions for the existence of at most two Nash equilibria as well as conditions for the uniqueness of the Nash equilibrium. Our numerical experiments demonstrate the tightness of our bounds in addition to showcasing the difficulty of extending the approach and result to settings with more than two agents.

Joint work with Reda Ouhamma and Maryam Kamgarpour.