Essays in learning, optimization and game theory

Essays in Learning, Optimization and Game theory

Benoit G.P. Duvocelle

This book was typeset by the author using LaTeX, and the template of Robbert Harms.

Published by Benoit Duvocelle,
19 allée de l'orée du lac, 64200 Biarritz, France
In 2021
Cover from Jo Design || http://joanasalvado.com
ISBN: 979-10-699-6556-0
Printed in France by Imprimerie Artisanale, Bayonne
Legal submission, February 2021

ESSAYS IN LEARNING, OPTIMIZATION AND GAME THEORY

Dissertation

To obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. Dr. Rianne Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public
on Monday 22 of February 2021, at 10.00 hours

by

Benoit Georges Philippe Duvocelle

Promotor
Prof. Dr. Dries Vermeulen

Co-supervisors
Dr. János Flesch
Dr. Mathias Staudigl

Assessment Committee
Prof. Dr. Frank Thuijsman (Chair)
Prof. Dr. Marco Scarsini (LUISS University, Roma)
Prof. Dr. Tristan Tomala (HEC, Paris)
Dr. Anna Zseleva (Maastricht University)

This research was financially supported by the Graduate School of Business and Economics (GSBE).

To my family, for their eternal support
Du fond du cœur, merci.

Acknowledgments

First and Foremost, I would like to thank my supervisors, János Flesch, Mathias Staudigl and Dries Vermeulen. János, you have been during those three years not only an excellent supervisor but also a very good friend. I remember that the first person I announced that you were one of my supervisors was Rida Laraki, and he said "János, il a de la bouteille". This means someone who has a lot of maturity and experience in what he does, and it literally means "the person who can hold his liquor". He knows you really well. I enjoyed a lot our meetings when we could talk in a very informal way with a lot of drawings and intuition, and the improvement of the quality of the paper at the end of each meeting was an incredible source of happiness and satisfaction to me. Other than work, I really enjoyed the time spent with you and the other PhD students playing Klaverjassen and boards games, watching movies, going to conferences, and drinks of course. You are very humble, incredibly smart, maybe the most open-minded person I know and a very nice person. Thank you for everything. Mathias, thanks a lot for the opportunity you gave me to do a PhD. You are fascinated by research and have deep knowledge in optimization. You were willing to help me a lot when I was lost, and I was often lost during the first year of the PhD. For all of that, thank you. Dries, it was a real pleasure to share with you two of the courses you coordinated, even though some students were complaining, for whatever reasons. They always find a reason to complain. Al, ways. I really enjoyed our talks in game theory, auction theory, and mechanism design, even though I still have difficulties to understand the last one. You really helped me a lot to improve the quality of the papers, their content, and even with hundreds of courses and meetings in your calendar you always have found spots to talk about the projects or solve administrative issues quickly. I really enjoyed our board games sessions and the dinners at your place, where you have beaten us to the ground at every single game, especially at Foosball. It was always a nice moment. For all of

that, thank you.

Then, I would like to thank the assessment committee for taking the time to carefully read this thesis, Marco Scarsini, Frank Thuijsman, Tristan Tomala and Anna Zseleva. I would also like to thank Karin van den Boorn, Yolanda Paulissen and Vera Hoekstra, who are the most efficient and friendly secretarial officers I ever met. It was a real pleasure to work with you. Dank je wel !

As I have an online Defense, I do not officially have paranymphs, but there are many people to who I would have loved to ask this request. More precisely I would have loved Aditya, Niels, Luca, Caterina, Li, Diogo, Farzaneh, Niloufar, Dewi, Ilona, Shash, Tim, Veerle, Jasmine, Aida, Thijs, Elisa, Moritz, Francesco and Son to be my paranymphs.

I have spent an amazing time as a PhD student in the KE department, on the 4th floor of the SBE. I really enjoyed all the events we did together: hiking sessions, movies, week-end trips, board games, Sinterklaas secret presents, cooking sessions, Mario Kart sessions, D&D sessions, the bridge and the Klaverjassen clubs, the many lunches and dinners at Ginger, Cato by Cato, Piano B and Preuverij, tennis, running sessions, the uncountable number of beers at the Tribunal, at De Gouverneur and at Café Frape, to name a few. Even during Corona outbreak we kept having online PhD meetings. And of course, the PhD defense parties were always nice. My PhD would not have been the same without the numerous people with whom I spent those three amazing years.

Aditya, we started exactly at the same time. I remember that the first week I had to ask two to three times people to repeat in order to understand what they tried to tell me. But with you, I had to ask ten times. Fortunately, beer helps. I really enjoyed your calm, your sarcasm, and being positive and motivated in every circumstances. You are incredibly smart, you have always been humble with it, and have always been patient to explain. Well, except with one or two IB students. And except with me at chess and at English grammar, but I'm a hopeless case. I really enjoyed all the time spent with you and all the talks I could have

with you. It was a pleasure to host you in France, and to travel with you to conferences and week-end trips. Thanks a lot for everything.

Luca, I was really happy to have you as my office mate. You brought a nice atmosphere in the office. First by the nice geometric drawings you made on the board, second by introducing cat pictures and Italian puns on the walls, and of course by bringing a botanic garden in the office. That was a deep change with your predecessor who spent 90% of his office hours watching Twitch videos. I really enjoyed our talks in maths, stats, time-series and politics, even if I still don't understand much in time series, I know that this is cool and full of geometry and linear algebra, and I am grateful to have a free year to finally have time to read an introduction to these subjects. I really liked to teach with you, to watch movies, hike and have drinks to relax. I also thank you a lot for the organization of the two nice week-ends we had in Berlin and Prague, and your idea of investing in controllers to play Mario Kart on the beamer. You are an amazing cook, and I thank you for all the things I learned related to linear algebra, geometry, statistics and cooking during those three amazing years. Thanks also for all the nice words I learned in Italian, I'm sure that will help me to have new italian friends/girlfriends ! For all of that and for all the things I forgot, thanks a lot.

Niels, I realized how open-minded were Dutch people a week after I came in the Netherlands, but you kept impressing me the three last years. Your knowledge in History and Geography would make von Neumann feel jealous. You knew French history more than any of my French friends and likely some of my former history teachers, and you liked to make historical anecdotes while hiking, or during coffee breaks. And I loved it. Thank you for all the hiking sessions you organized, for your amazing talks, and for your outstanding PhD defense I keep laughing about.

Anna, you were the third person from Maastricht I have talked to, after the interview with Mathias and János. I remember the Skype call we had with Caterina, when my poor English level allowed me to barely

understand 20% of what you said (which was more than enough to get your opinion about the Dutch cuisine). You sent me an email more lengthy than this thesis with all the different nice activities to do in Maastricht, housing tips, and cool people to hang out with, in and outside the department, before I came to Maastricht. Furthermore, you also provided me all your notes of the Game Theory course of Dries. You have always been super nice with me, and you took care of everybody. You are an amazing teacher, an excellent researcher, a very nice person, and you are incredibly guifted in languages. I don't know anybody else who speaks fluently Bulgarian, Hungarian, English, Dutch, Russian, Hebrew and Italian. To make it short, I'm glad we are friends. Or shall I say, that you are a highly esteemed friend ? Thank you for everything.

Caterina we have started the PhD more or less at the same time. You have always been very nice, always in for a coffee break to gossip about the students or to think about a better world (not necessarily communist, but quite a bit). Furthermore, I really enjoyed playing tennis with you, even if we lost couple of balls, and the risk sessions, in which you have won way too often (I mean: every, single, time). I was also really impressed by your organization and your hard work. You could work on three projects at a time, while teaching, and still having your notes prepared for the next DnD session. You are funny, nice, really open-minded, and I really enjoyed spending those three years with you as a close friend. I wish you all the best. Thanks a lot.

Farzaneh, you are a really nice person. You have always been kind and friendly with everybody, and you proposed your help spontaneously all the time we needed some. I cannot imagine how would have been my moving without your help. You are an amazing cook, a very smart person, a great artist, you speak 103975430975 languages, and you have great heart. I am really happy to have you as a friend. Thank you for everything.

Niloufar, you have a huge personality. You have an impressive culture in cinema, literature and philosophy. You have opinions in everything,

and you like to defend them. You are full of life, determined, and a hard worker. I remember a couple of sessions we were working at the office until late evening. Thank you for everything.

Dewi, you are the most Dutch person I know. And it is a compliment. You are very organized, very direct, and you like to party. I liked your motivation for every event, your open mind on many subjects that French people do not like to talk about, and your enthusiasm about your work and teaching. For all the Sinterklaas gathering events you have organized, for all the cool running sessions, and for all the rest, thanks a lot, mum :)

Ilona, you are the very first person I talked to when I came to Maastricht, a bit lost. It was a great pleasure to talk to you, first about housing, and then politics, economics, bridge, DnD, solidarity, cooking, shampoos and teaching. You are smart, passionate, and open-minded person, even though a very slightly little bit stubborn sometimes. I learned a lot with you. Thank you for all the time we shared together.

Li, you are amazing. You are spontaneous, funny, very sociable, a great cook, and you are always in for games. Thanks, you a great friend, and a strong woff !

Diogo, it was a great pleasure to meet you. You taught me a lot about the mysteries of academic standards which I couldn't go though, and it is always a great time to see you, for work, drinks and parties. You also improved a lot my English skills, and made me discover many places close-by I didn't know anything about. I am happy to share two projects with you, and maybe more. Thanks for everything meu amigo.

Roland, you are a nice guy, ma gueule! Shashwat, you are awesome, it was great to party and meet you those during those three years. Aida, thank you so much for everything, tu paella catalana está incredible! I wish you all the best, and hope to see you soon. Son, you are a nice guy. I wish you all the best, and hope to play Mario Kart again. Francesco, it was great to meet you, I hope you enjoyed the wine bottle, and I hope to see you after the pandemic. Thijs, thank you man, it was great to

talk to you and complain about the many terrible things in academia. I wish you all the best. Tuomas, thank you for all the nice bridge games we have played, and for your amazing home made honey! Anastas, thank you for all the seminar you have organized. It was always a pleasant time. Andrés, thank you for your relevant questions during the talks, and for your amazing summer course on game theory. It was a real pleasure. Elisa, dit Lili, it was great to have a French buddy in the department. Thank you for the nice atmosphere you brought to the floor. Jasmine, thank you for your help and your support. You are very humble and very smart. It was a real pleasure to share some French-English sessions. Sorry for the fact that I gave up in Dutch.. I hope to see you soon to play board games, and wish you all the best in Gent! Tim, Veerle and Marc, thank you both of you for the nice atmosphere you brought to the department, especially for lunchtime. You are super smart, fantastic teachers and researchers. I hope to see again each time I come back to Maastricht. Dank je well!

I would also like to thank Waldemaar, Alex, Yicong, Waldemar, Hanno, Rasmus, Adam, Sean, Qian Qian, Lee Meng, Li Yang, Arkadi, Verena.

I would like to thank my co-authors. Hui Min, you were an excellent master student when we did the project with you. You were very enthusiastic, you had a great intuition, and you never complained during the corona outbreak when we asked you a lot to do, not even any complaint when we started to put meetings during week-ends. When I wrote those lines, the project just received very good reviews from the International Journal of Game Theory, and it is mostly thanks to you. I wish you all the best in your job in Eindhoven, and I wish we can have dinner with János this summer when the outbreak will be over and the paper published. Panayotis, thank you for your help, your patience and your hard work. You did almost everything in the first article when I was totally lost. Dennis Meier, thank you. You are a nice guy and a very smart researcher, it was a pleasure to meet you.

I would also like to thank all the friends I met during conferences. Emilien, it's always a pleasure to see you and talk to you, and it was

a great time to meet you again in Cracow, and I hope to see you soon. Tristan Garrec, thanks for your advice when I wanted to apply for a PhD in Toulouse, and your very good remarks while for the poster in Cracow. Bas Dietzenbacher, I'm glad we met in Paris, and many times after. I really like your topics, and your addiction to board games. Congratulations for your position in Maastricht, I wish you will have a great time there, even if I have no doubt you will. Also thank you Bruno Ziliotto, Xavier Venel and Miquel Oliu-Barton, Marion Hallet and Adriana Alventosa Baños.

Outside the PhD life, I had (a few) friends that I had a great time with in Maastricht. Aiste, I improved a loooot my skills in English at the beginning of my PhD thanks to our exchanges, and I'm really grateful to you. I really enjoyed all the events we have attended together, except maybe this weird Reggae concert, but it was still very funny. I'm really happy to count you as a friend, and I wish you will enjoy the next few years in Germany. Thank you for everything, and I see you soon. Marc, thank you for organizing all the amazing Dungeons & Dragons sessions we had. It was always a nice moment, and it would not have been possible without you. Mehrad and Adam, it was always a pleasure to see both of you at a party. Thank you for all those great times. I would also like to thank all the amazing farmers from the organic market of Maastricht. You are all super nice, and you do a great job. Keep it up! And of course, the cooks of Cato by Cato, the Preuverij and Ginger.

I am grateful to all the teachers, tutors and professors I had the chance to meet. Especially, I would like to thank Mrs. Caillaud. I was 11 years old, and you were the first teacher who ever gave me this fascination for maths. You explained all the concepts in a very informal way, with a lot of drawings and funny problems with dogs, pyramids and cinema subscriptions. That year, I received my very first award in maths. Two years later, you were my teacher again, and I was in the top 100 in France in the concours Kangourou. It is not false to say that you contributed a lot to what I am today, thank you so much. In preparatory school, I also got two great teachers in maths, Mr. Prézaut and Mrs. Tardif, and an amazing teacher of French and Latin, Mrs. Hiribarren. In

my high-school I again had three brilliant and nice teachers in maths: Mr. Bertrand, Mr. Péant and Mrs. Lézé. Other than maths, I had amazing teachers in biology (Mr. Bessouet), in physics (Mrs. Navarro), in history (Mrs. Delorge), and in philosophy (Mrs. Haran). In my two first years of bachelor, I really enjoyed the lectures of my two physics and chemistry teachers, Mrs. Brillant and Mr. Ranz, and my teacher of philosophy, Mrs. Séguier-Leblanc. All my professors from my third year Bachelor were extraordinary. I especially enjoyed the lectures of Frédéric Bourgeois in dynamic systems, Dominique Hulin in complex analysis, Jean-Michel Bismuth in Lebesgue integral theory and Pierre-Guy Plamandon in Group theory. And of course, the best lectures I have ever attended the course of Measure Theory taught by Laurent Moonens, who became a great friend. Thank you for everything. I would like to thank more personally Tristan Tomala, who wrote the very first chapter of the very first book I have read in Game Theory, who was my first lecturer in a Game theory course, who supervised my master thesis, and who accepted to be in the committee of my PhD defense. You were fascinated by the results you presented, gave a lot of intuition and invested a lot of time and effort to make it understandable, and you wrote everything in a very precise mathematical way. I enjoyed doing my master thesis with you. You gave me a lot of freedom, and I really enjoyed working on revision games. For all of that, thank you. I also really enjoyed the lectures of Blanche Buet in Basic Optimization, Rida Laraki in Dynamic Games, Vianney Perchet in Incomplete Information Games, Olivier Hudry in Complexity Theory and Christophe Picouleau in Graphs and Approximated Complexity. I would like to thank more personally Rida Laraki for a couple of things. First, I realized only once you were my lecturer that you created the Majority Judgment, which changed totally my point of view on politics and democracy. Second, you helped me a lot to find a PhD position, and you even recommended me for a PhD position to Hervé Moulin. Thank you for everything. I also thank Jérôme Renault, who answered to the numerous questions I had while reading the French book of Game Theory you wrote with Rida and Tristan when I was a master student, for the time you spent for considering my application for a PhD thesis and finally to aware me

about the PhD position Mathias proposed online, who led me where I spent three wonderful years. Thank you very much, all of you. I am happy, and proud, of being an element of the set of math teachers.

During those three years in Maastricht, I was happy to keep in touch with friends from my years as a master student in Orsay. Benoît Tran, merci pour tout mec, c'était un plaisir de partager ces deux années de maths avec toi. Bravo pour ton postdoc au Brésil, et j'espère qu'on aura d'autres occasions de se capter sur Panam ou ailleurs. Francois, alias Mr. Delgove, alias Chef, merci de m'avoir introduit à ce merveilleux groupe de thésards. Toutes les pauses cafés à faire des pendus, des Citadelles et, quelques fois, à parler de maths. Merci également à toute la bande, Benoît Robert, Benjamin, Jacques, Raphy et Samuel pour les quelques très bons moments qu'on a passé ensemble. Nhi, merci pour toutes nos conversations et ton implication incroyable pour la place des femmes dans la recherche, pour les séminaires en ligne ainsi que pour tout ce que tu fais pour rendre la vie des thésards d'Orsay plus supportable. Tu es quelqu'un d'exceptionnel! Pierre, merci pour toutes nos discussions de maths, de Magic et de politique, surtout au niveau progressisme. J'ai notamment adoré nos discussions sur la prostitution, les drogues et les revendications féministes. Je te souhaite de tout cœur bon courage et j'espère qu'on se reverra vite. Le LABO, merci pour ces trois années dans cette super colloc, c'est toujours un plaisir de vous retrouver. Célina, dit Dame Thérèse, merci pour tous nos voyages ensemble, perdus dans la forêt des Fagnes, à Boulogne, à Maastricht ou à Paris. T'as bonne humeur inconditionnelle fait chaud au cœur. Merci pour tous ces moments passés avec toi, et j'espère qu'il y en aura beaucoup d'autres. Merci à tous les amis du master d'optimisation de Paris Sud, Amin, Ayoub, Benoît, Daniel, Mehdi et Sélim, c'est toujours un plaisir de vous revoir.

From the time I left the Bask country, in 2013, there are not many friends from there I kept in touch with. Those ones, for sure, are friends for life. Adri, on se connaît depuis le CE2, et hormi quelques cheveux blancs, t'as pas changé. C'est toujours un plaisir de te voir. Bon courage pour fuir la capitale, et j'espère qu'on pourra continuer à se faire des

escapades en Europe pour les cinquante années à venir. Flo (Cabail), t'es comme un frère, je suis super fier de ce que t'es devenu, de tes choix, de ton ouverture d'esprit, et c'est toujours un plaisir de savoir qu'on peut se confier. Merci pour tout. Flo (Seyer), Alrick, Benji, Vincent, Benito, Charlène, Amandine, Cyril c'est toujours un bonheur de savoir que je vous retrouverai quand je rentre au pays. Votre sympathie est plus contagieuse que le covid et c'est toujours un bonheur de vous retrouver, simplement, autour d'une bière, d'un café, d'une pala, d'un Catane ou d'un tarot. Pourvu que ça dure. Merci pour tout. Merci à Joana et Charly pour tous les bons souvenirs passés ensemble au lycée, ainsi que les retrouvailles plus ou moins tous les étés, et plus particulièrement merci Jo pour la merveilleuse couverture de cette thèse, qui jusqu'à présent fait l'hunanimité. Merci Romain, dit Trinch, pour tous ces cafés à Bordeaux à parler recherche, et à tes blagues qui ne font rire que moi. Merci Gabriel, dit Gabi, pour toutes ces conversations politiques, même avec des avis divergents, ainsi que sur l'Education et la Recherche. Ton humour noir me fera toujours rire. Merci à la bande des 5/2, Loris Jeff et Thomas pour les quelques occasions qu'on a de se revoir. Enfin merci Antoine et Vincent Da Costa pour toutes ces mémorables branlées sur Age of Empires. Je vous kiffe.

Last but not least, I would like to thank my family. Papa, merci pour tout ce que tu nous as appris, en commençant par le ski, le tennis, les échecs, la politique, les cartes, ta passion pour la randonnée, parmi tant d'autres, ainsi que tous les bons moments qu'on a pu partager ensemble, notamment dans les Pyrénées que tu aimes tant. Maman, merci pour toute l'éducation que tu nous as inculquée, pour tous les sacrifices que tu as fait pour que nous arrivions aussi loin tous les trois, pour nous avoir fait découvrir la musique, le dessin, et toutes les activités qu'on a pu faire durant notre enfance. Tu nous as également inculqué des valeurs: celles du partage, de l'empathie et de la solidarité, et qui sont aujourd'hui encrées dans nos vies. Merci pour tout. Florian, merci pour tous les délirs qu'on partage sur des choses aussi improbables que Palace, Jimmy Neutron, les Guignols de l'info et tant d'autres. À qui d'autre pourrais-je crier "oh regarde: IL DANSE !" en faisant tourner

un objet λ sans passer pour un taré? Tu as subi une évolution remarquable ces dernières années, par ton ouverture d'esprit, ta tolérance, ton altruisme. Ton culot aussi. Tu as tenté bien des concours, sans que jamais les échecs des uns n'entravent ta détermination pour les autres, ce qui t'as valu l'une des meilleures écoles commerce, un poste haut placé dans une grande banque beaucoup de responsabilité au Secours Populaire. Tu es cultivé, bien plus que je ne le serai jamais, et tu sauras toujours nous surprendre. Tu es libre de tes choix, tu es charismatique, et bien que j'avais des doutes jusque récemment, je suis sûr que tu en feras bon usage. Enfin, tu as un humour de merde. Assez merdique en tous cas pour que j'en sois le premier adepte. Pour tout ça et pour tant d'autres choses, je suis fier de t'avoir pour frère. Merci gp. Agnès je ne sais pas par où commencer. Tu es brillante, sincère, directe, ouverte d'esprit, empathique, humaine. Tu assumes tes choix, et vit pleinement ta liberté. Tu as vécu tant d'aventures, et en vivras tant d'autres. Tu es un exemple, et je suis fier d'être ton frère. Merci pour tout. Merci à vous tous, je vous aime.

To all of you mentionned above, all the ones who attended my defense, and to you, who is reading those lines, thank you.

Benoit Duvocelle
Maastricht
22/02/2021

Contents

1
Introduction

In various situations such as on a market, a soccer match, or in klaverjassen, many decisions have to be taken, and those decisions will induce an output: the profit of a firm, or the identity of the winning team. And sometimes, decisions have to be taken regularly (each day, each month, each year), and the outcomes of those decisions might not be clear yet when the players take these decisions.

In this thesis, we will use Game Theory to analyse such situations in which firms or individuals interact repeatedly with each other. In Game Theory such a setting is called a repeated game, and the decision makers are called players. The possible decisions that players can make in a game are called strategies.

We are going to use two solution concepts to analyse repeated games. The first concept is based on the notion of regret, and the second is Nash equilibrium.

Regret is a notion based on an ex-post analysis. The regret induced by a strategy is the difference between the expected payoff of this strategy and the maximal payoff that the player could have gotten in retrospect.

Then, each player is interested in choosing a strategy that minimizes his regret. This notion of (dynamic) regret was introduced by Zinkevich in 2013. Before that, the notion of regret was based on the best static strategy, where a strategy is static if it uses the same action at every time period.

Another central solution concept in Game Theory is the Nash equilibrium, named after its inventor, the Nobel prize winner John Nash. A Nash equilibrium is a situation in which every player uses a strategy in such a way that, knowing the strategies of the other players, has no incentive to deviate from his strategy for another one.

In Chapter 2, we introduce a game over infinitely many periods, in which players have to take decisions at each period. Players have imperfect information with respect to their payoff and the strategies of their opponents: they have an indication about what they should play today based on the decisions taken in the previous days in order to maximize their payoff, but do not know exactly how much they will earn. Moreover, they do not observe the actions chosen by their opponents at any time.

In such an environment, the computation of equilibrium is hard, as players are uncertain about the past choices of the other players, and therefore they do not know their own payoff. For this reason we use regret as our solution concept in this chapter.

We study an algorithm that gives the players an advice on which choice to make based on what the players learnt yesterday. We show that if the players follow the advice of the algorithm, they are guaranteed a low regret against a specifically chosen dynamic strategy that does not change too much over time. Moreover, if the sequence of stage games does not vary too much over time, and if the players follow the advice of the algorithm, then the sequence of strategies converges to an equilibrium in the long run.

In Chapter 3, we present an algorithm in a general framework. We present known convergence results of this algorithm, and very recent

results of strong convergence. Our main focus in this chapter is to present its application to solve traffic network problems. We assume that commuters are on a network, and that each of them wants to travel from one node (his origin) to another (his destination). A commuter wants to reach his destination as close as possible to a specific time : he can arrive earlier or later than the designated time, but the bigger the gap is, the worse it is for the commuter.

The algorithm, taking into account all these parameters, derives the best flow of commuters over the network, recommends to each commuter the path he should take, and decides at which time he should start form his origin. A relevant solution concept in this setting is the Nash equilibrium: the output of the algorithm is publicly observed, and given that other commuters are obediently follow the recommendation of the algorithm, it is a best response to do so as well.

In a Nash equilibrium, commuters with identical characteristics (same origin, same destination and same desired arrival time) may have to take different paths in order not to overload the network. We finally compare the performance of this algorithm to other algorithms that were previously designed for finding the equilibrium of a congested network. In several cases, the algorithm we propose outperforms other algorithms.

While in Chapters 2 and 3 the main focus is to find an algorithm which implements an equilibrium, in Chapter 4 the main focus is on existence of Nash equilibria, and the stability of Nash equilibria under perturbations of the parameters of the game. Existence of Nash equilibrium is a fundamental problem in Game Theory, and establishing the existence can be a challenging problem.

In Chapter 4 we study a two-player game, played over infinitely many periods. An object is hidden in one of finitely many locations, called states, and two players compete to find the object. The players sequentially choose one of the different states. If the object is there, then the active player wins and the game stops. Otherwise, the object moves randomly to another state and the next player is asked to choose.

One can imagine, as on the cover of the thesis, two pirates looking for a treasure based on information written on a treasure map, but at the same time, the annoying Molly the mole moves the treasure below the ground. A more concrete (and more serious) application would be to consider two laboratories competing to find a cure for a viral disease. At each period, one of them receives a research grant that allows to invest in one of many different technologies in order to find a cure. However, the virus mutates over time, and a good technology today might become useless in a few months.

There are two relevant notions of equilibrium that we use as a solution concept in this chapter: the Nash equilibrium, and its refinement, called subgame perfect equilibrium. The latter solution concept has the advantage that it takes into account mistakes of the players. More precisely, a subgame perfect equilibrium is a strategy profile such that whatever happened in the past, it induces an equilibrium in the remaining game.

We identify conditions under which subgame perfect equilibria do exist. We also show that Nash equilibrium does not always exist, but that a slightly weaker solution concept, called ε-equilibrium, always exists. We show that those equilibria are stable to perturbations, such as time discounting, or if the game ends in a long but finite horizon. Finally we study structural properties of the game.

Detailed discussion of the Chapters

Chapter 2

In Chapter 2 we examine the long-term behavior of regret-minimizing agents in time-varying games with continuous action spaces. In its most basic form, (external) regret minimization guarantees that an agent's cumulative payoff is no worse in the long run than that of the agent's best fixed action in hindsight. Going beyond this worst-case guarantee, we consider a dynamic regret variant that compares the agent's accrued

rewards to those of *any* sequence of play. By properly adapting a restart procedure pioneered by Besbes et al. [7], we show that players are able to achieve no dynamic regret against any test sequence whose total variation grows sublinearly with the horizon of play. In particular, specializing to a wide-class of no-regret strategies based on mirror descent, we derive explicit rates of dynamic regret minimization, both in expectation and with high probability. We then leverage these results to show that players are able to stay close to Nash equilibrium in time-varying monotone games – and even converge to equilibrium if the sequence of stage games admits a limit.

Chapter 3

In Chapter 3 we use a class of strongly convergent primal-dual schemes for solving variational inequalities defined by a Lipschitz continuous and pseudo-monotone map in infinite-dimensional Hilbert spaces, which have been studied by Dennis Meier in [22]. This novel numerical scheme is based on Tseng's forward-backward-forward scheme, which is known to display weak convergence, unless very strong global monotonicity assumptions are made on the involved operators. We test the performance of the algorithm in the computationally challenging task to find dynamic user equilibria in traffic networks and verify that our scheme is at least competitive to state-of-the-art solvers, and in some cases even improves upon them.

Chapter 4

In Chapter 4 we introduce a discrete-time search game, in which two players compete to find an object first. The object moves according to a time-varying Markov chain on finitely many states. The players know the Markov chain and the initial probability distribution of the object, but do not observe the current state of the object. The players are active in turns. The active player chooses a state, and this choice is observed by the other player. If the object is in the chosen state, this player wins and

the game ends. Otherwise, the object moves according to the Markov chain and the game continues to the next period. We show that these games admit a value, and for any error-term $\varepsilon > 0$, each player has a pure (subgame-perfect) ε-optimal strategy. Interestingly, a 0-optimal strategy does not always exist. The ε-optimal strategies are robust in the sense that they are 2ε-optimal on all finite but sufficiently long horizons, and also 2ε-optimal in the discounted version of the game provided that the discount factor is close to 1. We derive results on the analytic and structural properties of the value and the ε-optimal strategies. Moreover, we examine the performance of the finite truncation strategies, which are easy to calculate and to implement. We devote special attention to the important time-homogeneous case.

2

Learning in time-varying games

In this Chapter [1], we examine the long-term behavior of regret-minimizing agents in time-varying games with continuous action spaces.

2.1 Introduction

A key requirement for decision-making in unknown, non-stationary environments is the minimization of *regret:* no rational agent would want to realize in hindsight that the decision policy they employed was strictly inferior to a crude policy prescribing the same action at each stage. Of course, depending on the context, this minimal worst-case guarantee admits several refinements. For starters, agents could tighten their comparison benchmarks and, instead of comparing their accrued rewards to those of the best *fixed* action, they could compare them to general test sequences that evolve over time. Moreover, if agents

[1]This chapter is based on [23]. I would like to thank the referees for their helpful comments and discussion. This research has been partially supported by the COST Action CA16228 "European Network for Game Theory" (GAMENET)

interact with one another and their rewards are determined by a fixed underlying mechanism – that of a *non-cooperative game* – there are much finer criteria that apply, chief among them that of convergence to a Nash equilibrium.

Since real-world scenarios are rarely stationary and typically involve several interacting agents, both issues are of high practical relevance and should be treated in tandem. With this in mind, the central question that we seek to address in this chapter is as follows: *What is the long-run behavior of strategic agents that follow an adaptive no-regret policy when the underlying game evolves over time in an unknown, unpredictable manner?*

Our contributions and related work

Our analysis revolves around two main axes, as outlined below:

Dynamic regret minimization

First, we examine the case where an agent seeks to minimize his regret against a fixed (but otherwise arbitrary) stream of payoff functions. As a comparison benchmark, we posit that the agent compares the rewards accrued by their chosen sequence of play to any other test sequence (as opposed to a fixed action). In particular, as a special case, this definition of regret also includes the agent's best *dynamic* policy in hindsight, i.e., the sequence of actions that maximizes the payoff function encountered at each stage of the process.

This measure of regret is considerably more ambitious than the standard definition of external regret (which only considers constant sequences as performance benchmarks). One of its antecedents is the notion of *shifting regret* which considers piecewise constant benchmark sequences and keeps track of the number of "shifts" relative to the horizon of play [13, 44]. Much closer in spirit is the dynamic regret definition of Besbes et al. [7] which takes as a benchmark the sequence of instantaneous payoff maximizers (individual best responses) of each stage. The analysis of

[7] shows that, in full generality, it is *not* possible to achieve no dynamic regret if the total variation of the sequence of payoff functions grows linearly with the horizon of play. However, if this growth is *sublinear*, attaining no dynamic regret *becomes* possible by means of a restarting procedure that amortizes the dynamic regret incurred by a policy with no *static* regret over a sequence of time windows of increasing length.

Our first result is an extension of this heuristic to the context of arbitrary test sequences: As we show in Section 2.4 , by using a carefully crafted restart procedure in the spirit of [7], an agent is able to achieve no regret relative to *any* slowly-varying test sequence – i.e., any test sequence whose total variation grows sublinearly with the horizon of play. In particular, under mild regularity assumptions, a sequence of concave payoff functions with sublinear total variation admits a slowly-varying sequence of maximizers (so we recover the original result of Besbes et al. [7] as a special case). Results of this flavor are not entirely new in the learning literature. Indeed, looking at the variation of comparator sequences is also prominent in the classical online convex optimization framework, which is a zero sum game of a single decision maker against nature (see e.g. [44] and references therein). However, to the best of our knowledge, the general strategic setting has never been studied. An important benefit of our formulation is that it shifts weight from the total variation of the payoff functions encountered to that of their maximizers: trivially, if each payoff function is shifted by a random constant at each stage, the total variation of the stream of functions encountered might be linear even though the set of maximizers remains the same. Perhaps less trivially, our result shows that the change of a function away from its maximum set does not really matter for dynamic regret minimization: the key limitation is the variation of the maximizers.

To quantify the above, we focus on decision-making policies based on *online mirror descent* (OMD), a broad class of algorithms that includes as special cases the online gradient descent (OGD) method of Zinkevich [113], the well-known multiplicative weights (MW) algorithm [3], and

many others.[2] First introduced by Nemirovski and Yudin in the context of convex programming [87], mirror descent is one of the most widely used policies for achieving no *static* regret in online optimization and is well-known to be black-box optimal in that respect [100]. Our analysis in Section 2.6 extends these results to the context of dynamic regret minimization: by leveraging the restart procedure of Besbes et al. [7], we show that the class of policies under consideration attains a regret minimization rate of $\mathcal{O}(T^{2/3}V^{1/3})$ relative to test sequences with total variation V over T stages.

This regret minimization rate essentially coincides with the bound obtained by Besbes et al. [7] for slowly-varying streams of payoff functions. Going beyond this basic bound, and under mild assumptions for the information available to the agent at each stage, we also show that this rate holds not only on average, but also with overwhelming probability. We make all this precise in Section 2.6, where we establish a large deviations bound for the algorithm's dynamic regret.

Game-theoretic learning

The second element of our analysis concerns the underlying assumption that the sequence of payoff functions encountered by an agent is *oblivious*. Specifically, this means that, when calculating the payoffs that the agent would have obtained by employing a different sequence of actions, the stream of payoff functions encountered by the agent remains the same. This assumption is well-grounded in the literature of (adversarial) online optimization as a minimal worst-case guarantee; however, in a game-theoretic setting, it is more difficult to justify. For instance, if two regret-minimizing players are involved in a game, the payoff functions encountered by one player will be influenced by the action choices of the other. Thus, if one player were to employ a

[2] Above, the word "descent" should really be "ascent", because players are typically maximizers in game theory. To avoid this clash in terminology, we use the more neutral term "proximal method".

different sequence of actions, the other player would most likely respond differently, altering in this way the sequence of payoff functions encountered by the first player (and vice versa). As such, even *dynamic* regret minimization against a *fixed* stream of payoff functions does not provide strong optimality guarantees in a game-theoretic setting.

To address this issue, we consider a general multi-agent framework where, at every stage $t = 1, 2, \ldots$, each player's payoff function is determined by the action choices of all other players via a non-cooperative game $\mathcal{G}^t$. The stage game $\mathcal{G}^t$ may vary with time, but the rules governing its evolution are not a priori known to the players (so the rationalistic viewpoint of the literature on repeated/dynamic games does not apply). In this context, the main question we seek to address is as follows: *If all players follow a dynamic regret minimization policy, do their actions eventually get close to a Nash equilibrium at each stage?*

Without further assumptions, the answer to this question is "no", even when players face the same stage game throughout. Indeed, (external) regret minimization in finite games guarantees that the players' empirical frequencies of play converge to the game's *Hannan set* (also known as the set of coarse correlated equilibria) [50], but this set may contain strategies that assign positive weight *only* to dominated strategies [109] (and these cannot be supported at a Nash equilibrium). In fact, in two-player zero-sum games, no-regret learning may cycle indefinitely without converging, always remaining a bounded distance away from the game's Nash set [75, 78]. On the other hand, if the game satisfies a monotonicity condition known as *diagonal strict concavity* (DSC) [95], Mertikopoulos and Zhou recently showed that no-regret policies based on mirror descent converge to Nash equilibrium with probability 1, even with imperfect gradient information on the players' side [79].

Building on our dynamic regret analysis, our first result is that if *a*) the stage games are monotone; and *b*) they admit a slowly-varying sequence of Nash equilibria, no-regret learning with a judiciously chosen restart schedule as in [7] allows players to remain close to the game's evolving equilibrium (at least on average). More to the point, as a refinement

of this result, we show that if the sequence of stage games converges to a strictly monotone game, then the induced sequence of play converges to a Nash equilibrium thereof. Importantly, this last result holds globally (i.e., independently of the algorithm's initialization) and with probability 1, irrespective of the magnitude of the noise entering the players' gradient signals.

In our view, these results constitute a first step towards understanding the behavior of utility-maximizing agents in unknown, online environments, where the top-down, "rationalistic" viewpoint of the theory of repeated and dynamic games does not apply. Specifically, even though the standard rationality postulates do not hold in our setting (knowledge of the game being played, common knowledge of rationality, etc.), our results show that no-regret learning can still lead to equilibrium in dynamic environments. We find this property of regret minimization particularly appealing, as it provides an important link between online learning and the emergence of rational behavior in strategic environments that evolve over time.

Notation

Given a finite-dimensional vector space $\mathcal{V}$ with norm $\|\cdot\|$, we will write $\mathcal{V}^*$ for its (algebraic) dual, $\langle y, x\rangle$ for the duality pairing between $y \in \mathcal{V}^*$ and $x \in \mathcal{V}$, and $\|y\|_* = \sup\{\langle y, x\rangle : \|x\| \leq 1\}$ for the dual norm of $y \in \mathcal{V}^*$. If $\mathcal{X}$ is a closed convex subset of $\mathcal{V}$, we write $\mathrm{ri}(\mathcal{X})$ for its relative interior and $\mathrm{diam}(\mathcal{X}) = \sup\{\|x' - x\| : x, x' \in \mathcal{X}\}$ for its diameter. Finally, if x^t, $t = 1, 2, \dots$, is a sequence of elements of $\mathcal{X}$, we will write $x^{\mathcal{T}} \equiv (x^t)_{t\in\mathcal{T}}$ for the subfamily of elements indexed by a subset $\mathcal{T}$ of $\mathbb{N}$.

2.2 Preliminaries

2.2.1 Concave games

The focal point of our analysis will be games with a finite number of players and continuous action sets. Specifically, every player $i \in$

$\mathcal{N} \equiv \{1, \dots, N\}$ is assumed to select an *action* x_i from a compact convex subset $\mathcal{X}_i$ of a finite-dimensional normed space $\mathcal{V}_i$. Subsequently, based on each player's individual objective and the *action profile* $x = (x_i; x_{-i}) \equiv (x_1, \dots, x_N)$ of all players' actions, every player receives a *reward*, and the process repeats.

In more detail, writing $\mathcal{X} \equiv \prod_{i \in \mathcal{N}} \mathcal{X}_i$ for the game's *action space* and $\mathcal{V} \equiv \prod_{i \in \mathcal{N}} \mathcal{V}_i$ for its corresponding ambient space,[3] we assume that each player's reward is determined by an associated *payoff* (or *utility*) *function* $u_i \colon \mathcal{X} \to \mathbb{R}$.[4] Since players are not assumed to "know the game" (or even that they are involved in one) these payoff functions might be a priori unknown, especially with respect to the dependence on the actions of other players. Following [95], our only structural assumption for u_i will be that

$$u_i(x_i; x_{-i}) \text{ is concave in } x_i \text{ for all } x_{-i} \in \mathcal{X}_{-i}, \tag{2.1}$$

where, in obvious notation, $\mathcal{X}_{-i} = \prod_{j \neq i} \mathcal{X}_j$ denotes the action space of all players other than the i-th one. For regularity purposes, it will also be convenient (albeit not necessary) to assume that each u_i is C^1-smooth in x; to streamline our presentation, these will be our standing assumptions in what follows.

With all this in hand, a *concave game* will be a tuple $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ with players, action spaces and payoffs defined as above. Below, we briefly discuss some recurring examples of such games:

Example 2.2.1 (Mixed extension of finite games). In *finite games*, each player $i \in \mathcal{N}$ chooses an action (or *pure strategy*) α_i from a finite set $\mathcal{A}_i$. The players' payoffs are then determined by the pure strategy profile

[3] Unless explicitly mentioned otherwise, we will assume that $\mathcal{V}$ is endowed with the norm $\|x\|^2 = \sum_i \|x_i\|^2$. Also, to streamline our presentation, we will use the same notation for the norm of each factor space $\mathcal{V}_i$ and rely on the context to resolve any ambiguities.

[4] For book-keeping reasons, it will be convenient to assume that u_i is actually defined on an open neighborhood of $\mathcal{X}$ in $\mathcal{V}$. However, none of our calculations depend on this device, so we do not make this assumption explicit.

$\alpha = (\alpha_i)_{i\in\mathcal{N}}$ of all players' actions via a collection of payoff functions $u_i\colon \mathcal{A} \equiv \prod_j \mathcal{A}_j \to \mathbb{R}$.

In the *mixed extension* of a finite game, players are allowed to randomize their decisions by playing *mixed strategies*, i.e., probability distributions $x_i \in \Delta(\mathcal{A}_i)$ with the interpretation that $x_{i\alpha_i}$ is the probability of choosing action $\alpha_i \in \mathcal{A}_i$. In this case (and in a slight abuse of notation), the expected payoff to player i under the mixed strategy profile $x = (x_i)_{i\in\mathcal{N}}$ is

$$u_i(x_i; x_{-i}) = \sum_{\alpha_1\in\mathcal{A}_1} \cdots \sum_{\alpha_N\in\mathcal{A}_N} u_i(\alpha_1, \dots, \alpha_N)\; x_{1,\alpha_1} \cdots x_{N,\alpha_N}. \tag{2.2}$$

Since each player's mixed strategy space $\mathcal{X}_i = \Delta(\mathcal{A}_i)$ is convex and u_i is individually linear in x_i, it follows that mixed extensions of finite games are concave in the sense of (2.1).

Example 2.2.2 (Saddle-point problems). Consider a saddle-point problem of the general form

$$\min_{x_1\in\mathcal{X}_1} \max_{x_2\in\mathcal{X}_2} f(x_1, x_2) \tag{SP}$$

where each feasible region $\mathcal{X}_i$, $i = 1, 2$, is a compact convex subset of $\mathcal{V}_i \equiv \mathbb{R}^{d_i}$ and $f\colon \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ is assumed to be convex in x_1 and concave in x_2. Letting $u_1 = -f$ and $u_2 = f$, the saddle-point problem (SP) can be seen as a zero-sum game with player set $\mathcal{N} = \{1, 2\}$ and payoff functions u_i, $i = 1, 2$. Since f is convex-concave, the resulting game $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ is itself concave in the sense of (2.1).

Example 2.2.3 (Resource allocation auctions). Consider a service provider with a splittable *resource* (bandwidth, computing cores, ad display time, etc.). Fractions of this resource can be leased to a set of N bidders (players) who can place monetary bids $x_i \geq 0$ for the utilization of said resource up to each player's total budget b_i. Once all bids are in, resources are allocated proportionally to each player's bid, i.e., the i-th player gets $\rho_i = (qx_i)/(c + \sum_{j\in\mathcal{N}} x_j)$ units of the auctioned

resource, where q denotes the total amount of the resource and $c \geq 0$ represents an "entry barrier" for bidding on it. A simple model for the utility of player i is then given by

$$u_i(x_i; x_{-i}) = [g_i \rho_i - x_{is}], \tag{2.3}$$

with g_i denoting the marginal gain of player i from acquiring a unit slice of resources. Writing $\mathcal{X}_i = [0, b_i]$ for the action space of player i, it is easy to see that the resulting game $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ is concave in the sense of (2.1).

Many other important scenarios can be formulated as concave games; for an incomplete list, see [59, 79, 24, 92, 101, 74] and references therein.

2.2.2 Nash equilibrium

The most prevalent solution concept in game theory is that of a *Nash equilibrium* (NE), defined here as any action profile $\hat{x} \in \mathcal{X}$ that is resilient to unilateral deviations, i.e.,

$$u_i(\hat{x}_i; \hat{x}_{-i}) \geq u_i(x_i; \hat{x}_{-i}) \quad \text{for all } x_i \in \mathcal{X}_i \text{ and all } i \in \mathcal{N}. \tag{NE}$$

By the classical existence theorem of Debreu [17], concave games with compact action spaces always admit a Nash equilibrium. Moreover, thanks to the individual concavity of the game's payoff functions, Nash equilibria can also be characterized via the first-order optimality condition

$$\langle v_i(\hat{x}), x_i - \hat{x}_i \rangle \leq 0 \quad \text{for all } x_i \in \mathcal{X}_i,\ i \in \mathcal{N}, \tag{2.4}$$

where $v_i(x)$ denotes the individual payoff gradient of the i-th player, i.e.,

$$v_i(x) = \nabla_{x_i} u_i(x_i; x_{-i}), \tag{2.5}$$

and ∇_{x_i} denotes differentiation with respect to the variable x_i.[5]

Geometrically, this characterization of Nash equilibria simply means that $v_i(\hat{x})$ belongs to the polar cone

$$\mathrm{PC}_{\mathcal{X}_i}(\hat{x}_i) = \{y_i \in \mathcal{V}_i^* : \langle y_i, x_i - \hat{x}_i \rangle \leq 0 \text{ for all } x_i \in \mathcal{X}_i\}. \tag{2.6}$$

of $\mathcal{X}_i$ at $\hat{x}_i$, i.e., $v_i(\hat{x})$ forms an obtuse angle with any displacement vector of the form $z_i = x_i - \hat{x}_i$, $x_i \in \mathcal{X}_i$. By concavity, this means that $u_i(\hat{x}_i + tz_i; \hat{x}_{-i})$ is nonincreasing in t, so (NE) holds for all $x_i \in \mathcal{X}_i$. We will use this geometric intuition freely in what follows.

2.2.3 Variational inequalities and monotonicity

The first-order characterization (2.4) of Nash equilibria can be written more concisely (but otherwise equivalently) as a variational inequality of the form

$$\langle v(\hat{x}), x - \hat{x} \rangle \leq 0 \quad \text{for all } x \in \mathcal{X} \tag{VI}$$

where

$$v(x) = (v_1(x), \dots, v_N(x)) \tag{2.7}$$

denotes the players' individual gradient profile at $x \in \mathcal{X}$. As a result, finding a Nash equilibrium of a concave game amounts to solving the (Stampacchia) variational inequality problem (3.5). This important observation has been the starting point of an extensive literature at the interface of game theory and optimization; for an overview, we refer the reader to [89, 25, 99, 79] and references therein.

Most of this literature has focused on problems where the vector field $v(x)$ of individual payoff gradients satisfies the monotonicity condition

$$\langle v(x') - v(x), x' - x \rangle \leq 0 \quad \text{for all } x, x' \in \mathcal{X}. \tag{MC}$$

[5] We adopt here the established convention of treating $v_i(x)$ as an element of the dual space $\mathcal{V}_i^*$ of $\mathcal{V}_i$. We do so in order to emphasize the fact that $v_i(x)$ acts naturally on vectors $z_i \in \mathcal{V}_i$ via the (linear) directional derivative mapping $z_i \mapsto u_i'(x; z_i) = d/dt|_{t=0}\, u_i(x_i + tz_i; x_{-i})$.

Owing to the link between (MC) and the theory of monotone operators in optimization, games that satisfy (MC) are commonly referred to as *monotone games* [99, 79].[6] In particular, mirroring the corresponding terminology from operator theory, we will say that a game is:

a) *Strictly monotone* if (MC) holds as a strict inequality when $x' \neq x$.

b) *Strongly monotone* if there exists a positive constant $\beta > 0$ such that

$$\langle v(x') - v(x), x' - x \rangle \leq -\beta \|x' - x\|^2 \quad \text{for all } x, x' \in \mathcal{X}. \tag{2.8}$$

Obviously, we have the inclusions "strongly monotone" $\subsetneq$ "strictly monotone" $\subsetneq$ "monotone", mirroring the corresponding chain of inclusions "strongly convex" $\subsetneq$ "strictly convex" $\subsetneq$ "convex" for convex functions.

The set of Nash equilibria of a monotone game – which coincides with the solution set of (3.5) – is itself convex and compact; in particular, if the game is strictly or strongly monotone, its Nash set is a singleton [25, 4, 79]. Moreover, on account of (MC), Nash equilibria of monotone games can also be characterized in this case as solutions of the Minty variational inequality

$$\langle v(x), x - \hat{x} \rangle \leq 0 \quad \text{for all } x \in \mathcal{X}. \tag{MVI}$$

This property of Nash equilibria of monotone games will play a crucial role in our analysis and we will make free use of it in the rest of our paper; for a more detailed discussion, we refer the reader to [79].

In terms of applications, monotone games constitute a very rich and diverse class. Cournot oligopoly models [80], atomic splittable conges-

[6] Rosen [95] uses the name *diagonal strict concavity* (DSC) for a weighted variant of (MC) which holds as a strict inequality when $x' \neq x$. Hofbauer and Sandholm [54] use the term "stable" to refer to a class of population games that satisfy a condition similar to (MC), while Sandholm [97] and Sorin and Wan [101] respectively call such games "contractive" and "dissipative". We use the term "monotone" throughout to underline the connection of (MC) with operator theory and variational inequalities.

tion games in networks with parallel links [92, 101], signal covariance optimization problems in wireless communications [99, 74], and many other problems where online decision-making is the norm. In particular, the class of monotone games contains all games that admit a (strictly) concave *potential*, i.e., a function $f\colon \mathcal{X} \to \mathbb{R}$ such that

$$v_i(x) = \nabla_{x_i} f(x) \quad \text{for all } x \in \mathcal{X},\, i \in \mathcal{N}. \tag{2.9}$$

In view of all this, monotone games will comprise an important part of our analysis, especially in the context of convergence to Nash equilibrium.

2.3 Problem setup

We now turn to a detailed description of our model for time-varying games. In its most general form, this boils down to the following sequence of events:

Time-varying games: sequence of events

Require: set of players $i \in \mathcal{N}$, action spaces $\mathcal{X}_i \subseteq \mathbb{R}^{d_i}$

for $t = 1, 2, \dots$ **do**
  set $\mathcal{G} \leftarrow \mathcal{G}^t(\mathcal{N}, \mathcal{X}, u^t)$ # stage game
  for all $i \in \mathcal{N}$ **do**
    play $X_i^t \in \mathcal{X}_i$ # choose action
    receive $u_i^t(X_i^t; X_{-i}^t)$ # collect reward
    get signal Y_i^t # receive feedback
  end for
end for

The core ingredients of the above framework are *a*) the sequence of games $\mathcal{G}^t$, $t = 1, 2, \dots$, encountered by the players at each stage of the process; and *b*) the sequence of feedback signals Y^t. We discuss both in detail below.

Stage game sequence

Our standing assumptions for the sequence of stage games $\mathcal{G}^t \equiv \mathcal{G}^t(\mathcal{N}, \mathcal{X}, u^t)$ will be that *a*) they are concave (in the sense of 2.2.1); and *b*) only the players' payoff functions evolve over time. In particular, the set of players $\mathcal{N}$ and their action spaces $\mathcal{X}_i$, $i \in \mathcal{N}$, are assumed to remain the same for all t. Nonetheless, if the payoff function of some player $i \in \mathcal{N}$ is identically equal to zero at stage t and their actions have no impact on other players, this player is effectively removed from the stage in question. Similar devices can also account for action spaces that vary with time (at least, as long as they are contained in some compact set), so the model at hand is sufficiently general for our purposes.

Two important special cases of this framework are when:

1. There is a single player and the sequence of stage games is fixed in advance (but is otherwise arbitrary). This unilateral framework is the gold standard in *online learning* (cf. the various definitions of regret in the next section) and has the property of being *oblivious*, i.e., a different sequence of play would yield the same sequence of payoff functions.

2. The sequence of stage games is *constant*, i.e., $\mathcal{G}^t = \mathcal{G}$ for some fixed game $\mathcal{G}$. This case is the norm in *game-theoretic learning* and, in addition to comprising several players, its main difference with the online learning framework is that it is *not* oblivious. In general, given the dependence of the payoff function of player i on the actions of all other players, a different sequence of actions $X^t \equiv (X_i^t; X_{-i}^t)$ would yield a different sequence of payoff functions $u_i(\cdot, X_{-i}^t)$.

In view of the above, a time-varying game can be seen as an amalgamation of these two classical frameworks: in particular, the dependence of a player's payoff function on the stage index t is both *explicit* (via the sequence of stage games $\mathcal{G}^t$) and *implicit* (via the sequence of actions

chosen by all other players). This "dual" dependence on t will play a key role in what follows, especially in the equilibrium analysis of 2.7.

In terms of regularity, we will be assuming throughout that the players' individual payoff gradients are uniformly bounded, i.e., there exists some finite $G_i \geq 0$ such that

$$\|v_i^t(x)\|_* \leq G_i \quad \text{for all } t = 1, 2, \ldots, \text{and all } x \in \mathcal{X}, \tag{2.10}$$

where, in obvious notation, we have set

$$v_i^t(x) = \nabla_{x_i} u_i^t(x_i; x_{-i}). \tag{2.11}$$

Other than that, we make no prior assumptions about the process that defines each stage game. For instance, this evolution could be random (i.e., $\mathcal{G}^t$ could be determined by some randomly drawn parameter θ^t), it could depend on the players' actions (e.g., as in the literature on dynamic/repeated games), some underlying (hidden) Markov chain, or any other mechanism. We also do not assume that such information is available to the players: from their individual point of view, each player is involved in a repeated decision process where the choice of an action returns a reward, and they have no knowledge of the mechanism generating this reward. The reason for this "agnostic" approach is that, in many cases of practical interest, the standard rationality postulates (full rationality, common knowledge of rationality, etc.) are not realistic: for instance, a commuter choosing a route to work has no way of knowing how many commuters will be making the same choice, let alone how these choices might influence their thinking for the next day.

Signals and feedback

The second core ingredient of our model is the feedback available to each player after choosing an action. In tune with the "bounded rationality" framework outlined above, we do not assume that players

can observe the actions of other players, their payoffs, or any other such information. Instead, we take a "partial monitoring" approach – see e.g., [96, 15, 66, 68] and references therein – and we only posit that every player $i \in \mathcal{N}$ receives a (random) signal Y_i^t from space containing payoff-relevant information for each stage t. In particular, we will be assuming that the random signal received by player i at stage t is of the general form

$$Y_i^t = v_i^t(X^t) + U_i^t, \tag{2.12}$$

where $X^t = (X_i^t; X_{-i}^t) \in \mathcal{X}$ is the profile of actions at stage t (possibly random), and U_i^t is a stochastic perturbation of the realized payoff gradient, modeling observational noise in the feedback signal. As such, under this model, the signals of player $i \in \mathcal{N}$ are drawn from the dual space $\mathcal{Y}_i \equiv \mathcal{V}_i^*$ of the ambient space $\mathcal{V}_i$ of $\mathcal{X}_i$.

Remark. In optimization-theoretic terms, the signal model (2.12) means that each player has access to a (stochastic) *first-order oracle*, i.e., a black-box feedback mechanism providing (possibly noisy) gradient information at each stage. From a game-theoretic standpoint, the motivation for this signal model is rooted in the case where players can only observe their realized, in-game payoffs (the so-called *bandit* setting). In this extremely low-information environment, it is still possible to construct an oracle of the form (2.12) by means of a simultaneous perturbation stochastic approximation (SPSA) procedure as in [102, 26]; we defer the details of this analysis to a future paper.

In terms of measurability, it is tacitly assumed that both $X^t = (X_i^t)_{i\in\mathcal{N}}$ and $Y^t = (Y_i^t)_{i\in\mathcal{N}}$ are defined over a common (complete) probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and all expectations or probabilities will be taken with reference to this space. The *private history* of player i is then defined as the filtration $\mathbb{F}_i = (\mathcal{F}_i^t)_{t=0}^{\infty}$, where

$$\mathcal{F}_i^t = \sigma(X_i^1, Y_i^1, \dots, X_i^t, Y_i^t) \tag{2.13}$$

is the σ-algebra generated by the player's chosen actions and signals received up to stage t (inclusive), while $\mathcal{F}_i^0$ is chosen so as to complete

the filtration (not necessarily in a trivial way). Aggregating over all players, the *history of play* is likewise defined as the filtration $\mathbb{F} = (\mathcal{F}^t)_{t=0}^{\infty}$, where

$$\mathcal{F}^t = \sigma(X^1, Y^1, \dots, X^t, Y^t). \tag{2.14}$$

Given all this, we posit that a player's action at stage $t+1$ is determined by the player's private history up to stage t, i.e.,

$$X_i^{t+1} = \mathtt{Alg}_i(X_i^1, Y_i^1, \dots, X_i^t, Y_i^t) \tag{2.15}$$

for some measurable deterministic function $\mathtt{Alg}_i$, which will be referred to as an *algorithm* (or *repeated game strategy*).[7] This means that, for all $t = 1, 2, \dots$, X_i^{t+1} is $\mathcal{F}_i^t$-predictable – or, collectively, that X^{t+1} is $\mathcal{F}^t$-predictable.

In the rest of our paper, and unless explicitly mentioned otherwise, our blanket assumptions for the signal process Y_i^t will be as follows:

1. Y_i^t is an unbiased estimator of $v_i^t(X^t)$, i.e.,

$$\mathbb{E}[Y_i^t \,|\, \mathcal{F}^{t-1}] = v_i^t(X^t) \tag{2.16a}$$

   for all $t = 1, 2, \dots$ and all $i \in \mathcal{N}$.

2. Y_i^t has uniformly bounded second-order moments, i.e.,

$$\mathbb{E}[\|Y_i^t\|_*^2 \,|\, \mathcal{F}^{t-1}] \leq M_i^2 \tag{2.16b}$$

   for some $M_i < \infty$ and all $t = 1, 2, \dots$, $i \in \mathcal{N}$.

Alternatively, the above is equivalent to asking that the noise process U_i^t is zero-mean with finite mean square, i.e.,

$$\mathbb{E}[U_i^t \,|\, \mathcal{F}^{t-1}] = 0 \tag{2.17a}$$

[7] To avoid superfluous notation, we are omitting in (2.15) the dependence of X and Y on ω, and we are treating $\mathtt{Alg}_i$ as a function of variable arity (so as to drop its dependence on t).

and

$$\mathbb{E}[\|U_i^t\|_*^2 \,|\, \mathcal{F}^{t-1}] \leq \sigma^2 \tag{2.17b}$$

for some finite $\sigma < \infty$ and all $t = 1, 2, \ldots$, $i \in \mathcal{N}$. Both of these assumptions can be relaxed in various ways (e.g., by asking that Y_i^t is accurate on average only up to some bias term, or by considering higher-order moments of U_i^t), but it will be more convenient to state our results with both these assumptions in play.

As a special case, the "noiseless" regime $U_i^t = 0$ will be sometimes referred to as *perfect information*. However, to avoid clashes with existing terminology (especially within the literature on dynamic and repeated games), we stress here that players are never assumed to observe the actions of other players: even with perfect information, only gradient observations are available at each stage.

2.4 Regret minimization

2.4.1 Types of regret

As we discussed in the introduction, a minimal worst-case requirement for online decision-making is that of regret minimization. In the non-stationary framework of the previous section, the (external) regret of player $i \in \mathcal{N}$ over a window of play $\mathcal{T} \subseteq \mathbb{N}$ is defined as

$$\mathrm{Reg}_i(\mathcal{T}) \equiv \max_{x_i \in \mathcal{X}_i} \sum_{t \in \mathcal{T}} [u_i^t(x_i; X_{-i}^t) - u_i^t(X^t)], \tag{2.18}$$

i.e., as the cumulative difference between the cumulative payoff of the focal player $i \in \mathcal{N}$ under the sequence of play $X^t \in \mathcal{X}$, $t = 1, 2, \ldots$, and that of the player's best fixed action over the time window $\mathcal{T}$.[8]

[8] By "window" we refer here to an interval of successive positive integers, i.e., $\mathcal{T}$ is of the form $\mathcal{T} = \{a, a+1, \ldots, b\}$ for some $a, b \in \mathbb{N}$. Unless explicitly mentioned otherwise, we will only work with intervals of this type.

Then, specializing to the case where $\mathcal{T} = \{1, \dots, T\}$, we will say that the sequence X^t leads to *no* (*external*) *regret* if

$$\limsup_{T\to\infty} \frac{1}{T} \operatorname{Reg}_i(\mathcal{T}) \leq 0 \quad \text{for all } i \in \mathcal{N}, \tag{2.19}$$

i.e., if every player's regret grows at most sublinearly with the horizon of play:

$$\operatorname{Reg}_i(\mathcal{T}) = o(T) \quad \text{for all } i \in \mathcal{N}. \tag{2.20}$$

By the individual concavity of the players' payoff functions, the payoff difference in the definition of the regret can be bounded from above as

$$u_i^t(x_i; X_{-i}^t) - u_i^t(X^t) \leq \langle v_i^t(X^t), x_i - X_i^t \rangle, \tag{2.21}$$

for any reference action $x_i \in \mathcal{X}_i$ and all $t \in \mathcal{T}$. Consequently, a player's regret can be itself bounded from above as

$$\operatorname{Reg}_i(\mathcal{T}) \leq \operatorname{Gap}_i(\mathcal{T}), \tag{2.22}$$

where

$$\operatorname{Gap}_i(\mathcal{T}) = \max_{x_i \in \mathcal{X}_i} \sum_{t\in\mathcal{T}} \langle v_i^t(X^t), x_i - X_i^t \rangle \tag{2.23}$$

represents a linearized, player-specific regret measure that we call the *gap function* of player i. Hence, to achieve no regret, it suffices to design an algorithm guaranteeing that $\operatorname{Gap}_i(\mathcal{T}) = o(T)$ for every player $i \in \mathcal{N}$. This linearization device has been the starting point of most no-regret strategies in the literature (see e.g., [14, 100, 79, 88, 12] and references therein), and we will use it freely in the rest of our paper.

Of course, an important limitation in the definition (2.18) of a player's regret is that it compares the sequence of accrued rewards to that of the best *fixed* action in hindsight. Since the players' payoff functions evolve over time, a player following a policy satisfying (2.19) may still incur a substantial loss relative to a *non-constant* sequence of actions. Thus, to get a finer performance benchmark, consider instead the *dynamic regret*

of player i relative to a *test sequence* $x_i^t \in \mathcal{X}_i$, $t = 1, 2, \ldots$, defined as

$$\mathrm{DynReg}_i(\mathcal{T}; x_i^{\mathcal{T}}) = \sum_{t \in \mathcal{T}} [u_i^t(x_i^t; X_{-i}^t) - u_i^t(X^t)]. \tag{2.24}$$

Then, as in the static case, we say that a sequence of play X^t leads to no dynamic regret relative to the test sequence $x^t \in \mathcal{X}$, $t = 1, 2, \ldots$, if

$$\limsup_{T \to \infty} \frac{1}{T} \mathrm{DynReg}_i(\mathcal{T}; x_i^{\mathcal{T}}) \leq 0 \quad \text{for all } i \in \mathcal{N}, \tag{2.25}$$

i.e., if every player's dynamic regret relative to x^t grows at most sublinearly with the horizon of play T.

As a special case, if $x_i^t \in \operatorname{argmax}_{x_i \in \mathcal{X}_i} u_i^t(x_i; X_{-i}^t)$ is a sequence of individual best responses of player i to the realized sequence of play X_{-i}^t of all other players, we get

$$\mathrm{DynReg}_i(\mathcal{T}; x_i^{\mathcal{T}}) = \sum_{t \in \mathcal{T}} \max_{x_i \in \mathcal{X}_i} [u_i^t(x_i; X_{-i}^t) - u_i^t(X^t)]. \tag{2.26}$$

Comparing this expression to the definition (2.18) of the external regret of player i, we see that the order of summation and maximization have been exchanged. In this way, we recover the original definition of Besbes et al. [7], suitably extended to our multi-agent setting: in the long run, each player's accrued rewards are no worse than what the player would have obtained by best-responding to the sequence of play of all other players at each stage.

2.4.2 Dynamic regret minimization

Our main goal in the rest of this section will be to provide a universal bound for the players' dynamic regret relative to arbitrary test sequences. To do so, we will again rely on the individual concavity of

the players' payoff functions to write

$$\mathrm{DynReg}_i(\mathcal{T}; x_i^{\mathcal{T}}) \leq \sum_{t\in\mathcal{T}} \langle v_i^t(X^t), x_i^t - X_i^t\rangle, \tag{2.27}$$

just as in the static case. Then, motivated by the recent analysis of Besbes at al. [7], we will decompose a player's dynamic regret into two components: one driven by the gap function (3.13) over smaller windows of play, and the other measuring the *variation* of the test sequence x_i^t over time, as defined below:

Definition 2.4.1. The variation of a test sequence $x_i^t \in \mathcal{X}_i$, $t = 1, 2, \ldots$, over the window $\mathcal{T} \subset \mathbb{N}$ is defined as

$$\mathrm{Var}_i(\mathcal{T}; x_i^{\mathcal{T}}) = \sum_{t\in\mathcal{T}} \|x_i^{t+1} - x_i^t\|, \tag{2.28}$$

with the convention that $x^{t+1} = x^t$ if $t = T$.

To proceed with the decomposition outlined above, let $\mathcal{T}_{i,1}, \ldots, \mathcal{T}_{i,m_i}$ be a partition of the time window $\mathcal{T} = \{1, \ldots, T\}$ into m_i *batches*, each of size

$$\Delta_i = \lfloor T/m_i \rfloor, \tag{2.29}$$

with the possible exception of the last one (which might be smaller). We then have:

Lemma 2.4.2. *The dynamic regret of the i-th player relative to a test sequence $x_i^t \in \mathcal{X}_i$, $t = 1, 2, \ldots$, satisfies*

$$\mathrm{DynReg}_i(\mathcal{T}; x_i^{\mathcal{T}}) \leq \sum_{\ell=1}^{m_i} \mathrm{Gap}_i(\mathcal{T}_{i,\ell}) + G_i \Delta_i \,\mathrm{Var}(\mathcal{T}, x_i^{\mathcal{T}}). \tag{2.30}$$

Proof. The proof is an elementary computation building on an idea of [7]. Indeed, dropping the player index i for notational clarity, individual

concavity yields

$$\mathrm{DynReg}(\mathcal{T}; x^{\mathcal{T}}) = \sum_{\ell=1}^{m} \mathrm{DynReg}(\mathcal{T}_\ell; x^{\mathcal{T}_\ell}) \leq \sum_{\ell=1}^{m} \sum_{t \in \mathcal{T}_\ell} \langle v^t(X^t), x^t - X^t \rangle. \tag{2.31}$$

Now, fixing a reference action $p_\ell \in \mathcal{X}$ for each batch $\ell = 1, \ldots, m$, let

$$I_\ell = \sum_{t \in \mathcal{T}_\ell} \langle v^t(X^t), p_\ell - X^t \rangle, \tag{2.32a}$$

and

$$J_\ell = \sum_{t \in \mathcal{T}_\ell} \langle v^t(X^t), x^t - p_\ell \rangle, \tag{2.32b}$$

so that the linearized dynamic regret over each batch can be written as

$$\sum_{t \in \mathcal{T}_\ell} \langle v^t(X^t), x^t - X^t \rangle = I_\ell + J_\ell. \tag{2.33}$$

To bound I_ℓ, note that

$$I_\ell = \sum_{t \in \mathcal{T}_\ell} \langle v^t(X^t), p_\ell - X^t \rangle \leq \max_{x \in \mathcal{X}} \sum_{t \in \mathcal{T}_\ell} \langle v^t(X^t), x - X^t \rangle = \mathrm{Gap}(\mathcal{T}_\ell). \tag{2.34}$$

Subsequently, to bound J_ℓ, let p_ℓ be the first element of the test sequence x^t over the ℓ-th batch $\mathcal{T}_\ell$. Then,

$$\begin{aligned} \sum_{t \in \mathcal{T}_\ell} \langle v^t(X^t), x^t - p_\ell \rangle &\leq \sum_{t \in \mathcal{T}_\ell} \|v^t(X^t)\|_* \cdot \|x^t - p_\ell\| \\ &\leq G \sum_{t \in \mathcal{T}_\ell} \|x^t - p_\ell\| \\ &\leq G\Delta \max_{t \in \mathcal{T}_\ell} \|x^t - p_\ell\| \end{aligned}$$

$$\leq G\Delta \sum_{t\in\mathcal{T}_\ell} \|x^{t+1} - x^t\| = G\Delta \operatorname{Var}(\mathcal{T}_\ell; x^{\mathcal{T}_\ell}), \quad (2.35)$$

where we used Young's inequality in the first line and the triangle inequality in the last one. Our claim then follows by summing over each batch $\ell = 1, \dots, m$. □

Lemma 2.4.2 suggests that minimizing a player's regret relative to a rapidly-varying test sequence x_i^t may be difficult (if not downright impossible). On the other hand, if the test sequence under study is *slowly-varying* in the sense that

$$\operatorname{Var}(\mathcal{T}; x^{\mathcal{T}}) = o(|\mathcal{T}|), \quad (2.36)$$

then it might be feasible to attain no (dynamic) regret by properly tweaking the batch size Δ_i in the regret decomposition (2.30).

This observation was the starting point of the analysis of [7] who proposed breaking the horizon of play into batches of a carefully chosen size, and then running on each batch an algorithm that incurs low static regret (i.e., sublinear relative to the size of the batch). In our game-theoretic setting, this boils down to each player choosing a batch size Δ_i and breaking the horizon of play into $m_i = \lceil T/\Delta_i \rceil$ successive time windows $\mathcal{T}_{i,1}, \dots, \mathcal{T}_{i,m_i}$, each of size Δ_i (except possibly the last one). Then, at every window $\mathcal{T}_{i,\ell}$, each player $i \in \mathcal{N}$ updates their actions following an (as yet unspecified) algorithm $\texttt{Alg}_i$ which is restarted every Δ_i stages.

Formally, this restart procedure can be encoded in pseudocode form as follows:

Algorithm 1 Batch restart (player indices suppressed)

Require: Horizon T, batch size Δ, choice algorithm `Alg` as in (2.15)

set $t \leftarrow 1$ #step counter
choose $X^1 \in \mathcal{X}$ #initialization
repeat
  set $\tau \leftarrow \lfloor (t-1)/\Delta \rfloor \Delta + 1$ #augment every Δ stages
  play $X^t \in \mathcal{X}$ #play chosen action
  get signal Y^t #receive feedback
  set $X^{t+1} \leftarrow \mathtt{Alg}(X^\tau, Y^\tau, \dots, X^t, Y^t)$ #update action
  $t \leftarrow t+1$ #next stage
until $t > T$ #end play

In view of all this, if the restart frequency is chosen as a function of the variation of the test sequence under study, we have:

Theorem 2.4.3. *Consider a time-varying game $\mathcal{G}^t \equiv \mathcal{G}^t(\mathcal{N}, \mathcal{X}, u^t)$, $t = 1, 2, \dots$, and let $\mathtt{Alg}_i$ be an algorithm of the general form* (2.15) *such that*

$$\mathbb{E}[\mathrm{Gap}_i(\mathcal{T})] \leq C_i \sqrt{T} \tag{2.37}$$

for some $C_i > 0$, for all $T \in \mathbb{N}$ and all time intervals $\mathcal{T} \subseteq \mathbb{N}$ of length T. Suppose further that $x_i^t \in \mathcal{X}_i$, $t = 1, 2, \dots$, is a test sequence enjoying the variation bound

$$\mathrm{Var}_i(\mathcal{T}; x_i^{\mathcal{T}}) \leq V_i^T, \tag{2.38}$$

for some $V_i^T \geq 1$. Then, if $\mathtt{Alg}_i$ is rebooted every $\Delta_i = \lceil (T/V_i^T)^{2/3} \rceil$ stages we have

$$\mathbb{E}[\mathrm{DynReg}_i(\mathcal{T}; x_i^{\mathcal{T}})] \leq (2C_i + 3G_i) T^{2/3} (V_i^T)^{1/3}. \tag{2.39}$$

In particular, if x_i^t is slowly-varying (i.e., $V_i^T/T \to 0$ as $T \to \infty$), we have

$$\limsup_{T \to \infty} \mathbb{E}[\mathrm{DynReg}_i(\mathcal{T}; x_i^{\mathcal{T}})] \leq 0. \tag{2.40}$$

Remark. In the above, expectations are taken with respect to the randomness of the players' signals (and induced action sequences).

Proof. Taking expectations on both sides of the bound (2.30) yields

$$\mathbb{E}[\mathrm{DynReg}_i(\mathcal{T}; x_i^{\mathcal{T}})] \leq \sum_{\ell=1}^{m_i} \mathbb{E}[\mathrm{Gap}_i(\mathcal{T}_{i,\ell})] + G_i \Delta_i \,\mathrm{Var}(\mathcal{T}, x_i^{\mathcal{T}}), \tag{2.41}$$

where $m_i = \lceil T/\Delta_i \rceil$ is the number of restarts up to stage T (inclusive). Then, with $\mathrm{Var}_i(\mathcal{T}; x_i^{\mathcal{T}}) \leq V_i^T$ and $|\mathcal{T}_{i,\ell}| \leq \Delta_i = \lceil (T/V_i^T)^{2/3} \rceil$, this bound becomes:

$$\begin{aligned}
&\mathbb{E}[\mathrm{DynReg}_i(\mathcal{T}; x_i^{\mathcal{T}})] && (2.42)\\
&\leq m_i C_i \sqrt{\Delta_i} + G_i \Delta_i V_i^T \\
&\leq (T/\Delta_i + 1) C_i \sqrt{\Delta_i} + G_i \Delta_i V_i^T \\
&\leq C_i [1 + T^{2/3} (V_i^T)^{1/3} + (T/V_i^T)^{1/3}] + G_i [V_i^T + T^{2/3} (V_i^T)^{1/3}] \\
&\leq (3C_i + 2G_i) T^{2/3} (V_i^T)^{1/3} && (2.43)
\end{aligned}$$

where, in the last line, we used the fact that $V_i^T \geq 1$. □

2.5 Distributed learning

In this section we present a class of distributed learning algorithms based on *online mirror descent* (OMD), a family of algorithms which, together with the closely related "follow the regularized leader" (FTRL) protocol, comprise one of the most widely used algorithmic schemes for no-regret learning in online optimization – for a partial survey, see [87, 5, 57, 86, 106, 16, 88, 85, 100, 79] and references therein.

Viewed abstractly, the basic idea of mirror descent (or, in our case, "ascent") is as follows: if player $i \in \mathcal{N}$ plays $x_i \in \mathcal{X}_i$ and receives the gradient signal $y_i \in \mathcal{Y}_i$, the algorithm generates a new action x_i^+ by

taking an "approximate gradient" step from x_i along y_i. Formally, this can be written as

$$x_i^+ = P_i(x_i, \gamma y_i) \tag{2.44}$$

where

1. γ is a step-size parameter controlling the weight attributed to the signal y_i.
2. $P_i \colon \mathcal{X}_i \times \mathcal{Y}_i \to \mathcal{X}_i$ is a "proximal mapping" (discussed in detail below) which determines the exact way in which the step along y_i is taken.

Remark. Because the prox-mapping P_i plays a defining role in the players' action selection process, and to avoid clashes between the term "descent" and the fact that players are treated as maximizers in our setting, we will refer to (2.44) as a *prox-method* (PM).

Now, given a convex subset $\mathcal{C}$ of some ambient vector space $\mathcal{V} \cong \mathbb{R}^d$, the prototypical example of a prox-mapping is the Euclidean projector

$$\begin{aligned} P(x, y) = \Pi_{\mathcal{C}}(x + y) &\equiv \operatorname*{argmin}_{x' \in \mathcal{C}} \{\|x + y - x'\|_2^2\} \\ &= \operatorname*{argmin}_{x' \in \mathcal{C}} \{\langle y, x - x' \rangle + \tfrac{1}{2}\|x' - x\|_2^2\} \end{aligned} \tag{2.45}$$

i.e., the closest-point projection of $x + y$ onto $\mathcal{C}$.[9] Going beyond the Euclidean case, the key novelty of prox-methods is to replace the distance term $\frac{1}{2}\|x' - x\|_2^2$ in (2.45) with a (possibly non-symmetric) "divergence" defined by means of a *distance-generating function* (DGF) $h \colon \mathcal{C} \to \mathbb{R}$, itself assumed to be continuous and K-strongly convex, i.e.,

$$h(tx + (1-t)x') \leq th(x) + (1-t)h(x') - \frac{K}{2}t(1-t)\|x' - x\|^2 \tag{2.46}$$

[9] Note here that, in writing $x + y$, we are blurring the lines between primal vectors $x \in \mathcal{V}$ and dual vectors $y \in \mathcal{V}^*$. This distinction is reinstated in the second line of (2.45) where $y \in \mathcal{V}^*$ is paired properly to $x - x' \in \mathcal{V}$.

for all $x, x' \in \mathcal{C}$ and all $t \in [0,1]$. For technical reasons (and in a slight abuse of notation), we further assume that the subdifferential $\partial h(x) = \{y \in \mathcal{V}^* : h(x') \geq h(x) + \langle y, x' - x\rangle\}$ of h admits a *continuous selection*: specifically, letting $\mathcal{C}^\circ \equiv \operatorname{dom} \partial h = \{x \in \mathcal{C} : \partial h(x) \neq \varnothing\}$ denote the domain of subdifferentiability of h, we posit that there exists a continuous function $\nabla h : \mathcal{C}^\circ \to \mathcal{V}^*$ such that $\nabla h(x) \in \partial h(x)$ for all $x \in \mathcal{C}^\circ$.[10] The *Bregman divergence* induced by h is then defined as

$$D_h(x', x) = h(x') - h(x) - \langle \nabla h(x), x' - x\rangle \quad \text{for all } x' \in \mathcal{C},\, x \in \mathcal{C}^\circ, \tag{2.47}$$

and the associated *prox-mapping* $P : \mathcal{C} \times \mathcal{V}^* \to \mathcal{C}$ is given by

$$P(x, y) = \operatorname*{argmin}_{x' \in \mathcal{C}} \{\langle y, x - x'\rangle + D_h(x', x)\} \quad \text{for all } x \in \mathcal{C}^\circ,\, y \in \mathcal{V}^*. \tag{2.48}$$

Remark. Because P is defined only for states $x \in \mathcal{C}^\circ$, the set $\mathcal{C}^\circ$ will be sometimes referred to as the *prox-domain* of h.

Before continuing, it will be instructive to provide some standard examples of prox-mappings:

Example 2.5.1 (Euclidean projections). We begin by recovering the archetypal example of Euclidean projections. To do so, let $h(x) = \frac{1}{2}\|x\|^2$. Since h is convex and subdifferentiable throughout $\mathcal{C}$, we have $\mathcal{C}^\circ = \mathcal{C}$ and $\nabla h(x) = x$ is a continuous selection of $\partial h(x)$ for all $x \in \mathcal{C}$. Hence, the associated Bregman divergence is

$$D_h(x', x) = \tfrac{1}{2}\|x'\|_2^2 - \tfrac{1}{2}\|x\|_2^2 - \langle x, x' - x\rangle = \tfrac{1}{2}\|x' - x\|_2^2, \tag{2.49}$$

and the induced prox-mapping is given by (2.45) for all $x \in \mathcal{C}$, $y \in \mathcal{V}^*$.

Example 2.5.2 (Entropic regularization). Let $\mathcal{C} = \{x \in \mathbb{R}^d_+ : \sum_{j=1}^d x_j = 1\}$ denote the unit simplex of $\mathcal{V} = \mathbb{R}^d$. A very widely used distance-generating function for this geometry is the (negative)

[10] By standard results in convex analysis [93], we have $\mathcal{C}^\circ \subseteq \operatorname{ri} \mathcal{C} \subseteq \mathcal{C}$. Note also that we are making use of the standard convention that $h(x) = +\infty$ if $x \in \mathcal{V} \setminus \{\mathcal{C}\}$.

Gibbs-Shannon entropy $h(x) = \sum_{j=1}^{d} x_j \log x_j$. By inspection, the domain of (sub)differentiability of h is $\mathcal{C}^\circ = \operatorname{ri}\mathcal{C}$, and the resulting Bregman divergence is given by the *Kullback–Leibler* (KL) expression

$$D_h(x', x) = \sum_{j=1}^{d} x'_j \log\left(\frac{x'_j}{x_j}\right), \tag{2.50}$$

valid for all $x \in \mathcal{C}^\circ$, $x' \in \mathcal{C}$. In turn, this gives rise to the prox-mapping

$$P(x, y) = \frac{(x_j \exp(-y_j))_{j=1}^{d}}{\sum_{j=1}^{d} x_j \exp(-y_j)} \tag{2.51}$$

for all $x \in \mathcal{C}^\circ$, $y \in \mathcal{V}^*$. The update rule $x^+ = P(x, y)$ is widely known in the literature as the *multiplicative weights* (MW) algorithm and plays a central role for learning in multi-armed bandit problems and finite games [3, 30, 52, 90].

Example 2.5.3 (Fermi-Dirac regularization). Let $\mathcal{C} = [0, 1]$ and let $h(x) = x\log(x) + (1-x)\log(1-x)$ be the (negative) Fermi-Dirac entropy. Then, $\mathcal{C}^\circ = (0, 1)$ and the induced prox-mapping is given by the expression

$$P(x, y) = \frac{x \exp(-y)}{1 - x + x \exp(-y)}, \tag{2.52}$$

valid for all $x \in (0, 1)$, $y \in \mathbb{R}$.

With all this at hand, the general class of game-theoretic learning algorithms that we will consider in the rest of this chapter will be given by the recursion

$$X_i^{t+1} = P_i(X_i^t, \gamma_i^t Y_i^t) \tag{2.53}$$

where, in detail:

1. $t = 1, 2, \ldots$ denotes the stage of the process.

2. $X_i^t \in \mathcal{X}_i$ is the action played by player i at stage t.

Algorithm 2 Prox-method for distributed learning (player indices suppressed)

Require: prox-mapping $P\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$, step-size sequence $\gamma^t \geq 0$,
sequence of stage games $\mathcal{G}^t \equiv \mathcal{G}^t(\mathcal{N}, \mathcal{X}, u^t)$

initialize $X^1 \in \operatorname{argmin}_{x_i \in \mathcal{X}_i} h_i(x_i)$ # initialization
for $t = 1, 2, \ldots$ **do**
  set $\mathcal{G} \leftarrow \mathcal{G}^t(\mathcal{N}, \mathcal{X}, u^t)$ # stage game definition
  play $X^t \in \mathcal{X}$ # play chosen action
  get signal $Y^t \in \mathcal{Y}$ # receive feedback
  set $X^{t+1} \leftarrow P(X^t, \gamma^t Y^t)$ # update action
  $t \leftarrow t + 1$ # next stage
end for

3. $Y_i^t \in \mathcal{Y}_i$ is the signal received by player i at stage t, assumed throughout to satisfy the unbiasedness assumption $\mathbb{E}[Y_i^t \,|\, \mathcal{F}^{t-1}] = v_i^t(X^t)$ (cf. 2.3).

4. γ_i^t is a player-specific step-size sequence (assumed nonincreasing).

5. $P_i\colon \mathcal{X}_i \times \mathcal{Y}_i \to \mathcal{X}_i$ denotes the prox-mapping of player i, itself derived from some distance-generating function $h_i\colon \mathcal{X}_i \to \mathbb{R}$ as above.

In particular, unless explicitly mentioned otherwise, all repeated game strategies described in 2.3 will be henceforth assumed to be of the (Markovian) from

$$\mathtt{Alg}_i(X_i^1, Y_i^1, \ldots, X_i^t, Y_i^t) = P_i(X_i^t, \gamma_i^t Y_i^t). \tag{2.54}$$

For concreteness, we also provide a pseudocode implementation of this prox-based learning protocol as Algorithm 2 above.

2.6 Explicit regret bounds

We show in this section how Theorem 2.4.3 can be applied in the specific case where the algorithm of each player is the prox strategy described in the previous section. For this, we have to understand the static regret suffered by this strategy. Given the time varying game $\{u^t\}_{t\geq 1}$, the signal process $\{Y_i^t\}_{t\geq 1}$ of the form (2.12), as well as step-size sequence $\{\gamma_i^t\}_{t\geq 1}$ and a prox-mapping P_i. Suppose that each player executes Algorithm 2. This induces a repeated game strategy $\{X^t\}_{t\in\mathbb{N}}$. Our analysis of dynamic regret of this strategy starts with a basic inequality bounding the regret of any fixed action over a finite time horizon. Results of this flavor are well-known form the online learning literature. Still, for the sake of being self-contained, we present a quick proof.

Lemma 2.6.1. *Consider a time-varying game $\mathcal{G}^t$, $t = 1, 2, \ldots$. For any $T \geq 1$ and constant step size γ_i, the regret of Algorithm 2 relative to the fixed action $x_i \in \mathcal{X}_i$ is*

$$\sum_{t=1}^{T}[u_i^t(x_i; X_{-i}^t) - u_i^t(X^t)] \leq \frac{1}{\gamma_i}\mathcal{D}[\mathcal{X}_i, h_i] + \sum_{t=1}^{T}\langle U_i^t, X_i^t - x_i\rangle + \sum_{t=1}^{T}\frac{\gamma_i}{2K_i}\|Y_i^t\|_*^2, \tag{2.55}$$

where

$$\mathcal{D}[\mathcal{X}_i, h_i] := \max_{\mathcal{X}_i} h_i - \min_{\mathcal{X}_i} h_i. \tag{2.56}$$

Proof. The basic starting point of our analysis is the following inequality, taken from inequality 2.135 in Appendix 2.9.1 :

$$D_{h_i}(x_i, X_i^{t+1}) - D_{h_i}(x_i, X_i^t) \leq \langle -\gamma_i^t Y_i^t, x_i - X_i^t\rangle + \frac{(\gamma_i^t)^2}{2K_i}\|Y_i^t\|_*^2. \tag{2.57}$$

Using the decomposition of the signal at stage t and rearranging the above expression yields

$$\langle \gamma_i^t v_i^t(X^t), x_i - X_i^t\rangle \leq D_{h_i}(x_i, X_i^t) - D_{h_i}(x_i, X_i^{t+1})$$

$$- \langle \gamma_i^t U_i^t, x_i - X_i^t \rangle + \frac{(\gamma_i^t)^2}{2K_i} \|Y_i^t\|_*^2. \tag{2.58}$$

Taking a constant step-size $\gamma_i^t = \gamma_i$ and telescoping gives

$$\begin{aligned} \sum_{t=1}^{T} \langle v_i^t(X^t), x_i - X_i^t \rangle &\leq \frac{1}{\gamma_i} \Big(D_{h_i}(x_i, X_i^1) - D_{h_i}(x_i, X_i^{T+1}) \Big) \\ &\quad - \sum_{t=1}^{T} \langle U_i^t, x_i - X_i^t \rangle + \sum_{t=1}^{T} \frac{\gamma_i}{2K_i} \|Y_i^t\|_*^2. \end{aligned} \tag{2.59}$$

Since $X_i^1 \in \operatorname{argmin}_{x_i \in \mathcal{X}_i} h_i(x_i)$ and the Bregman divergence is non-negative, we arrive at

$$\begin{aligned} \sum_{t=1}^{T} \langle v_i^t(X^t), x_i - X_i^t \rangle &\leq \frac{1}{\gamma_i} D_{h_i}(x_i, X_i^1) + \sum_{t=1}^{T} \langle U_i^t, X_i^t - x_i \rangle + \sum_{t=1}^{T} \frac{\gamma_i}{2K_i} \|Y_i^t\|_*^2 \\ &\leq \frac{1}{\gamma_i} \mathcal{D}[\mathcal{X}_i, h_i] + \sum_{t=1}^{T} \langle U_i^t, X_i^t - x_i \rangle + \sum_{t=1}^{T} \frac{\gamma_i}{2K_i} \|Y_i^t\|_*^2 \end{aligned} \tag{2.60}$$

where the last line uses equality (2.127). □

This step of the proof is very much inspired from recent results on Mirror-Prox algorithms [86, 57]. As a result of this, we will easily obtain a bound on the static regret *in expectation*. In the next step we will use this bound, and some basic facts from martingale theory to prove a sequence of concentration result on the gap function, and a-fortiori, the static regret.

2.6.1 Bounding expected cumulative regret

In order to bound the static regret of the prox-strategy, we recall the basic assumptions we imposed on the observational noise of the players'

feedback signal. From (2.17b), we get the existence of a positive scalar σ_i such that

$$\mathbb{E}[\|U_i^t\|_*^2|\mathcal{F}^{t-1}] \leq \sigma_i^2, \tag{2.61}$$

holds for all $i \in \mathcal{N}$. Given the signal structure $Y_i^t = v_i^t(X^t) + U_i^t$, condition (2.61) and the gradient bound (2.10) implies that

$$\mathbb{E}[\|Y_i^t\|_*^2|\mathcal{F}^{t-1}] \leq 2G_i^2 + 2\sigma_i^2 =: M_i^2 \qquad \forall t \geq 1, i \in \mathcal{N}. \tag{2.62}$$

Under these assumptions we prove a sublinear bound on the *expected regret*.

Proposition 2.6.2. *For fixed $T \geq 1$, assume that player i executes Algorithm 2 , with constant step-size*

$$\gamma_i = 2\sqrt{\frac{\mathcal{D}[\mathcal{X}_i; h_i]K_i}{T(M_i^2 + \sigma_i^2)}}. \tag{2.63}$$

Then, we have the following regret bound over the time window $\mathcal{T} = \{1, \ldots, T\}$:

$$\mathbb{E}[\mathrm{Reg}_i(\mathcal{T})] \leq 2\sqrt{T(M_i^2 + \sigma_i^2)\mathcal{D}[\mathcal{X}_i; h_i]/K_i}. \tag{2.64}$$

In particular, $\limsup_{T\to\infty} \mathbb{E}[\mathrm{Reg}_i(\mathcal{T})/T] = 0$.

Before proving this result, let us emphasize that Proposition 2.6.2 provides an upper bound on the expected regret, and not on the *pseudo-regret*, as commonly done in the literature (see e.g. [12]). Pseudo-regret is a much weaker notion of regret, and usually it is much easier to obtain bounds on the pseudo-regret.

Proof. By (2.58), we have

$$\langle \gamma_i^t v_i^t(X^t), x_i - X_i^t \rangle \leq D_{h_i}(x_i, X_i^t) - D_{h_i}(x_i, X_i^{t+1}) - \langle \gamma_i^t U_i^t, x_i - X_i^t \rangle$$

$$+ \frac{(\gamma_i^t)^2}{2K_i}\|Y_i^t\|_*^2. \tag{2.65}$$

For each player i define the auxiliary process $\{Z_i^t\}_{t\in\mathbb{N}}$ by

$$Z_i^1 = X_i^1 \text{ and } Z_i^{t+1} = P_i(Z_i^t, \gamma_i^t U_i^t). \tag{2.66}$$

A simple induction argument shows that the process $\{Z_i^t\}_{t\geq 1}$ is $\{\mathcal{F}_i^t\}_{t\geq 1}$ measurable, for all $i \in \mathcal{N}$. Using this process, the previous inequality can be rewritten as

$$\begin{aligned}\langle \gamma_i^t v_i^t(X^t), x_i - X_i^t\rangle \leq D_{h_i}(x_i, X_i^t) - D_{h_i}(x_i, X_i^{t+1}) \\ - \langle \gamma_i^t U_i^t, Z_i^t - X_i^t\rangle + \frac{(\gamma_i^t)^2}{2K_i}\|Y_i^t\|_*^2 + \langle \gamma_i^t U_i^t, Z_i^t - x_i\rangle.\end{aligned} \tag{2.67}$$

Hence, after summing and telescoping, we arrive at the bound

$$\begin{aligned}\sum_{t=1}^T \langle \gamma_i^t v_i^t(X^t), x_i - X_i^t\rangle \leq D_{h_i}(x_i, X^1) - D_{h_i}(x_i, X_i^{T+1}) + \sum_{t=1}^T \langle \gamma_i^t U_i^t, X_i^t - Z_i^t\rangle \\ + \sum_{t=1}^T \frac{(\gamma_i^t)^2}{2K_i}\|Y_i^t\|_*^2 + \sum_{t=1}^T \langle \gamma_i^t U_i^t, Z_i^t - x_i\rangle.\end{aligned} \tag{2.68}$$

By Lemma 2.9.2 , proven in Appendix 2.9.1 , shows that

$$\sum_{t=1}^T \langle \gamma_i^t U_i^t, Z_i^t - x_i\rangle \leq D_{h_i}(x_i, X_i^1) + \frac{1}{2K_i}\sum_{t=1}^T \|\gamma_i^t U_i^t\|_*^2. \tag{2.69}$$

Combining these two inequalities gives

$$\sum_{t=1}^T \langle \gamma_i^t v_i^t(X^t), x_i - X_i^t\rangle \leq 2D_{h_i}(x_i, X_i^1) + \sum_{t=1}^T \langle \gamma_i^t U_i^t, X_i^t - Z_i^t\rangle$$

$$+\sum_{t=1}^{T}\frac{(\gamma_i^t)^2}{2K_i}\big(\|Y_i^t\|_*^2+\|U_i^t\|_*^2\big). \tag{2.70}$$

In the case of a constant step-size $\gamma_i^t=\gamma_i$, this implies

$$\sum_{t=1}^{T}\langle v_i^t(X^t),x_i-X_i^t\rangle \le \frac{2D_{h_i}(x_i,X_i^1)}{\gamma_i}+\sum_{t=1}^{T}\langle U_i^t,X_i^t-Z_i^t\rangle + \sum_{t=1}^{T}\frac{\gamma_i}{2K_i}\big(\|Y_i^t\|_*^2+\|U_i^t\|_*^2\big). \tag{2.71}$$

Since $X_i^1\in\operatorname{argmin}_{x_i\in\mathcal{X}_i}h_i(x_i)$, taking the supremum over actions $x_i\in\mathcal{X}_i$ on both sides of this inequality, and using (3.13), we conclude

$$\operatorname{Reg}_i(\mathcal{T}) \le \operatorname{Gap}_i(\mathcal{T}) \tag{2.72}$$

$$\le \frac{2}{\gamma_i}\mathcal{D}[\mathcal{X}_i,h_i]+\sum_{t=1}^{T}\langle U_i^t,X_i^t-Z_i^t\rangle+\sum_{t=1}^{T}\frac{\gamma_i}{2K_i}\big(\|Y_i^t\|_*^2+\|U_i^t\|_*^2\big). \tag{2.73}$$

The process $\sum_{t=1}^{T}\langle U_i^t,X_i^t-Z_i^t\rangle$ is a martingale with respect to the filtration $\mathbb{F}:=\{\mathcal{F}_t\}_{t\ge1}$, which is also bounded in $L^2(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$, thanks to (2.87). The process $\sum_{t=1}^{T}\big(\|Y_i^t\|_*^2+\|U_i^t\|_*^2\big)$ is a non-negative submartingale, with expected value bounded by $T(M_i^2+\sigma_i^2)$. Hence, taking expectations on both sides, we obtain

$$\mathbb{E}[\operatorname{Reg}_i(\mathcal{T})]\le\mathbb{E}[\operatorname{Gap}_i(\mathcal{T})]\le\frac{2}{\gamma_i}\mathcal{D}[\mathcal{X}_i,h_i]+\frac{T\gamma_i}{2K_i}(M_i^2+\sigma_i^2). \tag{2.74}$$

Optimizing with respect to γ_i yields $\gamma_i=2\sqrt{\frac{\mathcal{D}[\mathcal{X}_i,h_i]K_i}{T(M_i^2+\sigma_i^2)}}$. Using this step-size in the previous display, we get

$$\mathbb{E}[\operatorname{Reg}_i(\mathcal{T})]\le\mathbb{E}[\operatorname{Gap}_i(\mathcal{T})]\le 2\sqrt{\mathcal{D}[\mathcal{X}_i,h_i]T(M_i^2+\sigma_i^2)/K_i} \tag{2.75}$$

and our proof is complete. □

It is of course much more interesting to know how large the realized regret of a player actually is. Based on the bound on the expected regret derived in the proposition, a simple probabilistic bound on the magnitude of the realized regret can be obtained via Markov's inequality. Indeed, for all $\varepsilon > 0$, we have

$$\mathbb{P}(\mathrm{Reg}_i(\mathcal{T}) \geq \varepsilon T) \leq \frac{1}{T\varepsilon}\mathbb{E}[\mathrm{Reg}_i(\mathcal{T})] \leq \frac{2}{\varepsilon}\sqrt{\frac{\mathcal{D}[\mathcal{X}_i, h_i](M_i^2 + \sigma_i^2)}{K_i T}}. \quad (2.76)$$

In Section 2.6.3 , we obtain a more refined bound on the regret under slightly stronger assumptions on the distribution of the feedback signal.

We close this section by remarking that in case where the feedback signal is deterministic, the above bound on the expected regret immediately delivers an $O(\sqrt{T})$-bound for the realized regret.

Corollary 2.6.3. *Consider the time-varying game with perfect gradient feedback signals, $\sigma_i = 0$ for all $i \in \mathcal{N}$. Then*

$$\mathrm{Reg}_i(\mathcal{T}) \leq 2\sqrt{\mathcal{D}[\mathcal{X}_i, h_i] T M_i^2 / K_i}. \quad (2.77)$$

In particular, $\limsup_{T\to\infty} \mathrm{Reg}_i(\mathcal{T})/T = 0$ *for all* $i \in \mathcal{N}$.

2.6.2 Bounding expected dynamic regret

As an immediate consequence of Proposition 2.6.2, we are able to bound the players' expected *dynamic* regret via the decomposition Lemma 2.4.2. Recall that the key ingredients for this bound were (i) a restarting procedure (Algorithm 1), and (ii) an algorithm guaranteeing sublinear asymptotic behavior of the gap function. In this section we carry out this program, using Algorithm 2 as the driving algorithm in players' updating decisions.

Let $1 \leq \Delta_i \leq T$ be a given batch size for blocks $\mathcal{T}_{i,1}, \ldots, \mathcal{T}_{i,m_i}$ with $m_i = \lceil T/\Delta_i \rceil$ in order to realize Algorithm 1 , with Algorithm 2 to update the action within each block. Denote the resulting repeated game strategy by $\{X^t\}_{1 \leq t \leq T}$. Under this strategy, we can use Theorem 2.4.3 to obtain an immediate bound on the expected dynamic regret.

Proposition 2.6.4. *Let $\mathcal{T} = \{1, \ldots, T\}$, and let $x_i^t \in \mathcal{X}_i$, $t = 1, 2, \ldots$ be a test sequence enjoying the variation bound*

$$\mathrm{Var}_i(\mathcal{T}; x_i^{\mathcal{T}}) \leq V_i^T,$$

for some $V_i^T \geq 1$. Let $\{X_i^t\}_{1 \leq t \leq T}$ be the strategy defined by Algorithm 1 with batch size $\Delta_i = \lceil (T/V_i^T)^{2/3} \rceil$, and subroutine Algorithm 2 to update the actions within each block. Then we have

$$\mathbb{E}[\mathrm{DynReg}_i(\mathcal{T}, \{x_i^t\}_{k \in \mathbb{N}})] \leq (3C_i + 2G_i)(V_i^T)^{1/3} T^{2/3}, \tag{2.78}$$

where $C_i = 2\sqrt{\mathcal{D}[\mathcal{X}_i, h_i](M_i^2 + \sigma_i^2)K_i}$.

Proof. Partition the time window $\mathcal{T}$ into m_i blocks $\mathcal{T}_{i,1}, \ldots, \mathcal{T}_{i,m_i}$ of size at most Δ_i. Lemma 2.4.2 tells us that we can bound the dynamic regret suffered by player i against an arbitrary sequence of test actions $\{x_i^t\}_{t \geq 1}$ as

$$\mathrm{DynReg}_i(\mathcal{T}; \{x_i^t\}_{t \geq 1}) \leq \sum_{\ell=1}^{m_i} \mathrm{Gap}_i(\mathcal{T}_{i,\ell}) + G_i \Delta_i \, \mathrm{Var}(\mathcal{T}; \{x_i^t\}_{t \geq 1}). \tag{2.79}$$

On each time window $\mathcal{T}_{i,\ell}$ assume that player i uses the constant step size

$$\gamma_i^{(\ell)} = 2\sqrt{\frac{K_i \mathcal{D}[\mathcal{X}_i, h_i]}{\Delta_i (M_i^2 + \sigma_i^2)}} \qquad 1 \leq \ell < m_i. \tag{2.80}$$

Then, using equality (2.75) , we can bound the gap function of player i

on the time window $\mathcal{T}_{i,\ell}$ as

$$\mathbb{E}[\mathrm{Gap}_i(\mathcal{T}_{i,\ell})] \leq 2\sqrt{\mathcal{D}[\mathcal{X}_i, h_i]\Delta_i(M_i^2 + \sigma_i^2)/K_i}. \tag{2.81}$$

Let $\{x_i^t\}_{k\in\mathbb{N}}$ be a test sequence satisfying $\mathrm{Var}(\mathcal{T}; \{x_i^t\}_{t\geq 1}) \leq V_i^T$ for all $T \geq 1$. Adding the bounds of the gap function on each block, we can bound the expected dynamic regret under the prox-strategy with restart as

$$\mathbb{E}[\mathrm{DynReg}_i(\mathcal{T}; \{x_i^t\}_{t\geq 1})] \leq 2m_i\sqrt{\mathcal{D}[\mathcal{X}_i, h_i]\Delta_i(M_i^2 + \sigma_i^2)/K_i} + G_i\Delta_i V_i^T. \tag{2.82}$$

From here we continue as in the proof of Theorem 2.4.3. □

2.6.3 High-Probability bounds

The previous results all gave us bounds on the expected regret. Of course, in real-world applications of online learning, a player needs to employ a strategy which guarantees not only a low (dynamic) regret on average, but actually *with high probability*. To that end, in this section we derive some concentration inequalities of the (static) regret suffered by the prox-method. For the derivation of high-probability regret bounds, we need to make more restrictive assumptions on the players' observational noise. Specifically, we assume that there exists a constant $M_*^2 > 0$ such that

$$\mathbb{E}\left[\exp\left(\frac{\|Y_i^t\|_*^2}{M_*^2}\right)\right] \leq \exp(1). \tag{2.83}$$

A simple application of Jensen's inequality shows that this implies that the signal process $\{Y_i^t\}_{t\geq 1}$ is bounded in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, as in equality 2.62. Clearly, this assumption on the feedback imposes similar structure on the observational noise process. Indeed, by definition, we have

$$\|U_i^t\|_*^2 = \|Y_i^t - v_i^t(X^t)\|_*^2 \leq 2\|Y_i^t\|_*^2 + 2\|v_i^t(X^t)\|_*^2. \tag{2.84}$$

Since

$$\|v_i^t(X^t)\|_* = \|\mathbb{E}[Y_i^t \,|\, \hat{\mathcal{F}}_t]\|_* \leq \sqrt{\mathbb{E}[\|Y_i^t\|_*^2 \,|\, \hat{\mathcal{F}}_t]} \leq M_*, \tag{2.85}$$

we conclude that

$$\|U_i^t\|_*^2 \leq 2\|Y_i^t\|_*^2 + 2M_*^2, \tag{2.86}$$

and

$$\mathbb{E}\left[\exp\left(\frac{\|U_i^t\|_*^2}{4M_*^2}\right)\right] \leq \exp(1). \tag{2.87}$$

This sub-Gaussian assumption on the observation noise is quite common in stochastic optimization and statistics [86, 57], and therefore provides a reasonable assumption on the driving stochastic forces. It holds under isotropic Gaussian observation noise, but also for every bounded noise distribution. It is a stronger assumption than (2.16b), since it implies that the observational noise must have finite moments of all orders greater or equal than 2.

We will use these exponential bounds to derive a concentration inequality for the static regret, formulated in the following Proposition, whose proof is relegated to Appendix 2.9.2.

Proposition 2.6.5. *For all $T \geq 1$ and all $\varepsilon \in (0,1)$, with probability at least $1-\varepsilon$, there exists a step-size policy $\{\gamma_i^t\}_{1\leq k\leq T}$ such that*

$$\mathrm{Reg}_i(\mathcal{T}) \leq 2\sqrt{2T\mathcal{D}[\mathcal{X}_i; h_i]\Omega_i^\varepsilon} + 8M_*\sqrt{2T\log(2/\varepsilon)\mathcal{D}[\mathcal{X}_i, h_i]/K_i}, \tag{2.88}$$

where $\Omega_i^\varepsilon = 5(1+\log(2/\varepsilon))M_*^2/(2K_i)$.

2.7 Regret minimization and Nash equilibrium

In this section, we examine the equilibrium convergence properties of the players' long-run behavior in two distinct regimes: *a*) when the sequence of stage games $\mathcal{G}^t$ encountered by the players evolves over

time without converging; and *b*) when $\mathcal{G}^t$ converges to some limit game $\mathcal{G}$. In what follows, we will treat the process defining the time-varying game as a "black box" and we will not scrutinize its origins in detail; we do so in order to focus directly on the interplay between the fluctuations of the stage game and the induced sequence of play.

2.7.1 Tracking Nash equilibria

We begin by considering the case where $\mathcal{G}^t$ evolves without converging. Building on the discussion in 3.2 , we will assume in what follows that each $\mathcal{G}^t$ is β-strongly monotone in the sense of (2.8), i.e.,

$$\langle v^t(x') - v^t(x), x' - x \rangle \leq -\beta \|x' - x\|^2 \tag{2.89}$$

for all $t = 1, 2, \ldots$, and all $x, x' \in \mathcal{X}$. In particular, this implies that each stage game $\mathcal{G}^t$ admits a unique Nash equilibrium, which we will denote by $\hat{x}^t$. Then, to quantify the degree to which the players' chosen actions $X^t \in \mathcal{X}$ "track" the Nash equilibrium sequence $\hat{x}^t$ over the window of play $\mathcal{T} \subseteq \mathbb{N}$, we will use the *error function*

$$\mathrm{err}_i(\mathcal{T}) = \sum_{t \in \mathcal{T}} \|X_i^t - \hat{x}_i^t\|^2, \tag{2.90}$$

or, aggregating over all players $i \in \mathcal{N}$,

$$\mathrm{err}(\mathcal{T}) = \sum_{i \in \mathcal{N}} \mathrm{err}_i(\mathcal{T}) = \sum_{t \in \mathcal{T}} \|X^t - \hat{x}^t\|^2. \tag{2.91}$$

By construction, if this error function grows sublinearly with the size $T = |\mathcal{T}|$ of the window of play, the sequence X^t will be close to Nash equilibrium for most of the time (as determined by the asymptotic growth of $\mathrm{err}(\mathcal{T})$ over time). That being said, it should be intuitively clear that if the sequence of Nash equilibria varies arbitrarily from one

stage to the next, then there is no way of achieving $\mathrm{err}(\mathcal{T}) = o(T)$.[11] For this reason, we will focus in what follows on time-varying games with *slowly-varying equilibria*, i.e., such that

$$\sum_{t=1}^{T-1} \|\hat{x}^{t+1} - \hat{x}^t\| = o(T) \quad \text{as } T \to \infty. \tag{2.92}$$

Under this assumption we have:

Theorem 2.7.1. *Let $\mathcal{G}^t \equiv \mathcal{G}^t(\mathcal{N}, \mathcal{X}, u^t)$ be a sequence of strongly monotone games. Assume further that the variation of each player's equilibrium component over the window of play $\mathcal{T} = \{1, \dots, T\}$ satisfies $\sum_{t=1}^{T-1} \|\hat{x}_i^{t+1} - \hat{x}_i^t\| \leq V_i^T$ for some $V_i^T > 0$. Then, if players follow Algorithm 2 for batches of size $\Delta_i = \lceil (T/V_i^T)^{2/3} \rceil$ (as per Algorithm 1) with step-size defined by equation (2.80), we have*

$$\mathbb{E}[\mathrm{err}_i(\mathcal{T})] = \mathcal{O}(T^{2/3}(V_i^T)^{1/3}) \tag{2.93}$$

for all $i \in \mathcal{N}$. In particular, if the sequence of stage equilibria is slowly-varying in the sense of equation (2.92)*, we have $\mathbb{E}[\mathrm{err}(\mathcal{T})] = o(T)$ as $T \to \infty$.*

Proof. Our proof strategy will be to leverage the dynamic regret minimization properties of the restart schedule of Algorithm 1 (cf. Theorem 2.4.3). To that end, note first that, for every reference action $p_i \in \mathcal{X}_i$, (2.8) and (2.89) yield

$$\begin{aligned} \beta\|X_i^t - \hat{x}_i^t\|^2 &\leq \langle v_i^t(X^t), \hat{x}_i^t - X_i^t \rangle \\ &\leq \langle v_i^t(X^t), p_i - X_i^t \rangle + \langle v_i^t(X^t), \hat{x}_i^t - p_i \rangle. \end{aligned} \tag{2.94}$$

Now, letting $\mathcal{T}_{i,\ell}$ be a batch of size at most $\Delta_i = \lceil (T/V_i^T)^{2/3} \rceil$ (as per the restart procedure of Algorithm 1), we obtain the local error bound

$$\sum_{t \in \mathcal{T}_{i,\ell}} \beta\|X_i^t - \hat{x}_i^t\|^2 \leq \sum_{t \in \mathcal{T}_{i,\ell}} \langle v_i^t(X^t), p_i - X_i^t \rangle + \sum_{t \in \mathcal{T}_{i,\ell}} \langle v_i^t(X^t), \hat{x}_i^t - p_i \rangle$$

[11] For a rigorous statement and proof in the single-player setting, see the recent paper [7].

$$\leq \mathrm{Gap}_i(\mathcal{T}_{i,\ell}) + \sum_{t\in\mathcal{T}_{i,\ell}} \langle v_i^t(X^t), \hat{x}_i^t - p_i\rangle. \tag{2.95}$$

Hence, writing $\tau_{i,\ell} = \min \mathcal{T}_{i,\ell}$ for the first index of batch $\mathcal{T}_{i,\ell}$ and taking $p_i = \hat{x}_i^{\tau_{i,\ell}}$ as a reference action for the ℓ-th batch, we can bound the second term above as

$$\begin{aligned}\sum_{t\in\mathcal{T}_{i,\ell}} \langle v_i^t(X^t), \hat{x}_i^t - p_i\rangle &\leq G_i|\mathcal{T}_{i,\ell}| \max_{t\in\mathcal{T}_{i,\ell}} \|\hat{x}_i^t - \hat{x}_i^{\tau_{i,\ell}}\| \\ &\leq G_i\Delta_i \,\mathrm{Var}(\mathcal{T}_{i,\ell}, \hat{x}_i^{\tau_{i,\ell}}).\end{aligned} \tag{2.96}$$

as $|\mathcal{T}_{i,\ell}| \leq V_T$ and $\max_{t\in\mathcal{T}_{i,\ell}} \|\hat{x}_i^t - \hat{x}_i^{\mathcal{T}_{i,\ell}}\| \leq \mathrm{Var}(\mathcal{T}_{i,\ell}, \hat{x}_i^{\tau_{i,\ell}})$ by applying several times the triangular inequality over the batch $\mathcal{T}_{i,\ell}$. Then, taking expectations and summing over all batches as in the proof of Lemma 2.4.2 , we get

$$\mathbb{E}\left[\sum_{t=1}^{T} \|X_i^t - \hat{x}_i^t\|^2\right] \leq \frac{1}{\beta}\sum_{\ell=1}^{m_i} \mathbb{E}[\mathrm{Gap}_i(\mathcal{T}_{i,\ell})] + \frac{G_i}{\beta}\Delta_i V_i^T, \tag{2.97}$$

By Proposition 2.6.2 , we have $\mathrm{Gap}_i(\mathcal{T}_{i,\ell}) = \mathcal{O}(\Delta_i^{1/2})$. Thus, with $\Delta_i = \mathcal{O}((T/V_i^T)^{2/3})$ and $m_i = \mathcal{O}(T/\Delta_i) = \mathcal{O}(T^{1/3}(V_i^T)^{2/3})$, we finally get

$$\begin{aligned}\mathbb{E}[\mathrm{err}_i(\mathcal{T})] &= \mathcal{O}(m_i\Delta_i) + \mathcal{O}(\Delta_i V_i^T) \\ &= \mathcal{O}(T^{1/3}(V_i^T)^{2/3} \cdot T^{1/3}(V_i^T)^{-1/3}) + \mathcal{O}((T/V_i^T)^{2/3} \cdot V_i^T) \\ &= \mathcal{O}(T^{2/3}(V_i^T)^{1/3}),\end{aligned} \tag{2.98}$$

as claimed. Our second assertion then follows by noting that $T^{2/3}(V_i^T)^{1/3} = o(T)$ if $V_i^T = o(T)$. □

Note that the strategy used to bound the tracking error depends on the variation of the sequence of Nash equilibria of each stage game $\mathcal{G}^t$. We emphasize that this does not mean that the players actually *know* this precise variation: it suffices to have a bound thereof (even pessimistic). For instance, such information could be available to a

player who knows that the sequence of stage games encountered comes from a family of games that follow some sufficiently slow variation, but not the exact realization of the game (and, much less, their equilibria). It thus follows to reason that a sharper bound of this form leads to a better tracking error (as evidenced by the $(V_i^T)^{1/3}$ dependence of $\mathrm{err}_i(T)$ on the variation bound V_i^T).

2.7.2 Convergence to Nash equilibrium

We now turn to the case where the sequence of stage games $\mathcal{G}^t \equiv \mathcal{G}^t(\mathcal{N}, \mathcal{X}, u^t)$ converges to some (monotone) limit game $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$.[12] Formally, it will be convenient to characterize this convergence in terms of the quantity

$$B_i^t = \max_{x \in \mathcal{X}} \|v_i^t(x) - v_i(x)\|_*, \tag{2.99}$$

i.e., via the maximum difference in (individual) payoff gradients between stage t and the limit $t \to \infty$. We will then say that the sequence of games $\mathcal{G}^t$ *converges effectively* to $\mathcal{G}$ if

$$B^t \equiv \sum_{i \in \mathcal{N}} B_i^t \to 0 \quad \text{as } t \to 0. \tag{2.100}$$

The reason for defining the convergence of a sequence of games in terms of payoff gradients instead of payoff functions is twofold: First, if the payoff functions of a game are perturbed by arbitrary (player-specific) constants, the game's equilibrium points will remain unchanged, but the corresponding payoff differences may be large (so $\|u_i^t - u_i\|$ may fail to converge to 0 as $t \to \infty$). Second, the variational characterization (2.4) shows that the Nash equilibria of a (concave) game are actually determined by the players' individual payoff gradients, not their payoff functions; as such, characterizing the convergence of a sequence of

[12] To be clear, we are not assuming that each stage game $\mathcal{G}^t$ is a priori monotone.

stage games in terms of payoff gradients is closer to the true primitives that define equilibrium behavior in our setting.

Now, as in the previous section, we will focus on the prox-based learning protocol outlined in Algorithm 2. However, since we are now interested in the convergence of the generated sequence of play X^t to a specific target point in $\mathcal{X}$, we will require in what follows that the Bregman divergence of $h = \sum_i h_i$ satisfy the additional "reciprocity" condition

$$x^t \to p \quad \text{whenever} \quad D_h(p, x^t) \to 0, \tag{RC}$$

for every sequence of actions $x^t \in \mathcal{X}^\circ \equiv \prod_i \mathcal{X}_i^\circ$. This requirement is known in the literature as "Bregman reciprocity" [16, 76] and, essentially, it ensures that the sublevel sets of $D_h(p, \cdot)$ constitute a neighborhood basis for p in $\mathcal{X}$, i.e., every Bregman zone of the form $\mathbb{D}_\varepsilon(p) \equiv \{x \in \mathcal{X} : D_h(p, x) \leq \varepsilon\}$ contains some δ-ball $\mathbb{B}_\delta(p) = \{x \in \mathcal{X} : \|p - x\| \leq \delta\}$.[13]

With all this at hand, our main Nash equilibrium convergence result in this setting is as follows:

Theorem 2.7.2. *Let $\mathcal{G}^t \equiv \mathcal{G}^t(\mathcal{N}, \mathcal{X}, u^t)$ be a sequence of concave games converging to a strictly monotone limit game $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ in the sense of* (2.100)*. Assume further that each player follows Algorithm 2 with a prox-mapping satisfying* (RC) *and a step-size sequence γ^t such that*

$$\sum_{t=1}^{\infty} \gamma^t = \infty, \quad \sum_{t=1}^{\infty} (\gamma^t)^2 < \infty, \quad \textit{and} \quad \sum_{t=1}^{\infty} \gamma^t B^t < \infty. \tag{2.101}$$

Then, with probability 1*, the sequence of realized actions X^t converges to the* (*necessarily unique*) *Nash equilibrium $\hat{x}$ of the limit game $\mathcal{G}$.*

Before discussing the proof of 2.7.2 , some remarks are in order: First, the requirement $\sum_{t=1}^{\infty} \gamma_i^t B_i^t < \infty$ should be interpreted as a bound on

[13] The converse to this condition (i.e., that $X^t \to p$ whenever $D_h(p, X^t) \to 0$) holds automatically as a simple consequence of the fact that $D_h(p, x) \geq (K/2)\|p - x\|^2$ (cf. 2.9.1).

how slow the convergence of $\mathcal{G}^t$ can be in order for convergence to equilibrium to be guaranteed. For instance, as long as $B^t = \mathcal{O}(1/(\log t)^\varepsilon)$ for some $\varepsilon > 0$, the step-size conditions (2.101) can all be satisfied by taking $\gamma^t \propto 1/(t \log t)$. Second, as in Theorem 2.7.1 , the players of the game are not required to know the exact value of B_i^t (which would require a very detailed knowledge of the game at hand): as in all our variation results so far, it suffices to work with an upper bound thereof (even a loose, pessimistic one).

Our proof strategy will be based on two intermediate results, both of independent interest. First, we will show that the sequence of generated actions converges (a.s.) to a level set of the Bregman divergence $D_h(\hat{x}, \cdot)$ relative to $\hat{x}$. Subsequently, we show that X^t cannot remain a bounded distance away from $\hat{x}$ for all sufficiently large t. Combining these results will then suffice to show that X^t can only converge to the zero-level set of the Bregman divergence, i.e., $\lim_{t\to\infty} X^t = \hat{x}$.

We begin by establishing the convergence of X^t to a level set of the Bregman divergence:

Proposition 2.7.3. *With assumptions as in 2.7.2 , the Bregman divergence $D_h(\hat{x}, X^t)$ converges (a.s.) to a random variable D^∞ with $\mathbb{P}(D^\infty < \infty) = 1$.*

Proof. To begin, it will be convenient to decompose the signal process Y_i^t as

$$Y_i^t = v_i^t(X^t) + U_i^t = v_i(X^t) + U_i^t + b_i^t, \tag{2.102}$$

where we have set $b_i^t = v_i^t(X^t) - v_i(X^t)$. Then, letting $D_i^t = D_{h_i}(\hat{x}_i, X_i^t)$, the descent inequality (2.58) for the Bregman divergence readily yields

$$\begin{aligned} D_i^{t+1} &\leq D_i^t + \gamma^t \langle Y_i^t, X_i^t - \hat{x}_i \rangle + \frac{(\gamma^t)^2}{2K_i} \|Y_i^t\|_*^2 \\ &\leq D_i^t + \gamma^t \langle v_i(X^t), X_i^t - \hat{x}_i \rangle + \gamma^t \langle U_i^t, X_i^t - \hat{x}_i \rangle \qquad (2.103) \\ &\quad + \gamma^t \langle b_i^t, X_i^t - \hat{x}_i \rangle + \frac{(\gamma^t)^2}{2K_i} \|Y_i^t\|_*^2 \qquad (2.104) \end{aligned}$$

$$\leq D_i^t + \gamma^t \xi_i^t + \gamma^t \beta_i^t + \frac{(\gamma^t)^2}{2K_i} \|Y_i^t\|_*^2, \tag{2.105}$$

where, in the third line, we have set $\xi_i^t = \langle U_i^t, X_i^t - \hat{x}_i \rangle$ and $\beta_i^t = \langle b_i^t, X_i^t - \hat{x}_i \rangle$, and we used the fact that $\hat{x}$ is a Nash equilibrium of the limit game $\mathcal{G}$ (implying in turn that $\langle v_i(x_i), x_i - \hat{x}_i \rangle \leq 0$ for all $x_i \in \mathcal{X}_i$ and all $i \in \mathcal{N}$). Thus, taking expectations, we obtain:

$$\begin{aligned}
\mathbb{E}[D_i^{t+1} \,|\, \mathcal{F}^{t-1}] &\leq \mathbb{E}\left[D_i^t + \gamma^t \xi_i^t + \gamma^t \beta_i^t + \frac{(\gamma^t)^2}{2K_i} \|Y_i^t\|_*^2 \,\middle|\, \mathcal{F}^{t-1} \right] \\
&\leq D_i^t + \gamma^t \, \mathbb{E}[\|b_i^t\|_* \|X_i^t - \hat{x}_i\| \,|\, \mathcal{F}^{t-1}] + \frac{(\gamma^t)^2}{2K} \, \mathbb{E}[\|Y_i^t\|_*^2 \,|\, \mathcal{F}^{t-1}] \\
&\leq D_i^t + \gamma^t B_i^t \operatorname{diam}(\mathcal{X}_i) + \frac{1}{2K} (\gamma^t)^2 M_i^2
\end{aligned} \tag{2.106}$$

where *a*) in the second line, we used the fact that X^t is predictable relative to $\mathcal{F}^t$, the definition of β_i^t, and the fact that $\mathbb{E}[U^t \,|\, \mathcal{F}^{t-1}] = 0$; and *b*) in the last line, we used (2.17b) and the definition of B_i^t.

To proceed, let $\varepsilon_i^t = \gamma^t B_i^t \operatorname{diam}(\mathcal{X}_i) + \frac{1}{2K} (\gamma^t)^2 M_i^2$, so the last line of (2.106) can be written as

$$\mathbb{E}[D_i^{t+1} \,|\, \mathcal{F}^{t-1}] \leq D_i^t + \varepsilon_i^t. \tag{2.107}$$

Consider now the auxiliary process $\zeta_i^t = D_i^{t+1} + \sum_{s=t+1}^{\infty} \varepsilon_i^s$. Then, taking expectations yields

$$\mathbb{E}[\zeta_i^t \,|\, \mathcal{F}^{t-1}] \leq D_i^t + \varepsilon_i^t + \sum_{s=t+1}^{\infty} \varepsilon_i^s = D_i^t + \sum_{s=t}^{\infty} \varepsilon_i^s = \zeta_i^{t-1}, \tag{2.108}$$

i.e., ζ_i^t is a supermartingale relative to $\mathcal{F}^t$. Furthermore, since $\sum_{t=1}^{\infty} \varepsilon_i^t < \infty$ by the step-size assumption (2.101), we also get $\mathbb{E}[\zeta_i^t] \leq \mathbb{E}[\zeta_i^1] < \infty$, i.e., ζ_i^t is bounded in L^1. Thus, by Doob's (sub)martingale convergence theorem, it follows that ζ_i^t converges almost surely to some random variable ζ_i that is itself finite (almost surely and in L^1). Since $D_i^t = \zeta_i^{t-1} - \sum_{s=t}^{\infty} \varepsilon_i^s$ and $\lim_{t\to\infty} \sum_{s=t}^{\infty} \varepsilon_i^s = 0$ (again, by the step-size summability

assumption), we conclude that D_i^t converges itself to ζ_i. Our claim then follows by noting that $D_h(\hat{x}, X^t) = \sum_i D_{h_i}(\hat{x}_i, X_i^t)$. □

Moving on, our next result shows that the sequence of play X^t gets arbitrarily close to the Nash equilibrium $\hat{x}$ of the limit game:

Proposition 2.7.4. *With probability* 1*, there exists a* (*random*) *subsequence* X^{t_k} *of* X^t *which converges to* $\hat{x}$*.*

Proof. Our proof is by contradiction. To that end, suppose that, with positive probability, the sequence of play X^t does not admit $\hat{x}$ as a limit point. Conditioning on this event, there exists a ball $\mathbb{B}_\delta(\hat{x})$ such that $X^t \notin \mathbb{B}_\delta(\hat{x})$ for all sufficiently large t, implying in turn that X^t is contained in some compact set $\mathcal{K} \subseteq \mathcal{X}$ such that $\hat{x} \notin \mathcal{K}$. By (2.4) and because the game is strictly monotone, we have $\langle v(x), x - \hat{x}\rangle < 0$ whenever $x \in \mathcal{K}$. Therefore, by the continuity of v and the compactness of $\mathcal{K}$, there exists some $c > 0$ such that

$$\langle v(x), x - \hat{x}\rangle \leq -c \quad \text{for all } x \in \mathcal{K}. \tag{2.109}$$

To proceed, let $D^t = D_h(\hat{x}, X^t)$ as in the proof of 2.7.3. Then, telescoping (2.105) yields the estimate

$$D^{t+1} \leq D^1 + \sum_{s=1}^{t} \gamma^s \langle v(X^s), X^s - \hat{x}\rangle + \sum_{s=1}^{t} \gamma^s \xi^s + \sum_{s=1}^{t} \gamma^s \beta^s + \sum_{s=1}^{t} \frac{(\gamma^s)^2}{2K} \|Y^s\|_*^2, \tag{2.110}$$

where the strong convexity modulus K is defined as $K = \min_i K_i$, and, as in the proof of Proposition 2.7.3, we set $\xi^t = \langle U^t, X^t - \hat{x}\rangle$ and $\beta^t = \langle b^t, X^t - \hat{x}\rangle$. Hence, setting $S^t = \sum_{s=1}^t \gamma^s$ and using (2.109), we

obtain for all t large enough

$$D^{t+1} \leq D^1 - S^t \left[c - \frac{\sum_{s=1}^t \gamma^s \xi^s}{S^t} - \frac{\sum_{s=1}^t \gamma^s \beta^s}{S^t} - \frac{(2K)^{-1} \sum_{s=1}^t (\gamma^s)^2 \|Y^s\|_*^2}{S^t} \right]. \tag{2.111}$$

We proceed to analyze this bound term-by-term:

- First, by definition, we have $\mathbb{E}[\xi^t \,|\, \mathcal{F}^{t-1}] = 0$, so the second term in the brackets of (2.111) is itself a martingale. Furthermore, by (2.16b), we have

$$\begin{aligned} \sum_{t=1}^\infty (\gamma^t)^2 \, \mathbb{E}[(\xi^t)^2 \,|\, \mathcal{F}^{t-1}] &\leq \sum_{t=1}^\infty (\gamma^t)^2 \|X^t - \hat{x}\|^2 \, \mathbb{E}[\|U^t\|_*^2 \,|\, \mathcal{F}^{t-1}] \\ &\leq \operatorname{diam}(\mathcal{X})^2 \sigma^2 \sum_{t=1}^\infty (\gamma^t)^2 < \infty. \end{aligned} \tag{2.112}$$

  Therefore, by the law of large numbers for martingale difference sequences [45, Theorem 2.18], we conclude that $(1/S^t) \sum_{s=1}^t \gamma^s \xi^s$ converges to 0 with probability 1.

- For the third term in the brackets of (2.111), we have $\beta^t \leq \sum_{i \in \mathcal{N}} \operatorname{diam}(\mathcal{X}_i) B_i^t$, so $\beta^t \to 0$ as $t \to \infty$. Since $\sum_{t=1}^\infty \gamma^t = \infty$, it follows that $\sum_{s=1}^t \gamma^s \beta^s / S^t \to 0$.

- Finally, for the last term, let $R^t = (1/2K) \sum_{s=1}^t (\gamma^s)^2 \|Y^s\|_*^2$. We then have

$$\begin{aligned} \mathbb{E}[R^t \,|\, \mathcal{F}^{t-1}] &= \frac{1}{2K} \, \mathbb{E}\left[\sum_{s=1}^{t-1} (\gamma^s)^2 \|Y^s\|_*^2 + \frac{(\gamma^t)^2}{2K} \|Y^t\|_*^2 \,\middle|\, \mathcal{F}^{t-1} \right] \\ &= R^t + (\gamma^t)^2 \, \mathbb{E}[\|Y^t\|_*^2 \,|\, \mathcal{F}^{t-1}] \geq R^t, \end{aligned} \tag{2.113}$$

  i.e., R^t is a submartingale relative to $\mathcal{F}^t$. Furthermore, by the law

of total expectation, we also have

$$\mathbb{E}[R^t] = \mathbb{E}[\mathbb{E}[R^t \mid \mathcal{F}^{t-1}]] \leq \frac{M^2}{2K} \sum_{s=1}^{\infty} (\gamma^s)^2 < \infty, \tag{2.114}$$

where we set $M^2 = \sum_{i \in \mathcal{N}} M_i^2$. In turn, this implies that R^t is uniformly bounded in L^1 so, by Doob's (sub)martingale convergence theorem [45, Theorem 2.5], we conclude that R^t converges to some (almost surely finite) random variable R^∞ with $\mathbb{E}[R^\infty] < \infty$. Consequently, we get $\lim_{t \to \infty} R^t / S^t = 0$ with probability 1.

Combining all of the above, we infer that there exists some (possibly random, but almost surely finite) t_0 such that $D^t \leq D^1 - c/2 \cdot S^t$ for all $t \geq t_0$. In turn, this implies that $D_h(\hat{x}, X^t) \to -\infty$ with probability 1, a contradiction. Going back to our original assumption, this shows that, with probability 1, $\hat{x}$ is a limit point of X^t, so our proof is complete. □

With these two results at hand, we are finally in a position to prove our Nash equilibrium convergence theorem:

Proof of 2.7.2 . Proposition 2.7.4 shows that, with probability 1, there exists a (possibly random) subsequence t_k such that $X^{t_k} \to \hat{x}$. By the reciprocity condition (RC), this implies that $\liminf_{t \to \infty} D_h(\hat{x}, X^t) = 0$ (a.s.). However, since $\lim_{t \to \infty} D_h(\hat{x}, X^t)$ exists with probability 1 by Proposition 2.7.3, it follows that

$$\lim_{t \to \infty} D_h(\hat{x}, X^t) = \liminf_{t \to \infty} D_h(\hat{x}, X^t) = 0 \tag{2.115}$$

i.e., X^t converges to $\hat{x}$. □

2.7.3 Convergence in two-player zero-sum games

We close this section with a convergence result for two-player zero-sum games in the spirit of time-average guarantees that are common in

the online learning literature. To state it, assume as in 2.2.2 that the sequence of stage games encountered by the players is determined by a sequence of smooth, convex-concave saddle functions $f^t\colon \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ so that $u_1^t = -f^t = -u_2^t$. We then have:

Theorem 2.7.5. *Let f^t be a sequence of convex-concave saddle functions converging to $f\colon \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ in the sense of* (2.100)*. Assume further that each player follows Algorithm 2 with a step-size sequence γ^t such that*

$$\sum_{t=1}^{\infty} \gamma^t = \infty, \quad \sum_{t=1}^{\infty} (\gamma^t)^2 < \infty, \quad \textit{and} \quad \sum_{t=1}^{\infty} \gamma^t B^t < \infty. \tag{2.116}$$

Then, with probability 1*, the sequence of ergodic averages $\bar{X}^t = \sum_{s=1}^t \gamma^s X^s / \sum_{s=1}^t \gamma^s$ converges to the set of saddle-points of f.*

Before discussing the proof of 2.7.5 , some remarks are in order:

Remark 1. It should be noted that, if the limit game is strictly monotone (for instance, if f is *strictly* convex-concave), 2.7.5 is essentially subsumed by 2.7.2 : if the sequence of play X^t converges to the game's unique equilibrium, then so does the corresponding ergodic average $\bar{X}^t = \sum_{s=1}^t \gamma^s X^s / \sum_{s=1}^t \gamma^s$. On the other hand, this leaves open the non-strict case: for instance, if the game at hand is the mixed extension of a two-player zero-sum finite game (i.e., $f(x_1, x_2) = x_1^\top A x_2$ for some matrix A of appropriate dimensions), the limit game is *not* strictly monotone, so Theorem 2.7.2 does not apply (but Theorem 2.7.5 does).

Remark 2. We should also note that 2.7.5 does not invoke the Bregman reciprocity condition (RC). The reason for this is that the analysis of the ergodic average is not as delicate as that of the actual sequence of play, but this (technical) simplification comes at a price: specifically, 2.7.5 says little for the convergence of X^t. In fact, even in the static case ($f^t = f$ for all $t = 1, 2, \dots$), the sequence of chosen actions might be recurrent or cycle around the game's limit Nash equilibrium without

converging [75, 78]: this is a qualitative difference in behavior which cannot be detected by the convergence of $\bar{X}^t$.

We now proceed with the proof of 2.7.5:

Proof of 2.7.5 . Let $\hat{x} \in \mathcal{X}$ be a Nash equilibrium of the limit game induced by f, and, as in the proof of 2.7.4 , let $D^t = D_h(\hat{x}, X^t)$. Then, working as in 2.105 , we obtain the basic estimate

$$D^{t+1} \leq D^t + \gamma^t \langle Y^t, X^t - \hat{x} \rangle + \frac{(\gamma^t)^2}{2K} \|Y^t\|_*^2, \tag{2.117}$$

where we have set $K = \min\{K_1, K_2\}$. Then, decomposing the input signal Y^t as in (2.102) and rearranging, we get:

$$\gamma^t \langle v(X^t), \hat{x} - X^t \rangle \leq D^t - D^{t+1} + \gamma^t \langle U^t + b^t, X^t - \hat{x} \rangle + \frac{(\gamma^t)^2}{2K} \|Y^t\|_*^2. \tag{2.118}$$

Then, summing over t gives

$$\sum_{s=1}^{t} \gamma^s \langle v(X^s), \hat{x} - X^s \rangle \tag{2.119}$$

$$\leq D^1 - D^{t+1} + \sum_{s=1}^{t} \gamma^s \langle U^s + b^s, X^s - \hat{x} \rangle + \sum_{s=1}^{t} \frac{(\gamma^s)^2}{2K} \|Y^s\|_*^2$$

$$\leq D^1 + \sum_{s=1}^{t} \gamma^s \xi^s + \sum_{s=1}^{t} \gamma^s \beta^s + \frac{1}{2K} R^t \tag{2.120}$$

where, as in the proof of 2.7.4, we set $\xi^t = \langle U^t, X^t - \hat{x} \rangle$, $\beta^t = \langle b^t, X^t - \hat{x} \rangle$, and $R^t = \sum_{s=1}^{t} (\gamma^s)^2 \|Y^s\|_*^2$. Hence, letting $S^t = \sum_{s=1}^{t} \gamma^s$ and arguing in the same way as in the proof of 2.7.4 (which has the same step-size requirements), we deduce that

$$\lim_{t \to \infty} \frac{\sum_{s=1}^{t} \gamma^s \xi^s}{S^t} = 0, \quad \lim_{t \to \infty} \frac{\sum_{s=1}^{t} \gamma^s \beta^s}{S^t} = 0, \quad \text{and} \quad \lim_{t \to \infty} \frac{R^t}{S^t} = 0, \tag{2.121}$$

with probability 1. On the other hand, given that f is convex-concave, we also have

$$\begin{aligned}\frac{\sum_{s=1}^{t}\gamma^s\langle v(X^s),\hat{x}-X^s\rangle}{S^t} &\geq u_1(\hat{x}_1,\hat{x}_2)-u_1(\bar{X}_1^t,\hat{x}_2)+u_2(\hat{x}_1,\hat{x}_2)-u_2(\hat{x}_1,\bar{X}_2^t)\\ &= f(\bar{X}_1^t,\hat{x}_2)-f(\hat{x}_1,\bar{X}_2^t). \qquad (2.122)\end{aligned}$$

Therefore, combining all of the above, we conclude that

$$f(\bar{X}_1^t,\hat{x}_2)-f(\hat{x}_1,\bar{X}_2^t)\leq\frac{D^1}{S^t}+o(1), \qquad (2.123)$$

i.e., $f(\bar{X}_1^t,\hat{x}_2)-f(\hat{x}_1,\bar{X}_2^t)\to 0$ as $t\to\infty$. Since $\hat{x}$ is a Nash equilibrium, this shows that $\bar{X}^t$ attains the value of f, i.e., $\bar{X}^t$ converges itself to the set of saddle-points of f, as claimed. □

2.8 Concluding remarks

There are many interesting points for future research. First, we have been very agnostic towards the data generating process of the game problem. With an eye towards simulation based solution techniques, it is important to study time-varying games generated by ergodic processes in the spirit of [20]. For monotone variational inequality problems subjected to Brownian noise, a first step in this direction has been done by [77]. More effort is needed to understand the asymptotic properties of the repeated game process in this approach, possibly also using different assumptions on the driving random process. We are currently investigating this issue.

In view of applications to problems in engineering and control, the study of a bona fide continuous-time version of the present approach is also a priority. At a higher level, we have made many strong regularity assumptions on the time-varying games in this chapter (concavity in own action, and differentiability). Relaxing the smoothness of the individual player function is an important extension of the present

approach. Finally, introducing coupled constraints into the player's action set is an important and challenging extension of the present framework, which is also currently under investigation.

2.9 Appendix

2.9.1 Prox-Mappings

In this appendix we collect some basic technical facts on the prox-method. In the following we let $\mathcal{C}$ be a convex compact domain in a finite-dimensional normed vector space $(\mathcal{V}, \|\cdot\|)$, and $h{:}\mathcal{C} \to \mathbb{R}$ be a distance generating function with convexity parameter K. The corresponding Bregman divergence is

$$D_h(x, x') = h(x) - h(x') - \langle \nabla h(x'), x - x' \rangle \tag{2.124}$$

for $x \in \mathrm{dom}(h)$, $x' \in \mathrm{dom}(h)^\circ$. Define $\Theta_h(a) = \max_{x \in \mathcal{C}} D_h(a, x)$, and

$$x^h = \mathrm{argmin}\{h(x) : x \in \mathcal{C}\}. \tag{2.125}$$

Note that x^h is uniquely defined thanks to the strong-convexity of h, and we have

$$\langle \nabla h(x^h), a - x^h \rangle \geq 0 \tag{2.126}$$

for all $a \in \mathcal{C}$. From this it follows immediately that

$$\Theta(x^h) \leq \max_{x \in \mathcal{C}} h(x) - \min_{x \in \mathcal{C}} h(x) =: \mathcal{D}[\mathcal{C}; h]. \tag{2.127}$$

Furthermore, by K-strong convexity,

$$\frac{K}{2} \|x - x^h\|^2 \leq D_h(x, x^h) \leq \Theta_h(x^h). \qquad \forall x \in \mathcal{C}. \tag{2.128}$$

Hence,

$$\|x - x^h\| \leq \sqrt{\frac{2}{K} \mathcal{D}[\mathcal{C}, h]} \qquad \forall x \in \mathcal{C}, \tag{2.129}$$

and

$$\mathcal{C} \subseteq \left\{ a \in \mathcal{V} |\ \|a - x^h\| \leq \sqrt{2\mathcal{D}[\mathcal{C}, h]/K} \right\}. \tag{2.130}$$

Lemma 2.9.1. *Let*

$$P(x, y) = \operatorname*{argmin}_{a \in \mathcal{A}} \{\langle y, a - x\rangle + D_h(a, x)\}.$$

Then

1. *For all $x \in \mathcal{A}$, the map $\mathcal{V}^* \ni y \mapsto P(x, y)$ is single-valued.*
2. *For all $x \in \mathcal{A}$ and all $y, v \in \mathcal{V}^*$, $\|P(x, v) - P(x, y)\| \leq \frac{1}{K}\|v - y\|_*$.*
3. *For all $a, x \in \mathcal{A}$ and all $y \in \mathcal{V}^*$, we have*

$$D_h(a, P(x, y)) \leq D_h(a, x) + \langle y, a - P(x, y)\rangle - D_h(P(x, y), x). \tag{2.131}$$

Proof. 1. This is clear by strong convexity.

2. Let $a = P(x, y)$ and $b = P(x, v)$. The optimality conditions at these points are

$$\langle \nabla h(a) - \nabla h(x) + y, x' - a\rangle \geq 0, \tag{2.132}$$

and

$$\langle \nabla h(b) - \nabla h(x) + v, x' - b\rangle \geq 0 \tag{2.133}$$

for all $x' \in \mathcal{X}$. Evaluating the first inequality at the point $x' = b$ and the second inequality at the point $x' = a$ gives

$$\begin{aligned} \langle \nabla h(a) - \nabla h(x) + y, b - a\rangle &\geq 0, \\ \langle \nabla h(b) - \nabla h(x) + v, b - a\rangle &\leq 0. \end{aligned}$$

Hence,

$$\langle \nabla h(a) - \nabla h(x) + y, b - a\rangle \geq \langle \nabla h(b) - \nabla h(x) + v, b - a\rangle$$

$$\iff \langle y - v, b - a \rangle \geq \langle \nabla h(b) - \nabla h(a), b - a \rangle \geq K\|a - b\|^2 \tag{2.134}$$

where the last inequality uses the K-strong convexity of the distance generating function h. Hence,

$$\|y - v\|_* \geq K\|a - b\|.$$

3. The three-point identity [16] gives

$$D_h(a, x) - D_h(a, c) - D_h(c, x) = \langle \nabla h(c) - \nabla h(x), a - c \rangle$$

Combined with the optimality condition satisfied by the point $c = P(x, y)$, we get

$$\begin{aligned} & D_h(a, x) - D_h(a, P(x, y)) - D_h(P(x, y), x) \\ = & \langle \nabla h(P(x, y)) - \nabla h(x), a - P(x, y) \rangle \\ \geq & \langle -y, a - P(x, y) \rangle. \end{aligned}$$

Rearranging gives the desired inequality.

□

2.131 provides the first step to derive regret bounds for the prox-method as done in the main text. Using the Fenchel Young inequality

$$\langle y, a - b \rangle \leq \frac{1}{2K}\|y\|_*^2 + \frac{K}{2}\|a - b\|^2$$

this inequality can be refined to

$$\begin{aligned} D_h(a, P(x, y)) - D_h(a, x) &\leq \langle y, a - x \rangle + \langle y, x - P(x, y) \rangle - D_h(P(x, y), x) \\ &\leq \langle y, a - x \rangle + \frac{1}{2K}\|y\|_*^2. \end{aligned} \tag{2.135}$$

The next Lemma provides a slight refinement of the previous one.

Lemma 2.9.2. *Let $\{v^t\}_{t\in\mathbb{N}}$ be a sequence in $\mathcal{V}^*$. Define the process $\{Y^t\}_{t\in\mathbb{N}}$ by*

$$Y^{t+1} = P(Y^t, v^t), \quad Y^1 \in \mathcal{C}^\circ \text{ given.}$$

Then, for all $T \geq 1$ and all $a \in \mathcal{C}^\circ$, we have

$$\sum_{t=1}^{T} \langle v^t, Y^t - a\rangle \leq D_h(a, Y^1) + \frac{1}{2K}\sum_{t=1}^{T} \|v^t\|_*^2.$$

Proof. From 2.9.1 and 2.135 , we get

$$D_h(a, Y^{t+1}) \leq D_h(a, Y^t) + \langle v^t, a - Y^t\rangle + \frac{\|v^t\|_*^2}{2K}.$$

Rearranging and telescoping gives

$$\begin{aligned}\sum_{t=1}^{T} \langle v^t, Y^t - a\rangle &\leq D_h(a, Y^1) - D_h(a, Y^{T+1}) + \frac{1}{2K}\sum_{t=1}^{T} \|v^t\|_*^2 \\ &\leq D_h(a, Y^1) + \frac{1}{2K}\sum_{t=1}^{T} \|v^t\|_*^2.\end{aligned}$$

The last inequality uses the non-negativity of the Bregman divergence. □

2.9.2 Proof of Lemma

The purpose of this section is to prove 2.6.5 , under the assumption that the signal process satisfies 2.83 . We need two intermediate technical results.

Lemma 2.9.3. *Define*

$$\Phi_{i,T} = \sum_{t=1}^{T} \frac{(\gamma_i^t)^2}{2K_i} \big(\|Y_i^t\|_*^2 + \|U_i^t\|_*^2\big), \tag{2.136}$$

and let $\xi_i^t = \langle \gamma_i^t U_i^t, X_i^t - Z_i^t \rangle$. *Set*

$$\Xi_{i,T} = \sum_{t=1}^{T} \xi_i^t. \tag{2.137}$$

Then, for all constants $C_1, C_2 > 0$, *we have*

$$\mathbb{P}(\Phi_{i,T} + \Xi_{i,T} \geq (1 + C_1)\Gamma_{i,T} + C_2 \Psi_{i,T}) \leq \exp(-C_1) + \exp(-C_2^2/4). \tag{2.138}$$

where $\Gamma_{i,T} := \sum_{t=1}^{T} \gamma_i^t$, *and* $\Psi_{i,T} = 4M_* \sqrt{2\mathcal{D}_i[\mathcal{X}_i; h_i]/K_i} \sqrt{\sum_{t=1}^{T} (\gamma_i^t)^2}$.

Proof. By definition we have

$$\begin{aligned} \Phi_{i,T} &= \sum_{t=1}^{T} \frac{(\gamma_i^t)^2}{2K_i} \big(\|Y_i^t\|_*^2 + \|U_i^t\|_*^2\big) \\ &\leq \sum_{t=1}^{T} \frac{(\gamma_i^t)^2}{2K_i} \big(3\|Y_i^t\|_*^2 + 2M_*^2\big) \end{aligned} \tag{2.139}$$

Calling $\gamma_i^t = 5M_*^2 (\gamma_i^t)^2/(2K_i)$, gives

$$\mathbb{E}\left[\exp\left(\frac{3(\gamma_i^t)^2 \|Y_i^t\|_*^2}{2K_i \gamma_i^t}\right)\right] \leq \exp(3/5) \tag{2.140a}$$

and

$$\mathbb{E}\left[\exp\left(\frac{(\gamma_i^t)^2 M_*^2}{K_i \gamma_i^t}\right)\right] \leq \exp(2/5). \tag{2.140b}$$

Hence,

$$\mathbb{E}\left[\exp\left(\frac{3(\gamma_i^t)^2\|Y_i^t\|_*^2 + 2(\gamma_i^t)^2 M_*^2}{2K_i\gamma_i^t}\right)\right] \leq \exp(1).$$

Call $\Gamma_{i,T} := \sum_{t=1}^T \gamma_i^t$. Then, Jensen's inequality shows that[14]

$$\mathbb{E}[\exp(\Phi_{i,T}/\Gamma_{i,T})] \leq \exp(1).$$

Therefore, for all $C_1 > 0$, Markov's inequality readily implies

$$\begin{aligned}\mathbb{P}(\Phi_{i,T} \geq (1+C_1)\Gamma_{i,T}) &= \mathbb{P}(\exp(\Phi_{i,T}/\Gamma_{i,T}) \geq \exp(1+C_1)) \\ &\leq \exp(-1-C_1)\mathbb{E}\left[\exp(\Phi_{i,T}/\Gamma_{i,T})\right] \\ &\leq \exp(-C_1). \end{aligned} \tag{2.141}$$

Now, let $\xi_i^t = \langle \gamma_i^t U_i^t, X_i^t - Z_i^t\rangle$ and set $\Xi_{i,T} = \sum_{t=1}^T \xi_i^t$. Observe that $\mathbb{E}[\xi_i^t \,|\, \mathcal{F}_{t-1}] = 0$ for all $t \geq 1$. Therefore $\Xi_{i,T}$ is a martingale with respect to the filtration $\mathbb{F} := \{\mathcal{F}_t\}_{t\geq 1}$, which is also bounded in $L^2(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, thanks to (2.87).

Via the Cauchy-Schwarz inequality, the K_i-strong convexity of the distance generating function, as well as eqs. (2.125) and (2.130), we see that

$$\begin{aligned}|\xi_i^t| &\leq \gamma_i^t \|U_i^t\|_* \cdot \|X_i^t - Y_i^t\|_i \\ &\leq \gamma_i^t \|U_i^t\|_* \left[\|X_i^t - x^{h_i}\|_i + \|x^{h_i} - Y_i^t\|_i\right]\end{aligned}$$

[14] The convexity of the mapping $x \mapsto \exp(x)$ shows the following: Let $\{a_t\}_{t\geq 1}, \{b_t\}_{t\geq 1}$ be sequences in $(0,\infty)$. Then, by Jensen's inequality

$$\sum_{t=1}^T \frac{a_t}{\sum_{\ell=1}^T a_\ell} \exp\left(\frac{b_t}{a_t}\right) \geq \exp\left(\sum_{t=1}^T \frac{b_t}{\sum_{\ell=1}^T a_\ell}\right).$$

We apply this inequality with the identification $b_t = \frac{3(\gamma_i^t)^2}{2K_i}\|Y_i^t\|_*^2 + \frac{(\gamma_i^t)^2}{K_i}M_*^2$ and $a_t = \gamma_i^t$.

$$\leq 2\gamma_i^t \|U_i^t\|_* \sqrt{\frac{2}{K_i}\mathcal{D}[\mathcal{X}_i, h_i]}$$

Hence,

$$\|U_i^t\|_*^2 \geq \frac{K_i |\xi_i^t|^2}{8(\gamma_i^t)^2 \mathcal{D}[\mathcal{X}_i, h_i]}.$$

Consequently,

$$\mathbb{E}\left[\exp\left(\frac{K_i|\xi_i^t|^2}{32(\gamma_i^t)^2\mathcal{D}[\mathcal{X}_i,h_i]M_*^2}\right)\,\middle|\,\mathcal{F}_t\right] \leq \mathbb{E}\left[\exp\left(\frac{\|U_i^t\|_*^2}{4M_*^2}\right)\right].$$

For all $t \geq 1$, denote by $\tau_i^t := 4\gamma_i^t M_* \sqrt{2\mathcal{D}[\mathcal{X}_i, h_i]/K_i}$. Using (2.87), this shows that

$$\mathbb{E}\left[\exp\big((\xi_i^t/\tau_i^t)^2\big)\,\middle|\,\hat{\mathcal{F}}_t\right] \leq \exp(1).$$

Since $\Xi_{i,t} = \Xi_{i,t-1} + \xi_i^t$, we get for all $\delta > 0$

$$\mathbb{E}[\exp(\delta\Xi_{i,t})] = \mathbb{E}[\exp(\delta\Xi_{i,t-1})\exp(\delta\xi_i^t)] = \mathbb{E}[\exp(\delta\Xi_{i,t-1})\mathbb{E}[\exp(\delta\xi_i^t)\,|\,\hat{\mathcal{F}}_t]]$$

Following [86], we see that for all $\delta > 0$ and $t \geq 1$

$$\mathbb{E}[\exp(\delta\xi_i^t)\,|\,\hat{\mathcal{F}}_t] \leq \exp\left(\delta^2(\tau_i^t)^2\right).$$

Proceeding by induction, we observe that for all $\delta > 0$,

$$\mathbb{E}[\exp(\delta\Xi_{i,T})] \leq \exp\left(\delta^2 \sum_{t=1}^T (\tau_i^t)^2\right).$$

Therefore, if we set $\Psi_{i,T} = \sqrt{\sum_{t=1}^T (\tau_i^t)^2}$, Markov's inequality yields the immediate bound

$$\begin{aligned}\mathbb{P}(\Xi_{i,T} \geq C_2\Psi_{i,T}) &\leq \exp\left(-\delta C_2 \Psi_{i,T}\right)\mathbb{E}[\exp(\delta\Xi_{i,T})] \\ &\leq \exp\left(-\delta C_2\Psi_{i,T} + \delta^2\Psi_{i,T}^2\right).\end{aligned}$$

Then, setting $\delta = \frac{C_2}{2\Psi_{i,T}}$ yields

$$\mathbb{P}(\Xi_{i,T} \geq C_2\Psi_{i,T}) \leq \exp(-C_2^2/2 + C_2^2/4) = \exp(-C_2^2/4)$$

for all $C_2 > 0$. Combining this with the bound (2.141), we conclude that for all $C_1, C_2 > 0$,

$$\mathbb{P}(\Phi_{i,T} + \Xi_{i,T} \geq (1 + C_1)\Gamma_{i,T} + C_2\Psi_{i,T}) \leq \exp(-C_1) + \exp(-C_2^2/4).$$

To see this, introduce the events $E_3 = \{\Xi_{i,T} + \Phi_{i,T} \geq (1 + C_1)\Gamma_{i,T} + C_2\Psi_{i,T}\}$, $E_1 = \{\Phi_{i,T} \geq (1 + C_1)\Gamma_{i,T}\}$ and $E_2 = \{\Xi_{i,T} \geq C_2\Psi_{i,T}\}$. Then $E_3 \subseteq E_1 \cup E_2$, so that $\mathbb{P}(E_3) \leq \mathbb{P}(E_1 \cup E_2) \leq \mathbb{P}(E_1) + \mathbb{P}(E_2)$. □

Lemma 2.9.4. *For $C > 0$, define*

$$\begin{aligned}\mathcal{Q}_{i,T}(C) &:= 2\mathcal{D}[\mathcal{X}_i, h_i] + (1 + C)\Gamma_{i,T} + 2\sqrt{C}\Psi_{i,T} \\ &= 2\mathcal{D}[\mathcal{X}_i, h_i] + (1 + C)\frac{5}{2K_i}M_*^2\sum_{t=1}^{T}(\gamma_i^t)^2 \\ &+ 8\sqrt{2C\mathcal{D}[\mathcal{X}_i, h_i]/K_i}M_*\sqrt{\sum_{t=1}^{T}(\gamma_i^t)^2}.\end{aligned}$$

For all $T \geq 1$ and for all $\varepsilon \in (0, 1)$, with probability at least $1 - \varepsilon$, we have

$$\mathrm{Gap}_i(\mathcal{T}) \leq \mathcal{Q}_{i,T}(\log(2/\varepsilon)).$$

Proof. Observe that $\mathbb{E}[\xi_i^t \mid \mathcal{F}_{t-1}] = 0$ for all $t \geq 1$. Therefore $\Xi_{i,T}$, defined in (2.137), is a martingale with respect to the filtration $\mathbb{F} := \{\mathcal{F}_t\}_{t\geq 1}$, which is also bounded in $L^2(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, thanks to (2.87).

2.70 implies that

$$\mathrm{Gap}_i(\mathcal{T}) \leq 2\mathcal{D}[\mathcal{X}_i, h_i] + \Phi_{i,T} + \Xi_{i,T},$$

so $\{\mathrm{Gap}_i(\mathcal{T}) \geq \mathcal{Q}_{i,T}(C)\} \subseteq \{\Phi_{i,T} + \Xi_{i,T} \geq (1+C)\Gamma_{i,T} + 2\sqrt{C}\Psi_{i,T}\}$. Consequently, from 2.9.3 , we deduce that for all $C > 0$,

$$\mathbb{P}(\mathrm{Gap}_i(\mathcal{T}) \geq \mathcal{Q}_{i,T}(C)) \leq 2\exp(-C).$$

Choosing $C = \log(2/\varepsilon)$ proves our claim. □

Proof of 2.6.5 . From the variational characterization (3.13) for the external regret, the Prox-strategy with a constant step-size $\gamma_i^t \equiv \gamma_i$ gives

$$\mathrm{Reg}_i(\mathcal{T}) \leq \max_{x_i \in \mathcal{X}_i} \sum_{t=1}^{T} \langle v_i^t(X^t), x_i - X_i^t \rangle \leq \frac{2\mathcal{D}[\mathcal{X}_i, h_i]}{\gamma_i} + \frac{1}{\gamma_i}(\Phi_{i,T} + \Xi_{i,T}).$$

Hence, for all $\rho > 0$,

$$\{\mathrm{Reg}_i(\mathcal{T}) \geq \rho\} \subseteq \{\Phi_{i,T} + \Xi_{i,T} \geq \gamma_i \rho - 2\mathcal{D}[\mathcal{X}_i, h_i]\}.$$

Therefore, choosing $\rho = \mathcal{Q}_{i,T}(C)/\gamma_i$ we deduce from 2.138 that

$$\begin{aligned}
&\mathbb{P}(\mathrm{Reg}_i(\mathcal{T}) \geq \mathcal{Q}_{i,T}(C)/\gamma_i) \\
\leq &\mathbb{P}(\Phi_{i,T} + \Xi_{i,T} \geq (1+C)\Gamma_{i,T} + 2\sqrt{C}\Psi_{i,T}) \\
\leq &2\exp(-C).
\end{aligned}$$

Picking $C = \log(2/\varepsilon)$, for any $\varepsilon \in (0,1)$ fixed, we get the desired $(1-\varepsilon)$-probability bound. Now, observe that for a constant step-size, we have

$$\begin{aligned}
\frac{\mathcal{Q}_{i,T}(\log(2/\varepsilon))}{\gamma_i} = &\frac{2\mathcal{D}[\mathcal{X}_i, h_i]}{\gamma_i} + \frac{5(1+\log(2/\varepsilon))M_*^2}{2K_i}\gamma_i T \\
&+ 8M_*\sqrt{2\log(2/\varepsilon)\mathcal{D}[\mathcal{X}_i, h_i]/K_i}\sqrt{T}.
\end{aligned}$$

Call $\Omega_i^\varepsilon := \frac{5(1+\log(2/\varepsilon))M_*^2}{2K_i}$, and optimizing the above expression with respect to γ_i, gives the optimal constant step-size

$$\gamma_i = \sqrt{\frac{2\mathcal{D}[\mathcal{X}_i, h_i]}{\Omega_i^\varepsilon T}}.$$

Using this step size in the previous display gives

$$\frac{\mathcal{Q}_{i,T}(\log(2/\varepsilon))}{\gamma_i} = 2\sqrt{2T\mathcal{D}[\mathcal{X}_i, h_i]\Omega_i^\varepsilon} + 8M_*\sqrt{2T\log(2/\varepsilon)\mathcal{D}[\mathcal{X}_i, h_i]/K_i}.$$

This shows that, with probability at least $1 - \varepsilon$, we have

$$\mathrm{Reg}_i(\mathcal{T}) \leq 2\sqrt{2T\mathcal{D}[\mathcal{X}_i, h_i]\Omega_i^\varepsilon} + 8M_*\sqrt{2T\log(2/\varepsilon)\mathcal{D}[\mathcal{X}_i, h_i]/K_i}$$

and our proof is complete. □

3

Variational inequalities with applications to dynamic user equilibrium in traffic networks

In this Chapter[1], we use a class of strongly convergent primal-dual schemes for solving variational inequalities defined by a Lipschitz continuous and pseudo-monotone map in infinite-dimensional Hilbert spaces solve dynamic user equilibria network flows.

3.1 Introduction

Variational inequalities (VIs) are a flexible mathematical formulation of many equilibrium problems in engineering, machine learning, operations research and economics (see [25] for a masterful survey of theory and applications of finite-dimensional VIs). Formulated on an

[1]This chapter is based on [22]. I would like to thank the referees for their helpful comments and discussion.

infinite-dimensional real Hilbert space, variational inequalities also play a key role in the field of partial differential equations (PDEs) and optimal control, with important applications in imaging, differential equations, and network flows [60, 53]. This chapter is concerned with the computation of *dynamic user equilibria* in traffic networks using an algorithm which solves such variational inequalities.

3.1.1 Dynamic user equilibrium

Dynamic user equilibrium (DUE) is a widely studied form of *dynamic traffic assignment* (DTA), in which road travelers engage in a non-cooperative Nash game with departure time and route choices. One characteristic feature of DTA is that it provides a "general equilibrium" model whose aim is to predict departure rates, departure times and route choices of travelers over a given time horizon. Exact DTA models are built on two layers: (i) a game-theoretic formulation of trip assignment, such as the dynamic extension of Wardrop's first principle [111]; (ii) a network flow model, which captures the physical relationships between entry and exit flows, junction flows, link delay and path delay. The latter is referred in the literature as *dynamic network loading* (DNL). The DNL procedure is a manifestation of the physical principles of traffic flows, and various formulation of DNL exist in the literature, ranging from fluid models to differential equations. We refer the reader to the survey [10] and the book [38] for an in-depth treatment of this important subject. We focus in this paper on the computation of DUE, leaving the network loading in the back. Section 3.4 gives a precise explanation how this division between the two levels works. A key challenge in the algorithmic approach to DUE is the usual lack of a closed-form expression of the *delay operator*. The delay operator is the quantity of interest in DUE, since it informs us about the latencies on the individual paths of the traffic network. Indeed, as shown already in the seminal work [31], the delay operator is the defining map in the VI approach of dynamic user equilibrium. However, without detailed information on this map, it is impossible

to make a-priori monotonicity statements, which are crucial in the choice of numerical algorithms to solve the variational inequality. In fact, even if an explicit expression for the delay operator is available, it has been shown in [81] that strong monotonicity cannot hold for general networks and DNL models. Hence, any numerical algorithm guaranteeing *strong convergence* under a-priori *weak monotonicity* assumptions marks a breakthrough in the applicability of DUE as a predictive tool for traffic engineers.

The literature on DUE is huge, and naturally it is impossible to give a fair representation of all available results. We therefore only give a summary of those contributions which are the most related to our work. In finite-dimensions, the connection between VIs and traffic user equilibrium is classical (see e.g. [25]). Once the user equilibrium problem is put into a dynamic setting, the natural model domain is the space of path-flows, which are assumed to be square-integrable functions satisfying a natural conservation condition. To our knowledge, the VI formulation of dynamic flows over time has been first presented in [31]. Departing from that work, the field has grown substantially, and various numerical schemes have been constructed to solve the resulting VI under different global regularity and Lipschitz continuity assumptions on the involved operator.[2] A gradient projection method is studied in [48]. Weak convergence of this method is known if the operator is Lipschitz continuous and strongly monotone [4]. As noted in [51], relaxing strong monotonicity assumptions could even lead to divergence of the algorithm. [64] blue develops an alternating direction method under the assumption that the delay operator is *cocoercive*. Sufficient for cocoercivity is Lipschitz-continuity and monotonicity, so again we need to make rather restrictive global monotonicity assumption. Assuming weaker monotonicity conditions, the well-known extragradient scheme, due to Korpelevich and Antipin [62, 2], has been employed in [65] to solve for DUE. In [110] the weak convergence of the extragradient method

[2] In terms of numerical analysis, these papers can thus be seen to follow the classical philosophy to *first optimize, then discretize*.

is studied in some detail. A further drawback of the extragradient method is that it requires two costly projection steps at each iteration, making it a relatively unattractive method given our desire to have schemes with computationally cheap iterations. [48] also discusses a proximal-point algorithm, first studied by Martinet [71] and Rockafellar [94], and the self-adaptive projection scheme of [46]. Again, without assuming strong monotonicity, proximal-point methods are known to converge only weakly [43], and the self-adaptive projection scheme has been introduced in [46] in a finite-dimensional setting, making the distinction between weak and strong convergence meaningless. In light of the above survey, the following research question emerges:
Can we develop a numerical algorithm with computationally cheap iterations and exhibiting strong convergence of the iterates under mild monotonicity assumptions?
Dennis Meier gave in his PhD-thesis (the analysis is done in [22]) an affirmative answer to this question. We will use his algorithms in this chapter to compute dynamic user equilibria in traffic networks. The rest of this chapter is organized as follows. In Section 3.2 we introduce standard notation and concepts from variational analysis. Section 3.3 describes the numerical scheme. An extension of the basic scheme to an adaptive algorithm is also discussed in that Section, showing that we can even get rid of Lipschitz continuity assumptions when designing the algorithm's parameters. Section 3.4 reports numerical experiments in solving dynamic user equilibria in standard test instances, and compares our method with the projection-based algorithm described in [32, 48, 47].

3.2 Preliminaries

We follow the standard notation as in [4]. Let $\mathbb{N} := \{1, 2, \ldots\}$ be the set of positive integers and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$ the set of nonnegative integers. Let $\mathcal{H}$ be a real separable Hilbert space with inner product $\langle x, y\rangle$ and induced norm $\|x\| := \sqrt{\langle x, x\rangle}$. A sequence $(x_n)_{n\in\mathbb{N}}$ converges strongly to a point $x \in \mathcal{H}$ if $\lim_{n\to\infty}\|x_n - x\| = 0$. A sequence $(x_n)_{n\in\mathbb{N}}$ converges

weakly to a point $x \in \mathcal{H}$ if, for every $u \in \mathcal{H}$, $\langle x_n, u\rangle \to \langle x, u\rangle$; in symbols $x_n \rightharpoonup x$.

Let $\mathcal{X} \subseteq \mathcal{H}$ be a closed convex nonempty subset. Define the normal cone mapping by $\mathsf{NC}_{\mathcal{X}}(x) := \{u \in \mathcal{H} | \langle u, y - x\rangle \leq 0 \quad \forall y \in \mathcal{X}\}$ if $x \in \mathcal{X}$, and $\mathsf{NC}_{\mathcal{X}}(x) = \emptyset$ otherwise. The Euclidean projector onto $\mathcal{X}$ is defined as $P_{\mathcal{X}}(x) := \operatorname{argmin}_{y\in\mathcal{X}} \frac{1}{2}\|y - x\|^2$. It is well known that $P_{\mathcal{X}}$ is nonexpansive and the following property, taken from [42], hold.

Lemma 3.2.1. *Let $\mathcal{X}$ be a nonempty closed convex subset of a real Hilbert space $\mathcal{H}$. Given $x \in \mathcal{H}$ and $z \in \mathcal{X}$. Then*

$$z = P_{\mathcal{X}}(x) \iff \langle x - z, z - y\rangle \geq 0 \qquad \forall y \in \mathcal{X}. \tag{3.1}$$

Definition 3.2.2. A mapping $F : \mathcal{H} \to \mathcal{H}$ is *pseudo-monotone* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$ it holds

$$\langle F(x), y - x\rangle \geq 0 \Rightarrow \langle F(y), y - x\rangle \geq 0 \tag{3.2}$$

The mapping $F : \mathcal{H} \to \mathcal{H}$ is *monotone* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$ it holds

$$\langle F(x) - F(y), x - y\rangle \geq 0. \tag{3.3}$$

Clearly, pseudo-monotonicity is a weakened monotonicity assumption providing enough structure to derive provably strongly convergent algorithms. In particular, if $F = \nabla f$ is the gradient of a differentiable real-valued function $f : \mathcal{H} \to \mathbb{R}$, pseudo-monotonicty of F implies *pseudo-convexity* of the function f and vice versa. Pseudo-convexity is the classical assumption involved in existing proofs of economic equilibria and Nash equilibria in games with continuous action spaces [25].

3.3 A strongly convergent algorithm for pseudo-monotone VIs

We are given a mapping $F : \mathcal{H} \to \mathcal{H}$, satisfying the following assumptions:

Assumption 1. $F : \mathcal{H} \to \mathcal{H}$ is Lipschitz continuous with Lipschitz constant $L > 0$, and sequentially weak-to-weak continuous on bounded subsets of $\mathcal{H}$.

Recall that weak-to-weak continuity requires that for every weakly converging sequence $x_n \rightharpoonup x$, it follows that $F(x_n) \rightharpoonup F(x)$ [4]. In terms of regularity, we also rely on the following mild monotonicity assumption on the map F:

Assumption 2. $F : \mathcal{H} \to \mathcal{H}$ is *pseudo-monotone* on $\mathcal{X}$: For all $x, y \in \mathcal{X}$ it holds

$$\langle F(x), y - x\rangle \geq 0 \Rightarrow \langle F(y), y - x\rangle \geq 0. \tag{3.4}$$

We present an algorithm which solves the Hilbert-space valued variational inequality $\mathrm{VI}(\mathcal{X}, F)$:

$$\text{find } x^* \in \mathcal{H} \text{ such that } \langle F(x^*), x - x^*\rangle \geq 0 \qquad \forall x \in \mathcal{X}. \tag{3.5}$$

It is known that under Assumption 1 and Assumption 2, the solution set of $\mathrm{VI}(\mathcal{X}, F)$, which is denoted by $\mathcal{X}_*$, is a nonempty closed and convex set (see [82] and [4]).

3.3.1 Algorithmic Setting

In this section we present two strongly convergent numerical schemes for solving $\mathrm{VI}(\mathcal{X}, F)$ under Assumptions 1-2 introduced in the thesis of Meier. The building block of this construction is the classical forward-backward-forward algorithm proposed by Tseng [108], in the context

of solving monotone inclusions. As is well known, the advantage of Tseng's splitting technique is that it allows us to treat monotone inclusions for finding zeroes of the operator $A + B$, where $A : \mathcal{H} \to 2^{\mathcal{H}}$ and $B : \mathcal{H} \to \mathcal{H}$ are both maximally monotone and B is L-Lipschitz. Compared to the celebrated forward-backward splitting, Tseng's method does not require cocoercivity of the single-valued operator B. When applied to variational inequalities, the main advantage of the forward-backward-forward method is that it requires only a single projection step at each iteration, which makes the algorithm much more efficient in practice relative to its close competitor the extragradient method of [62].[3] We first present a non-adaptive version of our strongly convergent forward-backward-forward algorithm (Algorithm 3). This scheme iteratively constructs a sequence $(x^k, r^k, z^k)_{k\in\mathbb{N}_0} \subset \mathcal{H} \times \mathcal{H} \times \mathcal{X}$, where z^k and r^k are just the classical forward-backward-forward iterations. If we would run the scheme only with these two iterative steps, the best we can hope for is weak convergence of the iterates under the common hypothesis that the map F is monotone. The innovative element of the scheme of Meier is the additional extrapolation step generating x^{k+1}, which will be enforcing strong convergence of the trajectories to a minimum norm solution of $\mathrm{VI}(\mathcal{X}, F)$. We would like to point out that this modification of the forward-backward-forward scheme is much simpler than the one presented in [107] since no hyperplane projection subroutine is involved in our construction. In view of our objective to develop numerical methods with cheap iterations, this is a notable feature of the algorithmic approach of Meier, which is similar to that of [107]. In the literature, there is another approach called bilevel that is often used not only to obtain strong convergence to the least norm element, but also to handle ill-posed problems (see, e.g. [18]). The main theoretical result of the PhD-thesis of Meier reads then as follows:

Theorem 3.3.1. *Suppose that Assumptions 1-2 are satisfied. Let* $(\alpha_k)_{k\in\mathbb{N}_0}$ *and* $(\beta_k)_{k\in\mathbb{N}_0}$ *be two real sequences in* $(0, 1)$*, such that* $(\beta_k)_{k\in\mathbb{N}_0} \subset (\alpha, 1-\alpha_k)$

[3] See [8, 103, 9] for an in-depth discussion in stochastic and deterministic variational inequality problems.

Algorithm 3 FBF for VI($F, \mathcal{X}$).

Require: step-size sequence $\gamma \in (0, 1/L)$, map $F : \mathcal{H} \to \mathcal{H}$,
parameters $(\alpha_k)_{k\in\mathbb{N}_0}, (\beta_k)_{k\in\mathbb{N}_0} \subset (0,1)$,
Initial point $x^0 \in \mathcal{X}$ #initialization
while $k = 0, 1, \ldots, k_{\max}$ **do**
  obtain x^k
  if Stopping condition not satisfied **then**
    Compute $z^k = P_{\mathcal{X}}[x^k - \gamma F(x^k)]$
    Compute $r^k = z^k + \gamma(F(x^k) - F(z^k))$
    Update $x^{k+1} = (1 - \alpha_k - \beta_k)x^k + \beta_k r^k$.
  else
    Stop and report x^k as the solution
  end if
end while

for some $\alpha > 0$, *and*

$$\lim_{k\to\infty} \alpha_k = 0, \sum_{k=1}^{\infty} \alpha_k = \infty. \tag{3.6}$$

Then the sequence $(x^k)_{k\in\mathbb{N}_0}$ *generated by Algorithm 3 converges strongly to* $p \in \mathcal{X}_*$, *where* $p = \operatorname{argmin}\{\|z\|\colon z \in \mathcal{X}_*\}$.

Beside excellent convergence properties and computationally cheap iterations, Algorithm 3 requires knowledge of the Lipschitz constant of the map F. In practice we usually have no information about such a global quantity, making the applicability of Algorithm 3 questionable. Fortunately, we can circumvent this annoying strong assumption by constructing a simple adaptive step-size policy relying on evaluations of the function F only, without requesting explicit knowledge of the Lipschitz constant. Specifically, let us consider a sequence $(\gamma_k)_{k\in\mathbb{N}_0}$,

Algorithm 4 FBF for VI$(F, \mathcal{X})$ adaptative step-size.

Require: step-size sequence $\gamma_0 > 0$, map $F : \mathcal{H} \to \mathcal{H}$,
parameters $\rho \in (0,1)$, $(\alpha_k)_{k\in\mathbb{N}_0}, (\beta_k)_{k\in\mathbb{N}_0} \subset (0,1)$.
Ensure: Minimal norm solution $x^* \in \mathcal{X}_*$ of VI$(\mathcal{X}, F)$.
Initial point $x^0 \in \mathcal{X}$ #initialization
while $k = 0, 1, \ldots, k_{\max}$ **do**
  obtain x^k
  if Stopping condition not satisfied **then**
    Compute $z^k = P_{\mathcal{X}}[x^k - \gamma_k F(x^k)]$
    Compute $r^k = z^k + \gamma_k(F(x^k) - F(z^k))$
    Update $x^{k+1} = (1 - \alpha_k - \beta_k)x^k + \beta_k r^k$.
    Update new step-size γ_{k+1} by (3.7).
  else
    Stop and report x^k as the solution
  end if
end while

defined recursively by

$$\gamma_{k+1} := \begin{cases} \min\left\{\frac{\rho\|z^k - x^k\|}{\|F(z^k) - F(x^k)\|}, \gamma_k\right\} & \text{if } F(z^k) - F(x^k) \neq 0, \\ \gamma_k & \text{otherwise.} \end{cases} \tag{3.7}$$

The parameters $\rho \in (0,1)$ and γ_0 are chosen at the beginning of the scheme by the user. It is clear that $(\gamma_k)_{k\in\mathbb{N}_0}$ is non-increasing and bounded from above by $\min\left\{\gamma_0, \frac{\rho}{L}\right\}$. This implies that the sequence $(\gamma_k)_{k\in\mathbb{N}_0}$ has a limit point not smaller than $\left\{\gamma_0, \frac{\rho}{L}\right\}$. Replacing in Algorithm 3 the constant step-size γ by the sequence $(\gamma_k)_{k\in\mathbb{N}_0}$, leads us directly an adaptive forward-backward-forward scheme, precisely defined in Algorithm 4. This adaptive stepsizes were also proposed in [41] for solving maximal monotone inclusion problem.

Theorem 3.3.2. *Suppose that Assumptions 1-2 are satisfied. Let $(\alpha_k)_{k\in\mathbb{N}_0}$ and $(\beta_k)_{k\in\mathbb{N}_0}$ be two real sequences in $(0,1)$, satisfying the same conditions as in Theorem 3.3.1. Let $(\gamma_k)_{k\in\mathbb{N}_0}$ be designed by the adaptive rule* (3.7).

Then the sequence $(x^k)_{k\in\mathbb{N}_0}$ *generated by Algorithm 4 converges strongly to* $p = \operatorname{argmin}\{\|z\|\colon z \in \mathcal{X}_*\}$.

The proof of this Theorem only requires a simple twist of the proof of Theorem 3.3.1 , and is given at the end of the next Section.

The proof of those theorems can be foudn in [22].

3.4 Application to computing dynamic user equilibria

In this section we apply the strongly-convergent forward-backward-forward algorithm to compute dynamic user equilibria in two standard test examples taken from the literature. Our description follows the recent survey [32].

3.4.1 Problem Formulation

Let $[t_0, t_1]$ be a fixed planning horizon. We are given a connected directed graph $G = (V, A)$ with finite set of vertices V, representing traffic intersections (junctions) and arc set A, representing road segments. A path p in the graph G is identified with a non-repeating finite sequence of arcs which connect a sequence of different vertices. Hence, an arbitrary path p is identified with the list of edges incident to it, i.e. $p = \{a_1, a_2, \ldots, a_m\}$. The integer $m = m(p)$ denotes the length of the path p. We denote the set of all paths of interest by $\mathcal{P}$, and set $\mathcal{H} := \mathbb{R}^{|\mathcal{P}|}$. We are interested in paths which connect a set of distinguished vertices acting as the *origin-destination* (o/d) pairs in our graph. We are given N distinct o/d pairs denoted as $w_1, \ldots, w_N$, where each $w_i = (o_i, d_i) \in V$. Call $\mathcal{W} := \{w_1, \ldots, w_N\}$, and the set of paths connecting the o/d pair w is denoted by $\mathcal{P}_w \subseteq \mathcal{P}$. For each o/d pair $w \in \mathcal{W}$ we are given a *demand* $Q_w > 0$; This represents the number of drivers who have to travel from the origin to the destination described by w. For simplicity we assume that this demand is exogenously given. The list $Q = (Q_w)_{w\in\mathcal{W}}$ is often called the *trip table*. In DUE modeling, the single most crucial ingredient

is the path delay operator, which maps a given vector of departure rates (path flows) h to a vector of path travel times. We stipulate that path flows are square integrable functions over the planning horizon, so that $h_p \in L^2([t_0, t_1]; \mathbb{R}_+)$ and $h = (h_p; p \in \mathcal{P}) \in \mathcal{H} := L^2([t_0, t_1]; \mathcal{H})$. To measure the delay of drivers on paths, we introduce the operator $D : \mathcal{H} \to \mathcal{H}, h \mapsto D(h)$, with the interpretation that $D_p(t, h)$ is the path travel time of a driver departing at time t from the origin of path p, and following this path throughout. This operator is the result of a dynamic network loading procedure, which is an integrated subroutine in the dynamic traffic assignment problem. See [47] for further information.

On top of path delays, we consider penalty terms of the form $\rho(t + D_p(t, h) - T_A)$, penalizing all arrival times different from the target time T_A (i.e. the usual time of a trip on the o/d. pair w). The function $\rho : [-\infty, \infty) \to [0, \infty]$ should be monotonically increasing with $\rho(x) > 0$ for $x > 0$ and $\rho(x) = 0$ for $x \leq 0$. Define the effective delay operator as

$$\Psi_p(t, h) := D_p(t, h) + \rho(t + D_p(t, h) - T_A). \tag{3.8}$$

We thus obtain an operator $\Psi : \mathcal{H} \to \mathcal{H}$, mapping each profile of path departure rates h to effective delays $\Psi(h) \in \mathcal{H}$.

We follow the perceived DUE literature, and stipulate that Wardrop's first principle holds: Users of the network aim to minimize their own travel time, given the departure rates in the system. Thus, a user equilibrium is envisaged, where the delays (interpreted as costs) of all travelers in the same o/d pair are equal, and no traveler can lower his/her costs by unilaterally switching to a different route. To put this behavioral axiom into a mathematical framework, we first formulate the meaning of "minimal costs" in the present Hilbert space setting. Recall the essential infimum of a measurable function $g : [t_0, t_1] \to \mathbb{R}$ as $\text{essinf}\{g(t) | t \in [t_0, t_1]\} = \sup\{x \in \mathbb{R} | \mathsf{Leb}(\{s \in [t_0, t_1] : g(s) < x\}) = 0\}$, where $\mathsf{Leb}(\cdot)$ denoted the Lebesgue measure on the real line. Given a profile $h \in \mathcal{H}$, define

$$\nu_p(h) := \text{essinf}\{\Psi_p(t, h) | t \in [t_0, t_1]\} \qquad \forall p \in \mathcal{P}, \text{ and} \tag{3.9}$$

$$\nu_w(h) := \min_{p \in \mathcal{P}_w} \nu_p(h) \qquad \forall w \in \mathcal{W}. \tag{3.10}$$

On top of minimal costs, we have to restrict the set of departure rates to functions satisfying a basic flow conservation property. Specifically, insisting that all trips are realized, we naturally define the set of feasible flows as

$$\Lambda := \left\{ f \in \mathcal{H} \,\middle|\, \sum_{p \in \mathcal{P}_w} \int_{t_0}^{t_1} f_p(t)\, dt = Q_w \quad \forall w \in \mathcal{W} \right\}. \tag{3.11}$$

Definition 3.4.1. A profile of departure rates $h^* \in \mathcal{H}$ is a *dynamic user equilibrium* (DUE) if

(a) $h^* \in \Lambda$, and

(b) $h_p^*(t) > 0 \Rightarrow \Psi_p(t, h^*) = \nu_w(h^*)$.

In [31] it is observed that the definition of DUE can be formulated equivalently as a variational inequality $\mathrm{VI}(\Lambda, \Psi)$: A flow $h^* \in \Lambda$ is a DUE if

$$\langle \Psi(h^*), h - h^* \rangle \geq 0 \qquad \forall h \in \Lambda \tag{3.12}$$

3.4.2 A Strongly Convergent Forward-Backward-Forward Scheme for DUE

Departing from (3.12), our aim is to solve the DUE problem by using the strongly convergent forward-backward-forward scheme 3 . Adapting this scheme to the usual notation in DUE, we arrive at Algorithm 5.

Some remarks on the implementation of this algorithm are in order. First, it should be pointed out that Algorithm 5 requires two evaluations of the delay operator Ψ. As already said, this operator is the outcome of an inner procedure, solving the dynamic network loading part of the model. Dynamic network loading is a separate computational step in the dynamic traffic assignment problem. A very popular formulation

Algorithm 5 FBF algorithm for computing DUE

Require: Graph $G = (V, A)$ with o/d pairs $\mathcal{W} \subset V^2$ step-size $\gamma > 0$

Trip Table $(Q_w)_{w \in \mathcal{W}}$, parameters $(\alpha_k)_{k \geq 0}, (\beta_k)_{k \geq 0} \subset \mathbb{R}_+$

Ensure: An approximate DUE h^*

Initial path flow $h^0 \in \mathcal{H}$ #initialization

while $k = 0, 1, \ldots, k_{\max}$ **do**

  obtain h^k

Compute $\epsilon_k = \frac{\|h^{k+1} - h^k\|^2}{\|h^k\|^2}$

  if $\epsilon_k > 10^{-4}$ **then**

    Compute the effective path delays $\Psi_p(t, h^k)$

    Compute $z^k = P_\Lambda[h^k - \gamma \Psi(h^k)]$

    Update Compute the effective path delays $\Psi_p(t, z^k)$

    Compute $r^k = z^k + \gamma(\Psi(h^k) - \Psi(z^k))$

    Compute $h^{k+1} = (1 - \alpha_k - \beta_k)h^k + \beta_k r^k$;

  else

    Stop and report $h^k = h^*$ as the solution.

  end if

end while

of dynamic network loading is the fluid dynamic approximation of traffic flows, known as the Lighthill-Whitham-Richards (LWR) model. We refer the interested reader to [38] for modeling approaches of the dynamic network loading procedure. In case of the popular LWR model evaluating the delay operator requires solving a coupled system of hyperbolic partial differential equations for the traffic density. It is clear that this procedure is the most costly step in the implementation of Algorithm 5.

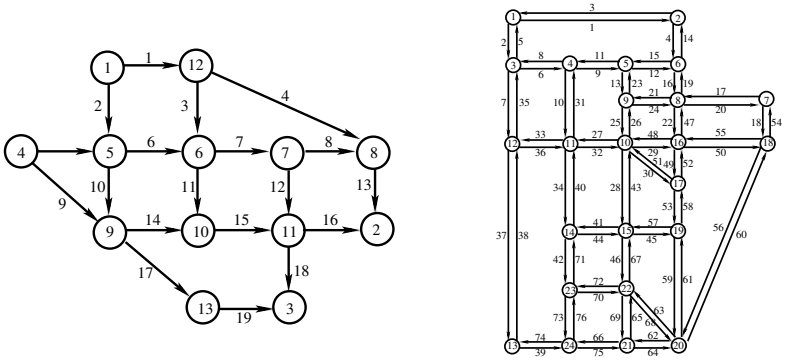
Algorithm 5 is, modulo the obvious change in notation, equivalent to Algorithm 3 if the delay operator Ψ is Lipschitz continuous and pseudo-monotone. Numerous computational algorithms for solving

DUE problem were summarized in [47, Table 1]. For each solution method, there are assumptions for which the convergence can be guaranteed. At the moment, the weakest conditions for DUE commuinity is pseudo-monotonicity and Lipschitz continuity of the cost operator Ψ. We refer the readers to [47] and extensive references quoted therein. Currently, weak-to-strong continuity of the delay operator has been established for many network loading models, including the LWR model. See [37] for a state-of-the-art summary. As for monotonicity, global Lipschitz continuity has not been established rigorously for dynamic traffic assignment models. Hence, like any other fixed-point algorithm applied to DUE, our application of the strongly convergence FBF scheme has to be understood as a heuristic. However, in all numerical studies we present below, the fixed-point iterations do meet the convergence criterion (3.13).

3.4.3 Numerical Experiments

We implemented Algorithm 5 in MATLAB, building on the open-source MATLAB package described in [47], and accessible under https://github.com/DrKeHan/DTA. As DNL subroutine a numerical implementation of the LWR model is used, generating the delay operator $\Psi(h)$ at flow profile $h \in \mathcal{H}$. By adapting this toolbox to Algorithm 5 , we compute dynamic user equilibria for the Nguyen and the Sioux fall network (see Figure 3.1) and compare our results with the projected gradient method. The parameters α_k, β_k and γ were chosen for each instance separately to guarantee the best convergence. The Nguyen network is a traffic network with 13 nodes connected by 19 links, and 4 o/d pairs. There are 24 paths to compute. The Sioux fall is a significantly larger instance, consisting of 76 links, 24 nodes, 530 o/d pairs and 6,180 paths. We stop the algorithm if the relative gap is smaller than a user defined tolerance, i.e.

$$\varepsilon_k := \frac{\|h^{k+1} - h^k\|^2}{\|h^k\|^2} \leq 10^{-4}. \tag{3.13}$$



Nguyen network
(13 nodes, 19 links, 4 zones)

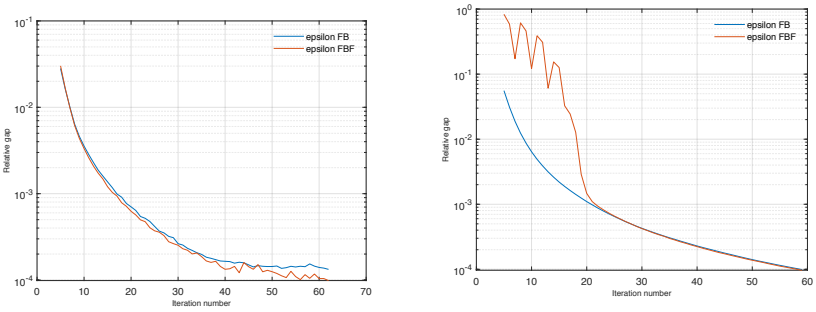
Sioux Falls network
(24 nodes, 76 links, 24 zones)

Figure 3.1: The Nguyen and Sioux Falls network.

This measure can be interpreted as the iteration complexity of the algorithm employed. Figure 3.2 shows the relative gaps for the Nguyen and the Sioux fall networks until the stopping criterion is reached. It can be seen from this Figure that both methods have a similar iteration complexity, with a slight tendency favoring our FBF approach. Figure 3.3 shows the path departure rates as well as the corresponding effective path delays. We observe that the departure rates are nonzero only when the corresponding effective delays are equal and minimum, which conforms to the notion of DUE. To rigorously assess the quality of obtained DUE solutions, we define the gap function between each o/d pair $w \in \mathcal{W}$ as

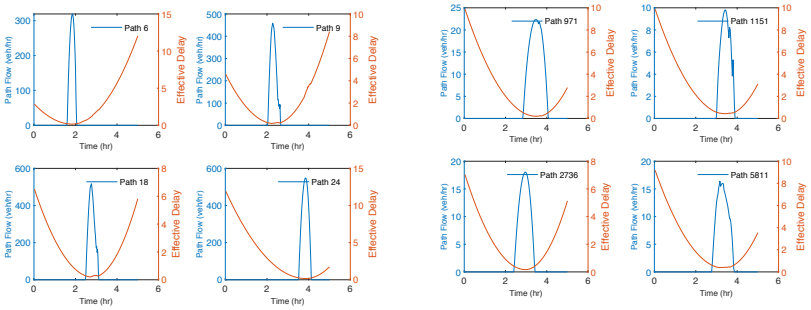
$$\begin{aligned}\Gamma_w = &\max\{\Psi_p(h^*, t), t \in [t_0, t_1], p \in \mathcal{P}_w \text{ such that } h_p^*(t) > 0\} \\ &- \min\{\Psi_p(h^*, t), t \in [t_0, t_1], p \in \mathcal{P}_w \text{ such that } h_p^*(t) > 0\} \quad (3.14)\end{aligned}$$

In an exact DUE, we should have $\Gamma_w = 0$ for all $w \in \mathcal{W}$. Figure 3.4 displays histograms of o/d gaps obtained by running FBF and the projection method of [47] until the stopping criterion is reached. It is



(a) Ngyuen Network **(b)** Sioux Fall Network

Figure 3.2: Relative gap (3.13) (called epsilon in the figure) computed under the forward-backward iteration of [47] and Algorithm 5 , using the same parameter values



(a) Nguyen Network **(b)** Siuox Fall network

Figure 3.3: Path departure rates and corresponding effective path delays of selected paths in the DUE solutions.

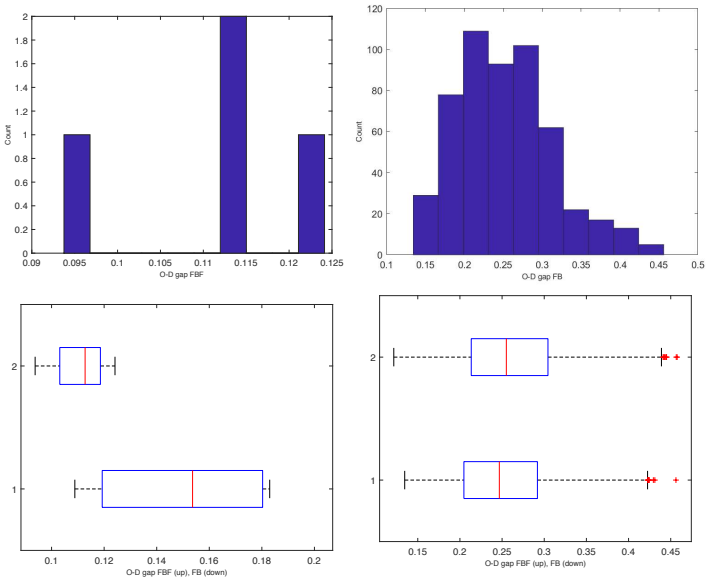


Figure 3.4: Distributions of O-D gaps corresponding to the DUE solutions. The O-D gap is calculated according to (3.14).

seen that most o/d gaps are varying between 0.1 and 0.3 for both test instances, reflecting the early stopping of the method. We highlight that Algorithm 5 beats the projection method in the Nguyen network significantly, while it is comparable in overall performance in the Sioux fall network, and at the same time is a strongly convergent method. This provides strong evidence for the good performance of our scheme.

3.5 Conclusions and perspectives

In this chapter, we used a recent strongly convergent numerical scheme for Hilbert-space valued variational inequality problems. We implemented this algorithm in order to solve a challenging class of dynamic

user equilibrium problems, and verified its competitiveness with state-of-the-art solvers used in the transportation science literature. It seems to be possible to extend the scheme to a larger class of variational problems, where distributed implementations are important, such as generalized Nash equilibrium.

4

Competitive search games with a moving target

In this Chapter [1] we introduce a discrete-time search game, in which two players compete to find an object first. The object moves according to a time-varying Markov chain on finitely many states. The players know the Markov chain and the initial probability distribution of the object, but do not observe the current state of the object. The players are active in turns. The active player chooses a state, and this choice is observed by the other player. If the object is in the chosen state, this player wins and the game ends. Otherwise, the object moves according to the Markov chain and the game continues at the next period.

We show that these games admit a value, and for any error-term $\varepsilon > 0$, each player has a pure (subgame-perfect) ε-optimal strategy. Interestingly, a 0-optimal strategy does not always exist. The ε-optimal strategies are robust in the sense that they are 2ε-optimal on all finite but sufficiently long horizons, and also 2ε-optimal in the discounted

[1] This chapter is based on [21]. I would like to thank Steve Alpern, Jérôme Renault and Miquel Oliu-Barton for their helpful comments and discussion.

version of the game provided that the discount factor is close to 1. We derive results on the analytic and structural properties of the value and the ε-optimal strategies. Moreover, we examine the performance of the finite truncation strategies, which are easy to calculate and to implement. We devote special attention to the important time-homogeneous case, where additional results hold.

4.1 Introduction

The field of search problems is one of the original disciplines of Operations Research. In the basic settings, the searcher's goal is to find a hidden object, also called the target, with maximal probability or as soon as possible. By now, the field of search problems has evolved into a wide range of models. The models in the literature differ from each other by the characteristics of the searchers and of the objects. Concerning objects, there might be one or several objects, mobile or not, and they might have no aim or their aim is to not be found. Concerning the searchers, there might be one or more. When there is only one searcher, the searcher faces an optimization problem. When there are more than one searcher, searchers might be cooperative or not. If the searchers cooperate, their aim is similar to the settings with one player: they might want to minimize the expected time of search, the worst time, or some search cost function. If the searchers do not cooperate, the problem becomes a search game with at least two strategic non-cooperative players, and hence game theoretic solution concepts and arguments will play an important role. For an introduction to search games, we refer to [1], [35], 2010, 2013, [39], and for surveys see [6] and [55].

We introduce a competitive search game, played at discrete periods in $\mathbb{N}$. An object is moving according to a time-varying Markov chain on finitely many states. Two players compete to find the object first. They both know the Markov chain and the initial probability distribution of the object, but do not observe the current state of the object. Player 1 is active at odd periods, and player 2 is active at even periods. The

active player chooses a state, and this choice is observed by the other player. If the object is in the chosen state, this player wins and the game ends. Otherwise, the object moves according to the Markov chain and the game continues at the next period. If the object is never found, the game lasts indefinitely. In that case, neither player wins.

When the active player chooses a state, he needs to take two opposing effects into account. First, if the object is at the chosen state, then he wins immediately. This aspect makes choosing states favorable where the object is located with a high probability. Second, if the object is not at the chosen state, then knowing this, the opponent gains information: the opponent can calculate the conditional probability distribution of the location of the object at the next period. This aspect makes choosing states favorable where, on condition that the object not being there, the induced conditional distribution at the next period disfavors the opponent. In particular, this conditional distribution should not be too informative, and for example it should not place too high a probability on a state. Clearly, in some cases there is no state that would be optimal for both scenarios at the same time, and hence the active player somehow needs to aggregate the two scenarios in order to make a choice.

Each player's goal is to maximize the probability to win the game, that is, to find the object first. In our model, we do not assume that the players take into account the period when the object is found. Of course, in most cases, maximizing the probability to win will entail at least partially that each player would prefer to find the object at earlier periods, thereby preventing the other player from finding the object. We refer to Section 4.5 on the finite horizon and on the discounted versions of the search game, where the period when the object is found also matters.

The two players have opposite interests, up to the event when the object is never found. More precisely, each player's preferred outcome is that he finds the object, but he is indifferent between the outcome that the other player finds the object and the outcome that the object is never

found. As we will see, the possibility that neither player finds the object will only have minor role, and hence the two players have essentially opposite interests in the search game.

Main results. Our main results can be summarized as follows.

[1] We study the existence of ε-equilibria. A strategy profile is called an ε-equilibrium if neither player can increase his expected payoff by more than ε with a unilateral deviation. We prove that each competitive search game admits an ε-equilibrium in pure strategies, for all error-terms $\varepsilon > 0$ (cf. Theorem 4.3.2 and for subgame-perfect ε-strategies cf. Proposition 4.5.1). The proof is based on topological properties of the game. Interestingly, a 0-equilibrium does not always exist, not even in mixed strategies. We demonstrate it with two different examples (cf. Examples 4.3.1 and 4.3.2).

[2] We examine the properties of ε-equilibria. We show that in each ε-equilibrium, the object is eventually found with probability at least $1 - \varepsilon \cdot |S|$, where $|S|$ is the number of states (cf. Lemma 4.4.1), and that the set of ε-equilibrium payoffs converge to a singleton $(v, 1-v)$, with $v \in [\frac{1}{|S|}, 1]$ as ε vanishes (cf. Proposition 4.4.2 and Theorem 4.4.3). This implies that, in such search games, the two players have essentially opposite interests, and that we may consider v to be the value of the game and the strategies of ε-equilibria as ε-optimal strategies (cf. Definition 4.4.4 and Proposition 4.4.5).

[3] We prove that the ε-optimal strategies are robust in the following sense: they are 2ε-optimal if the horizon of the game is finite but sufficiently long (cf. Theorem 4.6.2), and they are also 2ε-optimal in the discounted version of the game, provided that the discount factor is close to 1 (cf. Theorem 4.6.3).

[4] We investigate the functional and structural properties of the value and the ε-optimal strategies (cf. Theorems 4.8.3, 4.8.4 and 4.5.2). In particular, we consider the set of probability distributions for the location of the object where choosing a particular state is optimal, and show that this set is star-shaped.

[5] Since the ε-optimal strategies may have a complex structure and may be difficult to identify, we examine the finite truncation strategies, which maximize the probability to win in a finite number of periods. We show that each finite truncation strategy, provided that the horizon of the truncation is sufficiently long, is ε-optimal in the search game on the infinite horizon (cf Theorem 4.6.2). Note that the finite truncation strategies are easy to calculate by backward induction and only require finite memory.

[6] We devote attention to the special case when the Markov chain is time-homogenous (cf. Section 4.5.3), as time-homogenous Markov chains are well studied in the literature of Markov chains and frequently used in applications. For time-homogenous Markov chains, we prove additional results. In particular, if the initial probability distribution of the object is an invariant distribution of the time-homogenous Markov chain, then the value is at least $1/2$, so player 1 has a weak advantage (cf. Proposition 4.5.3). Moreover, if the time-homogenous Markov chain is irreducible and aperiodic, then the game admits a 0-equilibrium in pure strategies (cf. Theorem 4.5.4).

Related literature. Discrete search problems with a moving object have been widely investigated. [91], [98], [19] and [58] study the two-state problem. Assuming perfect detection, [84] investigates the three-state problem. [11] considers the search for a target with Markov motion in discrete time and space using an exponential detection function. He provides a necessary and sufficient condition for an optimal search plan and an efficient iterative algorithm for generating optimal plans. [112] studies a discrete effort analogue of [11], in which searchers decide the effort they want to invest in order to find the object at each location they visit. General necessary and sufficient conditions which extend Brown's results to an arbitrary stochastic process for any mixture of discrete and continuous time and space are given in [104]. More recently, [40] study a hide-search game in a random graph, that is a graph in which each edge is available at each period with a positive probability. For extensive surveys, see [6] and [55].

Most of the search games focus on the case of one searcher, or several cooperative searchers. Some problems with several cooperative searchers and one or several moving targets are mentioned in the book of [105], where some algorithms are also studied to solve those problems. To the best of our knowledge, only two models consider several competitive searchers. [83] investigates a non-zero-sum game in which two searchers compete with each other for quicker detection of an object hidden in one of n boxes, with exponential detection functions. Each player wishes to maximize the probability that he detects the object before the opponent detects it. The author shows the existence of an equilibrium point of the form of a solution of simultaneous differential equations, and gets explicit solution results showing that both players have the same equilibrium strategy even though the detection rates are different. [27] investigates the problem in which an agent has to find an object that moves between two locations according to a discrete Markov process, with the additional costless option to wait instead of searching. They find a unique optimal strategy characterized by two thresholds and show that, in a clear contrast with our model, it can never be optimal to search the location with the lower probability of containing the object. They also analyze the case of multiple agents, where the agents not only compete against time but also against each other in finding the object. They find different kinds of subgame perfect equilibria.

As in [84] we investigate functional and structural properties of the objective function. Nakai proved that the function that allocates to a probability distribution the average number of looks before finding the object is continuous, concave and enjoy some linear properties. They also show that the optimal decision regions (see Section 4.5.2) are star-convex. These properties have also been studied in [67] and in the PhD thesis of [56].

Structure of the paper. In Section 4.2, we present the model. In Section 4.3, we examine the existence of ε-equilibrium. In Section 4.4, we argue that the two players have essentially opposite interests, and we define the value and the notion of ε-optimal strategies. In Section 4.6,

we present two relevant strategies, namely the finite truncation strategy and the discounted strategy, and we prove payoff guarantees of those strategies. In Section 4.5 we present additional results related to the structural properties of the value, the subgame-perfect equilibria and the case in which the Markov chain is time-homogeneous. Functional properties of the value can be found in the Appendix. The conclusion is in Section 4.7.

4.2 The model

The Game. We study a competitive search game G played by two players. Let $\mathbb{N} = \{1, 2, 3, \ldots\}$. An object is moving according to a discrete-time Markov chain $(X_t)_{t\in\mathbb{N}}$ on a finite state space S. The initial probability distribution of the object over the set S is given by $p \in \Delta(S)$, and the transitions probabilities at period t are given by an $S \times S$ transition matrix $P_t = [P_t(i,j)]_{(i,j)\in S^2}$, where $P_t(i,j)$ is the probability for the object to move from state i to state j at period t.

At each period $t \in \mathbb{N}$, one of the players is active: At odd periods player 1 is the active player, and at even periods player 2 is the active player. The active player chooses a state $s_t \in S$, which we call the action at period t. If the object is at state $X_t = s_t$, then the active player finds the object and wins the game. Otherwise, the object moves according to the transition matrix P_t at time t and the game enters period $t+1$. We assume that each player observes the actions chosen by his opponent. The transition matrices $(P_t)_{t\in\mathbb{N}}$ and the initial distribution p are known to the players.

The aim of each player is to maximize the probability that he finds the object first.

Histories. A history at period $t \in \mathbb{N}$ is a sequence $h_t = (s_1, \ldots, s_{t-1}) \in S^{t-1}$ of past actions. By $H_t = S^{t-1}$ we denote the set of all histories at period t. Note that H_1 consists of the empty sequence. Let $\mathbb{N}^{\text{odd}} = \{1, 3, 5, \ldots\}$ and $\mathbb{N}^{\text{even}} = \{2, 4, 6, \ldots\}$. We denote by $H^{\text{odd}} = \cup_{t\in\mathbb{N}^{\text{odd}}} H_t$

the set of histories at odd periods, and by $H^{\text{even}} = \cup_{t \in \mathbb{N}^{\text{even}}} H_t$ the set of histories at even periods. Note that at each history h, the players can calculate the probability distribution for the current location of the object.

Strategies. The action sets for both players are $A_1 = A_2 = S$. A strategy $\sigma = (\sigma_t)_{t \in \mathbb{N}^{\text{odd}}}$ for player 1 is a sequence of functions $\sigma_t \colon H_t \to \Delta(S)$. The interpretation is that, at each period $t \in \mathbb{N}^{\text{odd}}$, given the history h_t, the strategy σ_t chooses to search state $s \in S$ with probability $\sigma_t(h_t)(s)$. Similarly, a strategy $\tau = (\tau_t)_{t \in \mathbb{N}^{\text{even}}}$ for player 2 is a sequence of functions $\tau_t \colon H_t \to \Delta(S)$. We denote by Σ and $\mathcal{T}$ the set of strategies for players 1 and 2, respectively. Note that $\Sigma = \prod_{h \in H^{\text{odd}}} \Delta(S)$ and $\mathcal{T} = \prod_{h \in H^{\text{even}}} \Delta(S)$. We say that a strategy is pure if, for any history, it places probability 1 on one action.

Winning probabilities. We define the stopping time[2] of the game by $\Theta = \min\{t \in \mathbb{N} |\ s_t = X_t\}$. Consider a strategy profile (σ, τ). The probability under (σ, τ) that player 1 wins is denoted by $u_1(\sigma, \tau) = \mathbb{P}_{\sigma,\tau}\left(\Theta \in \mathbb{N}^{\text{odd}}\right)$, and that player 2 wins is denoted by $u_2(\sigma, \tau) = \mathbb{P}_{\sigma,\tau}\left(\Theta \in \mathbb{N}^{\text{even}}\right)$. Note that $u_1(\sigma, \tau) + u_2(\sigma, \tau) = 1 - \mathbb{P}_{\sigma,\tau}(\Theta = \infty)$. If the object has not been found before period t, and the history is h_t, the continuation winning probabilities from period t onward are denoted by $u_1(\sigma, \tau)(h_t)$ for player 1 and $u_2(\sigma, \tau)(h_t)$ for player 2. [3]

ε-Equilibrium. Let $\varepsilon \geq 0$ be an error-term. A strategy σ for player 1 is an ε-best response against strategy τ for player 2 if $u_1(\sigma, \tau) \geq u_1(\sigma', \tau) - \varepsilon$ for every strategy σ' of player 1. Similarly, a strategy τ for player 2 is an ε-best response against strategy σ for player 1 if $u_2(\sigma, \tau) \geq u_2(\sigma, \tau') - \varepsilon$ for every strategy τ' of player 2. A strategy profile (σ, τ) is called an ε-equilibrium if σ is an ε-best response against τ and τ is an ε-best response against σ.

[2] With the convention that $\min\{\emptyset\} = +\infty$

[3] When we wish to emphasize the parameter p, we will write $u_1(\sigma, \tau)(p)$ and $u_2(\sigma, \tau)(p)$.

An alternative interpretation of the game. We call the previous game Model [1]. We present an alternative model of this game in perfect information. This model is useful in order to prove the existence of ε-equilibrium for all $\varepsilon > 0$ (cf. Theorem 4.3.2).

Model [2] One could imagine that the game consists of two phases. In the first phase the players choose actions. More precisely, in the first phase player 1 chooses an action at odd periods and player 2 chooses an action at even periods just as before. This results in an infinite sequence of states $(s_1, s_2, \ldots)$. The set of infinite histories is S^∞. Every pure strategy profile (σ, τ) induces a unique infinite history $h^\infty_{\sigma,\tau} \in S^\infty$. In a second phase, players receive a payoff. Now, for $i = 1, 2$, consider the payoff function $f_i : S^\infty \to [0, 1]$ defined as follows. Consider an infinite history $(s_1, s_2, \ldots)$. Take any pure strategy profile (σ, τ) such that $h^\infty_{\sigma,\tau} = (s_1, s_2, \ldots)$ and define $f_i(s_1, s_2, \ldots) = u_i(\sigma, \tau)$. Note that this definition does not depend on the choice of (σ, τ). The goal of each player is to maximize his payoff. Note that this is a game without an object. This way we obtain a two-player perfect information game.

Discussion. We briefly argue that the above descriptions are equivalent. For each pure strategy profile (σ, τ), for each player $i = 1, 2$, we have $u_i(\sigma, \tau) = f_i(h^\infty_{\sigma,\tau})$. Then, a strategy profile in one of the models leads to the same payoff in the other game. The difference is that Model [1] is in imperfect information, as players only know the probability distribution of the object, while Model [2] is in perfect information.

Model [1] gives a very clear, intuitive and concrete description of the game. This is the reason why we usually work with this model in the paper. Model [2] is used as a tool to prove existence of ε-equilibrium as in Theorem 4.3.2.

4.3 Existence of equilibrium

In this section, we examine equilibria in competitive search games. In the first subsection, we show that there are search games for which

there exist no 0-equilibrium, not even in mixed strategies. From a technical point of view, this is caused by discontinuity in the payoff functions of the players. In the second subsection, we focus on the notion of ε-equilibrium, where $\varepsilon > 0$ is an error-term, and prove that each search game admits an ε-equilibrium in pure strategies, for all $\varepsilon > 0$. We conclude the section by presenting an ε-equilibrium for the games introduced in the first subsection.

4.3.1 Search games with no 0-equilibrium

Theorem 4.3.1. *There exist time-homogeneous competitive search games which admit no 0-equilibrium, not even in mixed strategies.*

We provide two counter-examples: Example 4.3.1 and Example 4.3.2. A common property of these counter-examples is that during the game the players are forced to choose states where the probability of the object is positive but converges to zero when t goes to infinity. In Example 4.3.1, this happens within the class of transient states. In contrast, in Example 4.3.2, there are multiple ergodic sets in the Markov chain, and the players are forced to choose states in an ergodic set, even when the probability that the object is in this ergodic set is very small.

Example 4.3.1. Consider the game in Figure 1. In this game, $\eta \in (0, 1/4)$ and the initial probability distribution is $p = (q, q, 1/2 - q, 1/2 - q)$, where $q \in (0, 1/4)$. Notice that states 1 and 2 have the same transition probabilities, and so do states 3 and 4. States 1 and 2 are transient, whereas states 3 and 4 are absorbing.

We show that this game admits no 0-equilibrium. The intuition for this claim is as follows. As we will show, it is not optimal for either player to be the first one who chooses an absorbing state. As a consequence, both players prefer to choose state 1 or state 2 and wait until the other player chooses state 3 or state 4. However, if both players do so, they will choose state 1 and state 2 forever, which is not a 0-equilibrium.

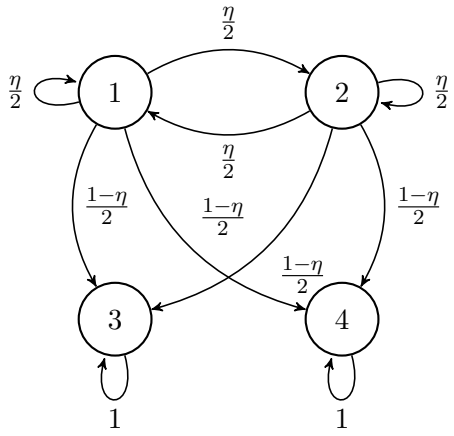


Figure 4.1: A game without 0-equilibrium.

Let $\sigma = (\sigma_t)_{t \in \mathbb{N}^{\text{odd}}}$ be the strategy of player 1 defined as follows. For all $t \in \mathbb{N}^{\text{odd}}$, for all $h_t \in H_t$,

$$\sigma_t(h_t) = \begin{cases} \text{state } 1 & & \text{if } t = 1, \\ \text{state } 3 & & \text{if } t \geq 3 \text{ and } h_t(t-1) = 4 \\ & \text{or} & \text{if } t \geq 3 \text{ and } h_t(t-1) \in \{1,2\} \text{ and } h_t(t-2) \neq 3, \\ \text{state } 4 & & \text{if } t \geq 3 \text{ and } h_t(t-1) = 3 \\ & \text{or} & \text{if } t \geq 3 \text{ and } h_t(t-1) \in \{1,2\} \text{ and } h_t(t-2) = 3, \end{cases}$$

where $h_t(t-2)$ and $h_t(t-1)$ are the second-to-last and the last actions chosen under history h_t, respectively. The idea is that from period 3 onward, σ chooses the most likely location of the object.

CLAIM 1: When player 1 uses σ he guarantees himself strictly more than 1/2: $u_1(\sigma, \tau) > 1/2$ for every τ.

PROOF OF CLAIM 1: Under σ, player 1 looks at state 1 at period 1 and finds the object with probability q at period 1. If the object is not found, there is a positive probability that it is in state 2 at period 1, in which

case it moves with positive probability to state 3 or state 4 at period 2. Then player 1 looks at state 3 or state 4 at period 3, depending on the action of player 2 at period 2, and finds the object with probability strictly greater than $1/2 - q$ at period 3 no matter the action of player 2 at period 2. □

CLAIM 2: Suppose that player 1 follows a strategy σ that looks at state 3 or state 4 at period 1. Then player 2 has a strategy τ such that $u_1(\sigma, \tau) \leq 1/2$.

PROOF OF CLAIM 2: Let $\tau = (\tau_t)_{t \in \mathbb{N}^{\text{even}}}$ be the strategy of player 2 defined as follows. For all $t \in \mathbb{N}^{\text{even}}$, for all $h_t \in H_t$,

$$\tau_t(h_t) = \begin{cases} \text{state } 1 & \text{if } h_t \in \{1,2\}^{t-1}, \\ \text{state } 3 & \text{if } h_t(t-1) = 4 \\ \quad \text{or} & \text{if } h_t(t-1) \in \{1,2\} \text{ and } h_t(t-2) = 4, \\ \text{state } 4 & \text{if } h_t(t-1) = 3 \\ \quad \text{or} & \text{if } h_t(t-1) \in \{1,2\} \text{ and } h_t(t-2) = 3. \end{cases}$$

The idea is that τ looks at state 1 if player 1 has never played state 3 or state 4, and plays the most likely state otherwise. Assume for simplicity that player 1 looks at state 3 at period 1. Assume that player 1 does not find the object at period 1. The conditional probability for the object of being in state 4 at period 2 is then equal to

$$p_2(4) = \frac{1/2 - q}{1/2 + q} + 2 \cdot \frac{1 - \eta}{2} \cdot \frac{q}{1/2 + q} = \frac{1/2 - \eta \cdot q}{1/2 + q},$$

which is strictly higher than $1/2$ by our assumption that $q < 1/4$ and $\eta < 1/4$. Then, in the continuation of the game, player 2 guarantees strictly more than $1/2$ if he looks at state 4 at period 2. If he does not, player 2 will get strictly less than $1/2$ if player 1 looks at state 4 at period 3. For similar reasons, if period 3 is reached, it is better for player 1 to look at state 3. By repeating this argument, it is better for player 1 to always look at state 3 against τ.

At period 1, player 1 finds the object with probability $1/2 - q$. At period

2, player 2 finds the object with probability $1/2 - q + q(1-\eta)$. At period 3, player 1 finds the object with probability $q(1-\eta) + q(1-\eta)\eta$. At period 4, player 2 finds the object with probability $q(1-\eta)\eta + q(1-\eta)\eta^2$. And so on. Then, player 1 finds the object with probability

$$\frac{1}{2} - q + q(1-\eta) + q(1-\eta)\eta + q(1-\eta)\eta^2 + q(1-\eta)\eta^3 + \dots$$

$$= \frac{1}{2} - q + q(1-\eta)\frac{1}{1-\eta} = \frac{1}{2}.$$

So, by playing state 3 or state 4 at period 1, player 1 gets at most $1/2$ against τ. □

CLAIM 3: There is no 0-equilibrium.

PROOF OF CLAIM 3: Assume by way of contradiction that there is a 0-equilibrium (σ', τ'). From CLAIM 1 and CLAIM 2, player 1 chooses state 1 or state 2 with probability 1 at period 1. In both cases, at period 2 the current probability distribution is $p_2 = (q\eta, q\eta, 1/2 - q\eta, 1/2 - q\eta)$. Then, at period 2, the game is similar to the original one, with a parameter $q' = q\eta$ instead of q, which still satisfies $q' \in (0, 1/4)$, and where the roles of the players are exchanged. Then, as τ is a 0-best response, it follows from the previous reasoning that player 2 plays state 1 or state 2 with probability 1. By following this process recursively, players will choose states 1 and 2 with probability 1 forever. This leads to the payoff $\frac{4}{4-\eta^2}$ for player 1. Then, player 1 has an incentive to deviate from σ' and to choose state 3 at period 1 to get a payoff of at least $1/2 - q > \frac{4}{4-\eta^2}$, a contradiction. □

Example 4.3.2. We present another game with time-homogeneous Markov chain without a 0-equilibrium. Consider the game in Figure 4.2. Let $\eta \in (0, 1/6)$ and $q \in (0, 1/3)$. Let $p = (0, 0, 0, 0, 0, q(1-\eta), q\eta, \frac{1-q}{2}, \frac{1-q}{2})$ be the initial probability distribution. Notice that in this example there is no transient state.

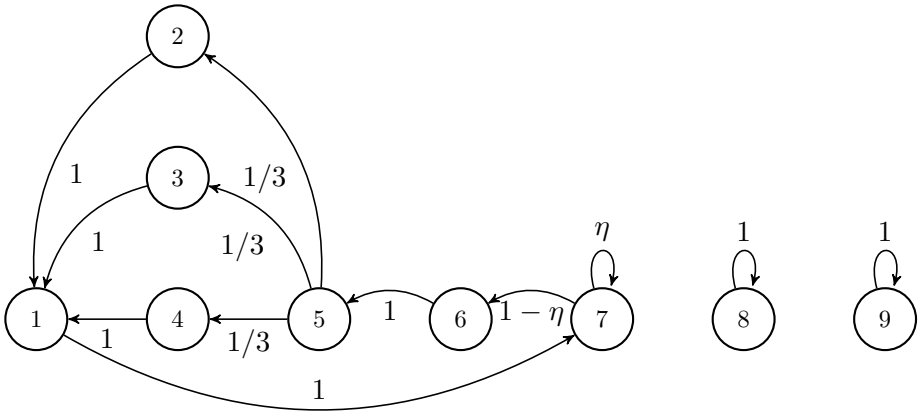


Figure 4.2: A game without a 0-equilibrium.

CLAIM 1: In any 0-equilibrium, at period 1 player 1 chooses state 6 with probability 1.

PROOF OF CLAIM 1. If at period 1 player 1 looks at state 6, he guarantees $q(1-\eta)+\frac{1-q}{2} > 1/2$ by looking at period 3 at state 8 or 9.

If at period 1 player 1 looks at state 1, 2, 3, 4, 5 or 7, then player 2 can find the object with probability $q(1-\eta)$ at period 2 by looking at state 5 and with probability $\frac{1-q}{2}$ at period 4 by looking at state 8 or 9. As $q(1-\eta)+\frac{1-q}{2} > 1/2$, player 1 cannot get more than $1/2$.

If at period 1 player 1 looks at state 8 (respectively, at state 9), then player 2 can guarantee $\frac{1-q}{2}$ by looking at state 9 (respectively, at state 8) at period 2 and then $q(1-\eta)\cdot\frac{2}{3}$ by looking at state 1 at period 4. As $\frac{1-q}{2}+q(1-\eta)\cdot\frac{2}{3} = \frac{1}{2}+q(1/6-\eta) > 1/2$ as $\eta < 1/6$, player 1 cannot get more than $1/2$.

So, there can be no 0-equilibrium in which at period 1 player 1 places a positive probability on a state different from state 6. □

CLAIM 2: In any 0-equilibrium, at period 2 player 2 chooses state 6 with probability 1.

PROOF OF CLAIM 2. From Claim 1, we know that in a 0-equilibrium, player 1 looks at state 6 at period 1 with probability 1. If he does so, he finds the object with probability $q(1-\eta)$ at period 1. Then, under the condition that the object is not found, the object was in state 7, 8 or 9 with probability 1 at period 1 and the updated probability distribution of the object is $\left(0,0,0,0,0,0,\frac{\eta q}{1-q(1-\eta)},\frac{1-q}{2[1-q(1-\eta)]},\frac{1-q}{2[1-q(1-\eta)]}\right)$. Then, the object follows the transition matrix and the probability distribution of the object at period 2 is

$$\left(0,0,0,0,0,0,\frac{\eta(1-\eta)q}{1-q(1-\eta)},\frac{\eta^2 q}{1-q(1-\eta)},\frac{1-q}{2[1-q(1-\eta)]},\frac{1-q}{2[1-q(1-\eta)]}\right)$$
$$=\left(0,0,0,0,0,0,q'(1-\eta),q'\eta,\frac{1-q'}{2},\frac{1-q'}{2}\right),$$

where $q'=\frac{\eta q}{1-q(1-\eta)}<\frac{q/6}{1-1/3(1-0)}=\frac{q}{4}<q<1/3$ as $0<\eta<1/6$ and $0<q<1/3$. Thus, at period 2 player 2 is facing a similar situation as player 1 at period 1. Claim 2 follows from Claim 1. □

CLAIM 3: This game has no 0-equilibrium.

PROOF OF CLAIM 3. Assume by way of contradiction that the game has a 0-equilibrium. From Claim 1, player 1 plays state 6 at period 1. From Claim 2, player 2 plays state 6 at period 2. By repeating the same reasoning as in Claim 2, in a 0-equilibrium, the active player looks at state 6 with probability 1 at each period. Under this strategy profile, the object is found with probability lower than $q<\frac{1-q}{2}$. Hence, it would be profitable for player 1 to deviate and look at state 6 at period 1. In conclusion, there is no 0-equilibrium. □

4.3.2 Existence of ε-equilibrium

In this subsection we are interested in the existence of ε-equilibrium, where $\varepsilon>0$. We show that there is an ε-equilibrium in pure strategies for every search game, and for each $\varepsilon>0$. The proof relies on existence results for ε-equilibria in games with Borel measurable payoff functions

(see the proof of Mertens and Neyman in [73]) and with lower semi-continous payoff functions (see [28] and [29]).

Theorem 4.3.2. *Each competitive search game admits an ε-equilibrium in pure strategies, for all $\varepsilon > 0$.*

Proof. Consider the Model [2] of a competitive search game in Section 4.2. Note that

1. this is a multiplayer perfect-information game,
2. from Proposition 4.8.2 it follows that the payoffs are bounded and lower semi-continuous.

Thus by applying Theorem 2.3 of [28] to the Model [2], the game admits an ε-equilibrium in pure strategies for every $\varepsilon > 0$. It also follows from [29] and [69]. □

Revisiting Examples 4.3.1 and 4.3.2. In view of Theorem 4.3.2, the game in Example 4.3.1 has an ε-equilibrium in pure strategies for every $\varepsilon > 0$. We now present an (subgame perfect) ε-equilibrium in pure strategies of this game, for all $\varepsilon > 0$.

Let $\varepsilon > 0$. The idea of the ε-equilibrium in pure strategies described here is to choose state 1 for a long time as long as the other player does the same, and then to choose the most likely between state 3 or state 4 in the remaining game. More formally, for each $n \in \mathbb{N}$, let (σ^n, τ^n) be the pure strategy profile defined as follows. For all $t \in \mathbb{N}$, for all history h_t at period t, for all $n \in \mathbb{N}$, let $f_t^n : H_t \to S$ be defined by

$$f_t^n(h_t) = \begin{cases} \text{state } 1 & \text{if } h_t \in \{1,2\}^{t-1} \text{ and } t < n, \\ \text{state } 3 & \text{if } h_t(t-1) = 4 \\ \quad \text{or} & \text{if } h_t(t-1) \in \{1,2\} \text{ and } h_t(t-2) = 4 \\ \quad \text{or} & \text{if } h_t \in \{1,2\}^{t-1} \text{ and } t \geq n, \\ \text{state } 4 & \text{if } h_t(t-1) = 3 \\ \quad \text{or} & \text{if } h_t(t-1) \in \{1,2\} \text{ and } h_t(t-2) = 3. \end{cases}$$

Then, we define $\sigma_t^n(h_t) = f_t^n(h_t)$ for all $t \in \mathbb{N}^{\text{odd}}$, and $\tau_t^n(h_t) = f_t^n(h_t)$ for all $t \in \mathbb{N}^{\text{even}}$ and all history h_t at time t. The idea of σ^n and τ^n is to look at state 1 until period n (if the other player does the same) and from period n onward (or before if the other player deviates) to look at the most likely state. We argue that if $n \geq \frac{\ln \eta q - \ln 4\varepsilon}{\ln 2 - \ln \eta}$ then (σ^n, τ^n) is an ε-equilibrium. For simplicity, we assume that n is odd.
It follows from the Claim 2 of the proof of Theorem 4.3.1 that τ^n is a 0-best-response against σ^n. It is then sufficient to show that σ^n is an ε-best response against τ^n when n is large enough. From Claim 2 of the proof of Theorem 4.3.1 it follows that a 0-best response against τ^n is to follow the strategy σ^{n+1}, which only differs from σ^n at period n. Under (σ^n, τ^n), player 1 finds the object at period 1 with probability q, player 2 finds the object at period 2 with probability $q.\left(\frac{\eta}{2}\right)$, player 1 finds the object at period 3 with probability $q.\left(\frac{\eta}{2}\right)^2$, and so on until period $n-1$ where player 2 finds the object with probability $q.\left(\frac{\eta}{2}\right)^{n-2}$. Then in the continuation game that starts at period n it follows from the proof of Claim 2 in Theorem 4.3.1 that both players find the object with probability 1/2. So, player 1 finds the object before period n with probability $q + \left(\frac{\eta}{2}\right)^2 \cdot q + \left(\frac{\eta}{2}\right)^4 \cdot q + \ldots + \left(\frac{\eta}{2}\right)^{n-3} \cdot q$, player 2 finds the object before period n with probability $q.\left(\frac{\eta}{2}\right) + \ldots + q.\left(\frac{\eta}{2}\right)^{n-2}$ and each player finds the object from period n with probability $\frac{1}{2} \cdot \left[1 - \left(q + \left(\frac{\eta}{2}\right) \cdot q + \left(\frac{\eta}{2}\right)^2 \cdot q + ... + \left(\frac{\eta}{2}\right)^{n-2} \cdot q\right)\right]$. This implies that under (σ_n, τ_n) the expected payoff of player 1 is

$$q + \left(\frac{\eta}{2}\right)^2 \cdot q + \left(\frac{\eta}{2}\right)^4 \cdot q + \ldots + \left(\frac{\eta}{2}\right)^{n-3} \cdot q$$

$$+ \left[1 - \left(q + \left(\frac{\eta}{2}\right) \cdot q + \left(\frac{\eta}{2}\right)^2 \cdot q + ... + \left(\frac{\eta}{2}\right)^{n-2} \cdot q\right)\right] \cdot \frac{1}{2}$$

and under $(\sigma^{n+1}, \tau^{n+1})$, the expected payoff of player 1 is

$$q + \left(\frac{\eta}{2}\right)^2 \cdot q + \left(\frac{\eta}{2}\right)^4 \cdot q + \ldots + \left(\frac{\eta}{2}\right)^{n-1} \cdot q$$

$$+\left[1-\left(q+\left(\frac{\eta}{2}\right)\cdot q+\left(\frac{\eta}{2}\right)^2\cdot q+...+\left(\frac{\eta}{2}\right)^{n-1}\cdot q\right)\right]\cdot\frac{1}{2}.$$

Those two terms converge to the same limit $\frac{q}{1-\left(\frac{\eta}{2}\right)^2}+\left[1-\frac{q}{1-\left(\frac{\eta}{2}\right)}\right]\cdot\frac{1}{2}$ which is the value of the game. Moreover, difference between these two expressions is

$$\left|\left(\frac{\eta}{2}\right)^{n-1}\cdot q-\frac{1}{2}\cdot\left(\frac{\eta}{2}\right)^{n-1}\cdot q\right|=\frac{1}{2}\cdot\left(\frac{\eta}{2}\right)^{n-1}\cdot q.$$

Hence, when $n\geq\frac{\ln\left(\frac{\eta q}{4\varepsilon}\right)}{\ln\left(\frac{2}{\eta}\right)}$, the difference between those two expressions is smaller than ε so (σ^n,τ^n) is an ε-equilibrium. □

With the same idea one can construct an ε-equilibrium in Example 4.3.2 where both players choose state 6 until for a long time and then switch to state 8 or 9.

4.4 Payoff properties under ε-equilibrium and existence of the value

Competitive search games are not constant-sum games, and the payoff functions are not continuous as mentioned in Proposition 4.8.2. Thus, the notion of value is not clear yet, and its existence is not immediate. We will first show that if a player chooses an ε-best response against the strategy of the other player, the payoffs almost add up to 1. Thus, the game is essentially constant-sum, so the notion of value becomes natural. Then, we show the existence of the value of these games, to finally prove existence of ε-optimal strategies for both players for all $\varepsilon>0$ and relate optimal strategies and equilibria.

Lemma 4.4.1. *Consider a strategy τ for player 2. Let $\varepsilon>0$. If the strategy σ of player 1 is an ε-best response against τ, then under (σ,τ) the object is*

found with probability at least $1 - \varepsilon \cdot |S|$*. In other words,*

$$u_1(\sigma, \tau) + u_2(\sigma, \tau) \geq 1 - \varepsilon \cdot |S|.$$

A similar statement holds with exchanged roles of the players.

Proof. Note that the sequence of events $([t < \Theta < +\infty])_{t \in \mathbb{N}}$ is decreasing and its limit is the empty set. Thus, $(\mathbb{P}_{(\sigma,\tau)}(t < \Theta < \infty))_t$ is decreasing and converges to 0 as t goes to ∞, by σ-additivity of probability measures.

Suppose that player 1 plays σ against τ. Then player 1 finds the object with probability $u_1(\sigma, \tau) = \mathbb{P}_{(\sigma,\tau)}(\Theta \in \mathbb{N}^{\text{odd}})$. Assume now that player 1 follows σ until a certain period $T \in \mathbb{N}$, and then deviates from σ by choosing a state uniformly from period $T + 2$ onward, and denote this strategy σ'. Then player 1 finds the object at period $T + 2$ with probability $(1 - \mathbb{P}_{(\sigma,\tau)}(\Theta \leq T + 1))/|S|$. Thus, $u_1(\sigma', \tau) \geq \mathbb{P}_{(\sigma,\tau)}(\Theta \in \mathbb{N}^{\text{odd}},\ \Theta \leq T) + (1 - \mathbb{P}_{(\sigma,\tau)}(\Theta \leq T + 1))/|S|$. As σ is an ε-best response against τ, it holds that

$$\mathbb{P}_{(\sigma,\tau)}(\Theta \in \mathbb{N}^{\text{odd}}) = u_1(\sigma, \tau) \geq u_1(\sigma', \tau) - \varepsilon = \mathbb{P}_{(\sigma',\tau)}(\Theta \in \mathbb{N}^{\text{odd}}) - \varepsilon.$$

So, since σ and σ' are identical for $\Theta \leq T + 1$ this implies

$$\begin{aligned}
&\mathbb{P}_{(\sigma,\tau)}(\Theta \in \mathbb{N}^{\text{odd}}, \Theta \geq T + 2) \\
&\geq \mathbb{P}_{(\sigma',\tau)}(\Theta \in \mathbb{N}^{\text{odd}}, \Theta \geq T + 2) - \varepsilon \\
&\geq (1 - \mathbb{P}_{(\sigma,\tau)}(\Theta \leq T + 1))/|S| - \varepsilon \\
&= \mathbb{P}_{(\sigma,\tau)}(\Theta \geq T + 2))/|S| - \varepsilon \\
&\geq \mathbb{P}_{(\sigma,\tau)}(\Theta \in \mathbb{N}^{\text{odd}}, \Theta \geq T + 2))/|S| + \mathbb{P}_{(\sigma,\tau)}(\Theta = \infty)/|S| - \varepsilon.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbb{P}_{(\sigma,\tau)}(\Theta = \infty) \\
&\leq \mathbb{P}_{(\sigma,\tau)}(\Theta \in \mathbb{N}^{\text{odd}}, \Theta \geq T + 2)) \cdot (|S| - 1) + \varepsilon \cdot |S|
\end{aligned}$$

$$\leq \mathbb{P}_{(\sigma,\tau)}(T+2 \leq \Theta < \infty)) \cdot (|S|-1) + \varepsilon \cdot |S|.$$

As $(\mathbb{P}_{(\sigma,\tau)}(T+2 \leq \Theta < \infty))_T$ converges to 0 when T goes to ∞, then $\mathbb{P}_{(\sigma,\tau)}(\Theta = \infty) \leq \varepsilon \cdot |S|$. Thus, $u_1(\sigma,\tau) + u_2(\sigma,\tau) = \mathbb{P}_{(\sigma,\tau)}(\Theta < \infty) = 1 - \mathbb{P}_{(\sigma,\tau)}(\Theta = \infty) \geq 1 - \varepsilon \cdot |S|$. □

We denote $v_1 = \sup_{\sigma\in\Sigma} \inf_{\tau\in\mathcal{T}} u_1(\sigma,\tau)$
and $v_2 = \sup_{\tau\in\mathcal{T}} \inf_{\sigma\in\Sigma} u_2(\sigma,\tau)$.

Proposition 4.4.2. *The following equalities hold:*

$$v_1 = \inf_{\tau\in\mathcal{T}} \sup_{\sigma\in\Sigma} u_1(\sigma,\tau), \tag{4.1}$$

$$v_2 = \sup_{\tau\in\mathcal{T}} \inf_{\sigma\in\Sigma} u_2(\sigma,\tau). \tag{4.2}$$

$$v_1 + v_2 = 1. \tag{4.3}$$

Proof. First we prove equality (4.1). In this equality, player 1 is maximizing $u_1(\sigma,\tau)$ and player 2 is minimizing the same expression. Note that $(\sigma,\tau) \mapsto u_1(\sigma,\tau)$ is bounded. Moreover, by Proposition 4.8.2, it is lower semi-continuous, and hence Borel measurable. Now, equality (4.1) follows from [69], [70] or Maitra and Sudderth (1998).

Equality (4.2) follows similarly.

We now show that $v_1 + v_2 \leq 1$. Let $\varepsilon > 0$ and let (σ,τ) be an ε-equilibrium. We have:

$$u_1(\sigma,\tau) \geq \sup_{\sigma'} u_1(\sigma',\tau) - \varepsilon \geq \inf_{\tau'} \sup_{\sigma'} u_1(\sigma',\tau) - \varepsilon = v_1 - \varepsilon.$$

Similarly, $u_2(\sigma,\tau) \geq v_2 - \varepsilon$. Then,

$$v_1 + v_2 \leq u_1(\sigma,\tau) + u_2(\sigma,\tau) + 2 \cdot \varepsilon \leq 1 + \varepsilon.$$

As $\varepsilon > 0$ is arbitrary, we get $v_1 + v_2 \leq 1$.

We now show that $v_1 + v_2 \geq 1$. Let $\varepsilon > 0$ and let (σ', τ) be a strategy profile where σ' is an ε best response against τ. Then by Lemma 4.4.1 we have $u_1(\sigma, \tau) \geq 1 - u_2(\sigma, \tau) - \varepsilon \cdot |S|$. Denote $B_\varepsilon^\tau \subseteq \Sigma$ the set of ε-best responses of player 1 against τ. We have

$$\begin{aligned}
v_1 &= \inf_{\tau \in \mathcal{T}} \sup_{\sigma \in \Sigma} u_1(\sigma, \tau) \\
&\geq \inf_{\tau \in \mathcal{T}} \sup_{\sigma \in B_\varepsilon^\tau} u_1(\sigma, \tau) \\
&\geq \inf_{\tau \in \mathcal{T}} \sup_{\sigma \in B_\varepsilon^\tau} [1 - u_2(\sigma, \tau) - \varepsilon \cdot |S|] \\
&= 1 - \sup_{\tau \in \mathcal{T}} \inf_{\sigma \in B_\varepsilon^\tau} u_2(\sigma, \tau) - \varepsilon \cdot |S| \\
&\geq 1 - \sup_{\tau \in \mathcal{T}} \inf_{\sigma \in \Sigma} u_2(\sigma, \tau) - \varepsilon \cdot |S| \\
&= 1 - v_2 - \varepsilon \cdot |S|.
\end{aligned}$$

As $\varepsilon > 0$ is arbitrary, we conclude that $v_1 + v_2 \geq 1$. □

The last theorem of this section shows that all ε-equilibria give almost the same payoffs, for small ε.

Theorem 4.4.3. *For each $\varepsilon \geq 0$, for each ε-equilibrium (σ, τ):*

1. *the object is found with probability at least $1 - \varepsilon \cdot |S|$,*
2. *$|u_1(\sigma, \tau) - v_1| \leq \varepsilon$ and $|u_2(\sigma, \tau) - v_2| \leq \varepsilon$, where v_1 and v_2 are characterised above Proposition 4.4.2.*

Proof.
1. It is a direct consequence from Lemma 4.4.1.

2. Let $\varepsilon \in (0, 1)$. Let (σ, τ) be an ε-equilibrium. As a consequence of Proposition 4.4.2,

$$u_1(\sigma, \tau) \geq \sup_{\sigma' \in \Sigma} u_1(\sigma', \tau) - \varepsilon \geq \sup_{\sigma' \in \Sigma} \inf_{\tau' \in T} u_1(\sigma', \tau') - \varepsilon = v_1 - \varepsilon.$$

Similarly, $u_2(\sigma, \tau) \geq v_2 - \varepsilon$. Thus

$$u_1(\sigma, \tau) \leq 1 - u_2(\sigma, \tau) \leq 1 - (v_2 - \varepsilon) = v_1 + \varepsilon.$$

Similarly, $u_2(\sigma, \tau) \leq v_2 + \varepsilon$. Those inequalities give [2]. □

A competitive search game is not a constant sum game in a strict sense. However, Proposition 4.4.2 and Theorem 4.4.3 show that, in essence, it has the same properties as a game in which the payoffs add up to 1 and thus the players have opposite interest. This leads to the following definition.

Definition 4.4.4. Consider a competitive search game, and let v_1 and v_2 be as above Proposition 4.4.2.

1. We call $v = v_1$ the value of the game.
2. For $\varepsilon \geq 0$, we say that $\sigma \in \Sigma$ is an ε-optimal strategy for player 1 if $u_1(\sigma, \tau) \geq v_1 - \varepsilon$ for every $\tau \in \mathcal{T}$. Similarly, we say that $\tau \in \mathcal{T}$ is an ε-optimal strategy for player 2 if $u_2(\sigma, \tau) \geq v_2 - \varepsilon$ for every $\sigma \in \Sigma$.

For ε-optimal strategies we obtain the following proposition.

Proposition 4.4.5. *Consider a competitive search game.*

1. *For all $\varepsilon \geq 0$, if (σ, τ) is an ε-equilibrium, then σ and τ are ε-optimal strategies.*
2. *For all $\varepsilon \geq 0$, if σ and τ are ε-optimal strategies, then (σ, τ) is a 2ε-equilibrium.*
3. *A strategy profile (σ, τ) is a 0-equilibrium if and only if σ and τ are 0-optimal strategies.*
4. *For all $\varepsilon > 0$, each player has a pure ε-optimal strategy.*

Proof.
1. Let (σ, τ) be an ε-equilibrium. Hence, $u_1(\sigma, \tau) \geq u_1(\sigma', \tau) - \varepsilon$ for all $\sigma' \in \Sigma$. Then, $u_1(\sigma, \tau) \geq v_1 - \varepsilon$, which means that σ is an ε-optimal strategy for player 1. Similarly, τ is an ε-optimal strategy for player 2.

2. Assume now that σ and τ are ε-optimal strategies for player 1 and player 2. Let $\sigma' \in \Sigma$. Then, $u_2(\sigma', \tau) \geq v_2 - \varepsilon$. By Proposition 4.4.2, we get that

$$u_1(\sigma', \tau) \leq 1 - u_2(\sigma', \tau) \leq 1 - (v_2 - \varepsilon) = v_1 + \varepsilon.$$

This implies that $u_1(\sigma, \tau) \geq v_1 - \varepsilon \geq u_1(\sigma', \tau) - 2\varepsilon$. Similarly, we obtain $u_2(\sigma, \tau) \geq u_2(\sigma, \tau') - 2\varepsilon$ for every $\tau' \in \mathcal{T}$ So, (σ, τ) is a 2ε-equilibrium.

3. This is a direct consequence of [1] and [2].

4. This is a consequence of [1] and Theorem 4.3.2. □

4.5 Additional results

4.5.1 Subgame optimal strategies

The notion of ε-optimal strategy is a relevant solution concept as it guarantees the maximal payoff against any strategy. However, this notion does not take into account eventual mistakes of the opponent. Hence, in this subsection we examine *subgame ε-optimal* strategies.

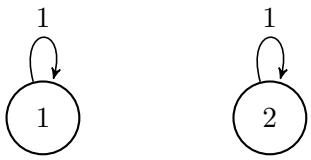
A strategy σ for player 1 is called subgame ε-optimal if, in each subgame, the continuation strategy of σ is ε-optimal. More precisely, for each history $h \in H^{\text{odd}}$ and strategy $\tau \in \mathcal{T}$ for player 2

$$u_1(\sigma, \tau)(h) \geq v_1(h) - \varepsilon.$$

The definition of a subgame ε-optimal strategy for payer 2 is similar.

Example 4.5.1. In this example, we show that there are ε-optimal strategies that are not subgame perfect ε-optimal strategies. The set of states

is $S = \{1, 2\}$, the transition matrix P is the identity over S and the initial probability distribution is $p = (1, 0)$.



The value of player 1 is $v_1 = 1$ and any optimal strategy of player 1 starts looking at state 1. Then, $v_2 = 0$ and all the strategies of player 2 are 0-optimal. In particular, it is optimal for player 2 to always choose state 2. Let τ denote this strategy.

Now suppose that player 1 makes a mistake and chooses state 2 at period 1. Then, the continuation strategy of τ from period 2 is not optimal. In fact, it would be the best for player 2 to choose state 1 at period 2 and win the game. □

Proposition 4.5.1. *Consider a competitive search game.*

1. *For every $\varepsilon > 0$, each player has a pure strategy which is subgame ε-optimal.*
2. *Let $\varepsilon \in (0, \frac{1}{|S|})$. If σ is a subgame ε-optimal strategy for player 1, then for every strategy τ of player 2, the object is found with probability 1 under the strategy profile (σ, τ). A similar statement holds for player 2.*

Proof. [1] Let $\varepsilon > 0$. In [28] and [29] it is shown that there exists a subgame perfect ε-equilibrium (σ, τ) in pure strategies. Now consider a subgame at a history h. Since the continuation strategies of σ and τ at h form an ε-equilibrium, it follows similarly to Proposition 4.4.5 that the continuation strategy of σ at h is ε-optimal in the subgame, and similarly the continuation strategy of τ at h is ε-optimal in the subgame. Hence, σ and τ are subgame ε-optimal.

[2] Let $\varepsilon \in (0, \frac{1}{|S|})$ and let σ be a subgame ε-optimal strategy. Consider a history h at an odd period. The strategy for player 1 which looks at a state with the highest probability guarantees $1/|S|$ in the subgame at h. So, $v(h) \geq 1/|S|$.

Now consider a strategy τ for player 2. Then, we have $u_1(\sigma, \tau)(h) \geq 1/|S| - \varepsilon > 0$. In particular, in the subgame at h, the object is found with probability at least $1/|S| - \varepsilon > 0$ under (σ, τ). Since this holds for every history h at an odd period, by Lévy's zero-one law, the object is found with probability 1 under (σ, τ). □

4.5.2 Structure of the optimal actions

In this subsection, we present some structural properties of the optimal actions. For all $s \in S$ and for all $p \in \Delta(S)$, we denote $v_1(p)$ the value of the game with initial probability distribution p, and $v_1(p, s)$ the expected payoff of player 1 if he chooses state s at period 1 when the initial distribution is p, assuming that both players will play optimally afterwards. In other words, $v_1(e^s) = v_1(e^s, s) = 1$ and for all $p \in \Delta(S) \backslash \{e^s\}$,

$$v_1(p, s) := p(s) + (1 - p(s)) \cdot (1 - v_1(p^{\neg s} P)) = 1 - (1 - p(s)) \cdot v_1(p^{\neg s} P),$$

where $p^{\neg s}$ is the probability distribution p conditional to the fact that the object is not in state s. In other words, $p^{\neg s}(s) = 0$ and $p^{\neg s}(j) = \frac{p(j)}{1-p(s)}$ for all $j \neq s$. Note that $v_1(p) = \max_{s \in S} v_1(p, s)$. We also denote for all $s \in S$ the set A_s of the probability distributions for which it is optimal for player 1 to look at state s at period 1. In other words, $A_s = \{p \in \Delta(S) \mid v_1(p, s) = v_1(p)\}$. Note that $\cup_{s \in S} A_s = \Delta(S)$.

Theorem 4.5.2. *The optimality regions A_s have the following properties.*

[1] If the initial probability p is sufficiently close to e^s, for some state s, then choosing state s is the only optimal action. That is, the region $A_s \backslash \cup_{j \neq s} A_j$ is a neighborhood of e^s in $\Delta(S)$.

[2] Looking at a state in which the object is with zero probability is never better than looking anywhere else. That is, for all states $s, s' \in S$, for all $p \in \Delta(S)$, if $p(s') = 0$ then $v_1(p, s') \leq v_1(p, s)$.

[3] For each subset $N \subseteq S$, the convex hull of the vertices e^s with $s \in N$ is included in the set $\cup_{s \in N} A_s$.

[4] There is an initial distribution at which choosing any state is optimal. That is, $\cap_{s \in S} A_s \neq \emptyset$.

[5] For all $s \in S$, the region A_s is star convex centered in e^s. That is, if $p \in A_s$ then the whole line segment between p and e^s is included in A_s.

Proof.

[1] The statement follows from the facts that each $v(p, s)$ is continuous (cf. Theorem 4.8.4) in p and that $v(e^s, s) = 1$ and $v(e^s, j) < 1$ for all $j \neq s$.

[2] Assume $p(s') = 0$ for some state $s' \in S$. Let $s \in S$. Let (σ, τ) be a strategy profile such that $\sigma_1(\emptyset) = s$ and σ and τ be Markov strategies : for each $t \in \mathbb{N}^{\text{odd}}$ (resp. $t \in \mathbb{N}^{\text{even}}$), σ_t (resp. τ_t) is constant over the set H_t. Let $\sigma' \in \Sigma$ be a Markov strategy of player 1 that starts looking at state s'. Let $p \in \Delta(S)$ and remark that $p = p(s) \cdot e^s + (1 - p(s)) \cdot p^{\neg s}$ for all $s \in S$, where $e^s \in \Delta(S)$ is the vector with $e^s(s) = 1$ and $e^s(j) = 0$ for all $j \neq s$. We have :

$$\begin{aligned} u_1(\sigma', \tau)(p) &= p(s) \cdot u_1(\sigma', \tau)(e^s) + (1 - p(s)) \cdot u_1(\sigma', \tau)(p^{\neg s}) \\ &\leq p(s) \cdot 1 + (1 - p(s)) \cdot u_1(\sigma', \tau)(p^{\neg s}) \\ &= p(s) \cdot 1 + (1 - p(s)) \cdot u_1(\sigma, \tau)(p^{\neg s}) \\ &= p(s) \cdot u_1(\sigma, \tau)(e^s) + (1 - p(s)) \cdot u_1(\sigma, \tau)(p^{\neg s}) \\ &= u_1(\sigma, \tau)(p). \end{aligned}$$

where the first equality comes from the linearity of the payoff function in respect of p (see Section 4.8.3), the first inequality comes from the fact that the payoffs are bounded from above by 1, the second equality

comes from the fact $[p^{\neg s}](s) = [p^{\neg s}](s') = 0$ and that the game played will be the same as σ, σ' and τ are not behavioral, the third equality comes from $u_1(\sigma, \tau)(e^s) = 1$ as $\sigma_1(\emptyset) = s$, and the fourth equality comes from the linearity of the payoff in respect of p. Taking the supremum over σ and the infimum over τ on both sides, we get $v_1(p, s') \leq v_1(p, s)$. □

[3] Let $p \in conv(\{e^s | s \in N\})$. Then $p(s) = 0$ for all $s \notin N$. By [2], there is an optimal action $j \in N$, and hence $p \in \cup_{s \in N} A_s$.

[4] We will use the Knaster-Kuratowski-Mazurkiewicz (KKM) theorem[4], see [61]. Note that by Theorem 4.8.4, the function $p \mapsto v(p, s)$ is continuous for all $s \in S$. Thus, each region A_s is closed. From this fact and from [3], we can apply the KKM theorem. We conclude from the KKM Theorem that $\cap_{s \in S} A_s \neq \emptyset$.

[5] Let $s \in S$, let $p \in A_s$ and let $\lambda \in [0, 1]$. We want to show that $\lambda e^s + (1 - \lambda)p \in A_s$. Let (σ, τ) be a strategy profile. By equation (4.6)

$$\begin{aligned}
&\sup_\sigma u_1(\sigma, \tau)(\lambda e^s + (1 - \lambda)p) \\
&= \sup_\sigma \left[\lambda \cdot u_1(\sigma, \tau)(e^s) + (1 - \lambda) \cdot u_1(\sigma, \tau)(p)\right] \\
&\leq \lambda \cdot \left[\sup_\sigma u_1(\sigma, \tau)(e^s)\right] + (1 - \lambda) \cdot \left[\sup_\sigma u_1(\sigma, \tau)(p)\right] \\
&= \lambda + (1 - \lambda) \cdot \left[\sup_\sigma u_1(\sigma, \tau)(p)\right],
\end{aligned}$$

where we used that $u_1(\sigma, \tau)(e^s) = 1$ for any strategy σ that looks at state s at period 1. Hence

$$v(\lambda e^s + (1 - \lambda)p) = \inf_\tau \sup_\sigma u_1(\sigma, \tau)(\lambda e^s + (1 - \lambda)p)$$

[4] The KKM theorem states: Let $n \in \mathbb{N}$ be the cardinal of the set of states S, in other words $|S| = n$. Let Δ^n be the simplex in $\mathbb{R}^n$. A KKM covering is defined as a collection $C_1, \ldots, C_n$ of closed sets such that for any $N \subseteq \{1, \ldots, n\}$, the convex hull of the vertices corresponding to N is covered by $\cup_{s \in N} C_s$. Then any KKM covering has a non-empty intersection, i.e.: $\cap_{s \in S} C_s \neq \emptyset$.

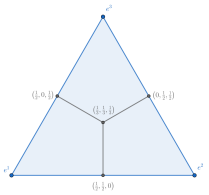


Figure 4.3: $P = I_3$

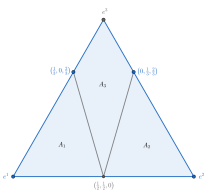


Figure 4.4: $P = Q$

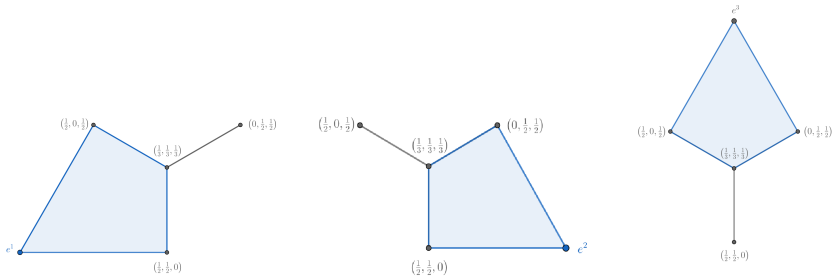


Figure 4.5: Regions A_1, A_2 and A_3 when $P = I_3$

$$\leq \lambda + (1-\lambda) \cdot \left[\inf_{\tau} \sup_{\sigma} u_1(\sigma, \tau)(p) \right]$$
$$= \lambda + (1-\lambda) \cdot v(\sigma, \tau).$$

On the other hand, by theorem 4.8.3, $v_1(\lambda e^s + (1-\lambda)p, s) = \lambda + (1-\lambda) \cdot v_1(p, s)$. So, choosing s when the initial probability distribution is $\lambda e^s + (1-\lambda)p$ is optimal. □

Example 4.5.2. Consider the case in which the set of states is $S = \{1, 2, 3\}$. Let $Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$. The sets A_1, A_2 and A_3 are represented in the time-homogeneous case where the transition matrix is the identity matrix in Figure 4.3 and in Figure 4.5, and the matrix Q in Figure 4.4.

Example 4.5.2 illustrates the statements of Theorem 4.5.2. It particular

here are some remarks.

- It makes intuitive sense that if the object is in a certain state with probability close to 1, then it is optimal to look at this state. Geometrically, this means that for all states $s \in S$, the set A_s contains a neighborhood of e^s in $\Delta(S)$.
- Looking at a state s' such that $p(s') = 0$ can still be (weakly) optimal. For example, in Figure 4.3 with initial probability distribution $p = (1/2, 1/2, 0)$, looking at state 3 is just as good as looking at either state 1 or state 2.
- Figure 4.5 illustrates that the intersection of the regions A_i can be more than a single point.
- Figure 4.5 illustrates the sets A_s are not always convex. However we conjecture that their relative interior is convex, in which case the closure of the relative interior of the sets A_s are polytopes.

4.5.3 Time-homogeneous Markov chains

In this subsection, we consider time-homogeneous competitive search games. A game is time-homogeneous when the transition matrix P_t at each period is the same. In this case, we will denote the transition matrix at each period by P. For all $r \in \mathbb{N}$ we denote P^r, the matrix P applied r times.

Recall that a transition matrix P is *irreducible* if for each entry (i, j), there exists $r \in \mathbb{N}$ such that the entry (i, j) of P^r is positive. A transition matrix P is *periodic* of period $r \geq 2$ if for all $k \in \mathbb{N}$, $P^k(x, x) > 0$ only if $k = r \cdot l$ for some $l \in \mathbb{N}$. If P is not periodic, we say that P is *aperiodic*. A subset $S' \subseteq S$ is *ergodic* if for $(i, j) \in S' \times (S \backslash S')$, $P(i, j) = 0$ and the transition matrix P restricted to the set S' is irreducible. A state $i \in S$ is called *absorbing* if $P(i, i) = 1$. A state $i \in S$ is *transient* if $\lim_{r \to \infty} P^r(i, i) = 0$.

A probability distribution $\pi \in \Delta(S)$ over the set S is called a *stationary distribution* for the transition matrix P if $\pi P = \pi$.

It is known that (see [63], Corollary 1.17 and Theorem 4.9) if the transition matrix P is irreducible, then there exists a unique stationary distribution $\pi \in \Delta(S)$. If P is also aperiodic, then there exist constants $\beta \in (0,1)$ and $c > 0$ such that for all $t \in \mathbb{N}$,

$$||pP^t - \pi||_{TV} \leq c \cdot \beta^t,$$

where $||p-q||_{TV} = \max_{A \subset S} \sum_{s \in A}(p(s) - q(s))$ is the total variation distance over $\Delta(S)$.

First we show that if the initial probability distribution is exactly an invariant distribution of the transition matrix P, then player 1 has a weak advantage.

Proposition 4.5.3. *Consider a time-homogenenous competitive search game. If π is an invariant distribution of P, then $v(\pi) \geq 1/2$.*

Proof. Assume first that there is a state $s \in S$ for which $\pi(s) = 0$. Then $\pi^{\neg s} P = \pi P = \pi$. Since $\pi(s) = 0$ we have $v(\pi, s) = 1 - v(\pi)$. As $v(\pi) \geq v(\pi, s)$, we obtain $v(\pi) \geq 1 - v(\pi)$. Hence, $v(\pi) \geq 1/2$.

Assume there is no state $s \in S$ for which $p(s) = 0$. Consider the game G' that arises by adding a state w to G. More precisely, G' is the game with set of states $S' = S \cup \{w\}$, initial probability distribution p' which places the same probabilities on states in S and probability zero on state w, and transition matrix P' that has the same transition probabilities between states in S and makes w absorbing. Then, the object will never be in w with probability 1. From Step 1 of the proof of [2] in Theorem 4.5.2, the players may ignore state w during the game. Let π' be the distribution on S' that coincides with π on S and $\pi'(w) = 0$. Then, π' is an invariant distribution of P', and hence by the first part we find $v(\pi) = v'(\pi') \geq 1/2$. □

Remarks. We conjecture that if P is irreducible and aperiodic, then $v_1(\pi) > 1/2$.

The value $v_1(p)$ can be smaller than $1/2$ if p is not the invariant distribution. Indeed, for example with three states, initial probability distribution $p = (1/3, 1/3, 1/3)$ and a transition matrix P such that at the second period the object is in state 1 with probability 1.

Theorem 4.5.4. *Consider a time-homogeneous competitive search game. Assume that the transition matrix P is irreducible and aperiodic. Then, no matter the initial probability distribution p, every strategy profile finds the object with probability 1. Hence, the payoff functions are continuous in this game, and there exists a 0-equilibrium in pure strategies.*

Proof. As mentioned, the transition matrix P has a unique stationary distribution $\pi \in \Delta(S)$ and $\pi(s) > 0$ for all $s \in S$. Moreover, there exist constants $c > 0$ and $\beta \in (0, 1)$ such that $|pP^t(s) - \pi(s)| \leq c \cdot \beta^t$ for all $t \in \mathbb{N}$, for all $s \in S$ and for all $p \in \Delta(S)$. Hence, there exists $t^* \in \mathbb{N}$ with the following property: for all $p \in \Delta(S)$, for all $s \in S$, for all $t \geq t^*$, we have $(pP^t)(s) > \frac{\delta}{2}$, where $\delta = \min_{s \in S} \pi(s)$. Without loss of generality we can assume that $t^* \geq 2$.

Let $\alpha = \frac{\delta}{4(t^*-1)}$. The proof is divided into four steps.

STEP 1: Let (σ, τ) be a pure strategy profile, and let $(s_t)_{t \in \mathbb{N}}$ denote the induced sequence of actions. We show that the object is found during the first t^* periods with probability at least α.

PROOF: For each $t \in \mathbb{N}$, let $p_t = (p_t(s))_{s \in S} \in \Delta(S)$ denote the probability distribution of the location of the object at period t, conditional on not being found through the history $(s_1, \ldots, s_{t-1})$.

If there is a period $t \leq t^*$ such that $p_t(s_t) \geq \alpha$, then under (σ, τ), the object is found at period t with probability at least α, if it has not been found before. Hence, the claim of step 1 is true.

Therefore, it suffices to show that if at each period $t \leq t^* - 1$ we have $p_t(s_t) < \alpha$, then $p_{t^*}(s_{t^*}) \geq \alpha$. So assume that at each period $t \leq t^* - 1$ we have $p_t(s_t) < \alpha$. The idea of the calculation below is that, since the object is found with low probabilities at the first $t^* - 1$ periods, the probability distribution for the object at period t^* on condition that it is not found during the first $t^* - 1$ periods is almost the same as the unconditioned probability distribution. That is, p_{t^*} is close to pP^{t^*-1}, which is in turn close to the invariant distribution.

Note that, if the players do not condition on the past, the probability distribution of the location of the object at period t^* is simply pP^{t^*-1}. We have

$$\begin{aligned}
||p_{t^*} - pP^{t^*-1}||_{TV} &\leq ||p_{t^*} - p_{t^*-1}P||_{TV} + ||p_{t^*-1}P - pP^{t^*-1}||_{TV} \\
&= ||p_{t^*-1}^{\neg s_{t^*-1}}P - p_{t^*-1}P||_{TV} + ||p_{t^*-1}P - pP^{t^*-1}||_{TV} \\
&\leq ||p_{t^*-1}^{\neg s_{t^*-1}} - p_{t^*-1}||_{TV} + ||p_{t^*-1} - pP^{t^*-2}||_{TV} \\
&= p_{t^*-1}(s_{t^*-1}) + ||p_{t^*-1} - pP^{t^*-2}||_{TV} \\
&< \alpha + ||p_{t^*-1} - pP^{t^*-2}||_{TV} \\
&< \alpha \cdot (t^* - 1) + ||p_1 - pP^0||_{TV} \\
&= \alpha \cdot (t^* - 1) \\
&= \frac{\delta}{4}.
\end{aligned}$$

Here, in the first inequality we used the triangle inequality. In the first equality, we used that $p_{t^*} = p_{t^*-1}^{\neg s_{t^*-1}}P$, as $p_{t^*-1}^{\neg s_{t^*-1}}$ is the probability distribution of the location of the object at period $t^* - 1$ conditional on the fact that the object has not been found before period $t^* - 1$ and that it is not in state s_{t^*-1} at period $t^* - 1$ after the history $(s_1, \ldots, s_{t^*-2})$ and not being in state s_{t^*-1} at period $t^* - 1$. The second inequality is true as $||qP - q'P||_{TV} \leq ||q - q'||_{TV}$ for all $q, q' \in \Delta(S)$. The second equality follows from the above interpretation of $p_{t^*-1}^{\neg s_{t^*-1}}$ and of the total variation norm. The third inequality is due to the assumption that at each period $t \leq t^* - 1$ we have $p_t(s_t) < \alpha$. The fourth inequality then follows by induction. The last two equalities are due to $p_1 = p$ and the

choice of α.

Therefore,

$$p_{t^*}(s_{t^*}) \geq (pP^{t^*-1})(s_{t^*}) - ||p_{t^*} - pP^{t^*-1}||_{TV} \geq \frac{\delta}{2} - \frac{\delta}{4} = \frac{\delta}{4} \geq \alpha.$$

This completes the proof of Step 1.

STEP 2: Consider any strategy profile (σ, τ). We show that the object is found during the first t^* periods with probability at least α.

PROOF: On the finite horizon t^*, each strategy can be equivalently represented as a mixed strategy, i.e. a probability distribution on the finite set of pure strategies on horizon t^* (see for example [72]). Hence, Step 2 follows from Step 1.

STEP 3: Consider any strategy profile (σ, τ). We show that the object is found with probability 1 under (σ, τ). By Proposition 4.8.2, this will imply that the payoff functions are continuous in this game.

PROOF: By Step 2, the object is found during the first t^* periods with probability at least α. Since t^* and therefore α do not depend on the initial distribution of the object, if the object is not found in the first t^* periods, then it will be found between periods $t^* + 1$ and $2t^*$ with probability at least α. By repeating this argument, the object is found with probability 1 under (σ, τ).

STEP 4: We show that there exists a 0-equilibrium in pure strategies.[5]

PROOF: In view of Theorem 4.3.2, for each $n \in \mathbb{N}$, there exists a $\frac{1}{n}$-equilibrium (σ^n, τ^n) in pure strategies. Since Σ and $\mathcal{T}$ are compact and metrizable, by taking a subsequence if necessary, we can assume that the sequence $(\sigma^n, \tau^n)_{n\in\mathbb{N}}$ converges to a strategy profile (σ, τ) in pure strategies as $n \to \infty$.

[5] By Step 3, the payoffs in the game are continuous. Since perfect info, it follows from [33] en [49] that there even exists a subgamne perfect 0-eq in pure strategies.

For each $n \in \mathbb{N}$, we have $u_1(\sigma^n, \tau^n) \geq u_1(\sigma', \tau^n) - \frac{1}{n}$ and $u_2(\sigma^n, \tau^n) \geq u_2(\sigma^n, \tau') - \frac{1}{n}$ for all $\sigma' \in \Sigma$ and $\tau' \in \mathcal{T}$. Since by Step 3 the payoff functions u_1 and u_2 are continuous, by taking the limits as $n \to \infty$, we obtain $u_1(\sigma, \tau) \geq u_1(\sigma', \tau)$ and $u_2(\sigma, \tau) \geq u_2(\sigma, \tau')$ for all $\sigma' \in \Sigma$ and $\tau' \in \mathcal{T}$. Hence, (σ, τ) is a 0-equilibrium in pure strategies. □

Remark 3. Consider a time-homogeneous search game. If this game does not satisfy the condition of Theorem 4.5.4, i.e. the transition matrix is not irreducible or not aperiodic, then the conclusion of Theorem 4.5.4 is no longer true, and there is even an initial probability distribution of the object and a strategy profile under which the object is found with probability zero. Indeed, if the transition matrix is not irreducible or not aperiodic, we distinguish the following three (not exclusive) situations: (i) If there is a transient state, then consider an initial probability distribution which places probability zero on every transient state and a strategy profile which always chooses a transient state. (ii) If there is more than 1 ergodic class, then consider an initial probability distribution which places probability 1 on an ergodic class and a strategy profile which always chooses a state in another ergodic class. (iii) If there is a periodic ergodic class, then consider an initial probability distribution which places probability 1 on a state. Then due to periodicity, at each period there is a state where the object is with probability zero (see Exercise 1.6 of [63]). So consider a strategy profile which always chooses such a state.

4.6 Variations

In this section we study two related versions of the search game: first where the horizon of the game is finite, and second through discounting when the players want to find the object as soon as possible. As we will see, the ε-optimal strategies of the original model are robust, in the sense that they are 2ε-optimal if the horizon of the game is finite but sufficiently long, and they are also 2ε-optimal in the discounted version of the game, provided that the discount factor is close to 1. Similarly,

each strategy that is optimal on a finite but sufficiently long horizon or for a high discount factor is also ε-optimal in the original search game. In particular, as the optimal strategies over the finite horizon games can be calculated easily, we obtain ε-optimal strategies in the original search game that are easy to calculate and to implement.

4.6.1 The finite horizon version of the search game

Suppose that the game ends at a specific period $T \in \mathbb{N}$, if it has not ended before. For simplicity, we will focus on player 1. Let

$$u_{1,T}(\sigma,\tau) = \mathbb{P}_{(\sigma,\tau)}(\Theta \in \mathbb{N}^{\text{odd}},\ \Theta \leq T)$$

denote the probability that player 1 finds the object within the T first periods under (σ,τ). We assume that player 1 is maximizing $u_{1,T}$ whereas player 2 is minimizing $u_{1,T}$. This is a zero-sum game which has value

$$v_{1,T} := \max_{\sigma}\min_{\tau} u_{1,T}(\sigma,\tau) = \min_{\tau}\max_{\sigma} u_{1,T}(\sigma,\tau).$$

Note that, with exchanged roles of the players, we could also define $v_{2,T}$. However, since the game has finite horizon, it may have a positive probability under each strategy profile that the object is not found, so it will not always be true that $v_{1,T} + v_{2,T} = 1$; in contrasts with Proposition 4.4.2 for the infinite horizon.

An advantage of the finite point of view is that unlike the infinite horizon, the value in finite horizon can be computed explicitly via the following dynamic programming equations:

$$v_{1,1}(p) = v_{1,2}(p) = ||p||_\infty,$$
$$v_{1,T}(p) = \max_{s_1}\min_{s_2} p(s_1) + (1 - p(s_1)(1 - [p^{\neg s_1}P_1](s_2)) \cdot v_{1,T-2}([p^{\neg s_1}P_1]^{\neg s_2}P_2).$$

As we mentioned in the beginning of this section, the finite horizon search game is strongly related to the original search game.

Definition 4.6.1. Let $\alpha \in (0,1)$. A transition matrix P is *α-strongly mixed* if for all $(i,j) \in S^2$, $P(i,j) \geq \alpha$.

Theorem 4.6.2. *Consider a competitive search game.*
[1] Let $\varepsilon > 0$. Let $\sigma^* \in \Sigma$ be an ε-optimal strategy for player 1 in the original search game, and for all $T \in \mathbb{N}$, let σ_T^* be a strategy for player 1 such that $u_{1,T}(\sigma_T^*, \tau) \geq v_{1,T}$ for each strategy τ of player 2. Then, there exists $\widetilde{T} \in \mathbb{N}$ such that for all $T \geq \widetilde{T}$, for all strategy $\tau \in \mathcal{T}$,

$$u_{1,T}(\sigma^*, \tau) \geq v_1 - 2\varepsilon \geq v_{1,T} - 2\varepsilon. \tag{4.4}$$

Consequently, $v_{1,T}$ converges to v_1 as T goes to ∞ and for all $T \geq \widetilde{T}$,

$$u_{1,T}(\sigma_T^*, \tau) \geq v_1 - \varepsilon \geq v_{1,T} - \varepsilon. \tag{4.5}$$

[2] If there exists a real number $\alpha \in (0,1)$ such that for all $T \in \mathbb{N}$ the transition matrix P_T at period T is α-strongly mixed, then for all $T \in \mathbb{N}$

$$v_1 \geq v_{1,T} \geq v_1 - (1-\alpha)^{T-1}.$$

[3] Analogous statements hold for player 2.

Proof.

PROOF OF [1]. The second inequality in (4.4) and the second inequality in (4.5) are trivial. We now prove that for large T the first inequality of (4.4) holds. Assume by way of contradiction that for every $\tilde{T} \in \mathbb{N}$, there exists $T \geq \tilde{T}$ and there is a strategy τ_T such that $u_{1,T}(\sigma^*, \tau_T) < v_1 - 2\varepsilon$. Since the set of strategies $\mathcal{T}$ for player 2 is compact, by taking a subsequence if necessary, we can assume that τ_T converges to some strategy τ as $T \to \infty$. Note that for every $T' \leq T$ we have

$$u_{1,T'}(\sigma^*, \tau_T) \leq u_{1,T}(\sigma, \tau_T) < v_1 - 2\varepsilon.$$

By taking the limit for $T \to \infty$, we find $u_{1,T'}(\sigma^*, \tau) \leq v_1 - 2\varepsilon$. Since this holds for all T', when taking the limit for $T' \to \infty$, we obtain $u_1(\sigma^*, \tau) \leq v_1 - 2\varepsilon < v_1 - \varepsilon$. This is a contradiction with the choice of σ^*. Thus, the inequality (4.4) holds.

As a consequence,

$$v_{1,T} = \max_{\sigma} \min_{\tau} u_{1,T}(\sigma, \tau) \geq \min_{\tau} u_{1,T}(\sigma^*, \tau) \geq v_1 - \varepsilon/2$$

for T large enough, where the last inequality comes from (4.4).

From this it follows that for all $\varepsilon > 0$, there exists a $\widetilde{T} \in \mathbb{N}$ such that for all $T \geq \widetilde{T}$, $u_{1,T}(\sigma_T^*, \tau) \geq v_{1,T} \geq v_1 - \varepsilon$. Hence, for large T the inequality (4.5) holds as well.

PROOF OF [2]. The first inequality is trivial. Assume that there exists a real number $\alpha \in (0, 1)$ such that for all $T \in \mathbb{N}$ the transition matrix P_T at period T is α-strongly mixed. Let $T \in \mathbb{N}$. We use the following notations:

- σ_T^* is an optimal strategy for player 1 in the zero-sum game with payoffs $(u_{1,T}, -u_{1,T})$,
- σ_T^- an optimal strategy for player 1 in the zero-sum game with payoffs $(-u_{2,T}, u_{2,T})$,
- τ_T^* an optimal strategy for player 2 in the zero-sum game with payoffs $(-u_{2,T}, u_{2,T})$,
- τ_T^- an optimal strategy for player 2 in the zero-sum game with payoffs $(u_{1,T}, -u_{1,T})$.

Let (σ, τ) be a strategy profile. We have:

$$\begin{aligned}
&u_{1,T}(\sigma, \tau) + u_{2,T}(\sigma, \tau) \\
=&u_{1,T-1}(\sigma, \tau) + u_{2,T-1}(\sigma, \tau) + \mathbb{P}_{(\sigma,\tau)}(\Theta = T) \\
\geq&u_{1,T-1}(\sigma, \tau) + u_{2,T-1}(\sigma, \tau) + [1 - u_{1,T-1}(\sigma, \tau) - u_{2,T-1}(\sigma, \tau)] \cdot \alpha \\
=&(1 - \alpha) \cdot [u_{1,T-1}(\sigma, \tau) + u_{2,T-1}(\sigma, \tau)] + \alpha.
\end{aligned}$$

Then,

$$u_{1,T}(\sigma,\tau)+u_{2,T}(\sigma,\tau)-1\geq(1-\alpha)\cdot\left[u_{1,T-1}(\sigma,\tau)+u_{2,T-1}(\sigma,\tau)-1\right],$$

which implies by induction

$$\begin{aligned}u_{1,T}(\sigma,\tau)+u_{2,T}(\sigma,\tau)-1&\geq(1-\alpha)^{T-1}\cdot[u_{1,1}(\sigma,\tau)+u_{2,1}(\sigma,\tau)-1]\\&=(1-\alpha)^{T-1}\cdot[p(\sigma(\emptyset))-1].\end{aligned}$$

Thus,

$$\begin{aligned}&u_{1,T}(\sigma,\tau)+u_{2,T}(\sigma,\tau)\\&\geq 1-(1-\alpha)^{T-1}\cdot[1-p(\sigma(\emptyset))]\geq 1-(1-\alpha)^{T-1}.\end{aligned}$$

In particular,

$$\begin{aligned}&u_{1,T}(\sigma_T^*,\tau_T^-)+u_{2,T}(\sigma_T^-,\tau_T^*)\\&\geq u_{1,T}(\sigma_T^-,\tau_T^-)+u_{2,T}(\sigma_T^-,\tau_T^-)\\&\geq 1-(1-\alpha)^{T-1}\\&=v_1+v_2-(1-\alpha)^{T-1}\end{aligned}$$

As $u_{2,T}(\sigma_T^-,\tau_T^*)=v_{2,T}\leq v_2$, it implies

$$v_{1,T}\geq u_{1,T}(\sigma_T^*,\tau_T^-)\geq v_1-(1-\alpha)^{T-1}.$$

□

4.6.2 The discounted version of the search game

Now we examine the discounted optimal strategies, once again with focus on player 1. For a discount factor $\beta\in(0,1)$ and strategy pair (σ,τ), let

$$u_{1,\beta}(\sigma,\tau)=\sum_{t\in\mathbb{N}^{\text{odd}}}\beta^{t-1}\mathbb{P}_{\sigma,\tau}(\Theta=t),$$

which is the expected discounted time that player 1 finds the object, not counting the instances where the object is not found. We assume that player 1 is maximizing $u_{1,\beta}$ whereas player 2 is minimizing $u_{1,\beta}$. This is a zero-sum game. The reason why we study this game which creates asymmetry between the players is because in the Theorem 4.6.3 we will use this lower bound on the payoff player 1 guarantees to show that player 1 when he plays an optimal best response in the β-discounted game, the payoff he guarantees converges to the value in the original game when β goes to 1. Let $v_{1,\beta}$ denote the value of the β-discounted game, and let σ_β denote a pure optimal [6] strategy of player 1. Note that the value and such a strategy σ_β exist, because the discounted payoff is continuous (cf. for example Fudenberg and Levine (1983)). With exchanged roles of the players, we can also define $v_{2,\beta}$, and due to discounting we do not have $v_{1,\beta} + v_{2,\beta} = 1$ for all $\beta \in (0,1)$ except in the case when there is a state $s \in S$ such that $p = e^s$.

As we mentioned in the beginning of this section, the discounted search game is strongly related to the original search game.

Theorem 4.6.3. *Consider a competitive search game.*
[1] Let $\varepsilon > 0$. Let $\sigma \in \Sigma$ be an ε-optimal strategy for player 1, and for all $\beta \in (0,1)$, let σ^*_β be a strategy for player 1 such that $u_{1,\beta}(\sigma^*_\beta, \tau) \geq v_{1,\beta}$ for each strategy τ of player 2. Then, there exists $\widetilde{\beta} \in (0,1)$ such that for all $\beta \in (\widetilde{\beta}, 1)$, for all strategies $\tau \in \mathcal{T}$,

$$u_{1,\beta}(\sigma, \tau) \geq v_1 - 2\varepsilon.$$

Consequently, $v_{1,\beta} \to v_1$ as $\beta \to 1$ and for all $\beta \in (\widetilde{\beta}, 1)$,

$$u_1(\sigma^*_\beta, \tau) \geq v_1 - \varepsilon.$$

[6] In discounted games, one usually considers stationary strategies. In our model, the natural state space would be the set $\Delta(S)$ of possible probability distributions for the location of the object (often called the belief space, as the players only have a belief where the object could be). Since this space is infinite, and states are often only visited once, we omit the detailed discussion of stationarity.

[2] Analogous statements hold for player 2.

Proof.

PROOF OF [1]. For every $T \in \mathbb{N}$ let $\delta(T) \in (0,1)$ such that $(\delta(T))^{T-1} \geq 1 - \frac{1}{T^2}$. Then, for every $\beta \in [\delta(T), 1)$ and every strategy profile (σ, τ)

$$u_{1,\beta}(\sigma,\tau) \geq \sum_{\substack{t=\mathbb{N}^{\text{odd}} \\ t\leq T}} \beta^{t-1} \cdot \mathbb{P}_{(\sigma,\tau)}(\Theta = t) \geq \sum_{\substack{t=\mathbb{N}^{\text{odd}} \\ t\leq T}} \left(1 - \tfrac{1}{T^2}\right) \cdot \mathbb{P}_{(\sigma,\tau)}(\Theta = t)$$
$$\geq u_{1,T}(\sigma,\tau) - \tfrac{1}{T}.$$

Hence, for all $\varepsilon > 0$, for all $T > \frac{1}{\varepsilon}$, the statements of the theorem follow from Theorem 4.6.2. The proof of [2] is similar. □

4.7 Concluding remarks and future work

We introduced an infinite horizon search game, in which two players compete to find an object that moves according to a time-varying Markov chain. We prove that these games always admit an ε-equilibrium in pure strategies, for all error-terms $\varepsilon > 0$, but not necessarily a 0-equilibrium. We showed that the ε-equilibrium payoffs converge to a singleton $(v, 1-v)$ as ε vanishes, and therefore the game is essentially a zero-sum game with value v. We examined the functional and structural properties of the solutions, and demonstrated that they are robust to having a finite but long horizon and respectively to having a sufficiently large discount factor. We devoted attention to the important special case when the Markov chain is time-homogeneous, where stronger results hold.

It would be interesting to generalize the results when the active player is chosen according to an arbitrary stochastic process. Also, one could introduce overlooking probabilities to the model. In that case, even if the active player chooses the state that currently contains the object, there is a positive probability that the player fails to find it.

4.8 Appendix

4.8.1 Topological properties of search games

We endow the strategy spaces $\Sigma = \prod_{h \in H^{\text{odd}}} \Delta(S)$ and $\mathcal{T} = \prod_{h \in H^{\text{even}}} \Delta(S)$ with the topology of pointwise convergence. This is identical with the product topology on Σ and the product topology on $\mathcal{T}$. Under this topology, the spaces Σ and $\mathcal{T}$ are compact, and as H^{odd} and H^{even} are countable, Σ and $\mathcal{T}$ are also metrizable.

Definition 4.8.1. Let X be a topological space. A function $f : X \to \mathbb{R}$ is called *lower semi-continuous at* $x \in X$ if, for every sequence $x^n \to x$, we have $\liminf_{n\to\infty} f(x^n) \geq f(x)$. A function $f : X \to \mathbb{R}$ is called *upper semi-continuous* at $x \in X$ if, for every sequence $x^n \to x$, we have $\limsup_{n\to\infty} f(x^n) \leq f(x)$. A function $f : X \to \mathbb{R}$ is called *continuous* at $x \in X$ if it is lower semi-continuous at x and upper semi-continuous at x.

A function $f : X \to \mathbb{R}$ is called *lower semi-continuous* (resp. *upper semi-continuous*, resp. *continuous*) if f is lower semi-continuous at all $x \in X$ (resp. upper semi-continuous at all $x \in X$, resp. continuous at all $x \in X$).

Proposition 4.8.2. *Take a player $i \in \{1, 2\}$.*

1. *The payoff function $u_i : \Sigma \times \mathcal{T} \to \mathbb{R}$ is lower semi-continuous.*
2. *Assume that (σ, τ) is a strategy profile under which the object is found with probability 1. Then, u_i is continuous at (σ, τ).*

Proof.
1. For each strategy profile $(\sigma, \tau) \in \Sigma \times \mathcal{T}$, for each period $n \in \mathbb{N}$, we denote by $u_i^n(\sigma, \tau)$ the probability that player i finds the object during the first n periods under the strategy profile (σ, τ). Note that $u_i^n(\sigma, \tau)$ is non-decreasing in n and converges to $u_i(\sigma, \tau)$ as $n \to \infty$.

Let $(\sigma^k, \tau^k)_{k\in\mathbb{N}}$ be a sequence in $\Sigma \times \mathcal{T}$ converging to a strategy profile (σ, τ). We have for each $n \in \mathbb{N}$

$$u_i^n(\sigma,\tau) = \lim_{k\to\infty} u_i^n(\sigma^k,\tau^k) = \liminf_{k\to\infty} u_i^n(\sigma^k,\tau^k) \le \liminf_{k\to\infty} u_i(\sigma^k,\tau^k).$$

Since $u_i^n(\sigma,\tau)$ converges to $u_i(\sigma,\tau)$ as $n \to \infty$, we obtain

$$u_i(\sigma,\tau) \le \liminf_{k\to\infty} u_i(\sigma^k,\tau^k),$$

which proves that u_i is lower semi-continuous.
2. Assume that under the strategy profile (σ,τ) the object is found with probability 1. Thus, $u_1(\sigma,\tau) + u_2(\sigma,\tau) = 1$. Due to part 1, we only need to show that u_1 and u_2 are upper semi-continuous at (σ,τ). We will prove it for u_1; the proof for u_2 is similar.

Let $(\sigma^k, \tau^k)_{k\in\mathbb{N}}$ be a sequence in $\Sigma \times \mathcal{T}$ converging to (σ,τ). Then

$$\begin{aligned}\limsup_{k\to\infty} u_1(\sigma^k,\tau^k) &= 1 - \liminf_{k\to\infty}(1 - u_1(\sigma^k,\tau^k)) \le 1 - \liminf_{k\to\infty} u_2(\sigma^k,\tau^k)\\ &\le 1 - u_2(\sigma,\tau) = u_1(\sigma,\tau),\end{aligned}$$

where the first equality is a classic supinf equality applied to a limit, the first inequality comes from $u_1 + u_2 \le 1$, the second inequality follows from part 1, and the second equality comes from the assumption we made on (σ,τ). Hence, u_1 is upper semi-continuous at (σ,τ), as desired. □

4.8.2 Functional properties of the value function

In this section we discuss some general functional properties of the value function $p \mapsto v(p)$. The first theorem is devoted to linear properties and the second theorem to Lipschitz-continuity. We remind that the function $p \mapsto v_1(p,s)$ was introduced at the beginning of the subsection 4.5.2.

Theorem 4.8.3. *Let (σ, τ) be a strategy profile. Then the expected payoff functions are linear in the initial probability distribution of the object: for every $\lambda \in [0,1]$, for every $p, q \in \Delta(S)$, for every player $i = 1, 2$,*

$$u_i(\sigma,\tau)(\lambda p + (1-\lambda)q) = \lambda \cdot u_i(\sigma,\tau)(p) + (1-\lambda) \cdot u_i(\sigma,\tau)(q). \quad (4.6)$$

Moreover, for every $s \in S$, the map $p \mapsto v(p,s)$ is linear over every line passing through e^s (the initial probability distribution having probability 1 on state s): for every $p \in \Delta(S)$, for every $\lambda \in (0,1)$

$$v(\lambda e^s + (1-\lambda)p, s) = \lambda + (1-\lambda) \cdot v(p,s).$$

Proof. First we prove equality (4.6). The probability distribution $\lambda \cdot p + (1-\lambda) \cdot q$ can be interpreted as follows: with probability λ the initial probability distribution is p and induces the expected payoff $u_i(\sigma,\tau)(p)$ for player i, and with probability $(1-\lambda)$ the probability distribution is q and induces the expected payoff $u_i(\sigma,\tau,q)$ for player i. Hence, the equality (4.6) holds.

Now we prove the second part of the theorem. Let $p \in \Delta(S)$, $p \neq e^s$, and let $\lambda \in (0,1)$, and denote $p^{\neg s}$ the linear projection of x from e^s to the face $\{y \in \Delta(S) | y_s = 0\}$. Then

$$(\lambda e^s + (1-\lambda)p)^{\neg s} = p^{\neg s}.$$

Indeed, $(\lambda e^s + (1-\lambda)p)^{\neg s}(s) = 0 = [p^{\neg s}](s)$ and for all $j \neq s$:

$$\begin{aligned}
(\lambda e^s + (1-\lambda)p)^{\neg s}(j) &= \frac{(\lambda e^s + (1-\lambda)p)(j)}{1 - (\lambda e^s + (1-\lambda)p)(s)} \\
&= \frac{(1-\lambda) \cdot p(j)}{1 - (\lambda + (1-\lambda) \cdot p(s))} \\
&= \frac{(1-\lambda) \cdot p(j)}{(1-\lambda) \cdot (1 - p(s))} = \frac{p(j)}{(1-p(s))} \\
&= \frac{p(j)}{1-p(s)} = [p^{\neg s}](j).
\end{aligned}$$

Hence, by using $(\lambda e^s + (1-\lambda)p)(s) = \lambda + (1-\lambda)\cdot p(s)$ we have

$$\begin{aligned}
&v(\lambda e^s + (1-\lambda)p, s)\\
=&(\lambda e^s + (1-\lambda)p)(s) + (1-(\lambda e^s + (1-\lambda)p)(s))\cdot(1 - v((\lambda e^s + (1-\lambda)p)^{\neg s}P))\\
=&(\lambda e^s + (1-\lambda)p)(s) + (1-(\lambda e^s + (1-\lambda)p)(s))\cdot(1 - v(p^{\neg s}P))\\
=&\lambda + (1-\lambda)(p(s) + (1-p(s))\cdot(1 - v(p^{\neg s}P)))\\
=&\lambda + (1-\lambda)\cdot v(p, s)
\end{aligned}$$

which completes the proof. □

Remark. For each line passing through e^s, the linearity of the function $p \mapsto v(p, s)$ relies on the fact that if by choosing state s player 1 does not find the object, then the conditional distribution of the location of the object, $p^{\neg s}$, stays on the same line. For lines not passing through e^s, this is no longer true, and the function $p \mapsto v(p, s)$ is generally non-linear. For example when $P = I_4$, $p = (1/3, 1/3, 1/3, 0)$ and $p' = (1/3, 1/3, 0, 1/3)$. In that case, $v(p, 1) = 2/3$ and $v(p', 1) = 2/3$, but $v_1(p/2 + p'/2, 1) = 1/2$.

Before introducing the next theorem, we recall the definition of the total variation distance: for $p, q \in \Delta(S)$, the total variation distance between p and q is the non-negative number

$$||p - q||_{TV} = \max_{S' \subset S} \sum_{s \in S'} [p(s) - q(s)].$$

Theorem 4.8.4. *Let $p, q \in \Delta(S)$. Let $T \in \mathbb{N}$ and let (σ, τ) be a strategy profile. Then, the functions $p \mapsto u_{1,T}(\sigma, \tau)(p)$, $p \mapsto u_1(\sigma, \tau)(p)$, $p \mapsto v_{1,T}(p)$, $p \mapsto v_1(p, s)$ and $p \mapsto v_1(p)$ are 1-Lipschitz continuous with respect to the total variation distance.*

Proof. By Theorem 4.8.3, we have

$$u_1(\sigma, \tau)(p) = \sum_{s \in S} p(s) \cdot u_1(\sigma, \tau)(e^s),$$

$$u_1(\sigma,\tau)(q) = \sum_{s\in S} q(s)\cdot u_1(\sigma,\tau)(e^s).$$

Then,

$$\begin{aligned} u_1(\sigma,\tau)(p) - u_1(\sigma,\tau)(q) &= \sum_{s\in S}[p(s)-q(s)]\cdot u_1(\sigma,\tau)(e^s) \\ &\leq \sum_{\substack{s\in S,\\ p(s)>q(s)}} [p(s)-q(s)] \\ &= ||p-q||_{TV}, \end{aligned}$$

and similarly

$$u_1(\sigma,\tau)(q) - u_1(\sigma,\tau)(p) \leq ||p-q||_{TV}.$$

Hence, $p \mapsto u_1(\sigma,\tau)(p)$ is 1-Lipschitz-continuous.

Taking the infimum over τ and the supremum over σ on both sides of the inequality $u_1(\sigma,\tau)(p) \leq u_1(\sigma,\tau)(q) + ||p-q||_{TV}$ gives $v_1(p) \leq ||p-q||_{TV} + v_1(q)$, which can be written $v_1(p) - v_1(q) \leq ||p-q||_{TV}$. Similarly, $v_1(q) - v_1(p) \leq ||p-q||_{TV}$. Hence, $p \mapsto v_1(p)$ is 1-Lipschitz-continuous too.

The proof for $p \mapsto u_{1,T}(\sigma,\tau)(p)$ and $p \mapsto v_{1,T}(\sigma,\tau)$ are similar. The proof for $p \mapsto v_1(p,s)$ is also similar, but the supremum in σ has to be taken over the strategies that look at state s at period 1. □

Summary and General Discussion

This thesis is divided into three Chapters. Chapter 2 deals with learning in games. Chapter 3 is devoted to optimization. Finally Chapter 4 focuses on game theory.

In Chapter 2 we examine the long-term behavior of regret-minimizing agents in time-varying games with continuous action spaces. In its most basic form, (external) regret minimization guarantees that an agent's cumulative payoff is no worse in the long run than that of the agent's best fixed action in hindsight. Going beyond this worst-case guarantee, we consider a dynamic regret variant that compares the agent's accrued rewards to those of *any* sequence of play. By properly adapting a restart procedure pioneered by Besbes et al. [7], we show that players are able to avoid dynamic regret against any test sequence whose total variation grows sublinearly with the horizon of play. In particular, specializing to a wide-class of no-regret strategies based on mirror descent, we derive explicit rates of dynamic regret minimization, both in expectation and in high probability. We then leverage these results to show that players are able to stay close to Nash equilibrium in time-varying monotone games – and even converge to equilibrium if the sequence of stage games admits a limit.

While the information structure is relevant in some contexts, as the reviewers of the article mentioned, this structure is not well-adapted

when the information has a different form than the gradient of a function. More precisely, in many applications, agents take decisions based on information (amount of money, quantity of products, ...). Further work has to be done in order to adapt this kind of information structure, and study properties of derivative-free algorithms. Another drawback is that all the agents have to follow the same algorithm. In real life, such an algorithm would probably be a mediator. Thus, it would be more natural to target a correlated equilibrium more than a Nash equilibrium. Such a study would be of great interest.

In Chapter 3 we use a class of strongly convergent primal-dual schemes for solving variational inequalities defined by a Lipschitz continuous and pseudo-monotone map in infinite-dimensional Hilbert spaces, which have been studied by Dennis Meier in [22]. This novel numerical scheme is based on Tseng's forward-backward-forward scheme, which is known to display weak convergence, unless very strong global monotonicity assumptions are made on the involved operators. We test the performance of the algorithm in the computationally challenging task to find dynamic user equilibria in traffic networks and verify that our scheme is at least competitive to state-of-the-art solvers, and in some cases even improves upon them.

In general, the commuting operator of a network is not monotone, but the theoretical convergence of the algorithm relies on this assumption. Further work should done, both in order to relax the strong assumption of monotonicity of the operator, and on finding classes of networks that have a monotone operator.

In Chapter 4 we introduce a discrete-time search game, in which two players compete to find an object first. The object moves according to a time-varying Markov chain on finitely many states. The players know the Markov chain and the initial probability distribution of the object, but do not observe the current state of the object. The players are active in turns. The active player chooses a state, and this choice is observed by the other player. If the object is in the chosen state, this

player wins and the game ends. Otherwise, the object moves according to the Markov chain and the game continues at the next period.

We show that this game admits a value, and for any error-term $\varepsilon > 0$, each player has a pure (subgame-perfect) ε-optimal strategy. Interestingly, a 0-optimal strategy does not always exist. The ε-optimal strategies are robust in the sense that they are 2ε-optimal on all finite but sufficiently long horizons, and also 2ε-optimal in the discounted version of the game provided that the discount factor is close to 1. We derive results on the analytic and structural properties of the value and the ε-optimal strategies. Moreover, we examine the performance of the finite truncation strategies, which are easy to calculate and to implement. We devote special attention to the important time-homogeneous case, where additional results hold.

Although many variations of this game are interesting (finitely many players, unknown probability distributions, players move over a graph, the object does not want to be found), two very relevant computational questions arise : how can we compute efficiently optimal strategies for the two players? Is it possible to achieve linear convergence rate of optimal strategies in the finite horizon game to the optimal ones in the infinite horizon game for general sequences of permutation matrices?

Bibliography

[1] Steve Alpern and Shmuel Gal. *The theory of search games and rendezvous*. Vol. 55. Springer Science & Business Media, 2006.

[2] AS Antipin. "Method of convex programming using a symmetric modification of lagrange function". In: *Matekon* 14.2 (1978), pp. 23–38.

[3] Sanjeev Arora, Elad Hazan, and Satyen Kale. "The multiplicative weights update method: a meta-algorithm and applications". In: *Theory of Computing* 8.1 (2012), pp. 121–164.

[4] Heinz H Bauschke, Patrick L Combettes, et al. *Convex analysis and monotone operator theory in Hilbert spaces*. Vol. 408. Springer, 2011.

[5] Amir Beck and Marc Teboulle. "Mirror descent and nonlinear projected subgradient methods for convex optimization". In: *Operations Research Letters* 31.3 (2003), pp. 167–175.

[6] Stanley J Benkoski, Michael G Monticino, and James R Weisinger. "A survey of the search theory literature". In: *Naval Research Logistics (NRL)* 38.4 (1991), pp. 469–494.

[7] Omar Besbes, Yonatan Gur, and Assaf Zeevi. "Non-stationary stochastic optimization". In: *Operations research* 63.5 (2015), pp. 1227–1244.

[8] Radu Ioan Bot, Ernö Robert Csetnek, and Phan Tu Vuong. "The Forward-Backward-Forward Method from discrete and continuous perspective for pseudo-monotone variational inequalities in Hilbert Spaces". In: *arXiv preprint arXiv:1808.08084* (2018).

[9] Radu Ioan Bot, Panayotis Mertikopoulos, Mathias Staudigl, and Phan Tu Vuong. "Forward-backward-forward methods with variance reduction for stochastic variational inequalities". In: *arXiv preprint arXiv:1902.03355* (2019).

[10] Alberto Bressan, Sunčica Čanić, Mauro Garavello, Michael Herty, and Benedetto Piccoli. "Flows on networks: recent results and perspectives". In: *EMS Surveys in Mathematical Sciences* 1.1 (2014), pp. 47–111.

[11] Scott Shorey Brown. "Optimal search for a moving target in discrete time and space". In: *Operations research* 28.6 (1980), pp. 1275–1289.

[12] Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. "Regret analysis of stochastic and nonstochastic multi-armed bandit problems". In: *Foundations and Trends® in Machine Learning* 5.1 (2012), pp. 1–122.

[13] Nicolo Cesa-Bianchi, Pierre Gaillard, Gábor Lugosi, and Gilles Stoltz. "Mirror descent meets fixed share (and feels no regret)". In: *Advances in Neural Information Processing Systems*. 2012, pp. 980–988.

[14] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[15] Nicolo Cesa-Bianchi, Gábor Lugosi, and Gilles Stoltz. "Regret minimization under partial monitoring". In: *Mathematics of Operations Research* 31.3 (2006), pp. 562–580.

[16] Gong Chen and Marc Teboulle. "Convergence analysis of a proximal-like minimization algorithm using Bregman functions". In: *SIAM Journal on Optimization* 3.3 (1993), pp. 538–543.

[17] Gerard Debreu. "A social equilibrium existence theorem". In: *Proceedings of the National Academy of Sciences* 38.10 (1952), pp. 886–893.

[18] Bui V Dinh, Pham G Hung, and Le D Muu. "Bilevel optimization as a regularization approach to pseudomonotone equilibrium problems". In: *Numerical Functional Analysis and Optimization* 35.5 (2014), pp. 539–563.

[19] James M Dobbie. "A two-cell model of search for a moving target". In: *Operations Research* 22.1 (1974), pp. 79–92.

[20] John C Duchi, Alekh Agarwal, Mikael Johansson, and Michael I Jordan. "Ergodic mirror descent". In: *SIAM Journal on Optimization* 22.4 (2012), pp. 1549–1578.

[21] Benoit Duvocelle, János Flesch, Mathias Staudigl, and Dries Vermeulen. "A competitive search game with a moving target". In: *arXiv preprint arXiv:2008.12032* (2020).

[22] Benoit Duvocelle, Dennis Meier, Mathias Staudigl, and Phan Tu Vuong. "Strong Convergence of Forward-Backward-Forward Methods for Pseudo-monotone Variational Inequalities with Applications to Dynamic User Equilibrium in Traffic Networks". In: *arXiv preprint arXiv:1908.07211* (2019).

[23] Benoit Duvocelle, Panayotis Mertikopoulos, Mathias Staudigl, and Dries Vermeulen. "Learning in time-varying games". In: *arXiv preprint arXiv:1809.03066* (2018).

[24] Francisco Facchinei, Andreas Fischer, and Veronica Piccialli. "On generalized Nash games and variational inequalities". In: *Operations Research Letters* 35.2 (2007), pp. 159–164.

[25] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.

[26] Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. "Online convex optimization in the bandit setting: gradient descent without a gradient". In: *arXiv preprint cs/0408007* (2004).

[27] János Flesch, Emin Karagözoğlu, and Andrés Perea. "Optimal search for a moving target with the option to wait". In: *Naval Research Logistics (NRL)* 56.6 (2009), pp. 526–539.

[28] János Flesch, Jeroen Kuipers, Ayala Mashiah-Yaakovi, Gijs Schoenmakers, Eilon Solan, and Koos Vrieze. "Perfect-information games with lower-semicontinuous payoffs". In: *Mathematics of Operations Research* 35.4 (2010), pp. 742–755.

[29] János Flesch and Arkadi Predtetchinski. "Subgame-perfect epsilon-equilibria in perfect information games with common preferences at the limit". In: *Mathematics of Operations Research* 41.4 (2016), pp. 1208–1221.

[30] Yoav Freund and Robert E Schapire. "Adaptive game playing using multiplicative weights". In: *Games and Economic Behavior* 29.1-2 (1999), pp. 79–103.

[31] Terry L Friesz, David Bernstein, Tony E Smith, Roger L Tobin, and Byung-Wook Wie. "A variational inequality formulation of the dynamic network user equilibrium problem". In: *Operations research* 41.1 (1993), pp. 179–191.

[32] Terry L Friesz and Ke Han. "The mathematical foundations of dynamic user equilibrium". In: *Transportation research part B: methodological* 126 (2019), pp. 309–328.

[33] Drew Fudenberg and David Levine. "Subgame-Perfect Equilibria of Finite-and Infinite-Horizon Games". In: *Journal of Economic Theory* 31.2 (1983), pp. 251–268.

[34] Shmuel Gal. "Search games". In: *Wiley Encyclopedia of Operations Research and Management Science* (2010).

[35] Shmuel Gal. "Search games with mobile and immobile hider". In: *SIAM Journal on Control and Optimization* 17.1 (1979), pp. 99–122.

[36] Shmuel Gal. "Search games: a review". In: *Search Theory*. Springer, 2013, pp. 3–15.

[37] Mauro Garavello, Ke Han, and Benedetto Piccoli. *Models for vehicular traffic on networks*. Vol. 9. American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2016.

[38] Mauro Garavello and Benedetto Piccoli. *Traffic flow on networks*. Vol. 1. American institute of mathematical sciences Springfield, 2006.

[39] Andrey Garnaev. *Search games and other applications of game theory*. Vol. 485. Springer Science & Business Media, 2012.

[40] Tristan Garrec and Marco Scarsini. "Search for an immobile hider on a stochastic network". In: *European Journal of Operational Research* 283.2 (2020), pp. 783–794.

[41] Aviv Gibali and Duong Viet Thong. "Tseng type methods for solving inclusion problems and its applications". In: *Calcolo* 55.4 (2018), p. 49.

[42] K Goebel and S Reich. *Uniform convexity, hyperbolic geometry, and nonexpansive mappings. 1984*. 1984.

[43] Osman Güler. "On the convergence of the proximal point algorithm for convex minimization". In: *SIAM Journal on Control and Optimization* 29.2 (1991), pp. 403–419.

[44] Eric C Hall and Rebecca M Willett. "Online convex optimization in dynamic environments". In: *IEEE Journal of Selected Topics in Signal Processing* 9.4 (2015), pp. 647–662.

[45] Peter Hall and Christopher C Heyde. *Martingale limit theory and its application*. Academic press, 2014.

[46] Deren Han and Hong K Lo. "Two new self-adaptive projection methods for variational inequality problems". In: *Computers & Mathematics with Applications* 43.12 (2002), pp. 1529–1537.

[47] Ke Han, Gabriel Eve, and Terry L Friesz. "Computing dynamic user equilibria on large-scale networks with software implementation". In: *Networks and Spatial Economics* 19.3 (2019), pp. 869–902.

[48] Ke Han, Terry L Friesz, WY Szeto, and Hongcheng Liu. "Elastic demand dynamic network user equilibrium: Formulation, existence and computation". In: *Transportation Research Part B: Methodological* 81 (2015), pp. 183–209.

[49] Christopher Harris. "Existence and characterization of perfect equilibrium in games of perfect information". In: *Econometrica: Journal of the Econometric Society* (1985), pp. 613–628.

[50] Sergiu Hart and Andreu Mas-Colell. "Uncoupled dynamics do not lead to Nash equilibrium". In: *American Economic Review* 93.5 (2003), pp. 1830–1836.

[51] BS He and Li-Zhi Liao. "Improvements of some projection methods for monotone nonlinear variational inequalities". In: *Journal of Optimization Theory and applications* 112.1 (2002), pp. 111–128.

[52] Amélie Heliou, Johanne Cohen, and Panayotis Mertikopoulos. "Learning with bandit feedback in potential games". In: *Advances in Neural Information Processing Systems*. 2017, pp. 6369–6378.

[53] Michael Hinze, René Pinnau, Michael Ulbrich, and Stefan Ulbrich. *Optimization with PDE constraints*. Vol. 23. Springer Science & Business Media, 2008.

[54] Josef Hofbauer and William H Sandholm. "Stable games and their dynamics". In: *Journal of Economic theory* 144.4 (2009), pp. 1665–1693.

[55] Ryusuke Hohzaki. "Search games: Literature and survey". In: *Journal of the Operations Research Society of Japan* 59.1 (2016), pp. 1–34.

[56] Benjamin Paul Jordan. "On optimal search for a moving target". PhD thesis. Durham University, 1997.

[57] Anatoli Juditsky, Arkadi Nemirovski, and Claire Tauvel. "Solving variational inequalities with stochastic mirror-prox algorithm". In: *Stochastic Systems* 1.1 (2011), pp. 17–58.

[58] YC Kan. "A Counterexample for an Optimal Search-and-Stop Model". In: *Operations Research* 22.4 (1974), pp. 889–892.

[59] Aswin Kannan and Uday V Shanbhag. "Distributed computation of equilibria in monotone Nash games via iterative regularization techniques". In: *SIAM Journal on Optimization* 22.4 (2012), pp. 1177–1205.

[60] David Kinderlehrer and Guido Stampacchia. *An introduction to variational inequalities and their applications*. Vol. 31. Siam, 1980.

[61] Bronisław Knaster, Casimir Kuratowski, and Stefan Mazurkiewicz. "Ein Beweis des Fixpunktsatzes für n-dimensionale Simplexe". In: *Fundamenta Mathematicae* 14.1 (1929), pp. 132–137.

[62] GM Korpelevich. "The extragradient method for finding saddle points and other problems". In: *Matecon* 12 (1976), pp. 747–756.

[63] David A Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, 2017.

[64] Hong K Lo and Wai Yuen Szeto. "A cell-based variational inequality formulation of the dynamic user optimal assignment problem". In: *Transportation Research Part B: Methodological* 36.5 (2002), pp. 421–443.

[65] Jiancheng Long, Hai-Jun Huang, Ziyou Gao, and Wai Yuen Szeto. "An intersection-movement-based dynamic user optimal route choice problem". In: *Operations Research* 61.5 (2013), pp. 1134–1147.

[66] Gábor Lugosi, Shie Mannor, and Gilles Stoltz. "Strategies for prediction under imperfect monitoring". In: *Mathematics of Operations Research* 33.3 (2008), pp. 513–528.

[67] IM MacPhee and BP Jordan. "Optimal search for a moving target". In: *Probability in the Engineering and Informational Sciences* 9.2 (1995), pp. 159–182.

[68] Shie Mannor, Vianney Perchet, and Gilles Stoltz. "Set-valued approachability and online learning with partial monitoring". In: *The Journal of Machine Learning Research* 15.1 (2014), pp. 3247–3295.

[69] Donald A Martin. "Borel determinacy". In: *Annals of Mathematics* 102.2 (1975), pp. 363–371.

[70] Donald A Martin. "The determinacy of Blackwell games". In: *The Journal of Symbolic Logic* 63.4 (1998), pp. 1565–1581.

[71] Bernard Martinet. “Régularisation d’inéquations variationnelles par approximations successives. Rev. Française Informat”. In: *Recherche Opérationnelle* 4 (1970), pp. 154–158.

[72] M Maschler, Eilon Solan, and Shmuel Zamir. “Game Theory”. In: *Cambridge University Press, Cambridge* (2013).

[73] Jean-François Mertens. “Repeated games”. In: *Game Theory and Applications*. Elsevier, 1990, pp. 77–130.

[74] Panayotis Mertikopoulos, E Veronica Belmega, Romain Negrel, and Luca Sanguinetti. “Distributed stochastic optimization via matrix exponential learning”. In: *IEEE Transactions on Signal Processing* 65.9 (2017), pp. 2277–2290.

[75] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. “Cycles in adversarial regularized learning”. In: *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2018, pp. 2703–2717.

[76] Panayotis Mertikopoulos and Mathias Staudigl. “On the convergence of gradient-like flows with noisy gradient input”. In: *SIAM Journal on Optimization* 28.1 (2018), pp. 163–197.

[77] Panayotis Mertikopoulos and Mathias Staudigl. “Stochastic mirror descent dynamics and their convergence in monotone variational inequalities”. In: *Journal of optimization theory and applications* 179.3 (2018), pp. 838–867.

[78] Panayotis Mertikopoulos, Houssam Zenati, Bruno Lecouat, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. “Mirror descent in saddle-point problems: Going the extra (gradient) mile”. In: *arXiv preprint arXiv:1807.02629* (2018).

[79] Panayotis Mertikopoulos and Zhengyuan Zhou. “Learning in games with continuous action sets and unknown payoff functions”. In: *Mathematical Programming* 173.1-2 (2019), pp. 465–507.

[80] Dov Monderer and Lloyd S Shapley. “Potential games”. In: *Games and economic behavior* 14.1 (1996), pp. 124–143.

[81] Richard Mounce and Mike Smith. "Uniqueness of equilibrium in steady state and dynamic traffic networks". In: *Transportation and Traffic Theory 2007. Papers Selected for Presentation at ISTTT17Engineering and Physical Sciences Research Council (Great Britain) Rees Jeffreys Road FundTransport Research FoundationTMS ConsultancyOve Arup and Partners, Hong KongTransportation Planning (International) PTV AG*. 2007.

[82] Le Dung Muu. "Stability property of a class of variational inequalities". In: *Optimization* 15.3 (1984), pp. 347–351.

[83] Teruhisa Nakai. "A search game with one object and two searchers". In: *Journal of applied probability* 23.3 (1986), pp. 696–707.

[84] Teruhisa Nakai. "Model of search for a target moving among three boxes: Some special cases". In: *Journal of Operations Research Society of Japan* 16 (1973), pp. 151–162.

[85] Arkadi Nemirovski. "Prox-method with rate of convergence O $(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems". In: *SIAM Journal on Optimization* 15.1 (2004), pp. 229–251.

[86] Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. "Robust stochastic approximation approach to stochastic programming". In: *SIAM Journal on optimization* 19.4 (2009), pp. 1574–1609.

[87] Arkadi Semenovich Nemirovsky and David Borisovich Yudin. "Problem complexity and method efficiency in optimization." In: *Wiley-Interscience Series in Discrete Mathematics* (1983).

[88] Yurii Nesterov. "Primal-dual subgradient methods for convex problems". In: *Mathematical programming* 120.1 (2009), pp. 221–259.

[89] Hukukane Nikaidô, Kazuo Isoda, et al. “Note on non-cooperative convex games”. In: *Pacific Journal of Mathematics* 5.Suppl. 1 (1955), pp. 807–815.

[90] Gerasimos Palaiopanos, Ioannis Panageas, and Georgios Piliouras. “Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 5872–5882.

[91] Stephen M Pollock. “A simple model of search for a moving target”. In: *Operations Research* 18.5 (1970), pp. 883–903.

[92] Ariel Orda Raphael, Raphael Rom, and Nahum Shimkin. “Competitive Routing in Multi-User Communication Networks”. In: *IEEE/ACM Transactions on Networking*. Citeseer. 1993.

[93] R Tyrrell Rockafellar. *Convex analysis*. 28. Princeton university press, 1970.

[94] R Tyrrell Rockafellar. “Monotone operators and the proximal point algorithm”. In: *SIAM journal on control and optimization* 14.5 (1976), pp. 877–898.

[95] J Ben Rosen. “Existence and uniqueness of equilibrium points for concave n-person games”. In: *Econometrica: Journal of the Econometric Society* (1965), pp. 520–534.

[96] Aldo Rustichini. “Minimizing regret: The general case”. In: *Games and Economic Behavior* 29.1-2 (1999), pp. 224–243.

[97] William H Sandholm. “Population games and deterministic evolutionary dynamics”. In: *Handbook of game theory with economic applications*. Vol. 4. Elsevier, 2015, pp. 703–778.

[98] Paul J Schweitzer. “Threshold probabilities when searching for a moving target”. In: *Operations Research* 19.3 (1971), pp. 707–709.

[99] Gesualdo Scutari, Daniel P Palomar, Francisco Facchinei, and Jong-Shi Pang. “Convex optimization, game theory, and variational inequality theory”. In: *IEEE Signal Processing Magazine* 27.3 (2010), pp. 35–49.

[100] Shai Shalev-Shwartz et al. “Online learning and online convex optimization”. In: *Foundations and Trends in Machine Learning* 4.2 (2012), pp. 107–194.

[101] Sylvain Sorin and Cheng Wan. “Finite composite games: Equilibria and dynamics”. In: *arXiv preprint arXiv:1503.07935* (2015).

[102] James C Spall. “A one-measurement form of simultaneous perturbation stochastic approximation”. In: *Automatica* 33.1 (1997), pp. 109–112.

[103] Mathias Staudigl and Panayotis Mertikopoulos. “Convergent Noisy forward-backward-forward algorithms in non-monotone variational inequalities”. In: *IFAC-PapersOnLine* 52.3 (2019), pp. 120–125.

[104] Lawrence D Stone. *Theory of optimal search*. Vol. 118. Elsevier, 1976.

[105] Lawrence D Stone, Johannes O Royset, and Alan R Washburn. “Optimal Search for Moving Targets (International Series in Operations Research & Management Science 237)”. In: *Cham, Switzerland: Springer* (2016).

[106] Marc Teboulle. “Entropic proximal mappings with applications to nonlinear programming”. In: *Mathematics of Operations Research* 17.3 (1992), pp. 670–690.

[107] Duong Viet Thong and Dang Van Hieu. “Strong convergence of extragradient methods with a new step size for solving variational inequality problems”. In: *Computational and Applied Mathematics* 38.3 (2019), p. 136.

[108] Paul Tseng. “A modified forward-backward splitting method for maximal monotone mappings”. In: *SIAM Journal on Control and Optimization* 38.2 (2000), pp. 431–446.

[109] Yannick Viossat and Andriy Zapechelnyuk. “No-regret dynamics and fictitious play”. In: *Journal of Economic Theory* 148.2 (2013), pp. 825–842.

[110] Phan Tu Vuong. “On the weak convergence of the extragradient method for solving pseudo-monotone variational inequalities”. In: *Journal of Optimization Theory and Applications* 176.2 (2018), pp. 399–409.

[111] John Glen Wardrop and James Ivor Whitehead. “Correspondence. some theoretical aspects of road traffic research.” In: *Proceedings of the Institution of Civil Engineers* 1.5 (1952), pp. 767–768.

[112] Alan R Washburn. “Search for a moving target: The FAB algorithm”. In: *Operations research* 31.4 (1983), pp. 739–751.

[113] Martin Zinkevich. “Online convex programming and generalized infinitesimal gradient ascent”. In: *Proceedings of the 20th international conference on machine learning (icml-03)*. 2003, pp. 928–936.

Impact of the thesis

Chapter 2 of this thesis deals with decision making when several agents interact in an environment with a low amount of information. Such situations are ubiquitous in economics. One can think of companies on a market, political parties that have to form a coalition, behavior in a group of humans, animals, to name a few. The agents, based on the information they received yesterday, can take better individual decisions today. Consider two factories on a market. If there is overproduction the previous month, they know they have to reduce their production during the coming month. The algorithm we studied gives precise production figures each month, and guarantees that, if both factories follow this algorithm, they will reach a stable market.

Chapter 3 of this thesis deals with traffic flows. We discuss an algorithm, known to have good properties, and we run this algorithm with, as an instance, a network with commuters. The algorithm takes into consideration all the information of the users, and in order to reduce the traffic time, it recommends to each agent which road to take. Such algorithms are applied for example in route planners, such as Google Maps, Waze, Coyote, Tom-Tom, to name a few.

Chapter 4 of this thesis deals with a game in which two agents interact in turns to find an object first, where the object is moving over time. Such a game can be used to model for example patent races, and help to find good investment strategies. Consider the following situation. Two laboratories are competing in order to find a vaccine to Covid 19. To do so, they have to invest in one of different technologies. The model takes into account that the virus mutates over time, and a vaccine today might not be effective next year.

Research Projects

Duvocelle, B., Mertikopoulos, P., Staudigl, M., & Vermeulen, D. (2018). "Learning in time-varying games". *arXiv preprint arXiv:1809.03066*. Proceeding for Mathematics of Operations Research.

Duvocelle, B., Meier, D., Staudigl, M., & Vuong, P. T. (2019). "Strong Convergence of Forward-Backward-Forward Methods for Pseudo-monotone Variational Inequalities with Applications to Dynamic User Equilibrium in Traffic Networks". *arXiv preprint arXiv:1908.07211*. Proceeding for Optimization and Engineering.

Duvocelle, B., Flesch, J., Staudigl, M., & Vermeulen, D. (2020). "A competitive search game with a moving target". *arXiv preprint arXiv:2008.12032*. Submitted for publication.

About the author

Benoit Duvocelle was born on January 28, 1992 in Bayonne, France. He started his Bachelor in Mathematics, Physics and Engineering Science at the Lycée Louis Barthou, Pau, (classe préparatoire, MPSI-MP) in 2010. He obtained a Bachelor degree of Fundamental and Applied Mathematics at Université Paris-Saclay, France, in 2014. He received a Master degree of Higher Education of Mathematics at Université Paris-Saclay in 2016. The same year, he obtained the Agrégation de mathématiques and a scolarship from the Fondation Mathématiques Jacques Hadamard (FMJH). He obtained a Master degree of Optimization at Université Paris-Saclay in 2017, and received the degree of Magistère de Mathématiques the same year. Afterwards, he joined Maastricht University as a PhD candidate under the supervision of Dr. János Flesch, Dr. Mathias Staudigl and Prof. Dr. Dries Vermeulen. The outcome of his research are presented in this thesis. Some parts of this thesis have been presented at various European conferences. His research interests are Game Theory, Operations Research, Optimization and Contests.