Robustness of Solutions in Game Theory

Values and Strategies in Partially Observable, Perturbed, Stochastic, and Infinite Games

by

Raimundo Julián Saona Urmeneta

July, 2025

A thesis submitted to the
Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Committee in charge:

Maksym Serbyn, Chair
Krishnendu Chatterjee
Miquel Oliu-Barton
Eilon Solan
Bruno Ziliotto

The thesis of Raimundo Julián Saona Urmeneta, titled *Robustness of Solutions in Game Theory*, is approved by:

Supervisor: Krishnendu Chatterjee, ISTA, Klosterneuburg, Austria

Signature: ________________________________

Committee Member: Miquel Oliu-Barton, Paris Nanterre University, Paris, France

Signature: ________________________________

Committee Member: Eilon Solan, Tel-Aviv University, Tel-Aviv, Israel

Signature: ________________________________

Committee Member: Bruno Ziliotto, CNRS, Toulouse, France

Signature: ________________________________

Defense Chair: Maksym Serbyn, ISTA, Klosterneuburg, Austria

Signature: ________________________________

Signed page is on file

ISTA Thesis, ISSN: 2663-337X

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I accept full responsibility for the content and factual accuracy of this work, including the data and their analysis and presentation, and the text and citation of other work.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: ______________________________

Raimundo Julián Saona Urmeneta
July, 2025

Signed page is on file

Abstract

Game Theory is the mathematical formalization of social dynamics – systems where agents interact over time and the evolution of the state of the system depends on the decisions of every player. This thesis takes the perspective of a single player and focuses on what they can guarantee in the worst case over the behavior of other players. In other words, we consider that the objective of every other player in the game is exactly the opposite to the player. We focus on sustained interactions over time, where the players repeatedly obtain quantitative rewards over time, and they are interested in maximizing their long-term performance. Formally, this thesis focuses on zero-sum games with the liminf average objective. Two fundamental questions that Game Theory aims to answer are the following.

1. How much can a player guarantee to obtain after the interaction?
2. How to act in order to obtain the previously mentioned guarantee?

These questions are formalized by the concepts of *value* and *optimal strategies*. We study their properties on games that exhibit one or more of the following properties.

1. Partial Observation: the players can not perfectly observe the current state of the system during the game. We consider the model of (finite) Partially Observable Markov Decision Processes and prove that finite-memory strategies are sufficient to approximately guarantee the value.
2. Perturbed Description: the formal description of the game is perturbed by a small parameter. We consider the model of (finite) Perturbed Matrix Games, and provide algorithms to check various robustness properties and to compute the parameterized value and optimal strategies.
3. Stochastic Transitions: the actions of the players determine the behavior of the evolution of the system, described as a probability distribution over the next state. We consider the model of (finite) Perturbed Stochastic Games and provide formulas for the marginal value.
4. Infinite States: the system can be in infinitely many states. We consider the model of Random Dynamic Games on a class of infinite graphs, prove the existence of the value, and quantify the concentration of finite-horizon values.

Acknowledgements

I thank my advisor, Krishnendu Chatterjee, for the opportunity to develop as a researcher during this PhD. While always being available to provide inspiring guidance, he provided all the means and freedom I required. Moreover, the working environment he created was something I hold close to my heart. Thank you Krish.

I particularly thank the guidance of and collaboration with Miquel Oliu-Barton and Bruno Ziliotto. They have been invaluable to my development in Game Theory. Furthermore, I would like to thank Eilon solan for being a part of my thesis committee, Maksym Serbyn for chairing my thesis defense, and Maximilian Jösch for chairing my qualifying exam. In general, this PhD thesis and my development as a researcher would not have been possible without all the amazing people in the community as a whole. Thank you everyone.

Lastly, I would like to thank my partner Delfina Krause for her support during the time of my PhD and beyond.

Funding sources The works included in this thesis were partially supported by:

- Austrian Science Fund (FWF), grants 10.55776/COE12 and No RiSE/SHiNE S11407,
- French Agence Nationale de la Recherche (ANR), grants ANR-21-CE40-0020 (CONVERGENCE) and ANR-20-CE40-0002 (GrHyDy),
- Fondation Mathématique Jaques Hadamard, grant PGMO RSG 2018-0031H,
- European Research Council (ERC), Consolidator grant 863818 (ForM-SMArt),
- Agencia Nacional de Investigación y Desarrollo (ANID Chile), grant ACT210005,
- Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt Chile), grant 1220174,
- Comisión Nacional de Investigación Científica y Tecnológica (CONICYT Chile), grant PII 20150140,
- Evaluation-orientation de la Coopération Scientifique and Comisión Nacional de Investigación Científica y Tecnológica (ECOS-CONICYT), grant C15E03,
- European Cooperation in Science and Technology (E-COST), grants CA16228 - European Network for Game Theory (GAMENET) and E-COST-GRANT-CA16228-c5a69859.

About the Author

Raimundo Julián Saona Saona Urmeneta grew up in Chile where he completed a bachelor and master of science in Mathematics at the University of Chile before joining ISTA in September 2019 for the PhD. His research focuses on theoretical and algorithmic aspects of Game Theory. His main research interests include developing tools to further analyze situations involving uncertainty. He has a broad publication record in mathematical journals and computer science conferences. He has the following major journal papers: three papers at Mathematics of Operations Research (MOR), one paper at Mathematical Programming. In addition, his work covers many different areas in computer science, with two papers at SODA (the flagship conference in algorithms), one paper at LICS (the flagship conference in logic in computer science), one paper at AAAI (premier AI conference), one paper at TACAS (a top verification conference), one paper in ICALP (a top theoretical computer science conference), and one paper in UAI (premier AI conference).

List of Collaborators and Publications

The main text of this thesis contains the following works, where all authors have equal contribution and are ordered alphabetically.

- Chapter 2 is based on "Krishnendu Chatterjee, Raimundo Saona, and Bruno Ziliotto. Finite-Memory Strategies in POMDPs with Long-Run Average Objectives. *Mathematics of Operations Research*, 47(1):100–119, 2021", which is referred to as [CSZ21].
- Chapter 3 is based on "Krishnendu Chatterjee, Miquel Oliu-Barton, and Raimundo Saona. Value-Positivity for Matrix Games. *Mathematics of Operations Research*, page 1–24, 2024", which is referred to as [COBS24].
- Chapter 4 is based on "Luc Attia, Miquel Oliu-Barton, and Raimundo Saona. Marginal Values of a Stochastic Game. *Mathematics of Operations Research*, 50(1):482–505, 2025", which is referred to as [AOBS25].
- Chapter 5 is based on "Luc Attia, Lyuben Lichev, Dieter Mitsche, Raimundo Saona, and Bruno Ziliotto. Random Zero-Sum Dynamic Games on Infinite Directed Graphs. *Dynamic Games and Applications*, 2025", which is referred to as [ALM$^+$25].

The following list contains other works published during the PhD period that are not discussed in this thesis, ordered by the time the project started.

- Raimundo Saona, Fyodor A. Kondrashov, and Ksenia A. Khudiakova. Relation Between the Number of Peaks and the Number of Reciprocal Sign Epistatic Interactions. *Bulletin of Mathematical Biology*, 84(8):74, 2022[1].
- Raimundo Saona, Fyodor A. Kondrashov, and Ksenia A. Khudiakova. Correction to: Relation Between the Number of Peaks and the Number of Reciprocal Sign Epistatic Interactions. *Bulletin of Mathematical Biology*, 85(3):17, 2023[2].
- Krishnendu Chatterjee, Tobias Meggendorfer, Raimundo Saona, and Jakub Svoboda. Faster Algorithm for Turn-based Stochastic Games with Bounded Treewidth. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, page 4590 – 4605, 2023.
- Krishnendu Chatterjee, Mona Mohammadi, and Raimundo Saona. Repeated Prophet Inequality with Near-optimal Bounds, 2022.

[1] This is a first-author paper, as used in the biological sciences, i.e., I was the main contributor to the development of the content of this work.

[2] This is a first-author paper, as used in the biological sciences, i.e., I was the main contributor to the development of the content of this work.

- Giordano Giambartolomei, Frederik Mallmann-Trenn, and Raimundo Saona. Prophet Inequalities: Separating Random Order from Order Selection, 2023.

- Krishnendu Chatterjee, Mahdi JafariRaviz, Raimundo Saona, and Jakub Svoboda. Value Iteration with Guessing for Markov Chains and Markov Decision Processes. In *Tools and Algorithms for the Construction and Analysis of Systems*, volume 15697, page 217–236. Springer, 2025.

- Ali Asadi, Krishnendu Chatterjee, Jakub Svoboda, and Raimundo Saona Urmeneta. Deterministic Sub-exponential Algorithm for Discounted-sum Games with Unary Weights. In *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science*, page 1–12, 2024.

- Giordano Giambartolomei, Frederik Mallmann-Trenn, and Raimundo Saona. IID Prophet Inequality with Random Horizon: Going beyond Increasing Hazard Rates. In *52nd International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 334, pages 87:1–87:21, 2025.

- Ali Asadi, Krishnendu Chatterjee, Raimundo Saona, and Jakub Svoboda. Concurrent Stochastic Games with Stateful-Discounted and Parity Objectives: Complexity and Algorithms. *Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, 323:5:1–5:17, 2024.

- Krishnendu Chatterjee, David Lurie, Raimundo Saona, and Bruno Ziliotto. Ergodic Unobservable MDPs: Decidability of Approximation, 2024.

- Andrea Davini, Raimundo Saona, and Bruno Ziliotto. Stochastic Homogenization of HJ Equations: a Differential Game Approach, 2024.

- Krishnendu Chatterjee, Ruichen Luo, Raimundo Saona, and Jakub Svoboda. Linear Equations with Min and Max Operators: Computational Complexity. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(11):11150–11157, 2025.

- Ali Asadi, Krishnendu Chatterjee, Raimundo Saona, and Ali Shafiee. Limit-sure reachability for small memory policies in POMDPs is NP-complete, 2024.

Table of Contents

List of Figures

List of Algorithms

CHAPTER 1

Introduction

1.1 Prologue

Game In this work, we consider dynamic games where two opponents interact repeatedly over time. The state of the dynamic is jointly controlled by the players. At each stage, players chose actions and the state of the system evolves depending only on the current state and the actions of the players, but not, for example, on time. Whenever the dynamic is in a given state and players choose a given profile of actions, one player must pay to the other a fixed reward. Note that the action of players affects both their next reward and the evolution of the system, and therefore their future rewards. These two effects are sometimes contradictory, making the analysis of a game interesting and challenging. Because players play over time, they obtain a sequence of rewards. Players are assumed to be rational and maximize their long-term average reward. A strategy for a player describes how to act in each possible situation of a game. In other words, given a partial development of the game, the strategy determines which action to take.

Value and Optimal Strategies Each strategy of a player is related to a worst-case long-term average reward, i.e., what it guarantees irrespective of the actions of the opponent. We study the best long-term average reward a player can guarantee, which we call the *value* of the game. Related to the value, there are (approximately) optimal strategies. While an optimal strategy guarantees the value, an approximately optimal strategy guarantees approximately the value. We study these solution concepts in games with the following characteristics.

Partial Observation Partial observation refers to the fact that, at each step, the players control the dynamic but can not observe the underling state directly. Instead, correlated signals about it are revealed to them. We consider the model of finite Partially Observable Markov Decision Process (POMDP). In other words, we consider the basic case where there is only one player. In this model, at each stage, the player chooses an action that determines, together with the current state, the distribution over the next state. The player receives a stage signal about the current state before acting. POMDPs generalize the model of Markov Decision Process (MDP) introduced by Bellman [Bel57]. We focus on the complexity required by approximately optimal strategies.

Perturbed Description The description of the game is given by the evolution of the system and the rewards obtain whenever the players take an action profile in each state. The simplest dynamic game has only one state, making the dynamic trivial. In this case, the game is fully determined by the rewards related to each action profile. Because we consider zero-sum games, this model corresponds to matrix games, which is the most basic model in Game Theory. We consider polynomial matrix games, that is, matrix games where the rewards are polynomial functions of a real parameter $\varepsilon \geq 0$. The game corresponding to $\varepsilon = 0$ is interpreted as unperturbed, as opposed to the case $\varepsilon > 0$ which is interpreted as perturbed. We focus on the existence of robust strategies, and the parameterized description of optimal strategies and the value.

Stochastic Transitions Zero-sum, finite actions, finite states, and stochastic transitions describe the model of stochastic games, introduced by Shapley [Sha53]. Like in the original model, we consider that there are finitely many actions and states. Stochastic games are a central model in dynamic games and represent situations where the evolution of the state is controlled by the players and the environment, and the environment affects the state in a stationary manner. They generalize matrix games, repeated games, stopping games, and Markov Decission Processes. Stochastic games are parameterized by a reward function, a transition function, and possibly a discount rate. We focus on the marginal value with respect to perturbations on the description of the game.

Infinite States All models previously mentioned are finite in nature, and only the objective of the players considers infinitely many stages. Moreover, a central assumption is that the evolution of the system depends only on the current state and actions of the players, but not on time. Note that the full generality of infinite states allows to keep track of time, which leads to models that are far too general. Therefore, we consider a restricted class of infinite-state games. Moreover, for simplicity, we consider games where that players play in turns instead of simultaneously, leading to the model of zero-sum dynamic games with perfect information played on graphs, introduced by Ehrenfeucht and Mycielski [EM79], where the dynamic is as follows. Given a graph where every vertex has at least one outgoing edge, the game starts at a given vertex. Two players, who both know the position of the state at all times, alternate in moving the state along an outgoing edge to a neighboring vertex. Each vertex is assigned a uniformly bounded reward and every time a vertex is visited, Player 2 pays Player 1 a fixed reward. In the n-stage game, the objective of Player 1 is to maximize the average reward over the first n stages. The value of the n-stage game is the maximum average reward Player 1 can guarantee irrespective of the actions of Player 2. We focus on the existence of a limit of the sequence of n-stage values as n grows to infinity.

1.2 Prior work

Partially Observable Markov Decision Processes The liminf average reward is a standard long-run objective, see [ABFG$^+$93] for a survey. The value is known to coincide with several classic definitions of long-run values (asymptotic value, uniform value, general uniform value, long-run average value, uncertain-duration process value [RSV02, Ren10, RV17, NS10, VZ16]) and has been characterized in [RV17]. While, strong results are available concerning the existence and characterization of the value, little was known about the sophistication of approximately optimal strategies. In the full-observation model of Markov Decision Processes (MDPs), there exist optimal stationary strategies [Bla62]. In the no-observation model of blind

MDPs, there exist approximately optimal belief-stationary strategies [RSV02]. In POMDPs with an ergodic structure, there also exist approximately optimal belief-stationary strategies [Bor00]. In the full generality of POMDPs, a natural question is the existence of approximately optimal strategies that make use of only finite memory. Previous undecidability results show that, even if they exist, an upper bound on the required memory in terms of the data of the POMDP can not exist [MHC03].

Perturbed Matrix Games A matrix game is represented by a reward matrix M_0 where rows (respectively columns) correspond to possible actions for the row- (respectively column-) player, and the entry $(\mathsf{M}_0)_{i,j}$ is the reward the column-player pays the row-player when the pair of actions (i, j) is chosen. The existence of the value and optimal strategies of matrix games was established nearly 100 years ago [vN28]. The continuity and Lipschitz property of the value of matrix games is well known. The value of a matrix game can be computed by solving a linear program (LP), which can be solved in polynomial time [Kha80]. In fact, LPs are equivalent to matrix games [Adl13, Dan51].

Because the value function is not Fréchet-differentiable, the stability analysis of matrix games is usually investigated through directional derivatives. Mills [Mil56] considered a linearly perturbed matrix game, that is, $\varepsilon \mapsto \mathsf{M}_0 + \mathsf{M}_1\varepsilon$, also framed as a perturbation of M_0 in the direction of M_1. A characterization was obtained for the right derivative of the value, together with a polynomial-time algorithm, based on the reduction to an auxiliary LP of similar size. This characterization has been extended to a broader class of games [RS01]. Perturbed games can be seen as a set of parametrized games, and the regularity of the value function and the set of ε-optimal strategies has been studied in [TV80] for a broad class of games. The asymptotic behavior of perturbed games has been studied in [AFFG01]. Lastly, polynomial matrix games arise naturally in the study of stochastic games [AOB19].

More generally, perturbation theory is a broad field and perturbed linear programs have been studied [AFH13]. Previous research has focused on perturbations only to the objective function or the right-hand side of the constraints of an LP [Fia83, Fia97, Gal94, GGH97], while research on general perturbation that affects the entries of the matrix is sparse [Jer73b, Jer73a]. LPs with general perturbations can exhibit very different behaviors, for example, feasibility can change for small perturbations or the limit of optimal strategies as the perturbation vanishes is no longer feasible. Therefore, assumptions on the effect of the perturbation are usually taken. Checking whether these conditions are satisfied may take exponential time [AFH13]. A particular class of perturbed LPs with efficient algorithms is described in [Ruh91] studying parametric flows, which boils down to studying perturbed LPs where only the objective function depends on the perturbation.

Perturbed Stochastic Games The literature on stochastic games is abundant both in theory and applications. We refer the reader to [SV15] for a summary of the historical context and the impact of Shapley's seminal work, and to [Ami03] for a review of applications, which include resource economics, industrial organization, market games, and empirical economics among others.

Together with the value, an important solution concept of stochastic games is the discounted value. Its existence and characterization go back to [Sha53]. The convergence of the discounted values as the discount rate vanishes was established by [BK76] using the theory of semi-algebraic sets. An alternative, probabilistic proof was obtained by [OB14]. The existence of the undiscounted value was established by [MN81].

The perturbation analysis goes back to [FV97], who provided a modulus of continuity for the discounted value function in terms of the parameters of the game (i.e., rewards, transition probabilities, and discount rate). [Sol03] obtained an analogous result for the undiscounted value function under certain conditions on the perturbation of the transition probabilities. Formulas for the discounted and undiscounted values were obtained by [AOB19], together with algorithms to compute them [OB20] whose runtime is polynomial in the number of pure stationary strategies. Robustness results for the undiscounted value function were provided, among others, by [NS10, Zil16a, COBZ21, OB18] and [OB22].

Of particular interest are the directional derivatives of the discounted and undiscounted value functions with respect to a given perturbation, referred to as the *marginal values* of the game. As already mentioned, this concept was introduced by [Mil56] in the context of matrix games and linear programming, and extended to compact-continuous games [RS01]. Notably, this result provides a formula for the marginal value of a stochastic game in the discounted case when only the rewards are perturbed.

Random Dynamic Games on Infinite Graphs A seminal result of Bewley and Kohlberg [BK76] shows that the sequence of n-stage values $(v_n)_{n\geq 1}$ converges in every finite stochastic game. This result has been extended in many directions, including partial observations [LS15, LS17, Sol22, SZ16]. When the state space is infinite, positive results are scarce (see [LR20, Zil24] for some recent advances) and counterexamples have been found [Zil16b]. When the action space is infinite, there are positive and negative results [Gar19].

Recently, a class of random zero-sum dynamic games on infinite graphs with vertices in $\mathbb{Z}^d$ has been introduced under the name of percolation games [GZ23]. In this model, as in usual dynamic games, each vertex is assigned a uniformly bounded reward, and these are known by the players before the game starts. The authors study a distribution over such games given by assigning i.i.d. random rewards to the vertices. Then, the asymptotic behavior of the random n-stage value, denoted by V_n, is studied. It is shown that, under the assumption that every action increases the projection of the state onto some axis, the random sequence $(V_n)_{n\geq 1}$ converges almost surely to a deterministic limit value. Moreover, they provide exponential concentration estimates for $(V_n)_{n\geq 1}$ around the limit. One important takeaway message of this result is that equipping the state space with a particular graph structure (e.g., $\mathbb{Z}^d$) and assuming that rewards have some probabilistic regularity (e.g., i.i.d. random variables) can ensure the existence of the limit of $(V_n)_{n\geq 1}$ even for an infinite graph.

Other distributions over games have been studied. For example, studying random zero-sum games on graphs and their long-term behavior goes beyond the game theory community [ARY21, ACSZ21, FPS23, HJM$^+$23]. Indeed, on the one hand, a class of random games has been used to solve a well-known open problem regarding Probabilistic Finite Automata [HMM19] (see [BKPR23] for an extension). On the other hand, the class of games studied in [GZ23] has served as a toy model for contributing to a well-studied problem in the theory of Partial Differential Equations (see [GZ23, Section 4] and [Zil17, DSZ24]).

1.3 Contributions

Partially Observable Markov Decision Processes Chapter 2 presents our contributions on POMDPs. The main result is that, for every POMDP with long-run average objectives and $\varepsilon > 0$, there exists a finite-memory strategy, i.e., implementable by a finite state automaton, that guarantees the value minus ε. Moreover, finite recall, i.e. decisions are defined using only the last actions, cannot achieve ε-approximations in general POMDPs. Many consequences of this result are discussed.

Perturbed Matrix Games Chapter 3 presents our contributions on Perturbed Matrix Games. We consider polynomial matrix games where each entry of the reward matrix is a polynomial on a single parameter $\varepsilon \geq 0$, and study how the value and the optimal strategies vary with the parameter. We introduce new properties to capture the existence of strategies that are robust to perturbations. More precisely, we consider two properties: (i) *value-positivity*, that asks whether there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in [0, \varepsilon_0]$ there is a strategy that guarantees the value of the unperturbed game, and (ii) *uniform value-positivity* asks whether there exists a fixed strategy that guarantees the value of the unperturbed game for all $\varepsilon \in [0, \varepsilon_0]$. Lastly, we consider the *functional form* problem, which consists of giving the analytical form of the value, and of an optimal strategy, as a function of ε in some right-neighborhood of zero.

The main result is polynomial time algorithms to check value-positivity and uniform value-positivity, and to solve the functional form problem. Given the connection between matrix games and linear programming, and polynomial matrix games and stochastic games, we discuss the consequences of applying our approach in these settings.

Perturbed Stochastic Games Chapter 4 presents our contributions on Perturbed Stochastic Games. The main results are the following.

1. A formula for the marginal discounted value of stochastic games with perturbations on any of its defining parameters, which extends to the compact-continuous case.
2. A formula for the limit of the marginal discounted value, as the discount rate vanishes.
3. A formula for the marginal (undiscounted) value, under mild regularity assumptions.

Random Dynamic Games on Infinite Graphs Chapter 5 presents our contributions on Random Dynamic Games on Infinite Graphs. The main results are the following. We extend the results in [GZ23] to a class of graphs that is fairly more general than $\mathbb{Z}^d$. In more detail, we introduce *directed games*, a class of games on acyclic directed graphs. On the one hand, under certain assumptions of weak transitivity and sub-exponential growth of the graph, we prove that $(V_n)_{n\geq 1}$ is exponentially concentrated around a given deterministic limit value (so, in particular, $(V_n)_{n\geq 1}$ converges almost surely) and relate the convergence rate to the expansion of the graph. On the other hand, we consider the infinite d-ary tree where each vertex has exactly $d \geq 2$ children and every edge is directed from the parent to the child. These graphs do not belong to the previous class of games due to their exponential expansion away from the root. In this case, we show a stronger double-exponential concentration of the random variables $(V_n)_{n\geq 1}$ around the limit.

CHAPTER 2

Finite-Memory Strategies in POMDPs with Long-Run Average Objectives

This chapter is based on [CSZ21], i.e., the following publication. Krishnendu Chatterjee, Raimundo Saona, and Bruno Ziliotto. Finite-Memory Strategies in POMDPs with Long-Run Average Objectives. *Mathematics of Operations Research*, 47(1):100–119, 2021.

Partially Observable Markov Decision Processes (POMDPs) is a standard model for dynamic systems with probabilistic and nondeterministic behaviour in uncertain environments. We prove that in POMDPs with long-run average objective, the decision-maker has approximately optimal strategies with finite memory. This implies notably that approximating the long-run value is recursively enumerable, as well as a characterization of the continuity property of the value with respect to the transition function.

2.1 Introduction

In a Partially Observable Markov Decision Process (POMDP), at each stage, the decision-maker chooses an action that determines, together with the current state, a stage reward and the distribution over the next state. The state dynamic is imperfectly observed by the decision-maker, who receives a stage signal on the current state before playing. Thus, POMDPs generalize the Markov Decision Process (MDP) model of Bellman [Bel57].

POMDPs are widely used in prominent applications such as in computational biology [DEKM98], software verification [vCH$^+$11], reinforcement learning [KLM96], to name a few. Even special cases of POMDPs, namely, probabilistic automata or blind MDPs, where there is only one signal, is also a standard model in several applications [Rab63, Paz71, Buk80].

In many of these applications, the duration of the problem is huge. Thus, considerable attention has been devoted to the study of POMDPs with long duration. A standard way is to consider the long-run objective criterion, where the total reward is the expectation of the inferior limit average reward (see [ABFG$^+$93] for a survey). The value for this problem is known to coincide with several classic definitions of long-run values (asymptotic value, uniform value, general uniform value, long-run average value, uncertain-duration process value [RSV02, Ren10, RV17, NS10, VZ16]) and has been characterized in [RV17]. We will simply name this common object *value*. Thus, strong results are available concerning the existence and characterization of the value.

This is in sharp contrast with the study of long-run optimal strategies. Indeed, before our work, little was known about the sophistication of strategies that approximate the value. It has been shown that:(i) stationary strategies approximate the value in MDPs [Bla62]; and (ii) belief-stationary strategies approximate the value in blind MDPs [RSV02] and POMDPs with an ergodic structure [Bor00].

Our main contributions are:

- *Strategy complexity.* We show that for every POMDP with long-run average objectives, for every $\varepsilon > 0$, there is a finite-memory strategy (i.e. generated by a finite state automaton) that achieves an expected reward within ε of the optimal value. In the case of blind MDP finite memory is equivalent to finite recall (i.e. decisions are defined using only the last actions), but finite recall cannot achieve ε-approximations in general POMDPs.
- *Computational complexity.* An important consequence of our above result is that the decision version of the approximation problem for POMDPs with long-run average objectives (see Definition 2.4.1) is *recursively enumerable (r.e.)* but not decidable. Our results on strategy complexity imply the recursively enumerable upper bound and the lower bound is a consequence of [MHC03].
- *Value property.* The long-run reward of a finite-memory strategy is robust upon small perturbations of the transition function, where the notion of perturbation over the transition function is defined as in Solan [Sol03] and Solan and Vieille [SV10]. This implies lower semi-continuity of the value function upon such small perturbations. This result is tight in the sense that there is an example with a discontinuous value function (see Example 2.5.4).

A natural question would be to ask for an upper bound on the size of the memory needed to generate ε-optimal strategies, in terms of the data of the POMDP. In fact, a previous *undecidability* result [MHC03] shows that such an upper bound can not exist (see Subsection 2.4.1). Thus, the existence of ε-optimal strategies with finite memory is, in some sense, the best possible result one can have in terms of strategy complexity.

Outline Section 2.2 presents the notation and basic definitions. Section 2.3 presents our main contributions. Section 2.4 explains the consequences of our result in terms of complexity and model robustness. Section 2.5 introduces examples used to prove negative results and to illustrate our techniques. Section 2.6 introduces two key lemmata, and shows that they imply Theorem 2.3.1. Section 2.7 proves one of the two lemmata and develops what we call *super-support* based strategies in details. Section 2.8 presents postponed proofs.

2.2 Preliminaries

We mostly use the following notation: (i) sets are denoted by calligraphic letters, e.g. $\mathcal{A}, \mathcal{H}, \mathcal{K}, \mathcal{S}$; (ii) elements of these sets are denoted by lowercase letters, e.g. a, h, k, s; and (iii) random elements with values in these sets are denoted by uppercase letters, e.g. A, H, K, S. For a set $\mathcal{C}$, denote $\Delta(\mathcal{C})$ the set of probability measure distributions over $\mathcal{C}$, and δ_c the Dirac measure at some element $c \in \mathcal{C}$. We will slightly abuse notation by not making a distinction between a probability measure (which can be evaluated on events) and its corresponding probability density (which can be evaluated on elements).

Consider a POMDP $\Gamma = (\mathcal{K}, \mathcal{A}, \mathcal{S}, q, g)$, with finite state space $\mathcal{K}$, finite action set $\mathcal{A}$, finite signal set $\mathcal{S}$, transition function $q \colon \mathcal{K} \times \mathcal{A} \to \Delta(\mathcal{K} \times \mathcal{S})$ and reward function $g \colon \mathcal{K} \times \mathcal{A} \to [0, 1]$.

Given $p_1 \in \Delta(\mathcal{K})$, called *initial belief*, the POMDP starting from p_1 is denoted by $\Gamma(p_1)$ and proceeds as follows:

- An initial state K_1 is drawn from p_1. The decision-maker knows p_1 but does not know K_1.
- At each stage $m \geq 1$, the decision-maker takes some action $A_m \in \mathcal{A}$. This action determines a stage reward $G_m := g(K_m, A_m)$, where K_m is the (random) state at stage m. Then, the pair (K_{m+1}, S_m) is drawn from $q(K_m, A_m)$. The next state is K_{m+1} and the decision-maker is informed of the signal S_m, but neither of the reward G_m nor of the state K_{m+1}.

At stage m, the decision-maker remembers all the past actions and signals, which is called *history before stage* m. Let $\mathcal{H}_m := (\mathcal{A} \times \mathcal{S})^{m-1}$ be the set of histories before stage m, with the convenient notation $(\mathcal{A} \times \mathcal{S})^0 := \{\emptyset\}$. A strategy is a mapping $\sigma \colon \cup_{m \geq 1} \mathcal{H}_m \to \mathcal{A}$. The set of strategies is denoted by Σ. The randomness introduced by the transition function, $q \colon \mathcal{K} \times \mathcal{A} \to \Delta(\mathcal{K} \times \mathcal{S})$, suggests that a history $h_m \in \mathcal{H}_m$ can occur under many sequences of states $(k_1, k_2, \ldots, k_{m-1})$. The infinite sequence $(k_1, a_1, s_1, k_2, a_2, s_2, \ldots)$ is called a *play*, and the set of all plays is denoted by Ω.

For $p_1 \in \Delta(\mathcal{K})$ and $\sigma \in \Sigma$, define $\mathbb{P}_\sigma^{p_1}$ the law induced by σ and the initial belief p_1 on the set of plays of the game $\Omega = (\mathcal{K} \times \mathcal{A} \times \mathcal{S})^{\mathbb{N}}$, and $\mathbb{E}_\sigma^{p_1}$ the expectation with respect to this law. For simplicity, identify $\mathcal{K}$ with the set of extremal points of $\Delta(\mathcal{K})$.

Let

$$\gamma_\infty^{p_1}(\sigma) := \mathbb{E}_\sigma^{p_1} \left(\liminf_{n \to +\infty} \frac{1}{n} \sum_{m=1}^{n} G_m \right) ,$$

and

$$v_\infty(p_1) := \sup_{\sigma \in \Sigma} \gamma_\infty^{p_1}(\sigma) .$$

The term $\gamma_\infty^{p_1}(\sigma)$ is the long-term reward given by strategy σ and $v_\infty(p_1)$ is the optimal long-term reward, called *value*, defined as the supremum long-term reward over all strategies.

Remark 2.2.1. It has been shown that v_∞ coincides with the limit of the value of the n-stage problem and λ-discounted problem, as well as the uniform value and weighted uniform value (see [RSV02, Ren10, RV17, VZ16]). In particular, we have:

$$\begin{aligned} v_\infty(p_1) &= \lim_{n \to +\infty} \sup_{\sigma \in \Sigma} \mathbb{E}_\sigma^{p_1} \left(\frac{1}{n} \sum_{m=1}^{n} G_m \right) \\ &= \lim_{\lambda \to 0} \sup_{\sigma \in \Sigma} \mathbb{E}_\sigma^{p_1} \left(\sum_{m \geq 1} \lambda (1 - \lambda)^{m-1} G_m \right) \\ &= \sup_{\sigma \in \Sigma} \liminf_{n \to +\infty} \mathbb{E}_\sigma^{p_1} \left(\frac{1}{n} \sum_{m=1}^{n} G_m \right) . \end{aligned}$$

Remark 2.2.2. In the literature, the concept of strategy that we defined is often called *pure strategy*, by contrast with *behavior strategies* that use randomness by allowing strategies of the form $\sigma \colon \cup_{m \geq 1} \mathcal{H}_m \to \Delta(\mathcal{A})$. By Kuhn's theorem, enlarging the set of pure strategies to behaviour strategies does not change v_∞ (see [VZ16, Fei96]), and thus does not change our results.

Definition 2.2.3 (Blind MDP)**.** A POMDP is called *blind MDP* if the signal set is a singleton.

Note that in a blind MDP, signals do not convey any relevant information. Therefore, a strategy is simply an infinite sequence of actions $(a_1, a_2, \dots) \in \mathcal{A}^{\mathbb{N}}$. We define several classes of strategies. Recall that $\Gamma(p_1)$ is the POMDP Γ starting from p_1, which is known to the player.

Definition 2.2.4 (ε-optimal strategy)**.** Let $p_1 \in \Delta(\mathcal{K})$ and $\varepsilon > 0$. A strategy $\sigma \in \Sigma$ is *ε-optimal* in $\Gamma(p_1)$ if

$$\gamma_\infty^{p_1}(\sigma) \geq v_\infty(p_1) - \varepsilon \,.$$

Definition 2.2.5 (finite-memory strategy)**.** A strategy σ is said to have *finite memory* if it can be modeled by a finite-state transducer. Formally, $\sigma = (\sigma_u, \sigma_a, \mathcal{M}, m_0)$, where $\mathcal{M}$ is a finite set of memory states, m_0 is the initial memory state, $\sigma_a : \mathcal{M} \to \mathcal{A}$ is the action selection function and $\sigma_u \colon \mathcal{M} \times \mathcal{A} \times \mathcal{S} \to \mathcal{M}$ is the memory update function.

Definition 2.2.6 (Finite-recall strategy)**.** A strategy σ is said to have *finite recall* if there exists a constant $M > 0$ such that for all $h_M \in \mathcal{H}_M$, and for all $m > M$ and $h_{m-M} \in \mathcal{H}_{m-M}$, we have that $\sigma(h_{m-M}, h_M)$ does not depend on h_{m-M}.

Remark 2.2.7. For blind MDPs, finite-recall and finite-memory strategies coincide with the set of *eventually periodic strategies*: a strategy $\sigma = (a_1, a_2, \ldots)$ is eventually periodic if there exists $T \geq 1$ and $N \geq 1$ such that for all $m \geq N$, $a_{m+T} = a_m$. This property does not extend to general POMDPs (see Proposition 2.3.4): any finite-recall strategy has finite-memory, but the inverse is not true.

Remark 2.2.8. Finite-memory strategies and finite-recall strategies have been investigated in the Shapley zero-sum stochastic game model [Sha53]. In this framework, none of these strategies is enough to approximate the value, and a long-standing open problem is whether finite-memory strategies *with a clock* are good enough (see [HIJN23, HIJN21] for more details on this topic).

2.3 Main Contributions

Our main contribution is the following theorem.

Theorem 2.3.1. *For every POMDP Γ, initial belief p_1 and $\varepsilon > 0$, there exists an ε-optimal finite-memory strategy in $\Gamma(p_1)$.*

Remark 2.3.2. A previous complexity result [MHC03] shows that the size of the memory can not be bounded from above in terms of the data of the POMDP (see Subsection 2.4.1).

Corollary 2.3.3. *For every blind MDP Γ, initial belief p_1 and $\varepsilon > 0$, there exists an ε-optimal finite-memory strategy in $\Gamma(p_1)$, and thus the strategy is eventually periodic and has finite recall.*

Lastly, finite-recall is not enough to ensure ε-optimality in general POMDPs.

Proposition 2.3.4. *There exists a POMDP, and $\varepsilon > 0$, with no ε-optimal finite-recall strategy.*

2.4 Consequences of the Results

We explain several consequences of our results.

2.4.1 Complexity

Decidability A decision problem consists in deciding between two options given an input (accepting or rejecting) and its complexity is characterized by Turing machines. A Turing machine takes an input and, if it halts, it either accepts or rejects it. If it halts for all possible inputs in a finite number of steps, then the Turing machine is considered an algorithm. An algorithm solves a decision problem if it takes the correct decision for all inputs. The class of decision problems that are solvable by an algorithm is called *decidable*. Two natural generalizations of decidable problems are: *recursively enumerable (r.e.)* and *co-recursively enumerable (co-r.e.)*. The decision problems in r.e. (resp., co-r.e.) are those for which there is a Turing machine that accepts (resp., rejects) every input that should be accepted (resp., rejected) according to the problem, but, on other inputs, it needs not to halt.

Notice that the class of decidable problems is the intersection of r.e. and co-r.e. In this work, the algorithmic problem of interest is the following.

Definition 2.4.1 (Decision version of approximating the value)**.** Let $p_1 \in \Delta(\mathcal{K})$. Given $x \in [0,1]$, $\varepsilon > 0$ such that $v_\infty(p_1) > x+\varepsilon$ or $v_\infty(p_1) < x-\varepsilon$, the problem consists in deciding which one is the case: to accept means to prove that $v_\infty(p_1) > x+\varepsilon$ holds, while to reject means to prove the opposite.

Previous Results and Implication of our Results It is known that the decision version of the approximation problem is not decidable [MHC03] (even for blind MDPs). However, the complexity characterization has been open. Thanks to Theorem 2.3.1, we can design a Turing machine that accepts every input that should be accepted for this problem.

Consider playing a finite-memory strategy σ. Then, the dynamics of the game can be described by a finite Markov chain. Therefore, the reward obtained by playing σ (i.e. $\gamma_\infty^{p_1}(\sigma)$) can be deduced from its stationary measure, which can be computed in polynomial time by solving a linear programming problem [FV97, Section 2.9, page 70]. Our protocol checks the reward given by every finite-memory strategy to approximate the value of the game $v_\infty(p_1)$. By Theorem 2.3.1, if $v_\infty(p_1) > x+\varepsilon$ holds, a finite-memory strategy that achieves a reward strictly greater than $(x+\varepsilon)$ will be eventually found and our protocol will accept the input. On the other hand, if $v_\infty(p_1) < x-\varepsilon$, the protocol will never find out that this is the case because there are infinitely many finite-memory strategies, so it will not halt. Thus, our result establishes that the approximation version of the problem is in r.e., and the previously known results imply that the problem is not decidable. Formally, we have the following result.

Corollary 2.4.2. *The decision version of approximating the value is r.e. but not decidable.*

Remark 2.4.3. The former paragraph shows that no upper bound on the size of the memory used by ε-optimal strategies can be proved. Indeed, if such a bound existed, one could modify the previous algorithm in the following way: reject the input if every finite-memory strategy of size lower than the bound has been enumerated. This would imply that the decision version of approximating the value is decidable, which is a contradiction.

2.4.2 Comparison of Objectives

In this section, we contrast our results with other natural objectives.

Recall that the value of $\Gamma(p_1)$ is defined as

$$v_\infty(p_1) = \sup_{\sigma \in \Sigma} \mathbb{E}_\sigma^{p_1} \left(\liminf_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} G_m \right) .$$

We say this is a *liminf-average* objective. Consider replacing $\liminf_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} G_m$ by: (i) $\limsup_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} G_m$, which we call *limsup-average* objective; (ii) $\limsup_{n\to\infty} G_n$, which we call *limsup* objective.

Proposition 2.4.4. *For both limsup-average and limsup objective, there exists a POMDP, and $\varepsilon > 0$, with no ε-optimal finite-memory strategy.*

This negative result, proved in Section 2.5.1, does not imply any computational complexity characterization for the limsup-average or limsup objective, and whether the approximation of the value problem for limsup-average objectives is recursively enumerable remains open. However, it shows that any approach based on finite-memory strategies cannot establish recursively enumerable bounds for the approximation problem.

Let us focus on the limsup objective. Limsup objective is arguably simpler than the liminf-average objective and, to formalize this statement, we can compare the complexity of the objects themselves irrespective of any particular context or model (such as POMDPs). The *Borel hierarchy* describes the complexity of an objective by the number of quantifier alternations needed to describe it. Its construction is similar to that of the Borel σ-algebra, or σ-field, and is defined as follows.

Definition 2.4.5 (Borel hierarchy)**.** Consider $h_m \in \mathcal{H}_m = (\mathcal{A} \times \mathcal{S})^{m-1}$ a finite history of the game. The cylinder set generated by h_m is given by $\{h_m\} \times (\mathcal{A} \times \mathcal{S})^{\mathbb{N}}$. Finite intersection, unions and complements of the cylinder sets generated by finite histories form the first level in the hierarchy. Countable unions of the first level form Σ_1 and countable intersections form Π_1. The next level is always obtained from the previous one: countable unions of Π_i give Σ_{i+1} and countable intersections of Σ_i give Π_{i+1}. The nested sequence of family of problems $\{\Sigma_i \cup \Pi_i\}_{i\geq 1}$ is called *Borel hierarchy*.

For example, limsup objective can be described as countable intersection of countable unions of rewards: given a family of sets $(\mathcal{C}_n)_{n\geq1}$, $\limsup_{n\to\infty} \mathcal{C}_n = \cap_{n\geq1} \cup_{m\geq n} \mathcal{C}_m$. The formal result is the following (see [Cha07]).

Proposition 2.4.6. *The limsup objective is Π_2-complete, i.e. complete for the second level of the Borel hierarchy, whereas the liminf-average objective is Π_3-complete, i.e. complete for the third level of the Borel hierarchy.*

While the notion of Borel hierarchy characterizes the topological complexity for objectives, a similar notion of *Arithmetic hierarchy* characterizes the computational complexity for decision problems.

Definition 2.4.7 (Arithmetic hierarchy)**.** Denote Σ_0^1 the class of r.e. problems and Π_0^1 the co-r.e. problems. For $i > 1$, define Σ_0^i as the class of problems solved by Turing machines with access to oracles for Π_0^{i-1} and Π_0^i is similarly defined with oracles for Σ_0^{i-1}. The nested sequence of family of problems $\{\Sigma_0^i \cup \Pi_0^i\}_{i \geq 1}$ is called *Arithmetic hierarchy*.

By Corollary 2.4.2, we have that POMDPs with a liminf-average objective is in $\Sigma_0^1 \setminus (\Sigma_0^1 \cap \Pi_0^1)$. On the other hand, it was shown in [BGB12, BG09] that POMDPs with limsup objective with boolean rewards is Σ_0^2-complete.

We conclude this section with a summary chart contrasting liminf-average and limsup objectives. The surprising result is the complexity switch: limsup objective has lower Borel hierarchy complexity but higher Arithmetic hierarchy complexity in the context of POMDPs.

Objective comparison in POMDPs		
Objective	Borel hierarchy	Arithmetic Hierarchy
limsup	Π_2-complete	Σ_0^2-complete
liminf-average	Π_3-complete	$\Sigma_0^1 \setminus (\Sigma_0^1 \cap \Pi_0^1)$

Figure 2.1: Objective comparison in POMDPs

2.4.3 Robust ε-Optimal Strategies

Consider a POMDP $\Gamma = (\mathcal{K}, \mathcal{A}, \mathcal{S}, q, g)$. It is well known that the value function is continuous with respect to perturbations of the reward function g and the initial belief p_1. Now, we show a robustness result concerning the transition function q.

In applications, just as in any stochastic model, the structure of the model is decided first, and then the specific probabilities are either estimated or fixed. The values of transition probabilities are approximations: an ε-perturbation of these probabilities are not expected to have an impact on the modelling. In our setting, the transitions are encoded in the function $q \colon \mathcal{K} \times \mathcal{A} \to \Delta(\mathcal{K} \times \mathcal{S})$ and we would expect some robustness against perturbations of the values it takes.

The notion of perturbation over q is measured as in Solan [Sol03] and Solan and Vieille [SV10], where perturbations in each transition probability are measured as relative differences, not additive differences. Formally, define the semimetric

$$d(q, q') = \max_{\substack{k \in \mathcal{K},\ a \in \mathcal{A} \\ k' \in \mathcal{K},\ s \in \mathcal{S}}} \left\{ \frac{q(k,a)(k',s)}{q'(k,a)(k',s)}, \frac{q'(k,a)(k',s)}{q(k,a)(k',s)} \right\} - 1 \,.$$

Under this notion, and taking q and q' close to each other, we can prove the existence of strategies which are approximately optimal for the POMDP corresponding to q and perform almost as well when they are applied to the POMDP corresponding to q'. To formally state this notion of robustness, let us give the following definition.

Definition 2.4.8 (Robust strategies)**.** Given a POMDP Γ with transition function q, an initial belief $p_1 \in \Delta(\mathcal{K})$, we say that σ is a *robust strategy* for $\Gamma(p_1)$ if the following condition holds: $\forall \eta > 0\ \exists \delta > 0$ such that

$$d(q, q') \leq \delta \quad \Rightarrow \quad \gamma'^{p_1}_\infty(\sigma) \geq \gamma^{p_1}_\infty(\sigma) - \eta \,,$$

where γ_∞ is the long-term reward in Γ and γ'_∞ is the long-term reward in $\Gamma' = (\mathcal{K}, \mathcal{A}, \mathcal{S}, g, q')$.

Lemma 2.4.9. *Any finite-memory strategy is robust. Thus, in any POMDP and for any $\varepsilon > 0$, there exists a robust ε-optimal finite-memory strategy.*

Proof. Let $\Gamma = (\mathcal{K}, \mathcal{A}, \mathcal{S}, g, q)$ be a POMDP. Consider $\sigma = (\sigma_u, \sigma_a, \mathcal{M}, m_0)$ a finite-memory strategy for $\Gamma(p_1)$. Playing σ from p_1 induces a Markov Chain $(Y_n)_{n\geq 1}$ on $\mathcal{K} \times \mathcal{S} \times \mathcal{M}$. Define $\tilde{g} \colon \mathcal{K} \times \mathcal{M} \to [0,1]$ by $\tilde{g}(k,m) := g(k, \sigma_a(m))$.

Now, consider the 0-Player stochastic game with reward $\tilde{g}$ and transitions given by the kernel of the Markov Chain $(Y_n)_{n\geq 1}$. Let $s_0 \in \mathcal{S}$ be any signal. By definition, for any $k \in \mathcal{K}$, the value of this stochastic game starting from (k, s_0, m_0) coincides with $\gamma_\infty^k(\sigma)$. Using [Sol03][Theorem 6, page 841], we deduce that

$$\gamma_\infty'^k(\sigma) \geq \gamma_\infty^k(\sigma) - 4|\mathcal{K}||\mathcal{S}||\mathcal{M}|d(q,q') \,.$$

Integrating k over p_1 yields

$$\gamma_\infty'^{p_1}(\sigma) \geq \gamma_\infty^{p_1}(\sigma) - 4|\mathcal{K}||\mathcal{S}||\mathcal{M}|d(q,q') \,.$$

Taking $\delta = \eta(4|\mathcal{K}||\mathcal{S}||\mathcal{M}|)^{-1}$, we conclude that σ is a robust strategy. By Theorem 2.3.1, for all $\varepsilon > 0$, there exists an ε-optimal finite-memory strategy, which is thus robust. □

Corollary 2.4.10. *Let $\mathcal{K}, \mathcal{A}, \mathcal{S}$ be finite sets, $g \colon \mathcal{K} \times \mathcal{A} \to \mathbb{R}$ a reward function, and $p_1 \in \Delta(\mathcal{K})$ an initial belief. The mapping from $(\Delta(\mathcal{K} \times \mathcal{S})^{\mathcal{K}\times\mathcal{A}}, d)$ to $\mathbb{R}$ that maps each transition function q to the value at p_1 of the POMDP $(\mathcal{K}, \mathcal{A}, \mathcal{S}, g, q)$ is lower semi-continuous.*

Proof. Let $q \in \Delta(\mathcal{K} \times \mathcal{S})^{\mathcal{K}\times\mathcal{A}}$. Let $\Gamma = (\mathcal{K}, \mathcal{A}, \mathcal{S}, q, g)$. By the previous lemma, for all $\varepsilon > 0$, there exists σ_ε a robust ε-optimal strategy in $\Gamma(p_1)$. Take $\eta = \varepsilon$, by robustness of σ_ε, there exists $\delta > 0$ such that, for all $q' \in \Delta(\mathcal{K} \times \mathcal{S})^{\mathcal{K}\times\mathcal{A}}$, we have that if $d(q,q') \leq \delta$, then $\gamma_\infty'^{p_1}(\sigma_\varepsilon) \geq \gamma_\infty^{p_1}(\sigma_\varepsilon) - \varepsilon$. Also, by ε-optimality of σ_ε, we have that $\gamma_\infty^{p_1}(\sigma_\varepsilon) \geq v_\infty(p_1) - \varepsilon$. Then,

$$v'_\infty(p_1) \geq \gamma_\infty'^{p_1}(\sigma_\varepsilon) \geq v_\infty(p_1) - 2\varepsilon \,.$$

Taking $\varepsilon \to 0$, we conclude that

$$\liminf_{q' \to q} v'_\infty(p_1) \geq v_\infty(p_1) \,,$$

and thus v_∞ is lower semi-continuous with respect to q. □

Lower semi-continuity of the value function is the best result one can achieve in the following sense.

Proposition 2.4.11. *There is a POMDP such that the mapping from $(\Delta(\mathcal{K} \times \mathcal{S})^{\mathcal{K}\times\mathcal{A}}, d)$ to $\mathbb{R}$ that maps each transition function q to the value at p_1 of the POMDP $(\mathcal{K}, \mathcal{A}, \mathcal{S}, g, q)$ is discontinuous.*

2.5 Examples

In this section, we introduce examples to prove negative results (Propositions 2.3.4, 2.4.4 and 2.4.11) and to illustrate our techniques later on.

2.5.1 Negative Results

Let us prove Propositions 2.3.4 and 2.4.4 by presenting an example for each statement.

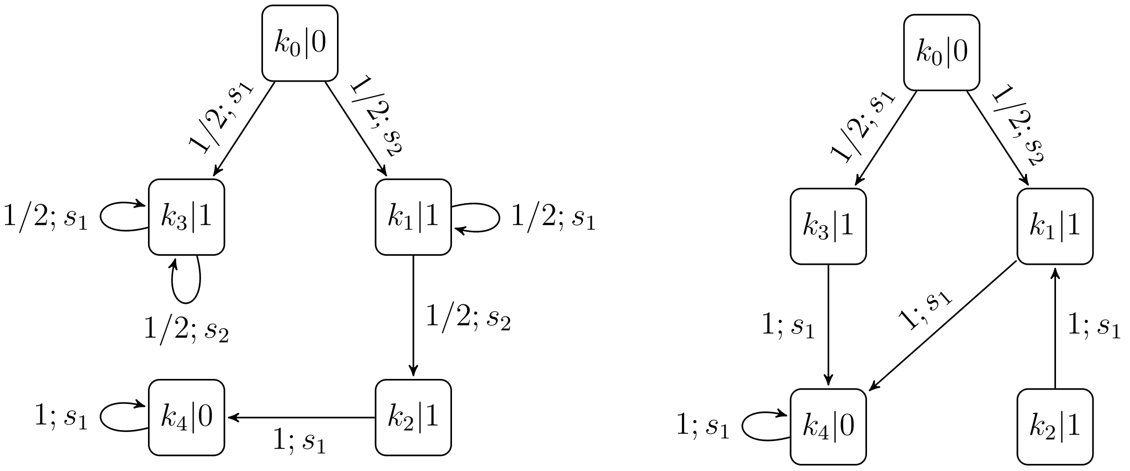
Proof of Proposition 2.3.4

We will prove that there exists a POMDP and $\varepsilon > 0$ with no ε-optimal finite-recall strategy by an explicit construction. Recall that a strategy has finite recall if it uses only a finite number of the last stages in the current history to decide the next action (see Definition 2.2.6). Therefore, our construction should have the property that, for any finite-recall strategy, there is a pair of finite histories such that:

1. The last stages are identical, i.e., the player did the same actions and received the same signals in the last part of both histories (but the starting point was different).
2. Taking the same decision in both histories leads to losing some reward that can not be compensated in the long-run.
3. The previous loss does not decrease to zero by increasing the amount of memory.

Example 2.5.1. Consider the POMDP $\Gamma = (\mathcal{K}, \mathcal{A}, \mathcal{S}, q, g)$ with five states: $k_0, k_1, \ldots, k_4$. The initial state is k_0 and players know it (formally, the initial belief is δ_{k_0}). The state k_4 is an absorbing state from where it is impossible to get out and rewards are zero. The states k_1 and k_2 form a sub-game where the optimal strategy is trivial. This is the same for the state k_3. From k_0 a random initial signal is given indicating which sub-game the state moved to. The key idea is that there is an arbitrarily long sequence of actions and signals which can be gotten in both sub-games, but the optimal strategy behaves differently in each of them. Therefore, to forget the initial signal of the POMDP leads to at most half of the optimal value.

Figure 2.2 is a representation of Γ: first under action a and then action b. Each state is followed by the corresponding reward, and the arrows include the probability for the corresponding transition along with the signal obtained.



(a) Action a (b) Action b

Figure 2.2: Finite recall is not enough for POMDPs

The sub-game of k_1 and k_2 has a unique optimal strategy: play action a until receiving signal s_2, then play action b once and repeat. The value of this sub-game is 1 and deviating from the prescribed strategy would lead to a long-run reward of 0. Similarly, the value of the sub-game of k_3 has a unique optimal strategy: to always play action a. Again, the value of this sub-game is 1 and playing any other strategy leads to a long-run reward of 0.

By the previous discussion, the value of this game starting from k_0 is 1. On the other hand, the maximum value obtained by strategies with finite recall is only $1/2$, by playing, for example, always action a. Finite-recall strategies achieve at most $1/2$ because, no matter how much finite recall there is, by playing the game the decision-maker faces a history of having played always action a and always receiving a signal s_1, except for the last signal which is s_2. Then, if action b is played, the second sub-game is lost; if action a is played, the first sub-game is lost. That is why, for any $0 < \varepsilon < 1/2$, there is no ε-optimal finite-recall strategy for this POMDP.

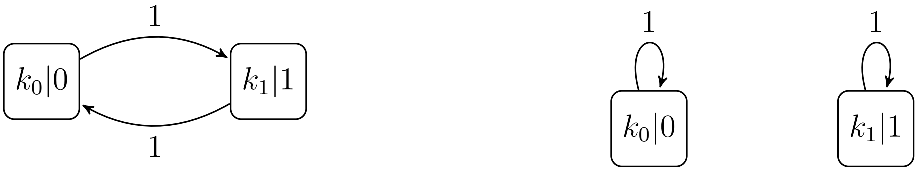
Proof of Proposition 2.4.4

We will show that for the limsup-average and limsup objectives there is a blind MDP where there is no ε-optimal finite-memory strategy. For both cases, the example is constructed with the following idea in mind. To achieve the optimal value, the decision-maker needs to play an action a_1 for some period, then play another action a_2 and repeat the process. The key is to require that the length of the period gets longer as the game progresses. This kind of strategy can not be achieved with finite-memory strategies.

For the limsup-average objective, the blind MDP example is due to Venel and Ziliotto [VZ21] and is presented below.

Example 2.5.2. Consider two states k_0 and k_1 and the player receives a reward only when the state is k_1. To reach k_1, the decision-maker can play action *change* and move between the two states. By playing action *wait*, the state does not change.

Figure 2.3 is a representation of the game.



(a) Action *change*

(b) Action *wait*

Figure 2.3: Finite recall is not enough for POMDPs

Consider the initial belief $p_1 = \frac{1}{2} \cdot \delta_{k_0} + \frac{1}{2} \cdot \delta_{k_1}$, the uniform distribution. It is easy to see that finite-memory strategies (or equivalently finite-recall strategies) can not achieve more than $1/2$. On the other hand, the value of this game with the *limsup-average* objective is 1, and is guaranteed by the following strategy:

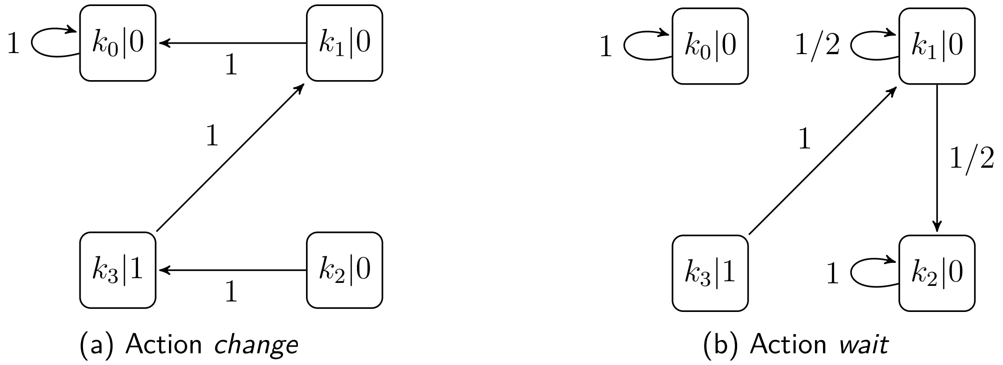
$$\sigma = (wait)^{2^{0^2}} (change)(wait)^{2^{1^2}} \cdots (change)(wait)^{2^{N^2}} \cdots$$

Hence, finite-memory strategies do no guarantee any approximation for POMDPs with *limsup-average* objective.

For the *limsup* objective, the blind MDP example is the following.

Example 2.5.3. Consider four states (k_0, k_1, k_2 and k_3) and two actions: *wait* (w) and *change* (c). The initial state is k_1 and players know it. In k_1, if w is played, then the state moves to k_2 with probability $1/2$ and stays with probability $1/2$; if c is played, then the absorbing state k_0 is reached. From state k_2, if we play w, we stay in the same state; if we play c, we move to state k_3. From k_3, the only state that has a positive reward, if we play any action, we return to the initial state k_1.

Figure 2.4 is a representation of the game.



(a) Action *change*

(b) Action *wait*

Figure 2.4: Finite-memory is not enough for *limsup* objective

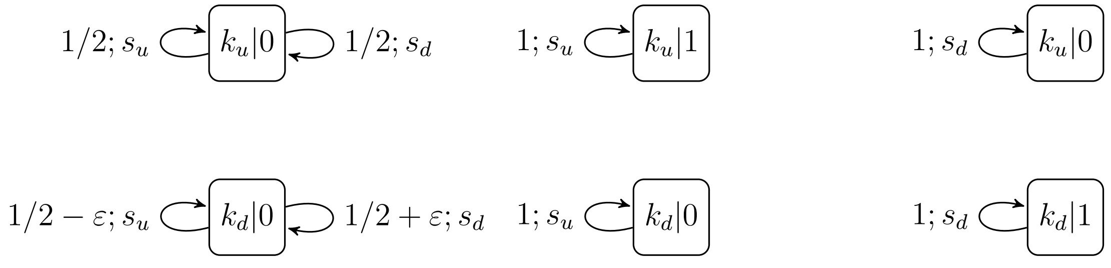
In this blind MDP (see [BG09, BGB12]), for the *limsup* objective, for any $\varepsilon > 0$, there is an infinite-memory strategy that guarantees $1 - \varepsilon$, so the value of the game is 1. On the other hand, applying any finite-memory strategy (or equivalently finite-recall strategies) yields a *limsup* reward of 0. Hence, finite-memory strategies do no guarantee any approximation for POMDPs with *limsup* objective.

Proof of Proposition 2.4.11

We will show that there is a POMDP with discontinuous value with respect to the transition function. The idea is to have two possible scenarios where signals are slightly different. By analyzing a long sequence of signals, the player is able to identify which is the scenario of the current state and so take a better strategy. The following example considers a transition function parameterized by $\varepsilon \geq 0$.

Example 2.5.4. Consider two states (k_u, k_d) and three actions: *up* (a_u), *down* (a_d) and *wait* (a_w). Signals are relevant only for action a_w: under a non-symmetric transition function, they inform about the underlying state. More concretely, there are two signals s_u and s_d. Playing actions a_u or a_d will give signal s_u or s_d respectively, adding no information. Playing action a_w leads to signals s_u and s_d with slightly different probabilities if the state is k_u or k_d. In terms of actions and rewards, a_u leads to positive reward only if the state is k_u, similarly, a_d leads to positive reward only if the state is k_d. Finally, a_w leads to null reward in both states. Figure 2.5 is a representation of this POMDP where transitions are specified.

Consider an initial belief $p_1 = 1/2 \cdot \delta_{k_u} + 1/2 \cdot \delta_{k_d}$. If $\varepsilon = 0$, the value is $1/2$, achieved for example by the constant strategy $\sigma \equiv a_u$. Note that playing action a_w leads to no information since the random signal the player receives is independent of the underlying state. In contrast, if $\varepsilon > 0$, playing action a_w reveals information about the underlying state. If action a_w is played sufficiently many times, the player can estimate the state by comparing the number of signals



(a) Action a_w (b) Action a_u (c) Action a_d

Figure 2.5: Discontinuous value POMDP

s_d against s_u: if s_d appear more than s_u, then it is more probable that the underlying state is k_d. Therefore, by playing a_w, the player can estimate the state with increasing probability. That is why the value for $\varepsilon > 0$ is 1. This proves that this POMDP is discontinuous with respect to the transition function since

$$1 = \liminf_{\varepsilon \to 0} v_\infty(p_1 \mid \varepsilon) > v_\infty(p_1 \mid \varepsilon = 0) = \frac{1}{2}\,.$$

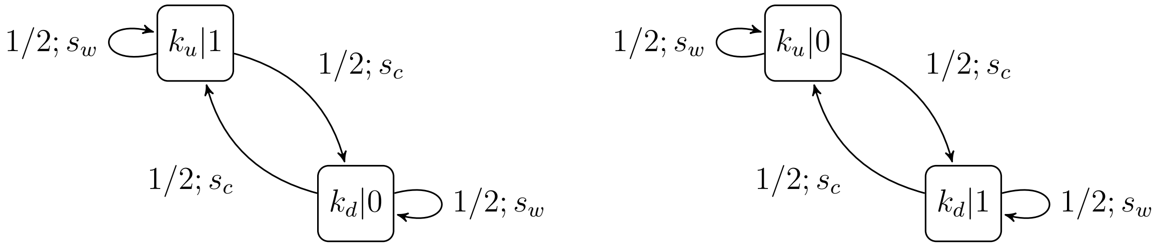
2.5.2 Illustrative Examples

We show an example of POMDP that will be analyzed in Section 2.7.2 in light of our technique. This example comes in two variants differing in sophistication.

Simple Version

Let us explain the easiest version.

Consider two states (k_u, k_d) and two actions: *up* (a_u) and *down* (a_d). All transitions are possible (including loops) and they do not depend on the action. Signals inform the player when the state changes. In terms of actions and rewards, by playing a_u the player obtains a reward of 1 only if the current state is k_u. Similarly, by playing a_d the player obtains a reward of 1 only if the state is k_d. Figure 2.6 is a representation of the game with specific transition probabilities.



(a) Action a_u (b) Action a_d

Figure 2.6: Simple POMDP

Consider an initial belief $p_1 = 1/4 \cdot \delta_{k_u} + 3/4 \cdot \delta_{k_d}$. During a play, the decision-maker can have two beliefs, $1/4 \cdot \delta_{k_u} + 3/4 \cdot \delta_{k_d}$ or $3/4 \cdot \delta_{k_u} + 1/4 \cdot \delta_{k_d}$, because the signals notify when there has been a change. The value of this game is $3/4$. An optimal strategy is to play action a_d until getting a signal s_c, then playing action a_u until getting a signal s_c, and repeat.

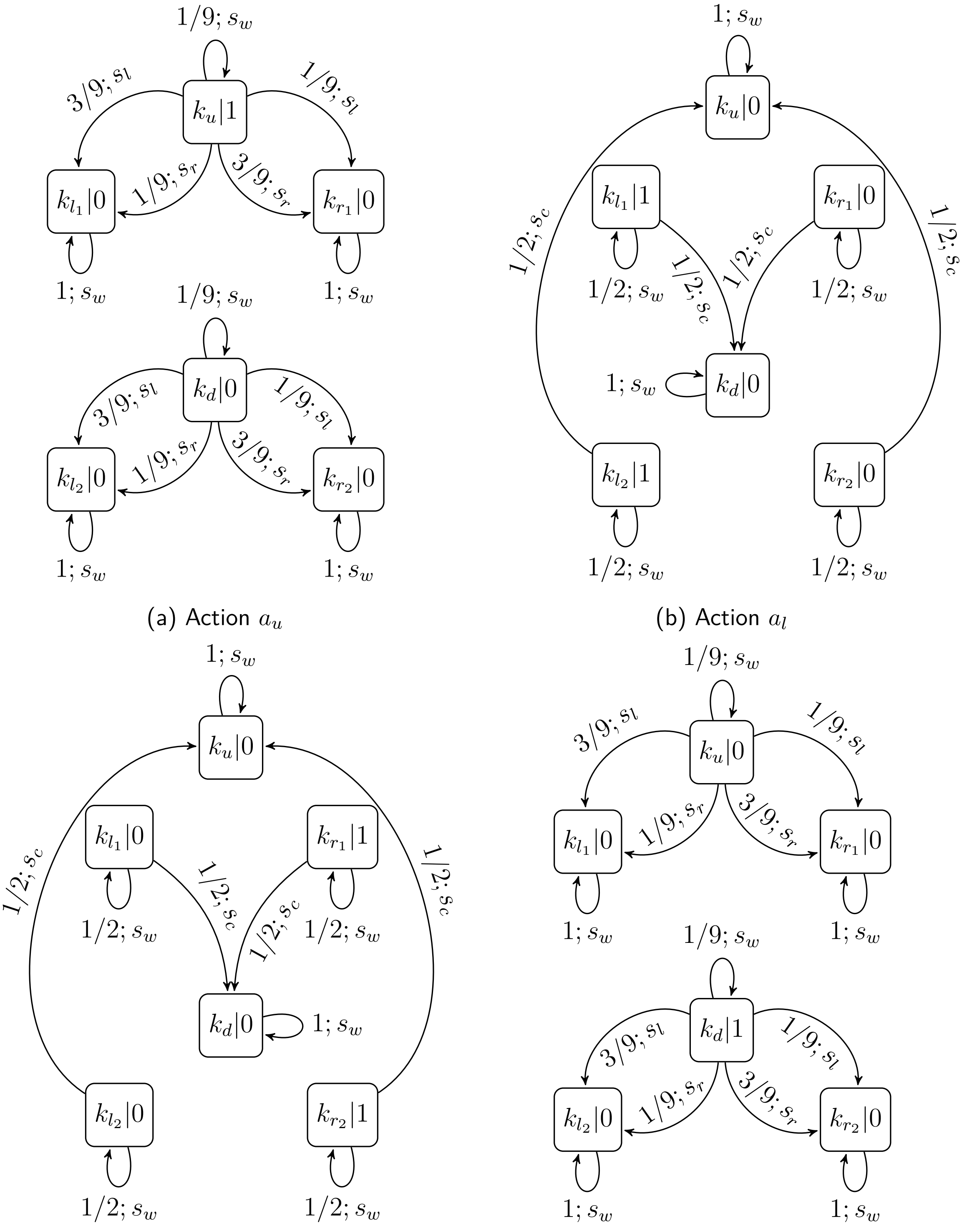
Involved Version

Let us go to the more complex version. Now the transition between the two extremes includes more states, instead of being a direct jump.

Consider six states and four actions: *up* (a_u), *down* (a_d), *left* (a_l) and *right* (a_r). States can be separated into two groups: *extremes* (k_u and k_d) and *transitional* ($k_{l_1}, k_{r_1}, k_{l_2}$ and k_{r_2}). Furthermore, transitional states can be divided into two groups: *left states* (k_{l_1} and k_{l_2}) and *right states* (k_{r_1} and k_{r_2}). Transitions are from extreme states to transitional states and from transitional to extremes. More precisely, excluding loops, only the following transitions are possible: from k_u to either k_{l_1} or k_{r_1}, then from these two to k_d, from k_d to either k_{l_2} or k_{r_2} and then back to k_u. Signals are such that the player knows: (i) the state changed to an extreme state, or (ii) the state changed to a transitional state and the new state is with higher probability a left state or a right state. In terms of actions and rewards, each action has an associated set of states in which the reward is 1 and the rest is 0: by playing a_u the reward is 1 only if the current state is k_u, playing a_d rewards only state k_d, a_l rewards states k_{l_1} and k_{l_2}, and a_r rewards states k_{r_1} and k_{r_2}. Figure 2.7 is a representation of this game with specific transition probabilities. Consider an initial belief $p_1 = 1/4 \cdot \delta_{k_u} + 3/4 \cdot \delta_{k_d}$. The value of the game is $21/32$. An optimal strategy is given by playing action a_d until getting a signal s_l or s_r. If the decision-maker got signal s_l, then play action a_l, otherwise, play action a_r. Repeat action a_l or a_r until getting the signal s_c. Then, play a_u until getting a signal s_l or s_r. When this happens, play a_l or a_r accordingly until getting signal s_c. And so, repeat the cycle.

The belief dynamic under this optimal strategy is the following. The initial belief is p_1, supported in the extreme states. By getting a signal s_w, the belief does not change. By getting signal s_l, the weight on k_u distributes between states k_{l_1} and k_{r_1} in a proportion $3 : 1$ and the weight on k_d distributes between k_{l_2} and k_{r_2} in the same way. By getting signal s_r, the distribution is similar, but the role of left states are interchanged with right states. Once the belief is in the transitional states, by playing the respective action (either a_l or a_r), the belief does not change while receiving signal s_w. Upon receiving the signal s_c, the new belief is $3/4 \cdot \delta_{k_u} + 1/4 \cdot \delta_{k_d}$. By symmetry of the POMDP, the dynamic is then similar until getting signal s_c for a second time. At that time, the belief is equal to the initial distribution, namely $1/4 \cdot \delta_{k_u} + 3/4 \cdot \delta_{k_d}$.

Remark 2.5.5. For the decision-maker to have a finite-memory strategy, some quantity with finitely many options must be updated over time. A tentative idea is to compute the posterior belief, but it can take infinitely many values. In this example, using a belief partition is enough to encode an optimal strategy. In general, it is an open question if a belief partition is sufficient to achieve ε-optimal strategies.



(a) Action a_u

(b) Action a_l

(c) Action a_r

(d) Action a_d

Figure 2.7: Complex POMDP

2.6 Structure of the Proof

In this section, we introduce two key lemmas and derive from them the proof of Theorem 2.3.1. We first define the history at stage m, which is all the information the decision-maker has at stage m.

Definition 2.6.1 (m-stage history)**.** Given a strategy $\sigma \in \Sigma$ and an initial belief p_1, denote the (random) history at stage m by

$$H_m \coloneqq ((A_1, S_1), (A_2, S_2), \dots, (A_{m-1}, S_{m-1}))\,.$$

The random variable H_m takes values in $\mathcal{H}_m = (\mathcal{A} \times \mathcal{S})^{m-1}$.

Recall that we denote the state at stage m by K_m, which takes values in $\mathcal{K}$; the signal at stage m by S_m, which takes values in $\mathcal{S}$; and the action at stage m by A_m, which takes values in $\mathcal{A}$. Note that the history at stage m does not contain direct information about the states $K_1, \dots, K_m$.

The belief of the player at stage m plays a key role in the study of POMDPs, and we formally define it as follows.

Definition 2.6.2 (m-stage belief)**.** Given a strategy $\sigma \in \Sigma$ and an initial belief p_1, denote the belief at stage m by P_m, which is given by, for all $k \in \mathcal{K}$,

$$P_m(k) \coloneqq \mathbb{P}^{p_1}_{\sigma}(K_m = k \mid H_m)\,.$$

For fixed σ and p_1, one can use Bayes rule to compute P_m. To avoid heavy notations, we omit the dependence of P_m on σ and p_1. For $p \in \Delta(\mathcal{K})$, denote the support of p by $\operatorname{supp}(p)$, which is the set of $k \in \mathcal{K}$ such that $p(k) > 0$.

The first ingredient of the proof of Theorem 2.3.1 is the following lemma.

Lemma 2.6.3. *For any initial belief p_1 and $\varepsilon > 0$, there exists $m_\varepsilon \geq 1$, $\sigma^\varepsilon \in \Sigma$ and a (random) belief $P^* \in \Delta(\mathcal{K})$ (which depends on the history before stage m_ε) such that:*

1.
$$\mathbb{P}^{p_1}_{\sigma^\varepsilon}(\|P_{m_\varepsilon} - P^*\|_1 \leq \varepsilon) \geq 1 - \varepsilon\,.$$

2. *There exists $\sigma \in \Sigma$, which depends on P^*, such that for all $k \in \operatorname{supp}(P^*)$*
$$\left(\frac{1}{n}\sum_{m=1}^{n} G_m\right) \xrightarrow[n\to\infty]{} \gamma^k_\infty(\sigma) \quad \mathbb{P}^k_\sigma - a.s.$$
Moreover, $\gamma^{P^*}_\infty(\sigma) = v_\infty(P^*)$ and $\mathbb{E}^{p_1}_{\sigma^\varepsilon}(v_\infty(P^*)) \geq v_\infty(p_1) - \varepsilon$.

This result is a consequence of Venel and Ziliotto [VZ16, Lemma 33]. This previous work states the existence of elements $\mu^* \in \Delta(\Delta(\mathcal{K}))$ and $\sigma^* \in \Delta(\Sigma)$ with similar properties to those of $P^* \in \Delta(\mathcal{K})$ and $\sigma \in \Sigma$. In this sense, the present lemma can be seen as a deterministic version of this previous result. To focus on the new tools we introduce to prove Theorem 2.3.1, we relegate the proof and the explanation of the differences between the two lemmata to Section 2.8.

Remark 2.6.4. The first property of Lemma 2.6.3 follows immediately from [VZ16, Lemma 33] by the type of convergence in this previous result. On the other hand, the second property requires the introduction of a certain Markov chain on $\mathcal{K} \times \mathcal{A} \times \Delta(\mathcal{K})$. This Markov chain is already present in the work [VZ16] but was used for other purposes. Therefore, the proof consists mainly of recalling previous results and constructions.

Remark 2.6.5. Note that $\mathbb{P}^k_\sigma$ represents the law on plays induced by the strategy σ, conditional on the fact that the initial state is k. This does not mean that we consider the decision-maker to know k. In the same fashion, $\gamma^k_\infty(\sigma)$ is the reward given by the strategy σ, conditional on the fact that the initial state is k. Even though σ is optimal in $\Gamma(P^*)$, this does not imply that σ is optimal in $\Gamma(\delta_k)$: we may have $\gamma^k_\infty(\sigma) < v_\infty(\delta_k)$.

The importance of Lemma 2.6.3 comes from the fact that the average rewards converge almost surely to a limit that only depends on the initial state k. Intuitively, this result means that for any initial belief p_1, after a finite number of stages, we can get ε-close to a belief P^* such that the optimal reward from P^* is, in expectation, almost the same as from p_1, and moreover from P^* there exists an optimal strategy that induces a strong ergodic behavior on the state dynamics. Thus, there is a natural way to build a 3ε-optimal strategy $\tilde{\sigma}$ in $\Gamma(p_1)$: first, apply the strategy σ^ε for m_ε stages, then apply σ. Since after m_ε steps the current belief P_{m_ε} is ε-close to P^* with probability higher than $1-\varepsilon$, the reward from playing $\tilde{\sigma}$ is at least the expectation of $\gamma^{P^*}_\infty(\sigma) - 2\varepsilon$, which is greater than $v_\infty(p_1) - 3\varepsilon$. Therefore, this procedure yields a 3ε-optimal strategy. Nonetheless, σ may not have finite memory, and thus $\tilde{\sigma}$ may not have either. The main difficulty of the proof is to transform σ into a finite-memory strategy. We formalize this discussion below.

Definition 2.6.6 (ergodic strategy)**.** Let $p^* \in \Delta(\mathcal{K})$. We say that a strategy σ is *ergodic* for p^* if the following holds for all $k \in \operatorname{supp}(p^*)$

$$\left(\frac{1}{n}\sum_{m=1}^{n} G_m\right) \xrightarrow[n\to\infty]{} \gamma^k_\infty(\sigma) \quad \mathbb{P}^k_\sigma - a.s.$$

From the previous discussion, we aim at proving the following result.

Lemma 2.6.7. *Let $p^* \in \Delta(\mathcal{K})$ and σ be an ergodic strategy for p^*. For all $\varepsilon > 0$, there exists a finite-memory strategy σ' such that*

$$\gamma^{p^*}_\infty(\sigma') \geq \gamma^{p^*}_\infty(\sigma) - \varepsilon\,.$$

This is our key lemma and the main technical contribution. The next section is devoted to explaining the technique used and proving it.

Proof of Theorem 2.3.1 assuming Lemmas 2.6.7 and 2.6.3. Let p_1 be an initial belief and $\varepsilon > 0$. Let m_ε, σ^ε, P^* and σ be given by Lemma 2.6.3. Define the strategy σ^0 by: playing σ^ε until stage m_ε, then switch to the strategy σ' given by Lemma 2.6.7 for σ and $p^* = P^*$. Note that σ^0 has finite memory. We have

$$\begin{aligned}
\gamma^{p_1}_\infty(\sigma^0) &= \mathbb{E}^{p_1}_{\sigma^\varepsilon}\left(\gamma^{P_{m_\varepsilon}}_\infty(\sigma')\right) && ;\text{def } \sigma^0\\
&\geq \mathbb{E}^{p_1}_{\sigma^\varepsilon}\left(\gamma^{P^*}_\infty(\sigma')\right) - 2\varepsilon && ;\text{Lemma } 2.6.3\\
&\geq \mathbb{E}^{p_1}_{\sigma^\varepsilon}\left(\gamma^{P^*}_\infty(\sigma)\right) - 3\varepsilon && ;\text{Lemma } 2.6.7\\
&= \mathbb{E}^{p_1}_{\sigma^\varepsilon}\left(v_\infty(P^*)\right) - 3\varepsilon && ;\text{Lemma } 2.6.3\\
&\geq v_\infty(p_1) - 4\varepsilon && ;\text{Lemma } 2.6.3\,,
\end{aligned}$$

and the theorem is proved.

□

2.7 Super-Support and Finite-Memory Strategies

We present the proof of Lemma 2.6.7. In this entire section, fix $p^* \in \Delta(\mathcal{K})$, which will be used as an initial belief, and σ an ergodic strategy for p^*.

2.7.1 Notation

For $a, b \in \mathbb{R}$, denote the set $[a, b] \cap \mathbb{Z}$ by $[a \,..\, b]$.

Definition 2.7.1 (Value partition)**.** Let $\sim$ be the equivalence relationship on $\operatorname{supp}(p^*)$ defined by $k \sim k'$ if and only if $\gamma^k_\infty(\sigma) = \gamma^{k'}_\infty(\sigma)$. Let $\{\mathcal{K}_1, \dots, \mathcal{K}_I\}$ be the corresponding *value partition*.

Definition 2.7.2 (Super-support)**.** For $i \in [1 \,..\, I]$ and $m \geq 0$, define, for all $h_m \in \mathcal{H}_m$,

$$\mathcal{B}^i_m(h_m) := \bigcup_{k_1 \in \mathcal{K}_i} \left\{k \in \mathcal{K} : \mathbb{P}^{k_1}_\sigma(H_m = h_m) > 0, \mathbb{P}^{k_1}_\sigma(K_m = k | H_m = h_m) > 0\right\}.$$

In other words, $\mathcal{B}^i_m(h_m)$ is the set of all reachable states at stage m starting from some state in $\mathcal{K}_i$, playing the strategy σ and obtaining history h_m (if $H_m = h_m$ is possible). Denote

$$B^i_m := \mathcal{B}^i_m(H_m),$$

the random set associated with H_m, and $B_m := (B^1_m, \dots, B^I_m)$ the *super-support* at stage m.

Remark 2.7.3. On the one hand, the set of reachable states at stage m is

$$\operatorname{supp}(P_m) = \bigcup_{i \in [1 \,..\, I]} B^i_m.$$

Therefore, the support of P_m can be deduced from the super-support B_m. On the other hand, B_m can not be deduced from P_m. In particular, B_m can not be deduced from the support of P_m. This justifies the vocabulary.

We will build a finite-memory ε-optimal strategy that plays by blocks. Each block has fixed finite length and, within each block, the strategy depends only on the history in the block and on the super-support at the beginning of the block. At the end of the block, the automaton computes the new super-support according to the block history and the previous super-support. Thus, the only difference with a bounded recall strategy is that our strategy keeps track of the super-support. Super-support is a type of origin information: it is related to the value partition, and therefore to where the current mass distribution comes from.

Definition 2.7.4 (h_m-shift)**.** Let $m \geq 1$ and $h_m \in \mathcal{H}_m$. The *h_m-shift* of σ is the strategy $\sigma[h_m]$ defined by, for all $m' \geq 1$,

$$\sigma[h_m](h_{m'}) := \sigma(h_m, h_{m'}).$$

We denote $\sigma_m := \sigma[H_m]$, the corresponding random shift at stage m.

In other words, $\sigma[h_m]$ corresponds to the continuation of the strategy σ conditional on the fact that the history of the first m stages was h_m.

2.7.2 Illustration

The super-support captures specific information related to the beginning of the game: the origin of the current mass distribution (given by P_m) in terms of the initial value partition $(\mathcal{K}_i)_{i\in[1\,..\,I]}$. There are finitely many possible super-supports and it is possible to keep track of the current super-support using Bayesian updating. Therefore, it is a good variable to be used in finite-memory strategies.

Let us recall our simple example of a POMDP, Example 2.5.2.

Consider two states (k_u, k_d) and two actions: *up* (a_u) and *down* (a_d). All transitions are possible (including loops) and they do not depend on the action. Signals inform the player when the state changes. In terms of actions and rewards, by playing a_u the player obtains a reward of 1 only if the current state is k_u. Similarly, by playing a_d the player obtains a reward of 1 only if the state is k_d. Figure 2.6 is a representation of the game with specific transition probabilities.

Finite-recall is enough to approximate the value of this POMDP: the decision-maker can recall the last action. Then, upon seeing the signal s_c, the player has to change actions. Recall that $p_1 = 1/4 \cdot \delta_{k_u} + 3/4 \cdot \delta_{k_d}$. Therefore, an optimal strategy is given by playing a_d until getting signal s_c, then playing a_u until getting signal s_c and repeat. This strategy is ergodic for p_1 and the corresponding value partition is given by $(\mathcal{K}_1 = \{k_u\}, \mathcal{K}_2 = \{k_d\})$, because, if $K_1 = k_u$, the long-run reward is 0 and, if $K_1 = k_d$, the long-run reward is 1. In this case, the super-support describes completely the belief P_m since it keeps track of which state has the highest (or lowest) probability.

Although the example is simple, we can already see the difference between support strategies and super-support strategies. In this case, all strategies based on the current support (the support of P_m) are constant and therefore can achieve a long-run reward of at most $1/2$. On the other hand, super-support strategies can be optimal and achieve a long-run reward of $3/4$.

This example also shows that playing by blocks and defining the behaviour in each block by the current support (instead of the super-support) is not enough.

Let us analyze now our more complex POMDP example, Example 2.5.2.

Consider six states and four actions: *up* (a_u), *down* (a_d), *left* (a_l) and *right* (a_r). States can be separated into two groups: *extremes* (k_u and k_d) and *transitional* ($k_{l_1}, k_{r_1}, k_{l_2}$ and k_{r_2}). Furthermore, transitional states can be divided into two groups: *left states* (k_{l_1} and k_{l_2}) and *right states* (k_{r_1} and k_{r_2}). Transitions are from extreme states to transitional states and from transitional to extremes. More precisely, excluding loops, only the following transitions are possible: from k_u to either k_{l_1} or k_{r_1}, then from these two to k_d, from k_d to either k_{l_2} or k_{r_2} and then back to k_u. Signals are such that the player knows: (i) the state changed to an extreme state, or (ii) the state changed to a transitional state and the new state is with higher probability a left state or a right state. In terms of actions and rewards, each action has an associated set of states in which the reward is 1 and the rest is 0: by playing a_u the reward is 1 only if the current state is k_u, playing a_d rewards only state k_d, a_l rewards states k_{l_1} and k_{l_2}, and a_r rewards states k_{r_1} and k_{r_2}. Figure 2.7 is a representation of this game with specific transition probabilities.

Recall that $p_1 = 1/4 \cdot \delta_{k_u} + 3/4 \cdot \delta_{k_d}$ and that an optimal strategy is given by playing action a_d until getting a signal s_l or s_r. If the decision-maker gets signal s_l, then play action a_l, otherwise, play action a_r. Repeat action a_r until getting the signal s_c. Then, play a_u until getting a signal s_l or s_r. When this happens, play a_l or a_r accordingly until getting signal s_c. And so, repeat the cycle.

This optimal strategy is ergodic for p_1 and the corresponding value partition is given by $(\mathcal{K}_1 = \{k_u\}, \mathcal{K}_2 = \{k_d\})$ because, if $K_1 = k_u$, the long-run reward is 0 and, if $K_1 = k_d$, the long-run reward is $7/8$. Contrary to the previous example, the super-support does not describe completely the belief P_m. Indeed, consider the initial belief p_1, which is supported on the extreme states, and that the decision-maker gets either signals s_l or s_r. Then, the new belief is supported in all the transitional states and the super-support is the same under any of these two histories, and equal to: $B = (B^1 = \{k_{l_1}, k_{r_1}\}, B^2 = \{k_{l_2}, k_{r_2}\})$. Based on this super-support one can not reconstruct the current belief, but one knows more than only the support: we can differentiate the origin (k_u or k_d) of the current belief distribution.

Notice that using the super-support alone is not enough to get ε-optimal strategies. Indeed, in transitional states, the decision-maker needs to know whether the state is more likely to be in a left state or a right state in order to play well, and the super-support does not contain such information. That is why, in the proof of Lemma 2.6.7, we consider a more sophisticated class of strategies, that combine super-support and bounded recall. For the moment, let us describe such a strategy for this example. Choose n_0 very large, and for each $\ell \geq 1$, play the following strategy in the time block $[\ell n_0 + 1 \mathinner{..} (\ell + 1)n_0]$:

- *Case 1: the super-support at stage $\ell n_0 + 1$ is $(\{k_u\}, \{k_d\})$.* Play the previous 0-optimal strategy, that is: play action a_d until getting a signal s_l or s_r. If the decision-maker gets signal s_l, then play action a_l, otherwise, play action a_r. Repeat action a_r until getting the signal s_c. Then, play a_u until getting a signal s_l or s_r. When this happens, play a_l or a_r accordingly until getting signal s_c. And so, repeat the cycle.
- *Case 2: the super-support at stage $\ell n_0 + 1$ is $(\{k_d\}, \{k_u\})$.* Play the same strategy as in Case 1, except that the roles of a_u and a_d are switched.
- *Case 3: the super-support at stage $\ell n_0 + 1$ is $(\{k_{l_1}, k_{r_1}\}, \{k_{l_2}, k_{r_2}\})$.* Play a_l (or a_r) until getting the signal s_c. At this point, the super-support is $(\{k_d\}, \{k_u\})$. Then, play as in Case 2.
- *Case 4: the super-support at stage $\ell n_0 + 1$ is $(\{k_{l_2}, k_{r_2}\}, \{k_{l_1}, k_{r_1}\})$.* Play a_l (or a_r) until getting the signal s_c. At this point, the super-support is $(\{k_u\}, \{k_d\})$. Then, play as in Case 1.

This strategy is sub-optimal during the first phase of Case 3 and Case 4, until the decision-maker receives signal s_c. As n_0 grows larger and larger, this part becomes negligible. Thus, for any $\varepsilon > 0$, there exists n_0 such that this strategy is ε-optimal (but not optimal).

2.7.3 Properties

Now we can state properties of super-supports when the strategy σ is ergodic for p^* and explain how rich is the structure of the random sequence of beliefs $(P_m)_{m\geq 1}$. By definition of ergodic strategies, the map $k \mapsto \gamma_\infty^k(\sigma)$ is constant on $\mathcal{K}_i$, and we denote its value by γ_∞^i.

Lemma 2.7.5 (Continuation value). *For all $m \geq 1$ and $i \in [1 \mathinner{..} I]$ it holds $\mathbb{P}_\sigma^{p^*}$-a.s. that*

$$\forall k \in B_m^i \quad \gamma_\infty^k(\sigma_m) = \gamma_\infty^i$$

Consequently, $B_m^1, \ldots, B_m^I$ are disjoint $\mathbb{P}_\sigma^{p^*}$-a.s.

Proof. Let $i \in [1\,..\,I]$. Considering the law given by $\mathbb{P}_\sigma^{p^*}$, fix a realization $K_m = k \in B_m^i$. By definition of super-support, there exists $k' \in \mathcal{K}_i \subseteq \operatorname{supp}(p^*)$ such that k can be reached from k' in m steps. Recall that, since σ is ergodic for p^*,

$$\left(\frac{1}{n}\sum_{m'=m}^{m+n-1} G_{m'}\right) \xrightarrow[n\to\infty]{} \gamma_\infty^{k'}(\sigma) = \gamma_\infty^i \quad \mathbb{P}_\sigma^{k'} - a.s.\,.$$

In particular, the convergence holds when $K_m = k$. Then,

$$\left(\frac{1}{n}\sum_{m'=1}^{n} G_{m'}\right) \xrightarrow[n\to\infty]{} \gamma_\infty^{k'}(\sigma) = \gamma_\infty^i \quad \mathbb{P}_{\sigma_m}^{k} - a.s.\,,$$

and therefore $\gamma_\infty^k(\sigma_m) = \gamma_\infty^i$. □

Another property of the super-support is concerned with consecutive conditioning and is fairly intuitive. We formally state it in the following lemma and show the proof for completeness.

Lemma 2.7.6 (Continuation super-support). *Let $i \in [1\,..\,I]$, $m, m' \geq 0$. For all realizations $H_{m+m'} = h = (h_m, h_{m'}) \in \mathcal{H}_{m+m'}$, denoting $\mathcal{C} = \mathcal{B}_{m+m'}^i(h_{m+m'})$, we have that, for all $k \in \mathcal{B}_m^i(h_m)$,*

$$\mathbb{P}_{\sigma[h_m]}^k(K_{m'} \in \mathcal{C} | H_{m'} = h_{m'}) = 1\,.$$

In other words, the super-support that arises at stage $m + m'$, $B_{m+m'}$, coincides with the super-support that would arise from a two-step procedure: first, advancing m stages; and then, applying the continuation of the strategy, σ_m, for m' more stages.

Proof. Fix a realization $H_{m+m'} = h = (h_m, h_{m'})$ and let $k \in \mathcal{B}_m^i(h_m)$. Recall that, by definition of super-support,

$$\mathcal{B}_m^i(h_m) = \bigcup_{\substack{\bar{k}_1 \in \mathcal{K}_i \\ \mathbb{P}_\sigma^{\bar{k}_1}(h_m > 0)}} \operatorname{supp}\left(\mathbb{P}_\sigma^{\bar{k}_1}(K_m = \cdot \mid H_m = h_m)\right)\,.$$

Therefore, there exists $\bar{k}_1 \in \mathcal{K}_i$ such that $k \in \operatorname{supp}(\mathbb{P}_\sigma^{\bar{k}_1}(K_m = \cdot \mid H_m = h_m))$. In particular, we have that $\mathbb{P}_\sigma^{\bar{k}_1}(K_m = k) > 0$.

Consider k' such that $\mathbb{P}_{\sigma[h_m]}^k(K_{m'} = k' \mid H_{m'} = h_{m'}) > 0$. By a semi-group property, we deduce that

$$\mathbb{P}_\sigma^{\bar{k}_1}(K_{m+m'} = k' \mid H_{m+m'} = h) > 0\,,$$

which implies that $k' \in \mathcal{B}_m^i(h) = \mathcal{C}$, and thus the lemma is proved. □

Remark 2.7.7. This property does not depend on the fact that σ is ergodic for p^*.

2.7.4 Proof of Lemma 2.6.7

Fix $p^* \in \Delta(\mathcal{K})$ such that σ is ergodic for p^*. Note that, for all $m \geq 1$, $B_m \in \{(\mathcal{C}_1, \ldots, \mathcal{C}_I) : \mathcal{C}_1, \ldots, \mathcal{C}_I \subseteq \mathcal{K}\}$, which is a finite set. Denote all different super-supports that can occur with positive probability by $\mathcal{D}^1, \mathcal{D}^2, \ldots, \mathcal{D}^J$, i.e.

$$\left\{\mathcal{D}^1, \mathcal{D}^2, \ldots, \mathcal{D}^J\right\} \coloneqq \bigcup_{m\geq 1} \operatorname{supp} B_m\,.$$

Moreover, since $\mathcal{D}^j$ corresponds to a super-support that occurs at some stage and under some history, there exists h^j and m_j such that $h^j \in \mathcal{H}_{m_j}$ and $\mathcal{D}^j = \mathcal{B}_m^i(h^j)$. In other words, $\mathcal{D}^j$ is the realization of the super-support at stage m_j under history h^j and $\left\{\mathcal{D}^1, \mathcal{D}^2, \ldots, \mathcal{D}^J\right\}$ contains all super-supports that can occur.

Definition of the Strategy σ' Let $\varepsilon > 0$. By Lemma 2.7.5, there exists $n_0 \in \mathbb{N}^*$ such that for all $i \in [1 \,..\, I]$, $j \in [1 \,..\, J]$ and $k \in \mathcal{D}_i^j$,

$$\mathbb{E}^k_{\sigma[h^j]}\left(\frac{1}{n_0}\sum_{m=1}^{n_0} G_m\right) \geq \gamma^i_\infty - \varepsilon\,.$$

Define the strategy σ' by blocks, and characterize each block by induction. For each $\ell \geq 0$, the block number ℓ consists in the stages m such that $\ell n_0 + 1 \leq m \leq (\ell+1)n_0$. We characterize the behavior in block ℓ by a variable $J_\ell \in [1 \,..\, J]$ in the following way. For stage m inside block ℓ, the strategy σ' plays according to J_ℓ and the history between stages $\ell n_0 + 1$ and m. Each block is characterized by induction because the variable J_ℓ is computed at stage $\ell n_0 + 1$ according to $J_{(\ell-1)}$ and the history in the last n_0 stages. Thus, σ' can be seen as mapping from $\cup_{m=1}^{n_0}\mathcal{H}_m \times [1 \,..\, J]$ to $\mathcal{A}$.

Consider $\ell = 0$, i.e. the first block. The strategy σ' is defined on the first n_0 stages as follows. Consider the value partition $\{\mathcal{K}_1, \ldots, \mathcal{K}_I\}$ given by p^* and σ. By definition of $\mathcal{D}^1, \mathcal{D}^2, \ldots, \mathcal{D}^J$, there exists $j \in [1 \,..\, J]$ such that $B_1 = \mathcal{D}^j$. Set $J_0 = j$, and define $\sigma'(h, J_0) := \sigma(h^{J_0}, h)$ for all $h \in \mathcal{H}_m$ and $m \leq n_0$.

Let us proceed to the induction step. Consider $\ell \geq 1$ and assume that we have defined $J_{(\ell-1)}$ and σ' up to stage ℓn_0. Denote the history between stages $(\ell-1)n_0 + 1$ and $\ell n_0 + 1$ by $h \in \mathcal{H}_{n_0+1}$ and define J_ℓ such that, for all $i \in [1 \,..\, I]$,

$$\mathcal{D}_i^{J_\ell} = \mathcal{B}^i_{m_{J_{(\ell-1)}}+n_0+1}(h^{J_{(\ell-1)}}, h)\,.$$

Then, extend σ' for n_0 additional stages as before: $\sigma'(h, J_\ell) := \sigma(h^{J_\ell}, h)$ for all $h \in \mathcal{H}_m$ and $m \leq n_0$. Thus, we have defined J_ℓ and extended σ' up to stage $(\ell+1)n_0$.

To summarize our construction in words, during stages $\ell n_0 + 1, \ell n_0 + 2, \ldots, (\ell+1)n_0$, the decision-maker plays as if he was playing σ from history h^{J_ℓ}. Notice that the indexes $J_0, J_1, \ldots$ depend on the history, and therefore are random. Now we will connect the strategy σ' with the super-support given by p^* and σ.

Lemma 2.7.8. *For all $i \in [1 \,..\, I]$, $k \in \mathcal{K}_i$ and $\ell \geq 0$, we have that*

$$\mathbb{P}^k_{\sigma'}(K_{\ell n_0+1} \in \mathcal{D}_i^{J_\ell}) = 1\,.$$

Consequently, $\mathcal{D}^{J_\ell} = (\mathcal{D}_1^{J_\ell}, \mathcal{D}_2^{J_\ell}, \ldots, \mathcal{D}_I^{J_\ell})$ is a partition of the support of $P_{\ell n_0+1}$. Moreover,

$$\mathbb{P}^{p^*}_{\sigma'}(K_{\ell n_0+1} \in \mathcal{D}_i^{J_\ell}) = p^*(\mathcal{K}_i)\,.$$

Proof. Fix $i \in [1 \,..\, I]$ and $k \in \mathcal{K}_i$. We will prove the result by induction on $\ell \geq 0$. For $\ell = 0$, $\mathcal{D}^{J_0} = (\mathcal{K}_1, \ldots, \mathcal{K}_I)$ and $\mathbb{P}^k_{\sigma'}(K_1 \in \mathcal{D}_i^{J_0} = \mathcal{K}_i) = 1$. Thus, the result holds.

Assume $\ell \geq 1$. Note that $H_{(\ell-1)n_0+1}$ determines the value of $J_0, \ldots, J_{(\ell-1)}$. Therefore, $\mathcal{D}_i^{J_{(\ell-1)}}$ is also determined by $H_{(\ell-1)n_0+1}$. By induction hypothesis,

$$\mathbb{P}^k_{\sigma'}(K_{(\ell-1)n_0+1} \in \mathcal{D}_i^{J_{(\ell-1)}}) = 1\,.$$

We must prove that, under these circumstances, $K_{\ell n_0+1} \in \mathcal{D}_i^{J_\ell}$.

Indeed, index $J_{\ell-1}$ defines strategy σ' for stages $\ell n_0 + 1, \ell n_0 + 2, \ldots, (\ell+1)n_0$: the decision-maker will play according to $\sigma[h^{J_{(\ell-1)}}]$. By playing σ' during this block, a history $h \in \mathcal{H}_{n_0+1}$

will be collected. Let $m := J_{(\ell-1)}$ and $m' := n_0 + 1$. By Lemma 2.7.6, we have that, starting from $K_{(\ell-1)n_0+1} \in \mathcal{D}_i^{J_{(\ell-1)}}$, playing $\sigma[h^{J_{(\ell-1)}}]$ during n_0 stages and collecting history $h \in \mathcal{H}_{n_0+1}$ leads to a state that, by definition of J_ℓ, is in $\mathcal{D}_i^{J_\ell}$. Therefore,

$$\mathbb{P}^k_{\sigma'}(K_{\ell n_0+1} \in \mathcal{D}_i^{J_\ell}) \geq \mathbb{P}^k_{\sigma'}(K_{(\ell-1)n_0+1} \in \mathcal{D}_i^{J_{(\ell-1)}}) = 1\,,$$

which proves the first result.

Now we know that the union of $\mathcal{D}_1^{J_\ell}, \mathcal{D}_2^{J_\ell}, \ldots, \mathcal{D}_I^{J_\ell}$ covers the support of $\mathbb{P}^{p^*}_{\sigma'}(K_{\ell n_0+1} = \cdot)$. Moreover, by Lemma 2.7.5, $\mathcal{D}_1^{J_\ell}, \mathcal{D}_2^{J_\ell}, \ldots, \mathcal{D}_I^{J_\ell}$ are disjoint. Since $(\mathcal{K}_1, \ldots, \mathcal{K}_I)$ partitions the support of p^*, we have that, for all $i \in [1\,..\,I]$,

$$\begin{aligned}\mathbb{P}^{p^*}_{\sigma'}(K_{\ell n_0+1} \in \mathcal{D}_i^{J_\ell}) &= \sum_{i'=1}^{I} \sum_{k \in \mathcal{K}_{i'}} p^*(k)\mathbb{P}^k_{\sigma'}(K_{\ell n_0+1} \in \mathcal{D}_i^{J_\ell}) \\ &= \sum_{i'=1}^{I} \sum_{k \in \mathcal{K}_{i'}} p^*(k)\mathbb{1}_{k \in \mathcal{K}_i} \\ &= p^*(\mathcal{K}_i)\,,\end{aligned}$$

which proves the second property. □

To finish the proof of Lemma 2.6.7, we must show that the finite-memory strategy σ' guarantees the reward obtained by σ up to ε. The idea is that in each block we are playing some shift of σ for n_0 stages. The shift is chosen so that information about the initial belief is correctly updated, while the number n_0 is chosen so that the expected average reward of the whole block is close to the expected limit average reward. Then, since all blocks have the same approximation error, the average considering all blocks yields approximately $\gamma^{p^*}_\infty(\sigma)$. This is the intuition behind the following lemma.

Lemma 2.7.9. *Let $L \in \mathbb{N}^*$. The following inequality holds:*

$$\mathbb{E}^{p^*}_{\sigma'}\left(\frac{1}{Ln_0}\sum_{m=1}^{Ln_0} G_m\right) \geq \gamma^{p^*}_\infty(\sigma) - \varepsilon.$$

Proof. We have, for all $\ell \geq 0$,

$$\begin{aligned}\mathbb{E}^{p^*}_{\sigma'}\left(\frac{1}{n_0}\sum_{m=\ell n_0+1}^{(\ell+1)n_0} G_m\right) &= \mathbb{E}^{p^*}_{\sigma'}\left(\mathbb{E}^{K_{\ell n_0+1}}_{\sigma[h^{J_\ell}]}\left(\frac{1}{n_0}\sum_{m=1}^{n_0} G_m\right)\right) && ;\text{def. } \sigma' \\ &= \mathbb{E}^{p^*}_{\sigma'}\left(\sum_{i=1}^{I}\sum_{k \in \mathcal{D}_i^{J_\ell}} \mathbb{P}^{p^*}_{\sigma'}(K_{\ell n_0+1} = k)\mathbb{E}^k_{\sigma[h^{J_\ell}]}\left(\frac{1}{n_0}\sum_{m=1}^{n_0} G_m\right)\right) && ;\text{Lemma 2.7.8} \\ &\geq \mathbb{E}^{p^*}_{\sigma'}\left(\sum_{i=1}^{I}\sum_{k \in \mathcal{D}_i^{J_\ell}} \mathbb{P}^{p^*}_{\sigma'}(K_{\ell n_0+1} = k)\left[\gamma^i_\infty(\sigma) - \varepsilon\right]\right) && ;\text{def. } n_0 \\ &= \left(\sum_{i=1}^{I} \mathbb{P}^{p^*}_{\sigma'}(K_{\ell n_0+1} \in \mathcal{D}_i^{J_\ell})\left[\gamma^i_\infty(\sigma) - \varepsilon\right]\right) \\ &= \sum_{i=1}^{I} p^*(\mathcal{K}_i)\left[\gamma^i_\infty(\sigma) - \varepsilon\right] && ;\text{Lemma 2.7.8} \\ &= \gamma^{p^*}_\infty(\sigma) - \varepsilon\,.\end{aligned}$$

It follows that

$$\mathbb{E}_{\sigma'}^{p^*}\left(\frac{1}{Ln_0}\sum_{m=1}^{Ln_0} G_m\right) = \frac{1}{L}\sum_{\ell=0}^{L-1}\mathbb{E}_{\sigma'}^{p^*}\left(\frac{1}{n_0}\sum_{m=\ell n_0+1}^{(\ell+1)n_0} G_m\right) \geq \gamma_\infty^{p^*}(\sigma) - \varepsilon\,.$$

□

To conclude, since σ' has finite memory, we have

$$\lim_{L\to+\infty}\mathbb{E}_{\sigma'}^{p^*}\left(\frac{1}{Ln_0}\sum_{m=1}^{Ln_0} G_m\right) = \mathbb{E}_{\sigma'}^{p^*}\left(\liminf_{n\to+\infty}\frac{1}{n}\sum_{m=1}^{n} G_m\right) = \gamma_\infty^{p^*}(\sigma')\,,$$

and the above lemma implies that $\gamma_\infty^{p^*}(\sigma') \geq \gamma_\infty^{p^*}(\sigma) - \varepsilon$, which proves Lemma 2.6.7, i.e., for each ergodic strategy σ and $\varepsilon > 0$, one can construct a finite-memory strategy σ' that guarantees the reward obtained by σ up to ε.

2.8 Postponed Proofs

In this section we present the proof of Lemma 2.6.3, which is a consequence of [VZ16, Lemma 33]. Therefore, we start by introducing some of the terms used in [VZ16], namely: n-stage game, invariant measure, occupation measure and the Kantorovich-Rubinstein distance.

Definition 2.8.1 (n-stage game)**.** Given a POMDP $\Gamma = (\mathcal{K}, \mathcal{A}, \mathcal{S}, q, g)$, we denote Γ_n the *n-stage game* with a value defined by

$$v_n(p) := \sup_{\sigma\in\Sigma}\gamma_n^p(\sigma)\,,$$

where $\gamma_n^p(\sigma) := n^{-1}\mathbb{E}_\sigma^p\left(\sum_{m=1}^n G_m\right)$.

Remark 2.8.2. As the notation suggests, it was proven in [VZ16] that for any finite POMDP $(v_n) \xrightarrow[n\to\infty]{} v_\infty$ uniformly. The fact that $(v_n)_{n\geq 1}$ converges was proven in [RSV02].

The set $\Delta(\mathcal{K})$ is equipped with its Borelian σ-algebra $\mathcal{B}(\Delta(\mathcal{K}))$, and $\mathcal{C}(\Delta(\mathcal{K}), [0,1])$ denotes the set of continuous functions from $\Delta(\mathcal{K})$ to $[0,1]$.

Definition 2.8.3 (Invariant measure)**.** Given a POMDP $\Gamma = (\mathcal{K}, \mathcal{A}, \mathcal{S}, q, g)$, $\mu \in \Delta(\Delta(\mathcal{K}))$ and $\sigma\colon \Delta(\mathcal{K}) \to \Delta(\mathcal{A})$ measurable, we say that μ is *σ-invariant* if $\forall f \in \mathcal{C}(\Delta(\mathcal{K}), [0,1])$ we have that

$$\int_{\Delta(\mathcal{K})}\mathbb{E}\left[f(\tilde{q}[p,\sigma(p)])\right]\mu(dp) = \int_{\Delta(\mathcal{K})} f(p)\mu(dp)\,,$$

where $\tilde{q}\colon \Delta(\mathcal{K}) \times \mathcal{A} \to \Delta(\Delta(\mathcal{K}))$ is the natural transition in $\Delta(\mathcal{K})$ from one belief to another, given by Bayes rule.

The above definition can be intuitively understood in the following way: if the initial belief is distributed according to μ, and the decision-maker plays the stationary strategy σ at stage 1, then the belief at stage 2 is distributed according to μ too.

Remark 2.8.4. Since $v_\infty\colon \Delta(\mathcal{K}) \to [0,1]$ is a continuous function, one can replace f by v_∞ in the previous definition. Moreover, interpreting σ as a (mixed) stationary strategy, we would have that the sequence $(\mathbb{E}_\sigma^\mu[v_\infty(P_m)])_{m\geq 1}$ is constant.

Definition 2.8.5 (m-stage occupation measure)**.** Given a POMDP $\Gamma = (\mathcal{K}, \mathcal{A}, \mathcal{S}, q, g)$, a measure $\mu \in \Delta(\Delta(\mathcal{K}))$ and a strategy σ, consider the following dynamic over $\Delta(\mathcal{K})$. First, P_1 is drawn according to μ. Then, $(P_n)_{n \geq 1}$ is obtained by playing according to σ. This way, for each $m \geq 1$, we have that Γ, μ and σ induce a probability over $\Delta(\mathcal{K})$: for each measurable set $\mathcal{A} \subseteq \Delta(\mathcal{K})$, we can define $\mathbb{P}^{\mu}_{\sigma}(P_m \in \mathcal{A})$. Therefore, the m-stage belief, P_m, is a random belief.

We denote the *m-stage occupation measure* $z_m[\mu, \sigma] \in \Delta(\Delta(\mathcal{K}))$ by the law of P_m over $\Delta(\mathcal{K})$. Formally, $z_m[\mu, \sigma] \colon \mathcal{B}(\Delta(\mathcal{K})) \to [0, 1]$ is given by, for all $\mathcal{C} \in \mathcal{B}(\Delta(\mathcal{K}))$,

$$z_m[\mu, \sigma](\mathcal{C}) = \mathbb{P}^{\mu}_{\sigma}(P_m \in \mathcal{C}) \,.$$

For sake of notation, we identify $\Delta(\mathcal{K})$ with the extreme points of $\Delta(\Delta(\mathcal{K}))$.

Definition 2.8.6 (Kantorovich-Rubinstein distance)**.** For all $z, z' \in \Delta(\Delta(\mathcal{K}))$, define

$$d_{KR}(z, z') := \sup_{f \in \mathcal{E}_1} \left| \int_{\Delta(\mathcal{K})} f(p) z(dp) - \int_{\Delta(\mathcal{K})} f(p) z'(dp) \right| \,,$$

where $\mathcal{E}_1$ is the set of 1-Lipschitz functions from $\Delta(\mathcal{K})$ to $[0, 1]$.

Remark 2.8.7. The set $\Delta(\Delta(\mathcal{K}))$ equipped with distance d_{KR} is a compact metric space.

Now we can state [VZ16, Lemma 33].

Lemma 2.8.8. *Consider a POMDP Γ and let $p_1 \in \Delta(\mathcal{K})$. There exists $\mu^* \in \Delta(\Delta(\mathcal{K}))$ and a (mixed) stationary strategy $\sigma^* \colon \Delta(\mathcal{K}) \to \Delta(\mathcal{A})$ such that*

1. *μ^* is σ^*-invariant.*
2. *For all $\varepsilon > 0$ and $N \geq 1$, there exists $n_\varepsilon \geq N$ and σ^ε a pure strategy in Γ such that σ^ε is 0-optimal in the n_ε-stage game $\Gamma_{n_\varepsilon}(p_1)$ and*
$$d_{KR}\left(\frac{1}{n_\varepsilon} \sum_{m=1}^{n_\varepsilon} z_m[p_1, \sigma^\varepsilon], \mu^* \right) \leq \varepsilon \,.$$
3. $$\int_{\Delta(\mathcal{K})} g(p, \sigma^*(p)) \mu^*(dp) = \int_{\Delta(\mathcal{K})} v_\infty(p) \mu^*(dp) = v_\infty(p_1) \,.$$

Remark 2.8.9. Lemma 2.8.8 works with elements in $\Delta(\Delta(\mathcal{K}))$ and a mixed stationary strategy, while Lemma 2.6.3 deals with (random) elements in $\Delta(\mathcal{K})$ and a pure strategy. In this sense, we would like to "go down a level": moving from μ^* to a random P^*, from z_m to P_m and still preserve a relationship between $v_\infty(p_1)$ and $\mathbb{E}^{p_1}_{\sigma^\varepsilon}(v_\infty(P^*))$. The ergodic property 2 of Lemma 2.6.3 follows from the first and third property of Lemma 2.8.8.

Proof of Lemma 2.6.3. Consider $p_1 \in \Delta(\mathcal{K})$ and $\varepsilon > 0$ fixed. Since $(v_n)_{n \geq 1}$ converges uniformly to v_∞, consider $N \geq 1$ such that $\varepsilon \geq 1/N$ and $\forall n \geq N \quad ||v_n - v_\infty||_\infty \leq \varepsilon$. Now, using Lemma 2.8.8, there exists μ^* and σ^* such that μ^* is σ^*-invariant and, considering ε^4, $\exists n_\varepsilon \geq N$ such that

$$d_{KR}\left(\frac{1}{n_\varepsilon} \sum_{m=1}^{n_\varepsilon} z_m[p_1, \sigma^\varepsilon], \mu^* \right) \leq \varepsilon^4 \,,$$

with $\sigma^\varepsilon \in \Gamma_{n_\varepsilon}(p_1)$ an optimal pure strategy for the n_ε-stage game starting in p_1.

We claim that $\exists m_\varepsilon \leq \lceil \varepsilon n_\varepsilon \rceil$ such that

$$\mathbb{P}^{p_1}_{\sigma^\varepsilon}(P_{m_\varepsilon} \in \operatorname{supp}(\mu^*) + B(0,\varepsilon)) > 1 - \varepsilon. \tag{2.1}$$

Proceeding by contradiction, assume that $\forall m \leq \lceil \varepsilon n_\varepsilon \rceil$, we have $\mathbb{P}^{p_1}_{\sigma^\varepsilon}(P_m \in \operatorname{supp}(\mu^*) + B(0,\varepsilon)) \leq 1 - \varepsilon$. Define the function $f \colon \Delta(\Delta(\mathcal{K})) \to [0,1]$ by $f(p) = d_\infty(p, \operatorname{supp}(\mu^*))$, the supremum distance from $\operatorname{supp}(\mu^*)$. Clearly, $f \in \mathcal{E}_1$. Moreover,

$$\begin{aligned}
\varepsilon^4 &\geq d_{KR}\left(\frac{1}{n_\varepsilon}\sum_{m=1}^{n_\varepsilon} z_m[p_1,\sigma^\varepsilon], \mu^*\right) \\
&\geq \left| \int_{\Delta(\mathcal{K})} f(p) \frac{1}{n_\varepsilon}\sum_{m=1}^{n_\varepsilon} z_m[p_1,\sigma^\varepsilon](dp) - \int_{\Delta(\mathcal{K})} f(p)\mu^*(dp) \right| \\
&\geq \left| \frac{1}{n_\varepsilon}\sum_{m=1}^{n_\varepsilon} \int_{\Delta(\mathcal{K})} f(p) z_m[p_1,\sigma^\varepsilon](dp) \right| && ; f(p) = 0, p \in \operatorname{supp}(\mu^*) \\
&\geq \frac{1}{n_\varepsilon}\sum_{m=1}^{\lceil \varepsilon n_\varepsilon \rceil} \int_{\Delta(\mathcal{K})\setminus(\operatorname{supp}(\mu^*)+B(0,\varepsilon))} f(p) z_m[p_1,\sigma^\varepsilon](dp) && ; f, z_m[p_1,\sigma^\varepsilon] \geq 0 \\
&\geq \varepsilon \frac{1}{n_\varepsilon}\sum_{m=1}^{\lceil \varepsilon n_\varepsilon \rceil} z_m[p_1,\sigma^\varepsilon](\Delta(\mathcal{K}) \setminus (\operatorname{supp}(\mu^*) + B(0,\varepsilon))) && ; \text{def. of } f \\
&\geq \varepsilon^3 && ; \text{contradiction hypothesis,}
\end{aligned}$$

which is a contradiction for $\varepsilon < 1$. Thus, we have proven (2.1).

Take $P^* \in \operatorname{argmin}_{p \in \operatorname{supp}(\mu^*)} \|p - P_{m_\varepsilon}\|_1$. By equation (2.1), the first property of Lemma 2.6.3 is satisfied.

For the second property of Lemma 2.6.3, note that, with probability higher than $1 - \varepsilon$,

$$\begin{aligned}
v_\infty(P^*) &\geq v_{n_\varepsilon}(P^*) - \varepsilon && ; ||v_{n_\varepsilon} - v_\infty||_\infty \leq \varepsilon \\
&\geq v_{n_\varepsilon}(P_{m_\varepsilon}) - 2\varepsilon && ; v_{n_\varepsilon} \text{ is 1-Lipschitz} .
\end{aligned}$$

On the other hand, taking expectation we get that

$$\begin{aligned}
\mathbb{E}^{p_1}_{\sigma^\varepsilon}\left(v_{n_\varepsilon}(P_{m_\varepsilon})\right) &\geq \mathbb{E}^{p_1}_{\sigma^\varepsilon}\left(v_{n_\varepsilon - m_\varepsilon}(P_{m_\varepsilon})\right) - \varepsilon && ; m_\varepsilon \leq \lceil \varepsilon n_\varepsilon \rceil \\
&\geq v_{n_\varepsilon}(p_1) - 2\varepsilon && ; \sigma^\varepsilon \text{ is 0-optimal in } \Gamma_{n_\varepsilon}(p_1) \\
&\geq v_\infty(p_1) - 3\varepsilon && ; \forall n \geq N \quad ||v_n - v_\infty||_\infty \leq \varepsilon .
\end{aligned}$$

Therefore, we conclude that

$$\mathbb{E}^{p_1}_{\sigma^\varepsilon}\left(v_\infty(P^*)\right) \geq v_\infty(p_1) - 6\varepsilon .$$

To complete the proof of Lemma 2.6.3, given P^*, we need a (pure) strategy σ such that

$$\forall\ k \in \operatorname{supp}(P^*) \quad \left(\frac{1}{n}\sum_{m=1}^{n} G_m\right) \xrightarrow[n\to\infty]{} \gamma^k_\infty(\sigma) \quad \mathbb{P}^k_\sigma - a.s. \tag{2.2}$$

and such that

$$\gamma^{P^*}_\infty(\sigma) = v_\infty(P^*). \tag{2.3}$$

Consider the random process $(Y_m)_{m\geq 1}$ on $\mathcal{Y} := \mathcal{K} \times \mathcal{A} \times \Delta(\mathcal{K})$ defined by $Y_m := (K_m, A_m, P_m)$. We claim that, under σ^*, the process $(Y_m)_{m\geq 1}$ is a Markov chain. Indeed, given $m \geq 1$ and $(Y_1, \ldots, Y_m) \in \mathcal{Y}^m$, Y_{m+1} is generated by the following procedure:

1. Draw a pair (K_{m+1}, S_m) according to $q(K_m, A_m)$,
2. Compute P_{m+1} using Bayes rule according to P_m and S_m,
3. Draw the next action A_{m+1} according to $\sigma^*(P_{m+1})$.

By construction, the law of Y_{m+1} depends only on Y_m and therefore $(Y_m)_{m\geq 1}$ is a Markov chain.

Define $\nu^* \in \Delta(\mathcal{Y})$ by fixing the third marginal to μ^* and for all $p \in \Delta(\mathcal{K})$, $\nu^*(\cdot \mid p) \in \Delta(\mathcal{K}\times\mathcal{A})$ is $p \otimes \sigma^*(p)$. We claim that ν^* is an invariant measure for $(Y_m)_{m\geq 1}$. Indeed, fixing σ^* as the strategy for the player, if P_1 is drawn according to μ^*, then, since μ^* is σ^*-invariant, the third marginal of Y_m follows μ^*, for all $m \geq 1$. Moreover, conditional on P_m, the random variables K_m and A_m are independent: the conditional distribution of K_m is P_m and the one of A_m is $\sigma^*(P_m)$. Thus, ν^* is an invariant measure of $(Y_m)_{m\geq 1}$.

The strategy $\sigma^* \colon \Delta(\mathcal{K}) \to \Delta(\mathcal{A})$ is a (stationary) mixed strategy, and we are looking for a deterministic strategy $\sigma \in \Sigma$. To derandomize this strategy, note that σ^* starting from any $p \in \Delta(\mathcal{K})$ is strategically equivalent to a p-dependent element of $\Delta(\Sigma)$, that is, a distribution over pure (not necessarily stationary) strategies (Kuhn's theorem, see [Fei96]). To simplify notations, we still denote this equivalent strategy σ^*, and omit its dependence in p.

Define $f\colon \mathcal{Y} \to [0,1]$ by $f(k,a,p) := g(k,a)$, a measurable function. Applying an ergodic theorem in Hernández-Lerma and Lasserre [HLL03, Theorem 2.5.1, page 37], we know that $\exists f^*$ integrable with respect to μ^* such that, for all $p \in \operatorname{supp}(\mu^*)$ and $\sigma \in \operatorname{supp}(\sigma^*)$,

$$\left(\frac{1}{n}\sum_{m=1}^{n} f(K_m, A_m, P_m) = \frac{1}{n}\sum_{m=1}^{n} G_m\right) \xrightarrow[n\to\infty]{} f^*(K_1, a_1, p) \quad \mathbb{P}^p_\sigma - a.s.\,,$$

where f^* satisfies that $\int_{\Delta(\mathcal{K})} f^*(y)\nu^*(dy) = \int_{\Delta(\mathcal{K})} f(y)\nu^*(dy)$.

We claim that, for all $p \in \operatorname{supp}(\mu^*)$ and $\sigma \in \operatorname{supp}(\sigma^*)$, we have that, for all $k \in \operatorname{supp}(p)$,

$$f^*(k, a_1, p) = \gamma^k_\infty(\sigma)\,,$$

where a_1 is the first action according to σ (formally, $a_1 = \sigma(\emptyset) = \sigma^*(p)$).

Indeed, take $p \in \operatorname{supp}(\mu^*)$, $\sigma \in \operatorname{supp}(\sigma^*)$ and $k \in \operatorname{supp}(p)$, then

$$\gamma^k_\infty(\sigma) = \mathbb{E}^k_\sigma\left(\liminf_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n} G_m\right) = \mathbb{E}^k_\sigma\left(f^*(K_1, a_1, p)\right) = f^*(k, a_1, p)\,.$$

Since $\operatorname{supp}(P^*) \subseteq \{k \in \mathcal{K} : \exists p \in \operatorname{supp}(\mu^*) \text{ s.t. } k \in \operatorname{supp}(p)\}$, property (2.2) is satisfied.

Let us now turn to property (2.3). We claim that, $\mu^* - a.s.$ and $\sigma^* - a.s.$,

$$\gamma^p_\infty(\sigma) = v_\infty(p)\,.$$

Indeed, note that

$$\begin{aligned}
\int_{\Delta(\mathcal{K})}\int_\Sigma \gamma^p_\infty(\sigma)\sigma^*(d\sigma)\mu^*(dp) &= \int_{\Delta(\mathcal{K})} f^*(y)\nu^*(dy) && ;\text{def. of } \nu^* \\
&= \int_{\Delta(\mathcal{K})} f(y)\nu^*(dy) \\
&= \int_{\Delta(\mathcal{K})} g(p, \sigma^*(p))\mu^*(dp) && ;\text{def. of } \nu^* \\
&= \int_{\Delta(\mathcal{K})} v_\infty(p)\mu^*(dp) && ;\mathit{Lemma}\ 2.8.8\,.
\end{aligned}$$

By definition of v_∞, for all $\sigma \in \Sigma$, $\gamma_\infty^{\cdot}(\sigma) \leq v_\infty(\cdot)$. Therefore, by positivity, we can conclude that, $\mu^* - a.s.$ and $\sigma^* - a.s.$, $\gamma_\infty^p(\sigma) = v_\infty(p)$. Since the support of P^* is included in $\operatorname{supp}(\mu^*)$, we conclude that

$$\gamma_\infty^{P^*}(\sigma) = v_\infty(P^*)\,,$$

and property (2.3) is satisfied. □

CHAPTER 3

Value-Positivity for Matrix Games

This chapter is based on [COBS24], i.e., the following publication. Krishnendu Chatterjee, Miquel Oliu-Barton, and Raimundo Saona. Value-Positivity for Matrix Games. *Mathematics of Operations Research*, page 1–24, 2024.

Matrix games are the most basic model in game theory, and yet robustness with respect to small perturbations of the matrix entries is not fully understood. We introduce value-positivity and uniform value-positivity, two properties that refine the notion of optimality in the context of polynomially perturbed matrix games. The first concept captures how the value depends on the perturbation parameter, and the second consists of the existence of a fixed strategy that guarantees the value of the unperturbed matrix game for every sufficiently small positive parameter. We provide polynomial-time algorithms to check whether a polynomially perturbed matrix game satisfies these properties. We further provide the functional form for a parameterized optimal strategy and the value function. Finally, we translate our results to linear programming and stochastic games, where value-positivity is related to the existence of robust solutions.

3.1 Introduction

Matrix games are the most basic model in game theory, where two opponents face each other over finitely many actions. The existence of the value and optimal strategies was established nearly 100 years ago [vN28] and yet, surprisingly, robustness with respect to small perturbations of the matrix entries is not fully understood.

We consider *polynomial matrix games*, that is, matrix games where the entries are polynomial functions of some real parameter $\varepsilon \geq 0$. The matrix corresponding to $\varepsilon = 0$ is interpreted as the unperturbed game, as opposed to the other matrices which are perturbed games. Recently, Attia and Oliu-Barton [AOB19] pointed out that polynomial matrix games arise naturally in the study of stochastic games, the first dynamic game introduced in the literature [Sha53, SV15].

In the context of polynomially perturbed matrix games, we introduce *value-positivity* and *uniform value-positivity*, two properties that refine the notion of optimality in matrix games. Whereas the first captures the dependency of optimal strategies with respect to perturbations, the second consists of the existence of a fixed strategy guaranteeing the value of the unperturbed matrix game for all sufficiently small positive parameters. Applied to linear programs (LPs), which are known to be equivalent to matrix games [Adl13, Dan51], and stochastic games, these properties imply the existence of robust solutions and optimal strategies, respectively.

Matrix Games A matrix game is a game played between two opponents. It is represented by a matrix M_0 where rows (respectively columns) correspond to possible actions for the row- (respectively column-) player, and the entry $(\mathsf{M}_0)_{i,j}$ is the reward the column-player pays the row-player when the pair of actions (i, j) is chosen. The two players choose actions simultaneously and independently, possibly randomizing their choices. Randomized strategies are called mixed strategies. The value of a matrix game, denoted $\operatorname{val} \mathsf{M}_0$, is the maximum amount the row-player can guarantee, that is, the amount one can obtain regardless of the column-player's strategy. By the min-max theorem [vN28], it is also the minimum amount the column-player can guarantee, and both players have optimal mixed strategies. The value can be computed by solving a linear program, which can be solved in polynomial time [Kha80].

Polynomial Matrix Games A polynomial matrix game M is described by $(K + 1)$ matrices, $\mathsf{M}_0, \mathsf{M}_1, \mathsf{M}_2, \ldots, \mathsf{M}_K$. We associate to each $\varepsilon \geq 0$ the reward $\mathsf{M}(\varepsilon) = \mathsf{M}_0 + \mathsf{M}_1\varepsilon + \ldots + \mathsf{M}_K\varepsilon^K$. The parameter ε represents the magnitude of a perturbation. The matrix game $M(0) = M_0$ is referred to as the unperturbed matrix game, whereas M_1 is the first-order perturbation, M_2 is the second-order perturbation, and so on.

Value-Positivity Problems We aim to understand how the value and the optimal strategies of a polynomial matrix game vary with the parameter. For every mixed strategy p of the row-player, we denote the amount that the row-player guarantees in the matrix game $M(\varepsilon)$ when playing the strategy p by $\operatorname{val}(\mathsf{M}(\varepsilon); p)$. We introduce value-positivity to capture the existence of strategies that are robust to perturbations. More precisely, we consider two properties: (i) *value-positivity* asks whether there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in [0, \varepsilon_0]$ there is a strategy p_ε such that $\operatorname{val}(\mathsf{M}(\varepsilon); p_\varepsilon) \geq \operatorname{val}(\mathsf{M}_0)$, and (ii) *uniform value-positivity* asks whether there exists a fixed strategy p and $\varepsilon_0 > 0$ such that $\operatorname{val}(\mathsf{M}(\varepsilon); p) \geq \operatorname{val}(\mathsf{M}_0)$ holds for all $\varepsilon \in [0, \varepsilon_0]$. Lastly, we consider the *functional form* problem, which consists of giving the analytical form of the function $\varepsilon \mapsto \operatorname{val} \mathsf{M}(\varepsilon)$, and of an optimal strategy, in some right-neighborhood of zero. Without loss of generality, we assume $\operatorname{val} \mathsf{M}_0 = 0$, as the value operator translates constants, that is, for every matrix A and constant $c \in \mathbb{R}$, we have that $\operatorname{val}(\mathsf{A} + c\mathsf{U}) = \operatorname{val}(\mathsf{A}) + c$, where U is a matrix of ones of the same size as A. This jusitfies the terminology of value-positivity and uniform value-positivity.

Relevance Both value-positivity and uniform value-positivity capture the existence of *robust strategies*, where robustness is to be understood as guaranteeing the value of the unperturbed value for all sufficiently small perturbations. Uniform value-positivity is a stronger notion where the strategy is fixed, as opposed to value-positivity where the strategy can depend on ε. Relying on the equivalence between matrix games and linear programming, we translate the notion of value-positivity to perturbed linear programming. Further, we exploit the connection between stochastic games and polynomial matrix games from [AOB19] to connect value-positivity with lower bounds on the discounted and undiscounted values of a stochastic game and uniform value-positivity with the existence of an optimal stationary strategy in undiscounted stochastic games.

Linear Programming Linear programming is an optimization problem where both the objective function and the (finitely many) constraints are linear [DT03, KT14, Roc70]. An

LP (P) is represented by a matrix A and vectors b and c; the objective is to maximize the function $x \mapsto c^\top x$ subject to the constraints $\mathsf{A}x \le b$ and $x \ge 0$. The value of the LP, denoted $\mathrm{val}(P)$, is $+\infty$ if the maximum is unbounded and $-\infty$ if constraints are infeasible, or it belongs to $\mathbb{R}$ otherwise. A vector is feasible in (P) if it satisfies the constraints. A solution of (P) is a feasible vector $x^* \in \mathbb{R}^n$ such that $\mathrm{val}(P) = c^\top x^*$. By the strong duality theorem [Adl13, DT03], LPs have a related dual LP. Moreover, as [BR23] puts it, LPs are equivalent to matrix games "both in the sense that a solution to any matrix game can be obtained by solving a suitably chosen LP problem and vice versa, as well as in the sense that their fundamental theorems, minimax and strong duality, each follow from the other".

Perturbed LPs Analogous to our approach for matrix games, we consider that A, b, c allow perturbations modeled by polynomials, that is, $\mathsf{A}(\varepsilon) = \mathsf{A}_0 + \mathsf{A}_1\varepsilon + \ldots + \mathsf{A}_k\varepsilon^k$, $b(\varepsilon) = b_0 + b_1\varepsilon + \ldots + b_k\varepsilon^k$, $c(\varepsilon) = c_0 + c_1\varepsilon + \ldots + c_k\varepsilon^k$. Denote $(P(\varepsilon))$ the LP given by the $\mathsf{A}(\varepsilon), b(\varepsilon)$ and $c(\varepsilon)$. We aim to understand how the value and the optimal strategies vary with the parameter, so we look into optimization-based interpretations of robustness. We consider how feasibility, optimal points, and the value function $\varepsilon \mapsto \mathrm{val}(P(\varepsilon))$ change for sufficiently small positive parameters. Similar to the value-positivity problems, we consider the following robustness problems: (i) the *weak robustness* problem asks whether there exists a sufficiently small $\varepsilon_0 > 0$ such that, for all $\varepsilon \le \varepsilon_0$, the corresponding LP is feasible and bounded, that is, if $\mathrm{val}(P(\varepsilon)) \in \mathbb{R}$; and (ii) the *strong robustness* problem asks whether there exists a fixed vector x^* and sufficiently small $\varepsilon_0 > 0$ such that, for all $\varepsilon \le \varepsilon_0$, the vector x^* is a solution of $(P(\varepsilon))$. Also, we consider the *functional form* problem, which consists of computing the analytical form of the functions $\varepsilon \mapsto \mathrm{val}(P(\varepsilon))$ and $\varepsilon \mapsto x^*_\varepsilon$.

Previous Results The continuity and Lipschitz property of the value function of matrix games is well known. Because the value function is not Fréchet-differentiable, the stability analysis of matrix games is a classic problem usually investigated through directional derivatives. Mills [Mil56] considered a linearly perturbed matrix game, that is, $\varepsilon \to \mathsf{M}_0 + \mathsf{M}_1\varepsilon$, also framed as a perturbation of M_0 in the direction of $\mathsf{M}_1\varepsilon$. A characterization was obtained for the right derivative of this function, together with a polynomial-time algorithm, based on the reduction to an auxiliary LP of similar size. Value-positivity and uniform value-positivity were never addressed before.

Our Contributions Our main contributions are the following.

1. We show that value-positivity and uniform value-positivity problems differ and cannot be derived from Mills [Mil56], even for linear perturbations, because the value can have a nonlinear dependency on the perturbation parameter (see Example 3.4.2).

2. We present polynomial-time algorithms to solve value-positivity, uniform value-positivity, and functional form problems for polynomial matrix games.

3. We apply this approach to perturbed linear programming to derive similar robustness results and show that weak (resp., strong) robustness for perturbed LPs has a polynomial-time reduction to value-positivity (resp., uniform value-positivity) for polynomial matrix games.

Related Work Perturbed matrix games go back to [Mil56], who characterized the right derivative of the value function. This result was extended to a broader class of games in [Sor03], using an operator approach. Perturbed games can be seen as a set of parametrized games; the regularity of the value function and the set of ε-optimal strategies were studied in [TV80] for a broad class of games. The asymptotic behavior of perturbed games has been studied in [AFFG01].

More generally, perturbation theory is a broad field and general perturbed linear programs have been studied [AFH13]. Previous research has focused on perturbations only to the objective function or the right-hand side of the constraints [Fia83, Fia97, Gal94, GGH97], while research on general perturbation that affects the entries of the matrix is sparse [Jer73b, Jer73a]. LPs with general perturbations can exhibit very different behaviors, for example, feasibility can change for small perturbations or the limit of optimal strategies as the perturbation vanishes is no longer feasible. Therefore, assumptions on the effect of the perturbation are usually taken. Checking whether these conditions are satisfied may take exponential time. In our case, we make the following structural assumption: both the primal and dual perturbed problems have uniformly bounded feasible regions. Under this class of perturbed LPs, our approach yields polynomial-time algorithms. For a general introduction on perturbation theory, see [AFH13] which also presents general computational results where the best algorithm so far takes exponential time in the worst case.

A particular class of perturbed LPs with efficient algorithms is described in [Ruh91] studying parametric flows, which boils down to studying perturbed LPs where only the objective function depends on the perturbation and therefore efficient methods can be applied. Imprecise matrix games include interval matrix games that have been studied, for example, in [VK20] where an exponential time algorithm is given to compute the value and optimal strategies.

Outline First, we present the analysis for value-positivity problems: Section 3.2 introduces basic definitions and classic results; Section 3.3 presents our main contribution; Section 3.4 clarifies the relationship between value-positive and uniform value-positive and their connection with the right derivative; and Section 3.5 presents polynomial-time algorithms for the value-positivity, uniform value-positivity, and functional form problems. Second, we present an analysis for perturbed LPs: Section 3.6 defines robustness problems for perturbed LPs and connect them to value-positivity for polynomial matrix games. Third, we present an analysis for stochastic games: Section 3.7 recalls the connection between stochastic games and polynomial matrix games, and connects value-positivity with lower bounds on the discounted and undiscounted values of a stochastic game and uniform value-positivity with the existence of an optimal stationary strategy in undiscounted stochastic games. Section 3.8 discusses a natural extensions of our results.

3.2 Preliminaries

In this section, we introduce basic definitions, classic results, and our main contribution.

3.2.1 Basic Definitions

For an integer $m \in \mathbb{N}$, we denote the set of integers from 1 to m by $[m] := \{1, 2, \dots, m\}$ and the set of probability distributions over $[m]$ by $\Delta[m]$. In the sequel, we consider matrices with rational entries. Let us start with basic definitions such as matrix games, linear programs, and some of their connections.

Definition 3.2.1 (Matrix Games)**.** A matrix game is given by a matrix M of size $m \times m$. Both players have the same strategy set $\Delta[m]$. Given $p \in \Delta[m]$, denote $(p^\top \mathsf{M})_j$ the reward the row-player obtains by using strategy p if the column-player plays action $j \in [m]$ in the matrix game M. The value of the matrix game M is

$$\operatorname{val} \mathsf{M} := \max_{p \in \Delta[m]} \min_{j \in [m]} (p^\top \mathsf{M})_j \,.$$

An *optimal strategy* for the row-player is a strategy $p \in \Delta[m]$ such that, for all $j \in [m]$, we have that $(p^\top \mathsf{M})_j \geq \operatorname{val} \mathsf{M}$. Properties of the value for matrix games, as well as the continuity of the value function, go back to von Neumann [vN28].

Definition 3.2.2 (Linear Program)**.** A Linear program (*LP*) of size $m \times m$ is given by

$$(P) \begin{cases} \max_x & c^\top x \\ s.t. & \mathsf{A}x \;\; \leq b \\ & \;\;\; x \;\; \geq 0\,, \end{cases}$$

where the matrix A has size $m \times m$. The value of the LP, denoted $\operatorname{val}(P)$, is $+\infty$ if the maximum is unbounded and $-\infty$ if constraints are infeasible, and belongs to $\mathbb{R}$ otherwise. We say a vector $x \in \mathbb{R}^n$ is feasible for (P) if $\mathsf{A}x \leq b$ and $x \geq 0$. An LP is feasible if it has a feasible vector. For a feasible LP with bounded value, a feasible vector $x^* \in \mathbb{R}^n$ is a solution if $\operatorname{val}(P) = c^\top x^* \geq c^\top x$ for every feasible vector x.

LP for a Matrix Game Computing the value of a matrix game M can be done in polynomial time by solving a linear program such that $\operatorname{val}(P_{\mathsf{M}}) = \operatorname{val} \mathsf{M}$ given as follows.

$$(P_{\mathsf{M}}) \begin{cases} \max_{p,z} & z & & \\ s.t. & (p^\top \mathsf{M})_j & \geq z, & \forall j \in [m] \\ & p & \in \Delta([m])\,. & \end{cases}$$

Simplifications For our purposes, it is enough to consider matrices M and A, and vectors b and c that have integer entries. This is without loss of generality by standard linear transformations. For example, for matrix games, the value function satisfies: (i) for all positive scalars λ, we have $\operatorname{val}(\lambda \mathsf{M}) = \lambda \operatorname{val} \mathsf{M}$; and (ii) for all scalars r, we have $\operatorname{val}(r + \mathsf{M}) = r + \operatorname{val} \mathsf{M}$. Also, for LPs, inequalities can be multiplied by positive constants and the objective function can be multiplied by positive scalars, as its value follows a similar relationship to the value of matrix games. Note that such linear transformations have no consequences in the complexity of algorithms, as the value of a matrix game with rational entries is the solution of an LP with rational entries and therefore has polynomial binary size. Finally, the fact that M and A are square is also without loss of generality, as, for matrix games, rows and columns can be duplicated and, for LPs, constraints can be duplicated and slack variables can be introduced.

We now turn to perturbed matrix games and the value-positivity problems.

Definition 3.2.3 (Polynomial Matrix Games)**.** A *polynomial matrix game* $\mathsf{M}(\cdot)$ of degree K and size $m \times m$ is a square matrix whose entries are polynomials of degree K with integer coefficients, that is, for all $\varepsilon \geq 0$,

$$\mathsf{M}(\varepsilon) = \mathsf{M}_0 + \mathsf{M}_1 \varepsilon + \ldots + \mathsf{M}_K \varepsilon^K \,,$$

where $\mathsf{M}_0, \mathsf{M}_1, \ldots, \mathsf{M}_K \in \mathbb{Z}^{m \times m}$. If $K = 1$, then $\mathsf{M}(\cdot)$ is called a *linear matrix game*.

Informally, the value-positivity problems consider a polynomial matrix game $\mathsf{M}(\cdot)$ and the goal is to compare $\operatorname{val}\mathsf{M}(\varepsilon)$ with $\operatorname{val}\mathsf{M}(0)$ for small positive ε. Intuitively, ε stands for the magnitude of small perturbations in a known direction. For simplicity and without loss of generality, we assume that $\operatorname{val}\mathsf{M}(0) = \operatorname{val}\mathsf{M}_0$ is zero. We study several variants of value-positivity of polynomial matrix games and the functional form problem.

Definition 3.2.4 (Value-Positivity Problem)**.** Given a polynomial matrix $\mathsf{M}(\cdot)$, with $\operatorname{val}\mathsf{M}(0) = 0$, the *value-positivity problem* is to determine, for each $\varepsilon \geq 0$ small enough, the existence of a strategy that ensures a reward greater or equal to zero for the row-player when playing $\mathsf{M}(\varepsilon)$. Formally, it consists of the following decision problem:

$$\exists \varepsilon_0 > 0 \quad \forall \varepsilon \in [0, \varepsilon_0] \quad \exists p_\varepsilon \in \Delta[m] \quad \forall j \in [m] \qquad (p_\varepsilon^\top \mathsf{M}(\varepsilon))_j \geq 0\,,$$

or equivalently, deciding if there exists ε_0 such that for all $\varepsilon \in [0, \varepsilon_0]$ we have that $\operatorname{val}\mathsf{M}(\varepsilon) \geq 0$. If the answer is yes, then $\mathsf{M}(\cdot)$ is called *value-positive*.

Definition 3.2.5 (Uniform Value-Positivity Problem)**.** Given a polynomial matrix $\mathsf{M}(\cdot)$, the *uniform value-positivity problem* is to determine the existence of a fixed strategy that ensures a reward greater or equal to zero for the row-player when playing $\mathsf{M}(\varepsilon)$ for all sufficiently small $\varepsilon \geq 0$. Formally, it consists of the following decision problem:

$$\exists p_0 \in \Delta[m] \quad \exists \varepsilon_0 > 0 \quad \forall \varepsilon \in [0, \varepsilon_0] \quad \forall j \in [m] \qquad (p_0^\top \mathsf{M}(\varepsilon))_j \geq 0\,.$$

If the answer is yes, then $\mathsf{M}(\cdot)$ is called *uniform value-positive*.

Definition 3.2.6 (Functional Form Problem)**.** Given a polynomial matrix $\mathsf{M}(\cdot)$, the *functional form problem* is to compute the function $\varepsilon \mapsto \operatorname{val}\mathsf{M}(\varepsilon)$ and a selection of optimal strategies for the row-player $\varepsilon \mapsto p^*(\varepsilon)$ in an interval of the form $[0, \varepsilon_0]$, for some $\varepsilon_0 > 0$.

Note that the value function of a polynomial matrix game is known to be a continuous piecewise rational function (see Theorem 3.2.9 below, by Shapley and Snow [SS50]). Therefore, a solution to the functional form problem is encoded by rational functions R and p^* such that, for all $\varepsilon \in [0, \varepsilon_0]$,

$$\operatorname{val}\mathsf{M}(\varepsilon) = R(\varepsilon)\,,$$

and $p^*(\varepsilon)$ is an optimal strategy of the matrix game $\mathsf{M}(\varepsilon)$.

The functional form of the value function $\varepsilon \mapsto \operatorname{val}\mathsf{M}(\varepsilon)$ has been studied by Mills [Mil56]. Specifically, Mills focused on computing the right derivative of the value of a perturbed matrix game, and therefore only linear matrix games were considered.

Definition 3.2.7 (Right Derivative of the Value Problem)**.** Given a linear matrix $\mathsf{M}(\varepsilon) = \mathsf{M}_0 + \mathsf{M}_1\varepsilon$, the *right derivative of the value problem* is to compute the right derivative of the function $\varepsilon \mapsto \operatorname{val}\mathsf{M}(\varepsilon)$ at zero. Formally,

$$\mathcal{D}\operatorname{val}\mathsf{M}(0^+) := \lim_{\varepsilon \to 0^+} \frac{\operatorname{val}\mathsf{M}(\varepsilon) - \operatorname{val}\mathsf{M}(0)}{\varepsilon}\,.$$

3.2.2 Classic Results

Mills states the following result.

Theorem 3.2.8 ([Mil56, Theorem 1])**.** *For linear matrix games $M(\varepsilon) = M_0 + M_1(\varepsilon)$ the following assertions hold.*

1. Right derivative characterization.

$$\mathcal{D}\,\mathrm{val}\,M(0^{+}) = \max_{p\in P(M_0)} \min_{q\in Q(M_0)} p^{\top} M_1 q\,,$$

where $P(M_0)$ and $Q(M_0)$ are the set of optimal strategies for the row- and column-player respectively for the matrix game M_0.

2. Right derivative computation. *We obtain $\mathcal{D}\,\mathrm{val}\,M(0^{+})$ by solving an explicit LP twice the size of $M(\cdot)$.*

Recall that, for a matrix $M \in \mathbb{R}^{m\times m}$, its cofactor of index $(i,j) \in [m]\times[m]$ is the determinant of the matrix obtained by deleting row i and column j from M, if $m > 1$, and 1 otherwise. The following result by Shapley and Snow is fundamental in our setting.

Theorem 3.2.9 (Value characterization [SS50])**.** *Consider a matrix game M. There exists a square submatrix $\overline{M}$ such that $\mathbb{1}^{\top}\,\mathrm{co}(\overline{M})\mathbb{1} \neq 0$ and*

$$\mathrm{val}\,M = \frac{\det\,\overline{M}}{\mathbb{1}^{\top}\,\mathrm{co}(\overline{M})\mathbb{1}}\,, \qquad \bar{p}^{*\top} = \frac{\mathbb{1}^{\top}\,\mathrm{co}(\overline{M})}{\mathbb{1}^{\top}\,\mathrm{co}(\overline{M})\mathbb{1}}\,,$$

where $\mathbb{1}$ is the vector of ones of the corresponding size, $\mathrm{co}(\overline{M})$ is the matrix of the cofactors of $\overline{M}$, and $\bar{p}^{*}$ is an optimal strategy of M when extended by zeros in the missing coordinates.

3.3 Main Contributions

Our main contribution is the following.

Theorem 3.3.1. *We present polynomial-time algorithms for the value-positivity, uniform value-positivity, and functional form problems.*

Significance By computing the functional form, we extend the result of Mills (which computes only the right derivative) and Shapley and Snow (which computes the value of unperturbed matrix games). Moreover, we solve the above-mentioned value-positivity problems which have a clear game-theoretical interpretation.

3.4 Examples and Implications

We clarify the relationship between value-positive and uniform value-positive and their connection with the right derivative.

Theorem 3.4.1 (Strict Implications of Problems)**.** *Consider a polynomial matrix game $M(\cdot)$.*

1. *If $M(\cdot)$ is uniform value-positive, then it is value-positive, but the converse does not hold.*
2. *For linear matrices $M(\varepsilon) = M_0 + M_1\varepsilon$,*
    a) *if $M(\cdot)$ is value-positive, then $\mathcal{D}\,\mathrm{val}\,M(0^{+}) \geq 0$, but the converse does not hold.*
    b) *if $M(\cdot)$ is uniform value-positive, then $\mathcal{D}\,\mathrm{val}\,M(0^{+}) \geq 0$, but the converse does not hold.*

c) if $\mathcal{D}\operatorname{val} M(0^+) > 0$, then $M(\cdot)$ is value-positive, but the converse does not hold.

d) $\mathcal{D}\operatorname{val} M(0^+) > 0$ does not imply that $M(\cdot)$ is uniform value-positive.

Proof. We prove the affirmative implications and the following examples show the implications that do not hold.

1. Assume that $\mathsf{M}(\cdot)$ is uniform value-positive. Then, there exists a strategy $p_0 \in \Delta[m]$ that ensures a positive value for small $\varepsilon \geq 0$. Because p_0 is one out of all possible strategies for the row-player, we have that $\operatorname{val} \mathsf{M}(\varepsilon) \geq \min_{j\in[m]}(p_0^\top \mathsf{M}(\varepsilon))_j \geq 0$, for ε small enough. Therefore, $\mathsf{M}(\cdot)$ is value-positive.
2. Consider a linear matrix $\mathsf{M}(\varepsilon) = \mathsf{M}_0 + \mathsf{M}_1\varepsilon$.
   a) Assume that $\mathsf{M}(\cdot)$ is value-positive. Then, for ε small enough, we have that $\operatorname{val} \mathsf{M}(\varepsilon) \geq 0$. Therefore, taking the limit as ε goes to zero, we have that
   $$\mathcal{D}\operatorname{val} \mathsf{M}(0^+) = \lim_{\varepsilon\to 0^+} \frac{\operatorname{val} \mathsf{M}(\varepsilon)}{\varepsilon} \geq 0\,.$$
   This proves that value-positivity implies a positive right derivative of the value function at zero.
   b) Assume that $\mathsf{M}(\cdot)$ is uniform value-positive. By Theorem 3.4.1-1, we have that $\mathsf{M}(\cdot)$ is value-positive. By Theorem 3.4.1-2-(a), we have that $\mathcal{D}\operatorname{val} \mathsf{M}(0^+) \geq 0$.
   c) Assume that $\mathcal{D}\operatorname{val} \mathsf{M}(0^+) > 0$. Because $\operatorname{val} \mathsf{M}(\cdot)$ is a smooth function in some right neighborhood of zero, $\mathcal{D}\operatorname{val} \mathsf{M}(0^+) > 0$ implies that $\operatorname{val} \mathsf{M}(\varepsilon) > 0$ for ε small enough. Therefore, $\mathsf{M}(\cdot)$ is value-positive.

□

A single example shows that the implications in items 1, 2-(b), and 2-(c) of Theorem 3.4.1 are tight. It consists of a linear matrix game such that: (i) the right derivative of the value function is zero, that is, $\mathcal{D}\operatorname{val} \mathsf{M}(0^+) = 0$; (ii) it is value-positive; and (iii) it is not uniform value-positive.

Example 3.4.2 (One-Way Implications)**.**

$$\mathsf{M}(\varepsilon) = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & -3 \\ 0 & 2 \end{pmatrix} \varepsilon\,.$$

For every $\varepsilon > 0$, the unique optimal strategy for the row-player is given by

$$p_\varepsilon = \left(\frac{1+\varepsilon}{2+3\varepsilon}, \frac{1+2\varepsilon}{2+3\varepsilon} \right)^\top\,.$$

Therefore,

$$\operatorname{val} \mathsf{M}(\varepsilon) = \frac{\varepsilon^2}{2+3\varepsilon}\,.$$

Hence, $\mathcal{D}\operatorname{val} \mathsf{M}(0^+) = 0$ and $\mathsf{M}(\cdot)$ is value-positive. However, note that there is no fixed strategy p_0 such that, for all small $\varepsilon > 0$, p_0 guarantees a positive value for the row-player. Indeed, consider some strategy $p \in \Delta[2]$. Then, either the probability of playing the top row under p is strictly less than $1/2$, in which case the column-player can choose the left column to ensure a negative payoff, or the probability of playing the top row is at least $1/2$, in which case the column-player can choose the right column to ensure a negative payoff for every $\varepsilon > 0$. Therefore, the polynomial matrix game $\mathsf{M}(\cdot)$ is not uniform value-positive.

The following example shows that the implication in item 2-(a) of Theorem 3.4.1 is tight. It consists of a linear matrix game such that: (i) the right derivative is zero; and (ii) it is not value-positive.

Example 3.4.3 (Zero Derivative but not Value-Positive)**.** Consider the following linear matrix game with two actions per player.

$$\mathsf{M}(\varepsilon) = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} \varepsilon \,.$$

For every $\varepsilon > 0$, the unique optimal strategy for the row-player is given by

$$p_\varepsilon = \left(\frac{1-\varepsilon}{2-\varepsilon}, \frac{1}{2-\varepsilon} \right)^\top \,.$$

Therefore,

$$\operatorname{val} \mathsf{M}(\varepsilon) = \frac{-\varepsilon^2}{2-\varepsilon} \,.$$

Hence, $\mathcal{D} \operatorname{val} \mathsf{M}(0^+) = 0$ and $\mathsf{M}(\cdot)$ is not value-positive.

The following example proves item 2-(d) of Theorem 3.4.1. It consists of a linear matrix game such that: (i) the right derivative is strictly positive; and (ii) it is not uniform value-positive.

Example 3.4.4 (Tightness of results)**.**

$$\mathsf{M}(\varepsilon) = \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix} + \begin{pmatrix} 2 & -1 \\ 0 & 0 \end{pmatrix} \varepsilon \,.$$

For every $\varepsilon > 0$, the unique optimal strategy for the row-player is given by

$$p_\varepsilon = \left(\frac{2}{2+3\varepsilon}, \frac{3\varepsilon}{2+3\varepsilon} \right)^\top \,,$$

and the value function is

$$\operatorname{val} \mathsf{M}(\varepsilon) = \frac{\varepsilon}{2+3\varepsilon} \,.$$

Hence, $\mathsf{M}(\cdot)$ is value-positive and $\mathcal{D} \operatorname{val} \mathsf{M}(0^+) = 1/2 > 0$. However, $\mathsf{M}(\cdot)$ is not uniform value-positive. This completes the proof of Theorem 3.4.1.

Implications Examples 3.4.2, 3.4.4 and 3.4.3 show the following interesting implications.

1. The value-positivity and uniform value-positivity problems are different and they are not captured by the right derivative provided by Mills [Mil56].
2. Even for linear matrix games the value function can be quadratic, see Example 3.4.2. In fact, for linear matrix games of size $m \times m$, the value function can be of order m. Therefore, again, the right derivative solves neither the value-positivity nor the uniform value-positivity problems.

Because the existing approaches of Mills [Mil56] do not solve the value-positivity problems, we consider algorithms for them in Section 3.5.

3.5 Algorithms

We present polynomial-time algorithms for the value-positivity, uniform value-positivity, and functional form problems.

3.5.1 Value-Positivity

The value-positivity problem consists of recognizing polynomial matrix games whose value is positive in some right neighborhood of zero.

Main Algorithmic Idea The key insight is that because the polynomials have integer coefficients, the value function is a continuous, piecewise rational function, see Theorem 3.2.9. In particular, there is a root separation so it is sufficient to know the sign of the value function at some $\varepsilon_0 > 0$ small enough. Algorithm 3.1 computes the small parameter ε_0, the value $\operatorname{val} \mathsf{M}(\varepsilon_0)$ and decides whether the polynomial matrix game is value-positive or not.

Algorithm 3.1: Value-Positivity Algorithm

```
Input: Polynomial matrix game $\mathsf{M}(\cdot)$ of degree $K$ and size $m \times m$
Output: true if the polynomial matrix game is value-positive; false otherwise
$B \leftarrow \max_{i,j,k}\{|\mathsf{M}_k^{i,j}|\}$ ;                    #Maximum absolute entry
$\varepsilon_0 \leftarrow (2m^5K^2B)^{-11m^2K}$ ;     #The sign of the value function does
  not change in $[0, \varepsilon_0]$
if $\operatorname{val} M(\varepsilon_0) \geq 0$ then
    return true;
end
return false;
```

Lemma 3.5.1 (Correctness and Complexity of Algorithm 3.1)**.** *Given a polynomial matrix game $M(\cdot)$, Algorithm 3.1 returns* true *if and only if $M(\cdot)$ is value-positive. Moreover, it runs in polynomial time.*

The proof Lemma 3.5.1 is given in Section 3.5.4.

3.5.2 Functional Form

The functional form problem consists of computing the functions $\varepsilon \mapsto \operatorname{val} \mathsf{M}(\varepsilon)$ and $\varepsilon \mapsto p^*(\varepsilon)$, where $p^*(\varepsilon)$ is an optimal strategy of the matrix game $\mathsf{M}(\varepsilon)$, in some right-neighborhood of zero. Note that, if there is more than one optimal strategy, then the problem imposes no restriction on which optimal strategy should be returned.

Main Algorithmic Idea By Theorem 3.2.9, these two functions are piecewise rational of bounded degree. Therefore, we only need to compute their coefficients. One can recover their coefficients from evaluations by solving a system of linear equations [Jac46]. Indeed, consider a rational function where the numerator and denominator have degrees N_d and D_d respectively,

$$f(\varepsilon) = \frac{a_0 + a_1\varepsilon + \ldots + a_{N_d}\varepsilon^{N_d}}{1 + b_1\varepsilon + \ldots + b_{D_d}\varepsilon^{D_d}} .$$

Then, the coefficients $a_0, a_1, \dots a_{N_d}$ and $b_1, b_2, \dots b_{D_d}$ can be computed by fitting a linear model of the form

$$y = a_0 + a_1 x + \ldots + a_{N_d} x^{N_d} - b_1 xy - \ldots - b_{D_d} x^{D_d} y \,.$$

This model has $(N_d + 1 + D_d)$ parameters. To fit this model and obtain the coefficients of the rational function, we only require $(N_d + 1 + D_d)$ evaluations, that is, pairs of points $(x_i, f(x_i))_{i \in [N_d+1+D_d]}$. On the one hand, for the value function, this reconstruction can be done from sufficiently many pairs of points $(\varepsilon, \operatorname{val} \mathsf{M}(\varepsilon))$. On the other hand, for optimal strategies, to compute points of a single rational function, we have to deal with one more problem, namely, the selection of the optimal strategy. We solve this problem by computing Shapley-Snow kernels, which also will provide us bounds on the degree and size of the coefficients of the model we fit.

Definition 3.5.2 (Shapley-Snow Kernel)**.** Consider a matrix game M of size $m \times m$. Then, every pair of sets of actions for each player $(\bar{I}, \bar{J})$, where $\bar{I} \subseteq [m]$ and $\bar{J} \subseteq [m]$, induces a smaller matrix game $\overline{\mathsf{M}}$ where players are restricted to choose an action from $\bar{I}$ and $\bar{J}$ respectively. A pair $(\bar{I}, \bar{J})$ is called a *Shapley-Snow kernel* if it induces a matrix game $\overline{\mathsf{M}}$ such that $\mathbb{1}^\top \operatorname{co}(\overline{\mathsf{M}}) \mathbb{1} \neq 0$ and

$$\operatorname{val} \mathsf{M} = \frac{det\, \overline{\mathsf{M}}}{\mathbb{1}^\top \operatorname{co}(\overline{\mathsf{M}}) \mathbb{1}} \,, \qquad \bar{p}^* = \frac{\operatorname{co}(\overline{\mathsf{M}})}{\mathbb{1}^\top \operatorname{co}(\overline{\mathsf{M}}) \mathbb{1}} \mathbb{1} \,,$$

where $\mathbb{1}$ is the vector of ones of the corresponding size, $\operatorname{co}(\overline{\mathsf{M}})$ is the matrix of cofactors of $\overline{\mathsf{M}}$, and $\bar{p}^*$ is the unique optimal strategy for the matrix game $\overline{\mathsf{M}}$.

Asymptotic Kernel By Theorem 3.2.9, every matrix game has at least one Shapley-Snow kernel. Moreover, there exists some pair of subsets that is a Shapley-Snow kernel for every $\varepsilon > 0$ sufficiently small. Indeed, consider the perturbed matrix game $\mathsf{M}(\cdot)$. For each $\varepsilon > 0$, the matrix game $\mathsf{M}(\varepsilon)$ has a Shapley-Snow kernel associated to a submatrix $\overline{\mathsf{M}(\varepsilon)}$ and

$$\operatorname{val} \mathsf{M}(\varepsilon) = \frac{\det\, \overline{\mathsf{M}(\varepsilon)}}{\mathbb{1}^\top \operatorname{co}(\overline{\mathsf{M}(\varepsilon)}) \mathbb{1}} \,.$$

Note that there are at most $(2^m - 1) \cdot (2^m - 1)$ rational functions of the form $\varepsilon \mapsto \frac{det\, \overline{\mathsf{M}(\varepsilon)}}{\mathbb{1}^\top \operatorname{co}(\overline{\mathsf{M}(\varepsilon)}) \mathbb{1}}$, one for each possible Shapley-Snow kernel. Moreover, two different such functions can coincide in only finitely many points. Therefore, because the function $\operatorname{val} \mathsf{M}(\cdot)$ is continuous, there exists a fixed pair $(\bar{I}, \bar{J})$ that is a Shapley-Snow kernel for every $\varepsilon > 0$ sufficiently small. By computing this pair, we obtain a selection of optimal strategies for every $\varepsilon \geq 0$ sufficiently small.

Computing a Kernel If the unperturbed matrix game has several optimal strategies, then it also necessarily has several Shapley-Snow kernels. Computing a Shapley-Snow kernel of a matrix game reduces to finding a basic solution for an LP. A polynomial-time algorithm that computes a basic solution from an arbitrary solution of an LP is presented in [Kar91, proof of Theorem 10, pages 18–20]. Furthermore, computing an optimal basis for an LP given a pair of optimal primal and dual solutions is strongly polynomial [Meg91]. Therefore, a basic solution of an LP, and so a Shapley-Snow kernel, can be found in polynomial time.

Efficiency of Fitting To solve the functional form problem, we need to compute the coefficients of the corresponding rational functions. We do so by fitting a rational function to sufficiently many points. This is formalized in Algorithm 3.2. An alternative procedure to solve the functional form problem would be to first obtain a pair $(\bar{I}, \bar{J})$ that is a Shapley-Snow kernel for all ε small enough and then use Definition 3.5.2 to compute it. This alternative procedure involves computing the determinant of a matrix of polynomials, which requires exponential time in general. Therefore, fitting a rational function is more efficient.

Algorithm 3.2: Functional Form Algorithm

```
Input: Polynomial matrix game $\mathsf{M}(\cdot)$ of degree $K$ and size $m \times m$
Output: Analytical form of the value function and an optimal strategy in some right
        neighborhood of zero
$B \leftarrow \max_{i,j,k}\{|\mathsf{M}_k^{i,j}|\}$;
$\varepsilon_0 \leftarrow (2m^5 K^2 B)^{-11m^2 K}$;
$x \leftarrow \left(\varepsilon_0, \frac{\varepsilon_0}{2}, \dots, \frac{\varepsilon_0}{2mK+1}\right)$;
$y \leftarrow \left(\mathrm{val}\,\mathsf{M}\left(\varepsilon_0\right), \mathrm{val}\,\mathsf{M}\left(\frac{\varepsilon_0}{2}\right), \dots, \mathrm{val}\,\mathsf{M}\left(\frac{\varepsilon_0}{2mK+1}\right)\right)$;
val $\leftarrow$ Rational fit with numerator and denominator of degree at most $mK$ from
  $(x, y)$;
$(\bar{I}, \bar{J}) \leftarrow$ Shapley-Snow kernel of $\mathsf{M}(\varepsilon_0)$ ;
$\overline{\mathsf{M}}(\cdot) \leftarrow$ square submatrix corresponding to $(\bar{I}, \bar{J})$;
for $i \in [2mK + 1]$ do
    $z_i \leftarrow$ Optimal solution of $P_{\overline{\mathsf{M}}(\frac{\varepsilon_0}{i})}$;
end
$\bar{p}^* \leftarrow$ Rational fit with numerator and denominator of degree at most $mK$ from $(x, z)$;
$p^* \leftarrow$ extension of $\bar{p}^*$ by zero;
return $\mathrm{val}, p^*$;
```

Lemma 3.5.3 (Correctness and complexity of Algorithm 3.2)**.** *Given a polynomial matrix game M$(\cdot)$, Algorithm 3.2 returns the functional form of the value and an optimal strategy of M$(\cdot)$ in some right neighborhood of zero. Moreover, it runs in polynomial time.*

The proof of Lemma 3.5.3 is given in Section 3.5.4.

3.5.3 Uniform Value-Positivity

The uniform value-positivity problem consists of recognizing polynomial matrix games with a fixed strategy p_0 for the row-player that guarantees a positive value in some right neighborhood of zero. In other words, decide if the row-player has a strategy that, for all $\varepsilon \geq 0$ small enough, guarantees at least the value of the unperturbed matrix game.

Leading Coefficient An equivalent characterization of a strategy p_0 that guarantees uniform value-positivity is as follows. First, note that a strategy p_0 for the row-player and a pure action of the column-player $j \in [m]$ determine a polynomial payoff $\varepsilon \mapsto (p_0^\top \mathsf{M}(\varepsilon))_j$. The sign of this polynomial in some right neighborhood of zero is determined by its *leading coefficient*, that is, the first nonzero coefficient ordered by the degree of the corresponding monomial it multiplies. Therefore, a witness of uniform value-positivity is a strategy p_0 for which all the corresponding polynomial payoffs (one for each action of the column-player) either have a strictly positive leading coefficient or are constant to zero.

Frontier Consider a polynomial matrix game $\mathsf{M}(\cdot)$ and assume it is uniform value-positive. Note that, if for some action j of the column-player the corresponding polynomial payoff $\varepsilon \mapsto (p_0^\top \mathsf{M}(\varepsilon))_j$ has a leading coefficient of order, for example, 3, then changing the perturbation coefficients at column j of order 4 or more does not change the fact that the polynomial matrix game is uniform value-positive. Therefore, for each witness strategy p_0 for $\mathsf{M}(\cdot)$, there is a sequence of indices (one per column) containing the corresponding leading coefficients' order of perturbation. We represent these indices by a vector $\boldsymbol{k} \in [K+1]^m$, where the index $(K+1)$ corresponds to the polynomial payoff being constant to zero. We call such a sequence a *frontier*.

A useful visualization of the frontier of a strategy p of the row-player is the following. Consider the matrix given by $\left((p^\top \mathsf{M}_k)_j \right)_{k,j\in[m]}$. In other words, each column corresponds to an action of the column-player and each row corresponds to the coefficient of the respective polynomial when the row-player plays according to p. The leading coefficient of the strategy p on a column corresponds to the first (as k increases) nonzero entry, if there is some. The frontier of p represents the indices where the leading coefficients are in this matrix. Note that, if all columns of this matrix have a strictly positive leading coefficient, or are full of zeros, then p guarantees the uniform value-positivity of $\mathsf{M}(\cdot)$. Figure 3.1 visualizes the matrix given by $\left((p^\top \mathsf{M}_k)_j \right)_{k,j\in[m]}$ in a hypothetical example.

$$\begin{array}{cccc} & 1 & 2 & 3 \\ M_0 & 0 & 0 & 0 \\ M_1 & 0 & \boxed{2} & 0 \\ M_2 & 0 & -1 & \boxed{-1} \\ M_3 & \boxed{0} & 0 & 0 \end{array} .$$

Figure 3.1: Illustration of $\left((p^\top \mathsf{M}_k)_j \right)_{k,j\in[m]}$ for a hypothetical example with $K = 2$ and $m = 3$. Highlighted entries represent the frontier of this strategy $\boldsymbol{k} = (3, 1, 2)$. The first column has no leading coefficient as it is full of zeros. The second column has a leading coefficient of 2. The third column has a leading coefficient of -1.

Verification of a Frontier Given a frontier, deciding if there is a strategy p_0 with this frontier and positive leading coefficients is polynomial-time: it consists of solving at most m LPs. Indeed, for a frontier $\boldsymbol{k} \in [K+1]^m$ and an action of the column-player $j_0 \in [m]$, which we call *focus action*, define the following LP.

$$(P_{\boldsymbol{k},j_0}) \begin{cases} \max_p & (p^\top \mathsf{M}_{\boldsymbol{k}_{j_0}})_{j_0} \\ s.t. & (p^\top \mathsf{M}_k)_j \geq 0, \qquad \forall j \in [m], k \leq \boldsymbol{k}_j \\ & p \in \Delta([m]) \end{cases}$$

Note that, if $(P_{\boldsymbol{k},j_0})$ is feasible, then its value is at least zero. Figure 3.2 visualizes the constraints and objective function in $(P_{\boldsymbol{k},j_0})$. The LP $(P_{\boldsymbol{k},j_0})$ computes a strategy whose frontier is at least $\boldsymbol{k}$ and maximizes the candidate leading coefficient of the polynomial corresponding to the focus action j_0. To verify if a given frontier has a witness strategy, we do as follows.

1. Check if there is a strategy making all the coefficients up to the frontier at least zero, discarding the frontier if there is none.

2. Iterate over all focus actions, maximizing the corresponding coefficient in the frontier. For each focus action, the maximum value of the corresponding coefficient is either zero or strictly positive.

3. If all focus actions returned a strictly positive value, then there is a witness for the frontier. If some focus action returned zero (and $\boldsymbol{k}_{j_0} < K+1$), then there is no witness with this frontier.

Lexicographic Search The problem with frontiers is that there are $(K+1)^m$ of them, that is, exponentially many. Therefore, we must search efficiently among all possible frontiers. Algorithm 3.3 implements a search pattern that considers at most $(K+1)m$ frontiers. During this search pattern, the coordinates of the candidate frontier are always increasing, exploiting properties of the verification of a frontier. Indeed, while verifying a frontier, if a focus action returns zero, then the search can safely continue to the next frontier in our search pattern. Altogether, the algorithm decides whether the polynomial matrix game is uniform value-positive or not by solving at most $(K+1)m^2$ LPs. This is formally presented in Algorithm 3.3.

$$\begin{array}{c c}
 & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\
\begin{array}{c} M_0 \\ M_1 \\ M_2 \\ M_3 \end{array} &
\begin{pmatrix}
\geq 0 & \geq 0 & \geq 0 \\
\geq 0 & \boxed{\geq 0} & \geq 0 \\
\geq 0 & & \geq 0 \\
\geq 0 & &
\end{pmatrix}
\end{array}$$

Figure 3.2: Illustration of $(P_{\boldsymbol{k},j_0})$ for a hypothetical frontier $\boldsymbol{k} = (3,1,2)$ and focus column $j_0 = 2$. The coordinates with entries ≥ 0 correspond to the constraints of the program, that is, coordinates $(k,j) \in \{1,2,3\} \times \{0,1,2,3\}$ such that $k \leq \boldsymbol{k}_j$. The highlighted entry at the coordinate $(\boldsymbol{k}_{j_0}, j_0)$ corresponds to the objective function of the program.

Lemma 3.5.4 (Correctness and Complexity of Algorithm 3.3)**.** *Given a polynomial matrix game M(·), Algorithm 3.3 returns $true$ if and only if M(·) is uniform value-positive. Moreover, it runs in polynomial time.*

The proof Lemma 3.5.4 is given in Section 3.5.4.

Example 3.5.5 (Running the Uniform Value-Positivity Algorithm)**.** Consider the following polynomial matrix game.

$$\mathsf{M}(\varepsilon) = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & -3 \\ 0 & 2 \end{pmatrix} \varepsilon\,.$$

This polynomial matrix game is not uniform value-positive. Indeed, M_0 has a unique optimal strategy $p^* = (1/2, 1/2)^\top$ which, for all $\varepsilon > 0$, does not guarantee zero in $M(\varepsilon)$. Let us show how Algorithm 3.3 arrives at this conclusion. Recall that Algorithm 3.3 iterates over frontiers to make a decision. For each frontier, it makes a feasibility check and then checks each focus action. For each focus action, if the value of the related LP is zero, then it advances the frontier. If the value is strictly positive, it continues to the next focus action. During this procedure, each coordinate of the frontier increases. We now go over this procedure explicitly.

Algorithm 3.3: Uniform Value-Positivity Algorithm

```
   Input: Polynomial matrix game M(·) of degree $K$ and size $m \times m$
   Output: true if the polynomial matrix game is uniform value-positive; false otherwise
1  $\boldsymbol{k} \leftarrow (0,0,\dots,0) \in [K+1]^m$ ;              #Initialize the frontier
2  while true do
3      if $(P_{\boldsymbol{k},1})$ is infeasible ;        #Test feasibility of the frontier
4       then
5          return false ;                         #OUPUT: There is no witness
6      end
7      if $(P_{\boldsymbol{k},1})$ is infeasible ;        #Test feasibility of the frontier
8       then
9          return false ;                         #OUPUT: There is no witness
10     end
11     for $j_0 \in [m]$ ;            #Iterate over column-player's actions
12      do
13         if $\boldsymbol{k}_{j_0} = K+1$ ;  #This action has been fully investigated
14          then
15             continue the for loop ;    #Consider the next focus action
16         end
17         if $\mathrm{val}(P_{\boldsymbol{k},j_0}) == 0$ ;  #Higher order terms need investigation
18          then
19             $\boldsymbol{k} \leftarrow \boldsymbol{k} + e_{j_0}$ ;                     #Advance the frontier
20             goto to Line 3 ;              #Investigate the new frontier
21         end
22         if $\mathrm{val}(P_{\boldsymbol{k},j_0}) > 0$ ;              #The focus action is covered
23          then
24             continue the for loop ;    #Consider the next focus action
25         end
26     end
27     return true ;    #OUPUT: The frontier has a witness strategy
28 end
```

Frontier $(0,0)$**, Action** 1 The initial frontier is $\boldsymbol{k} = (0,0)$. Write a strategy p as $(x, 1-x)^\top$. The first LP to check feasibility is the following:

$$(P_{(0,0),1}) \begin{cases} \begin{array}{lll} \max_x & x \cdot 1 + (1-x) \cdot (-1) & \\ s.t. & x \cdot 1 + (1-x) \cdot (-1) & \geq 0 \\ & x \cdot (-1) + (1-x) \cdot 1 & \geq 0 \\ & x & \in [0,1] \,. \end{array} \end{cases}$$

This LP is feasible, so we proceed to check each focus action. The first column-player's action we investigate is $j_0 = 1$. We compute the value of $(P_{(0,0),1})$. Its only solution is $p^* = (1/2, 1/2)^\top$ and its value is zero. Therefore, we increase the frontier by one in its j_0-th coordinate. The new frontier is $\boldsymbol{k} = (1,0)$.

Frontier $(1,0)$**, Action** 1 The second LP to consider is the following:

$$(P_{(1,0),1}) \begin{cases} \max_x & x \cdot 1 \\ s.t. & x \cdot 1 + (1-x) \cdot (-1) \geq 0 \\ & x \cdot 1 \geq 0 \\ & x \cdot (-1) + (1-x) \cdot 1 \geq 0 \\ & x \in [0,1] \,. \end{cases}$$

Again, it is feasible, so we check each focus action. Considering $j_0 = 1$ and the corresponding LP, the only solution is $p^* = (1/2, 1/2)^\top$ and its value is strictly positive. Therefore, we continue with the next action of the column-player $j_0 = 2$.

Frontier $(1,0)$**, Action** 2 The third LP to solve is the following:

$$(P_{(1,0),2}) \begin{cases} \max_x & x \cdot (-1) + (1-x) \cdot 1 \\ s.t. & x \cdot 1 + (1-x) \cdot (-1) \geq 0 \\ & x \cdot 1 \geq 0 \\ & x \cdot (-1) + (1-x) \cdot 1 \geq 0 \\ & x \in [0,1] \,. \end{cases}$$

Note that only the objective function changed with respect to $(P_{(1,0),1})$. Therefore, its only solution is still $p^* = (1/2, 1/2)^\top$ and its value is zero, so we increase the frontier by one in its j_0-th coordinate. The new frontier is $\boldsymbol{k} = (1,1)$.

Frontier $(1,1)$**, Action** 1 The fourth LP to solve is the following:

$$(P_{(1,1),1}) \begin{cases} \max_x & p \cdot 1 \\ s.t. & x \cdot 1 + (1-x) \cdot (-1) \geq 0 \\ & x \cdot 1 \geq 0 \\ & x \cdot (-1) + (1-x) \cdot 1 \geq 0 \\ & x \cdot (-3) + (1-x) \cdot 2 \geq 0 \\ & x \in [0,1] \,. \end{cases}$$

This last LP is infeasible. Therefore, the algorithm outputs *false*, correctly indicating that this polynomial matrix game is not uniform value-positive.

3.5.4 Detailed Proofs

In this section, we prove Lemma 3.5.1, Lemma 3.5.3 and Lemma 3.5.4, which jointly prove our main result Theorem 3.3.1. We start by recalling classic results on polynomials and follow with the technical proofs.

Classic Results

We recall a classic result on the minimal separation of roots of polynomials.

Lemma 3.5.6 (Bounding Polynomial Roots [Cau32, Mah64])**.** *Consider a nonzero polynomial with integer coefficients $r(\varepsilon) = a_1\varepsilon + a_2\varepsilon^2 + \ldots + a_L\varepsilon^L$. Then, the first strictly positive root is bounded away from zero. Formally, if $r(\varepsilon_0) = 0$ and $\varepsilon_0 > 0$, then*

$$\varepsilon_0 \geq L^{-(L+2)/2} \|r\|_2^{1-L} \,,$$

where $\|r\|_2^2 := \sum_{i \in [L]} a_i^2$.

We deal with polynomials that correspond to the determinant of a matrix with polynomial entries and, to apply Lemma 3.5.6, we need to bound their degree and the size of their coefficients. Denote the binary size of x by $\operatorname{bit}(x)$, that is, if x is a number, it is $\lceil \log_2(|x|+1) \rceil$, if x is a collection of numbers, then it is the sum of their binary size. We recall the following classic result, which is a consequence of [BPR06, Proposition 8.12].

Lemma 3.5.7 (Order of Matrix Determinant [BPR06])**.** *Consider a polynomial matrix game $\mathsf{M}(\cdot)$ of size $m \times m$, order K, and integer entries of size bounded by B. The function $det\, \mathsf{M}(\cdot)$ is a polynomial of degree at most mK and coefficients of binary size at most $m \operatorname{bit}(\mathsf{M}) + m \operatorname{bit}(m) + \operatorname{bit}(mK+1)$.*

Technical Proofs

We show a characterization of the value function of polynomial matrix games in an explicit, at most exponentially small, right neighborhood of zero. Although similar results can be found in the literature (see [SCFV97, Lemma 5.1, page 874] and [OB20]), we remark on the importance of the numerical bounds for our algorithms.

Lemma 3.5.8 (Value Characterization Close to Zero)**.** *Consider a polynomial matrix game $\mathsf{M}(\cdot)$ of size $m \times m$, order K, and integer entries of size bounded by B. Then, there exists a rational function R_1, a rational vector function R_2, and a threshold $\varepsilon_0 \geq (2m^5K^2B)^{-11m^2K}$ such that, for all $\varepsilon \in [0, \varepsilon_0]$,*

$$\operatorname{val} \mathsf{M}(\varepsilon) = R_1(\varepsilon) \qquad p^*(\varepsilon) = R_2(\varepsilon) \, .$$

Moreover, the functions R_1 and each coordinate of R_2 can be expressed as the quotient of two polynomials r/s, where r and s are at most of degree mK, the numerator r has no roots in $(0, \varepsilon_0]$ unless it is constant to zero, s is strictly positive in $[0, \varepsilon_0]$, and $s(0) = 1$.

Proof. Let $\mathsf{M}(\cdot)$ be a polynomial matrix game of size $m \times m$, order K, and integer entries of size bounded by B. By Theorem 3.2.9, for every $\varepsilon \geq 0$, there exists a square submatrix $\overline{\mathsf{M}(\varepsilon)}$ such that $\mathbb{1}^\top \operatorname{co}(\overline{\mathsf{M}(\varepsilon)})\mathbb{1} \neq 0$ and

$$\operatorname{val} \mathsf{M}(\varepsilon) = \frac{det\, \overline{\mathsf{M}(\varepsilon)}}{\mathbb{1}^\top \operatorname{co}(\overline{\mathsf{M}(\varepsilon)})\mathbb{1}} \, . \tag{3.1}$$

Because there are finitely many square submatrices and $\operatorname{val} \mathsf{M}(\cdot)$ is continuous, $\operatorname{val} \mathsf{M}(\cdot)$ is a piecewise rational function. Define the threshold $\varepsilon_0 := (2m^5K^2B)^{-11m^2K}$. We prove that $\operatorname{val} \mathsf{M}(\cdot)$ is a rational function in $[0, \varepsilon_0]$.

Denote $\mathcal{R}$ the set of all rational functions obtained as one varies the submatrix $\overline{\mathsf{M}}$ in equation (3.1). Note that two different such rational functions intersect only finitely many times. We claim that ε_0 is a lower bound on the closest parameter to zero where an intersection happens. Indeed, by Lemma 3.5.7, $\det \overline{\mathsf{M}(\varepsilon)}$ and the coordinates of $\operatorname{co}(\overline{\mathsf{M}(\varepsilon)})$ are polynomials of degree at most mK and coefficients of binary size at most $m \operatorname{bit}(\mathsf{M}) + m \operatorname{bit}(m) + \operatorname{bit}(mK+1)$. Therefore, for all $R \in \mathcal{R}$, we have that, $R = r/s$ where r and s are polynomials of degree at most mK and coefficients of binary size at most $m \operatorname{bit}(\mathsf{M}) + (m+2) \operatorname{bit}(m) + \operatorname{bit}(mK+1)$. Then, we have that if $R \neq R' \in \mathcal{R}$ and $\varepsilon_0 > 0$ are such that $R(\varepsilon_0) = R'(\varepsilon_0)$, then ε_0 is the root of a polynomial of degree at most $2mK$ whose coefficients have binary size at most

$2m\,\mathrm{bit}(\mathsf{M}) + 2(m+2)\,\mathrm{bit}(m) + 4\,\mathrm{bit}(mK+1)$. Then, applying Lemma 3.5.6 to $L = 2mK$ and $\log(\|r\|_2) \le \log(2mK) + 2m\,\mathrm{bit}(\mathsf{M}) + 2(m+2)\,\mathrm{bit}(m) + 4\,\mathrm{bit}(mK+1)$, we get that

$$\begin{aligned} \log(\varepsilon_0) &> -(mK+1)\log(2mK) \\ &\qquad - 4mK(m+2)\left(\log(2mK) + \mathrm{bit}(\mathsf{M}) + \mathrm{bit}(m) + \mathrm{bit}(mK+1)\right) , \\ &\ge -(mK+1)\log(2mK) - 9m^2K\log(2m^5K^2B) , \\ &\ge -11m^2K\log(2m^5K^2B) , \end{aligned}$$

that is, $\varepsilon_0 \ge (2m^5K^2B)^{-11m^2K}$. In conclusion, there exists $R_1 \in \mathcal{R}$ such that, for all $\varepsilon \in [0, \varepsilon_0]$,

$$\mathrm{val}\,\mathsf{M}(\varepsilon) = R_1(\varepsilon) .$$

Moreover, either R_1 is constant to zero or it has no roots in $(0, \varepsilon_0]$. The rest of the properties of R_1 are achieved by renormalization.

For the optimal strategy p^*, recall that the submatrix $\overline{\mathsf{M}}$ is given by a pair of indices $(\bar{I}, \bar{J})$. By Theorem 3.2.9, an optimal strategy p^* is recovered as the extension by zeros outside the rows of $\bar{I}$ of the rational vector function

$$\varepsilon \mapsto \frac{\mathrm{co}(\overline{\mathsf{M}}(\varepsilon))}{\mathbb{1}^\top \mathrm{co}(\overline{\mathsf{M}}(\varepsilon))\mathbb{1}}\mathbb{1} .$$

Therefore, p^* can be given by another rational vector function $R_2 = r/s$. The properties of r and s are proven the same way as for the value function.

□

Lemma 3.5.8 allow us to prove Lemma 3.5.1 as follows.

Proof of Lemma 3.5.1. We show the correctness and complexity of Algorithm 3.1. By Lemma 3.5.8, the value function does not change its sign between zero and $\varepsilon_0 = (2m^5K^2B)^{-11m^2K}$. Therefore, the sign of $\mathrm{val}\,\mathsf{M}(\varepsilon_0)$ is the sign of the value function in some right neighborhood of zero. This proves that Algorithm 3.1 is correct.

Computing the value of a matrix game takes polynomial time in terms of the binary size of the matrix game. In this case, the binary size of $\mathsf{M}(\varepsilon_0)$ is at most the binary size of the polynomial matrix game plus the binary size of ε_0^K, which is at most $11m^2K^2\lceil\log_2(2m^5K^2B)\rceil$. Therefore, computing the value of the matrix game $\mathsf{M}(\varepsilon_0)$ takes only polynomial time and so Algorithm 3.1 runs in polynomial time.

□

Before we prove Lemma 3.5.3, we clarify how Algorithm 3.2 performs Line 6, namely, computing a Shapley-Snow kernel of a matrix game in polynomial time.

Lemma 3.5.9. *Consider a matrix game M of size $m \times m$ and integer entries of size bounded by B. A Shapley-Snow kernel can be computed in polynomial time in terms of m and B.*

Recall that Shapley-Snow kernels are supports of optimal strategies of minimal support size and are closely related to basic solutions of LPs [SS50], that is, solutions uniquely defined by a subset of linearly independent constraints. We compute a Shapley-Snow kernel by computing basic solutions of LPs.

Proof of Lemma 3.5.9. Consider a matrix game M of size $m \times m$ and integer entries of size bounded by B. First, we relate a basic solution of an LP to the support of an optimal strategy in the matrix game. Then, we give an elimination procedure to reduce the size of the support until we find a Shapley-Snow kernel.

Basic Solutions as the Support of Optimal Strategies For the LP (P_{M}), constraints correspond to either: (i) restrictions given by actions of the column-player; or (ii) restrictions forcing p to be a distribution over rows. Therefore, a basic solution relates to a subset of actions of the column-player $\bar{J}$, which contains the support of an optimal strategy for the column-player, as these actions can force the value. Because $\bar{J}$, contains the support of an optimal strategy, there is a Shapley-Snow kernel $(\bar{I}', \bar{J}')$ such that $\bar{J}' \subseteq \bar{J}$. Similarly, from a basic solution of the LP $(P_{-\mathsf{M}^\top})$, we obtain a candidate set $\bar{I}$ for the row-player. Although $(\bar{I}, \bar{J})$ might not be a Shapley-Snow kernel, it always contains one. Therefore, through a suitable elimination procedure, we extract a Shapley-Snow kernel.

Elimination Procedure Consider a pair $(\bar{I}, \bar{J})$ candidate for a Shapley-Snow kernel. While eliminating actions from $\bar{I}$ or $\bar{J}$, we distinguish two cases: (a) $|\bar{I}| \neq |\bar{J}|$; and (b) $|\bar{I}| = |\bar{J}|$ but $\mathbb{1}^\top \operatorname{co}(\bar{\mathsf{M}})\mathbb{1} = 0$. If none of these cases hold, then $(\bar{I}, \bar{J})$ is a Shapley-Snow kernel. Therefore, we show how to eliminate actions in each case until we find a Shapley-Snow kernel.

For case (a), without loss of generality, consider $|\bar{I}| > |\bar{J}|$. Recall that $\bar{I}$ corresponds to the support of an optimal strategy. Therefore, we correctly eliminate an action from $\bar{I}$ to find a smaller support also containing an optimal strategy as follows. We solve $|\bar{I}|$ many LPs, one for each action, where the corresponding action is forced to have probability zero (and therefore eliminated from the support). Because $\bar{I}$ is not of minimal size, at least one of these LPs has the same value as the original LP. Thus, we obtain a new candidate set $\bar{I}' \subsetneq \bar{I}$. Repeating this procedure we obtain a candidate pair $(\bar{I}, \bar{J})$ such that $|\bar{I}| = \bar{J}|$.

For case (b), recall that $(\bar{I}, \bar{J})$ contains a Shapley-Snow kernel. Therefore, we eliminate one of the actions from $\bar{I}$ as done in case (a) to obtain a new candidate set $\bar{I}' \subsetneq \bar{I}$ that contains an optimal strategy and go back to case (a).

Iterating this procedure, we obtain a Shapley-Snow kernel $(\bar{I}, \bar{J})$. Following the approach in the proof of [Kar91, Theorem 10, pages 18–20], a basic solution of (P_{M}) can be computed in polynomial time. Therefore, computing a Shapley-Snow kernel also takes only polynomial time.

□

We now proceed to the proof of Lemma 3.5.3, that is, the correctness and complexity of Algorithm 3.2. Recall that Algorithm 3.2 computes the value function by fitting $(2mK+1)$ points and an optimal strategy function by: (i) computing a Shapley-Snow kernel of the matrix game $\mathsf{M}(\varepsilon_0)$; (ii) computing optimal strategies for the corresponding subgame at various values of ε; (iii) fitting a rational function to these optimal strategies; and (iv) extending by zeros to a strategy of the original game.

Proof of Lemma 3.5.3. We consider first the value function and then the optimal strategy function.

Value Function By Lemma 3.5.8, the functional form of the value function is given by a rational function with a numerator and denominator of degree at most mK. Therefore, its coefficients are uniquely determined by $(2mK+1)$ values, for example, its value at the points $\varepsilon_0, \varepsilon_0/2, \ldots, \varepsilon_0/(2mk+1)$. The computation of this fitting is polynomial-time as it corresponds to solving a system of linear equations, which can be done in polynomial time.

Optimal Strategy By Lemma 3.5.9, $(\bar{I}, \bar{J})$ given in Line 6 is a Shapley-Snow kernel of $\mathsf{M}(\varepsilon_0)$ obtained in polynomial time. By Lemma 3.5.8, for all $\varepsilon \in [0, \varepsilon_0]$, the pair $(\bar{I}, \bar{J})$ is a Shapley-Snow kernel and an optimal strategy of the reduced game $\overline{\mathsf{M}(\varepsilon)}$ is optimal in $\mathsf{M}(\varepsilon)$ when extended by zeros and each of its coordinates is a rational function with a numerator and denominator of degree at most mK. Therefore, their coefficients are uniquely determined by $(2mK+1)$ values, for example, their values at the points $\varepsilon_0, \varepsilon_0/2, \ldots, \varepsilon_0/(2mk+1)$. Algorithm 3.2 computes these values as optimal strategies of $\overline{\mathsf{M}(\varepsilon)}$ are unique by definition of Shapley-Snow kernel. Therefore, Algorithm 3.2 recovers an optimal strategy of the subgame by fitting a rational function to each coordinate. Lastly, as already mentioned, extending it by zeros in the coordinates outside of $\bar{I}$ leads to an optimal strategy $p^*(\varepsilon)$.

□

The proof of Lemma 3.5.4 is more involved and requires a few more technical lemmas. We start by recalling the LP used in Algorithm 3.3. For a vector $\boldsymbol{k} \in [K]^m$ and an index $j_0 \in [m]$, we consider the following LP.

$$(P_{\boldsymbol{k},j_0}) \begin{cases} \max_p & (p^\top \mathsf{M}_{\boldsymbol{k}_{j_0}})_{j_0} & & \\ s.t. & (p^\top \mathsf{M}_k)_j & \geq 0, & \forall j \in [m],\, k \leq \boldsymbol{k}_j \\ & p & \in \Delta([m])\,. & \end{cases}$$

Algorithm 3.3 implements a search over frontiers and determines whether the polynomial matrix game $\mathsf{M}(\cdot)$ is uniformly value-positive or not. Lemma 3.5.10 proves sufficient conditions for uniform value-positivity (the same conditions that are verified by Algorithm 3.3). Lemma 3.5.11 relates the search over frontiers with all possible witnesses of uniform value-positivity. Finally, Lemma 3.5.12 proves sufficient conditions to determine that $\mathsf{M}(\cdot)$ is not uniformly value-positive. Together, these lemmas prove the correctness of Algorithm 3.3.

Lemma 3.5.10. *Consider a polynomial matrix game $\mathsf{M}(\cdot)$ of size $m \times m$ and order K. Fix $\boldsymbol{k} \in [(K+1)]^m$. Assume that: (1) the LP $(P_{\boldsymbol{k},1})$ is feasible; and (2) for all $j_0 \in [m]$, if $\boldsymbol{k}_{j_0} < K+1$, then the value of $(P_{\boldsymbol{k},j_0})$ is strictly positive. Then, $\mathsf{M}(\cdot)$ is uniform value-positive.*

Proof. We construct the witness of uniform value-positivity as a convex combination of the solutions of the LPs $(P_{\boldsymbol{k},j_0})$. For all $j_0 \in [m]$, because $(P_{\boldsymbol{k},1})$ is feasible, $(P_{\boldsymbol{k},j_0})$ is feasible as only the objective function changes. Consider $p^*(j_0)$ a solution of $(P_{\boldsymbol{k},j_0})$ and define

$$p := \frac{1}{m} \sum_{j_0=1}^{m} p^*(j_0)\,.$$

We claim that p is a witness of uniform value-positivity. Indeed, on the one hand, if $\boldsymbol{k}_{j_0} < K+1$, then the leading coefficient of the polynomial $(p^\top M(\varepsilon))_{j_0}$ is strictly positive because it is lower bounded by the corresponding coefficient of $p^*(j_0)$ divided by m, which is strictly positive by assumption. On the other hand, if $\boldsymbol{k}_{j_0} = K+1$, then all coefficients of $(p^\top M(\varepsilon))_{j_0}$ are zero. Overall, for all $j_0 \in [m]$ and $\varepsilon \geq 0$ sufficiently small, $(p^\top M(\varepsilon))_{j_0}$ is positive, so M$(\cdot)$ is uniform value-positive.

□

Algorithm 3.3 maintains an invariant while increasing the coordinates of $\boldsymbol{k}$ formalized as follows.

Lemma 3.5.11. *Consider a polynomial matrix game $M(\cdot)$ of size $m \times m$ and order K that is uniform value-positive with a witness strategy p. Then, for all frontiers $\boldsymbol{k} \in [(K+1)]^m$ obtained during the execution of Algorithm 3.3, we have that, for all $j \in [m]$ and $0 \leq k < \boldsymbol{k}_j$, it holds that $(p^\top M_k)_j = 0$.*

Proof. The proof is by induction on the frontier $\boldsymbol{k} \in [K+1]^m$. For the initial case, we have $\boldsymbol{k} = (0, 0, \dots, 0)$ and the statement is trivially true because the condition is vacuous: for all $j \in [m]$, there is no k such that $0 \leq k$ and $k < \boldsymbol{k}_j = 0$. We continue to the inductive step.

Assume the statement is true for $\boldsymbol{k}$. We will prove that the statement holds for the next frontier $\boldsymbol{k}'$ obtained during the execution of Algorithm 3.3. The only way to obtain $\boldsymbol{k}'$ is that there exists $j_0 \in [m] \setminus \{j : \boldsymbol{k}_j = K+1\}$ such that the value of $(P_{\boldsymbol{k},j_0})$ is zero. In this case, the next frontier is $\boldsymbol{k}' := \boldsymbol{k} + e_{j_0}$, the indicator vector of the coordinate j_0. Consider p a witness of the uniform value-positivity of M$(\cdot)$, that is, for all $j \in [m]$ and all ε small enough, we have that $(p^\top M(\varepsilon))_j \geq 0$. Equivalently, for all $j \in [m]$, the leading coefficient of $(p^\top M(\varepsilon))_j$ is strictly greater than zero, or the polynomial is constantly zero. By the inductive hypothesis, we know that, for all $j \in [m]$ and all $0 \leq k < \boldsymbol{k}_j$ we have that $(p^\top M_k)_j = 0$. Therefore, we have that $(p^\top M_{\boldsymbol{k}_{j_0}})_{j_0} \geq 0$. We only need to prove that $(p^\top M_{\boldsymbol{k}_{j_0}})_{j_0} = 0$.

Note that p is a feasible vector of $(P_{\boldsymbol{k},j_0})$, whose value is zero. Therefore, $(p^\top M_{\boldsymbol{k}_{j_0}})_{j_0} = 0$ must hold. This concludes the inductive step.

□

Lastly, we study what happens when $(P_{\boldsymbol{k},1})$ is infeasible.

Lemma 3.5.12. *Consider a polynomial matrix game M$(\cdot)$ of size $m \times m$ and order K, and $\boldsymbol{k} \in [(K+1)]^m$ obtained during the execution of Algorithm 3.3. If $(P_{\boldsymbol{k},1})$ is infeasible, then M$(\cdot)$ is not uniform value-positive.*

Proof. Assume that $(P_{\boldsymbol{k},1})$ is infeasible. By contradiction, assume that M$(\cdot)$ is uniform value-positive and consider a witness strategy p. By Lemma 3.5.11, we have that, for all $j \in [m]$ and all $0 \leq k < \boldsymbol{k}_j$, it holds that $(p^\top M_k)_j = 0$. Because the strategy p is a witness of uniform value-positivity, for all $j \in [m]$, the polynomial $(p^\top M(\varepsilon))_j$ is positive in some right neighborhood of zero. Therefore, for all $j \in [m]$, it holds that $(p^\top M_{\boldsymbol{k}_j})_j \geq 0$. In conclusion, p is a feasible vector of $(P_{\boldsymbol{k},1})$, which is a contradiction.

□

We now prove Lemma 3.5.4.

Proof of Lemma 3.5.4. Note that Algorithm 3.3 solves at most $(K+1)m^2$ LPs, terminating in polynomial time. Moreover, by Lemma 3.5.12, we have that whenever Algorithm 3.3 returns *false*, the polynomial matrix game M$(\cdot)$ is not uniform value-positive.

Lastly, note that, if during the execution of Algorithm 3.3 there exists $j_0 \in [m]$ such that $\boldsymbol{k}_{j_0} = K+1$, then $(P_{\boldsymbol{k},j_0})$ is feasible because M_{K+1} is the matrix of only zeros. Therefore, whenever Algorithm 3.3 returns *true*, all conditions of Lemma 3.5.10 are satisfied, so $\mathsf{M}(\cdot)$ is uniform value-positive.

□

3.6 Value-Positivity and Perturbed LPs

We define *robustness problems* for perturbed LPs and connect them to value-positivity for polynomial matrix games.

3.6.1 Perturbed LPs and Robustness Problems

The notion of robust LPs has been considered in many different contexts and some examples are the following. First, LPs are closely related to Markov Decision Processes (MDPs), and robust MDPs consider that the transition function is not known but comes from an uncertain set, which leads to a notion of robust LPs [NEG05]. Second, analytical perturbation of LPs has been considered in [AFH13, Chapter 5]. Finally, considering perturbations of inputs in a neighborhood and efficient algorithmic solutions of LPs via Simplex leads to the notion of smoothed complexity analysis [ST01]. There is a vast literature on robust LPs and, in this work, we focus on LPs in connection to matrix games. Whereas Mills [Mil56] considered only linear perturbations, we consider general polynomial perturbations.

Definition 3.6.1 (Perturbed LPs)**.** A perturbed LP of degree K and size $m \times m$ has the following form:

$$(P(\varepsilon)) \begin{cases} \max_x & (c_0 + c_1\varepsilon + \ldots + c_K\varepsilon^K)^\top x \\ s.t. & (\mathsf{A}_0 + \mathsf{A}_1\varepsilon + \ldots + \mathsf{A}_K\varepsilon^K)x \quad \leq b_0 + b_1\varepsilon + \ldots + b_K\varepsilon^K \\ & x \quad \geq 0\,, \end{cases}$$

where A_i, b_i and c_i have integer entries and A_i is of size $m \times m$, for all $i \in [K]$.

Robustness Problems The stability of LPs concerns the robustness of the solutions and the value upon perturbations. Assuming that (P_0) is a feasible and bounded LP, *robustness problems* consist in determining how its feasibility, optimal solutions, and value vary for small positive ε.

Perturbed LPs can lead to irregularities, namely discontinuities may arise at $\varepsilon = 0$ even when perturbations are regular and continuous, see [AFH13, Chapter 5, page 111] for examples. Therefore, studying robustness on LPs usually requires structural assumptions on the perturbations, such as the rank of $\mathsf{A}(\varepsilon)$ being constant near 0. In our case, we focus on perturbations where both primal and dual admit uniformly bounded feasible regions. We formalize this assumption in Section 3.6.2. In the sequel, (Q) denotes the dual LP of (P), given by

$$(Q) \begin{cases} \min_y & y^\top b \\ s.t. & y^\top \mathsf{A} \quad \geq c^\top \\ & y \quad \geq 0\,. \end{cases}$$

For a perturbed LP $(P(\cdot))$, the corresponding dual perturbed LP is denoted $(Q(\cdot))$. We now define the relevant robustness problems for perturbed LPs.

Definition 3.6.2 (Weak Robustness)**.** A perturbed LP $(P(\cdot))$ is *weakly robust* if there exists $\varepsilon_0 > 0$ such that, for all $\varepsilon \in [0, \varepsilon_0]$, the LP $(P(\varepsilon))$ is feasible and has a bounded value. In other words, $\mathrm{val}(P(\varepsilon)) \in \mathbb{R}$ or, equivalently, both $(P(\varepsilon))$ and its dual $(Q(\varepsilon))$ are feasible.

Definition 3.6.3 (Strong Robustness)**.** A perturbed LP $(P(\cdot))$ is *strongly robust* if there is a constant solution for both the primal and dual. Formally, there exist two vectors $x^*, y^* \in \mathbb{R}^n$ and a threshold $\varepsilon_0 > 0$ such that x^* is a solution of $(P(\varepsilon))$ and y^* is a solution of $(Q(\varepsilon))$ for all $\varepsilon \in [0, \varepsilon_0]$.

Definition 3.6.4 (Functional Form)**.** Consider a perturbed LP $(P(\cdot))$ that is weakly robust. The *functional form* of this perturbed LP is given by the maps

$$\begin{aligned}\varepsilon &\mapsto \mathrm{val}(P(\varepsilon))\\ \varepsilon &\mapsto x^*(\varepsilon)\,,\end{aligned}$$

where $x^*(\varepsilon)$ is an optimal vector for $(P(\varepsilon))$, for all $\varepsilon \in (0, \varepsilon_0]$ and some threshold $\varepsilon_0 > 0$.

Strong robustness can be interpreted as a property of an optimal basis as follows. A perturbed LP is strongly robust if there is an optimal basis, that is, a subset of inequalities that uniquely determine a solution, of the unperturbed primal and dual LP that remains optimal upon small positive perturbations. On top of this robustness of the basis, the corresponding optimal solutions defined by the basis, for the primal and dual LPs, do not depend on the parameter ε of the perturbation.

Note that the marginal value of Mills [Mil56] depends only on the first order of the perturbations, while weak robustness, strong robustness, and the functional form problem depend on all higher terms.

3.6.2 Value-Positivity and Robustness Problems

Before describing the connection between robustness for perturbed LPs and value-positivity for polynomial matrix games, we formalize the assumption we consider over the perturbed LPs.

Definition 3.6.5 (Apriori Bound)**.** A perturbed LP $(P(\cdot))$ has an *apriori bound* if there exists $\beta > 0$ and $\varepsilon_0 > 0$ such that, for all $\varepsilon \in [0, \varepsilon_0]$, if x_ε and y_ε are feasible for $(P(\varepsilon))$ and $(Q(\varepsilon))$ respectively, then $x_\varepsilon^\top 1 + y_\varepsilon^\top 1 \le \beta$. An apriori bound is said to have polynomial binary size if both β and ε_0 have polynomial binary size.

Note that an apriori bound is not concerned with the feasibility of the perturbed LP. Indeed, a perturbed LP whose primal and dual LPs are both infeasible also has an apriori bound, by vacuity. Nonetheless, if a perturbed LP $(P(\cdot))$ has an apriori bound and $(P(0))$ is feasible, then the perturbed LP is weakly robust. Indeed, if $(P(0))$ is feasible and bounded, then so is $(Q(0))$. By [Mar75, Theorem 1.1], the value function $\mathrm{val}(P(\cdot))$ is continuous at 0. In particular, $(P(\varepsilon))$ is feasible for ε small enough, that is, it is weakly robust.

The existence of an apriori bound is derived from the unperturbed LP as follows.

Lemma 3.6.6 (Sufficient conditions for an apriori bound)**.** *Consider a perturbed LP $(P(\cdot))$. If the unperturbed LP $(P(0))$ and its dual $(Q(0))$ have nonempty and bounded feasible regions, then $(P(\cdot))$ has an apriori bound of polynomial binary size which can be computed in polynomial time.*

Lemma 3.6.6 is closely related to [Mar75, Lemma 3.1] which states that, if the unperturbed LP has a nonempty and bounded region, then, for each possible continuous perturbation (not necessarily polynomial), the corresponding perturbed LP has a uniformly bounded feasible region for all ε small enough. The main difference is that, because we focus on polynomial perturbations, we prove that the apriori bound has polynomial binary size and a polynomial-time procedure to compute it.

Proof of Lemma 3.6.6. Consider a perturbed LP $(P(\cdot))$ of size $m \times m$, degree K, and integer entries of size bounded by B such that $(P(0))$ and its dual $(Q(0))$ have nonempty and bounded feasible regions. We closely follow the proof of [Mar75, Lemma 3.1] while keeping track of explicit numerical bounds.

Consider the LP with the same feasible region as $(P(0))$ but different objective given by

$$\begin{cases} \max_x & 1^\top x \\ s.t. & \mathsf{A}(0)x \;\le b(0) \\ & \qquad x \;\ge 0\,. \end{cases}$$

Because the feasible region of $(P(0))$ is bounded, it has a finite value. By duality, its dual LP, that is,

$$\begin{cases} \min_y & y^\top b(0) \\ s.t. & y^\top \mathsf{A}(0) \;\ge 1^\top \\ & \qquad y \;\ge 0\,, \end{cases}$$

is feasible. In particular, it has a basic solution y_0 which has polynomial binary size and can be computed in polynomial time. Indeed, using Cramer's rule on a basic solution, we deduce that the coordinates of y_0 have size at most the determinant of a matrix involving entries in $\mathsf{A}(0)$ and 1. Therefore, because all coefficients of $\mathsf{A}(0)$ are bounded by B, the size of each coordinate of y_0 is bounded by $m!B^m \le (Bm)^m$, so their binary size is at most $m\,\mathrm{bit}(Bm)$.

Define $\beta_1 := y_0^\top b(0) + 1$ and $\varepsilon_1 := (2(Bm)^{m+1}K)^{-1}$. We claim that, for all $\varepsilon \in [0, \varepsilon_1]$, every feasible vector x_ε of $(P(\varepsilon))$ satisfies that $x_\varepsilon^\top 1 \le 2\beta_1$. Indeed, first note that, because ε_1 is sufficiently small, we have that, for all $\varepsilon \in [0, \varepsilon_1]$,

$$y_0^\top \mathsf{A}(\varepsilon) \ge \frac{1}{2} 1^\top \quad \text{and} \quad y_0^\top b(\varepsilon) \le \beta_1\,.$$

This is because

$$y_0^\top \mathsf{A}(\varepsilon) \ge y_0^\top \mathsf{A}(0) - y_0^\top 1 B K \varepsilon_1 \ge 1 - (Bm)^{m+1} K \varepsilon_1 \ge \frac{1}{2}$$

and also

$$y_0^\top b(\varepsilon) \le y_0^\top b(0) + y_0^\top 1 B K \varepsilon_1 \le y_0^\top b(0) + (Bm)^{m+1} K \varepsilon_1 \le y_0^\top b(0) + 1 = \beta_1\,.$$

Then, consider a feasible vector x_ε of $(P(\varepsilon))$. Because $\mathsf{A}(\varepsilon)x_\varepsilon \le b(\varepsilon)$ and $x_\varepsilon \ge 0$, we have that

$$x_\varepsilon^\top 1 = 1^\top x_\varepsilon \le 2 y_0^\top \mathsf{A}(\varepsilon) x_\varepsilon \le 2 y_0^\top b(\varepsilon) \le 2\beta_1\,.$$

Using a similar construction with the dual, we obtain another pair (β_2, ε_2). Then, the pair $\beta := 2\beta_1 + 2\beta_2$ and $\varepsilon_0 := \min(\varepsilon_1, \varepsilon_2)$ is an apriori bound for $(P(\cdot))$ of polynomial binary size. Because we only require to solve two LPs for its construction, we can compute it in polynomial time.

□

We now describe the connection between robustness for perturbed LPs and value-positivity for polynomial matrix games.

Theorem 3.6.7. *There is a polynomial-time reduction from weak (respectively strong) robustness and the functional form of perturbed LPs with an apriori bound to value-positivity (respectively uniform value-positivity) and the functional form of polynomial matrix games.*

On the one hand, the strong robustness of an LP implies the existence of a vector x^* that is optimal for all positive small parameters. On the other hand, uniform value-positivity of polynomial matrix games implies the existence of a strategy p_0 that ensures a positive value for all small parameters, but p_0 does not need to be optimal for every small parameter. Therefore, at least intuitively, it is not immediately clear that strong robustness of perturbed LPs reduces to uniform value-positivity of polynomial matrix games. Theorem 3.6.7 shows that this is indeed the case.

Note that robustness problems of perturbed LPs with an apriori bound can be solved by computing the value of unperturbed LPs. Indeed, by Theorem 3.6.7, robustness problems of perturbed LPs reduce to value-positivity problems of polynomial matrix games. In turn, value-positivity problems are solved by computing the value of various unperturbed matrix games. Because computing the value of matrix games consists of solving an unperturbed LP, we conclude that robustness problems of perturbed LPs can be solved by computing the value of unperturbed LPs.

Reduction The reduction is based on [VS24], a simplification of Adler's work [Adl13] to fill the gap left by Dantzig [Dan51]. The reduction of [VS24] presents a single matrix game with entries computed from the matrix and vectors of the LP being reduced. The matrix game is an extension of Dantzig's matrix game [Dan51]. We accommodate this reduction in the context of polynomial coefficients instead of fixed coefficients.

Main Idea A key insight the so-called *Karp-reduction* of [VS24], also present in Adler [Adl13], is the introduction of slack variables and bounds. For unperturbed LPs, these bounds are computable in strongly polynomial time when coefficients are algebraic, see [Adl13]. Our assumption of an apriori bound for perturbed LPs is crucial to extend the reduction to the perturbed case. A detailed construction is given in Section 3.6.3.

3.6.3 Detailed Proofs

In this section, we prove Theorem 3.6.7. We start by recalling how solving an LP can be reduced to solving a matrix game according to [VS24].

Previous Results

A convenient restatement of the reduction from LPs to matrix games given in [VS24, Theorem 7, page 18] is the following.

Lemma 3.6.8 (LP Reduction to Matrix Games)**.** *Consider a primal LP (P), its dual LP (Q), and a bound β such that, if x is feasible in (P) and y is feasible in (Q), then $x^\top \mathbb{1} + y^\top \mathbb{1} \leq \beta$. The matrix game*

$$M := \begin{pmatrix} 0 & A & -b & -1 \\ -A^\top & 0 & c & -1 \\ b^\top & -c^\top & 0 & \beta \end{pmatrix}$$

satisfies the following properties.

1. *The value of the game is less or equal to zero, that is,* $\mathrm{val}(M) \leq 0$*.*
2. *We have that* $\mathrm{val}(M) = 0$ *if and only if the LPs* (P) *and* (Q) *are feasible.*
3. *If* $\mathrm{val}(M) = 0$ *with optimal strategies* $p^* = (y^*, x^*, t^*)$ *and* $q^* = (y^*, x^*, t^*, 0)$*, then the LPs* (P) *and* (Q) *have optimal solutions* x^*/t^* *and* y^*/t^* *respectively.*

Remark 3.6.9. Some differences between Lemma 3.6.8 and [VS24, Theorem 7, page 18] are the following: (1) the payoff matrix M corresponds to the payoff matrix of the column-player in [VS24]; and (2) we omit the conclusions of the case $\mathrm{val}(\mathsf{M}) < 0$ because we will not use them.

Technical Proofs

We now turn to the proof of Theorem 3.6.7.

Proof of Theorem 3.6.7. We proceed in four steps: first, we introduce a useful polynomial matrix game; second, we relate weak robustness to value-positivity; third, we relate strong robustness to uniform value-positivity; and fourth, we relate the functional forms.

Step 1: Introduction of a Polynomial Matrix Game Fix a perturbed LP $(P(\cdot))$ of degree K given by

$$(P(\varepsilon))\begin{cases} \max_x & c(\varepsilon)^\top x \\ s.t. & \mathsf{A}(\varepsilon)x \leq b(\varepsilon) \\ & x \geq 0\,. \end{cases}$$

By assumption $(P(\cdot))$ has an apriori bound β which bounds all primal and dual feasible points in an interval $[0, \varepsilon_0]$. For all $\varepsilon \in [0, \varepsilon_0]$, consider the following polynomial matrix game:

$$\mathsf{M}(\varepsilon) := \begin{pmatrix} 0 & \mathsf{A}(\varepsilon) & -b(\varepsilon) & -1 \\ -\mathsf{A}^\top(\varepsilon) & 0 & c(\varepsilon) & -1 \\ b^\top(\varepsilon) & -c^\top(\varepsilon) & 0 & \beta \end{pmatrix}.$$

Because β is an apriori bound, Lemma 3.6.8 applies, so that, for all $\varepsilon \in [0, \varepsilon_0]$, the following properties hold.

1. The value of the matrix game $\mathsf{M}(\varepsilon)$ is less or equal to zero, that is, $\mathrm{val}(\mathsf{M}(\varepsilon)) \leq 0$.
2. Moreover, $\mathrm{val}(\mathsf{M}(\varepsilon)) = 0$ if and only if the LPs $(P(\varepsilon))$ and $(Q(\varepsilon))$ are feasible.
3. If $\mathrm{val}(\mathsf{M}(\varepsilon)) = 0$ with optimal strategies $p_\varepsilon^{*\top} = (y_\varepsilon^*, x_\varepsilon^*, t_\varepsilon^*)$ and $q_\varepsilon^* = (y_\varepsilon^*, x_\varepsilon^*, t_\varepsilon^*, 0)$, then the LPs $(P(\varepsilon))$ and $(Q(\varepsilon))$ have optimal solutions $x_\varepsilon^*/t_\varepsilon^*$ and $y_\varepsilon^*/t_\varepsilon^*$ respectively.

Note that $\mathrm{val}(\mathsf{M}(0)) = 0$ because the LP $(P(0))$ is feasible and bounded by assumption.

Step 2: Weak Robustness to Value-Positivity We claim that the perturbed LP $(P(\cdot))$ is weakly robust if and only if the polynomial matrix game $\mathsf{M}(\cdot)$ is value-positive. Indeed, if $\mathsf{M}(\cdot)$ is value-positive, then there exists $\varepsilon_0 > 0$ such that, for all $\varepsilon \in [0, \varepsilon_0]$, we have that $\operatorname{val} \mathsf{M}(\varepsilon) \geq 0$. Because $\operatorname{val} \mathsf{M}(\varepsilon) \leq 0$ by construction, for all $\varepsilon \in [0, \varepsilon_0]$, we have that $\operatorname{val} \mathsf{M}(\varepsilon) = 0$. Moreover, by Lemma 3.6.8, we conclude that $(P(\varepsilon))$ and $(Q(\varepsilon))$ are feasible, that is, $(P(\cdot))$ is weakly robust.

Conversely, if $\mathsf{M}(\cdot)$ is not value-positive, then, for all $\varepsilon_0 > 0$, there exists $\varepsilon \in [0, \varepsilon_0]$ such that $\operatorname{val} \mathsf{M}(\varepsilon) < 0$. By Lemma 3.6.8, this implies that one of $(P(\varepsilon))$ or $(Q(\varepsilon))$ is not feasible. In other words, $(P(\cdot))$ is not weakly robust.

Step 3: Strong Robustness to Uniform Value-Positivity We now claim that the perturbed LP $(P(\cdot))$ is strongly robust if and only if the polynomial matrix game $\mathsf{M}(\cdot)$ is uniform value-positive. Indeed, if $\mathsf{M}(\cdot)$ is uniform value-positive, then, there exists a fixed strategy $p^* = (x^*, y^*, t^*)$ and another threshold $\varepsilon_1 > 0$ such that ${p^*}^\top \mathsf{M}(\varepsilon) \geq 0$, for all $\varepsilon \in [0, \varepsilon_1]$. Because $\operatorname{val} \mathsf{M}(\varepsilon) \leq 0$ for all $\varepsilon \in [0, \varepsilon_0]$, then $\operatorname{val} \mathsf{M}(\varepsilon) = 0$ and the strategy p^* is optimal for $\mathsf{M}(\varepsilon)$ for all $\varepsilon \in [0, \min(\varepsilon_0, \varepsilon_1)]$. By Lemma 3.6.8, p^* encodes a fixed optimal solution x^*/t^* for $(P(\varepsilon))$ and also a fixed optimal solution y^*/t^* for $(Q(\varepsilon))$, for all $\varepsilon \in [0, \min(\varepsilon_0, \varepsilon_1)]$, so $(P(\cdot))$ is strongly robust.

Conversely, assume that $(P(\cdot))$ is strongly robust. We will show that $\mathsf{M}(\cdot)$ is uniform value-positive. Because $(P(\cdot))$ is strongly robust, consider the corresponding (fixed) optimal solutions x^* and y^* of $(P(\varepsilon))$ and $(Q(\varepsilon))$ respectively, for all $\varepsilon \in [0, \varepsilon_0]$. Then, define $t := (\sum_i x_i^* + \sum_j y_j^* + 1)^{-1}$, which is a strictly positive scalar, and construct a strategy for the row-player given by

$$p^* := (tx^*, ty^*, t) \,.$$

We claim that p^* is a witness of uniform value-positivity for $\mathsf{M}(\cdot)$. Indeed, fix $\varepsilon \in [0, \varepsilon_0]$. We will show that the strategy p^* guarantees a payoff greater or equal to zero in the matrix game $\mathsf{M}(\varepsilon)$.

1. Because x^* is feasible in $(P(\varepsilon))$, we have that $\mathsf{A}(\varepsilon)x^* \leq b(\varepsilon)$. Therefore, $-(tx^*)^\top \mathsf{A}^\top(\varepsilon) + tb^\top(\varepsilon) \geq 0$.

2. Because y^* is feasible in $(Q(\varepsilon))$, we have that $(y^*)^\top \mathsf{A}(\varepsilon) \geq c^\top(\varepsilon)$. Therefore, $(ty^*)^\top \mathsf{A}(\varepsilon) - tc^\top(\varepsilon) \geq 0$.

3. Because x^* and y^* are optimal, by strong duality $c^\top(\varepsilon)\, x^* = (y^*)^\top\, b(\varepsilon)$. In particular, $-(ty^*)^\top b(\varepsilon) + (tx^*)^\top c(\varepsilon) \geq 0$.

4. Because x^* and y^* are feasible, and the perturbed LP has an apriori bound of β, we have that $(x^*)^\top 1 + (y^*)^\top 1 \leq \beta$. Therefore, because $t > 0$, one has that $-(tx^*)^\top 1 - (ty^*)^\top 1 + t\beta \geq 0$.

5. Because x^* and y^* are feasible, and by the choice of t, we have that p^* is a probability measure.

In conclusion, p^* is a fixed strategy that guarantees a payoff greater or equal to zero in the matrix game $\mathsf{M}(\varepsilon)$. This concludes the proof that strong robustness of perturbed LPs reduces to uniform value-positivity of polynomial matrix games.

Step 4: Functional Form to Functional Form Assume $(P(\cdot))$ is weakly robust. Then, the functional form of the polynomial matrix game $\mathsf{M}(\cdot)$ encodes the functional form of $(P(\cdot))$. Indeed, because the perturbed LP is weakly robust, there exists $\varepsilon_0 > 0$ such that, for all $\varepsilon \in [0, \varepsilon_0]$, we have that $\operatorname{val}(\mathsf{M}(\varepsilon)) = 0$. Therefore, every pair of optimal strategies $p_\varepsilon^{*\top} = (y_\varepsilon^*, x_\varepsilon^*, t_\varepsilon^*)$ and $q_\varepsilon^* = (y_\varepsilon^*, x_\varepsilon^*, t_\varepsilon^*, 0)$ encode optimal solutions $x_\varepsilon^*/t_\varepsilon^*$ and $y_\varepsilon^*/t_\varepsilon^*$ for $(P(\varepsilon))$ and $(Q(\varepsilon))$ respectively. Lastly, the functional form of the value of $(P(\cdot))$ is given by

$$\operatorname{val}(P(\varepsilon)) = c(\varepsilon)^\top \left(\frac{x_\varepsilon^*}{t_\varepsilon^*} \right) .$$

Therefore, it is enough to solve the functional form problem for polynomial matrix games to solve the functional form problem for perturbed LPs.

□

3.7 Value-Positivity and Stochastic Games

We recall the connection between stochastic games [Sha53] and polynomial matrix games developed in [AOB19, AOB21, AOB24, OB20], and connect value-positivity with lower bounds on the discounted and undiscounted values of a stochastic game and uniform value-positivity with the existence of an optimal stationary strategy in undiscounted stochastic games.

Stochastic games [Sha53] extend matrix games to a dynamic interaction between two opponents. They are played by stages, and described by a finite set of states, finite sets of actions, a reward function, and a transition function. At each stage, knowing the current state, both players simultaneously choose an action; this choice determines a stage reward and the probability distribution for the next state. The notion of value in stochastic games depends on how the stage rewards are aggregated. In the discounted value, they are weighted decreasingly at a constant rate. The undiscounted value corresponds to the limit of the discounted values as the discount rate vanishes [BK76, MN81].

The characterization and computation of the discounted and undiscounted values of a stochastic game is a central problem in game theory [SV15]. In a series of works [AOB19, AOB21, AOB24, OB20] a new approach was proposed, where the key ingredient is a *family of polynomial matrix games*, denoted $\{M[z] : z \in \mathbb{R}\}$. For each $z \in \mathbb{R}$ and each $\varepsilon \in (0, 1]$, the rows (respectively columns) of the matrix $M[z](\varepsilon)$ correspond to a *pure stationary* strategies of the row-player (respectively column-player) in the stochastic game, that is, an action choice for each state. For a stochastic game with n states and m actions per state, these matrices are of size m^n. The main connection between the parameterized polynomial matrix games and stochastic games is the following.

Theorem 3.7.1 ([AOB19, Theorem 1])**.** *Consider a stochastic game and let* $M[z]$ *be the corresponding polynomial matrix game for a fixed* $z \in \mathbb{R}$*. Then, for all* $\varepsilon \geq 0$*, we have that* $\operatorname{val} M[z](\varepsilon) \geq 0$ *if and only if the* ε*-discounted value of the stochastic game is at least* z*.*

Applied to this family of polynomial matrix games, value-positivity and uniform value-positivity provide useful insights for stochastic games.

Lemma 3.7.2 (Stochastic games and value-positivity)**.** *Consider a stochastic game and let* $M[z]$ *be the corresponding polynomial matrix game for a fixed* $z \in \mathbb{R}$*. Then,* $M[z]$ *is value-positive if and only if the discounted value of the stochastic game is at least* z *for all sufficiently small discount rates.*

Proof. This follows directly from Theorem 3.7.1.

□

Lemma 3.7.3 (Stochastic Games and Uniform Value-Positivity). *Consider a stochastic game and let M$[z]$ be the corresponding polynomial matrix game for a fixed $z \in \mathbb{R}$. If there exists a fixed stationary strategy that guarantees z in all discounted stochastic games with sufficiently small discount rates, then M$[z]$ is uniform value-positive.*

Proof. Again, this follows from Theorem 3.7.1 and the definition of uniform value-positivity of polynomial matrix games.

□

Limitations The reverse implication of Lemma 3.7.3 is not known to be true. This is the case as the connection between M$[z]$ and the stochastic game is via their values and a restricted set of strategies [AOB21].

Blackwell Optimality It is worth noting that, by instantiating z with the undiscounted value v of the stochastic game, Lemma 3.7.3 connects the uniform value-positivity of $M[v]$ with a weak version of Blackwell optimality (which we recall is the existence of a fixed stationary strategy that is optimal in all ε-discounted stochastic games for ε sufficiently small). Indeed, if p is a stationary strategy that guarantees v in all ε-discounted stochastic games for ε small enough, then M$[v]$ is uniform value-positivity. Note that, in this case, p is an optimal stationary strategy in the undiscounted stochastic game as it guarantees v, but not necessarily Blackwell optimal as the ε-discounted might be larger than v.

We now illustrate Lemmas 3.7.2 and 3.7.3 with an example.

Example 3.7.4 (Big Match [BF68]). Consider the following stochastic game, popularized by [BF68]:

$$\begin{pmatrix} 1^* & 0^* \\ 0 & 1 \end{pmatrix} .$$

In this game, the $*$ indicates an absorbing payoff. In other words, when the top row is played, the corresponding stage payoff is fixed for all future stages, and when the bottom row is played, the stage payoff is recorded once and the game starts over again. Its value, both in the discounted and undiscounted cases, is $1/2$.

Value-Positivity Analysis By analyzing the value-positivity and uniform value-positivity of the corresponding polynomial matrix games, we recover the following properties of this game: (i) for all sufficiently small discount rates the discounted value is $1/2$; and (ii) while the row-player has a unique optimal stationary strategy in every ε-discounted game, there is no fixed optimal stationary strategy in the undiscounted game. To see this, fix $z = 1/2$. Then, up to a positive constant,

$$\mathsf{M}[z](\varepsilon) = \begin{pmatrix} 1 & -1 \\ -\varepsilon & \varepsilon \end{pmatrix} .$$

Note that, M$[z]$ is value-positive but not uniform value-positive. Indeed, for all $\varepsilon \in [0,1]$, we have $\operatorname{val} \mathsf{M}[z](\varepsilon) = 0$ and the unique optimal strategy of the row-player is $p^*(\varepsilon) = (\varepsilon/(1+\varepsilon), 1/(1+\varepsilon))^\top$, while, for each fixed stationary strategy $p \in \Delta[m]$ of the row-player (that is, not depending on ε), there exists a stationary strategy q of the column-player that leads to a strictly negative payoff as ε goes to 0.

By Lemma 3.7.2, the discounted value of the Big Match is at least z for all sufficiently small discount rates, and, by Lemma 3.7.3, there is no fixed stationary strategy for the row-player that guarantees z in all discounted games with a sufficiently small discount rate. Similarly, by transposing our results to the column-player, we obtain the existence of a fixed strategy of the column-player which guarantees at most z for all sufficiently small discount rates, implying that the discounted value is at most z in these games. We have thus shown the desired properties (i) and (ii).

Comments We finish this section with a few comments on the applicability of value-positivity and uniform value-positivity to stochastic games.

- *Size of the polynomial matrix games.* By construction, the polynomial matrix games are of size m^n, hence exponential and thus less tractable. A notable exception is absorbing games, a well-studied class of stochastic games (which includes the Big Match) where the state changes at most once during the game. In this class, the corresponding polynomial matrices are of size m.

- *Complexity of the undiscounted value*. Lemmas 3.7.2 and 3.7.3 are particularly relevant when replacing z by the undiscounted value of the stochastic game. However, the undiscounted value is an algebraic number whose degree can be as high as m^n [OB20, Proposition 1]. Because our algorithms require rational entries in the polynomial matrix game, they may not be directly applicable in this case. A notable exception is the class of stochastic games where at most one player controls the transition in each state. In this class, which includes turned-based stochastic games and Markov decision processes, the discounted and undiscounted values are rational expressions of the data.

- *Stationary ε-optimal strategies.* While discounted stochastic games admit optimal stationary strategies [Sha53], undiscounted games generally do not admit ε-optimal stationary strategies. Some notable exceptions are the following. First, ergodic games, where every pair of stationary strategies induces an ergodic Markov chain over the states [HK66], admit optimal stationary strategies. Second, recursive or terminal reward games [Eve57], where the outcome of a single turn is either a real number (the terminal reward) or another state of the game (but not both), and reachability or safety games [DAHK07], where the objective of one player is to reach a set of states and the opponent wants to prevent it, admit ε-optimal stationary strategies. Therefore, the notion of uniform value-positivity is particularly relevant in these classes of stochastic games.

3.8 Extensions

We discuss natural extensions of our results.

3.8.1 Nonpolynomial Perturbations

We explain the extent to which our algorithms extend to nonpolynomial perturbations of a matrix game.

Value-positivity Algorithm 3.1 requires a lower bound on some right neighborhood of zero where the sign of $\operatorname{val} \mathsf{M}(\cdot)$ remains constant. Lemma 3.5.6 provides a lower bound for polynomial perturbations. In contrast, general analytical functions do not allow such a lower bound. For example, consider the function $x \mapsto x \sin(1/x)$ which has infinitely many roots arbitrarily close to zero. Therefore, to extend our algorithm we need to restrict the family of perturbations allowed and ensure a strictly positive lower bound on root separation.

Functional form Algorithm 3.2 requires a bound on some right neighborhood of zero where the functional form of $\operatorname{val} \mathsf{M}(\cdot)$ and $p^*(\cdot)$ belongs to a particular family of functions. In the case of polynomial perturbations, it is the family of rational functions. Moreover, this family of functions must allow an efficient identification procedure, that is, recovering a particular function of this family with little information. In the case of rational functions, because the degree is bounded, we can identify a function by sufficiently many evaluations.

Uniform value-positivity Algorithm 3.3 requires a finite expansion of the perturbation to terminate. Moreover, the functions in the expansion need to be ordered in the following sense. Denote the general perturbed matrix game

$$\mathsf{M}(\varepsilon) = \sum_{k=0}^{K} \mathsf{M}_k f_k(\varepsilon) \,.$$

Then, it must hold that $\lim_{\varepsilon \to 0^+} f_k(\varepsilon)/f_{k-1}(\varepsilon) = 0$ for all $k \in [K]$. Under this condition, Algorithm 3.3 can be extended to more general perturbations with no changes.

3.8.2 Constrained Matrix Games

Players may be constrained in the mixed strategies they are allowed to play. A classic restriction over players' strategies is that, for each player, their mixed strategy must belong to a polytope strictly included in the set of all mixed strategies, that is, they must satisfy finitely many linear inequalities. We can incorporate these restrictions in the computation of the value of a matrix game. Indeed, there are two changes one must make to the LP that computes the value. First, the constraints over the row-player's strategies must be incorporated as new constraints on the feasible strategies p. Second, instead of having an inequality for each action of the column-player, we must have an inequality for each vertex of the polytope of feasible strategies for the column-player. To maintain the polynomial-time complexity of solving the LP, the polytope must have a polynomial number of vertices and its inequalities must be given by rational coefficients of polynomial binary size. Under these conditions, our results extend to polynomial matrix games where strategies satisfy linear constraints.

3.8.3 Perturbed Stochastic Games

Consider a *polynomial stochastic game*, that is, where the payoff and transition functions are polynomials in some parameter $\delta \in \mathbb{R}$. This model extends linearly perturbed stochastic games, studied in [AOBS25]. We are interested in small positive parameters and focus on the discounted case. For each discount rate $\varepsilon \in (0, 1]$ and each $\delta \geq 0$, denote the corresponding discounted value by $v(\varepsilon, \delta)$. As explained in Section 3.7, each stochastic game has a corresponding family of polynomial matrix games. Consider the family $\{M[z, \varepsilon] : (z, \varepsilon) \in \mathbb{R} \times (0, 1]\}$ of polynomial matrix games with parameter δ. The following statements are similar to Lemmas 3.7.2 and 3.7.3 but fixing the discount rate ε and reformulated in terms of the parameter δ.

Lemma 3.8.1. *Consider a polynomial stochastic game and let M$[z,\varepsilon]$ be the corresponding polynomial matrix game for some fixed parameter $z \in \mathbb{R}$ and discount rate $\varepsilon \in (0,1]$. Then, M$[z,\varepsilon]$ is value-positive if and only if $v(\varepsilon,\delta) \geq z$ for all sufficiently small δ.*

Lemma 3.8.2. *Consider a stochastic game and let M$[z,\varepsilon]$ be the corresponding polynomial matrix game for a fixed parameter $z \in \mathbb{R}$ and discount rate $\varepsilon \in (0,1]$. If there exists a fixed stationary strategy that guarantees z in all ε-discounted polynomial stochastic games with sufficiently small δ, then M$[z,\varepsilon]$ is uniform value-positive.*

3.9 Conclusion

This work contributes to the stability of matrix games, linear programs, and stochastic games. We introduced value-positivity and uniform value-positivity for matrix games, providing polynomial-time algorithms to check these properties. Further, we provided the functional form for a parameterized optimal strategy and the value function. Finally, we translated our results to linear programming and stochastic games, where value-positivity is related to the existence of robust solutions.

Future Directions Extending the concept of value-positivity to the undiscounted value of polynomial stochastic games is very interesting. In [BR14, NR93, RS02, RTV85, TR97], it is shown that some classes of stochastic games allow a characterization of the undiscounted value similar to Theorem 3.2.9 and Lemma 3.5.8. For example, when optimal pure stationary strategies exist, the value of the game corresponds to one of finitely many candidates. Such property would imply algorithms for the value-positivity and functional form problems, while the uniform value-positivity would require an efficient search in the strategy space.

CHAPTER 4

Marginal Values of a Stochastic Game

This chapter is based on [AOBS25], i.e., the following publication. Luc Attia, Miquel Oliu-Barton, and Raimundo Saona. Marginal Values of a Stochastic Game. *Mathematics of Operations Research*, 50(1):482–505, 2025.

Zero-sum stochastic games are parameterized by payoffs, transitions, and possibly a discount rate. We study how the main solution concepts, the discounted and undiscounted values, vary when these parameters are perturbed. We focus on the marginal values, introduced by Mills (1956) in the context of matrix games, that is, the directional derivatives of the value along any fixed perturbation. We provide a formula for the marginal values of a discounted stochastic game. Further, under mild assumptions on the perturbation, we provide a formula for their limit as the discount rate vanishes, and for the marginal values of an undiscounted stochastic game. Finally, we show via an example that the two latter differ in general.

4.1 Introduction

Stochastic Games Introduced by [Sha53], finite zero-sum stochastic games are a central model in dynamic games and model situations where the environment evolves in response to players' actions in a stationary manner. They are parameterized by a payoff function, a transition function, and possibly a discount rate which describe, respectively, the current payoff and transition probabilities for each state and each action profile, and the probability that the game stops after each stage. The number of states and possible actions at each state are assumed to be finite.

References The literature on stochastic games is abundant both in theory and applications. For this reason, we focus on the finite zero-sum case and present only the results that are most directly connected to our findings. We refer the reader to [SV15] for a summary of the historical context and the impact of Shapley's seminal work, and to [Ami03] for a review of applications, which include resource economics, industrial organization, market games, and empirical economics among others.

Value The main solution concept of finite zero-sum stochastic games, henceforth stochastic games, is the discounted value. Its existence and characterization go back to [Sha53]. The convergence of the discounted values as the discount rate vanishes was established by [BK76] using the theory of semi-algebraic sets. An alternative, probabilistic proof was obtained by [OB14]. The existence of the undiscounted value was established by [MN81].

Perturbations The perturbation analysis goes back to [FV97], who provided a modulus of continuity for the discounted value function in terms of the parameters of the game (i.e., the payoffs, the transition probabilities, and the discount rate). [Sol03] obtained an analogous result for the undiscounted value function under certain conditions on the perturbation of the transition probabilities. A formula for the discounted and undiscounted values was recently obtained by [AOB19], together with algorithms to compute them [OB20] which are polynomial in the number of pure stationary strategies. Robustness results for the undiscounted value function were provided, among others, by [NS10, Zil16a, COBZ21, OB18] and [OB22].

Marginal Value Of particular interest are the directional derivatives of the discounted and undiscounted value functions with respect to a given perturbation, referred to as the *marginal values* of the game. This concept was introduced by [Mil56] in the context of matrix games and linear programming, together with an elegant characterization: for all two matrices A and B of the same size, the directional derivative of the value of A in the direction B, i.e., $\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(\mathrm{val}(A+\varepsilon B) - \mathrm{val}(B))$, is equal to the value of the matrix game B where players are constrained to play optimal strategies of A. An extension of this result to compact-continuous games, i.e., games where the action sets are arbitrary compact metric sets and the payoff and transition functions are continuous, was proved by [RS01]. Notably, this result provides a formula for the marginal value of a stochastic game in the discounted case when only the payoffs are perturbed.

Our Contributions We investigate the marginal values of stochastic games. Our contributions are the following.

1. We obtain a formula for the marginal values of a discounted stochastic game, when whichever combination of its defining parameters are perturbed. Further, we show that this formula also holds in the compact-continuous framework, i.e., when action sets are compact and the payoff and transition functions are continuous.

2. We obtain a formula for the limit of the marginal value of a discounted stochastic game as the discount rate vanishes. Further, we show by example that the marginal value of an undiscounted stochastic game can exist and differ from the limit of the marginal value of the discounted stochastic game as the discount rate vanishes.

3. We provide a formula for the marginal value of an undiscounted stochastic game, under mild regularity assumptions.

These characterizations provide additional quantitative and qualitative insights into the model of stochastic games. They describe the dependence of the value function on the various parameters of the game, requiring no specific functional form for the payoff or the transition functions. Moreover, the formula for the marginal discounted value implies tractable algorithms for its approximation. Indeed, we show that the complexity of computing an approximation of the marginal value is at most polynomial-time in the number of pure stationary strategies of the game. This complexity coincides with the algorithms described in [OB20] for computing an approximation of the discounted value.

Outline Section 4.2 presents basic definitions and our notation. Section 4.3 presents our main contributions. Section 4.4 illustrates our results through two examples of perturbed stochastic games. Section 4.5 recalls some useful results from the literature, which will be used to prove our results. Section 4.6 presents the proofs of our main contributions. Section 4.7 presents the computational complexity of Theorem 4.3.1. Section 4.8 provides proofs that follow from previous results, for completeness. Section 4.9 considers the extension of Theorem 4.3.1 to the compact-continuous case, i.e., to discounted stochastic games over a finite state space but where the sets of actions are compact metric sets, and the payoff and the transition functions are continuous.

4.2 Preliminaries

We start by presenting the standard model of finite two-player zero-sum stochastic games, henceforth stochastic games, as introduced by [Sha53]. Then, we recall the definition and some properties of the auxiliary matrices introduced by [AOB19]. Lastly, we introduce perturbed stochastic games.

4.2.1 Classic Framework

A stochastic game is described by a tuple $\Gamma = (K, k, I, J; g, q, \lambda)$, where K is a finite set of states, $k \in K$ is the initial state, I and J are the finite action sets respectively of Player 1 and 2, $g \colon K \times I \times J \to \mathbb{R}$ is the payoff function, $q \colon K \times I \times J \to \Delta(K)$ is the transition function, and $\lambda \in [0, 1]$ is the discount rate. We refer to the case $\lambda \in (0, 1]$ as the *discounted case* and $\lambda = 0$ as the *undiscounted case*. Both cases are described below. In the sequel, K, k, I, and J will be fixed, while g, q, λ will be parameters. When useful, we will highlight the parameters of Γ, especially the discount rate λ, by using the notation Γ_λ.

Game The game proceeds in stages as follows. At each stage $m \geq 1$, both players are informed of the current state $k_m \in K$. Then, independently and simultaneously, Player 1 chooses an action $i_m \in I$ and Player 2 chooses an action $j_m \in J$. Both players may choose their actions using randomization. The pair (i_m, j_m) is then observed by both players, from which they can infer the stage payoff $g_m = g(k_m, i_m, j_m)$. A new state k_{m+1} is then chosen according to the probability distribution $q(\,\cdot \mid k_m, i_m, j_m)$, and the game proceeds to stage $m + 1$. A play thus produces a sequence of payoffs $(g_m)_{m \geq 1}$, and the aim of Player 1 is to maximize, in expectation,

- $\sum_{m \geq 1} \lambda (1 - \lambda)^{m-1} g_m$ in the λ-*discounted case*;
- $\liminf_{L \to \infty} \frac{1}{L} \sum_{m=1}^{L} g_m$ in the *undiscounted case*.

Note that, by the Tauberian theorem for real sequences ([HL14]), the undiscounted objective is equal to $\liminf_{\lambda \to 0} \sum_{m \geq 1} \lambda (1 - \lambda)^{m-1} g_m$. The game is zero-sum so Player 2 minimizes the same amount.

Strategies A *strategy* is a decision rule from the set of possible observations of a player to the set of probabilities over their action set. A strategy for Player 1 is thus a sequence of mappings $\sigma = (\sigma_m)_{m \geq 1}$, where $\sigma_m \colon (K \times I \times J)^{m-1} \times K \to \Delta(I)$. Similarly, a strategy for Player 2 is a sequence of mappings $\tau = (\tau_m)_{m \geq 1}$, where $\tau_m \colon (K \times I \times J)^{m-1} \times K \to \Delta(J)$.

The sets of strategies are denoted, respectively, by Σ and $\mathcal{T}$. A *stationary strategy* depends only on the current state, so $x\colon K \to \Delta(I)$ is a stationary strategy for Player 1 and $y\colon K \to \Delta(J)$ is a stationary strategy for Player 2. The sets of stationary strategies are $\Delta(I)^K$ and $\Delta(J)^K$, respectively. A *pure stationary strategy* is deterministic, so the sets of pure stationary strategies are I^K and J^K, respectively.

Value For all pairs of strategies $(\sigma,\tau) \in \Sigma \times \mathcal{T}$, the unique probability distribution over the sets of plays $(K \times I \times J)^{\mathbb{N}}$ induced by (σ,τ), the initial state k, and the transition function q is denoted by $\mathbb{P}_{\sigma,\tau}$. The existence and uniqueness of this probability follow from Kolmogorov's extension theorem on the sigma-algebra generated by cylinders. The expectation with respect to this probability is denoted by $\mathbb{E}_{\sigma,\tau}$. Recall that we distinguish discounted and undiscounted stochastic games, depending on whether $\lambda > 0$ or $\lambda = 0$.

- *Discounted stochastic games.* Let $\Gamma_\lambda = (K,k,I,J;g,q,\lambda)$ be a stochastic game with $\lambda > 0$. For all pairs of strategies (σ,τ) set
$$\gamma_\lambda(\sigma,\tau) := \mathbb{E}_{\sigma,\tau}\left[\sum\nolimits_{m\geq 1} \lambda(1-\lambda)^{m-1} g(k_m,i_m,j_m)\right] .$$
By [Sha53], the discounted stochastic game admits a value, $v(\Gamma_\lambda)$, that is:
$$v(\Gamma_\lambda) := \sup_{\sigma\in\Sigma}\inf_{\tau\in\mathcal{T}} \gamma_\lambda(\sigma,\tau) = \inf_{\tau\in\mathcal{T}}\sup_{\sigma\in\Sigma} \gamma_\lambda(\sigma,\tau) .$$
Furthermore, the vector of discounted values, as the initial state ranges over the set K, was shown to be the unique fixed point of a contracting map of $\mathbb{R}^K$. This characterization implies, in particular, that both players have optimal stationary strategies, denoted by $O_1(\Gamma)$ and $O_2(\Gamma)$.

- *Undiscounted stochastic games.* Let $\Gamma_0 = (K,k,I,J;g,q,\lambda=0)$ be an *undiscounted* stochastic game. For all pairs of strategies (σ,τ), set
$$\gamma_0(\sigma,\tau) := \liminf_{\lambda\to 0} \gamma_\lambda(\sigma,\tau) .$$
By [MN81] the undiscounted stochastic games has a value, $v(\Gamma_0)$, that is:
$$v(\Gamma_0) := \sup_{\sigma\in\Sigma}\inf_{\tau\in\mathcal{T}} \gamma_0(\sigma,\tau) = \inf_{\tau\in\mathcal{T}}\sup_{\sigma\in\Sigma} \gamma_0(\sigma,\tau) .$$
Moreover, the equality $\lim_{\lambda\to 0} v(\Gamma_\lambda) = v(\Gamma_0)$ holds.

4.2.2 Auxiliary Matrices

We now introduce two auxiliary matrices and a parameterized matrix game that can be associated to the discounted stochastic game $\Gamma_\lambda = (K,k,I,J;g,q,\lambda)$ with $\lambda > 0$. Their definition goes as follows.

If the players play a fixed pair of pure stationary strategies $(\mathbf{i},\mathbf{j}) \in I^K \times J^K$, then the state follows a Markov chain. Its transition matrix is denoted by $Q(\mathbf{i},\mathbf{j}) \in \mathbb{R}^{K\times K}$. Furthermore, the stage payoffs depend only on the current state, so let $g(\mathbf{i},\mathbf{j}) \in \mathbb{R}^K$ denote this fixed payoff vector. Let $\rho_\lambda(\mathbf{i},\mathbf{j}) \in \mathbb{R}^K$ be the vector of expected λ-discounted payoffs as the initial state ranges over the set of states. In particular, $\rho^k_\lambda(\mathbf{i},\mathbf{j}) = \gamma_\lambda(\mathbf{i},\mathbf{j})$ since k is the initial state. By stationarity, $Q(\mathbf{i},\mathbf{j})$, $g(\mathbf{i},\mathbf{j})$ and $r_\lambda(\mathbf{i},\mathbf{j})$ satisfy the following recursive relation:

$$\rho_\lambda(\mathbf{i},\mathbf{j}) = \gamma_\lambda(\mathbf{i},\mathbf{j}) = \lambda g(\mathbf{i},\mathbf{j}) + (1-\lambda)Q(\mathbf{i},\mathbf{j})\rho_\lambda(\mathbf{i},\mathbf{j})\,.$$

The matrix $\mathrm{Id}-(1-\lambda)Q(\mathbf{i},\mathbf{j})$ is invertible because $Q(\mathbf{i},\mathbf{j})$ is a stochastic matrix and $\lambda \in (0,1]$. Then, by Cramer's rule, for the initial state k one has:

$$\gamma_\lambda(\mathbf{i},\mathbf{j}) = \rho^k_\lambda(\mathbf{i},\mathbf{j}) = \frac{\Delta^k(\mathbf{i},\mathbf{j})}{\Delta^0(\mathbf{i},\mathbf{j})}\,,$$

where $\Delta^0(\mathbf{i},\mathbf{j}) = \det(\mathrm{Id}-(1-\lambda)Q(\mathbf{i},\mathbf{j}))$ and $\Delta^k(\mathbf{i},\mathbf{j})$ is the determinant of the matrix obtained by replacing the k-th column of $\mathrm{Id}-(1-\lambda)Q(\mathbf{i},\mathbf{j})$ with the vector $\lambda g(\mathbf{i},\mathbf{j})$. Ranging over $(\mathbf{i},\mathbf{j})$ one thus defines the two auxiliary matrices Δ^0 and Δ^k of size $|I^K| \times |J^K|$ whose entries are polynomials in (λ, q) and (λ, g, q) respectively. In addition, every entry of Δ^0 is larger or equal to $\lambda^{|K|}$. This is the case as for any stochastic matrix $M \in \mathbb{R}^{d\times d}$ and any $r \in (0,1]$, the matrix $S := \mathrm{Id}-(1-r)M$ satisfies $S_{\ell,\ell} \geq \sum_{\ell'\neq\ell}|S_{\ell,\ell'}| + r$ for every $1 \leq \ell \leq d$, which, by [Ost37, Ost52], implies $\det(S) \geq r^d$. Finally, for every $z \in \mathbb{R}$, we define the following matrix game:

$$W(z) := \Delta^k - z\,\Delta^0\,.$$

This parameterized matrix game characterizes the value according to [AOB19, Theorem 1], i.e., $v(\Gamma_\lambda)$ is the unique $z \in \mathbb{R}$ satisfying $\mathrm{val}(W(z)) = 0$. It is worth noting that this characterization holds for a fixed initial state and involves matrices of size $|I|^K \times |J|^K$. In contrast, [Sha53]'s approach characterizes the vector of values (that is, for all possible initial positions) via a $|K|$-dimensional system of equations involving matrices of size $I \times J$. Lastly, it is worth noting that the use of these auxiliary matrices was refined in [AOB21], and generalized to the non-zero sum case in [AOB24].

4.2.3 Perturbed Games and Marginal Value

Consider a stochastic game $\Gamma = (K, k, I, J; g, q, \lambda)$ with $\lambda > 0$ (discounted) or $\lambda = 0$ (undiscounted). An *admissible perturbation* is a triplet $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$ where

- $\tilde{g}\colon K \times I \times J \to \mathbb{R}$;
- $\tilde{q}\colon K \times I \times J \to \mathbb{R}$ and, for all $\varepsilon \geq 0$ sufficiently small, $q + \varepsilon\tilde{q}$ is a transition function;
- for all $\varepsilon \geq 0$ sufficiently small, $\lambda + \varepsilon\tilde{\lambda} \geq 0$.

The *perturbed stochastic game* Γ *in the direction* H, denoted by $\Gamma + \varepsilon H$, is defined as

$$\Gamma + \varepsilon H := \left(K, k, I, J; g + \varepsilon\tilde{g}, q + \varepsilon\tilde{q}, \lambda + \varepsilon\tilde{\lambda}\right)\,.$$

Its value is denoted by $v(\Gamma + \varepsilon H)$. Note that the perturbed game $\Gamma + \varepsilon H$ is a discounted stochastic game if $\lambda + \tilde{\lambda} > 0$. In this case, the auxiliary matrices can be defined as in Section 4.2.2, and are denoted by Δ^0_ε, Δ^k_ε, and $W_\varepsilon(z)$, respectively.

The *marginal value of the stochastic game* Γ *in the perturbation direction* H is then defined as follows, provided that the limit exists:

$$\partial_H v(\Gamma) := \lim_{\varepsilon\to 0^+} \frac{v(\Gamma + \varepsilon H) - v(\Gamma)}{\varepsilon}\,.$$

This notion was introduced by [Mil56] in the context of matrix games and extended by [RS01] to more general two-player games, i.e., with arbitrary compact action sets, and a continuous payoff function.

The marginal value of stochastic games has not been characterized yet. However, their existence can be deduced using the theory of semi-algebraic sets, and bounds can be derived from [FV97, Chapter 4] for the discounted case, and from [Sol03, Theorem 6] for the undiscounted case under mild assumptions. The mild assumptions in [Sol03, Theorem 6] consist of $\tilde{\lambda} = 0$ and that $\tilde{q}$ *does not introduce new transitions*, i.e., there is no $(\ell, \ell', i, j) \in K^2 \times I \times J$ such that $\ell \neq \ell'$ and $q(\ell' \,|\, \ell, i, j) = 0$ and $\tilde{q}(\ell' \,|\, \ell, i, j) > 0$. The regularity of the perturbed value function as well as the existence of the marginal values under mild conditions is recalled in Section 4.5.3 for completeness.

4.2.4 Notation

The following notation is used in the sequel.

Simplices For any finite set E, the set of probabilities over E is denoted by $\Delta(E)$. Formally,

$$\Delta(E) \coloneqq \left\{ x : E \to [0,1], \; \sum\nolimits_{e \in E} x(e) = 1 \right\} .$$

Matrix Games For any matrix $M \in \mathbb{R}^{I \times J}$ and any pair of mixed strategies $(x, y) \in \Delta(I) \times \Delta(J)$, the expected payoff in the matrix game M when players play (x, y) is given by $x^\top M y$. The value of M exists by the minmax theorem and is denoted by

$$\operatorname{val}(M) \coloneqq \max_{x \in \Delta(I)} \min_{y \in \Delta(J)} x^\top M y \, .$$

The sets of optimal strategies are denoted by $O_1(M) \subset \Delta(I)$ and $O_2(M) \subset \Delta(J)$ respectively, which are then non-empty, convex, compact sets. Finally, we set $O^*(M) \coloneqq O_1(M) \times O_2(M)$.

Marginal Value For any pair of matrices $M, N \in \mathbb{R}^{I \times J}$, we consider the matrix game N in which the players are restricted to play an optimal strategy of M. Its value exists by [Sio58], and is denoted by $\operatorname{val}_{O^*(M)} N$. [Mil56] proved that

$$\operatorname{val}_{O^*(M)} N = \max_{x \in O_1(M)} \min_{y \in O_2(M)} x^\top N y \, .$$

Canonical Inclusion For any two finite sets E and F, the product set $\Delta(E)^F$ can be injected into the simplex $\Delta(E^F)$ via the product measure map as follows:

$$w = (w^1, \dots, w^{|F|}) \in \Delta(E)^F \mapsto \widehat{w} = w^1 \otimes \cdots \otimes w^{|F|} \in \Delta(E^F) \, ,$$

where $\otimes$ denotes the direct product. That is, $\widehat{w}(e_1, \dots, e_{|F|}) = \prod_{\ell=1}^{|F|} w^\ell(e_\ell)$, for all $(e_1, \dots, e_{|F|}) \in E^F$. The canonical inclusion is strict as soon as $|E| \geq 2$ and $|F| \geq 2$.

Strategies in Stochastic Game For a stochastic game $\Gamma = (K, k, I, J, g, q, \lambda)$ and a stationary strategy $x = (x^\ell)_{\ell \in K} \in \Delta(I)^K$, the canonical inclusion gives $\widehat{x} \coloneqq \otimes_{\ell \in K} x^\ell \in \Delta(I^K)$, which is a probability distribution over their set of pure stationary strategies. The same is true for Player 2.

Important Set of Strategies Consider a discounted stochastic game Γ. Its auxiliary matrix game is of size $|I^K| \times |J^K|$ and plays an important role in the sequel. We set:

$$\begin{cases} O_1^*(\Gamma) := O_1(W(v(\Gamma))), \\ O_2^*(\Gamma) := O_2(W(v(\Gamma))), \\ O^*(\Gamma) := O_1^*(\Gamma) \times O_2^*(\Gamma) . \end{cases}$$

These sets are non-empty, convex, and compact. Moreover, $O_1^*(\Gamma) \subset \Delta(I^K)$ and $O_2^*(\Gamma) \subset \Delta(J^K)$. Hence, the elements of $O^*(\Gamma)$ are probabilities over the sets of pure stationary strategies, rather than (mixed) stationary strategies, of the game Γ. However, by [AOB19], $(\widehat{x}, \widehat{y}) \in O^*(\Gamma)$ for all pairs of optimal stationary strategies $(x, y) \in \Delta(I)^K \times \Delta(J)^K$. In this sense, $O(\Gamma) \subset O^*(\Gamma)$.

Perturbed Value If the stochastic game Γ and admissible perturbation H are clear from the context, we use the notation $v_\varepsilon := v(\Gamma + \varepsilon H)$ in general, and $v_{\lambda,\varepsilon}$ if the discount rate λ is fixed.

4.3 Main Contributions

In the sequel, we consider a stochastic game $\Gamma = (K, k, I, J; g, q, \lambda)$ for some fixed $\lambda \in [0, 1]$ and an admissible perturbation $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$. Our main contributions are the characterizations of the marginal discounted values, of their limits as the discount rate vanishes, and of the undiscounted marginal values.

4.3.1 Formula for the Discounted Marginal Value

Our first result is a characterization of the marginal values of a discounted stochastic game (when $\lambda > 0$). By Section 4.2.2, to every discounted stochastic game one can associate auxiliary matrices Δ^0 and Δ^k, so in particular this is the case for the perturbed game $\Gamma + \varepsilon H$. We set

$$\begin{cases} \partial_H \Delta^0 := \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}(\Delta^0_\varepsilon - \Delta^0) , \\ \partial_H \Delta^k := \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}(\Delta^k_\varepsilon - \Delta^k) , \end{cases}$$

where the limits exist since, by construction, all entries of Δ^k_ε and Δ^0_ε are polynomial in ε. We can now state our first result.

Theorem 4.3.1. *Consider a stochastic game $\Gamma = (K, k, I, J; g, q, \lambda)$ with $\lambda > 0$ and an admissible perturbation $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$. Then, $\partial_H v(\Gamma)$ is the unique $z \in \mathbb{R}$ satisfying*

$$D(z) := \operatorname{val}_{O^*(\Gamma)} \left(\partial_H \Delta^k - v(\Gamma)\, \partial_H \Delta^0 - z\, \Delta^0 \right) = 0 .$$

We now discuss the intuition behind this result, its implications regarding the computation complexity of the discounted marginal value, the particular case where only the payoffs are perturbed, and its extension to infinite action sets.

1. *Heuristics.* The formula of Theorem 4.3.1 can be interpreted as a derivation under the value operator of the formula $\operatorname{val}(\Delta^k - v(\Gamma)\Delta^0) = 0$ obtained in [AOB19]. Indeed,

it follows that $\mathrm{val}(\Delta^k_\varepsilon - v_\varepsilon \Delta^0_\varepsilon) = 0$ for all ε sufficiently small. If allowed, entry-wise differentiation inside the value operator would yield

$$\begin{aligned} 0 = \frac{\partial}{\partial \varepsilon}\left(\mathrm{val}(\Delta^k_\varepsilon - v_\varepsilon \Delta^0_\varepsilon)\right)_{|\varepsilon=0} &= \mathrm{val}\left(\frac{\partial}{\partial \varepsilon}\left(\Delta^k_\varepsilon - v_\varepsilon \Delta^0\right)_{|\varepsilon=0}\right), \\ &= \mathrm{val}\left(\partial_H \Delta^k - v(\Gamma)\, \partial_H \Delta^0 - \partial_H v(\Gamma)\, \Delta^0\right). \end{aligned}$$

It thus makes sense to have $z = \partial_H v(\Gamma)$ as the unique solution of $\mathrm{val}(\partial_H \Delta^k - v(\Gamma)\partial_H \Delta^0 - z\Delta^0) = 0$. This heuristic can be turned into a formal statement by restricting the strategy domain in the right-hand side expressions to $O^*(\Gamma)$.

2. *Computational complexity.* From Theorem 4.3.1 one can derive an algorithm to approximate the marginal discounted values, whose computational complexity is polynomial $|I^K|$ and $|J^K|$ for rational data. To do so, we proceed in two steps. First, compute $v(\Gamma)$ using the algorithm from [OB20]. Second, use the map $z \mapsto D(z)$ to do a dichotomic search and find an approximation of the marginal value. At every step, $D(z)$ is the value of a linear program proposed by [Mil56] and can be computed efficiently by [Bel01]. See Section 4.7 for more details.

3. *Perturbation of the payoff only.* When only the payoffs are perturbed, i.e., for $H = (\tilde{g}, 0, 0)$, Theorem 4.3.1 boils down to the following alternative formula, already obtained in [RS01]:
$$\partial_H v(\Gamma) = \mathrm{val}_{O(\Gamma)}\, \tilde{\gamma}(x, y)\,,$$
where $O(\Gamma) := O_1(\Gamma) \times O_2(\Gamma)$ is the set of optimal stationary strategies of Γ and $\tilde{\gamma}_\lambda$ is the discounted payoff function of the game $\tilde{\Gamma} = (K, k, I, J; \tilde{g}, q, \lambda)$. Indeed, the condition for $\partial_H v(\Gamma)$ simplifies to $\mathrm{val}_{O^*(\Gamma)}(\tilde{\Delta}^k - \partial_H v(\Gamma)\tilde{\Delta}^0) = 0$. Therefore, considering optimal strategies for each player separately one can reverse the linearization used in constructing the auxiliary matrices, and thus recover $\tilde{\gamma}(x, y)$.

4. *Extension to the compact-continuous case.* Theorem 4.3.1 can be extended to the compact-continuous case, i.e., when action sets are compact and the payoff and transition functions are continuous. This extension is proved in Section 4.9.

4.3.2 Formula for the Limit of the Discounted Marginal Value

Let $\Gamma_\lambda = (K, k, I, J; g, q, \lambda)$ be a discounted stochastic game, where the sub-index $\lambda > 0$ is used to better track the variations in λ, while the rest of the parameters remain fixed. Similarly, we denote the auxiliary matrix games by Δ^0_λ and Δ^k_λ.

Theorem 4.3.2. *Let Γ_λ be a λ-discounted stochastic game and $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$ an admissible perturbation. Assume that $(\partial_H v(\Gamma_\lambda))_\lambda$ remains bounded. Then, $\lim_{\lambda \to 0} \partial_H v(\Gamma_\lambda)$ exists and is the unique $w \in \mathbb{R}$ where the following map (over the extended reals) changes sign:*

$$z \in \mathbb{R} \;\mapsto\; F(z) := \lim_{\lambda \to 0} \lambda^{-|K|}\, \mathrm{val}_{O^*(\Gamma_\lambda)}\left(\partial_H \Delta^k_\lambda - v_\lambda\, \partial_H \Delta^0_\lambda - z\, \Delta^0_\lambda\right) \in [-\infty, +\infty]\,.$$

This result calls for several observations too.

1. *The boundedness assumption.* Sufficient conditions ensuring that $(\partial_H v(\Gamma_\lambda))_\lambda$ remains bounded follow from [Sol03, Theorem 6]. For example, it is enough that $\tilde{q}$ *does not introduce new transitions* in the following sense: for all $(i, j) \in I \times J$ and $\ell \neq \ell'$, if $q(\ell' \,|\, \ell, i, j) = 0$, then $\tilde{q}(\ell' \,|\, \ell, i, j) = 0$. See Section 4.5.3 for details.

2. *Applicability.* Similarly to Theorem 4.3.1, the present formula suggests a dichotomic algorithm to compute an approximation of the limit of the discounted marginal values, based on the successive computation of the sign of $F(z)$. To compute the latter, we would need to determine an explicit $\lambda_0 > 0$ such that the sign of $F(z)$ is given by the sign of $\mathrm{val}_{O^*(\Gamma_\lambda)}(\partial_H \Delta_\lambda^k - v_\lambda\, \partial_H \Delta_\lambda^0 - z\, \Delta_\lambda^0)$ at $\lambda = \lambda_0$, for all the considered z. This approach was followed in [OB20] where the similar map $z \mapsto \lim_{\lambda \to 0} \lambda^{|K|}\, \mathrm{val}(\Delta_\lambda^k - z\Delta_\lambda^0)$ was considered to compute the undiscounted value of a stochastic game, and where an explicit lower bound for λ_0 was obtained whose size is polynomial in $|I^K|$ and $|J^K|$. Note, however, that additional difficulties may arise in determining an explicit lower bound for λ_0 in the present case, notably because of the presence of the term v_λ, which is a Puiseux series near 0 unlike all other terms which are polynomials in λ.

3. *Extension.* This result does not extend to the compact-continuous case, as explained in Section 4.9.

At this stage, a natural question is whether the operators $\lim_{\lambda \to 0}$ and ∂_H commute, so that the limit of the marginal discounted values is equal to the marginal undiscounted value. The answer is no, except in very particular cases. More precisely, the following assertions hold (see Section 4.6.4 for more details).

- The operators ∂_H and $\lim_{\lambda \to 0}$ commute when either $|K| = 1$, or $|I| = |J| = 1$ and $H = (\tilde{g}, 0, 0)$.
- There exists a (minimal) example with $|K| = |I| = 2$, $|J| = 1$ and $H = (\tilde{g}, 0, 0)$ where
$$\lim_{\lambda \to 0} \partial_H v(\Gamma_\lambda) \neq \partial_H \lim_{\lambda \to 0} v(\Gamma_\lambda) = \partial_H v(\Gamma_0)\,.$$

A minimal example is given by the perturbed stochastic game represented in Figure 4.1. Indeed, for $\lambda \in [0, 1]$ and $\varepsilon \geq 0$, a direct computation yields:

$$v_{\lambda,\varepsilon} = \begin{cases} (1-\lambda)(1+\varepsilon) & \text{if } \varepsilon \geq \frac{\lambda}{1-\lambda}, \\ 1 & \text{otherwise.} \end{cases}$$

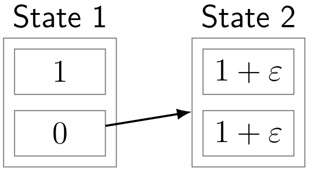


Figure 4.1: A perturbed stochastic game with $\lim_{\lambda \to 0} \partial_H v(\Gamma_\lambda) \neq \partial_H v(\Gamma_0)$. The arrow indicates a deterministic transition from state 1 to state 2 when the bottom action is played, and where there are no other positive transition probabilities.

Hence, for all $\lambda > 0$, there exists $\varepsilon_0 > 0$ small enough such that $v_{\lambda,\varepsilon} = 1$ for all $\varepsilon \in [0, \varepsilon_0]$. Consequently,
$$\lim_{\lambda \to 0} \partial_H v(\Gamma_\lambda) = \lim_{\varepsilon \to 0} \frac{v_{\lambda,\varepsilon} - v_{\lambda,0}}{\varepsilon} = 0\,.$$
On the other hand, for all $\varepsilon \geq 0$, $v_{0,\varepsilon} = \lim_{\lambda \to 0} v_{\lambda,\varepsilon} = 1 + \varepsilon$, so

$$\partial_H v(\Gamma_0) = \lim_{\varepsilon \to 0} \frac{v_{0,\varepsilon} - v_{0,0}}{\varepsilon} = 1\,.$$

Consequently, both $\lim_{\lambda\to 0}\partial_H v(\Gamma_\lambda)$ and $\partial_H v(\Gamma_0)$ exist, but differ.

In view of this (minimal) example, it is worth noting that Theorem 4.3.2 characterizes the limit of the marginal discounted values, but not the marginal undiscounted values.

4.3.3 Formula for the Marginal Undiscounted Value

We explore the undiscounted case, $\lambda = 0$. We assume that $H = (\tilde{g}, \tilde{q}, 0))$, since for $\tilde{\lambda} > 0$ the marginal undiscounted value may fail to exist. Indeed, [Koh74, page 120] exhibited an example of a discounted stochastic game where $v(\Gamma_\lambda) = \frac{1-\sqrt{\lambda}}{1-\lambda} = 1 - \sqrt{\lambda} + o(\sqrt{\lambda})$, so that for $H = (0,0,1)$ the marginal undiscounted value $\partial_H v(\Gamma_0) = \lim_{\lambda\to 0}\frac{1}{\lambda}(v_\lambda - v_0)$ exists but is unbounded.

We introduce a family of perturbed discounted stochastic games and define a finite set of bivariate polynomials denoted $\mathcal{P}_{(\Gamma,H)}$. For every $\beta > 0$ and $\varepsilon \geq 0$ sufficiently small, let $\Gamma_{\beta,\varepsilon} := (K, k, I, J; g + \varepsilon\tilde{g}, q + \varepsilon\tilde{q}, \beta)$ and let $W_{\beta,\varepsilon}(z)$ denote its corresponding parameterized matrix game. For every square sub-matrix of $W_{\beta,\varepsilon}(z)$, consider the polynomial $P(\beta,\varepsilon,z)$ obtained by taking the determinant of this sub-matrix. There exist $(R_m)_m \in \mathbb{R}[\varepsilon, z]$ such that $P(\beta,\varepsilon,z) = \sum_{m\geq 0} R_m(\varepsilon,z)\beta^m$, and let $s \geq 0$ be the smallest integer such that $R_s \neq 0$. Define then the projection $\Phi(P) := R_s \in \mathbb{R}[\varepsilon, z]$. Running over all possible square sub-matrices of $W_{\beta,\varepsilon}(z)$, one thus defines the desired finite set of polynomials:

$$\mathcal{P}_{(\Gamma,H)} := \left\{\Phi(P) : P(\beta,\varepsilon,z) = \det(\overline{W}_{\beta,\varepsilon}(z)),\ \overline{W}_{\beta,\varepsilon}(z) \text{ square sub-matrix of } W_{\beta,\varepsilon}(z)\right\}.$$

Our next results will justify the introduction of this set.

Proposition 4.3.3. *Let $\Gamma = (K, k, I, J; g, q, 0)$ be an undiscounted stochastic game and let $H = (\tilde{g}, \tilde{q}, 0)$ be an admissible perturbation (not perturbing the discount rate). Then, there exists $\varepsilon_0 > 0$ and $P \in \mathcal{P}_{(\Gamma,H)}$ such that*

$$P \not\equiv 0, \quad \textit{and} \quad P(\varepsilon, v(\Gamma + \varepsilon H)) = 0, \quad \forall \varepsilon \in (0, \varepsilon_0].$$

This result may call for some clarifications, notably as the existence of *some* polynomial P satisfying $P(\varepsilon, v_\varepsilon) = 0$ for all ε near 0 follows from the theory of semi-algebraic sets (see Section 4.5.2 for more details).

1. *Novelty of the result.* The novelty of Proposition 4.3.3 is the identification of a finite, computable set of candidates for the polynomial P. This step is crucial to obtain an explicit formula for the undiscounted marginal values in Theorem 4.3.4 below.

2. *Computability of P.* While in general a suitable polynomial P may be hard to find, information about (the support of) optimal stationary strategies in the auxiliary discounted stochastic games $\Gamma_{\beta,\varepsilon}$ greatly simplifies the search. For example, if both players have a unique optimal stationary strategy whose support is constant for all (ε, λ) sufficiently small, and of equal size, then it is enough to define the sub-matrix $\overline{W}_{\beta,\varepsilon}(z)$ over these supports, and set $P(\varepsilon, z) = \Phi(\det(\overline{W}_{\beta,\varepsilon}(z))$. This result, proved in Lemma 6 of [OBV23], gives a practical way to determine P in (generic) applications.

Our last result is an explicit formula for the undiscounted marginal values, relying on the set of polynomials identified in Proposition 4.3.3.

Theorem 4.3.4. *Let $\Gamma = (K, k, I, J; g, q, 0)$ be an undiscounted stochastic game and $H = (\tilde{g}, \tilde{q}, 0)$. Let $P \in \mathbb{R}[\varepsilon, z]$ be a polynomial satisfying $P(\varepsilon, v_\varepsilon) = 0$ for all $\varepsilon > 0$ sufficiently small (which can be determined using Proposition 4.3.3). Suppose that $\frac{\partial P}{\partial z}(0, v(\Gamma)) \neq 0$ and that $\varepsilon \mapsto v(\Gamma + \varepsilon H)$ is continuous at $\varepsilon = 0$. Then, the marginal undiscounted value exists and is given by*

$$\partial_H v(\Gamma) = -\frac{\frac{\partial P}{\partial \varepsilon}(0, v(\Gamma))}{\frac{\partial P}{\partial z}(0, v(\Gamma))}\,.$$

Note that Theorem 4.3.4 requires the continuity of $\varepsilon \mapsto v(\Gamma + \varepsilon H)$ at 0. Like for Theorem 4.3.2, a simple sufficient condition ensuring this property is that $\tilde{q}$ does not introduce new transitions.

4.4 Examples

We now illustrate our contributions via two known examples. The first one is a perturbed Big Match, the second is a perturbed version of a game introduced by [BK78]. For both examples we will compute the discounted marginal values, their limit as the discount rate goes to 0, and the undiscounted marginal values, using Theorem 4.3.1, Theorem 4.3.2 and Theorem 4.3.4.

4.4.1 The Big Match

We start with the classical Big Match introduced by [Gil58]. This is a three-state stochastic game, where two states are absorbing with payoffs of 0 and 1 respectively. The Big Match is thus an absorbing game. It can be represented by Figure 4.2, where a^* indicates a stage payoff of a followed by a deterministic transition to an absorbing state with payoff a. For all $\lambda \in (0, 1]$, we denote by Γ_λ the λ-discounted Big Match with initial state $k = 1$. Its value satisfies $v_\lambda = 1/2$ for all λ.

State 1

1^*	0^*
0	1

Figure 4.2: Big Match stochastic game.

Auxiliary Matrices Because the game is absorbing, the set of pure stationary strategies can be identified with the set $I \times J$. Consequently, up to removing redundant rows and columns, the auxiliary matrices are of size $I \times J$. Let states 2 and 3 be, respectively, absorbing states with payoff 1 and 0. We follow Section 4.2.2 to compute the auxiliary matrices, denoted by Δ_λ^0, Δ_λ^k, and the parameterized matrix game $W_\lambda(z) := \Delta_\lambda^k - z\Delta_\lambda^0$. Set $I = \{T, B\}$ and $J = \{L, R\}$, where T, B, L, and R refer to 'top', 'bottom', 'left', and 'right', respectively. When the pure stationary strategy (T, L) is played, every time states 1, 2, and 3 are visited, the stage payoffs are $g(1, T, L) = 1$, $g(2, T, L) = 1$, and $g(3, T, L) = 0$, respectively. On the other hand, the transitions are δ_2 in states 1 and 2, and δ_3 in state 3. Hence, in matrix form, one has:

$$g(T, L) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad Q(T, L) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Recall that $\Delta^0_\lambda(T,L) = \det(\mathrm{Id} - (1-\lambda)Q(T,L))$ and $\Delta^k_\lambda(I,J)$ is the determinant of the matrix obtained by replacing the k-th column of $\mathrm{Id} - (1-\lambda)Q(T,L)$ with $\lambda g(T,L)$. Hence, for all $\lambda \in (0,1]$,

$$\begin{cases} \Delta^0_\lambda(T,L) = \det \begin{pmatrix} 1 & -(1-\lambda) & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} = \lambda^2, \\ \Delta^k_\lambda(T,L) = \det \begin{pmatrix} \lambda & -(1-\lambda) & 0 \\ \lambda & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} = \lambda^2 . \end{cases}$$

Hence, $W_\lambda(z)(T,L) = \lambda^2(1-z)$ for all $z \in \mathbb{R}$. Computations for the other action profiles are similar. Overall one obtains:

$$\Delta^0_\lambda = \lambda^2 \begin{pmatrix} 1 & 1 \\ \lambda & \lambda \end{pmatrix}, \Delta^k_\lambda = \lambda^2 \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}, \quad \text{and} \quad W_\lambda(z) = \lambda^2 \begin{pmatrix} 1-z & -z \\ -\lambda z & \lambda(1-z) \end{pmatrix}.$$

In particular, for $z = v_\lambda = 1/2$,

$$\mathrm{val}(W_\lambda(1/2)) = \lambda^2 \,\mathrm{val} \begin{pmatrix} 1/2 & -1/2 \\ -\lambda/2 & \lambda/2 \end{pmatrix} = 0 .$$

Like for the Big Match itself, the auxiliary game $W_\lambda(1/2)$ has a unique pair of optimal strategies, $x_\lambda = (\frac{\lambda}{1+\lambda}, \frac{1}{1+\lambda})$ and $y_\lambda = (\frac{1}{2}, \frac{1}{2})$. This is not a coincidence: for any discounted absorbing game Γ, the set of optimal stationary strategies of the game coincides with $O^*(\Gamma)$.

Marginal Discounted Value and its Limit As already argued, $O^*(\Gamma_\lambda) = O(W_\lambda(v_\lambda)) = \{(x_\lambda, y_\lambda)\}$ is a singleton, so the value operator $\mathrm{val}_{O^*(\Gamma_\lambda)}$ is trivialized. By Theorem 4.3.1, for any perturbation $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$ the marginal value is thus the unique $z \in \mathbb{R}$ satisfying

$$D_\lambda(z) \coloneqq x_\lambda^\top \left(\partial_H \Delta^k_\lambda - v_\lambda \, \partial_H \Delta^0_\lambda - z \Delta^0_\lambda \right) y_\lambda = 0 .$$

Or, equivalently, the marginal discounted values satisfy the following explicit formula:

$$\partial_H v_\lambda = \frac{x_\lambda^\top \left(\partial_H \Delta^k_\lambda - v_\lambda \partial_H \Delta^0_\lambda \right) y_\lambda}{x_\lambda^\top \Delta^0_\lambda y_\lambda} . \tag{4.1}$$

The computation of the marginal discounted values is thus straightforward for any perturbation H, as it boils down to computing the matrices $\partial_H \Delta^k_\lambda$ and $\partial_H \Delta_0$. Similarly, their limit can be obtained directly from Theorem 4.3.2 where again the value operator is trivialized.

Remark 4.4.1. The previous explicit formula (4.1) holds for every discounted absorbing game where both players have a unique pair of optimal stationary strategies, denoted by (x_λ, y_λ).

Undiscounted Marginal Value Consider now a perturbed version of the Big Match, for an admissible perturbation $H = (\tilde{g}, \tilde{q}, 0)$. The assumption $\tilde{\lambda} = 0$ is not needed here to ensure the existence of the marginal values, but is kept for simplicity and to apply Theorem 4.3.4. The perturbed game is illustrated in Figure 4.3 below, where the arrows indicate transitions from state 1 to states 2 and 3, respectively, with the indicated probabilities: More precisely, we consider an arbitrary perturbation of the stage payoffs at state 1, and an arbitrary perturbation of the two positive transition probabilities of the game. To be more precise of the latter, we consider $\tilde{q} \equiv 0$ except for $\tilde{q}_{12} \coloneqq \tilde{q}(2 \,|\, 1, T, L) \le 0$ and $\tilde{q}_{12} \coloneqq \tilde{q}(2 \,|\, 1, T, L) \le 0$, and also

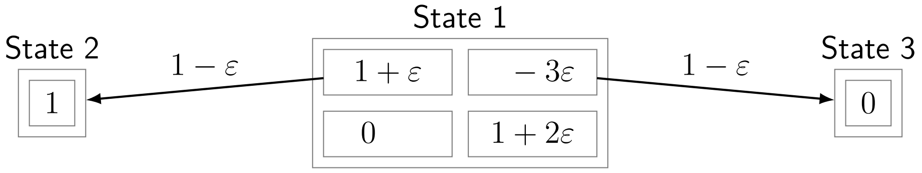


Figure 4.3: Perturbed Big Match stochastic game.

$\tilde{q}(1 \,|\, 1, T, L) = -\tilde{q}_{12}$ and $\tilde{q}(3 \,|\, 1, T, L) = -\tilde{q}_{13}$ so that $q + \varepsilon\tilde{q}$ is a transition probability for all ε sufficiently small.

Let v_0 and v_ε denote the value of the undiscounted Big Match and its perturbed version respectively. To apply Theorem 4.3.4, we proceed as follows. First, we check the continuity of $\varepsilon \mapsto v_\varepsilon$ at 0. Second, we find a polynomial P such that $P(\varepsilon, v_\varepsilon) = 0$ for all ε sufficiently small. Lastly, we check $\frac{\partial P}{\partial z}(0, v_0) \neq 0$.

Step 1: Regularity of the perturbed value functions. Note that the perturbation $\tilde{q}$ does not introduce new transitions. The continuity of $\varepsilon \mapsto v_\varepsilon$ at $\varepsilon = 0$ follows then directly from [Sol03, Theorem 6]. See also Section 4.5.3 for more details.

Step 2: Determine the desired polynomial. By Proposition 4.3.3 we have a finite set of candidates, $\mathcal{P}_{(\Gamma,H)}$, obtained by applying the operator $\Phi\colon \mathbb{R}[\beta, \varepsilon, z] \to \mathbb{R}[\varepsilon, z]$ to the determinants of all possible square sub-matrices of $W_{\beta,\varepsilon}(z)$ where $\beta > 0$ is an auxiliary discount rate. Since this matrix is of size 2×2, the set $\mathcal{P}_{(\Gamma,H)}$ contains five polynomials, the four corresponding to the entries of the matrix, denoted by $P_{T,L}, P_{T,R}, P_{B,L}$ and $P_{B,R}$, and the one corresponding to the entire matrix, denoted by P. To compute them explicitly, one needs to find $W_{\beta,\varepsilon}(z)$. We proceed as for the classical Big Match, one strategy profile at a time, to obtain the auxiliary matrices for all $\beta > 0$ and $\varepsilon \geq 0$:

$$\begin{cases} \Delta^0_{\beta,\varepsilon} = \beta^2 \begin{pmatrix} 1 + (1-\beta)\tilde{q}_{12}\varepsilon & 1 + (1-\beta)\tilde{q}_{13}\varepsilon \\ \beta & \beta \end{pmatrix}, \\ \Delta^k_{\beta,\varepsilon} = \beta^2 \begin{pmatrix} 1 + \beta\varepsilon(\tilde{g}_{11} - \tilde{q}_{12}) + \varepsilon\tilde{q}_{12} & \beta\varepsilon\tilde{g}_{12} \\ \beta\varepsilon\tilde{g}_{21} & \beta(1 + \varepsilon\tilde{g}_{22}) \end{pmatrix}. \end{cases}$$

Set then $W_{\beta,\varepsilon}(z) := \Delta^k_{\beta,\varepsilon} - z\Delta^0_{\beta,\varepsilon}$ for all $z \in \mathbb{R}$. The set $\mathcal{P}_{(\Gamma,H)}$ is then composed of the following five polynomials:

$$\begin{cases} P_{T,L}(\varepsilon, z) & = \Phi(W_{\beta,\varepsilon}(z)(T, L)) = 1 - z + \varepsilon\tilde{q}_{12}(1 - z)\,, \\ P_{T,R}(\varepsilon, z) & = \Phi(W_{\beta,\varepsilon}(z)(T, R)) = -z + \varepsilon\tilde{q}_{13}z\,, \\ P_{B,L}(\varepsilon, z) & = \Phi(W_{\beta,\varepsilon}(z)(B, L)) = -z + \varepsilon\tilde{g}_{12}\,, \\ P_{B,R}(\varepsilon, z) & = \Phi(W_{\beta,\varepsilon}(z)(B, R)) = 1 - z + \varepsilon\tilde{g}_{22}\,, \\ P(\varepsilon, z) & = \Phi(\det(W_{\beta,\varepsilon}(z)) = 1 - 2z + \varepsilon\left(\tilde{g}_{22}(1 - z) + \tilde{g}_{21}z + \tilde{q}_{12}(1 - z)^2 - \tilde{q}_{13}z^2\right) + o(\varepsilon)\,. \end{cases}$$

Note that we have abbreviated P by grouping the multiple of ε^2, as these terms do not matter for the sequel. To determine the appropriate polynomial we use the practical remark after Theorem 4.3.4. For every $\beta > 0$ and $\varepsilon \geq 0$, the optimal strategy of Player 2 in the perturbed discounted Big Match, $\Gamma_{\beta,\varepsilon}$, has full support, by continuity. Hence, as already noted, this property implies that P is the desired polynomial by Lemma 6 in [OBV23].

Step 3: A formula for the undiscounted marginal values. One easily checks that $\frac{\partial P}{\partial z}(0, v_0) = -2 \neq 0$ for all the considered perturbations $(\tilde{g}, \tilde{q})$. Theorem 4.3.4 can thus be applied to obtain the desired formula:

$$\partial_H v_0 = -\frac{\frac{\partial P}{\partial \varepsilon}(0, v_0)}{\frac{\partial P}{\partial z}(0, v_0)} = \frac{\tilde{q}_{12} - \tilde{q}_{13}}{8} + \frac{\tilde{g}_{21} + \tilde{g}_{22}}{4}\,.$$

Final Remark Consider the perturbation $H = (\tilde{g}, \tilde{q}, 0)$ described in Figure 4.3: do the undiscounted marginal values and the limit discounted marginal values coincide in this example? From the expressions of $\Delta^0_{\beta,\varepsilon}$ and $\Delta^k_{\beta,\varepsilon}$ one easily determines $\partial_H \Delta^0_\lambda$ and $\partial_H \Delta^0_\lambda$ by setting $\beta := \lambda$ and then taking the entry-wise derivatives in ε at 0. Consequently,

$$\partial_H \Delta^0_\lambda = \lambda^2 \begin{pmatrix} (1-\lambda)\tilde{q}_{12} & (1-\lambda)\tilde{q}_{13} \\ 0 & 0 \end{pmatrix},$$

$$\partial_H \Delta^k_\lambda = \lambda^2 \begin{pmatrix} \lambda(\tilde{g}_{11} - \tilde{q}_{12}) + \tilde{q}_{12} & \lambda\tilde{g}_{12} \\ \lambda\tilde{g}_{21} & \lambda\tilde{g}_{22} \end{pmatrix}.$$

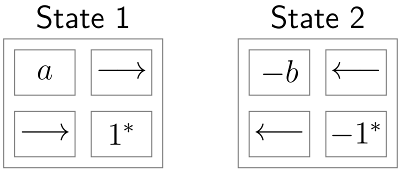
Replacing these matrices in (4.1), together with (x_λ, y_λ) and v_λ one thus obtains

$$\partial_H v_\lambda = \frac{(1\,\breve{}\,\lambda)(\tilde{q}_{12} - \tilde{q}_{13})}{8} + \frac{\lambda\tilde{g}_{11} + \lambda\tilde{g}_{12} + \tilde{g}_{21} + \tilde{g}_{22}}{4}\,.$$

Therefore, the equality $\partial_H v_0 = \lim_{\lambda\to 0} \partial_H v_\lambda$ holds for this example.

4.4.2 The Bewley-Kohlberg Stochastic Game

Our next example is a (non-absorbing) stochastic game introduced by [BK78, page 120], which was also considered in [Vig13]. A representation of this game is given in Figure 4.5, where $a, b \geq 0$ are fixed parameters. The Bewley-Kohlberg stochastic game has four states, two of which are absorbing with payoffs 1 and -1. We label these states as states 3 and 4, respectively. In Figure 4.4, the arrows indicate a deterministic transition from state 1 to state 2, and vice versa, and a stage payoff of 0. The initial state is fixed to state $k = 1$. Let v_λ denote the value of the λ-discounted game, and let $v_0 := \lim_{\lambda\to 0} v_\lambda$ be the undiscounted value. From the literature we know that, for each $\lambda \in (0, 1]$, both players have a unique optimal stationary strategy in the λ-discounted game, which is of full support. We also know that $v_0 = \frac{a-b}{a+b+2}$.



State 1

a	$\longrightarrow$
$\longrightarrow$	1^*

State 2

$-b$	$\longleftarrow$
$\longleftarrow$	-1^*

Figure 4.4: The Bewley-Kohlberg stochastic game with parameters $a, b \geq 0$. The originally published game corresponds to the case $a = b = 1$.

Marginal Discounted Value For every admissible perturbation $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$, the marginal discounted values can be obtained from Theorem 4.3.1, for example using the algorithm we described after this statement to obtain an approximation of $\partial_H v$ to a desired level of accuracy. Similarly, their limit as λ goes to 0 can be derived from Theorem 4.3.2.

On the other hand, it is worth noting that an explicit expression like (4.1) can not be obtained here, even when both players have a unique optimal stationary strategy for all $\lambda > 0$. This is because the game is not absorbing, and thus $O^*(\Gamma_\lambda)$ contains, but may not be equal to, the set of optimal stationary strategies of Γ_λ.

Marginal Undiscounted Value We now focus on the undiscounted marginal values for a given admissible perturbation $H = (\tilde{g}, \tilde{q}, 0)$. For the sake of simplicity, we assume that some of the entries of $\tilde{g}$ and $\tilde{q}$ are 0, namely we consider four perturbation parameters, denoted by $(\tilde{g}_1, \tilde{g}_2) \in \mathbb{R}^2$ and $(\tilde{q}_1, \tilde{q}_2) \in \mathbb{R}^2_+$, and set:

- $\tilde{g}(1, T, L) := \varepsilon\tilde{g}_1$ and $\tilde{g}(2, T, L) := \varepsilon\tilde{g}_2$;
- $\tilde{q}(1 \,|\, 1, B, R) := \varepsilon\tilde{q}_1$ and $\tilde{q}(3 \,|\, 1, B, R) := -\varepsilon\tilde{q}_1$;
- $\tilde{q}(2 \,|\, 2, B, R) := \varepsilon\tilde{q}_2$ and $\tilde{q}(4 \,|\, 2, B, R) := -\varepsilon\tilde{q}_2$;
- All other entries of $\tilde{g}$ and $\tilde{q}$ are equal to 0.

In other words, the perturbations consist in replacing the stage payoffs a and $-b$ with $a + \varepsilon\tilde{g}_1$ and $-b + \varepsilon\tilde{g}_2$, respectively, and the deterministic absorption probabilities with $1 - \varepsilon\tilde{q}_1$ and $1 - \varepsilon\tilde{q}_2$. This perturbed Bewley-Kolhberg stochastic game is illustrated in the Figure 4.5 below: The value of this perturbed undiscounted game is denoted by $v_{0,\varepsilon}$. Note that $\tilde{q}$ does not introduce new (non-loop) transitions, so that the map $\varepsilon \mapsto v_{0,\varepsilon}$ is continuous at $\varepsilon = 0$ and the marginal undiscounted values $\partial_H v_0$ exist. We determine $\partial_H v_0$ in three steps.

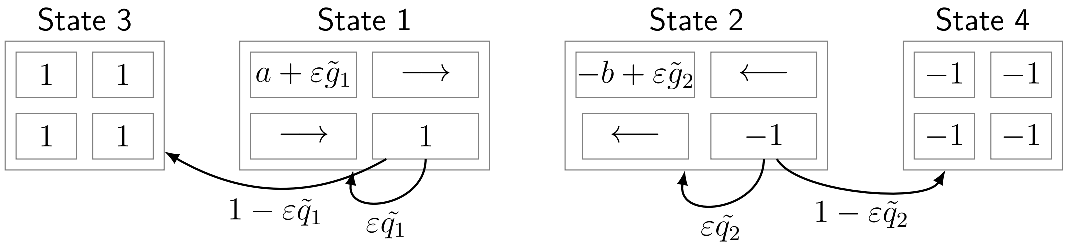


Figure 4.5: The perturbed Bewley-Kohlberg stochastic game with parameters $a, b \geq 0$.

Step 1: Auxiliary Matrices We consider the auxiliary matrices of a β-discounted version of the game, denoted by $\Delta^0_{\beta,\varepsilon}$ and $\Delta^k_{\beta,\varepsilon}$. To compute them, we follow the construction described in Section 4.2.2. A direct calculation gives that $\Delta^0_{\beta,\varepsilon}(z)$ is given by

$$\beta^2 \begin{pmatrix} \beta^2 & \beta & \beta & \beta(2-\beta) \\ \beta & \beta(1-\varepsilon\tilde{q}_2(1-\beta)) & \beta(2-\beta) & 1-\varepsilon\tilde{q}_2(1-\beta) \\ \beta & \beta(2-\beta) & \beta(1-\varepsilon\tilde{q}_1(1-\beta)) & 1-\varepsilon\tilde{q}_1(1-\beta) \\ \beta(2-\beta) & 1-\varepsilon\tilde{q}_2(1-\beta) & 1-\varepsilon\tilde{q}_1(1-\beta) & (1-\varepsilon\tilde{q}_1(1-\beta))(1-\varepsilon\tilde{q}_2(1-\beta)) \end{pmatrix},$$

and $\Delta^1_{\beta,\varepsilon}(z)$ is given by

$$\beta^2 \begin{pmatrix} \beta^2(a+\varepsilon\tilde{g}_1) & \beta(a+\varepsilon\tilde{g}_1) & \beta(1-\beta)(\varepsilon\tilde{g}_2-b) & 0 \\ \beta(a+\varepsilon\tilde{g}_1) & \beta(1-\varepsilon\tilde{q}_2(1-\beta))(a-\varepsilon\tilde{g}_1) & 0 & (1-\beta)(1-\varepsilon\tilde{q}_2(1-\beta)) \\ \beta(1-\beta)(\varepsilon\tilde{g}_2-b) & 0 & \beta(1-\varepsilon\tilde{q}_1(1-\beta)) & 1-\varepsilon\tilde{q}_1(1-\beta) \\ 0 & -(1-\beta)(1-\varepsilon\tilde{q}_2(1-\beta)) & 1-\varepsilon\tilde{q}_1(1-\beta) & (1-\varepsilon\tilde{q}_1(1-\beta))(1-\varepsilon\tilde{q}_2(1-\beta)) \end{pmatrix}.$$

One then sets $W_{\beta,\varepsilon}(z) := \Delta^k_{\beta,\varepsilon} - z\Delta^0_{\beta,\varepsilon}$, for all $z \in \mathbb{R}$.

Step 2: Desired Polynomial We now look for a polynomial P satisfying $P(\varepsilon, v_{0,\varepsilon}) = 0$ for all $\varepsilon \geq 0$ small enough. To do so, we consider the possible candidates, which are in the set $\mathcal{P}_{(\Gamma,H)}$ by Proposition 4.3.3. To determine the suitable one, we first note that for every $\beta \in (0,1]$, the unperturbed β-discounted stochastic game admits a unique pair of optimal stationary strategies which is of full support. This property was already noted by [BK78]. By continuity, this property still holds in the perturbed auxiliary discounted game $\Gamma_{\beta,\varepsilon}$ for $\varepsilon \geq 0$ sufficiently small. As already noted (see Comment 3 after Proposition 4.3.3), this property implies that the desired polynomial is $P(\varepsilon, z) := \Phi(\det(W_{\beta,\varepsilon}(z)))$. A direct computation using Matlab gives then $P(\varepsilon, z) = p(\varepsilon, z)^2$, where

$$p(\varepsilon, z) = a - b - z(a+b+2) + \varepsilon\left((1-z)(\tilde{g}_1 - (z-a)\tilde{q}_1) + (1+z)(\tilde{g}_2 + (z+b)\tilde{q}_2\right) + o(\varepsilon)\,.$$

Here again, we have abbreviated the polynomial, as the terms in ε^2 play no role in the sequel.

Step 3: Formula for the Undiscounted Marginal Value Lastly, we use Theorem 4.3.4 to determine the undiscounted marginal values. By construction p satisfies $p(\varepsilon, v_{0,\varepsilon}) = 0$ for all ε sufficiently small. We then check that $\frac{\partial p}{\partial z}(0, v_0) = -(a+b+2) \neq 0$. As the map $\varepsilon \mapsto v_\varepsilon$ is continuous at $\varepsilon = 0$ by the choice of $\tilde{q}$, we can apply Theorem 4.3.4 to obtain the following explicit expression (recall that $v_0 = \frac{a-b}{a+b+2}$):

$$\partial_H v_0 = -\frac{\frac{\partial p}{\partial \varepsilon}(0, v_0)}{\frac{\partial p}{\partial z}(0, v_0)} = \frac{(1-v_0)\tilde{g}_1 + (1+v_0)\tilde{g}_2 + (v_0 - a)(1-v_0)\tilde{q}_1 + (v_0+b)(1+v_0)\tilde{q}_2}{a+b+2}\,.$$

In particular, for $a = b = 1$ (and thus $v_0 = 0$) the following simpler formula holds:

$$\partial_H v_0 = \frac{\tilde{g}_1 + \tilde{g}_2 - \tilde{q}_1 + \tilde{q}_2}{4}\,.$$

Remark 4.4.2. In the Bewley-Kohlberg game, the polynomial $P(\varepsilon, z) := \Phi(\det(W_{\beta,\varepsilon}(z)))$ satisfies $\frac{\partial P}{\partial z}(0, v_0) = 0$. However, we could replace P by its divider, p, to obtain the desired formula from Theorem 4.3.4.

Remark 4.4.3. The two considered examples share a common regularity property: there exists a polynomial $P(\varepsilon, z)$ such that $P(\varepsilon, v_\varepsilon) = 0$ for all $\varepsilon \geq 0$ and $\frac{\partial P}{\partial z}(0, v_0)$ is independent from the perturbation (though it could be equal to 0). This property holds as long as there exists a pair of optimal stationary strategies in the perturbed discounted stochastic game $\Gamma_{\beta,\varepsilon}$ with a fixed and equal support for all $\beta > 0$ and $\varepsilon \geq 0$ sufficiently small. The reason for this stability is as follows: the fixed support implies there exists a constant sub-matrix that defines the optimal strategies, and this sub-matrix can then be chosen to define P, independently of the perturbation.

4.5 Previous Results

Before going into the proofs of our main contributions we provide some useful preliminary results. We start with the theory of matrix games developed by [SS50] to then briefly present the theory of semi-algebraic sets. Lastly, we present some known results regarding the regularity of the perturbed value function of a stochastic game following [FV97, Chapter 4] for the discounted case, and [Sol03] for the undiscounted case.

4.5.1 Shapley-Snow Theory

We follow the notation used in [SS50]. For any square matrix M, we denote $S(M)$, and $\mathrm{co}(M)$ the sum of its entries and its co-factor matrix respectively.

Definition 4.5.1. For any matrix M, a *Shapley-Snow kernel* is a square submatrix $\overline{M}$ satisfying:

- $S\left(\mathrm{co}(\overline{M})\right) \neq 0$.
- The strategies $\overline{x} = \frac{\mathrm{co}(\overline{M})}{S\left(\mathrm{co}(\overline{M})\right)}\overline{1}$ and $\overline{y} = \frac{\mathrm{co}(\overline{M})^\top}{S\left(\mathrm{co}(\overline{M})\right)}\overline{1}$, when completed by zeros, are optimal strategies of M, respectively for Player 1 and Player 2.

Shapley-Snow kernels characterize the extreme points of the set of optimal strategies, denoted $O^*(M)$. In particular, every matrix admits at least one and at most finitely many of these kernels. Moreover, for any Shapley-Snow kernel $\overline{M}$ of M the following properties hold:

(i) $\mathrm{val}(M) = \frac{\det \overline{M}}{S\left(\mathrm{co}(\overline{M})\right)}$, so in particular if $\mathrm{val}(M) = 0$, then $\det(\overline{M}) = 0$.

(ii) All entries of $\mathrm{co}(\overline{M})$ are of same sign, and not all 0.

4.5.2 Semi-Algebraic Sets

A set $A \subset \mathbb{R}^d$ is basic semi-algebraic if it is defined by finitely many polynomial equalities or strict inequalities, i.e., there exists $L \in \mathbb{N}$ and polynomials p_0, $p_1, \dots, p_L$ in $\mathbb{R}[x_1, \dots, x_d]$ such that

$$A = \{(x_1, \dots, x_d) \in \mathbb{R}^d : p_0(x_1, \dots, x_d) = 0,\ p_1(x_1, \dots, x_d) > 0,\ \dots\ , p_L(x_1, \dots, x_d) > 0\}\,.$$

A semi-algebraic set is the finite union or intersection of basic semi-algebraic sets. A semi-algebraic function is a function whose graph is a semi-algebraic set.

The following result, referred to as the Tarski-Seidenberg elimination theorem, establishes the stability of semi-algebraic by projection to lower dimensions.

Theorem 4.5.2 (Tarski-Seidenberg theorem)**.** *Consider $d \in \mathbb{N}$ and $A \subset \mathbb{R}^{d+\ell}$ a semi-algebraic set. Then, $\{x \in \mathbb{R}^d : \exists w \in \mathbb{R}^\ell \quad (x, w) \in A\}$ is a semi-algebraic set of $\mathbb{R}^d$.*

As a consequence, any set that can be expressed using first-order formulas over the real field is semi-algebraic.

Local expansions of semi-algebraic functions were investigated by [Pui50]. A Puiseux series is a function of the form

$$f(\varepsilon) = \sum_{m \geq m_0} c_m \varepsilon^{m/M}\,,$$

where $M \in \mathbb{N}$, $m_0 \in \mathbb{Z}$ and coefficients $(c_m)_{m \geq m_0} \subset \mathbb{R}$. Then, we have the following result.

Theorem 4.5.3 ([Pui50])**.** *Let $P \in \mathbb{R}[x, y]$ be a nonzero polynomial and $a > 0$. Suppose that $f \colon (0, a) \to \mathbb{R}$ is a continuous function satisfying $P(\varepsilon, f(\varepsilon)) = 0$, for all $\varepsilon \in (0, a)$. Then $\varepsilon \mapsto f(\varepsilon)$ admits a Puiseux expansion near 0.*

4.5.3 Regularity of the Perturbed Value Function

Consider a discounted stochastic game Γ with $\lambda > 0$ and an admissible perturbation $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$. Then $v_\varepsilon := \mathrm{val}(\Gamma + \varepsilon H)$ is the $(\lambda + \varepsilon\tilde{\lambda})$-discounted value of some stochastic game. For any map $\phi \colon K \times I \times J \to \mathbb{R}^K$, let

$$\|\phi\|_1 := \max_{\ell \in K, (i,j) \in I \times J} \sum_{\ell' \in K} |\phi(\ell' \mid \ell, i, j)| \; .$$

Recall that $\tilde{q} \colon K \times I \times J \to \mathbb{R}^K$, but it is not a transition function (rather, $q + \varepsilon\tilde{q}$ is a transition function for all ε sufficiently small), so $\|\tilde{q}\|_1$ is not necessarily equal to 1. Let $\|\cdot\|_\infty$ denote the standard L^∞-norm of $\mathbb{R}^d$, and let $z_+ := \max(0, z)$ for all $z \in \mathbb{R}$. The following property is a rephrasing of Equation (4.19) in [FV97, Chapter 4] using the current notation.

Lemma 4.5.4. *The following inequality holds for all $\varepsilon, \varepsilon' \geq 0$ sufficiently small:*

$$|v_\varepsilon - v_{\varepsilon'}| \leq |\varepsilon - \varepsilon'| \left(\|\tilde{g}\|_\infty + \frac{1 - (\lambda + \varepsilon\tilde{\lambda})}{\lambda + \varepsilon\tilde{\lambda}} \|\tilde{q}\|_1 \, \|g + \varepsilon'\tilde{g}\|_\infty + 2\frac{|\tilde{\lambda}|}{\lambda + \varepsilon\tilde{\lambda}} \|g + \varepsilon'\tilde{g}\|_\infty \right) .$$

Next, consider an undiscounted stochastic game Γ and an admissible perturbation $H = (\tilde{g}, \tilde{q}, 0)$. By definition, $v_\varepsilon := \mathrm{val}(\Gamma + \varepsilon H)$ is then the undiscounted value of some stochastic game. For all pairs of transition functions q_1 and q_2 with the same domain, i.e., $q_1, q_2 \colon K \times I \times J \to \Delta(K)$, let

$$r(q_1, q_2) := \max_{\substack{(i,j) \in I \times J \\ (\ell, \ell') \in K^2, \ell \neq \ell'}} \left\{ \max\left\{ \frac{q_1(\ell' \mid \ell, i, j)}{q_2(\ell' \mid \ell, i, j)}, \frac{q_2(\ell' \mid \ell, i, j)}{q_1(\ell' \mid \ell, i, j)} \right\} - 1 \right\} ,$$

with the convention $z/0 = +\infty$ for all $z > 0$, and $0/0 = 1$.

The next result is a rephrasing of [Sol03, Theorem 6] using the current notation.

Lemma 4.5.5. *Let Γ be a discounted or undiscounted stochastic game, and let $H = (\tilde{g}, \tilde{q}, 0)$ be an admissible perturbation. Then, for all $\varepsilon, \varepsilon' > 0$ sufficiently small,*

$$|v_{\varepsilon'} - v_\varepsilon| \leq \frac{4|K| \, r(q + \varepsilon\tilde{q}, \, q + \varepsilon'\tilde{q})}{(1 - 2|K| \, r(q + \varepsilon\tilde{q}, \, q + \varepsilon'\tilde{q}))_+} \|g + \varepsilon\tilde{g}\|_\infty + |\varepsilon - \varepsilon'| \, \|\tilde{g}\|_\infty .$$

Note that the perturbed value function is Lipschitz continuous as soon as $r(q, q + \varepsilon\tilde{q}) = O(\varepsilon)$. By [Sol03], this condition holds if $\tilde{q}$ does not introduce new transitions, i.e., if there is no $(i, j) \in I \times J$ and $(\ell, \ell') \in K^2$ with $\ell \neq \ell'$ so that $q(\ell' \mid \ell, i, j) = 0$ and $(q + \varepsilon\tilde{q})(\ell' \mid \ell, i, j) > 0$, or the converse, for all ε sufficiently small. Before we continue, a technical observation on the definition of r is needed.

Remark 4.5.6. In the original statement of [Sol03], the definition of r was slightly different: there was no condition $\ell \neq \ell'$ in the maximum. However, this additional condition comes for free as the bound in [Sol03, Theorem 6] relies on the limit discounted occupation times for a Markov chain, which by [FW12] depend only on the transitions between pairs of different states. Therefore, we have added the condition $\ell \neq \ell'$ in the definition of r.

The following three results are direct consequences of Lemma 4.5.4, Lemma 4.5.5, and the theory of semi-algebraic sets (see Section 4.5.2). Their proof is provided in Section 4.8 for completeness.

Proposition 4.5.7. *Consider a discounted or undiscounted stochastic game Γ and an admissible perturbation $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$. Then, there exists $\varepsilon_0 > 0$ so that $\varepsilon \mapsto v(\Gamma + \varepsilon H)$ is continuous on $(0, \varepsilon_0]$. Moreover, continuity at $\varepsilon = 0$ holds if either: (i) $\lambda > 0$; or (ii) $\lambda = 0$ and $r(q, q + \varepsilon \tilde{q}) = O(\varepsilon)$.*

Proposition 4.5.8. *Consider a discounted or undiscounted stochastic game Γ and an admissible perturbation $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$. The marginal values exist if either (i) $\lambda > 0$; or (ii) $\lambda = 0$ and $r(q, q + \varepsilon \tilde{q}) = O(\varepsilon)$.*

Proposition 4.5.9. *Let Γ_λ be a discounted stochastic game, and $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$ an admissible perturbation. Then $(\partial_H v(\Gamma_\lambda))_\lambda$ is uniformly bounded if $r(q, q + \varepsilon \tilde{q}) = O(\varepsilon)$.*

4.6 Proofs of Main Contributions

We are now ready to prove our main results. In the sequel, consider a fixed stochastic game $\Gamma = (K, k, I, J; g, q, \lambda)$ and an admissible perturbation $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$. To alleviate the notation we set $v := v(\Gamma)$ and $\partial_H v := \partial_H v(\Gamma)$ throughout this section. Also, note that we omitted the sub-index λ in Theorem 4.3.1, as the discount rate is fixed. We follow the same convention in its proof.

4.6.1 Proof of Theorem 4.3.1

The proof is divided into three steps. First, note that the map $z \mapsto D(z)$ is well-defined. This is the case as the set of admissible strategies $O^*(\Gamma)$ is compact and convex (by the minmax theorem), and the payoff function is bilinear. Hence, $D(z)$ is well-defined by [Sio58, Theorem 3.4]. Second, we argue the existence of at most one $z \in \mathbb{R}$ such that $D(z) = 0$. On the one hand, all the entries of Δ^0 are positive as noted in Section 4.2.2. On the other hand, the monotonicity and continuity of the value operator imply that $z \mapsto D(z)$ is a strictly decreasing, continuous, and bijective map from $\mathbb{R}$ to $\mathbb{R}$. Our second step follows. Lastly, we claim that $D(\partial_H v) = 0$. We present two alternative proofs.

First approach. Let $(\varepsilon_m, x_m, y_m) \in (0,1) \times \Delta(I)^K \times \Delta(J)^K$ be a sequence converging to $(0, x_0, y_0)$ such that, for every m, $\varepsilon_m > 0$ and (x_m, y_m) is a pair of optimal stationary strategies of the perturbed game $\Gamma + \varepsilon_m H$, whose value is denoted by v_{ε_m}. To establish the desired result, i.e., $D(\partial_H v) = 0$, it is enough to prove the following relations, where, as before, $\widehat{x_0} = \otimes_{\ell \in K} x_0^\ell \in \Delta(I^K)$ and $\widehat{y_0} = \otimes_{\ell \in K} y_0^\ell \in \Delta(J^K)$:

$$\begin{cases} \widehat{x_0}^\top (\partial_H \Delta^k - v\, \partial_H \Delta^0 - (\partial_H v)\, \Delta^0)\, y \geq 0, & \forall y \in O_2^*(\Gamma), \\ x^\top (\partial_H \Delta^k - v\, \partial_H \Delta^0 - (\partial_H v)\, \Delta^0)\, \widehat{y_0} \leq 0, & \forall x \in O_1^*(\Gamma)\,. \end{cases}$$

For all $m \in \mathbb{N}$ and $y \in O_2^*(\Gamma)$, the following holds:

$$\begin{aligned} 0 &\leq \widehat{x}_m^\top \left(\Delta^k_{\varepsilon_m} - v_{\varepsilon_m} \Delta^0_{\varepsilon_m} \right) y\,, \\ &= \widehat{x}_m^\top \Big(\Delta^k - v\Delta^0 + \Big(\partial_H \Delta^k - v\, \partial_H \Delta^0 - (\partial_H v) \Delta^0 \Big) \varepsilon_m + o(\varepsilon_m) \Big) y\,, \\ &= \widehat{x}_m^\top \Big(\Delta^k - v\Delta^0 \Big) y + \widehat{x}_m^\top \Big(\partial_H \Delta^k - v\, \partial_H \Delta^0 - (\partial_H v) \Delta^0 \Big) y\, \varepsilon_m + o(\varepsilon_m), \\ &\leq \varepsilon_m\, \widehat{x}_m^\top \Big(\partial_H \Delta^k - v\, \partial_H \Delta^0 - (\partial_H v) \Delta^0 \Big) y + o(\varepsilon_m)\,. \end{aligned}$$

Indeed, the first inequality follows from Theorem 1 in [AOB19], i.e., for any discounted stochastic game with value v one has $\mathrm{val}(\Delta^k - v\Delta^0) = 0$, and from the fact that the optimality of x_m in the stochastic game $\Gamma + \varepsilon_m H$ implies the optimality of $\widehat{x}_m$ in the auxiliary matrix game $\Delta^k_{\varepsilon_m} - v_{\varepsilon_m}\Delta^0_{\varepsilon_m}$. The second line is a standard asymptotic development for $\Delta^k_{\varepsilon_m}$, v_{ε_m} and $\Delta^0_{\varepsilon_m}$ at 0. The third line is a consequence of linearity, as we are dealing with mixed strategies and matrix games. The fourth line follows from the choice of y as an optimal strategy of $\Delta^k - v\Delta^0$, and from the fact that the latter has value 0 by [AOB19, Theorem 1]. Dividing the last inequality by ε_m and then taking $m \to \infty$, one thus obtains

$$\widehat{x_0}^\top \Big(\partial_H \Delta^k - v\, \partial_H \Delta^0 - (\partial_H v)\Delta^0\Big)\, y \geq 0, \quad \forall y \in O_2^*(\Gamma)\,.$$

By reversing the roles of the players, one similarly obtains the desired relation for $\widehat{y}_0$, which proves the claim. □

Second approach. Consider a matrix game $M(\varepsilon)$ whose entries depend on ε and are differentiable in a neighborhood of 0. Let $\frac{\partial}{\partial\varepsilon}M(\varepsilon)$ be its entry-wise derivative, and let $O^*(M)$ be the set of optimal strategies of $M(0)$. By [Mil56], if there exist two matrices of the same size M and N such that $M(\varepsilon) = M + \varepsilon N$, then the value operator commutes with differentiation up to a restriction of the strategy domain, namely:

$$\frac{\partial}{\partial\varepsilon}\Big(\mathrm{val}\ M(\varepsilon)\Big)_{|\varepsilon=0} = \mathrm{val}_{O^*(M)}\left(\frac{\partial}{\partial\varepsilon}M(\varepsilon)_{|\varepsilon=0}\right)\,.$$

The extension of this result to a non-linear differentiable dependency on ε is straightforward. Indeed, in this case $N := \frac{\partial}{\partial\varepsilon}M(\varepsilon)_{|\varepsilon=0}$ satisfies $M(\varepsilon) = M(0) + \varepsilon N + E(\varepsilon)$, where $E(\varepsilon)$ is a matrix of same size as $M(\varepsilon)$ such that all its entries are $o(\varepsilon)$ as ε goes to 0. The extension of Mills' result follows from the monotonicity and continuity of the value operator. Next, set $M(\varepsilon) := W^k(v_\varepsilon) = \Delta^k_\varepsilon - v_\varepsilon\Delta^0_\varepsilon$ for all ε. Then, by definition, $O^*(M) = O(W(v)) = O^*(\Gamma)$. On the other hand, $\mathrm{val}(M(\varepsilon)) = 0$ for all ε sufficiently small by Theorem 1 in [AOB19]. The desired result follows then from the differentiable extension of Mills' result, applied to $M(\varepsilon)$. Indeed,

$$\begin{aligned} 0 = \frac{\partial}{\partial\varepsilon}\Big(\mathrm{val}(\Delta^k_\varepsilon - v_\varepsilon\Delta^0_\varepsilon)\Big)_{|\varepsilon=0} &= \mathrm{val}_{O^*(\Gamma)}\left(\frac{\partial}{\partial\varepsilon}\Big(\Delta^k_\varepsilon - v_\varepsilon\Delta^0_\varepsilon\Big)_{|\varepsilon=0}\right) \\ &= \mathrm{val}_{O^*(\Gamma)}\Big(\partial_H\Delta^k - v\,\partial_H\Delta^0 - (\partial_H v)\Delta^0\Big)\,, \\ &= D(\partial_H v)\,. \end{aligned}$$

These three steps give the desired result. □

4.6.2 Simplified Formula for Perturbed Payoffs

We consider Theorem 4.3.1 for the particular case of a perturbation of the payoffs only, i.e., $H = (\tilde{g}, 0, 0)$. We prove that, in this case, our formula is equivalent to the following simple expression:

$$\partial_H v(\Gamma) = \mathrm{val}_{O(\Gamma)}\,\tilde{\gamma}(x, y)\,,$$

where $\tilde{\gamma}$ is the expected payoff function of $\tilde{\Gamma} = (K, k, I, J; \tilde{g}, q, \lambda)$, and $O(\Gamma)$ is the set of optimal stationary strategies of Γ (not to be confused with $O^*(\Gamma)$, which is the set of optimal strategies of the matrix game $W(v)$).

Proof. We proceed in three steps. First, let $\tilde{\Delta}^0$ and $\tilde{\Delta}^k$ denote the auxiliary matrices of $\tilde{\Gamma}$. By construction, $\Delta^0 = \tilde{\Delta}^0$ because the 0-auxiliary matrix depends on (q, λ), but not on the payoffs as noted in Section 4.2.2. For the same reason $\partial_H \Delta^0 = 0$. Second, by the linearity of the determinant, the definition of the k-auxiliary matrix implies $\partial_H \Delta^k = \tilde{\Delta}^k$. Theorem 4.3.1 thus boils down to $\partial_H v(\Gamma)$ being the unique solution to

$$z \in \mathbb{R}, \quad \mathrm{val}_{O^*(\Gamma)}(\tilde{\Delta}^k - z\,\tilde{\Delta}^0) = 0\,.$$

Third, we prove that $O^*(\Gamma)$ can be replaced with $O(\Gamma)$ in the previous equation. In the proof of Theorem 4.3.1 (first approach, replacing $\partial_H \Delta^k - v \partial_H \Delta^0 - (\partial_H v) \Delta^0$ with $\tilde{\Delta}^k - (\partial_H v) \tilde{\Delta}^0$) we proved the existence of $x_0 \in \Delta(I)^K$ such that

$$\widehat{x}_0^\top \left(\tilde{\Delta}^k - \partial_H v\, \tilde{\Delta}^0\right) y \geq 0, \quad \forall y \in O_2^*(\Gamma)\,.$$

Using the positivity of $\tilde{\Delta}^0$ and taking the minimum over y one thus obtains:

$$\min_{y \in O_2^*(\Gamma)} \frac{\widehat{x}_0^\top \tilde{\Delta}^k\, y}{\widehat{x}_0^\top \tilde{\Delta}^k\, y} \geq \partial_H v\,.$$

By [AOB19], $O_2(\Gamma) \subset O_2^*(\Gamma)$ holds via the canonical inclusion $y \mapsto \widehat{y}$. Hence,

$$\min_{y \in O_2(\Gamma)} \frac{\widehat{x}_0^\top \tilde{\Delta}^k\, \widehat{y}}{\widehat{x}_0^\top \tilde{\Delta}^k\, \widehat{y}} \geq \partial_H v\,.$$

On the other hand, the definition of the auxiliary matrices (see Section 4.2.2) and the multilinearity of the following map:

$$\begin{array}{ccc} \Delta(I)^K \times \Delta(J)^K \times \mathbb{R} & \to & \mathbb{R} \\ (x, \quad y, \quad z) & \mapsto & \widehat{x}^\top \tilde{\Delta}^k\, \widehat{y} - z\, \widehat{x}^\top \tilde{\Delta}^0\, \widehat{y}\,, \end{array}$$

implies that, for any pair of stationary strategies (x, y),

$$\tilde{\gamma}(x, y) = \frac{\widehat{x}^\top \tilde{\Delta}^k\, \widehat{y}}{\widehat{x}^\top \tilde{\Delta}^0\, \widehat{y}}\,.$$

Putting the last two equations together one thus obtains

$$\min_{y \in O_2(\Gamma)} \tilde{\gamma}(x_0, y) \geq \partial_H v\,.$$

Therefore, $\max_{x \in O_1(\Gamma)} \min_{y \in O_2(\Gamma)} \tilde{\gamma}(x, y) \geq \partial_H v$. Finally, we reverse the roles of the players to obtain the reverse inequality and conclude because the $\max\min$ is always smaller or equal to the $\min\max$, i.e.,

$$\partial_H v \geq \min_{y \in O_2(\Gamma)} \max_{x \in O_1(\Gamma)} \tilde{\gamma}(x, y) \geq \max_{x \in O_1(\Gamma)} \min_{y \in O_2(\Gamma)} \tilde{\gamma}(x_0, y) \geq \partial_H v\,.$$

This concludes the proof. □

4.6.3 Proof of Theorem 4.3.2

Proof. For all $\lambda \in (0,1]$, recall that $\Gamma_\lambda := (K,k,I,J;g,q,\lambda)$, and let $v_\lambda \in \mathbb{R}$ denote its value. Let Δ_λ^0, Δ_λ^k, be the corresponding auxiliary matrices, and $W_\lambda(z) := \Delta_\lambda^k - z\Delta_\lambda^0$ for all $z \in \mathbb{R}$. Lastly, for all $z \in \mathbb{R}$, we set

$$\begin{cases} D_\lambda(z) := \mathrm{val}_{O^*(\Gamma_\lambda)} \left(\partial_H \Delta_\lambda^k - v_\lambda \, \partial_H \Delta_\lambda^0 - z\, \Delta_\lambda^0 \right), \\ F(z) := \lim\limits_{\lambda \to 0} \dfrac{1}{\lambda^{|K|}} D_\lambda(z) \in [-\infty, +\infty] \,. \end{cases}$$

We start by assuming the following claims: (1) F is well-defined, (2) F is strictly decreasing provided it is not constant, and (3) F is not constant. Together, these statements prove that there is a unique point where F changes sign. Let w be the unique point where $F(z) > 0$ for all $z < w$ and $F(z) < 0$ for all $z > w$. Then, for all $\delta > 0$, we have that $F(w-\delta) > 0$. Therefore, by the definition of F, there exists $\lambda_0 > 0$ so that

$$D_\lambda(w-\delta) > 0, \quad \forall \lambda \in (0, \lambda_0)\,.$$

By Theorem 4.3.1, this implies $\partial_H v_\lambda > w - \delta$ for all $\lambda \in (0,\lambda_0)$. Since δ is arbitrary, we deduce that $\liminf_{\lambda\to 0} \partial_H v_\lambda \geq w$. Similarly, we deduce that $\limsup_{\lambda \to 0} \partial_H v_\lambda \leq w$. Therefore, $\lim_{\lambda\to 0} \partial_H v_\lambda$ exists and satisfies the desired property. The rest of the proof proves the three above-mentioned claims.

Claim 1. Recall that, by [Pui50], v_λ is a Puiseux series near 0. Also, the entries of $\partial_H \Delta_\lambda^k$, $\partial_H \Delta_\lambda^0$, and Δ_λ^0 are polynomials in λ. Fix $z \in \mathbb{R}$. By [Mil56], $D_\lambda(z)$ is the value of a linear program (see Section 4.7 for more details). Therefore, as λ goes to 0, all the rest of parameters being fixed, $D_\lambda(z)$ is a piece-wise rational fraction in λ and v_λ. Hence, $\lambda \mapsto D_\lambda(z)$ is a Puiseux series near 0, which gives the desired result.

Claim 2. Consider $(z_1, z_2) \in \mathbb{R}^2$ such that $z_2 > z_1$. Then, for all $\lambda \in (0,1]$ and $(x,y) \in O^*(\Gamma_\lambda)$,

$$x^\top \left(\partial_H \Delta_\lambda^k - v_\lambda \partial_H \Delta_\lambda^0 - z_1 \Delta_\lambda^0 \right) y - x^\top \left(\partial_H \Delta_\lambda^k - v_\lambda \partial_H \Delta_\lambda^0 - z_2 \Delta_\lambda^0 \right) y = (z_2 - z_1) x^\top \Delta_\lambda^0 y \,.$$

Moreover, as all the entries of Δ_λ^0 are larger or equal to $\lambda^{|K|}$, as noted in Section 4.2.2, and because x and y are probabilities, one has

$$x^\top \Delta_\lambda^0 y \geq \lambda^{|K|}\,.$$

Maximization over $x \in O^*(\Gamma)$, and minimization over $y \in O_2^*(\Gamma)$ gives then

$$D_\lambda(z_1) \geq D_\lambda(z_2) + (z_2 - z_1)\lambda^{|K|}\,.$$

Division by $\lambda^{|K|}$ and then taking λ to 0 yields then:

$$F(z_1) \geq F(z_2) + z_2 - z_1\,,$$

The function F is thus strictly decreasing, provided that it is not constant and equal to $+\infty$ or $-\infty$.

Claim 3. By assumption, $\partial_H v_\lambda$ is uniformly bounded, so there exists $C > 0$ such that $-C \leq \partial_H v_\lambda \leq C$ for all $\lambda \in (0,1]$. By Theorem 4.3.1, one then has $D_\lambda(C) \leq 0 \leq D_\lambda(-C)$ for all λ. Division by $\lambda^{|K|}$ and then taking λ to 0 yields $F(C) \leq 0 \leq F(-C)$, so in particular F is not constant. □

4.6.4 Differentiation and Limit Operators Do Not Commute

When both the limit of the marginal discounted values and the marginal undiscounted values exist, it is natural to ask whether they are equal. In this section, first, we show this is the case in two very particular cases: (a) when $|K| = 1$; and (b), when $|I| = |J| = 1$ and $H = (\tilde{g}, 0, 0)$. Then, we show by example that these two notions differ in general.

In the sequel, consider a discounted stochastic game $\Gamma_\lambda = (K, k, I, J; g, q, \lambda)$ with $\lambda > 0$ and an admissible perturbation $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$ and set $v_{\lambda,\varepsilon} := \mathrm{val}(\Gamma_{\lambda,\varepsilon})$ for all (λ, ε) sufficiently small.

(a) Assume $|K| = 1$. The transition function is trivial and g and $\tilde{g}$ are matrix games of size $I \times J$. Moreover, $v_{\lambda,\varepsilon} = \mathrm{val}(g + \varepsilon\tilde{g})$ for all ε, so in particular it is independent from λ, and thus equal to $\lim_{\lambda\to 0} v_{\lambda,\varepsilon}$ too. By [Mil56], the marginal value is equal to $\partial_H v(\Gamma_\lambda) = \mathrm{val}_{O^*(g)}\,\tilde{g}$, and this expression is both equal to $\partial_H \lim_{\lambda\to 0} v(\Gamma_\lambda)$ and to $\lim_{\lambda\to 0} \partial_H v(\Gamma_\lambda)$.
(b) Assume $|I| = |J| = 1$ and $H = (\tilde{g}, 0, 0)$. The game $\Gamma_{\lambda,\varepsilon}$ is then a discounted Markov chain with perturbed payoffs. Indeed, the transition function can be identified with a stochastic matrix $Q \in \mathbb{R}^{K\times K}$, and the payoffs are state-dependent and given by $g_\varepsilon := g + \varepsilon\tilde{g} \in \mathbb{R}^K$. Let Π_λ be the k-th column of $\sum_{m\geq 1} \lambda(1-\lambda)^{m-1} Q^{m-1}$, and $\Pi_0 := \lim_{\lambda\to 0} \Pi_\lambda$ where the limit exists by semi-algebraicity. Then, $v_{\lambda,\varepsilon} = \Pi_\lambda g_\varepsilon$ for all $\lambda \in (0, 1]$ and $\varepsilon \geq 0$ sufficiently small, and $v_{0,\varepsilon} := \lim_{\lambda\to 0} v_{\lambda,\varepsilon} = \Pi_0 g_\varepsilon$. Hence,

$$\partial_H \lim_{\lambda\to 0} v(\Gamma_\lambda) = \lim_{\varepsilon\to 0} \frac{\Pi_0(g + \varepsilon\tilde{g}) - \Pi_0 g}{\varepsilon} = \Pi_0\tilde{g}\,.$$

On the other hand, for all fixed λ,

$$\partial_H v(\Gamma_\lambda) = \lim_{\varepsilon\to 0} \frac{v_{\lambda,\varepsilon} - v_{\lambda,0}}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{\Pi_\lambda(g + \varepsilon\tilde{g}) - \Pi_\lambda g}{\varepsilon} = \Pi_\lambda\tilde{g}\,.$$

Taking λ to 0 gives the desired result, i.e., $\lim_{\lambda\to 0} \partial_H v(\Gamma_\lambda) = \partial_H \lim_{\lambda\to 0} v(\Gamma_\lambda)$.

Minimal Counterexample Consider now the example of Section 4.3.2. For $\lambda \in [0, 1]$ and $\varepsilon \geq 0$, a direct computation yields:

$$v_{\lambda,\varepsilon} = \begin{cases} (1-\lambda)(1+\varepsilon) & \text{if } \varepsilon \geq \frac{\lambda}{1-\lambda}, \\ 1 & \text{otherwise.} \end{cases}$$

Hence, for all $\lambda > 0$, there exists $\varepsilon_0 > 0$ small enough such that $v_{\lambda,\varepsilon} = 1$ for all $\varepsilon \in [0, \varepsilon_0]$. Consequently,

$$\lim_{\lambda\to 0} \partial_H v(\Gamma_\lambda) = \lim_{\varepsilon\to 0} \frac{v_{\lambda,\varepsilon} - v_{\lambda,0}}{\varepsilon} = 0\,.$$

On the other hand, for all $\varepsilon \geq 0$, $v_{0,\varepsilon} = \lim_{\lambda\to 0} v_{\lambda,\varepsilon} = 1 + \varepsilon$, so

$$\partial_H v(\Gamma_0) = \lim_{\varepsilon\to 0} \frac{v_{0,\varepsilon} - v_{0,0}}{\varepsilon} = 1\,.$$

In other words, $\lim_{\lambda\to 0} \partial_H v(\Gamma_\lambda)$ and $\partial_H v(\Gamma_0)$ both exist and differ.

4.6.5 Proof of Proposition 4.3.3

We start by proving an analog version of Proposition 4.3.3 for the case $\lambda + \tilde{\lambda} > 0$, i.e., that there exists a square sub-matrix of $W_\varepsilon(z)$, denoted by $\overline{W}_\varepsilon(z)$, so that $P(\varepsilon, z) := \det(\overline{W}_\varepsilon(z))$

is non-zero and $P(\varepsilon, v_\varepsilon) = 0$ for all $\varepsilon \geq 0$ sufficiently small. As in Proposition 4.3.3, the novelty of this remark is not the existence of such polynomial, which follows from the theory of semi-algebraic sets (see Section 4.5.2), but the identification of a tractable set of possible candidates.

For every $\varepsilon > 0$ sufficiently small, the assumption $\lambda + \tilde{\lambda} > 0$ implies that the perturbed game $\Gamma + \varepsilon H$ is a $(\lambda + \varepsilon\tilde{\lambda})$-discounted stochastic game. Consider a Shapley-Snow kernel $\overline{W}_\varepsilon(v_\varepsilon)$ of $W_\varepsilon(v_\varepsilon)$, and let $\overline{\Delta}^k_\varepsilon$ and $\overline{\Delta}^0_\varepsilon$ be the corresponding auxiliary matrices so that $\overline{W}_\varepsilon(z) = \overline{\Delta}^k_\varepsilon - z\overline{\Delta}^0_\varepsilon$ for all z. Let us show that $P(\varepsilon, z) := \det\left(\overline{W}_\varepsilon(z)\right)$ is non-zero and satisfies $P(\varepsilon, v_\varepsilon) = 0$. For the former, we use Jacobi's formula (i.e., for all two square matrices M and N of same size, the derivative of $z \mapsto \det(M - zN)$ is $-\operatorname{tr}(\operatorname{co}(M)^\top N)$) and the definition of $W_\varepsilon(z) = \Delta^k_\varepsilon - z\Delta^0_\varepsilon$, to get:

$$\frac{\partial P}{\partial z}(\varepsilon, v_\varepsilon) = -\operatorname{tr}\left(\operatorname{co}\left(\overline{W}_\varepsilon(v_\varepsilon)\right)^\top \overline{\Delta}^0_\varepsilon\right).$$

On the other hand, by [SS50], for any Shapley-Snow Kernel $\overline{M}$ of M, all the entries of $\operatorname{co}(\overline{M})$ are of same sign, and not all 0 (see Section 4.5.1). Since all entries of $\overline{\Delta}^0_\varepsilon$ are strictly positive, as noted in Section 4.2.2, it follows that $\frac{\partial P}{\partial z}(\varepsilon, v_\varepsilon) \neq 0$ so in particular $P \not\equiv 0$. The relation $P(\varepsilon, v_\varepsilon) = 0$ follows from [AOB19, Theorem 1], which implies $\operatorname{val}(W(v_\varepsilon)) = 0$, and from the theory of Shapley and Snow (see Section 4.5.1), which implies that for any Shapley-Snow kernel $\overline{W}_\varepsilon(v_\varepsilon)$ of $W(v_\varepsilon)$ one has $\det(\overline{W}_\varepsilon(v_\varepsilon) = 0$. As ε goes to 0, the Shapley-Snow kernels may vary, as well as the finitely many polynomials defined as their determinants. Since they are finitely many, there exists $\delta > 0$ so that for all two polynomials in this set, either they are equal or they never cross in the interval $(0, \delta]$. If $\overline{W}_\delta(v_\delta)$ is a Shapley-Snow kernel of $W_\delta(v_\delta)$, then $P(\varepsilon, z) := \det(\overline{W}_\delta(z))$ is the desired polynomial satisfying $P(\varepsilon, v_\varepsilon) = 0$ for all $\varepsilon \in (0, \delta]$. Moreover, the equality at $\varepsilon = 0$ follows from continuity.

We now prove Proposition 4.3.3. In this case $\lambda = 0$, and also $\tilde{\lambda} = 0$ by the choice of $H = (\tilde{g}, \tilde{q}, 0)$. The perturbed game $\Gamma + \varepsilon H$ is thus an undiscounted stochastic game for all ε sufficiently small. Let $\varepsilon > 0$ be fixed. For all $\beta \in (0, 1]$, consider the auxiliary β-discounted stochastic game $\Gamma_{\beta,\varepsilon} := (K, k, I, J; g + \varepsilon\tilde{g}, q + \varepsilon\tilde{q}, \beta)$, whose value and auxiliary parameterized game are denoted, respectively, by $v_{\beta,\varepsilon}$ and $W_{\beta,\varepsilon}(z)$. To this game, we can apply the property that we just proved, i.e., Statement (i) at the end of Section 4.3.2. Hence, there exists $\delta > 0$ and a fixed square sub-matrix $\overline{W}_{\beta,\varepsilon}(v_{\beta,\varepsilon})$ of $W_{\beta,\varepsilon}(v_{\beta,\varepsilon})$ so that $\det(\overline{W}_{\beta,\varepsilon}(z))$ is a non-zero polynomial in the variables (β, z) and $\det(\overline{W}_{\beta,\varepsilon}(v_{\beta,\varepsilon})) = 0$ for all $\beta \in [0, \delta]$. For this fixed sub-matrix, as ε varies, $P(\varepsilon, \beta, z) := \det(\overline{W}_{\beta,\varepsilon}(z))$ defines a polynomial in (β, ε, z). Its projection $\Phi(P) \in \mathbb{R}[\varepsilon, z]$ belongs to the set $\mathcal{P}_{(\Gamma,H)}$ by the definition of this set. By definition of $\Phi(P)$, there exist $s \geq 0$ and polynomials $(R_m)_{m>s}$ in $\mathbb{R}[\varepsilon, z]$ such that $P(\varepsilon, \beta, z) = \Phi(P)(\varepsilon, z)\beta^s + \sum_{m>s} R_m(\varepsilon, z)\beta^m$. This equality holds, as well as $P(\varepsilon, \beta, v_{\beta,\varepsilon}) = 0$, hold for all $\beta \in [0, \delta]$. Hence, dividing by β^s and letting β go to zero, we have that $\Phi(P)(\varepsilon, v_{0,\varepsilon}) = 0$, where $v_{0,\varepsilon} := \lim_{\beta\to 0} v_{\beta,\varepsilon}$ exists as the limit value of a discounted stochastic game by [BK76]. Let $\varepsilon > 0$ go to 0. While the choice of the polynomial P, and thus $\Phi(P)$ varies with ε, since there are only finitely many candidates for this polynomial, there exists $\varepsilon_0 > 0$ and $\Phi(P) \in \mathcal{P}_{(\Gamma,H)}$ such that $\Phi(P)(\varepsilon, v_{0,\varepsilon}) = 0$ for all $\varepsilon \in (0, \varepsilon_0]$. Note that this relation can be extended to $\varepsilon = 0$ as soon as the map $\varepsilon \mapsto v_{0,\varepsilon}$ is continuous at 0, which gives the desired result.

4.6.6 Proof of Theorem 4.3.4

By Proposition 4.3.3, there exists a polynomial P such that, for $\varepsilon > 0$ sufficiently small, $P(\varepsilon, v_\varepsilon) = 0$. Moreover, the map $\varepsilon \mapsto v_\varepsilon$ is continuous at $\varepsilon = 0$ by assumption, so $P(0, v_0) = 0$ holds too. The desired result follows then from the implicit function theorem (see [KP13]) applied to $P \in \mathbb{R}[\varepsilon, z]$ at $(0, v_0)$, which holds as P is continuous and differentiable, and satisfies $P(0, v_0) = 0$ and $\frac{\partial}{\partial z} P(0, z)\Big|_{z=v_0} \neq 0$.

4.7 Computational Complexity of the Discounted Marginal Value

We present the computational complexity of Theorem 4.3.1. Recall that Theorem 4.3.1 gives a characterization of the marginal values of a discounted stochastic game as the unique $z \in \mathbb{R}$ such that $D(z) = 0$, where $D(z)$ is a constrained value problem. We now explain how to derive an algorithm from this characterization to compute an approximation of the marginal values to a desired level of accuracy. The algorithm is a dichotomic search on $D(z)$, which is possible because $z \mapsto D(z)$ is strictly monotonous. Consequently, its computational complexity is reduced to (1) bounds on the marginal values, and (2), the complexity of the query $D(z)$.

In the sequel, we assume that the input data are all rational numbers, that is all entries of $(g, q, \tilde{g}, \tilde{q})$ as well as λ and $\tilde{\lambda}$ are rational numbers, and their complexity refers to the size of their binary representation.

Step 1: A bound for the marginal values. By Lemma 4.5.4,

$$|\partial_H v(\Gamma)| \leq \|\tilde{g}\|_\infty + \frac{1-\lambda}{\lambda} \|\tilde{q}\|_1 \, \|g\|_\infty + 2\frac{|\tilde{\lambda}|}{\lambda} \|g\|_\infty \, .$$

We can therefore start the dichotomic search in the interval $[-C, C]$, where C is an integer upper bound of the previous expression. This choice ensures that z is a rational number during the dichotomic search, whose size is polynomial in the input.

Step 2: Computational complexity of $D(z)$. Consider the monotonic operator

$$z \mapsto D(z) = \mathrm{val}_{O^*(\Gamma)} \left(\partial_H \Delta^k - v(\Gamma)\, \partial_H \Delta^0 - z\, \Delta^0 \right) \, .$$

The computation of $D(z)$ calls for two key observations: first, $D(z)$ is the value of a matrix game where players are restricted to playing optimal strategies of another matrix game; second, the input of $D(z)$ involves rationals numbers and the algebraic number $v(\Gamma)$. To address them we use, respectively, [Mil56], who proved that a constrained value problem can be written as a linear program, and [Bel01] who provided a polynomial-time algorithm to solve linear programs involving algebraic numbers.

Step 2a: Writing $D(z)$ as a linear program. Let us start by considering a standard linear program, i.e., maximize $X^\top c$ subject to $b + AX \geq 0$ and $X \geq 0$ where $X, c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$, and its canonical representation:

$$\left[\begin{array}{c|c} 0 & c \\ \hline b & A \end{array} \right] .$$

The value of a matrix game $M \in \mathbb{R}^{I \times J}$ is the solution of the following linear program, where $\mathbf{1}_J$ and $\mathbf{1}_J$ denote column vectors of ones in $\mathbb{R}^I$ and $\mathbb{R}^J$ respectively:

$$\begin{aligned} \max_{x,z} \quad & z \\ \text{s.t.} \quad & x^\top M \geq z \mathbf{1}_J \,, \\ & x^\top \mathbf{1}_I = 1 \,, \\ & (z, x) \in \mathbb{R} \times \mathbb{R}^I_+ \,. \end{aligned}$$

Its canonical representation is then as follows:

$$\widetilde{M} := \left[\begin{array}{r|rrrcr} 0 & 1 & -1 & 0 & \dots & 0 \\ \hline 1 & 0 & 0 & -1 & \dots & -1 \\ -1 & 0 & 0 & 1 & \dots & 1 \\ 0 & -1 & 1 & & & \\ \vdots & \vdots & \vdots & & M^\top & \\ 0 & -1 & 1 & & & \end{array}\right] .$$

By [Mil56], for all matrices $M, N \in \mathbb{R}^{I \times J}$ with $\operatorname{val}(M) = 0$, the constrained value problem $\operatorname{val}_{O(M)} N$ is a linear program whose canonical representation can be easily derived from the following block matrix:

$$\left[\begin{array}{c|c} \widetilde{N} & \widetilde{M} \\ \hline \widetilde{M} & 0 \end{array}\right] .$$

The first line and column of this block matrix correspond, respectively, to the objective vector $c \in \mathbb{R}^{2(|I|+3)}$, and to the constraint vector $b \in \mathbb{R}^{2(|J|+3)}$. Explicitly, the constrained value problem $\operatorname{val}_{O(M)} N$ is the solution of the following linear program:

$$\begin{aligned} \max_{z,x,u_0,u,\sigma} \quad & z + u \\ \text{s.t.} \quad & x^\top N + \sigma^\top M \geq (z+u) \mathbf{1}_J \,, \\ & x^\top M \geq z \mathbf{1}_J \,, \\ & x^\top \mathbf{1}_I = 1 \,, \\ & \sigma^\top \mathbf{1}_I = u_0 \,, \\ & (z, x, u_0, u, \sigma) \in \mathbb{R} \times \mathbb{R}^I_+ \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^I_+ \,. \end{aligned}$$

To compute $D(z)$ is thus sufficient to set $M := \Delta^k - v(\Gamma)\Delta^0$ and $N := \partial_H \Delta^k - v(\Gamma)\, \partial_H \Delta^0 - z\, \Delta^0$, so that $D(z) = \operatorname{val}_{O(M)} N$. The matrices M and N are of equal size, and $\operatorname{val} M = 0$. Consequently, we can compute $D(z)$ by solving the previously described linear program.

Step 2b: Computational complexity of $D(z)$. We claim that the computational complexity of $D(z)$ is polynomial in the input size, notably on the size of λ, and on $|I^K|$ and $|J^K|$.

Consider an approximation of the marginal $\partial_H v(\Gamma)$ up to some additive $\delta > 0$. A dichotomic search requires computing $D(z)$ on values of z that are rational numbers whose size is polynomial in the input description and $1/\delta$. On the other hand, $D(z)$ is a linear program involving the value $v(\Gamma)$. By [OB20, Proposition 1], this is an algebraic number of degree at most $d := \min(|I^K|, |J^K|)$ and whose defining polynomial has coefficients whose size is polynomial in the input size and d. By [Bel01, Theorem 21], since all the numbers involved in the program belong to the extension of the rational numbers by the algebraic number $v(\Gamma)$, this linear program can be solved in polynomial time with respect to the input size, $1/\delta$, and its size $2(|I^K| + 3)$ and $2(|J^K| + 3)$. We thus conclude that a δ-approximation of $\partial_H v(\Gamma)$ can be obtained in polynomial time with respect to the input size, $1/\delta$, $|I^K|$ and $|J^K|$. This proves the desired result.

4.8 Regularity of the Perturbed Value Function

For completeness, we provide proofs for Proposition 4.5.7 and Proposition 4.5.8. As already mentioned, these two results follow from [FV97] (i.e., Lemma 4.5.4), [Sol03] (i.e., Lemma 4.5.5), and the theory of semialgebraic sets.

4.8.1 Proof of Proposition 4.5.7

Proof. Recall that this statement is about the continuity of the perturbed value function. Consider $\varepsilon_0 > 0$ such that the perturbation is well-defined on $[0, \varepsilon_0]$. We distinguish three cases, $\lambda > 0$, $\lambda = \tilde{\lambda} = 0$, and $\tilde{\lambda} > \lambda = 0$.

Assume $\lambda > 0$. The Lipschitz continuity of $\varepsilon \mapsto v_\varepsilon$ in the interval $[0, \varepsilon_0]$ follows directly from Lemma 4.5.4.

Assume $\lambda = \tilde{\lambda} = 0$ and $\lim_{\varepsilon \to 0} r(q, q + \varepsilon \tilde{q}) = 0$. By Lemma 4.5.5, for all $(\varepsilon, \varepsilon') \in [0, \varepsilon_0]^2$,

$$|v_{\varepsilon'} - v_\varepsilon| \leq \frac{4|K|\, r(q + \varepsilon\tilde{q},\, q + \varepsilon'\tilde{q})}{(1 - 2|K|\, r(q + \varepsilon\tilde{q},\, q + \varepsilon'\tilde{q}))_+} \|g + \varepsilon\tilde{g}\|_\infty + |\varepsilon - \varepsilon'|\, \|\tilde{g}\|_\infty \,.$$

For all fixed $\varepsilon > 0$, the map $\varepsilon' \mapsto r(q + \varepsilon\tilde{q},\, q + \varepsilon'\tilde{q})$ is continuous and equal to zero at $\varepsilon' = \varepsilon$. Therefore, $\varepsilon \mapsto v_\varepsilon$ is continuous for all $\varepsilon > 0$ sufficiently small. The continuity at $\varepsilon = 0$ follows from $\lim_{\varepsilon \to 0} r(q, q + \varepsilon\tilde{q}) = 0$.

Assume $\lambda = 0 < \tilde{\lambda}$ and $\lim_{\varepsilon \to 0} r(q, q + \varepsilon\tilde{q}) = 0$. Fix $\eta \in [0, \varepsilon_O] > 0$. For all $\varepsilon \in [\eta, \varepsilon_0]$, one can write $\Gamma + \varepsilon H = \Gamma + \eta H + (\varepsilon - \eta)H$, where the game $\Gamma + \eta H$ is a discounted stochastic game with discount rate $\varepsilon\tilde{\lambda}$. Therefore, the case $\lambda > 0$ can be applied to the game $\Gamma + \eta H$ which is perturbed by $(\varepsilon - \eta)H$. We thus deduce that $\varepsilon \mapsto v_\varepsilon$ is right-continuous at η. The left-continuity can be obtained similarly by writing $\Gamma + \varepsilon H = \Gamma + \eta H + (\varepsilon_1 - \varepsilon)(-H)$ for all $\varepsilon \in (0, \eta]$. Hence, $\varepsilon \to v_\varepsilon$ is continuous for all $\varepsilon \in (0, \varepsilon_0]$. To establish the continuity at 0, consider, for all $\varepsilon \in [0, \varepsilon_0]$,

$$v_\varepsilon - v_0 = (v_\varepsilon - w_\varepsilon) + (w_\varepsilon - v_0)\,,$$

where w_ε is the value of the game $(K, k, I, J; g, q, \varepsilon\tilde{\lambda})$. The first term, $v_\varepsilon - w_\varepsilon$, converges to 0 when ε goes to 0 by [Sol03, Theorem 6]). The second term, $w_\varepsilon - v_0$, converges to 0 when ε goes to 0 because $w_\varepsilon \to v_0$ by the theory of semi-algebraic sets, as shown by [BK76]. Consequently, $\lim_{\varepsilon \to 0+} v_\varepsilon = v_0$, which proves the desired continuity at $\varepsilon = 0$. □

4.8.2 Proof of Proposition 4.5.8

Proof. Recall that this statement is about the existence of the marginal values. Distinguish the cases $\lambda + \tilde{\lambda} > 0$ and $\lambda + \tilde{\lambda} = 0$, although the proof relies on both cases in the theory of semi-algebraic sets, namely in the fact that $\varepsilon \mapsto v_\varepsilon$ admits a Puiseux series expansion near 0. Note that the marginal value $\partial_H v(\Gamma)$ is the right-derivative of this map.

Assume $\lambda + \tilde{\lambda} > 0$. Then, $\Gamma + \varepsilon H$ is a discounted stochastic game for all $\varepsilon \geq 0$. Let $(\varepsilon, x, y, v) \in \mathbb{R}^d$ be such that $\varepsilon \in [0, \varepsilon_0]$ is the perturbation parameter, (x, y) is a pair of optimal stationary strategies in the discounted game $\Gamma + \varepsilon H$, and v is its value. These conditions define finitely many polynomial equalities and inequalities whose coefficients depend on the entries of $(g, \tilde{g}, q, \tilde{q}, \lambda, \tilde{\lambda})$. Hence, the set of such tuples is semialgebraic. By Tarksi-Seidenberg (see Section 4.5.2), the projection onto the set (ε, v) is semialgebraic, so the map

$\varepsilon \mapsto v_\varepsilon$ is a semialgebraic function. By Proposition 4.5.7 it is also continuous, so it admits a Puiseux expansion near 0. In particular, $\varepsilon \mapsto v_\varepsilon$ is right-differentiable at zero and thus the marginal value exists (possibly being unbounded). A bound on the marginal value follows from [FV97, Equation (4.19)], i.e., Lemma 4.5.4, as this result implies that, for all $\varepsilon \in (0, \varepsilon_0]$,

$$\frac{|v_\varepsilon - v_0|}{\varepsilon} \leq \|\tilde{g}\|_\infty + \frac{1-\lambda}{\lambda}\|\tilde{q}\|_1 \|g + \varepsilon\tilde{g}\|_\infty + 2\frac{|\tilde{\lambda}|}{\lambda}\|g + \varepsilon\tilde{g}\|_\infty .$$

Assume $\lambda = \tilde{\lambda} = 0$ and $\lim_{\varepsilon\to 0} r(q, q + \varepsilon\tilde{q}) = 0$. Consider the pairs $(\varepsilon, w) \in \mathbb{R}^2$ so that w is the value of $\Gamma + \varepsilon H$. These pairs can be expressed using first-order formulas over the real field, as follows: $\forall \delta > 0$, $\exists \lambda_0 > 0$, $\exists x$, $\exists v$ so that (x, y) and v are, respectively, a pair of optimal stationary strategies and the value of the λ-discounted version of $\Gamma + \varepsilon H$, and $|v - w| \leq \delta$ for all $\lambda \in (0, \lambda_0)$. These conditions require only finitely many polynomial equations. By the theory of semi-algebraic sets, the set containing such pairs is semi-algebraic, so, by Tarski-Seidenberg, the graph of the map $\varepsilon \mapsto v_\varepsilon$ is semi-algebraic. By Proposition 4.5.7, this map is also continuous so it admits a Puiseux expansion near 0. In particular, $\varepsilon \mapsto v_\varepsilon$ is right-differentiable at zero, so the marginal value exists. Set $r_\varepsilon := r(q,\, q + \varepsilon\tilde{q})$. Then, by [Sol03, Theorem 6], i.e., Lemma 4.5.5 setting $\varepsilon' = 0$, for all $\varepsilon \in [0, \varepsilon_0]$ one has

$$|v_\varepsilon - v_0| \leq \frac{4|K|r_\varepsilon}{(1 - 2|K|r_\varepsilon)_+} \|g\|_\infty + \varepsilon \|\tilde{g}\|_\infty .$$

A bound for $\varepsilon \mapsto \frac{v_\varepsilon - v_0}{\varepsilon}$ near 0 follows then from the assumption $r_\varepsilon = O(\varepsilon)$. □

4.8.3 Proof of Proposition 4.5.9

Proof. Recall that this result is about a uniform bound for the discounted marginal values. Let us show that this follows directly from [Sol03, Theorem 6]. By assumption one has $r_\varepsilon = O(\varepsilon)$. Therefore, there exists $L > 0$ such that, for all $\lambda \in (0, 1]$ and all ε sufficiently small,

$$\frac{|v_{\lambda,\varepsilon} - v_{\lambda,0}|}{\varepsilon} \leq \frac{4|K|L}{(1 - 2|K|L\varepsilon)_+} \|g\|_\infty + \|\tilde{g}\|_\infty .$$

Taking ε to 0 gives the desired result, i.e.,

$$|\partial_H v_\lambda| \leq 4|K|L + \|\tilde{g}\|_\infty, \quad \forall \lambda \in (0, 1] .$$

□

4.9 Compact-Continuous Stochastic Games

We consider discounted stochastic games over a finite state space where the sets of actions are compact metric sets, and the payoff and the transition functions are continuous. The existence of a value for compact-continuous discounted stochastic games follows from [Sha53] and [Sio58]. We extend the fromuka for the discounted marginal value, Theorem 4.3.1, to the compact-continuous case. Note that the discounted values of these games may fail to have a limit [Vig13]. Therefore, Theorem 4.3.2 and Theorem 4.3.4 can not be extended to the compact-continuous case.

4.9.1 Preliminaries

In the sequel, let $\Gamma = (K, k, I, J; g, q, \lambda)$ be a discounted compact-continuous stochastic game with value v. Since action sets I and J are compact metric spaces and $\Delta(I)$ and $\Delta(J)$ denote, respectively, the sets of probability distributions over I and J, these sets are compact and convex when endowed with the weak* topology. For any $\varepsilon \geq 0$ sufficiently small, and any perturbation $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$, let $\Gamma_\varepsilon := \Gamma + \varepsilon H$ denote the perturbed game, and let v_ε denote its value.

The auxiliary games Δ^0_ε and Δ^k_ε and $W_\varepsilon(z)$ are defined as in the finite case (see Section 4.2.2). In other words, recall that each pair of pure stationary strategies $(\mathbf{i}, \mathbf{j}) \in I^K \times J^K$ induces a Markov chain $Q_\varepsilon(\mathbf{i}, \mathbf{j}) \in \mathbb{R}^{K \times K}$ with payoffs $g_\varepsilon(\mathbf{i}, \mathbf{j}) \in \mathbb{R}^K$. Then, $\Delta^0_\varepsilon(\mathbf{i}, \mathbf{j}) := \det(\mathrm{Id} - (1-\lambda) Q_\varepsilon(\mathbf{i}, \mathbf{j}))$ and $\Delta^k_\varepsilon(\mathbf{i}, \mathbf{j})$ is the determinant of the matrix obtained by replacing the k-th column of $\mathrm{Id} - (1-\lambda) Q_\varepsilon(\mathbf{i}, \mathbf{j})$ with the vector $\lambda g_\varepsilon(\mathbf{i}, \mathbf{j})$. Finally, for each $z \in \mathbb{R}$, one sets:

$$W_\varepsilon[z](\mathbf{i}, \mathbf{j}) := \Delta^k_\varepsilon(\mathbf{i}, \mathbf{j}) - z\, \Delta^0_\varepsilon(\mathbf{i}, \mathbf{j})\,.$$

For $\varepsilon = 0$, we use the notation Δ^0, Δ^k, and $W(z)$.

Contrary to the finite case, the auxiliary games are no longer matrix games, but compact-continuous zero-sum games. By [Sio58, Theorem 3.4], they have a value (in mixed strategies) and the sets of optimal strategies are compact and convex. In particular, this is the case for $O_1(\Gamma^*)$ and $O_2(\Gamma^*)$, the set of optimal strategies of $W(v)$.

Note that the games $\partial_H \Delta^0$ and $\partial_H \Delta^k$ are well-defined, as for each $(\mathbf{i}, \mathbf{j})$, the maps $\varepsilon \mapsto \Delta^0_\varepsilon(\mathbf{i}, \mathbf{j})$ and $\varepsilon \to \Delta^k_\varepsilon(\mathbf{i}, \mathbf{j})$ are polynomials in ε. Hence, the following Taylor expansions hold:

$$\Delta^0_\varepsilon = \Delta^0 + \varepsilon\, \partial_H \Delta^0 + o(\varepsilon) \quad \text{and} \quad \Delta^k_\varepsilon = \Delta^k + \varepsilon\, \partial_H \Delta^k + o(\varepsilon)\,,$$

where, by a slight abuse in the notation, $o(\varepsilon)$ denotes a matrix of size $I^K \times J^K$ that depends on ε, satisfying $\lim_{\varepsilon \to 0} \|o(\varepsilon)\|_\infty / \varepsilon = 0$. The two appearences of $o(\varepsilon)$ do not denote the same matrix, but possibly two some matrices with the same asymptotic property.

4.9.2 Extension of Theorem 4.3.1

We prove the following result.

Theorem 4.9.1. *Consider a compact-continuous stochastic game* $\Gamma = (K, k, I, J; g, q, \lambda)$ *with* $\lambda > 0$ *and an admissible perturbation* $H = (\tilde{g}, \tilde{q}, \tilde{\lambda})$*. Then,* $\partial_H v(\Gamma)$ *exists and is the unique* $z \in \mathbb{R}$ *satisfying*

$$D(z) := \mathrm{val}_{O^*(\Gamma)} \left(\partial_H \Delta^k - v(\Gamma)\, \partial_H \Delta^0 - z\, \Delta^0 \right) = 0\,.$$

In the finite case, the existence proof of the marginal values relied on the theory of semi-algebraic sets, namely on the fact that the map $\varepsilon \mapsto v(\Gamma + \varepsilon H)$ is a Puiseux series near 0. However, this theory does not apply to the compact-continuous framework. Therefore, we describe an alternative approach not relying on the existence of the marginal value.

Proof. We first show the existence of a unique $z*$ such that $D(z*) = 0$, and then we show it is equal to the marginal value.

Existence Note that, for all two games M and N over the same action sets, the value function (unconstrained or constrained over some convex compact subset of strategies) satisfies:

- *Monotonicity*. $|\operatorname{val}(M) - \operatorname{val}(N)\| \leq \|M - N\|_\infty$.
- *Continuity*. The map $z \mapsto \operatorname{val}(M + zN)$ is continuous.

Further, like in the finite case described in Section 4.2.2, we have that $\|\Delta^0\|_\infty \geq \lambda^{|K|} > 0$ because, for every stationary action profile $(\mathbf{i}, \mathbf{j}) \in I^K \times J^K$, the matrix $Q(\mathbf{i}, \mathbf{j})$ is a Markov chain over a finite state space, and $\Delta^0(\mathbf{i}, \mathbf{j}) = \det(\operatorname{Id} - (1-\lambda) Q(\mathbf{i}, \mathbf{j})) \geq \lambda^{|K|}$. Therefore, the mapping $z \mapsto D(z)$ is strictly decreasing and bijective from $\mathbb{R}$ to $\mathbb{R}$, which proves the existence of a unique $z^* \in \mathbb{R}$ such that $D(z^*) = 0$.

For each $\varepsilon > 0$ set $t_\varepsilon := (v_\varepsilon - v)/\varepsilon$. We prove that $\lim_{\varepsilon \to 0^+} t_\varepsilon$ exists and is equal to z^* in the following steps.

Step 1 A simple algebraic rearrangement yields $v_\varepsilon \Delta^0_\varepsilon = v\Delta^0 + \varepsilon\, t_\varepsilon \Delta^0_\varepsilon + v\,(\Delta^0_\varepsilon - \Delta^0)$. Together with the Taylor expansions of Δ^0_ε and Δ^k_ε one obtains:

$$\begin{aligned}\Delta^k_\varepsilon - v_\varepsilon \Delta^0_\varepsilon &= \Delta^k + \varepsilon\, \partial_H \Delta^k + o(\varepsilon) - \left(v\,\Delta^0 + \varepsilon\, t_\varepsilon\, \Delta^0_\varepsilon + v\,(\varepsilon\, \partial_H \Delta^0 + o(\varepsilon)\right)\,, \\ &= \Delta^k - v\Delta^0 + \varepsilon\left(\partial_H \Delta^k - v\,\partial_H \Delta^0 - t_\varepsilon \Delta^0_\varepsilon\right) + o(\varepsilon)\,.\end{aligned}$$

Rearranging terms, and adding the term $-\varepsilon z^* \Delta^0$ to both sides, one thus obtains:

$$\Delta^k_\varepsilon - v_\varepsilon \Delta^0_\varepsilon + \varepsilon\left(t_\varepsilon \Delta^0_\varepsilon - z^* \Delta^0\right) + o(\varepsilon) = \Delta^k - v\Delta^0 + \varepsilon\left(\partial_H \Delta^k - v\,\partial_H \Delta^0 - z^* \Delta^0\right)\,.$$

Step 2 By the extension of [AOB19, Theorem 1] to the compact-continuous case, $\operatorname{val}(\Delta^k_\varepsilon - v_\varepsilon\, \Delta^0_\varepsilon) = 0$ for all $\varepsilon \geq 0$. Applying [RS01, Proposition 4] to the compact-continuous game $\Delta^k - v\Delta^0$ perturbed in the direction $\partial_H \Delta^k - v \partial_H \Delta^0 - z^* \Delta^0$ yields

$$\lim_{\varepsilon \to 0} \frac{\operatorname{val}\left(\Delta^k - v\Delta^0 + \varepsilon\left(\partial_H \Delta^k - v\partial_H \Delta^0 - z^* \Delta^0\right)\right)}{\varepsilon} = D(z^*)\,.$$

Step 3 By the definition of z^*, the right-hand side of the previous equation equals 0. Therefore,

$$\operatorname{val}\left(\Delta^k - v\Delta^0 + \varepsilon\left(\partial_H \Delta^k - v\partial_H \Delta^0 - z^* \Delta^0\right)\right) = o(\varepsilon)\,.$$

Together with the last equation of step 1, and the monotonicity of the value function, we have that

$$\operatorname{val}\left(\Delta^k_\varepsilon - v_\varepsilon \Delta^0_\varepsilon + \varepsilon\left(t_\varepsilon \Delta^0_\varepsilon - z^* \Delta^0\right)\right) = o(\varepsilon)\,. \tag{4.2}$$

Step 4 The asymptotic expansion $\Delta^0 = \Delta^0_\varepsilon + o(1)$ implies that

$$\begin{aligned}\varepsilon\left(t_\varepsilon \Delta^0_\varepsilon - z^* \Delta^0\right) &= \varepsilon\left(t_\varepsilon \Delta^0_\varepsilon - z^*\left(\Delta^0_\varepsilon + o(1)\right)\right) \\ &= \varepsilon\,(t_\varepsilon - z^*)\Delta^0_\varepsilon + o(\varepsilon)\,.\end{aligned}$$

Replacing in Equation (4.2), we get that

$$\operatorname{val}\left(\Delta^k_\varepsilon - v_\varepsilon \Delta^0_\varepsilon + \varepsilon\,(t_\varepsilon - z^*)\,\Delta^0_\varepsilon\right) = o(\varepsilon)\,. \tag{4.3}$$

Step 5 Lastly, $\|\Delta^0_\varepsilon\|_\infty \geq (\lambda + \varepsilon\tilde{\lambda})^{|K|}$ for any $\varepsilon \geq 0$ as long as $q + \varepsilon\tilde{q}$ is a transition function, as noted earlier in the proof (i.e., the construction of Section 4.2.2 applies to the function $(\mathbf{i}, \mathbf{j}) \in I^K \times J^K \mapsto \Delta^0_\varepsilon(\mathbf{i}, \mathbf{j}) \in \mathbb{R}$). Also, recall that $\mathrm{val}(\Delta^k_\varepsilon - v_\varepsilon \Delta^0_\varepsilon) = 0$ for all $\varepsilon \geq 0$ by Step 2. Hence,

$$\left|\mathrm{val}\left(\Delta^k_\varepsilon - v_\varepsilon \Delta^0_\varepsilon + \varepsilon\left(t_\varepsilon - z^*\right)\Delta^0_\varepsilon\right)\right| \geq \varepsilon\left|t_\varepsilon - z^*\right|(\lambda + \varepsilon\tilde{\lambda})^{|K|}.$$

In combination with Equation (4.3), one obtains that $|t_\varepsilon - z^*|\,(\lambda + \varepsilon\tilde{\lambda})^{|K|} = o(1)$. In particular, $\lim_{\varepsilon\to 0} t_\varepsilon = z^*$, which is the desired result. □

4.10 Conclusion

This work contributes to the theory of stochastic games by providing explicit formulas for the marginal values, in the discounted and the undiscounted case. We provided a formula for the marginal values of a discounted stochastic game. Under mild assumptions on the perturbation, we provided a formula for their limit as the discount rate vanishes, and for the marginal values of an undiscounted stochastic game. Lastly, we showed via an example that the two latter differ in general. On the way, some questions were raised but only partially solved. The following points, for example, deserve further research.

Complexity of Marginal Value We provided an algorithm to approximate the marginal discounted values whose runtime is polynomial in $|I^K|$ and $|J^K|$ (and the size of the input, which should be rational). When addressing the exact computation problem, the following questions must be investigated. Are the marginal discounted values algebraic? And, if so, what is a tight bound for their degree and for the size of coefficients of their minimal polynomial?

Computing the Limit Discounted Marginal Value Theorem 4.3.2 suggests a dichotomic search algorithm for the limit marginal values based on the computation of (the sign of) $F(z) = \lim_{\lambda\to 0} \lambda^{|K|} D_\lambda(z)$, where $D_\lambda(z) = \mathrm{val}_{O^*(\Gamma_\lambda)}(\partial_H \Delta^k_\lambda - v_\lambda\, \partial_H \Delta^0_\lambda - z\,\Delta^0_\lambda)$. The map $\lambda \mapsto D_\lambda(z)$ is a Puiseux series near 0 (see Claim 1 of Section 4.6.3), so there exists $\lambda_0 > 0$ such that its sign is constant in $(0, \lambda_0)$. Hence, determining an explicit lower bound for λ_0 is required to derive a bound on the computational complexity of the above-mentioned algorithm. A similar problem was solved in [OB20], but there the map $\lambda \mapsto D_\lambda(z)$ was a rational fraction near 0. Thus, additional research is required to extend this result.

Computing the Polynomial in Proposition 4.3.3 Considering Proposition 4.3.3, determining a non-zero polynomial $P \in \mathcal{P}_{(\Gamma,H)}$ such that $P(\varepsilon, v_\varepsilon) = 0$ for all ε sufficiently small may be hard in general. A practical condition was given, namely the existence of a pair of optimal stationary strategies in the auxiliary discounted game $\Gamma_{\beta,\varepsilon}$ whose support is fixed for all $\beta > 0$ and $\varepsilon \geq 0$ sufficiently small. Determining the desired polynomial efficiently in the absence of this simplifying property requires further research. Further, suppose that the desired polynomial satisfies $\frac{\partial}{\partial z}P(0,z)\Big|_{z=v_0} \neq 0$. Is it always the case that a divider p of P exists satisfying (1) $p(\varepsilon, v_\varepsilon) = 0$ for all ε sufficiently small, and (2) $\frac{\partial}{\partial z}p(0,z)\Big|_{z=v_0} \neq 0$, as was the case in the Bewley-Kohlberg example? This property would have the following two important implications. First, the formula of Theorem 4.3.4 can be applied to every stochastic game. Second, the undiscounted marginal values are algebraic of the same degree as the undiscounted values.

Derivation and Limit Operators As noted via a minimal example, the derivation and limit operators do not commute in general. However, they do commute in certain cases, beyond the ones identified. For example, they commute in the Big Match. It is thus natural to look for necessary and sufficient conditions for the two operators to commute.

CHAPTER 5

Random Zero-sum Games on Directed Graphs

This chapter is based on [ALM$^+$25], i.e., the following publication. Luc Attia, Lyuben Lichev, Dieter Mitsche, Raimundo Saona, and Bruno Ziliotto. Random Zero-Sum Dynamic Games on Infinite Directed Graphs. *Dynamic Games and Applications*, 2025.

We consider random two-player zero-sum dynamic games with perfect information on a class of infinite directed graphs. Starting from a fixed vertex, the players take turns to move a token along the edges of the graph. Every vertex is assigned a payoff known in advance by both players. Every time the token visits a vertex, Player 2 pays Player 1 the corresponding payoff. We consider a distribution over such games by assigning i.i.d. payoffs to the vertices. On the one hand, for acyclic directed graphs of bounded degree and sub-exponential expansion, we show that, when the duration of the game tends to infinity, the value converges almost surely to a constant at an exponential rate dominated in terms of the expansion. On the other hand, for the infinite d-ary tree (that does not fall into the previous class of graphs), we show convergence at a double-exponential rate.

5.1 Introduction

Dynamic Games on Graphs Zero-sum dynamic games with perfect information played on graphs [EM79] provide a powerful mathematical framework to analyze several important problems in mathematics and computer science. They correspond to stochastic games [Sha53] with the restrictions that the transitions are deterministic and players play in turns. Considering a graph where every vertex has an outgoing edge, the game starts with a token at a vertex. Then, two players, who both know the position of the token at all times, alternate in choosing to move it along an outgoing edge to a neighboring vertex. Each vertex is assigned a uniformly bounded number called *payoff* and every time the token visits a vertex, Player 2 pays Player 1 its payoff. As the game evolves, the token moves from one vertex to another indefinitely, thus generating a sequence of payoffs.

Limit Value In the n-stage game, the objective of Player 1 is to maximize the mean payoff over n stages. Similarly, Player 2 aims to minimize this mean. The *value* of the n-stage game, denoted by v_n, is the maximum mean payoff Player 1 can guarantee irrespective of the actions of Player 2. Studying the behavior of dynamic games, or in general stochastic games, when the duration of the game tends to infinity, has been the subject of intense research over the last fifty years. One question that has received particular interest is the existence of a limit of the sequence $(v_n)_{n\geq 1}$ as n tends to infinity. When it exists, this limit can be interpreted as the asymptotic mean payoff of Player 1 over an infinitely long game, and it thus stands out as a fundamental concept in the theory of stochastic games with long duration. A seminal result of Bewley and Kohlberg [BK76] states that, when the state space and the action sets are finite, the sequence $(v_n)_{n\geq 1}$ converges even for general stochastic games. This result has been extended in many directions, ranging from models with partial observation of the state and actions through models with unknown duration to models with objective functions other than mean payoff, see [LS15, LS17, Sol22, SZ16]. Proving convergence of $(v_n)_{n\geq 1}$ turns out to be a particularly delicate task when the state space is infinite. Indeed, positive results are scarce (see [LR20, Zil24] for some recent advances) and counterexamples have been found [Zil16b]. When the action space is infinite, there are positive and negative results [Gar19].

Random Dynamic Games Recently, a class of random zero-sum dynamic games on infinite graphs with vertices in $\mathbb{Z}^d$ has been introduced under the name of *percolation games* [GZ23]. In this model, as in usual dynamic games, each vertex is assigned a uniformly bounded payoff, and these values are known by the players before the game starts. The authors study a distribution over such games given by assigning i.i.d. random payoffs to the vertices. Then, the asymptotic behavior of the random value of the n-stage game, denoted by V_n, is studied. It is shown in [GZ23] that, under the assumption that payoffs are uniformly bounded and every action increases the projection of the position of the token onto some axis, $(V_n)_{n\geq 1}$ converges almost surely to a deterministic limit value. Moreover, they provide exponential concentration estimates for $(V_n)_{n\geq 1}$ around the limit of their expectation. One important takeaway message of this result is that equipping the state space with a particular graph structure (e.g., $\mathbb{Z}^d$) and assuming that payoffs have some probabilistic regularity (e.g., i.i.d. random variables) can ensure the existence of the limit of $(V_n)_{n\geq 1}$ even for an infinite graph.

Contributions We extend the results in [GZ23] to a class of graphs that is fairly more general than $\mathbb{Z}^d$. In more detail, we introduce *directed games*, a class of games on acyclic directed graphs Γ where players move the token along the edges of Γ. On the one hand, under certain assumptions of weak transitivity and sub-exponential growth of Γ, we prove that $(V_n)_{n\geq 1}$ is exponentially concentrated around a given deterministic limit value (so, in particular, $(V_n)_{n\geq 1}$ a.s. converges to that value) and relate the convergence rate to the expansion of the graph (see Theorem 5.3.1). On the other hand, we consider the infinite d-ary tree where each vertex has exactly $d \geq 2$ children and every edge is directed from the parent to the child. These graphs do not belong to the previous class of games due to their exponential expansion away from the root. In this case, we show a stronger double-exponential concentration of the random variables $(V_n)_{n\geq 1}$ around their respective expectations.

In contrast to [GZ23], the graphs that we consider here do not have to be transitive and, more importantly, they do not necessarily grow polynomially. This leads to several differences in the proof techniques. On a technical level, the novelty of our work lies in treating a significantly more general class of graphs than the class of percolation games [GZ23] while providing sharper concentration estimates. Note that the game being weakly transitive, and not simply transitive, introduces additional technicalities. Lastly, the proof of the second result on d-ary trees significantly differs from the arguments used in [GZ23]: it is based on a pruning argument to avoid certain subtrees.

Related Work In our model, the payoffs associated with the states are the only random variables, and their realizations are known to both players before the game starts. Once the payoffs are revealed, the players play a deterministic game. Distributions over games, such as in our model of random dynamic games, have recently received growing attention. For example, studying random zero-sum games on graphs and their long-term behavior goes beyond the game theory community [ARY21, ACSZ21, FPS23, HJM$^+$23]. Indeed, on the one hand, a class of random games has been used to solve a well-known open problem regarding Probabilistic Finite Automata [HMM19] (see [BKPR23] for an extension). On the other hand, the class of games studied in [GZ23] has served as a toy model for contributing to a well-studied problem in the theory of PDEs called *stochastic homogenization* (see [GZ23, Section 4] and [Zil17, DSZ24]).

Outline Section 5.2 formally defines our model. Section 5.3 presents our main contributions. Section 5.4 recalls preliminary results for later use. Then, Sections 5.5 and 5.6 are dedicated to the proofs of our main results on controlled expansion graphs and d-ary trees, respectively. Finally, Section 5.7 shows that our results cover all previously defined percolation games.

5.2 Preliminaries

We formalize the model we consider. A *directed game* is a dynamic game that consists of a locally finite directed graph Γ with infinite countable vertex set Z called the *state space*, an initial state $z_0 \in Z$ and a collection of independent and identically distributed (i.i.d.) random variables $(G_z)_{z\in Z}$ called *payoffs* defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that Γ has uniformly bounded degrees and contains neither directed cycles nor vertices with out-degree zero. The game is played by two players called Player 1 and Player 2. At the start of the game, the random payoffs $(G_z)_{z\in Z}$ are sampled and $(G_z(\omega))_{z\in Z}$ are presented to both players, who thus obtain perfect information. Then, a token is placed at the initial state z_0. For every integer $i \geq 0$, stage i proceeds as follows. Given that the token is positioned at a state $z \in Z$,

- if i is even, Player 1 moves the token to an out-neighbor z' of z in Γ, and
- if i is odd, Player 2 moves the token to an out-neighbor z' of z in Γ.

After the token has been moved from z to z', Player 1 receives the payoff $G_{z'}(\omega)$ from Player 2 and stage i ends. We are mostly interested in the n*-stage game* consisting of the first n stages for (typically large) integers n.

While the setting in [GZ23] also considers players making moves alternatively and with perfect information, one stage there consists of a move of the first and a move of the second player. This is not a fundamental difference and separating every move in a different stage is only done to enhance the clarity of the exposition.

A *strategy* of Player 1 (resp. Player 2) is a function $\sigma\colon \Omega \times \bigcup_{m\geq 0} Z^{2m+1} \to Z$ (respectively $\tau\colon \Omega \times \bigcup_{m\geq 0} Z^{2m+2} \to Z$) with the property that, for every $m \geq 0$ and $(z_0, z_1, \ldots, z_{2m+1}) \in Z^{2m+2}$, Γ contains the edge from z_{2m} to $\sigma(\omega, z_0, \ldots, z_{2m})$ (resp. from z_{2m+1} to $\tau(\omega, z_0, \ldots, z_{2m+1})$). We denote by Σ the collection of all strategies for Player 1 and by $\mathcal{T}$ the collection of all strategies for Player 2.

Given a pair of strategies $(\sigma, \tau) \in \Sigma \times \mathcal{T}$, we define inductively the *trajectory of the token* by setting $z_{2i+1} \coloneqq \sigma(\omega, z_0, \ldots, z_{2i})$ and $z_{2i+2} \coloneqq \tau(\omega, z_0, \ldots, z_{2i+1})$ for every $i \geq 0$. This allows us to define the *n-stage payoff function* $\gamma_n^{z_0} \colon \Omega \times \Sigma \times \mathcal{T} \to \mathbb{R}$ by setting

$$\gamma_n^{z_0}(\omega, \sigma, \tau) := \frac{1}{n} \sum_{i=1}^{n} G_{z_i}(\omega)\,.$$

Recall that the directed graph Γ is locally finite. Therefore, the n-stage game with initial state $z_0 = z \in Z$ is a perfect information finite game whose value $V_n \colon \Omega \times Z \to \mathbb{R}$ is defined as usual by

$$V_n(\omega, z) := \max_{\sigma\in\Sigma} \min_{\tau\in\mathcal{T}} \gamma_n^z(\omega, \sigma, \tau) = \min_{\tau\in\mathcal{T}} \max_{\sigma\in\Sigma} \gamma_n^z(\omega, \sigma, \tau)\,,$$

where the equality between maxmin and minmax is given by Kuhn's theorem [Kuh50, Theorem 1].

Moreover, we will say that a strategy $\sigma \in \Sigma$ (resp. $\tau \in \mathcal{T}$) is *optimal* for the n-stage game (starting from z) if, for all realization of the payoffs $\omega \in \Omega$, the strategy σ maximizes $\min_{\tau\in\mathcal{T}} \gamma_n^z(\omega, \cdot\,, \tau)$ over Σ (resp. if τ minimizes $\max_{\sigma\in\Sigma} \gamma_n^z(\omega, \sigma, \cdot\,)$ over $\mathcal{T}$).

A classic question in the game-theoretic literature initiated by [BK76] is to ask for the convergence of the n-stage value as n grows to infinity. Since payoffs are random, V_n is a random variable. Therefore, we are interested in whether the sequence $(V_n)_{n\geq 1}$ converges a.s. to a constant. If no further assumptions are imposed, $(V_n)_{n\geq 1}$ does not necessarily converge, as the following example shows.

Example 5.2.1. For all integers $m \geq 0$, set $n_m := 2^{2^{2m}}$ and $n'_m := 2^{2^{2m+1}}$ and consider the case where Γ is a directed tree (all edges being directed away from the root) where each node of even height has only one child, while each node with odd height k has two children if $k = 1$ or $k \in [n_m, n'_m)$ for some $m \geq 0$, and it has only one child if $k \in [n'_m, n_{m+1})$. Moreover, let the payoffs be i.i.d. Bernoulli random variables with parameter $1/2$. In particular, for every $m \geq 1$, in the n_m-stage game, Player 2 has only one choice most of the time, while in the n'_m-stage game, she has two choices most of the time. Since Player 2 can pick a vertex with payoff 0 (if it is present), and pick an available vertex otherwise, one can show that a.s.

$$\limsup_{m\to\infty} V_{n'_m} \leq \frac{3}{8} < \frac{1}{2} = \lim_{m\to\infty} V_{n_m}\,.$$

Indeed, while Player 1 never has a choice in the n'_m-stage game (implying that the mean payoff over the odd states visited by the token a.s. converges to $1/2$), Player 2 can ensure with the above strategy that the mean payoff over the even states visited by the token a.s. converges to $1/4$, which yields that a.s. $\limsup_{m\to\infty} V_{n'_m} \leq 3/8$. At the same time, for every $\varepsilon > 0$, Chernoff's bound for the Binomial distribution $\mathrm{Bin}(n_m, 1/2)$ and a union bound over the $O(2^{n'_{m-1}})$ vertices at level n_m in Γ shows that V_{n_m} is in the interval $[1/2 - \varepsilon, 1/2 + \varepsilon]$ with probability very close to 1. In particular, a.s. $(V_n)_{n\geq 1}$ does not converge. Therefore, to ensure convergence, we will need further structural assumptions on the graph.

Before turning to our results, we provide some vocabulary. Given a vertex $z \in Z$, a *descendant* of z (in Γ) is a vertex that can be reached from z by a directed path in Γ. We say that z and z' are *equivalent* if the two subgraphs of Γ induced by the descendants of z and by the descendants of z', respectively, are isomorphic (as directed graphs).

Definition 5.2.2. The graph Γ is *weakly transitive* if there is a state z^* and an integer $M \geq 0$ such that the following holds: for each state $z \in Z$, in the game with initial state $z_0 = z$, each player has a strategy that, independently of the moves of the opponent, ensures that the token is placed at a state equivalent to z^* after an even number of $\ell \leq M$ stages.

Note that all vertex-transitive graphs are weakly transitive with $M = 0$. We also remark that the parity of ℓ is important to ensure that, when the token reaches z^*, it is Player 1's turn to make a move. For similar notions of generalized vertex-transitive graphs, see for example [BLC$^+$18].

In the following sections, we assume that Γ is weakly transitive and drop the dependence of the random variables on ω. The next two subsections present two types of directed games used in our main results.

Weakly Transitive Games with Sub-Exponential Expansion

One of the main difficulties in the analysis of the asymptotic behavior of $(V_n)_{n\geq 1}$ is to make use of the independence of the payoffs $(G_z)_z$. To this end, we use the partial order introduced by the directed graph Γ on the state space. More formally, let $z_1, z_2 \in Z$. We say that $z_1 \preceq z_2$ if z_2 is a descendant of z_1. Then, for all $z \in Z$ and $n \geq 1$, denote $Z_n(z)$ the set of *reachable states* from z in exactly n steps, and $Z^{(n)}(z)$ the ones reachable in at most n steps. Since Γ is locally finite, $Z^{(n)}(z)$ is a finite partially ordered set. Denote $h(n, z)$ its height, that is, the longest path in $Z^{(n)}(z)$. Note that, starting from z, after $h(n, z)$ stages, the game never returns to a state in $Z^{(n)}(z)$ as Γ is acyclic, thus separating $Z^{(n)}(z)$ from the future of the game. Hence, $h(n, z)$ can be considered as the waiting time after which the future of the game becomes independent of $Z^{(n)}(z)$. In this regard, we define the largest possible waiting time, which we call the *transient speed function*, as

$$h\colon n \in \mathbb{N} \mapsto \sup_{z\in Z} h(n, z) \in \mathbb{N} \cup \{\infty\}\,.$$

We note that Mirsky's theorem (see [Mir71, Theorem 2]) gives us a dual interpretation of $h(n, z)$ as the minimum number of antichains needed to partition $Z^{(n)}(z)$. Thus, $h(n)$ may be seen as an upper bound on these minima. Following this point of view, we expect to obtain a concentration inequality parameterized by $h(n)$ since the payoffs are i.i.d. In our analysis, this concentration has to be sufficiently strong to overcome a union bound over all possible states from which the game may continue, which is quantified by the growth of $Z^{(n)}(z)$. Formally, we will handle concentration inequalities bounding from above $\mathbb{P}(|X - \mathbb{E}[X]| \geq t)$ for suitable random variables X, so we define the following key function $\psi\colon \mathbb{N} \times (0, \infty) \to \mathbb{R}$ by

$$\psi(n, t) \coloneqq \exp\left(-\frac{t^2 n^2}{2h(n)}\right) \max_{z\in Z} |Z^{(2n)}(z)|\,.$$

Lemma 5.5.1 later essentially proves that $\mathbb{P}(|V_n(z_0) - \mathbb{E}[V_n(z_0)]| \geq t) \leq 2\exp\left(-\frac{t^2n^2}{2h(n)}\right)$. When proving the convergence of the value, we need to consider every reachable state after at most $2n$ steps, which leads to an expression of the form of $\psi(n, \varepsilon_n)$.

Our main goal is to analyze directed games in which the size of the sets $Z^{(n)}(z)$ does not grow too fast as n grows to infinity, which we formalize as follows.

Definition 5.2.3 (δ-transient games)**.** For a fixed $\delta > 0$, a directed game on a graph Γ with vertex set Z is called δ-*transient* if there exists a sequence $(\varepsilon_n)_{n\geq 1}$ of positive real numbers such that

$$\varepsilon_n + \psi(n, \varepsilon_n) = O(n^{-\delta})\,,$$

where the asymptotic notation is with respect to $n \to \infty$. In other words, there is a sequence $(\varepsilon_n)_{n\geq 1}$ that decreases to zero at least as fast as $n^{-\delta}$ such that the expression $\psi(n, \varepsilon_n)$ also decreases to zero at least that fast. Recalling that $\psi(n, \varepsilon_n)$ will serve as an upper bound for expressions of the form $\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon_n)$, this speed of convergence quantifies the fact that the concentration is strong enough to make the gap ε_n small as long as n is big enough.

Remark 5.2.4. The concept of δ-transient games is only relevant for $\delta \in (0, 1/2)$ because there is no directed game that is δ-transient for some $\delta \geq 1/2$. Indeed, consider a δ-transient game. By Definition 5.2.3, we have that $(\psi(n, \varepsilon_n))_{n \geq 1}$ converges to zero. In particular, we have that $\varepsilon_n^2 n^2/h(n) \to \infty$. Since $h(n) \geq n$, this implies that $\varepsilon_n^2 n \to \infty$. Since $\psi(n, \varepsilon_n) \geq 0$, we have that $(\varepsilon_n + \psi(n, \varepsilon_n))/n^{-1/2} \to \infty$, i.e., the game is not $1/2$-transient. In conclusion, if a game is δ-transient, then $\delta < 1/2$.

Remark 5.2.5. A sufficient condition under which a directed game is δ-transient is the following: there exist real numbers $\alpha \in [0, 2 - 2\delta)$ and $\beta \in [0, 2 - 2\delta - \alpha)$ such that $h(n) = O(n^\alpha)$ and $\max_{z \in Z} |Z^{(n)}(z)| = \exp(O(n^\beta))$.

Note that the definition of a δ-transient game is independent of the payoffs and only makes assumptions on the state space. In Section 5.2, we give a few examples of δ-transient games.

Oriented Games Oriented games were introduced by Garnier and Ziliotto [GZ23] and are defined as follows. Fix an integer $d \geq 1$, and denote by e_i the d-dimensional vector with 1 in coordinate i and 0 in all other $d - 1$ coordinates. Given positive integers $n_1, \dots, n_d \geq 1$, a directed graph Γ with vertex set $Z \subseteq \mathbb{Z}^d$ is called $(n_1, \dots, n_d)$-*invariant* (or simply *invariant*) if, for every $i \in [1, d]$, the translation at vector $n_i e_i$ is a graph isomorphism for Γ. A directed weakly transitive game is called *oriented* if its underlying graph Γ is invariant and there exists $u \in \mathbb{R}^d \setminus \{0\}$ such that, for every directed edge (z, w) in Γ, we have $(w - z) \cdot u > 0$ (here, $\cdot$ denotes the usual scalar product of vectors in $\mathbb{R}^d$). We show the following proposition by providing a simple explicit construction.

Proposition 5.2.6. *Every oriented game is δ-transient for all $\delta \in (0, 1/2)$.*

The following two classes of games present particular examples of oriented games.

Example 5.2.7 (Games on tilings, see Figure 5.1)**.** A *tiling* is a periodic partition of the plane into translations of one or several polygonal shapes, called *tiles*, with vertices in $\mathbb{Z}^2$. Tilings naturally define planar graphs whose vertex set coincides with the corners of the tiles and two vertices are connected by an edge if these can be connected by following the boundary of a tile without meeting another vertex on the way. Equipping the edges of this graph with suitable orientations defines an oriented game.

Example 5.2.8 (Games on directed chains of graphs)**.** Fix a finite vertex-transitive graph H with vertex set $V(H)$ and edge set $E(H)$, and a bi-infinite sequence of copies $(H_i)_{i \in \mathbb{Z}}$ of H. For every $i \in \mathbb{Z}$ and $u \in V(H)$, denote by u_i the vertex in H_i corresponding to u. We call an H-*chain* the graph Γ_H with vertices $\bigcup_{i \in \mathbb{Z}} V(H_i)$ and edges $\{u_i v_{i+1} : i \in \mathbb{Z}, uv \in E(H)\}$.

Games on H-chains can be seen as instances of oriented games on $\mathbb{Z}$. Indeed, fixing $h = |V(H)|$, one may identify the vertices of H_i with the integers in the interval $[ih + 1, (i + 1)h]$ for all $i \in \mathbb{Z}$ in a translation-invariant way.

Weakly Transitive Games with Controlled Expansion In the following paragraphs, we show that one can construct a tree T with an arbitrary growth that is faster than linear but slower than exponential. In particular, we have the following result.

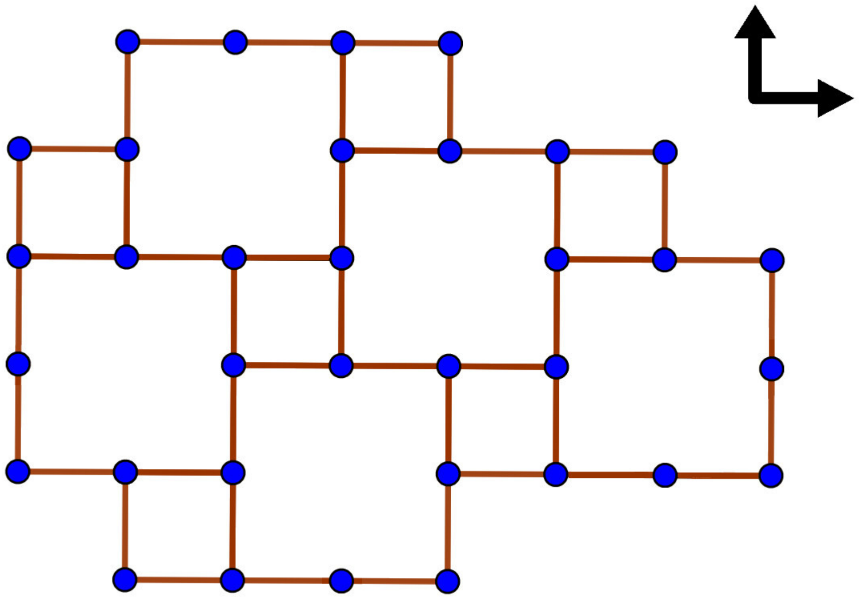


Figure 5.1: Part of a tiling with two types of square tiles. The vertices and the edges of the planar graph originating from the tiling are depicted in blue and red, respectively. Each horizontal edge is oriented from left to right and every vertical edge is oriented from bottom to top. One may choose z^* to be the bottom left vertex of a small square and $M = 6$.

Proposition 5.2.9. *For every $\delta \in (0, 1/2)$, there exist games that are δ-transient but are not $\delta'-$transient for every $\delta' > \delta$.*

Proof. Fix an arbitrary infinite rooted tree T with root r and a family of vertex-disjoint infinite paths $(P_v)_{v\in V(T)}$ where the path P_v intersects T at a unique vertex v. Define $\Gamma = T \cup (\bigcup_{v\in V(T)} P_v)$ as the tree rooted in r and with all edges oriented away from r. Thus, Γ is a directed rooted tree larger than T. Since a single move of each player is sufficient to place the token at the second vertex of some infinite path among $(P_v)_{v\in V(T)}$, the game is weakly transitive.

Recall that $Z_n(z)$ consists of all vertices at distance n from z. Therefore, the sequence $(Z_n(r))_{n\geq 1}$ is a partition of the vertex set Z of Γ and $Z_n(r)$ can only be visited once. Since $Z^{(n)}(r)$ can be partitioned into n antichains $Z_1, Z_2, \ldots, Z_n$, we have that $h(n, r) \leq n$.

Let us show that we can control the growth speed of $\max_{z\in Z} |Z^{(2n)}(z)|$. Consider a set of non-negative integers $L = \{\ell_i : i \geq 1\}$ with $\ell_1 < \ell_2 < \ldots$ and let every vertex of T in level ℓ have two children if $\ell \in L$ and one child otherwise. Moreover, suppose that $\ell_1 = 0$ and $(\ell_i - \ell_{i-1})_{i\geq 1}$ is a non-decreasing sequence. Then, one can readily check that, for every $n \geq 1$, $\max_{z\in Z} |Z^{(n)}(z)| = |Z^{(n)}(r)|$, and $h(n, z) = h(n, r) = n$. Indeed, for every $k, n \geq 1$ and a vertex $z \in Z$ on level k, using the assumptions that $\ell_1 = 0$ and $(\ell_i - \ell_{i-1})_{i\geq 1}$ is a non-decreasing sequence, we get

$$|Z^{(n)}(z) \setminus Z^{(n-1)}(z)| = 2^{|L\cap\{k,\ldots,k+n-1\}|} \leq 2^{|L\cap\{0,\ldots,n-1\}|} = |Z^{(n)}(r) \setminus Z^{(n-1)}(r)|.$$

Thus, for every integer $n \geq 0$, $|Z^{(n)}(r)| = 1 + \sum_{i=0}^{n-1} 2^{|L\cap\{0,\ldots,i\}|}$. □

Directed Games on d-ary Trees

We turn our attention to a natural example of a directed game where the set of reachable states after n steps grows exponentially with n. Note that, for all $\delta > 0$, it is not a δ-transient game. Fix an integer $d \geq 2$ and let T be an *infinite d-ary tree*, that is, a tree where every vertex has d children, with vertex set Z where every edge is directed from the parent to the child. We fix an arbitrary initial vertex z_0 and, for every integer $i \geq 0$, we define Z_i to be the set of vertices in Z that can be reached from z_0 by exactly i steps and also denote $Z_{\text{even}} := \bigcup_{i\geq 0} Z_{2i}$ and $Z_{\text{odd}} := \bigcup_{i\geq 0} Z_{2i+1}$. Note that, for every $n \geq 1$, the random variables $(V_n(z))_{z\in Z}$ have the same distribution. Hence, in this setting, we often omit the dependence of V_n in z.

5.3 Main Contributions

Our first main result shows sharp concentration for the n-stage value of δ-transient games around a deterministic constant.

Theorem 5.3.1. *Fix $\delta \in (0, 1/2)$. Consider a δ-transient directed game with transient speed h and i.i.d. payoffs $(G_z)_{z\in Z}$ supported on the interval $[0, 1]$. Then, there exist constants $v_\infty \in [0, 1]$ and $K > 0$ such that, for all $n \geq 1$, $t \geq 0$, and $z \in Z$,*

$$\mathbb{P}\left(|V_n(z) - v_\infty| \geq t + Kn^{-\delta}\right) \leq 2\exp\left(-\frac{t^2n^2}{2h(n)}\right) .$$

Consequently, for all $z \in Z$, $(V_n(z))_{n\geq 1}$ converges almost surely to v_∞.

Our second main result shows that the n-stage value of the directed game on a d-ary tree is tightly concentrated around a constant.

Theorem 5.3.2. *Fix an integer $d \geq 2$. Consider a directed game on the d-ary tree with i.i.d. payoffs supported on the interval $[0, 1]$. Then, there exists a constant $v_\infty \in [0, 1]$ such that, for every $\delta \in (0, 1/2)$, there exists $K > 0$ such that, for all $n \geq 1$, $t \geq 0$, and $z \in Z$,*

$$\mathbb{P}(|V_n(z) - v_\infty| \geq t + 2t^2 + Kn^{-\delta}) \leq \exp\left(-\frac{1}{6}\exp\left(\frac{t^2n}{4}\right)\right) .$$

Consequently, for all $z \in Z$, $(V_n(z))_{n\geq 1}$ converges almost surely to v_∞.

On a high level, the proofs of both theorems contain two main steps.

1. The first step relies on concentration arguments showing that V_n is close to $\mathbb{E}[V_n]$ with high probability. While standard concentration tools are sufficient for our proof of Theorem 5.3.1, the stronger probabilistic bound in Theorem 5.3.2 requires an additional boosting obtained by dividing the first n levels of the d-ary tree into two groups of consecutive levels and treating the n-stage game as two consecutive games on k and $n - k$ stages, respectively.

2. The second step uses the structure of the underlying graph to show that $(\mathbb{E}[V_n])_{n\geq 1}$ satisfies a certain subadditivity assumption, which allows us to conclude that $(\mathbb{E}[V_n])_{n\geq 1}$ converges to a constant v_∞, and moreover, $|\mathbb{E}[V_n] - v_\infty|$ is polynomially small.

Perspectives The proofs of Theorems 5.3.1 and 5.3.2 have a similar structure but use different arguments. A challenging research question would be to prove convergence of $(V_n)_{n\geq 1}$ and concentration bounds in all weakly transitive directed games, irrespective of the expansion speed of the underlying graph, thus unifying Theorems 5.3.1 and 5.3.2.

5.4 Previous Results

In our proofs, we make use of the well-known *bounded difference inequality*, also known as *McDiarmid's inequality.*

Lemma 5.4.1 ([JLR00, Corollary 2.27]). *Fix a function $f\colon \Lambda_1 \times \cdots \times \Lambda_N \to \mathbb{R}$ and let $Y_1, \dots, Y_N$ be independent random variables taking values in $\Lambda_1, \dots, \Lambda_N$, respectively. Suppose that there are positive constants $c_1, \dots, c_N$ such that, for every two vectors $z, w \in \Lambda_1 \times \cdots \times \Lambda_N$ that differ only in the k-th coordinate, we have $|f(z) - f(w)| \leq c_k$. Then, for every $t \geq 0$, the random variable $X = f(Y_1, \dots, Y_N)$ satisfies*

$$\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^N c_i^2}\right).$$

$$\mathbb{P}(X - \mathbb{E}[X] \leq -t) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^N c_i^2}\right).$$

We also use the following result that states convergence of almost subadditive sequences.

Lemma 5.4.2 ([BE52, Theorem 23]). *Fix an increasing function $\phi\colon \mathbb{N} \to (0, \infty)$ such that the sum of $(\phi(n)/n^2)_{n\geq 1}$ is finite, and a function $f\colon \mathbb{N} \to \mathbb{R}$ such that, for all $n \in \mathbb{N}$ and all integers $m \in [n/2, 2n]$, $f(n+m) \leq f(n) + f(m) + \phi(n+m)$. Then, there exists $\ell \in \mathbb{R} \cup \{-\infty\}$ such that*

$$\left(\frac{f(n)}{n}\right) \xrightarrow[n\to\infty]{} \ell\,.$$

5.5 δ-Transient Games

We present the proof of Theorem 5.3.1. To begin with, by using a suitable partition of the state space into subsets that are visited at most once, we show that the value of the n-stage game is well concentrated around its expected value. Note that the next lemma holds for weakly transitive games in general and will be reused in Section 5.6.

Lemma 5.5.1. *For every $z_0 \in Z$, $n \geq 1$, and $t \geq 0$,*

$$\mathbb{P}(V_n(z_0) - \mathbb{E}[V_n(z_0)] \geq t) \leq \exp\left(-\frac{t^2 n^2}{2h(n)}\right),$$

$$\mathbb{P}(V_n(z_0) - \mathbb{E}[V_n(z_0)] \leq -t) \leq \exp\left(-\frac{t^2 n^2}{2h(n)}\right).$$

Proof. Let us fix $z_0 \in Z$ and abbreviate $V_n = V_n(z_0)$, $Z_n = Z_n(z_0)$ and $Z^{(n)} = Z^{(n)}(z_0)$. Define the (random) vectors $X_k = (G_z)_{z\in Z_k} \in [0,1]^{|Z_k(z_0)|}$. Then, since $Z^{(n)} \subseteq Z_{[h(n)]}$, V_n can be written as $f(X_1, \dots, X_{h(n)})$ for some function $f\colon [0,1]^{|Z_1|} \times \cdots \times [0,1]^{|Z_{h(n)}|} \to \mathbb{R}$.

Moreover, for every integer $k \in [1, h(n)]$, the token visits the set Z_k at most once and therefore, for every pair of strategies $(\sigma, \tau) \in \Sigma \times \mathcal{T}$, $\gamma_n^{z_0}(\sigma, \tau)$ varies by at most $1/n$ as a function of X_k. Hence, for every choice of vectors $(x_i)_{i=1}^{h(n)} \in [0,1]^{|Z_1|} \times \cdots \times [0,1]^{|Z_{h(n)}|}$ and $x'_k \in [0,1]^{|Z_k|}$,

$$|f(x_1, \dots, x_k, \dots, x_{h(n)}) - f(x_1, \dots, x'_k, \dots, x_{h(n)})| \leq \frac{1}{n}.$$

Lemma 5.4.1 applied to V_n finishes the proof. □

In the remainder of the proof, we show that the expected n-stage value converges to a constant polynomially fast. Next, we show how to control the difference of the values of games of different lengths.

Lemma 5.5.2. *Fix integers $n \geq 1$ and $k \in [1, n]$. Then, for every $z_0 \in Z$, $|V_n(z_0) - V_{n-k}(z_0)| \leq k/n$.*

Proof. Observe that $nV_n(z_0) \geq (n-k)V_{n-k}(z_0)$ and $nV_n(z_0) \leq (n-k)V_{n-k}(z_0) + k$. Indeed, in the n-stage game, Player 1 (respectively Player 2) may first play according to an optimal strategy for the $(n-k)$-stage game, and play arbitrarily during the last k stages. Hence, $|n(V_n(z_0) - V_{n-k}(z_0))| \leq \max(kV_{n-k}(z_0), k - kV_{n-k}(z_0)) \leq k$, which implies the statement of the lemma. □

The next lemma shows that starting from different initial states changes the expected n-stage value only slightly when n is large.

Lemma 5.5.3. *Consider a graph Γ and denote by z^* a state by which Γ is weakly transitive. For every initial state $z_0 \in Z$, we have that $|\mathbb{E}[V_n(z_0)] - \mathbb{E}[V_n(z^*)]| = O(n^{-\delta})$.*

The main ideas of the proof are as follows: Given states z, z^*, the assumption of weak transitivity allows us to bound from below (respectively from above) $V_n(z)$ using the minimum (respectively the maximum) of the n-stage values over the nearby states equivalent to z^*. Moreover, by δ-transience, the graph Γ does not expand too quickly, which implies that, by a union bound, the expectation of the minimum (respectively the maximum) of the said n-stage values is approximately equal to the minimum (respectively the maximum) of their expectations.

Proof of Lemma 5.5.3. Fix an initial state $z_0 \in Z$ and denote by E the set of states in $Z^{(M)}(z_0)$ that are equivalent to z^*. By Definition 5.2.2, $E \neq \emptyset$ and, independently of the moves of Player 2, Player 1 can ensure that the token is at a state in E after an even number of $\ell \leq M$ stages. Hence, using Lemma 5.5.2, we have

$$nV_n(z_0) \geq (n-M) \min_{z \in E} V_{n-M}(z) \geq (n-M) \min_{z \in E}(V_n(z) - M/n) \geq \min_{z \in E} nV_n(z) - 2M\,. \quad (5.1)$$

Now, we bound from below the expectation of the right-hand side. Combining Lemma 5.5.1 and the choice of ε_n from Definition 5.2.3, we have

$$\begin{aligned} \mathbb{E}\left[\min_{z \in E} V_n(z)\right] &\geq (\mathbb{E}[V_n(z^*)] - \varepsilon_n)(1 - \mathbb{P}(\exists z \in E : V_n(z) \leq \mathbb{E}[V_n(z)] - \varepsilon_n)) \\ &\geq (\mathbb{E}[V_n(z^*)] - \varepsilon_n)\left(1 - \exp\left(-\frac{\varepsilon_n^2 n^2}{2h(n)}\right) \max_{z \in Z} |Z^{(M)}(z)|\right) \\ &\geq (\mathbb{E}[V_n(z^*)] - \varepsilon_n)(1 - \psi(n, \varepsilon_n)) = \mathbb{E}[V_n(z^*)] - O(n^{-\delta})\,, \end{aligned} \quad (5.2)$$

where the second inequality comes from a union bound, the third one uses that $Z^{(M)}(z) \subseteq Z^{(2n)}(z)$ for every $z \in Z$, and the equality is implied by the fact that $\varepsilon_n + \psi(n, \varepsilon_n) = O(n^{-\delta})$. Thus, taking expectations on both sides of (5.1) and using (5.2) shows that

$$\mathbb{E}[V_n(z_0)] \geq \mathbb{E}[V_n(z^*)] - O(n^{-\delta} + 2M/n) = \mathbb{E}[V_n(z^*)] - O(n^{-\delta})\,. \tag{5.3}$$

Similarly, Player 2 can ensure that the token reaches a state in E after an even number of $\ell \leq M$ stages. Hence,

$$nV_n(z_0) \leq (n - M) \max_{z \in E} V_{n-M}(z) + M \leq \max_{z \in E} nV_n(z) + M\,. \tag{5.4}$$

At the same time, similarly to (5.2), $\mathbb{E}\left[\max_{z \in E} V_n(z)\right]$ is bounded from above by

$$(\mathbb{E}[V_n(z^*)] + \varepsilon_n)(1 - \mathbb{P}(\exists z \in E : V_n(z) \geq \mathbb{E}[V_n(z)] + \varepsilon_n)) \\ + \mathbb{P}(\exists z \in E : V_n(z) \geq \mathbb{E}[V_n(z)] + \varepsilon_n)\,,$$

which is at most $\mathbb{E}[V_n(z^*)] + (\varepsilon_n + \psi(n, \varepsilon_n)) = \mathbb{E}[V_n(z^*)] + O(n^{-\delta})$. Combining this with (5.4) shows that $\mathbb{E}[V_n(z_0)] \leq \mathbb{E}[V_n(z^*)] + O(n^{-\delta})$, and, together with the lower bound in (5.3), this finishes the proof. □

Next, we show that the expected value of V_n converges as $n \to \infty$.

Lemma 5.5.4. *There is a constant v_∞ such that, $|\mathbb{E}[V_n(z)] - v_\infty| = O(n^{-\delta})$, for every $z \in Z$.*

As before, we start with an outline of the proof. We first use Lemma 5.5.3 to show that the initial state of the game is not important. Then, we justify that the sequence $(-n\mathbb{E}[V_n(z)])_{n \geq 1}$ is approximately subadditive, and use Lemma 5.4.2. The subadditivity is based on the simple observation that, for every player, playing optimally in the $(m + n)$-stage game is better than playing optimally in the initial m-stage game (based only on the m-th neighborhood of the initial state), and then do the same in the subsequent n-stage game.

Proof of Lemma 5.5.4. By Lemma 5.5.3, it is sufficient to show the lemma when $z_0 = z^*$; we abbreviate $V_n = V_n(z^*)$ and $Z^{(n)} = Z^{(n)}(z^*)$ for convenience. First, we show that $\mathbb{E}[V_n]$ converges to a limit $v_\infty \in \mathbb{R}$ as $n \to \infty$. By Lemma 5.5.1 and a union bound, for all $t \geq 0$,

$$\begin{aligned}\mathbb{P}(\exists z \in Z^{(2n)} : |V_n(z) - \mathbb{E}[V_n(z)]| \geq t) &\leq \sum_{z \in Z^{(2n)}} \mathbb{P}(|V_n(z) - \mathbb{E}[V_n(z)]| \geq t) \\ &\leq 2 \exp\left(-\frac{t^2 n^2}{2h(n)}\right) \max_{z \in Z} |Z^{(2n)}(z)| = 2\psi(n, t)\end{aligned} \tag{5.5}$$

where the factor of 2 in the last inequality appears since we bound both the upper and the lower tail of $V_n(z)$.

By definition of δ-transient game, there exists $(\varepsilon_n)_{n \geq 1}$ such that $\varepsilon_n + \psi(n, \varepsilon_n) = O(n^{-\delta})$. Denote by E the set of vertices in $Z^{(2n)}$ that are equivalent to z^*. Now, Lemma 5.5.3 implies that there is a constant $K' > 0$ such that, for every $z \in Z$ and $n \geq 1$, $|\mathbb{E}[V_n(z)] - \mathbb{E}[V_n]| \leq K' n^{-\delta}$. Combining this with (5.5), we get that

$$\begin{aligned}\mathbb{P}\left(\min_{z \in Z^{(2n)}} V_n(z) \leq \mathbb{E}[V_n] - \varepsilon_n - K' n^{-\delta}\right) &\leq \mathbb{P}\left(\min_{z \in Z^{(2n)}} |V_n(z) - \mathbb{E}[V_n(z)]| \geq \varepsilon_n\right) \\ &\leq \mathbb{P}(\exists z \in E, |V_n(z) - \mathbb{E}[V_n(z)]| \geq \varepsilon_n) \\ &\leq 2\psi(n, \varepsilon_n) = O(n^{-\delta})\,.\end{aligned}$$

In particular, it follows directly that

$$\begin{aligned}\mathbb{E}\left[\min_{z\in Z^{(2n)}} V_n(z)\right] &\geq \left(\mathbb{E}[V_n]-\varepsilon_n-K'n^{-\delta}\right)\mathbb{P}\left(\min_{z\in Z^{(2n)}} V_n(z)\geq \mathbb{E}[V_n]-\varepsilon_n-K'n^{-\delta}\right)\\ &\geq \left(\mathbb{E}[V_n]-\varepsilon_n-K'n^{-\delta}\right)(1-2\psi(n,\varepsilon_n))\\ &\geq \mathbb{E}[V_n]-2(\psi(n,\varepsilon_n)+\varepsilon_n)-K'n^{-\delta}\,.\end{aligned}$$

Now, fix an integer $m\in[1,2n]$ and consider the $(m+n)$-stage game. Suppose that Player 1 plays according to an optimal strategy for the m-stage game up to stage m and, once the m-stage game terminates at a state z_m, continues to play according to an optimal strategy for the subsequent n-stage game. Note that $z_m\in Z^{(2n)}$, so the above strategy of Player 1 for the first $m+n$ steps guarantees a gain of $\frac{m}{m+n}V_m+\frac{n}{m+n}\min_{z\in Z^{(2n)}}V_n(z)$. Thus,

$$\begin{aligned}(m+n)\mathbb{E}[V_{m+n}] &\geq m\mathbb{E}[V_m]+n\mathbb{E}\left[\min_{z\in Z^{(2n)}} V_n(z)\right]\\ &\geq m\mathbb{E}[V_m]+n\mathbb{E}[V_n]-2n(\psi(n,\varepsilon_n)+\varepsilon_n)-K'n^{1-\delta}\,.\end{aligned}\tag{5.6}$$

Since $\psi(n,\varepsilon_n)+\varepsilon_n=O(n^{-\delta})$, there is a constant $K''>0$ such that, for all $n\geq 1$,

$$2n(\psi(n,\varepsilon_n)+\varepsilon_n)+K'n^{1-\delta}\leq 2K''n^{1-\delta}\,.$$

Thus, using Lemma 5.4.2 with $f\colon n\mapsto -n\mathbb{E}[V_n]$ and $\phi\colon n\mapsto 2K''n^{1-\delta}$ (note that ϕ is increasing and $\sum_{n\geq 1}\phi(n)/n^2=2K''\sum_{n\geq 1}1/n^{1+\delta}<\infty$) implies that $\mathbb{E}[V_n]$ converges to a limit $v_\infty\in\mathbb{R}\cup\{\infty\}$ as $n\to\infty$. Note that v_∞ is in $[0,1]$ since this is the support of all payoff variables.

Finally, using (5.6) with $m=n$, for every $n\geq 1$, we have that

$$\mathbb{E}[V_{2n}]\geq \mathbb{E}[V_n]-(\psi(n,\varepsilon_n)+\varepsilon_n)-\frac{K'n^{-\delta}}{2}\geq \mathbb{E}[V_n]-K''n^{-\delta}\,.$$

In particular, for all integers $\ell,n\geq 1$, iterating the above observation for n taking values $n,2n,\dots,2^{\ell-1}n$ gives that

$$\mathbb{E}[V_{2^\ell n}]\geq \mathbb{E}[V_n]-K''n^{-\delta}\sum_{j=0}^{\ell-1}2^{-\delta j}\geq \mathbb{E}[V_n]-\frac{K''}{1-2^{-\delta}}n^{-\delta}\,.\tag{5.7}$$

Taking $\ell\to\infty$, we conclude that $v_\infty\geq \mathbb{E}[V_n]-O(n^{-\delta})$. A similar reasoning exchanging Player 1 with Player 2 shows that $v_\infty\leq \mathbb{E}[V_n]+O(n^{-\delta})$ and concludes the proof of the lemma. □

Finally, we are ready to prove Theorem 5.3.1.

Proof of Theorem 5.3.1. Fix an arbitrary $\varepsilon>0$, $z\in Z$ and let $V_n=V_n(z)$. By Lemma 5.5.4, there is a constant $K>0$ such that $|v_\infty-\mathbb{E}[V_n]|\leq Kn^{-\delta}$ for all $n\geq 1$. Combining this with the triangle inequality and Lemma 5.5.1 shows that, for every $t\geq 0$,

$$\begin{aligned}\mathbb{P}(|V_n-v_\infty|\geq t+Kn^{-\delta}) &\leq \mathbb{P}(|V_n-\mathbb{E}[V_n]|\geq t+Kn^{-\delta}-|\mathbb{E}[V_n]-v_\infty|)\\ &\leq \mathbb{P}(|V_n-\mathbb{E}[V_n]|\geq t)\leq 2\exp\left(\frac{-t^2n^2}{2h(n)}\right)\,,\end{aligned}$$

which is the desired result. □

5.6 Directed Games on Trees

We present the proof of Theorem 5.3.2. The first lemma in this section bootstraps upon the conclusion of Lemma 5.5.1 (which still holds in this setting), thus deriving superexponential concentration for the value of the n-stage game. Below, $\log$ stands for the natural logarithm.

Lemma 5.6.1. *Fix $\delta \in (0, 1/2)$, $n \geq 1$ and $t \geq n^{-\delta}$. For every even integer $k \in [2, n]$ such that*

$$k \log d + 2 \log 2 \leq t^2(n-k), \tag{5.8}$$

we have

$$\begin{aligned}\mathbb{P}(nV_n - (n-k)\mathbb{E}[V_{n-k}] \geq& (n-k)t + k) \leq \exp\left(-\frac{d^{k/2}}{6}\right),\\ \mathbb{P}(nV_n - (n-k)\mathbb{E}[V_{n-k}] \leq& -(n-k)t - k) \leq \exp\left(-\frac{d^{k/2}}{6}\right).\end{aligned}$$

The proof goes roughly as follows. Given an intermediate stage $k = k(n)$ of the game such that $(n - k(n))$ grows to infinity, there typically exist only a few vertices such that the value of the $(n-k)$-stage game starting from them has a value that is very far from the expectation (we see these vertices as “bad”). Therefore, with foresight, both players can avoid such “bad” vertices, thus boosting the concentration of the n-stage value.

Proof of Lemma 5.6.1. First of all, since T is a transitive graph, for every fixed $n \geq 1$, the variables $(V_n(z))_z$ have the same distribution. For every even integer $k \in [n]$, denote

$$S_k := \{z \in Z_k : V_{n-k}(z) - \mathbb{E}[V_{n-k}] \geq t\}.$$

In other words, S_k is the set of vertices z that could be reached from z_0 after k stages, for which the value of the $(n-k)$-stage game starting at z is greater than or equal to $\mathbb{E}[V_{n-k}] + t$.

Define the event $\mathcal{E}_k := \{|S_k| \geq d^{k/2}\}$. We provide an upper bound for $\mathbb{P}(\mathcal{E}_k)$. Since the random variables $(V_{n-k}(z))_{z \in Z_k}$ are i.i.d., we have that $|S_k|$ follows a binomial distribution $\mathrm{Bin}(d^k, q)$ where $q := \mathbb{P}(V_{n-k} \geq \mathbb{E}[V_{n-k}] + t)$. Consequently, by Lemma 5.5.1 (where the transient speed h is the identity function on $\mathbb{N}$ since, for all $z \in Z$ and $k \geq 2$, $Z_k(z)$ consists of all descendants of z at distance k, which form an antichain), $|S_k|$ is stochastically dominated by a binomial random variable $\mathrm{Bin}(d^k, \tilde{q})$ where $\tilde{q} = \exp(-t^2(n-k)/2)$. In particular,

$$\mathbb{P}(\mathcal{E}_k) \leq \mathbb{P}\left(\mathrm{Bin}(d^k, \tilde{q}) \geq d^{k/2}\right).$$

The random variable $\mathrm{Bin}(d^k, \tilde{q})$ has mean $\mu := d^k \tilde{q}$. We define

$$\xi := \frac{d^{k/2}}{\mu} - 1 = \exp\left(\frac{t^2(n-k) - k\log(d)}{2}\right) - 1 \geq 1,$$

where the last inequality comes from (5.8). Since $d^{k/2} = (1+\xi)\mu$, we have that

$$\mathbb{P}(\mathrm{Bin}(d^k, \tilde{q}) \geq d^{k/2}) = \mathbb{P}(\mathrm{Bin}(d^k, \tilde{q}) \geq (1+\xi)\mu).$$

Therefore, since $\xi \geq 1$ (so $3\xi \geq 2+\xi$), by Chernoff's bound,

$$\begin{aligned}\mathbb{P}\left(\mathrm{Bin}(d^k,\tilde{q}) \geq d^{k/2}\right) &\leq \exp\left(-\frac{\xi^2\mu}{2+\xi}\right)\\ &\leq \exp\left(-\frac{\xi\mu}{3}\right)\\ &= \exp\left(-\frac{d^{k/2}}{3}\left(1-d^{k/2}\tilde{q}\right)\right).\end{aligned}$$

Since $\xi = 1/(d^{k/2}\tilde{q}) - 1 \geq 1$, we have that $1 - d^{k/2}\tilde{q} \geq 1/2$, which finally yields

$$\mathbb{P}(\mathcal{E}_k) \leq \exp\left(-\frac{d^{k/2}}{6}\right). \tag{5.9}$$

At the same time, on the event $|S_k| < d^{k/2}$ (that is, $\overline{\mathcal{E}_k}$), Player 2 can ensure that the token avoids ending up in S_k after k stages. Indeed, at each of the $k/2 \in \mathbb{N}$ turns corresponding to decisions of Player 2, by the pigeonhole principle, Player 2 can always move the token to a vertex having at most a $(1/d)$-fraction of all remaining elements in S_k among its descendants. Since Player 2 has $k/2$ turns and $d^{-k/2}|S_k| < 1$, Player 2 can safely avoid the set S_k at stage k.

Let us condition on the event $\overline{\mathcal{E}_k}$. Then, Player 2 can guarantee that the sum of the payoffs over the last $n-k$ stages is strictly smaller than $(n-k)(\mathbb{E}[V_{n-k}]+t)$. Moreover, the sum of the first k payoffs is at most k. Consequently, Player 2 can guarantee that, after n stages, the global mean payoff is strictly smaller than $k/n + (n-k)(\mathbb{E}[V_{n-k}]+t)/n$, in other words,

$$nV_n < (n-k)\mathbb{E}[V_{n-k}] + (n-k)t + k. \tag{5.10}$$

In particular, using (5.9) implies that

$$\mathbb{P}\left(nV_n - (n-k)\mathbb{E}[V_{n-k}] \geq (n-k)t + k\right) \leq \mathbb{P}(|S_k| \geq d^{k/2}) = \mathbb{P}(\mathcal{E}_k) \leq \exp\left(-\frac{d^{k/2}}{6}\right).$$

A similar reasoning for Player 1 (using the sets $\tilde{S}_k := \{z \in Z_k : V_{n-k}(z) - \mathbb{E}[V_{n-k}] \leq -t\}$ instead of S_k and replacing (5.10) with $nV_n > (n-k)\mathbb{E}[V_{n-k}] - (n-k)t$) yields

$$\mathbb{P}\left(nV_n - (n-k)\mathbb{E}[V_{n-k}] \leq -(n-k)t\right) \leq \exp\left(-\frac{d^{k/2}}{6}\right),$$

which implies the second statement. Note that the additional $-k$ in it is introduced for reasons of symmetry only. □

Next, we show that the expected value of the n-stage game converges rapidly as n grows to infinity.

Lemma 5.6.2. *There exists $v_\infty \in \mathbb{R}$ such that, for every $\delta \in (0,1/2)$, we have $|\mathbb{E}[V_n] - v_\infty| = O(n^{-\delta})$.*

The arguments in the proof are very similar to the ones from the proof of Theorem 5.5.4 but the stronger concentration derived in Lemma 5.6.1 replaces the standard bounded difference estimate from Lemma 5.5.1.

Proof. Fix $\delta' \in (0, 1/2)$, $n \geq 1$, and $t \geq n^{-\delta'}$. Set $k = k(n) := 2\left\lfloor n^{1-2\delta'}/4\log d\right\rfloor$. Then, $k\log d + 2\log 2 \leq t^2(n-k)$ for all large n. For every even integer $m \in [n/2, 2n]$ and large n, we have

$$\begin{aligned}
&\mathbb{P}\left(\min_{z\in Z_m} nV_n(z) \leq (n-k)(\mathbb{E}[V_{n-k}] - t) - k\right) \\
&\quad\leq \sum_{z\in Z_m} \mathbb{P}\left(nV_n(z) \leq (n-k)(\mathbb{E}[V_{n-k}] - t) - k\right) \\
&\quad\leq d^m \exp\left(-\frac{d^{k/2}}{6}\right) \\
&\quad\leq \exp\left(2n\log d - \frac{d^{\lfloor n^{1-2\delta'}/4\log d\rfloor}}{6}\right),
\end{aligned}$$

where the first inequality comes from a union bound and the second inequality comes from Lemma 5.6.1. Fix $\delta \in (0, \delta')$ and define, for all $n \geq 1$,

$$\varepsilon_n := n^{-\delta} \quad \text{and} \quad \psi(n) := \exp\left(2n\log d - d^{\lfloor n^{1-2\delta'}/4\log d\rfloor}/6\right).$$

For large n and every even integer $m \in [n/2, 2n]$, we have

$$\begin{aligned}
\mathbb{E}\left[\min_{z\in Z_m} V_n(z)\right] &\geq \left(\frac{n-k}{n}(\mathbb{E}[V_{n-k}] - \varepsilon_n) - \frac{k}{n}\right) \qquad (5.11) \\
&\quad\mathbb{P}\left(\min_{z\in Z_m} nV_n(z) > (n-k)(\mathbb{E}[V_{n-k}] - \varepsilon_n) - k\right) \\
&\geq \left(\frac{n-k}{n}(\mathbb{E}[V_{n-k}] - \varepsilon_n) - \frac{k}{n}\right)(1 - \psi(n)) \\
&\geq \left(\mathbb{E}[V_{n-k}] - \frac{k}{n}(1 + \mathbb{E}[V_{n-k}]) - \varepsilon_n\right)(1 - \psi(n)) \\
&\geq \left(\mathbb{E}[V_n] - \frac{3k}{n} - \varepsilon_n\right)(1 - \psi(n)) \geq \mathbb{E}[V_n] - (\psi(n) + 2\varepsilon_n), \qquad (5.12)
\end{aligned}$$

where in the fourth inequality we used that $\mathbb{E}[V_n] \leq \mathbb{E}[V_{n-k}] + k/n$ by Lemma 5.5.2 and $1 + \mathbb{E}[V_{n-k}] \leq 2$, and the last inequality is valid for large n because $k/n = o(\varepsilon_n)$.

Consider integers $n \geq 1$ and even $m \in [n/2, 2n]$. In the $(n+m)$-stage game, Player 1 can play according to an optimal strategy for the m-stage game starting at z_0, and then play according to an optimal strategy for the n-stage game starting from the state z reached after m stages. This guarantees that $(m+n)V_{m+n} \geq mV_m + \min_{z\in Z_m} nV_n(z)$. Taking expectations on both sides and using (5.12) yields

$$\begin{aligned}
(m+n)\mathbb{E}[V_{m+n}] &\geq m\mathbb{E}[V_m] + n\mathbb{E}\left[\min_{z\in Z_m} V_n(z)\right] \\
&\geq m\mathbb{E}[V_m] + n\mathbb{E}[V_n] - n(\psi(n) + 2\varepsilon_n).
\end{aligned}$$

We find a similar inequality for odd $m \in [n/2, 2n]$. In this case, $m+1$ is even and also in $[n/2, 2n]$. Then, the previous inequality applied to $m+1$ and n yields

$$(m+n+1)\mathbb{E}[V_{m+n+1}] \geq (m+1)\mathbb{E}[V_{m+1}] + n\mathbb{E}[V_n] - n(\psi(n) + 2\varepsilon_n). \qquad (5.13)$$

However,

$$(m+n)\mathbb{E}[V_{m+n}] \geq (m+n+1)\mathbb{E}[V_{m+n+1}] - 1 \quad \text{and} \quad (m+1)\mathbb{E}[V_{m+1}] \geq m\mathbb{E}[V_m]\,,$$

which combined with (5.13) gives

$$(m+n)\mathbb{E}[V_{m+n}] \geq m\mathbb{E}[V_m] + n\mathbb{E}[V_n] - n(\psi(n) + 2\varepsilon_n) - 1\,.$$

To sum things up, for large n and $m \in [n/2, 2n]$,

$$(m+n)\mathbb{E}[V_{m+n}] \geq m\mathbb{E}[V_m] + n\mathbb{E}[V_n] - n(\psi(n) + 2\varepsilon_n) - 1\,. \tag{5.14}$$

Recall that there is a constant $K' > 0$ such that, for all $n \geq 1$, $n(\psi(n) + 2\varepsilon_n) + 1 \leq K'n^{1-\delta}$. We define $\phi(n) := K'n^{1-\delta}$ and deduce from (5.14) that

$$(m+n)\mathbb{E}[V_{m+n}] \geq m\mathbb{E}[V_m] + n\mathbb{E}[V_n] - \phi(n+m)\,.$$

Moreover, ϕ is increasing and verifies $\sum_{n\geq 1} \phi(n)/n^2 < \infty$. Consequently, Lemma 5.4.2 applied to the function $f\colon n \in \mathbb{N} \mapsto -n\mathbb{E}[V_n]$ implies that that $\mathbb{E}[V_n]$ converges to a limit $v_\infty \in \mathbb{R} \cup \{\infty\}$ as $n \to \infty$. Note that $v_\infty \in [0, 1]$ since $V_n \in [0, 1]$ for all $n \geq 1$.

Finally, using (5.14) with $m = n$ and a telescopic summation shows that the inequality (5.7) still holds. In particular, we conclude that $v_\infty \geq \mathbb{E}[V_n] - O(n^{-\delta})$. A similar reasoning replacing Player 1 with Player 2 shows that $v_\infty \leq \mathbb{E}[V_n] + O(n^{-\delta})$ and concludes the proof of the lemma. □

We are now ready to prove Theorem 5.3.2.

Proof of Theorem 5.3.2. Fix $t \geq n^{-\delta}$ and let K' be a constant such that $|\mathbb{E}[V_n] - v_\infty| \leq K'n^{-\delta}$ for all large n. Using that, for all n and $k \leq n$, we have $|nV_n - (n-k)V_{n-k}| \leq k$, and fixing $k = 2\lceil \frac{t^2 n}{4 \log d} \rceil$ (which satisfies (5.8)), we get

$$\begin{aligned}
&\mathbb{P}(|V_n - v_\infty| \geq t + 2t^2 + K'n^{-\delta}) \\
&\quad \leq \mathbb{P}\Bigg(\left| V_n - \frac{n-k}{n}\mathbb{E}[V_{n-k}] \right| \geq t + 2t^2 + K'n^{-\delta} \\
&\qquad\qquad - \left| \frac{n-k}{n}\mathbb{E}[V_{n-k}] - \mathbb{E}[V_n] \right| - |\mathbb{E}[V_n] - v_\infty| \Bigg) \\
&\quad \leq \mathbb{P}\left(\left| V_n - \frac{n-k}{n}\mathbb{E}[V_{n-k}] \right| \geq t + t^2 \right) \\
&\quad \leq \mathbb{P}(|nV_n - (n-k)\mathbb{E}[V_{n-k}]| \geq (n-k)t + k) \\
&\quad \leq \exp\left(-\frac{d^{\lfloor k/2 \rfloor}}{6} \right) \leq \exp\left(-\frac{d^{t^2 n/(4\log d)}}{6} \right) = \exp\left(-\frac{1}{6}\exp\left(\frac{t^2 n}{4} \right) \right)\,,
\end{aligned}$$

where the first inequality comes from the triangle inequality, the second inequality comes from the definition of K' and the fact that $|nV_n - (n-k)V_{n-k}| \leq k \leq nt^2$, and the third inequality once again uses the fact that $k \leq nt^2$.

Finally, choosing $K \geq K'$ sufficiently large ensures that, first, the upper bound shown above holds for all $n \geq 1$ (and not only for large n), and second, the upper bound holds for all $t \geq 0$, which finishes the proof. □

5.7 Oriented Games

We present a simple and self-contained proof of Proposition 5.2.6.

Proof of Proposition 5.2.6. First, by density of the rational vectors in $\mathbb{R}^d$ and rescaling, we may assume that $u \in \mathbb{Z}^d$ is such that the greatest common divisor of its coordinates is 1. We provide an upper bound on $h(n)$. For every integer $i \geq 1$ and initial state $z_0 = z$, define $Z_{2i}(z) \coloneqq \{w \in Z : w \cdot u = z \cdot u + i\}$, $Z_{2i+1}(z) \coloneqq \{w \in Z : w \cdot u = z \cdot u - i\}$, and $Z_0(z) \coloneqq \{z\}$ and $Z_1(z) \coloneqq \{w \in Z \setminus \{z\} : w \cdot u = z \cdot u\}$. Then, for all $z \in Z$, $(Z_i(z))_{i \geq 1}$ form a partition of Z and each of them can be visited at most once, i.e., for all $n \geq 1$, the sets $(Z_1, Z_2, \ldots, Z_n)$ forms an antichain. Set $r \coloneqq \max_{(u,v) \in E(\Gamma)} \|v - u\|_2$. After n steps of the game, the position z_n satisfies $\|z_n - z\|_2 \leq nr$, and by the Cauchy-Schwarz inequality,

$$|(z_n - z) \cdot u| \leq \|z_n - z\|_2 \cdot \|u\|_2 \leq \lceil nr \cdot \|u\|_2 \rceil =: N = N(n)\,.$$

In particular, $Z^{(n)}(z)$ is contained in the ball with radius N around z, which itself is contained in $Z_1 \cup Z_2 \cup \ldots \cup Z_{2N+1}(z)$, so the transient speed of the process satisfies $h(n) \leq 2N(n) + 1$ for all $n \geq 1$.

Now, fix $\delta \in (0, 1/2)$ and $z_0 = z \in Z$. We show that the game is δ-transient. Set $\varepsilon_n \coloneqq n^{-\delta}$. Then,

$$\begin{aligned} \psi(n, \varepsilon_n) &= \exp\left(-\frac{\varepsilon_n^2 n^2}{2h(n)}\right) \max_{z \in Z} |Z^{(2n)}(z)| \\ &\leq \exp\left(-\frac{\varepsilon_n^2 n}{6r \cdot \|u\|_2}\right) (2nr \cdot \|u\|_2 + 1)^d \\ &= \exp\left(-\frac{n^{1-2\delta}}{6r \cdot \|u\|_2}\right) (2nr \cdot \|u\|_2 + 1)^d = O(n^{-\delta})\,. \end{aligned}$$

Hence, for all $\delta \in (0, 1/2)$, $\varepsilon_n + \psi(n, \varepsilon_n) = O(n^{-\delta})$, and therefore, the game is δ-transient. □

Bibliography

[ABFG$^+$93] Aristotle Arapostathis, Vivek S. Borkar, Emmanuel Fernández-Gaucherand, Mrinal K. Ghosh, and Steven I. Marcus. Discrete-Time Controlled Markov Processes with Average Cost Criterion: A Survey. *SIAM Journal on Control and Optimization*, 31(2):282–344, 1993.

[ACSS24a] Ali Asadi, Krishnendu Chatterjee, Raimundo Saona, and Ali Shafiee. Limit-sure reachability for small memory policies in POMDPs is NP-complete, 2024.

[ACSS24b] Ali Asadi, Krishnendu Chatterjee, Raimundo Saona, and Jakub Svoboda. Concurrent Stochastic Games with Stateful-Discounted and Parity Objectives: Complexity and Algorithms. *Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, 323:5:1–5:17, 2024.

[ACSSU24] Ali Asadi, Krishnendu Chatterjee, Jakub Svoboda, and Raimundo Saona Urmeneta. Deterministic Sub-exponential Algorithm for Discounted-sum Games with Unary Weights. In *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science*, page 1–12, 2024.

[ACSZ21] Ben Amiet, Andrea Collevecchio, Marco Scarsini, and Ziwen Zhong. Pure Nash Equilibria and Best-Response Dynamics in Random Games. *Mathematics of Operations Research*, 46(4):1552–1572, 2021.

[Adl13] Ilan Adler. The Equivalence of Linear Programs and Zero-Sum Games. *International Journal of Game Theory*, 42(1):165–177, 2013.

[AFFG01] Eitan Altman, Eugene A. Feinberg, Jerzy Filar, and Vladimir A. Gaitsgory. Perturbed Zero-Sum Games with Applications to Stochastic and Repeated Games. In *Advances in Dynamic Games and Applications*, Annals of the International Society of Dynamic Games, page 165–181. Birkhäuser, 2001.

[AFH13] Konstantin E. Avrachenkov, Jerzy A. Filar, and Phil G. Howlett. *Analytic Perturbation Theory and Its Applications*. Society for Industrial and Applied Mathematics, 2013.

[ALM$^+$25] Luc Attia, Lyuben Lichev, Dieter Mitsche, Raimundo Saona, and Bruno Ziliotto. Random Zero-Sum Dynamic Games on Infinite Directed Graphs. *Dynamic Games and Applications*, 2025.

[Ami03] Rabah Amir. Stochastic Games in Economics and Related Fields: An Overview. In *Stochastic Games and Applications*, page 455–470. Springer, 2003.

[AOB19] Luc Attia and Miquel Oliu-Barton. A Formula for the Value of a Stochastic Game. *Proceedings of the National Academy of Sciences*, 116(52):26435–26443, 2019.

[AOB21] Luc Attia and Miquel Oliu-Barton. Shapley–Snow Kernels, Multiparameter Eigenvalue Problems, and Stochastic Games. *Mathematics of Operations Research*, 46(3):1181–1202, 2021.

[AOB24] Luc Attia and Miquel Oliu-Barton. Stationary Equilibria in Discounted Stochastic Games. *Dynamic Games and Applications*, 14(2):271–284, 2024.

[AOBS25] Luc Attia, Miquel Oliu-Barton, and Raimundo Saona. Marginal Values of a Stochastic Game. *Mathematics of Operations Research*, 50(1):482–505, 2025.

[ARY21] Noga Alon, Kirill Rudov, and Leeat Yariv. Dominance Solvability in Random Games. *SSRN Electronic Journal*, page 1–45, 2021.

[BE52] N.G. de Bruijn and P. Erdös. Some Linear and Some Quadratic Recursion Formulas II. *Indagationes Mathematicae (Proceedings)*, 55:152–163, 1952.

[Bel57] Richard Bellman. A Markovian Decision Process. *Journal of Mathematics and Mechanics*, 6(5):679–684, 1957.

[Bel01] Beling. Exact Algorithms for Linear Programming over Algebraic Extensions. *Algorithmica*, 31(4):459–478, 2001.

[BF68] David Blackwell and T. S. Ferguson. The Big Match. *The Annals of Mathematical Statistics*, 39(1):159–163, 1968.

[BG09] Blai Bonet and Héctor Geffner. Solving POMDPs: RTDP-Bel vs. Point-based Algorithms. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, page 1641–1646, 2009.

[BGB12] Christel Baier, Marcus Grösser, and Nathalie Bertrand. Probabilistic

$omega$-automata. *Journal of the ACM*, 59(1):1–52, 2012.

[BK76] Truman Bewley and Elon Kohlberg. The Asymptotic Theory of Stochastic Games. *Mathematics of Operations Research*, 1(3):197–208, 1976.

[BK78] Truman Bewley and Elon Kohlberg. On Stochastic Games with Stationary Optimal Strategies. *Mathematics of Operations Research*, 3(2):104–125, 1978.

[BKPR23] Dhruv Bhasin, Sayar Karmakar, Moumanti Podder, and Souvik Roy. On a class of PCA with size-3 neighborhood and their applications in percolation games. *Electronic Journal of Probability*, 28(none):1–60, 2023.

[Bla62] David Blackwell. Discrete Dynamic Programming. *The Annals of Mathematical Statistics*, 33(2):719–726, 1962.

[BLC$^+$18] Kannan Balakrishnan, Divya Sindhu Lekha, Manoj Changat, Bijo S. Anand, and Prasanth G. Narasimha-Shenoi. Generalized Vertex Transitivity in Graphs, 2018.

[Bor00] Vivek S. Borkar. Average Cost Dynamic Programming Equations for Controlled Markov Chains with Partial Observations. *SIAM Journal on Control and Optimization*, 39(3):673–681, 2000.

[BPR06] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in Real Algebraic Geometry*. Springer, 2 edition, 2006.

[BR14] Matthew Bourque and T. E. S. Raghavan. Policy improvement for perfect information additive reward andadditive transition stochastic games with discounted and average payoffs. *Journal of Dynamics & Games*, 1(3):347–361, 2014.

[BR23] Benjamin Brooks and Philip J. Reny. A canonical game—75 years in the making—showing the equivalence of matrix games and linear programming. *Economic Theory Bulletin*, 11(2):171–180, 2023.

[Buk80] Rais G. Bukharaev. Probabilistic Automata. *Journal of Soviet Mathematics*, 13(3):359–386, 1980.

[Cau32] Augustin-Louis Cauchy. Calcul des Indices des Fonctions. *Journal de l'École Polytechnique*, 15(25):176–229, 1832.

[Cha07] Krishnendu Chatterjee. Concurrent Games with Tail Objectives. *Theoretical Computer Science*, 388(1):181–198, 2007.

[CJSS25] Krishnendu Chatterjee, Mahdi JafariRaviz, Raimundo Saona, and Jakub Svoboda. Value Iteration with Guessing for Markov Chains and Markov Decision Processes. In *Tools and Algorithms for the Construction and Analysis of Systems*, volume 15697, page 217–236. Springer, 2025.

[CLSS25] Krishnendu Chatterjee, Ruichen Luo, Raimundo Saona, and Jakub Svoboda. Linear Equations with Min and Max Operators: Computational Complexity. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(11):11150–11157, 2025.

[CLSZ24] Krishnendu Chatterjee, David Lurie, Raimundo Saona, and Bruno Ziliotto. Ergodic Unobservable MDPs: Decidability of Approximation, 2024.

[CMS22] Krishnendu Chatterjee, Mona Mohammadi, and Raimundo Saona. Repeated Prophet Inequality with Near-optimal Bounds, 2022.

[CMSS23] Krishnendu Chatterjee, Tobias Meggendorfer, Raimundo Saona, and Jakub Svoboda. Faster Algorithm for Turn-based Stochastic Games with Bounded Treewidth. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, page 4590 – 4605, 2023.

[COBS24] Krishnendu Chatterjee, Miquel Oliu-Barton, and Raimundo Saona. Value-Positivity for Matrix Games. *Mathematics of Operations Research*, page 1–24, 2024.

[COBZ21] Olivier Catoni, Miquel Oliu-Barton, and Bruno Ziliotto. Constant Payoff in Zero-Sum Stochastic Games. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 57(4):1888–1900, 2021.

[CSZ21] Krishnendu Chatterjee, Raimundo Saona, and Bruno Ziliotto. Finite-Memory Strategies in POMDPs with Long-Run Average Objectives. *Mathematics of Operations Research*, 47(1):100–119, 2021.

[DAHK07] Luca De Alfaro, Thomas A. Henzinger, and Orna Kupferman. Concurrent reachability games. *Theoretical Computer Science*, 386(3):188–217, 2007.

[Dan51] G.B. Dantzig. A Proof of the Equivalence of the Programming Problem and the Game Problem. In *Activity Analysis of Production and Allocation*, volume 13, page 330–338. John Wiley and Sons, 1951.

[DEKM98] Richard Durbin, Sean Roberts Eddy, Anders Krogh, and Graeme Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

[DSZ24] Andrea Davini, Raimundo Saona, and Bruno Ziliotto. Stochastic Homogenization of HJ Equations: a Differential Game Approach, 2024.

[DT03] George Bernard Dantzig and Mukund Narain Thapa. *Linear Programming: 2: Theory and Extensions*. Springer, 2003.

[EM79] A. Ehrenfeucht and J. Mycielski. Positional Strategies for Mean Payoff Games. *International Journal of Game Theory*, 8(2):109–113, 1979.

[Eve57] Hugh Everett. Recursive Games. In *Contributions to the Theory of Games III*, volume 39 of *Annals of Mathematical Studies*, page 47–78, 1957.

[Fei96] Eugene A. Feinberg. On Measurability and Representation of Strategic Measures in Markov Decision Processes. In *Statistics, Probability and Game Theory*, page 29–43. Institute of Mathematical Statistics, 1996.

[Fia83] Anthony V. Fiacco. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. Academic Press, 1 edition, 1983.

[Fia97] Anthony V. Fiacco, editor. *Mathematical Programming with Data Perturbations*. CRC Press, 1 edition, 1997.

[FPS23] János Flesch, Arkadi Predtetchinski, and Ville Suomala. Random Perfect Information Games. *Mathematics of Operations Research*, 48(2):708–727, 2023.

[FV97] Jerzy Filar and Koos Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.

[FW12] Mark I. Freidlin and Alexander D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, 3 edition, 2012.

[Gal94] Tomas Gal. *Postoptimal Analyses, Parametric Programming, and Related Topics: Degeneracy, Multicriteria Decision Making, Redundancy*. De Gruyter, 1994.

[Gar19] Tristan Garrec. Communicating zero-sum product stochastic games. *Journal of Mathematical Analysis and Applications*, 477(1):60–84, 2019.

[GGH97] Tomas Gal, Harvey J. Greenberg, and Frederick S. Hillier, editors. *Advances in Sensitivity Analysis and Parametric Programming*. Springer, 1997.

[Gil58] Dean Gillette. Stochastic Games with Zero Stop Probabilities. In *Contributions to the Theory of Games*, volume III, page 179–188. Princeton University Press, 1958.

[GMS25] Giordano Giambartolomei, Frederik Mallmann-Trenn, and Raimundo Saona. IID Prophet Inequality with Random Horizon: Going beyond Increasing Hazard Rates. In *52nd International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 334, pages 87:1–87:21, 2025.

[GMTS23] Giordano Giambartolomei, Frederik Mallmann-Trenn, and Raimundo Saona. Prophet Inequalities: Separating Random Order from Order Selection, 2023.

[GZ23] Guillaume Garnier and Bruno Ziliotto. Percolation Games. *Mathematics of Operations Research*, 48(4):1811–2382, 2023.

[HIJN21] Kristoffer Arnsfelt Hansen, Rasmus Ibsen-Jensen, and Abraham Neyman. Absorbing Games with a Clock and Two Bits of Memory. *Games and Economic Behavior*, 128:213–230, 2021.

[HIJN23] Kristoffer Arnsfelt Hansen, Rasmus Ibsen-Jensen, and Abraham Neyman. The Big Match with a Clock and a Bit of Memory. *Mathematics of Operations Research*, 48(1):419–432, 2023.

[HJM$^+$23] Torsten Heinrich, Yoojin Jang, Luca Mungo, Marco Pangallo, Alex Scott, Bassel Tarbush, and Samuel Wiese. Best-Response Dynamics, Playing Sequences, and Convergence to Equilibrium in Random Games. *International Journal of Game Theory*, 52(3):703–735, 2023.

[HK66] A. J. Hoffman and R. M. Karp. On Nonterminating Stochastic Games. *Management Science*, 12(5):359–370, 1966.

[HL14] G. H. Hardy and J. E. Littlewood. Tauberian Theorems Concerning Power Series and Dirichlet's Series whose Coefficients are Positive>. *Proceedings of the London Mathematical Society*, s2-13(1):174–191, 1914.

[HLL03] Onésimo Hernández-Lerma and Jean Bernard Lasserre. *Markov Chains and Invariant Probabilities*. Birkhäuser, 2003.

[HMM19] Alexander E. Holroyd, Irène Marcovici, and James B. Martin. Percolation games, probabilistic cellular automata, and the hard-core model. *Probability Theory and Related Fields*, 174(3-4):1187–1217, 2019.

[Jac46] Carl Gustav Jacob Jacobi. Über die Darstellung einer Reihe gegebner Werthe durch eine gebrochne rationale Function. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1846(30):127–156, 1846.

[Jer73a] Robert G. Jeroslow. Asymptotic Linear Programming. *Operations Research*, 21(5):1128–1141, 1973.

[Jer73b] Robert G. Jeroslow. Linear Programs Dependent on a Single Parameter. *Discrete Mathematics*, 6(2):119–140, 1973.

[JLR00] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random Graphs*. John Wiley & Sons, 2000.

[Kar91] Howard Karloff. *Linear programming*. Birkhäuser, 1991.

[Kha80] Leonid Genrikhovich Khachiyan. Polynomial Algorithms in Linear Programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53–72, 1980.

[KLM96] Leslie Pack Kaelbling, Michael Lederman Littman, and A. W. Moore. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[Koh74] Elon Kohlberg. Repeated Games with Absorbing States. *The Annals of Statistics*, 2(4):724–738, 1974.

[KP13] Steven G. Krantz and Harold R. Parks. *The Implicit Function Theorem: History, Theory, and Applications*. Birkhäuser, 1 edition, 2013.

[KT14] Harold W. Kuhn and Albert W. Tucker. Nonlinear Programming. In *Traces and Emergence of Nonlinear Programming*, page 247–258. Springer, 2014.

[Kuh50] Harold William Kuhn. Extensive Games. *Proceedings of the National Academy of Sciences*, 36(10):570–576, 1950.

[LR20] Rida Laraki and Jérôme Renault. Acyclic Gambling Games. *Mathematics of Operations Research*, 45(4):1237–1257, 2020.

[LS15] Rida Laraki and Sylvain Sorin. Advances in Zero-Sum Dynamic Games. In *Handbook of Game Theory with Economic Applications*, volume 4, page 27–93. Elsevier, 2015.

[LS17] Yehuda John Levy and Eilon Solan. Stochastic Games. In *Encyclopedia of Complexity and Systems Science*, page 1–23. Springer, 2017.

[Mah64] Kurt Mahler. An Inequality for the Discriminant of a Polynomial. *Michigan Mathematical Journal*, 11(3):257–262, 1964.

[Mar75] D. H. Martin. On the Continuity of the Maximum in Parametric Linear Programming. *Journal of Optimization Theory and Applications*, 17(3-4):205–210, 1975.

[Meg91] Nimrod Megiddo. On Finding Primal- and Dual-Optimal Bases. *ORSA Journal on Computing*, 3(1):63–65, 1991.

[MHC03] Omid Madani, Steve Hanks, and Anne Condon. On the Undecidability of Probabilistic Planning and Related Stochastic Optimization Problems. *Artificial Intelligence*, 147(1):5–34, 2003.

[Mil56] Harlan D. Mills. Marginal Values of Matrix Games and Linear Programs. In *Linear Inequalities and Related Systems*, volume 38, page 183–194. Princeton University Press, 1956.

[Mir71] Leon Mirsky. A Dual of Dilworth's Decomposition Theorem. *The American Mathematical Monthly*, 78(8):876–877, 1971.

[MN81] Jean-François Mertens and Abraham Neyman. Stochastic Games. *International Journal of Game Theory*, 10(2):53–66, 1981.

[NEG05] Arnab Nilim and Laurent El Ghaoui. Robust Control of Markov Decision Processes with Uncertain Transition Matrices. *Operations Research*, 53(5):780–798, 2005.

[NR93] A. S. Nowak and T. E. S. Raghavan. A finite step algorithm via a bimatrix game to a single controller non-zero sum stochastic game. *Mathematical Programming*, 59(1):249–259, 1993.

[NS10] Abraham Neyman and Sylvain Sorin. Repeated Games with Public Uncertain Duration Process. *International Journal of Game Theory*, 39:29–52, 2010.

[OB14] Miquel Oliu-Barton. The Asymptotic Value in Finite Stochastic Games. *Mathematics of Operations Research*, 39(3):712–721, 2014.

[OB18] Miquel Oliu-Barton. The Splitting Game: Value and Optimal Strategies. *Dynamic Games and Applications*, 8(1):157–179, 2018.

[OB20] Miquel Oliu-Barton. New Algorithms for Solving Zero-Sum Stochastic Games. *Mathematics of Operations Research*, 46(1):255–267, 2020.

[OB22] Miquel Oliu-Barton. Weighted-Average Stochastic Games with Constant Payoff. *Operational Research*, 22(3):1675–1696, 2022.

[OBV23] Miquel Oliu-Barton and Guillaume Vigeral. Absorbing Games with Irrational Values. *Operations Research Letters*, 51(6):555–559, 2023.

[Ost37] Alexander Ostrowski. Sur La Détermination Des Bornes Inférieures Pour Une Classe Des Déterminants;. *Bulletin des Sciences Mathématiques*, 61:19–32, 1937.

[Ost52] A. M. Ostrowski. Note on bounds for Ddeterminants with Dominant Principal Diagonal. *Proceedings of the American Mathematical Society*, 3(1):26–30, 1952.

[Paz71] Azaria Paz. *Introduction to Probabilistic Automata*. Elsevier, 1971.

[Pui50] Victor Puiseux. Recherches sur les Fonctions Algébriques. *Journal de Mathématiques Pures et Appliquées*, 15:365–480, 1850.

[Rab63] Michael Oser Rabin. Probabilistic Automata. *Information and Control*, 6(3):230–245, 1963.

[Ren10] Jérôme Renault. Uniform Value in Dynamic Programming. *Journal of the European Mathematical Society*, 13(2):309–330, 2010.

[Roc70] Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[RS01] Dinah Rosenberg and Sylvain Sorin. An Operator Approach to Zero-Sum Repeated Games. *Israel Journal of Mathematics*, 121(1):221–246, 2001.

[RS02] T. E. S. Raghavan and Zamir Syed. Computing Stationary Nash Equilibria of Undiscounted Single-Controller Stochastic Games. *Mathematics of Operations Research*, 27(2):384–400, 2002.

[RSV02] Dinah Rosenberg, Eilon Solan, and Nicolas Vieille. Blackwell Optimality in Markov Decision Processes with Partial Observation. *The Annals of Statistics*, 30(4):1178–1193, 2002.

[RTV85] T. E. S. Raghavan, S. H. Tijs, and O. J. Vrieze. On Stochastic Games with Additive Reward and Transition Structure. *Journal of Optimization Theory and Applications*, 47(4):451–464, 1985.

[Ruh91] G. Ruhe. Parametric Flows. In *Algorithmic Aspects of Flows in Networks*, page 125–153. Springer, 1991.

[RV17] Jérôme Renault and Xavier Venel. Long-Term Values in Markov Decision Processes and Repeated Games, and a New Distance for Probability Spaces. *Mathematics of Operations Research*, 42(2):349–376, 2017.

[SCFV97] W. W. Szczechla, S. A. Connell, J. A. Filar, and O. J. Vrieze. On the Puiseux Series Expansion of the Limit Discount Equation of Stochastic Games. *SIAM Journal on Control and Optimization*, 35(3):860–875, 1997.

[Sha53] Lloyd Stowell Shapley. Stochastic Games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953.

[Sio58] Maurice Sion. On General Minimax Theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

[SKK22] Raimundo Saona, Fyodor A. Kondrashov, and Ksenia A. Khudiakova. Relation Between the Number of Peaks and the Number of Reciprocal Sign Epistatic Interactions. *Bulletin of Mathematical Biology*, 84(8):74, 2022.

[SKK23] Raimundo Saona, Fyodor A. Kondrashov, and Ksenia A. Khudiakova. Correction to: Relation Between the Number of Peaks and the Number of Reciprocal Sign Epistatic Interactions. *Bulletin of Mathematical Biology*, 85(3):17, 2023.

[Sol03] Eilon Solan. Continuity of the Value of Competitive Markov Decision Processes. *Journal of Theoretical Probability*, 16(4):831–845, 2003.

[Sol22] Eilon Solan. *A Course in Stochastic Game Theory*. Cambridge University Press, 1 edition, 2022.

[Sor03] Sylvain Sorin. The Operator Approach to Zero-Sum Stochastic Games. In *Stochastic Games and Applications*, NATO Science Series, page 417–426, 2003.

[SS50] Lloyd Stowell Shapley and R. N. Snow. Basic Solutions of Discrete Games. In *Contributions to the Theory of Games*, page 27–36. Princeton University Press, 1950.

[ST01] Daniel Spielman and Shang-Hua Teng. Smoothed Analysis of Algorithms: Why the Simplex Algorithm Usually Takes Polynomial Time. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, page 296–305, 2001.

[SV10] Eilon Solan and Nicolas Vieille. Computing Uniformly Optimal Strategies in Two-Player Stochastic Games. *Economic Theory*, 42(1):237–253, 2010.

[SV15] Eilon Solan and Nicolas Vieille. Stochastic Games. *Proceedings of the National Academy of Sciences*, 112(45):13743–13746, 2015.

[SZ16] Eilon Solan and Bruno Ziliotto. Stochastic Games with Signals. In *Advances in Dynamic and Evolutionary Games: Theory, Applications, and Numerical Methods*, Annals of the International Society of Dynamic Games, page 77–94. Springer, 2016.

[TR97] Frank Thuijsman and Thirukkannamangai E. S. Raghavan. Perfect Information Stochastic Games and Related Classes. *International Journal of Game Theory*, 26(3):403–408, 1997.

[TV80] S. H. Tijs and O. J. Vrieze. Perturbation Theory for Games in Normal Form and Stochastic Games. *Journal of Optimization Theory and Applications*, 30(4):549–567, 1980.

[vCH$^+$11] Pavol Černý, Krishnendu Chatterjee, Thomas A. Henzinger, Arjun Radhakrishna, and Rohit Singh. Quantitative Synthesis for Concurrent Programs. In *Computer Aided Verification*, volume 6806, page 243–259. Springer, 2011.

[Vig13] Guillaume Vigeral. A Zero-Sum Stochastic Game with Compact Action Sets and no Asymptotic Value. *Dynamic Games and Applications*, 3(2):172–186, 2013.

[VK20] Tina Verma and Amit Kumar. Matrix Games with Interval Payoffs. In *Fuzzy Solution Concepts for Non-cooperative Games*, volume 383, page 1–36. Springer, 2020.

[vN28] John von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, 1928.

[VS24] Bernhard Von Stengel. Zero-Sum Games and Linear Programming Duality. *Mathematics of Operations Research*, 49(2):1091–1108, 2024.

[VZ16] Xavier Venel and Bruno Ziliotto. Strong Uniform Value in Gambling Houses and Partially Observable Markov Decision Processes. *SIAM Journal on Control and Optimization*, 54(4):1983–2008, 2016.

[VZ21] Xavier Venel and Bruno Ziliotto. History-dependent Evaluations in Partially Observable Markov Decision Process. *SIAM Journal on Control and Optimization*, 59(2):1730–1755, 2021.

[Zil16a] Bruno Ziliotto. A Tauberian Theorem for Nonexpansive Operators and Applications to Zero-Sum Stochastic Games. *Mathematics of Operations Research*, 41(4):1522–1534, 2016.

[Zil16b] Bruno Ziliotto. Zero-sum Repeated Games: Counterexamples to the Existence of the Asymptotic Value and the Conjecture $operatorname\{maxmin\} =$ $operatorname\{lim\}v\_\{n\}$. *The Annals of Probability*, 44(2):1107–1133, 2016.

[Zil17] Bruno Ziliotto. Stochastic Homogenization of Nonconvex Hamilton-Jacobi Equations: A Counterexample. *Communications on Pure and Applied Mathematics*, 70(9):1798–1809, 2017.

[Zil24] Bruno Ziliotto. Mertens Conjectures in Absorbing Games with Incomplete Information. *The Annals of Applied Probability*, 34(2):1948–1986, 2024.