Université Sorbonne Paris Cité

université Paris Diderot Paris 7

Université Paris Diderot (Paris 7)
Sorbonne Paris Cité

École doctorale Sciences Mathématiques de Paris Centre
Institut de Recherche en Informatique Fondamentale

Bruno Karelović

Thèse de Doctorat

Analyse Quantitative des Systèmes Stochastiques – Jeux de Priorité et Population de Chaînes de Markov

Quantitative Analysis of Stochastic Systems – Priority Games and Populations of Markov Chains

Présentée et soutenue le 7 juillet 2017 devant le jury composé de :

Thomas Brihaye	Mathématiques Effectives, Université de Mons	Rapporteur
Serge Haddad	LSV, ENS Cachan	Rapporteur
Blaise Genest	CNRS, IRISA	co-Directeur
Wiesław Zielonka	IRIF, Université Paris Diderot	Directeur
Eugène Asarin	IRIF, Université Paris Diderot	Président
Anne Bouillard	Nokia Bell Labs France	Examinatrice
Hugo Gimbert	LaBRI, Université de Bordeaux	Examinateur

Remerciements

Tout d'abord je tiens à remercier mes directeurs de thèse, pour leur énorme patience. En particulier à Blaise pour m'avoir guidé depuis mon stage à Singapour et à Wiesław ainsi que pour sa grande disponibilité, pour ses conseils et les nombreuses discussions intéressantes que nous avons eu.

Je suis reconnaissant à Serge Haddad et Thomas Brihaye d'avoir bien voulu être les rapporteurs de ce mémoire et dont les suggestions m'ont bien aidé à le rendre meilleur. Merci Thomas pour avoir fait le déplacement pour ma soutenance.

Je voudrais également remercier chaleureusement Eugène Asarin, Anna Bouillard et Hugo Gimbert d'avoir accepté de faire partie du jury de soutenance.

Je remercie également tous les membres et anciens membres de l'IRIF. En particulier, je remercie Amos, Gustavo, Jean-Éric, Michel et Olivier.

J'aimerais aussi remercier mes professeurs qui ont guidé mes premiers pas dans la recherche : Julio Aracena et Anahí Gajardo.

A mi familia que, a pesar de los 12.000 km que nos separan, me ha apoyado un montón.

À tous mes amis du Kremlin et sa banlieue (Ivry, Paris, etc).

Les déjeuners et cafés ont joué un rôle très important dans mon doctorat. Moments de réflexion et de discutions intéressantes qui ont donné lieu à de nombreux projets de recherche. Je remercie mes collègues actuels et anciens, que j'ai eu le plaisir de côtoyer quotidiennement durant ces années et qui ont contribué à faire du laboratoire un lieu agréable et convivial : les Power Rangers (Bredinho, Guedin, Piada Séria, Reiter), mes collègues des projets ANR et des déjeuners : Alexandre, Bruno (Mimmo), Khaled, Laurent, Simon ; mes maîtres de go : Benjamin et Victor ; mes collègues de bureau qui ont dû supporter mes discussions scientifiques avec Alex : Finn, Clément, Florent, Mengchuan ; Alkida Pëllumba, Anna Carla, Célia, Daniela, Dennis, François et Pablo (et les parties d'AoE2), Guillaume, Jehanne, Lucas, Nadia (pour son soutien et sa joie), Mathieu (mon camarade d'enseignement au Cameroun), Maria Rita, Olya, Svetlana, Tatiana, Vincent (pour ses jeux de mots qui m'ont aidé à apprendre le français et parfois à rigoler) et aussi les gens de PPS : Amina, Étienne, Jovana, Lourdes et Pierre.

Sans oublier Allumette, Brocha, Coxinha, le Dream Team, Ewok, Fils de la Terre, el Flaco, el Irresistible, J'aime-bien-le-jambon, John Bonachon, Koala, Michael Jack-

son, Plancton, Rat, Richard Quinet, Sanglier, Tonhão (merci beaucoup), Tropeço et Verganus.

Mes remerciements vont aussi aux gens du CROUS qui m'ont nourri pendant ces années.

Enfin, je remercie Becas Chile. Sans son financement, ma thèse n'aurait pas été possible.

Abstract

This thesis examines some quantitative questions in the framework of two different stochastic models. It is divided into two parts: the first part examines a new class of stochastic games with priority payoff. This class of games contains as proper subclasses the parity games extensively studied in computer science, and limsup and liminf games studied in game theory. The second part of the thesis examines some natural but involved questions about distributions, studied in the simple framework of finite state Markov chain.

In the first part, we examine two-player zero-sum games focusing on a particular payoff function that we call the priority payoff. This payoff function generalizes the payoff used in parity games. We consider both turn-based stochastic priority games and concurrent priority games. Our approach to priority games is based on the concept of the nearest fixed point of monotone nonexpansive mappings and extends the μ-calculus approach to priority games.

The second part of the thesis deals with *population* questions. Roughly speaking, we examine how a probability distribution over states evolves in time. More specifically, we are interested in questions like the following one: from an initial distribution, can the population reach at some moment a distribution with a probability mass exceeding a given threshold in state Goal? It turns out that this type of questions is much more difficult to handle than the questions concerning individual trajectories: it is not known for the simple model of Markov chains whether population questions are decidable. We study restrictions of Markov chains ensuring decidability of population questions.

Résumé

Cette thèse examine certaines questions quantitatives dans le cadre de deux modèles stochastiques différents. Il est divisé en deux parties : la première partie examine une nouvelle classe de jeux stochastiques avec une fonction de paiement particulière que nous appelons « de priorité ». Cette classe de jeux contient comme sous-classes propre les jeux de parité, largement étudiés en informatique, et les jeux de limsup et liminf, étudiés dans la théorie des jeux. La deuxième partie de la thèse examine certaines questions naturelles mais complexes sur les distributions, étudiées dans le cadre plus simple des chaînes de Markov à espace d'états fini.

Dans la première partie, nous examinons les jeux à somme nulle à deux joueurs en se centrant sur la fonction de paiement de priorité. Cette fonction de paiement génère le gain utilisé dans les jeux de parité. Nous considérons à la fois les jeux de priorité stochastiques à tour de rôle et les jeux de priorité simultanés. Notre approche des jeux de priorité est basée sur le concept du point fixe le plus proche (« nearest fixed point ») des applications monotones non expansives et étend l'approche mu-calcul aux jeux de priorité.

La deuxième partie de la thèse concerne les questions de population. De manière simplifiée, nous examinons comment une distribution de probabilité sur les états évolue dans le temps. Plus précisément, nous sommes intéressés par des questions comme la suivante : à partir d'une distribution initiale, la population peut-elle atteindre à un moment donné une distribution avec une probabilité dépassant un seuil donné dans l'état visé ? Il s'avère que ce type de questions est beaucoup plus difficile à gérer que les questions concernant les trajectoires individuelles : on ne connaît pas, pour le modèle des chaînes de Markov, si les questions de population soient décidables. Nous étudions les restrictions des chaînes de Markov assurant la décision des questions de population.

Contents

1 Introduction **11**
1.1 Contributions 12
1.1.1 Part I: Priority games 12
1.1.2 Part II: Population questions 13

I Priority games **15**

2 Introduction **17**
2.1 Context - the parity games and μ-calculus 18
2.2 From parity games to priority games 21

3 On fixed points of bounded monotone nonexpansive mappings **25**
3.1 Fixed points of monotone nonexpansive mappings 26
3.2 Nested fixed points of bounded monotone nonexpansive mappings . . 29
3.3 Duality for the bounded monotone nonexpansive mappings 30

4 Turn-based stochastic priority games **35**
4.1 Preliminaries 37
4.2 Bounding the rewards 41
4.3 The one-step game 42
4.4 Nested nearest fixed point solution to priority games 43
4.4.1 Optimal strategy for player Max 48
4.4.2 Dual games 61
4.4.3 The duality of value mappings meets the duality of games . . 61
4.5 Remarks on priority games with infinite action or state sets 63

5 Concurrent stochastic priority games **67**
5.1 Concurrent stochastic priority games 69
5.2 Concurrent one-step game 72
5.3 General concurrent stopping priority games 73
5.4 Constructing ε-optimal strategies 74

5.4.1 $\varepsilon/2$-optimal strategy $\sigma_\star$ for player Max when $r_k < w_k$ and k is the starting state. 76
5.4.2 $\varepsilon/2$-optimal strategy $\tau_\star$ for player Min when $r_k \leqslant w_k$ and k is the starting state. 81
5.4.3 $\varepsilon/2$-optimal strategies for the other cases when the starting state is k 84
5.4.4 ε-optimal strategies for the $\varphi_r^{[k]}$-game starting at states $< k$. . . 84
5.4.5 Dual game 87

6 Discussion and conclusions **89**

II Population questions **91**

7 Analysing population dynamics of Markov chains **93**
7.1 Preliminaries and definitions 94
7.1.1 Motivation 95
7.1.2 Relation with the Skolem problem 96
7.1.3 Simple MCs 97
7.1.4 Trajectories and ultimate periodicity 98
7.2 Language of a MC 101
7.2.1 Partition of the set Init of initial distributions 103
7.2.2 High level description of the proof 104
7.3 Ultimate language 105
7.3.1 Limited number of switches. 105
7.3.2 Characterization of the ultimate language. 110
7.4 Regularity of the language 112
7.5 Discussion and conclusions 116

Bibliography **120**

Chapter 1

Introduction

Discrete time stochastic finite state systems can be modelled in many different ways. The simplest framework is provided by discrete homogeneous Markov chains which model systems evolving in time according to a fixed probabilistic transition function without any external control.

The systems with a single controller are modelled as Markov Decision Processes (MDP). In MDPs, the controller chooses at each stage an action to execute. The transition probability, that depends on the current state and on the executed action, describes how the system evolves in time. Markov chains can be seen as degenerate MDPs with only one action available in each state.

The next level of complexity is attained by two-player zero-sum games. Such games correspond to systems that are controlled by two controllers or two agents that have strictly opposite goals. The performance of each agent is measured through the payoff that he obtains. Zero-sum refers to the fact that for each game outcome, the gain of one player is equal to the loss of the other player. Two-player games can have different flavours:

- deterministic turn-based games where each state is controlled by one player who chooses the action to execute at this state and the transitions are deterministic,
- turn-based stochastic games where, again, each state is controlled by one player, but the transitions are probabilistic,
- concurrent stochastic games where at each state both players choose simultaneously and independently the actions to execute, and the probabilistic transition depends on both actions selected by the players.

Independently of whether the system evolves without any external control, or it is controlled by one, two or more agents, we can examine its behaviour from two different perspectives.

One point of view is that the system is at each stage in some state and this state evolves in stages. We can represent this situation as a single particle that moves from state to state according to a transition law, the movements influenced or controlled by the actions executed by the players or by controlling agents. In this framework (that we call *pebble semantics*), we are interested in the trajectory of the particle. This point of view is adopted in the first part of the thesis which is devoted to stochastic games.

Another point of view, namely *population semantics*, consists in seeing the system as composed of a whole population of particles spread over the states. The trajectory of a single particle is of no interest in this case, we are interested in how the distribution of the population evolves in time. This is the framework adopted in the second part of the thesis which examines population questions in Markov chains.

What is common to both parts of the thesis is that we deal uniquely with *quantitative* questions:

- in the first, part we examine the game value and the optimal and ε-optimal strategies of the players, in some infinite stochastic game,
- in the second part of the thesis, we examine if the population can reach a configuration where the proportion of the population in some goal states exceeds a given threshold.

This contrast with *qualitative* questions examined in computer science literature like, for example, the question if the probability of winning is positive, without specifying any concrete probability threshold. Here, each play is either winning or losing and the literature examines the existence of strategies which are surely winning, almost surely winning or winning with probability arbitrarily close to 1. Qualitative questions are outside the scope of the thesis.

1.1 Contributions

As mentioned above, the thesis consists of two parts.

1.1.1 Part I: Priority games

In Part I we examine stochastic zero-sum games with *priority payoff*.

The priority payoff is defined in the following way.

We assume that there is a total priority order over the states (we consider only games with a finite set of states) and that each state is labelled with a real valued reward. The priority payoff obtained for an infinite play is equal to the reward of the highest priority state seen infinitely often along this play. The priority payoff extends the payoff used in the parity games, a class of games extensively studied in computer science. The parity games are priority games with rewards in the two element set $\{0, 1\}$ rather than $\mathbb{R}$.

Part I consists of five chapters. We present an introduction of this part in Chapter 2. Chapter 3 is a short technical introduction to monotone nonexpansive mappings and their properties. We rely heavily on properties of such mappings in Chapters 4 and 5.

In Chapters 4 and 5 we study two classes of priority games.

In Chapter 4 we examine turn-based stochastic priority games where players play in turns, one after another.

Chapter 5 is devoted to concurrent priority games where at each stage players choose their actions simultaneously and independently.

Finally, in Chapter 6 we present the conclusions of Part I.

For turn-based stochastic priority games, we prove that both players have optimal memoryless strategies.

For concurrent priority games, optimal strategies do not exist in general and we construct ε-optimal strategies. Unfortunately, such ε-optimal strategies are not simple, to implement them the players need unbounded memory.

However, the crux of Chapters 4 and 5 does not lie in the fact that finite state priority games have values or in the fact that we can construct optimal or ε-optimal strategies. The main technical contribution is the powerful technique based on fixed points developed to obtain these results. A more technical and detailed discussion is postponed to the introduction of Part I. Preliminary version of the results obtained in Chapter 5 appears in [KZ15].

1.1.2 Part II: Population questions

In Part II, we will consider *population questions*. Suppose that a continuous population of agents is spread over the states of the system. A configuration is thus a distribution over the states and actions transform one distribution into another one. The general problem is thus to bring, by choosing the actions, the initial distribution of the population into particular configurations. For example we could be interested to bring at least half of the population in a set of *Goal* states. The questions concerning global probability distributions of a population of, say, some particles are considerably harder to tackle than the questions related to individual trajectory of one particle.

For instance it can be relatively easy to select a sequence of actions such that each particle will individually pass through some *Goal* state (or visit some *Goal* state periodically). On the other hand, if we consider a whole population of particles, it is undecidable in general whether there exists a strategy such that at least a half of particles will visit the same *Goal* state *at the same moment* [CKV+11]. The reason of this difficulty is that this question is equivalent to a quantitative undecidable question for finite probabilistic automata, [Paz71, Ber74].

We are interested in the following question. Given some initial distribution, or more generally some family of distributions, and some threshold γ, will the distribution reach a configuration where the fraction of the population in the *Goal* states is greater than γ? We study this problem from the symbolic dynamic perspective. We consider symbolic trajectories over the two letter alphabet $\{A, B\}$ describing the evolution of the distributions, where A represents configurations satisfying the threshold condition while B represents all other configurations. In this way the evolution of the distribution in time gives rise to an infinite word over the alphabet $\{A, B\}$. We define the language of the Markov chain to be the set of symbolic trajectories. We prove that if the eigenvalues of the Markov chain are distinct and positive, its symbolic language is regular and can be effectively computed. The findings presented in Chapter 7 appears in [AGKV16].

Part I

Priority games

Chapter 2

Introduction

This part of the thesis is devoted to a special class of zero-sum two-player stochastic games that we call stochastic priority games.

Stochastic two-player zero-sum games model the long-term interactions between two players that have strictly opposite objectives.

The study of stochastic games starts with the seminal paper of Shapley [Sha53]. Since then, the subject was intensively studied in game theory where it is seen as a special case of a more general model of repeated games. Repeated games are exhaustively treated in two monographs [Sor02, JFM15], both of them contain chapters devoted to stochastic games. As the books specifically devoted to stochastic games we can mention [FV97, NS04].

In computer science stochastic games were first examined from the algorithmic point of view where the aim is to find an efficient algorithm that computes optimal or ε-optimal strategies for both players. In this line of research, initiated by the paper of Hoffman and Karp [HK66], we are interested in "algorithmically implementable" optimal strategies which means that the strategies should be either memoryless (i.e. stationary) or their implementation should use a bounded memory. One of the most challenging open questions in this domain concerns the existence of a polynomial time algorithm solving so-called simple stochastic games. This is the simplest class of turn-based stochastic games, examined already in [HK66]. The problem of finding optimal strategies for these games is known to be in $NP \cap coNP$, [Con92], but no polynomial time algorithm is known.

Since this part of the thesis concerns games that are closely related to the so called parity games we should mention here that most recent achievement in this domain is a quasi-polynomial time algorithm solving deterministic parity games [CJK+16].

Another track of research involving games is motivated by applications to automata theory, logic and verification. This can be traced down to the groundbreaking paper of Gurevich and Harrington [GH82], where games were used in order to simplify the solution to the important complementation problem for automata on infinite trees. Initially this research was limited to deterministic games [1], see the collective volume [GTW02] for a presentation of the field. Problems related to the verification of probabilistic programs and systems motivated subsequent extensions based on stochastic game models. First the verification problem for one-player stochastic systems (Markov Decision Processes) was considered, see [dA97], next turn-based stochastic two-player games were examined [MM02, CJH04] and finally concurrent stochastic games were explored [2] [dAM04].

In stochastic games the players preferences are expressed by means of a payoff mapping. The payoff mapping maps infinite plays (infinite sequences of states and actions) to real numbers. The payoff mappings used in computer science tend to be different from the traditional payoff mappings used in game theory. The payoffs prevalent in computer science are often expressed in some kind of logic and the corresponding payoff mappings take only two values, 1 for the winning plays and 0 for the losing plays.

On the other hand, the payoff mappings used in game theory are rather real valued: mean-payoff, discounted payoff, $\limsup$ and $\liminf$ payoffs are among the most popular ones.

In this thesis we define and examine the class of priority games. The priority games constitute a natural extension of parity games, this latter class is the class of games popular in computer science having applications in automata theory and verification.

To put the results of the thesis in the context let us recall the relevant results concerning the parity games.

2.1 Context - the parity games and μ-calculus

A stochastic zero-sum two-player game is an infinite game played by two players, player Max and player Min, on an arena with a finite set of states $\mathbf{S}$ and a finite set of actions $\mathbf{A}$ (the games where one or both of these sets are infinite are beyond the scope of the thesis). Turn-based stochastic games and concurrent stochastic games differ in the law of motion that specifies how the game moves from one state to another in function of the actions played by the players.

In turn-based stochastic games each state is controlled by one of the players. The dynamical aspect of the system is captured by the family of probability distributions

1. perfect information games with deterministic transitions

2. The terms "turn-based stochastic games" and "concurrent stochastic games" are commonly used in Computer Science. In game theory these classes of games are called respectively "perfect information stochastic games" and "stochastic games". Thus, in particular "stochastic games" without any other qualifier refers to concurrent stochastic games.

$p(\cdot|i,a)$, where for state $i \in \mathbf{S}$ and action $a \in \mathbf{A}$, $p(j|i,a)$ is the probability to move to state j when the player controlling the current state i executes a. It is assumed that both players know all the history (sequences of visited states and played actions) of the game up to the current moment.

In concurrent stochastic games it is rather the case that both players control collectively the transitions. More specifically, in concurrent stochastic games, for each state i, both Max and Min have nonempty sets of available actions, $\mathbf{A}(i)$ and $\mathbf{B}(i)$ respectively. At each stage, the players, knowing the current state and all the previous history, choose independently and simultaneously actions $a \in \mathbf{A}(i)$ and $b \in \mathbf{B}(i)$ respectively and the game moves to state j with probability $p(j|i,a,b)$. Immediately after each stage, and before the next one, both players are informed about the action played by the adversary player.

Thus in the concurrent stochastic games the transition mapping assigns to each state i and to actions $a \in \mathbf{A}(i)$, $b \in \mathbf{B}(i)$, a probability distribution $p(\cdot|i,a,b)$ over states.

We assume that players play an infinite game. At each stage either one of the players, in the case of the turn-based stochastic games, or both players, for the concurrent stochastic games, choose action and the game moves to another state according to the transition probability.

An infinite sequence of states and action occurring during the game is called a play.

Since we are interested in finite state games, without loss of generality we assume in the sequel that the set of states is $\mathbf{S} = [n] = \{1, \ldots, n\}$.

Parity games are endowed with the reward vector $r = (r_1, \ldots, r_n)$, where $r_i \in \{0, 1\}$ is the reward of state i. The parity payoff $\varphi(h)$ of an infinite play h is defined to be equal [3] to the reward of the maximal state visited infinitely often in h, i.e. the payoff is equal to r_i if i was visited infinitely often in h and all states $j, j > i$, were visited only a finite number of times. This definition of the parity payoff is the same for all classes of parity games: deterministic parity games, turn-based stochastic parity games and concurrent parity games, the only difference between these three types of games lies in their transition mappings.

A strategy of a player is a mapping $\sigma : H \to \Delta(\mathbf{A})$, where $\Delta(\mathbf{A})$ denotes the set of probability distributions over $\mathbf{A}$. We will define more precisely the strategies for turn-based stochastic games in Chapter 4 and for concurrent games in Chapter 5.

The set of all plays is endowed in the usual way with the Borel σ-algebra generated by the cylinders. Strategies σ, τ of players Max and Min and an initial state

3. The payoff of the parity game is usually formulated in a bit different way: The states are a finite subset of natural numbers and reward of state i is equal to 0 if i is even and 1 otherwise. However it is easy to see that our definition is equivalent to the usual one by just renaming the states.

$i \in \mathbf{S}$ give rise to a probability measure $\mathbf{P}_i^{\sigma,\tau}$ over the Borel σ-algebra. The aim of player Max (respectively Min) is to maximize (respectively minimize) the expected payoff

$$\mathbf{E}_i^{\sigma,\tau}(\varphi) = \int \varphi(h)\mathbf{P}_i^{\sigma,\tau}(dh)$$

for each initial state i.

Since the parity payoff is Borel measurable, by the result of Martin [Mar98], parity games have value v_i for each initial state i, i.e.

$$\sup_\sigma \inf_\tau \mathbf{E}_i^{\sigma,\tau}(\varphi) = v_i = \inf_\tau \sup_\sigma \mathbf{E}_i^{\sigma,\tau}(\varphi), \quad \forall i \in \mathbf{S}. \tag{2.1}$$

Moreover, for deterministic and for turn-based stochastic parity games both players have optimal pure memoryless strategies, see for example [EJ91, Zie98, Wal02], where the deterministic parity games are examined, and [CJH04] for turn-based stochastic parity games.

One of the techniques used to solve parity games relies on the μ-calculus. In this approach the point of departure is a simple one-step game[4] played at each state $i \in \mathbf{S}$. The one-step game has a value for each state $i \in \mathbf{S}$ and each reward vector $r = (r_1, \ldots, r_n)$. Let

$$f = (f_1, \ldots, f_n) \tag{2.2}$$

be the mapping that maps the reward vectors $r \in \{0,1\}^n$ to the vector of values of the one-step games, i.e. for $r = (r_1, \ldots, r_n)$ and $i \in \mathbf{S}$, $f_i(r)$ is the value of the one-step game played at state i given the reward vector r. We endow $[0,1]^n$ with the product order, $x = (x_1, \ldots, x_n) \leqslant (y_1, \ldots, y_n) = y$ if $x_i \leqslant y_i$ for all $i \in [n]$, which makes it a complete lattice. It is easy to see that

$$f : [0,1]^n \to [0,1]^n$$

is monotone under $\leqslant$, thus by Tarski's theorem [Tar55], f has the least and the greatest fixed points.

Then one defines the nested fixed point

$$\mathbf{Fix}^n(f)(r) = \mu_{r_n} x_n . \mu_{r_{n-1}} x_{n-1} . \ldots \mu_{r_2} x_2 . \mu_{r_1} x_1 . f(x_1, x_2, \ldots, x_{n-1}, x_n), \tag{2.3}$$

where $\mu_{r_i} x_i$ denotes either the greatest fixed point if $r_i = 1$ or the least fixed point if $r_i = 0$ and f is the one-step value function (2.2). The main result obtained in

4. The term "one-step game" is commonly used in game theory. In computer science one-step games are not named explicitly, but their value function f is used in the μ-calculus approach to parity games, where is often called the predecessor operator.

the μ-calculus approach to concurrent stochastic parity games due to de Alfaro and Majumdar [dAM04], is that

$$v = (v_1, \dots, v_n) = \mathbf{Fix}^n(f)(r),$$

where the left-hand side vector v is composed of the values v_i for the parity game starting at i, cf. (2.1). To summarize, the value vector of the parity game can be obtained by calculating the nested fixed point of the one-step value mapping[5].

Let us note that for deterministic parity games (turn-based games with deterministic transitions) the μ-calculus representation simplifies since the one-step value mappings f_i map the binary vectors $\{0,1\}^n$ to $\{0,1\}$ and the parity games can be treated in the framework of the boolean μ-calculus [Wal02, AN01]. Since in the thesis we do not consider the deterministic games we omit the more detailed discussion of deterministic parity games.

2.2 From parity games to priority games

The parity games (as well as other related classes of games like the games with the Muller or Rabin winning conditions) arose from the study of decidability questions in logic. In this framework the winning criteria are expressed in some kind of logic, where there is room for only two types of plays, the winning plays that satisfy a logical formula and the losing plays that do not satisfy the formula. For this reason the rewards in the parity games take only two values, 0 and 1, with the intuition that the reward 1 is favourable and the reward 0 unfavourable for our player (and the preferences are inverse for the adversary player).

However, the restriction to 0, 1 rewards does not allow to express finer player's preferences. This motivates the study of the games that allow any real valued rewards. We define the priority game as the game where each state $i \in [n] = \mathbf{S}$ is equipped with a reward $r_i \in \mathbb{R}$. Like in parity games the payoff $\varphi(h)$ of a play h is defined to be the reward r_i of the greatest state i that is visited infinitely often in h.

At first glance, the priority games are just a mild extension of parity games. This impression is reinforced by the fact that deterministic priority games, which we do not consider in the thesis, can be reduced to deterministic parity games. However,

5. The traditional presentation of this result is a bit different. Roughly speaking the variables are regrouped in blocks, each block consists of consecutive variables to which the same fixed point is applied. In this way the fixed points are applied to the groups of variables rather than to each variable separately. This allows to decreases the number of fixed points and the resulting formula alternates the least and the greatest fixed points. However, this is only a technical detail which has no bearing on the result. For our purposes it is more convenient to apply fixed points to variables rather than to groups of variables.

we do not know if such reduction is possible for stochastic (turn-based or concurrent) priority games.

The interest in priority games is twofold. First, the priority games allow to quantify players' preferences in a more subtle way than it is possible in parity games. While in parity games there are only two classes of plays, the plays with the parity payoff 1 and the plays with the parity payoff 0, in priority games we can distinguish many levels of preferences. As a motivating simple example consider the priority game with three states $\mathbf{S} = \{1, 2, 3\}$ and rewards $r_1 = 0, r_2 = 1, r_3 = \frac{3}{4}$. This game gives rise to three distinct classes of infinite plays: player Max highest preference is for the plays such that the maximal state visited infinitely often is state 2 (plays give him the payoff 1), his second preference is for the plays that visit state 3 infinitely often (these plays give him the payoff $\frac{3}{4}$), and his lowest preference is for the plays that from some moment onward stay forever in state 1 (they give him payoff 0). It is impossible to capture such a hierarchy of preferences when we limit ourselves to the parity payoff.

The second reason to be interested in priority games stems from the fact that not only they generalize parity games, but they contain as proper subclasses two other well known families of stochastic games: the lim sup and lim inf payoff games [MS04]. This point will be discussed in Section 5.1.

Our approach to priority games is inspired by the μ-calculus approach to parity games. There are two major differences however.

It is impossible to solve the priority games using only the least and the greatest fixed points, we need also other fixed points that we name "the nearest fixed points". To define this notion we use the well known fact that the one-step game value mapping (2.2) is not only monotone but it is also nonexpansive, which means that, for $x, y \in \mathbb{R}^n$, $\|f(x) - f(y)\|_\infty \leqslant \|x - y\|_\infty$, where $\|x\|_\infty = \sup_i |x_i|$ is the supremum norm. Let us note that this property of the one-step games is used in the study of stochastic mean-payoff games [BK76, Ney03].

In the study of parity games the fact that the one-step game value mapping f is nonexpansive is irrelevant, the monotonicity of f is all that we need in order to apply Tarski's fixed point theorem. When we study the priority games, when other fixed points enter into consideration, the monotonicity of f is not sufficient and the fact that f is nonexpansive becomes paramount.

Our study of priority games is organized as follows.

It turns out that the priority games with rewards in $\mathbb{R}$ can be reduced through a linear transformation to the priority games with rewards in the interval $[0, 1]$. Therefore in the sequel we assume that the reward vector $r = (r_1, \ldots, r_n)$ belongs to $[0, 1]^n$. Under this condition value mapping f of the one-step game (2.2) is a monotone nonexpansive mapping from $[0, 1]^n$ to $[0, 1]^n$. Since our study of priority games is based on the analysis of the fixed points of f, in Chapter 3 we prepare the background and present basic facts concerning fixed points of monotone nonexpansive mappings from $[0,1]^n$ to $[0,1]^n$. All the facts presented in Chapter 3 are either well known or are rather straightforward observations. The purpose of Chapter 3 is to regroup in one place all the facts that we need in the sequel and to introduce the notion of the r-nearest fixed point

$$\mu_r x.g(x)$$

of the monotone nonexpansive mapping $g : [0,1] \to [0,1]$. Intuitively, $\mu_r x.g(x)$ is the fixed point of g which is nearest to $r \in [0,1]$. Note that the least and the greatest fixed points of g are special cases of this notion, the greatest fixed point is the fixed point nearest to 1 and the least fixed point is the fixed point nearest to 0. We show that the notion of the nearest fixed point makes sense for monotone nonexpansive mappings from $[0,1]$ to $[0,1]$. In Chapter 3 we define also, for each vector $r = (r_1,\ldots,r_n) \in [0,1]^n$ and a monotone nonexpansive mapping $f : [0,1]^n \to [0,1]^n$, the nested r-nearest fixed point

$$\mathbf{Fix}^n(f)(r) = \mu_{r_n} x_n.\mu_{r_{n-1}} x_{n-1}.\ldots\mu_{r_2} x_2.\mu_{r_1} x_1.f(x_1,x_2,\ldots,x_{n-1},x_n), \tag{2.4}$$

which generalizes the nested least/greatest fixed point (2.3).

Chapter 4 is devoted to the study of turn-based stochastic priority games. The main result of this chapter is that, given the reward vector $r = (r_1,\ldots,r_n)$, the value vector $v = (v_1,\ldots,v_n)$ of the turn-based stochastic priority game can be expressed as the nested r-nearest fixed point

$$v = (v_1,\ldots,v_n) = \mathbf{Fix}^n(f)(r) \tag{2.5}$$

of the value mapping f of the one-step game. Moreover, we prove that both players have optimal pure memoryless strategies.

Chapter 5 examines concurrent stochastic priority games. We prove that the r-nearest fixed point characterization (2.5) of the value vector holds also for concurrent priority games. However, in general the players have only ε-optimal history dependent strategies.

Although the results of Chapters 4 and 5 can be seen as extensions of the μ-calculus characterization known for parity games [MM02, dAM04] there is one more point that distinguish our approach from the traditional μ-calculus approach to parity games. In the case of parity games, to the best of our knowledge, the μ-calculus proofs presented previously were not inductive. In previous proofs a formula similar to (2.3) was announced and it was shown, in one big step, that this formula yields the value of the parity game [6].

6. Such single big step proofs characterize also the μ-calculus approach to deterministic parity games [Wal02]. In retrospect, what was lacking in previous proofs was a game interpretation of the partial fixed point, where some variables remain free.

The fact that the nested fixed point formula (2.3) is in some sense recursive, was not exploited to the full extent in the proof.

The novelty of the proofs presented in Chapters 4 and 5 lies in the fact that they are genuinely inductive. We provide a clear game theoretic interpretation of the partial fixed point formula

$$\mathbf{Fix}^k(f)(r) = \mu_{r_k} x_k. \dots \mu_{r_1} x_1. f(x_1, \dots, x_k, r_{k+1}, \dots, r_n), \tag{2.6}$$

where the fixed points are applied only to the low priority variables $x_1, \dots, x_k$, while the free variables $x_{k+1}, \dots, x_n$ take values $r_{k+1}, \dots, r_n$ respectively.

Let $G(r)$ be the priority game endowed with the reward vector r. Let $G_k(r)$ be the priority game obtained from $G(r)$ by transforming all states $i, i > k$, into absorbing states[7]. On the other hand, the states $j, j \leqslant k$, have the same transitions in $G(r)$ as in $G_k(r)$.

It turns out that the partial nested fixed point (2.6) is equal to the value vector $v = (v_1, \dots, v_n)$ of the priority game $G_k(r)$. We prove this fact by induction, starting with the trivial priority game $G_0(r)$, where all states are absorbing. And the inductive step consist in showing that, if (2.6) is the value of the game $G_k(r)$, then adding the new fixed point $\mu_{r_{k+1}} x_{k+1}$ we obtain the value vector of the game $G_{k+1}(r)$. In other words, adding one fixed point corresponds to the transformation of an absorbing state into a nonabsorbing one. Note that in priority games the absorbing states are trivial, if a state m is absorbing then $v_m = r_m$, i.e. the value of m is equal to the reward r_m. Thus transforming an absorbing state into a nonabsorbing we convert a trivial state into a nontrivial one. The crucial point is that in the inductive proof given in the thesis we apply this transformation to just one state. And it is much easier to understand what happens if one state changes its quality from absorbing to nonabsorbing than when all states are nonabsorbing from the outset.

7. Recall that a state i is absorbing if it is impossible to leave i, i.e. for all possible actions executed in i the game remains in i with the probability 1.

Chapter 3

On fixed points of bounded monotone nonexpansive mappings

In this technical chapter, we introduce monotone nonexpansive mappings, that play a crucial role in the study of stochastic priority games. The solution to stochastic turn-based and concurrent priority games given in Chapters 4 and 5 relies heavily on fixed point properties of such mappings examined in Section 3.1. In Section 3.2 we define and examine the nested nearest fixed points of monotone nonexpansive mappings.

The duality of the nested nearest fixed points is studied in Section 3.3.

An element $x = (x_1, \ldots, x_n)$ of $\mathbb{R}^n$ will be identified with the mapping x from $[n] = \{1, \ldots, n\}$ to $\mathbb{R}$ and we can occasionally write $x(i)$ to denote x_i.

The set $\mathbb{R}^n$ is endowed with the natural componentwise order, for $x, y \in \mathbb{R}^n$, $x \leqslant y$ if $x_i \leqslant y_i$ for all $i \in [n]$.

A mapping $f : \mathbb{R}^n \to \mathbb{R}^k$ is *monotone* if for $x, y \in \mathbb{R}^n$, $x \leqslant y$ implies $f(x) \leqslant f(y)$ (we do not assume that $k = n$, thus $x \leqslant y$ and $f(x) \leqslant f(y)$ can relate to componentwise orders in two different spaces).

We assume that the Cartesian product $\mathbb{R}^n$ is endowed with the structure of a normed real vector space with the norm $\|\cdot\|_\infty$, for $x \in \mathbb{R}^n$, $\|x\|_\infty = \max_{i\in[n]}|x_i|$. Thus, for $x, y \in \mathbb{R}^n$, $\|x - y\|_\infty$ defines a distance between x and y.

We say that a mapping $f : \mathbb{R}^n \to \mathbb{R}^k$ is *nonexpansive* if, for all $x, y \in \mathbb{R}^n$, $\|f(x) - f(y)\|_\infty \leqslant \|x - y\|_\infty$.

Such a mapping f can be written as vector of k mappings $f = (f_1, \ldots, f_k)$, where $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, k$. Clearly, f is monotone nonexpansive iff all f_i are monotone nonexpansive.

We say that a mapping $f : \mathbb{R}^n \to \mathbb{R}^k$ is *additive homogeneous* if for all $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^n$

$$f(x + \lambda e_n) = f(x) + \lambda e_k,$$

where e_n and e_k are the vectors $(1, \ldots, 1)$ in $\mathbb{R}^n$ and $\mathbb{R}^k$ respectively having all components equal to 1.

Crandall and Tartar [CT80] proved the following result.

Example 3.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ the max function such that for all $x \in \mathbb{R}^n$, $\max(x) = \max(x_1, \ldots, x_n)$ and let $g : \mathbb{R}^n \to \mathbb{R}$ the zero function such that for all $x \in \mathbb{R}^n$, $g(x) = 0$.

Remark that both f and g are nonexpansive and f is also additive homogeneous, but g is not additive homogeneous because for any $x \in \mathbb{R}^n$ and $\lambda \neq 0$,

$$0 = g(x + \lambda e_n) \neq g(x) + \lambda > 0.$$

Lemma 3.2 (Crandall and Tartar [CT80])**.** *For additive homogeneous mappings $f : \mathbb{R}^n \to \mathbb{R}^k$ the following conditions are equivalent:*

(i) f is monotone,

(ii) f is nonexpansive.

We will need only the implication (i)→(ii) that we prove below for the reader's convenience. Moreover, if the result holds for mappings from $\mathbb{R}^n$ to $\mathbb{R}$ then it holds for mappings from $\mathbb{R}^n$ to $\mathbb{R}^k$. Thus we assume in the proof that that $f : \mathbb{R}^n \to \mathbb{R}$.

Proof. For $x, y \in \mathbb{R}^n$, $e_n = (1, 1, \ldots, 1) \in \mathbb{R}^n$ and $\lambda = \|x - y\|_\infty$ we have $y - \lambda e_n \leqslant x \leqslant y + \lambda e_n$. Thus for $f : \mathbb{R}^n \to \mathbb{R}$ monotone and additive homogeneous we obtain

$$f(y) - \lambda \leqslant f(x) \leqslant f(y) + \lambda.$$

Thus $|f(x) - f(y)| \leqslant \lambda = \|x - y\|_\infty$. □

3.1 Fixed points of monotone nonexpansive mappings

We say that a monotone mapping $f : \mathbb{R}^n \to \mathbb{R}^k$ is *bounded* if $f([0, 1]^n) \subseteq [0, 1]^k$.

The set of bounded monotone nonexpansive mappings will be denoted by $M_{n,k}[0, 1]$. Moreover BMN will stand for the abbreviation for "bounded monotone nonexpansive".

In this section we introduce the notion of the nearest fixed point of BMN mappings generalizing the least and greatest fixed points.

In the following lemma states basic properties of fixed points of BMN mappings.

Lemma 3.3. *Let $f \in M_{1,1}[0,1]$. Define by induction, $f^{(0)}(x) = x$, $f^{(1)}(x) = f(x)$, $f^{(i+1)}(x) = f(f^{(i)}(x))$, for $x \in [0,1]$.*

Then

(i) for each $x \in [0,1]$ the sequence $(f^{(i)}(x)), i = 0, 1, \ldots,$ is monotone and converges to some $x^\infty \in [0,1]$. The limit x^∞ is a fixed point of f, $f(x^\infty) = x^\infty$,

(ii) if $x \leqslant y$ are fixed points of f, $f(x) = x$ and $f(y) = y$, then for each z such that $x \leqslant z \leqslant y$, $f(z) = z$,

(iii) the sequence $(f^{(i)}(0)), i = 0, 1, 2, \ldots,$ converges to the least fixed point $\bot_f$ of f while the sequence $(f^{(i)}(1)), i = 0, 1, 2, \ldots,$ converges to the greatest fixed point $\top_f$ of f. The interval $[\bot_f, \top_f]$ is the set of all fixed points of f.
If $0 \leqslant x \leqslant \bot_f$ then the sequence $(f^{(i)}(x))$ converges to $\bot_f$.
If $\top_f \leqslant x \leqslant 1$ then the sequence $(f^{(i)}(x))$ converges to $\top_f$.
If $0 \leqslant x < \bot_f$ then $x < f(x)$.
If $\top_f < x \leqslant 1$ then $f(x) < x$.

Proof. (i) Suppose that $f(x) \leqslant x$. Then inductively, since f is non-increasing, $f^{(i+1)}(x) \leqslant f^{(i)}(x)$ for all i, i.e. the sequence $f^{(i)}(x)$ is non-increasing. Since this sequence is bounded from below by 0 it converges to some x^∞.

The case of $f(x) \geqslant x$ can be treated in a similar way.

Since f is nonexpansive $|f(x^\infty) - f^{(i+1)}(x)| \leqslant |x^\infty - f^{(i)}(x)|$. As the right-hand side tends to 0 we can see that $f^{(i)}(x)$ converges to $f(x^\infty)$. On the other hand, $f^{(i)}(x)$ converges to x^∞. Therefore $f(x^\infty) = x^\infty$.

(ii) Let $0 \leqslant x \leqslant z \leqslant y \leqslant 1$ and $f(x) = x$, $f(y) = y$. Since f is monotone, $x = f(x) \leqslant f(z) \leqslant f(y) = y$. Thus, since f is nonexpansive, $0 \leqslant f(y) - f(z) \leqslant y - z$ and $0 \leqslant f(z) - f(x) \leqslant z - x$. This implies that $f(z) = z$.

(iii) is a direct consequence of (i) and (ii). □

Let $f \in M_{1,1}[0,1]$. For $a \in [0,1]$ we define *the a-nearest fixed point* of f to be

$$\mu_a x.f(x) := \lim_i f^{(i)}(a).$$

Lemma 3.3 shows that this is really a fixed point of f which is closest to a, i.e. $|a - \mu_a x.f(x)| = \min_{z \in [0,1]}\{|a - z| \mid f(z) = z\}$.

Moreover, the least and the greatest fixed points of $f \in M_{1,1}[0,1]$ are respectively equal to $\mu_0 x.f(x)$ and $\mu_1 x.f(x)$.

We can see also that

$$\mu_a x.f(x) = \begin{cases} \mu_0 x.f(x) & \text{if } a \leqslant \mu_0 x.f(x), \\ a & \text{if } \mu_0 x.f(x) \leqslant a \leqslant \mu_1 x.f(x), \\ \mu_1 x.f(x) & \text{if } \mu_1 x.f(x) \leqslant a, \end{cases} \tag{3.1}$$

i.e. the fixed point nearest to a is equal either to the least or to the greatest fixed point or is equal to a itself.

Let $f : (x_1, \ldots, x_n) \mapsto f(x_1, \ldots, x_n)$ be a BMN mapping from $[0,1]^n$ to $[0,1]$. For each $(r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_n) \in [0,1]^{n-1}$ we obtain a BMN mapping

$$x_k \mapsto f(r_1, \ldots, r_{k-1}, x_k, r_{k+1}, \ldots, r_n).$$

from $[0,1]$ to $[0,1]$. This mapping belongs to $M_{1,1}[0,1]$ thus, given $r_k \in [0,1]$, we can calculate the r_k-nearest fixed point

$$\mu_{r_k} x_k . f(r_1, \ldots, r_{k-1}, x_k, r_{k+1}, \ldots, r_n).$$

This fixed point depends on $r = (r_1, \ldots, r_{k-1}, r_k, r_{k+1}, \ldots, r_n)$, thus we can define the mapping

$$[0,1]^n \ni (r_1, \ldots, r_{k-1}, r_k, r_{k+1}, \ldots, r_n) \mapsto \mu_{r_k} x_k . f(r_1, \ldots, r_{k-1}, x_k, r_{k+1}, \ldots, r_n) \in [0,1] \tag{3.2}$$

Lemma 3.4. *If $(x_1, \ldots, x_n) \mapsto f(x_1, \ldots, x_n)$ is BMN then the mapping (3.2) is BMN.*

Proof. Let $r = (r_1, \ldots, r_n), w = (w_1, \ldots, w_n) \in [0,1]^n$. Define two sequences $(r_k^i), i = 1, 2, \ldots$ and $(w_k^i), i = 1, 2, \ldots$, such that

$$r_k^1 = r_k \quad \text{and} \quad r_k^{i+1} = f(r_1, \ldots, r_{k-1}, r_k^i, r_{k+1}, \ldots, r_n)$$

and

$$w_k^1 = w_k \quad \text{and} \quad w_k^{i+1} = f(w_1, \ldots, w_{k-1}, w_k^i, w_{k+1}, \ldots, w_n).$$

By Lemma 3.3 both sequences converge to some r_k^∞ and w_k^∞ respectively and

$$r_k^\infty = \mu_{r_k} x_k . f(r_1, \ldots, r_{k-1}, x_k, r_{k+1}, \ldots, r_n)$$

and

$$w_k^\infty = \mu_{w_k} x_k . f(w_1, \ldots, w_{k-1}, x_k, w_{k+1}, \ldots, w_n).$$

We shall prove by induction that for all i, $|r_k^i - w_k^i| \leqslant \|r - w\|_\infty$.

Clearly, $|r_k^1 - w_k^1| = |r_k - w_k| \leqslant \max_i |r_i - w_i| = \|r - w\|_\infty$. Suppose that

$$|r_k^i - w_k^i| \leqslant \|r - w\|_\infty.$$

We have then

$$|r_k^{i+1} - w_k^{i+1}| = |f(r_1, \ldots, r_{k-1}, r_k^i, r_{k+1}, \ldots, r_n) - f(w_1, \ldots, w_{k-1}, w_k^i, w_{k+1}, \ldots, w_n)| \leqslant \max\{\max_{j \neq k} |r_j - w_j|, |r_k^i - w_k^i|\} \leqslant \max\{\max_{j \neq k} |r_j - w_j|, \|r - w\|_\infty\} = \|r - w\|_\infty.$$

Taking the limit $i \nearrow \infty$ we obtain $|r_k^\infty - w_k^\infty| \leqslant \|r - w\|_\infty$.

□

Lemma 3.5. *If $f \in M_{k,m}[0,1]$ and $g \in M_{m,n}[0,1]$ then $g \circ f \in M_{k,n}[0,1]$, i.e. the composition of BMN mappings is BMN.*

Proof. For $x, y \in [0,1]^k$, we have $\|g(f(x)) - g(f(y))\|_\infty \leqslant \|f(x) - f(y)\|_\infty \leqslant \|x - y\|_\infty$ i.e. composition of nonexpansive mappings is nonexpansive. Trivially, monotonicity is also preserved by composition. □

3.2 Nested fixed points of bounded monotone non-expansive mappings

In this section we define by induction, for each k, $0 \leqslant k \leqslant n$, the nested fixed point operator.

We define by induction for each k, $0 \leqslant k \leqslant n$, the *nested nearest fixed point operator*

$$\mathbf{Fix}^k : M_{n,n}[0,1] \to M_{n,n}[0,1].$$

Each $\mathbf{Fix}^k$ can be decomposed into n operators $\mathbf{Fix}^k_i$,

$$\mathbf{Fix}^k_i : M_{n,n}[0,1] \to M_{n,1}[0,1], \quad i \in [n],$$

such that, for $f \in M_{n,n}$,

$$\mathbf{Fix}^k(f) = (\mathbf{Fix}^k_1(f), \ldots, \mathbf{Fix}^k_n(f)).$$

Let $f = (f_1, \ldots, f_n) \in M_{n,n}[0,1]$, where $f_i \in M_{n,1}[0,1]$, for $i \in [n]$.

For all $r \in [0,1]^n$ we set $\mathbf{Fix}^0(f)$ to be such that

$$\mathbf{Fix}^0(f)(r) = r.$$

Thus $\mathbf{Fix}^0(f)$ is the identity mapping and does not depend of f. Note that $\mathbf{Fix}^0_i(f)(r) = r_i$, i.e. $\mathbf{Fix}^0_i(f)$ is the projection on the ith coordinate.

Now, inductively, given $\mathbf{Fix}^{k-1}(f)$ we define $\mathbf{Fix}^k(f)$.

For $r \in [0,1]^n$ and $\zeta \in [0,1]$ let us set

$$F^{k-1}_i(\zeta; r) := \mathbf{Fix}^{k-1}_i(f)(r_1, \ldots, r_{k-1}, \zeta, r_{k+1}, \ldots, r_n), \quad \text{for } i \in [k-1]. \tag{3.3}$$

Note that $F^{k-1}_i(\zeta; r)$ depends on ζ and on $(r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_n)$ but does not depend on r_k. Thus F^{k-1}_i is in fact a mapping from $[0,1]^n$ to $[0,1]$.

Then we define

$$\mathbf{Fix}^k_k(f)(r) := \mu_{r_k}\zeta.f_k(F^{k-1}_1(\zeta; r), \ldots, F^{k-1}_{k-1}(\zeta; r), \zeta, r_{k+1}, \ldots, r_n), \tag{3.4}$$

$$\mathbf{Fix}^k_i(f)(r) := \mathbf{Fix}^{k-1}_i(f)(r_1, \ldots, r_{k-1}, \mathbf{Fix}^k_k(f)(r), r_{k+1}, \ldots, r_n), \quad \text{for } i \in [k-1],$$

$$\mathbf{Fix}^k_i(f)(r) := r_i, \quad \text{for } i \in \{k+1, \ldots, n\}.$$

Since the definition of the nested fixed point mappings uses only the composition and the nearest fixed point operators, Lemmas 3.5 and 3.4 imply that

Corollary 3.6. *If* $f \in M_{n,n}[0,1]$ *then, for all* $k \in \{0\} \cup [n]$, $\mathbf{Fix}^k(f) \in M_{n,n}[0,1]$.

Let us note finally that $\mathbf{Fix}^k(f)$ depends only on $f_1, \ldots, f_k$ but is independent of $f_{k+1}, \ldots, f_n$.

Example 3.7. Let $n = 2$ and $f = (f_1, f_2) : M_{2,2}[0,1]$ such that for all $x = (x_1, x_2) \in [0,1]^2$, $f_1(x_1, x_2) = max(x_1, x_2)$, $f_2(x_1, x_2) = x_1$ and let $r = (r_1, r_2) = (0, 1)$.

Let us calculate the value of $\mathbf{Fix}^2$ inductively, for $k = 0$ we have $\mathbf{Fix}^0(f)(r) = (0, 1)$.

For $k = 1$,

$$\mathbf{Fix}^1_1(f)(r) = \mu_0\zeta.f_1(\zeta, 1) = \mu_0\zeta.\max(\zeta, 1) = 1, \text{ and}$$

$$\mathbf{Fix}^1_2(f)(r) = r_2 = 1.$$

Finally, with $k = 2$,

$$\mathbf{Fix}^2_2(f)(r) = \mu_1\zeta.f_2(F^1_1(\zeta, 1), \zeta).$$

So we need to calculate the value of $F^1_1(\zeta, 1)$:

$$F^1_1(\zeta, 1) = \mathbf{Fix}^1_1(f)(\zeta, 1) = \mu_0\zeta.f_1(\zeta, 1) = 1.$$

Then $\mathbf{Fix}^2_2(f)(r) = 1$ and $\mathbf{Fix}^2_1(f)(r) = \mathbf{Fix}^1_1(f)(0, 1) = \mu_0\zeta.f_1(\zeta, 1) = 1$.
Hence,

$$\mathbf{Fix}^2(f)(r) = (1, 1).$$

3.3 Duality for the bounded monotone nonexpansive mappings

In this chapter we define and examine the notion of duality for the BMN mappings.

For $r = (r_1, \ldots, r_n) \in [0,1]^n$ we set $1 - r := (1 - r_1, \ldots, 1 - r_n)$.

Given a BMN mapping $f : [0,1]^n \to [0,1]$ the *dual* of f is the mapping $\overline{f} : [0,1]^n \to [0,1]$ such that

$$\overline{f}(r_1, \ldots, r_n) = 1 - f(1 - r_1, \ldots, 1 - r_n).$$

The dual of $f = (f_1, \ldots, f_k) \in M_{n,k}[0,1]$ is defined as $\overline{f} = (\overline{f}_1, \ldots, \overline{f}_n)$.

We can write this in a more explicit way if for $f = (f_1, \ldots, f_k) \in M_{n,k}[0,1]$ we define $1 - f := (1 - f_1, \ldots, 1 - f_k)$.

Then using this notation, for $f \in M_{n,k}[0,1]$, we can write succinctly

$$\overline{f}(r) = 1 - f(1 - r).$$

Lemma 3.8. *If f is BMN then $\overline{f}$ is BMN.*

Proof. Let $(r_1, \ldots, r_n) \leqslant (w_1, \ldots, w_n)$.

Then $(1 - r_1, \ldots, 1 - r_n) \geqslant (1 - w_1, \ldots, 1 - w_n)$ and $f(1 - r_1, \ldots, 1 - r_n) \geqslant f(1 - w_1, \ldots, 1 - w_n)$.

Thus $\overline{f}(r_1, \ldots, r_n) = 1 - f(1 - r_1, \ldots, 1 - r_n) \leqslant 1 - f(1 - w_1, \ldots, 1 - w_n) \leqslant \overline{f}(w_1, \ldots, w_n)$, i.e. $\overline{f}$ is monotone.

Finally $\|\overline{f}(r) - \overline{f}(w)\|_\infty = \|(1 - f(1 - r)) - (1 - f(1 - w))\|_\infty \leqslant \|(1 - r) - (1 - w)\|_\infty = \|r - w\|_\infty$, i.e. $\overline{f}$ is nonexpansive. □

Lemma 3.9. *If $f \in M_{n,1}[0,1]$ then, for all $k \in [n]$ and $r = (r_1, \ldots, r_n) \in [0,1]^n$,*

$$\mu_{r_k} x_k . f(r_1, \ldots, r_{k-1}, x_k, r_{k+1}, \ldots, r_n) = \\ 1 - \mu_{1-r_k} x_k . \overline{f}(1 - r_1, \ldots, 1 - r_{k-1}, 1 - x_k, 1 - r_{k+1}, \ldots, 1 - r_n).$$

Proof. Let $\top_f$ and $\bot_f$ be respectively the greatest and the least fixed points of the mapping

$$x_k \mapsto (f r_1, \ldots, r_{k-1}, x_k, r_{k+1}, \ldots, r_n).$$

Similarly let $\top_{\overline{f}}, \bot_{\overline{f}}$ the greatest and the least fixed points of the mapping

$$x_k \mapsto \overline{f}(1 - r_1, \ldots, 1 - r_{k-1}, 1 - x_k, 1 - r_{k+1}, \ldots, 1 - r_n).$$

Since $\overline{f}(1 - r_1, \ldots, 1 - r_{k-1}, x_k, 1 - r_{k+1}, \ldots, r_n) = 1 - f(r_1, \ldots, r_{k-1}, 1 - x_x, r_{k+1}, \ldots, r_n)$ we have $\bot_{\overline{f}} = 1 - \top_f$ and $\top_{\overline{f}} = 1 - \bot_f$.

There are three possibilities concerning the position of r_k relative to $\bot_f$ and $\top_f$.

If $\top_f \leqslant r_k$ then

$$\mu_{r_k}(x_k . f r_1, \ldots, r_{k-1}, x_k, r_{k+1}, \ldots, r_n) = \top_f.$$

However, in this case we have also $1 - r_k \leqslant 1 - \top_f = \bot_{\overline{f}}$ implying that

$$\mu_{1-r_k} x_k . \overline{f}(1 - r_1, \ldots, 1 - r_{k-1}, x_k, 1 - r_{k+1}, \ldots, r_n) = \bot_{\overline{f}}.$$

In a similar way if $r_k \leqslant \bot_f$ then

$$\mu_{r_k} x_k . f(r_1, \ldots, r_{k-1}, x_k, r_{k+1}, \ldots, r_n) = \bot_f$$

and

$$\mu_{1-r_k} x_k . \overline{f}(1 - r_1, \ldots, 1 - r_{k-1}, x_k, 1 - r_{k+1}, \ldots, r_n) = \top_{\overline{f}}.$$

The last case to examine is when $\bot_f \leqslant r_k \leqslant \top_f$. Then

$$\mu_{r_k} x_k . f(r_1, \ldots, r_{k-1}, x_k, r_{k+1}, \ldots, r_n) = r_k$$

and, on the other hand,

$$\bot_{\overline{f}} \leqslant 1 - r_k \leqslant \top_{\overline{f}},$$

implying

$$\mu_{1-r_k} x_k . \overline{f}(1 - r_1, \ldots, 1 - r_{k-1}, x_k, 1 - r_{k+1}, \ldots, r_n) = 1 - r_k.$$

□

Lemma 3.10. *Let $g \in M_{m,k}[0,1]$ and $f \in M_{k,n}[0,1]$. Then $\overline{f \circ g} = \overline{f} \circ \overline{g}$, i.e. the dual of the composition of BMN mappings is equal to the composition of duals.*

Proof. For $r \in [0,1]^n$ we have $\overline{(f \circ g)}(r) = 1 - (f \circ g)(1 - r) = 1 - f(g(1 - r)) = 1 - f(1 - (1 - g(1 - r))) = 1 - f(1 - \overline{g}(r)) = (\overline{f}(\overline{g}(r))$. □

The following lemma examines the duality for the nested nearest fixed points.

Lemma 3.11. *Let $f = (f_1, \ldots, f_n) \in M_{n,n}[0,1]$. Then for all k, $0 \leqslant k \leqslant n$, and $r \in [0,1]^n$*

$$\mathbf{Fix}^k(f)(r) = 1 - \mathbf{Fix}^k(\overline{f})(1 - r). \tag{3.5}$$

Proof. Induction on k.

$r \mapsto \mathbf{Fix}^0(f)(r) = r$ is the identity mapping independently of f. Thus the left-hand side of (3.5) is equal to r and the right-hand side is $1 - (1 - r) = r$ as well.

For each $0 \leqslant k \leqslant n$, let us set

$$\mathbf{Fix}^k(f)(r) = H^k(r) = (H_1^k(r), \ldots, H_n^k(r))$$

and

$$\mathbf{Fix}^k(\overline{f})(r) = \overline{H}^k(r) = (\overline{H}_1^k(r), \ldots, \overline{H}_n^k(r)).$$

Using this notation (3.5) can be written as

$$\overline{H}^k(r) = 1 - H^k(1 - r). \tag{3.6}$$

Our aim is to prove the last equality for k under the assumption that it holds for $k-1$.

By definition

$$\begin{aligned}\overline{H}^k_k(1-r) = \mu_{1-r_k}x_k.\overline{f}_k(&\overline{H}^{k-1}_1(1-r_1,\dots,1-r_{k-1},x_k,1-r_{k+1},\dots,r_n),\\ &\dots,\\ &\overline{H}^{k-1}_{k-1}(1-r_1,\dots,1-r_{k-1},x_k,1-r_{k+1},\dots,r_n),\\ &x_k,1-r_{k+1},\dots,1-r_n).\end{aligned}$$

Let us define a mapping $G^k \in M_{n,n}[0,1]$:

$$G^k := (H^{k-1}_1,\dots H^{k-1}_{k-1},\pi_k,\pi_{k+1},\dots,\pi_n),$$

where $\pi_i(x_1,\dots,x_n) = x_i$, $i = k, k+1,\dots,n$, is the projection on the i-th coordinate. Since $\overline{\pi}_i = \pi_i$, i.e. the dual of the projection is equal the same projection mapping we can see that the dual to G^k is

$$\overline{G}^k = (\overline{H}^{k-1}_1,\dots \overline{H}^{k-1}_{k-1},\pi_k,\pi_{k+1},\dots,\pi_n).$$

Therefore, by Lemmas 3.10 and 3.9,

$$\begin{aligned}\overline{H}^k_k(1-r) &= \mu_{1-r_k}x_k.\overline{f}_k \circ \overline{G}^k(1-r_1,\dots,1-r_{k-1},x_k,1-r_{k+1}\dots,1-r_n)\\ &= \mu_{1-r_k}x_k.\overline{f_k \circ G^k}(1-r_1,\dots,1-r_{k-1},x_k,1-r_{k+1}\dots,1-r_n)\\ &= 1-\mu_{r_k}x_k.f_k \circ G^k(r_1,\dots,r_{k-1},x_k,r_{k+1},\dots,r_n) = 1-H^k_k(r)\end{aligned}$$

For $m \in [k-1]$,

$$\begin{aligned}\overline{H}^k_m(1-r) &= \overline{H}^{k-1}_m(1-r_1,\dots,1-r_{k-1},\overline{H}^k_k(1-r),1-r_{k+1},\dots,1-r_n)\\ &= \overline{H}^{k-1}_m(1-r_1,\dots,1-r_{k-1},1-H^k_k(r),1-r_{k+1},\dots,1-r_n)\\ &= 1-H^{k-1}_m(r_1,\dots,r_{k-1},H^k_k(r),r_{k+1},\dots,r_n)\\ &= 1-H^k_m(r).\end{aligned}$$

Finally, for $m > k$,

$$1-\overline{H}^k_m(1-r) = 1-(1-r_m) = r_m = H^k_m(r).$$

This terminates the proof of (3.6).

□

Chapter 4

Turn-based stochastic priority games

A turn-based stochastic priority game is played by two players on an arena with a finite set of states $\mathbf{S} = [n] = \{1, \dots, n\}$ partitioned into two sets $\mathbf{S}^{\text{Max}}$ and $\mathbf{S}^{\text{Min}}$, where $\mathbf{S}^{\text{Max}}$ and $\mathbf{S}^{\text{Min}}$ are the sets states controlled by player Max and player Min, respectively. For each state $i \in \mathbf{S}$, $\mathbf{A}(i)$ is a finite nonempty set of actions that are available in i. For $i, j \in \mathbf{S}$ and $a \in \mathbf{A}(i)$, $p(j|i, a)$ is the transition probability to move to state j if action a is played at state i.

The players play an infinite game, at each stage the player controlling the current state selects an action to execute and the game moves to a new state according to the transition probability.

The arena is endowed with a reward vector $r = (r_1, \dots, r_n)$, where $r_i \in \mathbb{R}$ is the reward of state i. The priority payoff of an infinite play is defined to be the reward of the maximal (in the usual integer order) state visited infinitely often during the play. The goal of player Max (respectively player Min) is to maximize (respectively minimize) the payoff.

There are two main results in this chapter:

- the value vector of the turn-based stochastic priority game can be obtained as a nested nearest fixed point of a monotone nonexpansive mapping f, where f is the value mapping of the one-step game, and
- both players have pure memoryless optimal strategies.

Note that the last point implies that, since the number of possible pure memoryless strategies is finite, we can find, although in a very inefficient way, optimal strategies for both players through the exhaustive search among all pure memoryless strategies.

The turn-based stochastic priority game with the rewards in the two element set $\{0, 1\}$ is known as the turn-based stochastic *parity* game. These games have been examined in several papers [MM02, CJH04]. In particular Chatterejee, Jurdziński and Henzinger [CJH04] proved that in turn-based stochastic parity games both players have pure memoryless optimal strategies, but their proof is quite different from the one presented in this chapter and relies on the non-trivial general result of Martin [Mar98] concerning the existence of the value for Blackwell games.

In our approach we proceed differently. First of all we show that, without loss of generality, we can limit ourselves to priority games having rewards in the interval $[0, 1]$.

Next for each state i we define a trivial one-step game. The value of the one-step game depends on the reward vector r. Thus the one-step game played at state i gives rise to a mapping f_i that maps the reward vector r to the value $f_i(r)$ of state i in the one-step game. The mappings f_i, called one-step value mappings, can be expressed as either the maximum (for the states controlled by player Max) or the minimum (for the states controlled by the player Min) of a finite number of linear functions.

It is immediate to see that f_i are monotone and nonexpansive.

Let $f = (f_1, \ldots, f_n)$ be the mapping from $[0, 1]^n$ to $[0, 1]^n$ such that, for each m, the coordinate mapping f_m is the value mapping for the one-step game played in m.

Let

$$\mathbf{Fix}^n(f)(r)$$

be the nth nested r-nearest fixed point of f as defined in Chapter 3.

The first main result of this chapter is that, for each $i \in [n]$, the ith coordinate $\mathbf{Fix}_i^n(f)(r)$ of this fixed point is the value of state i in the priority game for the given reward vector r.

The proof has a nice recursive structure. Instead of proving this result in one big step, we prove it by induction on nesting level of the fixed point [1].

In our approach we provide for all $k = 0, 1, \ldots, n$ a game interpretation of the partial fixed point formula

$$\mathbf{Fix}^k(f)(r). \tag{4.1}$$

We prove that (4.1) is equal to the value vector of the priority game with all states greater than k transformed into absorbing states [2].

The chapter is organized as follows. Section 4.1 provides some basic definitions. In Section 4.3 we define the one-step game. This is a very simple one-player game played at each state of the arena. We show, in Section 4.2, that without loss of generality we can limit ourselves to priority games with rewards in the interval $[0, 1]$. In Section 4.4 we give an inductive proof that priority games have optimal pure memoryless strategies and that the value of the priority game can be expressed as a nested fixed point of the value function of the one-step game.

1. This is the main departure from the traditional μ-calculus approach to parity games as for example in [Wal02] and [dAM04], where the proofs were not inductive in spite of the recursive structure of the μ-calculus formula.

2. Recall that a state i is absorbing if for all possible actions the probability to quit i is 0.

The chapter ends with Section 4.5 where we show that the results of Section 4.4 do not carry over to priority games with an infinite number of states or actions.

4.1 Preliminaries

An arena $\mathcal{A}$ of a two-player turn-based stochastic game is composed of the following ingredients:

- a nonempty countable set $\mathbf{S}$ of states partitioned onto the sets $\mathbf{S}^{\mathrm{Max}}$ of states controlled by player Max and the set $\mathbf{S}^{\mathrm{Min}}$ of states controlled by player Min,
- for each state i, a nonempty countable set $\mathbf{A}(i)$ of actions available at i,
- for all $i, j \in \mathbf{S}$ and $a \in \mathbf{A}(i)$, the probability $p(j|i,a)$ to move to state j when action a is executed in state i.

We assume that the sets $\mathbf{A}(i), i \in \mathbf{S}$, are pairwise disjoint.

An infinite game played by players Max and Min starts at some state $s_1 \in \mathbf{S}$. At each stage $t, t = 1, 2, \ldots$, the player controlling the current state s_t chooses an available action $a_t \in \mathbf{A}(s_t)$ and the game moves to a state s_{t+1} with probability $p(s_{t+1}|s_t, a_t)$.

Example 4.1. Figure 4.2 depicts a two-player arena with $\mathbf{S}^{\mathrm{Min}} = \{2, 3\}$, $\mathbf{S}^{\mathrm{Max}} = \{1\}$, action sets $\mathbf{A}(1) = \{a, b, c\}, \mathbf{A}(2) = \{d\}$ and $\mathbf{A}(3) = \{e\}$. The transition probabilities are given by $p(2|1,a) = 0.7, p(3|1,a) = 0.3, p(2|1,c) = p(3|1,b) = p(3|2,d) = p(2|3,e) = 1$. We represent the states controlled by player Max and Min as squares and circles respectively.

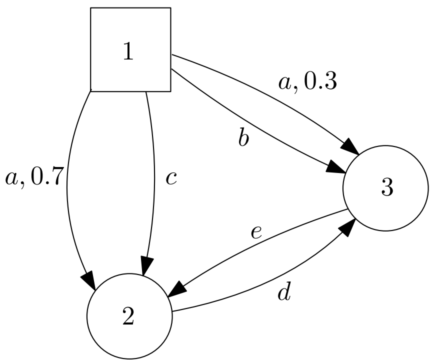

Figure 4.2 – A two-player stochastic arena

A *history* is a finite sequence $h = s_1, a_1, s_2, \ldots, s_{m-1}, a_{m-1}, s_m$, alternating states and actions which starts and ends in a state. The set of all histories is denoted H.

The length of a history h is the number of actions in h. Note that the histories of length 0 are of the form s_1 for $s_1 \in \mathbf{S}$, i.e. they consist of one state and no actions.

Let H_{Max} be the subset of H consisting of histories ending in a state controlled by player Max.

For a finite set A, by $\Delta(A)$ we will denote the set of probability distributions over A. The *support* of $\delta \in \Delta(A)$ is defined as $\text{supp}(\delta) = \{a \in A \mid \delta(a) > 0\}$.

A *strategy* of player Max is a mapping $\sigma : H_{\text{Max}} \to \Delta(\mathbf{A})$, such that $\text{supp}(\sigma(h)) \subseteq \mathbf{A}(s)$, where s is the last state of h.

A *selector* for player Max is a mapping $\sigma : \mathbf{S}^{\text{Max}} \to \mathbf{A}$ such that, for each $s \in \mathbf{S}^{\text{Max}}$, $\sigma(s) \in \mathbf{A}(s)$.

A strategy σ' of player Max is said to be *pure memoryless* if there exists a selector σ such that $\sigma'(h) = \sigma(s)$ for each history h ending in a state s controlled by Max. In the sequel we identify pure memoryless strategies with corresponding selectors.

The definitions of strategies, selectors and pure memoryless strategies carry over to player Min in the obvious way.

We write Σ and $\mathcal{T}$ to denote the sets of all strategies for player Max and Min respectively.

In the sequel σ, eventually with subscripts or superscripts, is used to denote strategies of player Max. Similarly, τ, with or without subscripts and superscripts is used to denote strategies of player Min.

An *infinite history* or a *play* is an infinite sequence $h = s_1, a_1, s_2, a_2, \ldots$ alternating states and actions. The set of plays is denoted H^∞.

Assuming that the sets $\mathbf{S}$ and $\mathbf{A}$ are equipped with the discrete topology we endow the set of plays H^∞ with the product topology. By $\mathcal{B}(H^\infty)$ we denote the σ-algebra of Borel subsets of $\mathbf{S}^\infty$.

Let $h = s_1, a_1, \ldots, a_{m-1}, s_m$ be a history. By h^+ we denote the cylinder generated by h, i.e. the set of plays (infinite histories) having prefix h.

Cylinders form the basis of the product topology on H^∞, and $\mathcal{B}(H^\infty)$ is the smallest σ-algebra generated by cylinders.

A strategy σ of player Max, a strategy τ of player Min and an initial state i determine a probability measure $\mathbf{P}_i^{\sigma,\tau}$ on $(H^\infty, \mathcal{B}(H^\infty))$.

We define inductively $\mathbf{P}_i^{\sigma,\tau}$ for cylinders in the following way. Let $\sigma \cup \tau$ be the mapping from H to $\Delta(\mathbf{A})$ defined in the following way, for $h \in H$,

$$(\sigma \cup \tau)(h) = \begin{cases} \sigma(h) & \text{if the last state of } h \text{ is controlled by Max,} \\ \tau(h) & \text{if the last state of } h \text{ is controlled by Min.} \end{cases}$$

If $h_0 = s_1$ is a finite history of length 0 then

$$\mathbf{P}_i^{\sigma,\tau}(h_0^+) = \begin{cases} 0 & \text{if } i \neq s_1, \\ 1 & \text{if } i = s_1. \end{cases}$$

Let $h_{t-1} = s_1, a_1, \ldots, s_{t-1}, a_{t-1} s_t$ and $h_t = h_{t-1}, a_t, s_{t+1}$. Then

$$\mathbf{P}_i^{\sigma,\tau}(h_t^+) = \mathbf{P}_i^{\sigma,\tau}(h_{t-1}^+) \cdot (\sigma \cup \tau)(h_{t-1})(a_t) \cdot p(s_{t+1}|s_t, a_t).$$

Note that the family of cylinders is closed under intersection, this family is a π-system of sets, which implies that a probability defined on cylinders extends in a unique way to all sets of $\mathcal{B}(H^\infty)$.

A *payoff mapping* is any bounded Borel measurable mapping

$$\varphi : H^\infty \to \mathbb{R}.$$

For each play $h \in H^\infty$, $\varphi(h)$ is the payoff that player Min pays to player Max if h is the play obtained during the game.

For each initial state i, the aim of the player Max (player Min) is to maximize (respectively minimize) the *expected payoff*:

$$\mathbf{E}_i^{\sigma,\tau}[\varphi] = \int_{H^\infty} \varphi(h) \mathbf{P}_i^{\sigma,\tau}(dh).$$

The game with payoff φ has value if, for each state i, there exist $v_i \in \mathbb{R}$, the value of state i, such that

$$\inf_{\tau \in \mathcal{T}} \sup_{\sigma \in \Sigma} \mathbf{E}_i^{\sigma,\tau}[\varphi] = v_i = \sup_{\sigma \in \Sigma} \inf_{\tau \in \mathcal{T}} \mathbf{E}_i^{\sigma,\tau}[\varphi].$$

Strategies $\sigma^\star$ and $\tau^\star$ are optimal for players Max and Min respectively if, for each state i,

$$\sup_{\sigma \in \Sigma} \mathbf{E}_i^{\sigma,\tau^\star}[\varphi] \leqslant v_i \leqslant \inf_{\tau \in \mathcal{T}} \mathbf{E}_i^{\sigma^\star,\tau}[\varphi],$$

for all strategies σ and τ of Max and Min.

In other words, given an initial state i, player Max using his optimal strategy can secure the expected payoff of at least v_i, while player Min using his optimal strategy ensures that he will pay no more than v_i.

Clearly if $\sigma^\star$ and $\tau^\star$ are optimal then $v_i = \mathbf{E}_i^{\sigma^\star,\tau^\star}[\varphi]$.

An arena is *finite* if the set of states $\mathbf{S}$ and all sets of actions $\mathbf{A}(s), s \in \mathbf{S}$, are finite.

Except in Section 4.5, all games considered in this chapter are played on finite arenas.

Thus, except in the last section, we will assume that the set of states is a finite initial segment of integers, i.e.

$$\mathbf{S} = [n] := \{1, \ldots, n\}.$$

To define the turn-based stochastic priority games we assume that $\mathbf{S} = [n]$ is endowed with the usual order relation $\leqslant$ over integers.

For two states $i, j \in [n]$ we shall say that j has a priority greater than i if $i < j$, in other words the natural order over integers will serve as a priority order over states.

A *reward mapping* is any mapping

$$r : \mathbf{S} \to \mathbb{R},$$

where, for $i \in \mathbf{S}$, the real number $r(i)$ is called the *reward* of i. Since $\mathbf{S} = [n]$ we will identify the reward mappings with the elements of the Cartesian product $\mathbb{R}^n$ and for $r \in \mathbb{R}^n$, we write $r = (r_1, \dots, r_n)$, where r_i is the reward of state i. In particular, we will often call r the reward vector rather than the reward mapping and r_i and $r(i)$ will be used interchangeably.

The stochastic priority game is the game played on arena $\mathcal{A}$ with the payoff mapping φ_r defined in the following way, for each play $h = s_1, a_1, s_2, \dots$,

$$\varphi_r(h) = r(\limsup_t s_t).$$

Note that since we assumed that the set of states is $\{1, \dots, n\}$, the sequence $s_1, s_2, s_3, \dots$ of visited states is a sequence of integers and lim sup is taken w.r.t. the natural order relation over integers. Thus $\limsup_t s_t$ is simply the maximal state appearing infinitely often in h and the payoff of the turn-based stochastic priority game is equal to the reward of the maximal state visited infinitely often.

Example 4.3. Let us take the arena $\mathcal{A}$ defined in Example 4.1. Let σ and τ be pure memoryless strategies for player Max and Min respectively such that $\sigma(1)(b) = 1/3$, $\sigma(1)(c) = 2/3$, $\tau(2)(d) = 1$ and $\tau(3)(e) = 1$. Once the memoryless strategies are fixed, we get a Markov chain, depicted in Figure 4.4. Let $r = (0, 1, 1/5)$ be the reward mapping.

Then if the initial state is 1, the game moves to state 2 with probability 2/3. In other words $\mathbf{P}_1^{\sigma,\tau}(1, b, 2) = 2/3$. Moreover, once the game is in state 2, it alternates between state 2 and 3, i.e., let $h_2 = 1, b, 2, d, 3, e, 2, d, 3, e, 2, \dots$ and $h_3 = 1, b, 3, e, 2, d, 3, e, 2, \dots$, hence we get $\mathbf{P}_1^{\sigma,\tau}(h_2) = 2/3$, $\mathbf{P}_1^{\sigma,\tau}(h_3) = 1/3$ and $\varphi_r(h_2) = \varphi_r(h_3) = 1/5$. The last equality is because im both histories the bigger state infinitely often visited is state 3 which has a reward $r_3 = 1/5$. Finally,

$$\mathbf{E}_1^{\sigma,\tau}[\varphi_r] = 1/5.$$

The aim of the rest of this chapter is to show that finite state turn-based stochastic priority games have value that can be expressed as a nested nearest fixed point of piecewise linear mappings (the value mappings of the one-day games) and that both players have optimal pure memoryless strategies.

The proof will carried out by induction on the number of absorbing states.

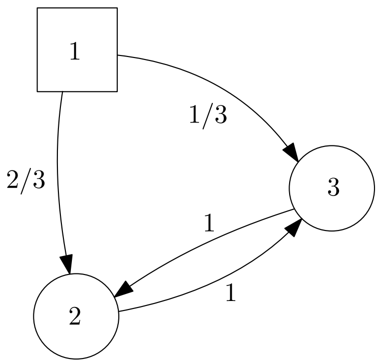

Figure 4.4 – Transition probabilities in $\mathcal{A}$ with strategies σ and τ.

Definition 4.5. *A state* $i \in \mathbf{S}$ *is called* absorbing *if, for each action* $a \in \mathbf{A}(i)$, $p(i|i,a) = 1$.

If the game enters an absorbing state i (in particular if it starts in an absorbing state i) then the game remains in i forever and the payoff is equal to the reward r_i. In particular, if all states are absorbing then the priority game is trivial, the value of each state i is equal to the reward r_i and all strategies are optimal.

In general, intuitively, a game with many absorbing states is simpler than a game with a few absorbing states. This observation leads to the inductive proof presented in this chapter. We start with the trivial priority game where all states are absorbing and next we transform the states, one by one, starting with state 1, next state 2 and so on, from absorbing to nonabsorbing.

4.2 Bounding the rewards

In the sequel it will be convenient to assume that all rewards belong to the interval $[0,1]$ rather than to $\mathbb{R}$. This can be achieved for each game without loss of generality by a simple linear transformation. Let $a = \min_{i \in \mathbf{S}} r_i$, $b = \max_{i \in \mathbf{S}} r_i$ and $g(x) = \frac{1}{b-a}x - \frac{a}{b-a}$. Then $0 = g(a) \leqslant g(x) \leqslant g(b) = 1$ for $x \in \{r_1, \ldots, r_n\}$. Changing the reward vector from $r = (r_1, \ldots, r_n)$ to $g(r) = (g(r_1), \ldots, g(r_n))$ transforms linearly the priority payoffs of all plays h since $\varphi_{g(r)}(h) = g(\varphi_r(h))$.

By the linearity of expectation, this implies that for all starting states i and all strategies σ and τ we have $g(\mathbf{E}_i^{\sigma,\tau}(\varphi_r)) = \mathbf{E}_i^{\sigma,\tau}(g(\varphi_r))$, in particular the priority games with the reward vectors r and $g(r)$ have the same optimal strategies.

4.3 The one-step game

For turn-based stochastic games the auxiliary one-step game is a simple one-player game played in each state. The one step games are an essential ingredient of our solution to the turn-based stochastic priority games.

Recall that we assume that the set of states is $\mathbf{S} = [n] = \{1, \ldots, n\}$.

Let $x \in \mathbb{R}^n$ be a reward vector. For each state k, we consider the following one-step game played:

- the player controlling k plays an action $a \in \mathbf{A}(k)$ and the game moves to state j with probability $p(j|k, a)$,
- this single move ends the one-step game and player Max obtains from player Min the payoff x_j.

If the player controlling k plays action $a \in \mathbf{A}(k)$ then the expected payoff obtained by player Max in the one-step game is equal to $\sum_i p(i|k, a) \cdot x_i$. As always, the aim of player Max (Min) is to maximize (minimize) this expected payoff.

As the game is finite, it is clear that the player controlling k has an optimal pure strategy in the one-step game, this strategy consists in playing an action a that either maximizes (if k is controlled by Max) or minimizes (if k is controlled by Min) the sum $\sum_i p(i|k, a) \cdot x_i$. Therefore, we can see that the value of the one-step game played at state $k \in [n]$ is equal to

$$f_k(x) := \begin{cases} \max_{a \in \mathbf{A}(k)} \sum_i p(i|k, a) \cdot x_i & \text{if } k \in \mathbf{S}^{\mathrm{Max}}, \\ \min_{a \in \mathbf{A}(k)} \sum_i p(i|k, a) \cdot x_i & \text{if } k \in \mathbf{S}^{\mathrm{Min}}. \end{cases} \tag{4.2}$$

In the sequel we consider the value of the one-step game as a function of the reward vector $x = (x_1, \ldots, x_n)$, i.e. f_k is considered as a function

$$f_k : [0, 1]^n \to \mathbb{R}$$

defined by (4.2).

We set

$$f = (f_1, \ldots, f_n),$$

i.e. $f : [0, 1]^n \to \mathbb{R}^n$ maps reward vectors $x \in [0, 1]^n$ to the vector of values of one-step games played in the states of $\mathbf{S}$.

Lemma 4.6. *The value mapping f of the one-step game is bounded monotone and nonexpansive.*

Proof. That f is monotone is obvious. It is bounded since the convex combination of elements belonging to $[0, 1]$ belongs to $[0, 1]$ as well. It is also evident that f is additively homogeneous, i.e. for each $x \in \mathbb{R}^n$ and each $\lambda \in \mathbb{R}$,

$$f(x + \lambda \cdot e_n) = f(x) + \lambda \cdot e_n,$$

where $e_n = (1, \ldots, 1) \in \mathbb{R}^n$ is the vector with 1 on all components. By Lemma 3.2 this implies that f is nonexpansive. □

4.4 Nested nearest fixed point solution to priority games

The priority game having all states absorbing is trivial, the value of state i, $i \in [n]$, is r_i, where $r \in [0,1]^n$ is the reward vector. Moreover, all strategies are optimal, in particular each pure memoryless strategy is optimal.

In this section we provide an inductive proof that all priority games have optimal pure memoryless strategies.

Moreover, we show that the value vector for the priority game with reward r is equal to $\mathbf{Fix}^n(f)(r)$ — the nested fixed point of the value mapping f of the one-step game defined Section 3.2.

The induction will be carried out on the number of nonabsorbing states. We show that if we can solve the priority game with states $k, k+1, \ldots, n$ absorbing then we can use this solution to solve the priority game with states $k+1, \ldots, n$ absorbing, i.e. we can decrease the number of absorbing states. Note that the order in which we transform the states from absorbing to nonabsorbing is essential, at each inductive step we transform the smallest absorbing state to a nonabsorbing one.

Although the idea of making some states absorbing in order to simplify the game is the one that is behind the proof, the direct application of this idea would lead to a cumbersome notation. For this reason we shall adopt another, equivalent, approach, where instead of modifying the transition probabilities of the arena we rather modify the payoff mapping.

By S_t and A_t, $t = 1, 2, \ldots$, we will denote two stochastic processes such that S_t is the state visited at time t and A_t is the action executed at stage t, i.e. for a play $h = s_1, a_1, s_2, a_2, s_3, \ldots$, $S_t(h) = s_t$ and $A_t(h) = a_t$.

For each state $k \in [n]$ we define the random variable

$$T_{>k} : H^\infty \to \mathbb{N} \cup \{\infty\}$$

such that

$$T_{>k} = \min\{t \mid S_t > k\}.$$

Thus $T_{>k}$ is the time of the first visit to a state greater than k. Since the minimum of the empty set is $+\infty$ we have $T_{>k} = \infty$ for the plays belonging to the event $\{\forall t, S_t \in [k]\}$, i.e. $T_{>k} = \infty$ if all visited states are in $[k]$.

Note that $T_{>k}$ is a stopping time with respect to $\{S_i\}_{i \geqslant 1}$. Indeed, for each time $t \in \mathbb{N}$,

$$\{T_{>k} = t\} = \{S_1 \leqslant k, \ldots, S_{t-1} \leqslant k, S_t > k\},$$

i.e. the event $\{T_{>k} = t\}$ belongs to the sigma algebra $\sigma(S_1, \dots, S_t)$ generated by $S_1, \dots, S_t$.

For each $k \in \{0\} \cup [n]$ we define the *stopped state process* $S_t^{[k]}, t \in \mathbb{N}$,

$$S_t^{[k]} = S_{t \wedge T_{>k}} = \begin{cases} S_t & \text{if } T_{>k} > t, \\ S_{T_{>k}} & \text{if } T_{>k} \leqslant t, \end{cases}$$

where $a \wedge b$ denotes the minimum of a and b.

Thus if all states visited up to the moment t belong to $\{1, \dots, k\}$ then $S_t^{[k]}$ is equal to the state S_t visited at the current epoch t. However, if at some previous epoch a state $> k$ was visited then $S_t^{[k]}$ is the first such state. In other words the process $S_t^{[k]}$ behaves as if the states $> k$ were absorbing.

For a given reward vector r, we define a new payoff mapping $\varphi_r^{[k]}$:

$$\varphi_r^{[k]} = r(\limsup_t S_t^{[k]}).$$

The game with payoff $\varphi_r^{[k]}$ will be called *stopped priority game* or simply $\varphi_r^{[k]}$-game.

Note that once a state m greater than k is visited, the game with payoff $\varphi_r^{[k]}$ is for all practical reasons over, independently of what can happen in the future the payoff is equal to the reward r_m of this state and the states visited after the moment $T_{>k}$ have no bearing on the payoff.

In the stopped priority $\varphi_r^{[k]}$-game the states $> k$ will be called *stopping states* while the states $\leqslant k$ will be called *non-stopping*.

Note that since we have assumed that $\mathbf{S} = [n]$, i.e. n is the greatest state, we have $\varphi_r^{[n]} = \varphi_r$.

Note also that solving games starting in stopping states is trivial. If $i > k$ then for all plays h starting in i, $\varphi_r^{[k]}(h) = r_i$, thus $\mathbf{E}_i^{\sigma,\tau}(\varphi_r^{[k]}) = r_i$ for all strategies σ, τ the value of a stopping state i, $i > k$, is r_i. In particular, the game with payoff $\varphi_r^{[0]}$ is trivial since all states of this game are stopping. Moreover, for the $\varphi_r^{[0]}$-game all strategies are optimal since the payoff does not depend on the strategy.

The main result of this chapter is

Theorem 4.7. *Let $f : [0,1]^n \to [0,1]^n$ be the value mapping of the one-step game defined in (4.2).*

Then, for each $r \in [0,1]^n$, the $\varphi_r^{[k]}$-game satisfies the following properties:

- *for each state $i \in [n]$, the value of i is equal to $\mathbf{Fix}_i^k(f)(r)$, where $\mathbf{Fix}_i^k(f)$ is the ith coordinate of the kth r-nearest fixed point $\mathbf{Fix}^k(f)$ of f,*
- *both players have optimal pure memoryless strategies.*

Theorem 4.7 holds trivially for i such that $i > k$. Indeed, in the $\varphi_r^{[k]}$-game all states $i > k$ are stopping thus $\varphi_r^{[k]}(h) = r_i$ for all plays h starting in a state $i > k$. On the other hand, we have also $\mathbf{Fix}_i^k(f)(r) = r_i$.

The recursive formula of the nested fixed points that, according to Theorem 4.7, represents the value of the stopping game has a natural game theoretic interpretation.

Let us consider the $\varphi^{k-1}_{(r_1,\ldots,r_{k-1},\zeta,r_{k+1},\ldots,r_n)}$-game. This is the priority game where the states $i \neq k$ have rewards r_i while the state k, the smallest stopping state, has reward ζ.

Suppose that Theorem 4.7 holds for $k-1$. Thus the value of state $i \in [k-1]$ in the $\varphi^{k-1}_{(r_1,\ldots,r_{k-1},\zeta,r_{k+1},\ldots,r_n)}$-game, seen as the function of the reward ζ of the state k, is

$$F_i^{k-1}(\zeta; r) = \mathbf{Fix}_i^{k-1}(f)(r_1, \ldots, r_{k-1}, \zeta, r_{k+1}, \ldots, r_n) \tag{4.3}$$

in the notation of (3.3).

Now let us consider the $\varphi_r^{[k]}$-game where the state k becomes the greatest non-stopping state. Let us note

$$\mathrm{val}_i(\varphi_r^{[k]})$$

the value of state i in the $\varphi_r^{[k]}$-game. Clearly for the stopping states we have

$$\mathrm{val}_i(\varphi_r^{[k]}) = r_i, \quad \text{for } i > k. \tag{4.4}$$

Suppose that

$$\mathrm{val}_k(\varphi_r^{[k]}) = \zeta, \tag{4.5}$$

i.e. the value of the state k in the $\varphi_r^{[k]}$-game is some unknown $\zeta \in [0, 1]$.

What are the values of the states $i < k$ in the $\varphi_r^{[k]}$-game? Let us start to play the $\varphi_r^{[k]}$-game starting at state $i < k$ and suppose that both players play optimally. When such a game hits the state k then in the auxiliary game starting at k, the payoff obtained will be equal to the value ζ of k. Thus it seems plausible that the value of state $i < k$ in the $\varphi_r^{[k]}$-game is equal to the value of this state in the $\varphi^{[k-1]}_{(r_1,\ldots,r_{k-1},\zeta,r_{k+1},\ldots,r_n)}$-game which stops at k with the payoff ζ, i.e.

$$\mathrm{val}_i(\varphi_r^{[k]}) = F_i^{k-1}(\zeta; r), \quad \text{for } i < k. \tag{4.6}$$

But what is the value of the state k in the $\varphi_r^{[k]}$-game? Suppose for example that k is controlled by player Max. When player Max executes action a at k, the game moves to state i with probability $p(i|k,a)$ and starting from i player Max can win at least the value $\mathrm{val}_i(\varphi_r^{[k]})$. Thus in the $\varphi_r^{[k]}$-game starting at k player Max can win

$$\max_{a\in\mathbf{A}(k)} \sum_i \mathrm{val}_i(\varphi_r^{[k]}) \cdot p(i|k,a).$$

We obtain a similar expression when k is controlled by Min with $\min_{a\in\mathbf{A}(k)}$ replacing $\max_{a\in\mathbf{A}(k)}$. Using the definition of the value function of the one-step game played at k, see (4.2), and (4.4),(4.5),(4.6), we obtain

$$\begin{aligned}\zeta = \mathrm{val}_k(\varphi_r^{[k]}) &= f_k(\mathrm{val}_1(\varphi_r^{[k]}),\dots,\mathrm{val}_n(\varphi_r^{[k]}))\\ &= f_k(F_1^{k-1}(\zeta;r),\dots,F_{k-1}^{k-1}(\zeta;r),\zeta,r_{k+1},\dots,r_n).\end{aligned}$$

Thus we can see that a natural candidate for the value of the state k in the $\varphi_r^{[k]}$-game is a fixed point of the mapping

$$\zeta \mapsto f_k(F_1^{k-1}(\zeta;r),\dots,F_{k-1}^{k-1}(\zeta;r),\zeta,r_{k+1},\dots,r_n).$$

This mapping can have many fixed points, however one of them seems more plausible than the others, this is the fixed point which is the nearest to the reward r_k of k, i.e. the natural conjecture is that

$$\mathrm{val}_k(\varphi_r^{[k]}) = \mu_{r_k}\zeta.f_k(F_1^{k-1}(\zeta;r),\dots,F_{k-1}^{k-1}(\zeta;r),\zeta,r_{k+1},\dots,r_n). \tag{4.7}$$

But (4.4),(4.5) and (4.7) and the definition of the nested nearest fixed point coincides with the inductive definition of the kth nested r-nearest fixed point,

$$\mathrm{val}(\varphi_r^{[k]}) = \mathbf{Fix}^k(f)(r)$$

i.e. the kth nested r-nearest fixed point of the value mapping of the one-step game is the natural candidate for the value of the $\varphi_r^{[k]}$-game.

Theorem 4.7 confirms these intuitions and the proof formalizes the reasoning given above.

Example 4.8. Let $\mathcal{A}$ be the arena defined as follows: let $\mathbf{S} = \{1,2,3\}$, $\mathbf{A}$ such that $\mathbf{A}(1) = \{a,b\}, \mathbf{A}(2) = \{c,d\}$ and $\mathbf{A}(3) = \{e\}$, such that $p(2|1,a) = p(3|1,b) = p(3,3,d) = 1$, $p(1|2,c) = 0.8$ and $p(3|2,c) = 0.2$ as shows Figure 4.9 and let $r = (0,1,1/2)$.

The stochastic priority game is the game played ion arena $\mathcal{A}$ with the priority payoff mapping φ_r defined above. We want to calculate the value of the φ_r-game, notice that, as state 3 is absorbing, $\varphi_r^{[2]}$-game and φ_r-game are equal. We start by calculating the value of state 2 in $\varphi_r^{[2]}$-game.

Recall the definitions in Section 3.2 and as state 1 is controlled by player Max and state 2 by player Min we have $f = (f_1,f_2,f_3) : [0,1]^3 \to [0,1]$ the value

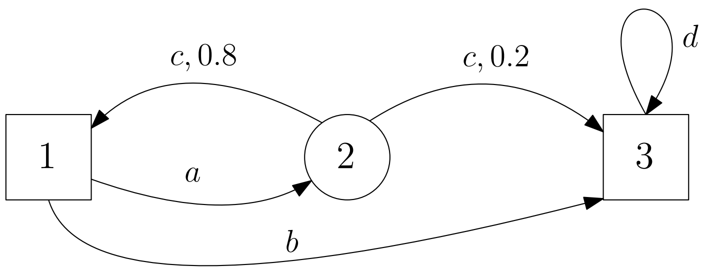

Figure 4.9 – Game with states $\mathbf{S} = \{1, 2, 3\}$ and actions $\mathbf{A}$as defined above.

mapping of the one-step game as defined in (4.2), i.e., $f_1 : (x_1, x_2, x_3) \mapsto \max(x_2, x_3)$, $f_2 : (x_1, x_2, x_3) \mapsto 0.8x_1 + 0.2x_3$ and $f_3 : (x_1, x_2, x_3) \mapsto x_3$.

Hence,

$$\mathbf{Fix}_2^2(f)(r) = \mu_1\zeta.f_2(F_1^1(\zeta; r), \zeta, 1/2) \tag{4.8}$$

and $F_1^1(\zeta; r) = \mathbf{Fix}_1^1(f)(0, \zeta, 1/2)$ that, by induction, it should be the value of state 1 in the $\varphi^{[1]}_{(0,\zeta,1/2)}$-game that is the max between ζ and $1/2$. In fact,

$$\begin{aligned} \mathbf{Fix}_1^1(f)(0, \zeta, 1/2) &= \mu_0\xi.f_1(\xi, \zeta, 1/2) \\ &= \mu_0\xi.\max(\zeta, 1/2) \\ &= \max(\zeta, 1/2). \end{aligned} \tag{4.9}$$

Then, retaking (4.8),

$$\begin{aligned} \mathbf{Fix}_2^2(f)(r) &= \mu_1\zeta.f_2(\max(\zeta, 1/2), \zeta, 1/2) \\ &= \mu_1\zeta.(0.8 \times \max(\zeta, 1/2) + 0.2 \times 1/2) \\ &= 1/2. \end{aligned}$$

And for state 1,

$$\begin{aligned} \mathbf{Fix}_1^2(f)(r) = \mathbf{Fix}_1^1(f)(0, \mathbf{Fix}_2^2(f)(r), 1/2) &= F_1^1(0, 1/2, 1/2) \\ &= \mathbf{Fix}_1^1(f)(0, 1/2, 1/2) \\ &= \max(1/2, 1/2) = 1/2. \end{aligned}$$

Last equality is due to (4.9). Finally, $\mathbf{Fix}_3^2(f)(r) = 1/2$ and hence the values of the game according to Theorem 4.7 are given by $(1/2, 1/2, 0)$ that match with the values of the game as the reader can easily verify.

4.4.1 Optimal strategy for player Max

The aim of this section is to construct an optimal pure memoryless strategy for Max in the $\varphi_r^{[k]}$-game.

Through the section we assume that Theorem 4.7 holds for $k-1$, i.e. for each reward vector $r \in [0,1]^n$, the $\varphi_r^{[k-1]}$-game satisfies the following properties:

(H.1) for each $i \in [n]$, the value of state i is equal to $\mathbf{Fix}_i^{k-1}(f)(r)$ and

(H.2) both players have optimal pure memoryless strategies.

We assume that

$$F_i^{k-1}(\zeta; r)$$

is defined as in (4.3) and we define

$$F_k^\sharp(\zeta; r) := f_k(F_1^{k-1}(\zeta; r), \dots, F_{k-1}^{k-1}(\zeta; r), \zeta, r_{k+1}, \dots, r_n).$$

Using this notation we have

$$\mathbf{Fix}_k^k(f)(r) = \mu_{r_k}\zeta . F_k^\sharp(\zeta; r).$$

Notation:

For a set of plays $C \subset H^\infty$, we will write $\mathbb{1}_C$ to denote the *indicator mapping* of the set C,

$$\mathbb{1}_C(h) = \begin{cases} 1 & \text{if } h \in C, \\ 0 & \text{otherwise.} \end{cases}$$

Thus for a mapping φ, $\mathbf{E}(\varphi \mathbb{1}_C) = \int_C \varphi(h) dh$.

Definition 4.10. *By $T_m^{[k]}$ we will denote the time of the mth visit to k of the stopped state process $S_i^{[k]}$, i.e.*

$$T_1^{[k]} = \min\{t \mid S_t^{[k]} = k\}$$

and

$$T_m^{[k]} = \min\{t \mid t > T_{m-1}^{[k]} \text{ and } S_t^{[k]} = k\}.$$

Note that since the minimum of the empty set is $+\infty$ we have $T_m^{[k]} = \infty$ if and only if the stopped state process $S_t^{[k]}$ visits state k less than m times.

Note also that if $T_m^{[k]} < \infty$ then the following conditions are satisfied:

— $S_{T_m^{[k]}} = k$ (the state k is visited at the time $T_m^{[k]}$),

— for all $1 \leqslant t \leqslant T_m^{[k]}$, $S_t \leqslant k$ (all states visited up to the time $T_m^{[k]}$ are non-stopping),

— $\sharp\{t \leqslant T_m^{[k]} \mid S_t = k\} = m$ (the number of visits of the state process S_t to k up to the moment $T_m^{[k]}$ included is m).

Lemma 4.11. *Suppose that for each reward vector $r \in [0,1]^n$, the $\varphi_r^{[k-1]}$-game satisfies **(H.1)** and **(H.2)**.*

Then for each $\zeta \in [0,1]$ such that

$$\zeta \leqslant F_k^\sharp(\zeta; r) \tag{4.10}$$

there exists a pure memoryless strategy σ_ζ^k for player Max *such that σ_ζ^k is optimal for* Max *in the $\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},\zeta,r_{k+1},\dots,r_n)}$-game and for each strategy τ of* Min *we have*

(C1) For all m,

$$F_k^\sharp(\zeta; r) \leqslant \zeta \cdot \mathbf{P}_k^{\sigma_\zeta^k,\tau}(T_{m+1}^{[k]} < \infty \mid T_m^{[k]} < \infty) + \mathbf{E}_k^{\sigma_\zeta^k,\tau}(\varphi^{[k]}_{(r_1,\dots,r_{k-1},\zeta,r_{k+1},\dots,r_n)} \mathbb{1}_{\{T_{m+1}^{[k]}=\infty\}} \mid T_m^{[k]} < \infty),$$

(C2)

$$F_k^\sharp(\zeta; r) \leqslant \mathbf{E}_k^{\sigma_\zeta^k,\tau}(\varphi^{[k]}_{(r_1,\dots,r_{k-1},\zeta,r_{k+1},\dots,r_n)}),$$

(C3) if the inequality (4.10) is strict then

$$\mathbf{P}_k^{\sigma_\zeta^k,\tau}(S_i^{[k]} = k \text{ for infinitely many } i) = 0.$$

Proof. We begin with the definition of the strategy σ_ζ^k. To simplify notation, we write

$$r_\zeta^{-k} := (r_1, \dots, r_{k-1}, \zeta, r_{k+1}, \dots, r_n).$$

By **(H.1)** and **(H.2)**, player Max has an optimal pure memoryless strategy σ_ζ^{k-1} in the $\varphi^{[k-1]}_{r_\zeta^{-k}}$-game such that for each strategy τ of player Min and each starting state $i < k$,

$$F_i^{k-1}(\zeta; r) \leqslant \mathbf{E}_i^{\sigma_\zeta^{k-1},\tau}(\varphi^{[k-1]}_{r_\zeta^{-k}}).$$

To define the strategy σ_ζ^k we should examine two cases.

Case 1: $k \in \mathbf{S}^{\text{Max}}$.

Then

$$F_k^\sharp(\zeta; r) = f_k(F_1^{k-1}(\zeta; r), \dots, F_{k-1}^{k-1}(\zeta; r), \zeta, r_{k-1}, \dots, r_n) = \max_{a \in \mathbf{A}(k)} \sum_{i<k} F_i^{k-1}(\zeta; r) \cdot p(i|k,a) + \zeta \cdot p(k|k,a) + \sum_{i>k} r_i \cdot p(i|k,a)$$

and selecting the action $a_\zeta \in \mathbf{A}(k)$ such that

$$a_\zeta := \arg\max_{a \in \mathbf{A}(k)} \sum_{i<k} F_i^{k-1}(\zeta; r) \cdot p(i|k,a) + \zeta \cdot p(k|k,a) + \sum_{i>k} r_i \cdot p(i|k,a)$$

we obtain

$$F_k^\sharp(\zeta; r) = \sum_{i<k} F_i^{k-1}(\zeta; r) \cdot p(i|k, a_\zeta) + \zeta \cdot p(k|k, a_\zeta) + \sum_{i>k} r_i \cdot p(i|k, a_\zeta). \tag{4.11}$$

We define the strategy σ_ζ^k in the following way, for each state $i \in [k] \cap \mathbf{S}^{\mathrm{Max}}$,

$$\sigma_\zeta^k(i) := \begin{cases} \sigma_\zeta^{k-1}(i) & \text{if } i < k, \\ a_\zeta & \text{if } i = k. \end{cases} \tag{4.12}$$

Case 2: $k \in \mathbf{S}^{\mathrm{Min}}$.
Then

$$F_k^\sharp(\zeta; r) = f_k(F_1^{k-1}(\zeta), \ldots, F_{k-1}^{k-1}(\zeta; r), \zeta, r_{k-1}, \ldots, r_n) = \min_{a \in \mathbf{A}(k)} \sum_{i<k} F_i^{k-1}(\zeta; r) \cdot p(i|k, a) + \eta \cdot p(k|k, a) + \sum_{i>k} r_i \cdot p(i|k, a_\zeta),$$

which implies that for each action $a \in \mathbf{A}(k)$ we have

$$F_k^\sharp(\zeta; r) \leqslant \sum_{i<k} F_i^{k-1}(\zeta; r) \cdot p(i|k, a) + \zeta \cdot p(k|k, a) + \sum_{i>k} r_i \cdot p(i|k, a) \tag{4.13}$$

and we define σ_ζ^k in the following way

$$\sigma_\zeta^k := \sigma_\zeta^{k-1}.$$

We will examine what happens in the $\varphi_{r_\zeta^{-k}}^{[k]}$-game starting in the state k when player Max plays using σ_ζ^k against any strategy τ of player Min.

Proof of (C1):

Before we start the proof of (C1) it is worthwhile to examine the intuitive meaning of this inequality. Suppose that $T_m^{[k]} < \infty$ and consider the moment $T_m^{[k]}$ when k is visited for the mth time. Let $(r_1, \ldots, r_{k-1}, \zeta, r_{k+1}, \ldots, r_n)$ be the reward vector. Consider the auxiliary game starting at time $T_m^{[k]}$ in k with the payoff defined in the following way:

- player Max receives from player Min the payoff ζ for the plays that return to k, i.e. for $h \in H^\infty$ such that $T_{m+1}^{[k]} < \infty$,
- for plays h that do not return to k, i.e. for plays h such that $T_{m+1}^{[k]}(h) = \infty$, the payoff is equal to $\varphi_{r_\zeta^{-k}}^{[k]}(h)$.

Then the right-hand side of (C1) is the expected payoff of such auxiliary game when player Max plays according to σ^k_ζ and inequality (C1) provides a lower bound for the payoff obtained in the auxiliary game.

To prove (C1), suppose that at the moment $T^{[k]}_m < \infty$, when the stopped state process $S^{[k]}_i$ visits the state k for the mth time, an action is played and this action is either the action $\sigma^k_\zeta(k)$ if k is controlled by Max or any action from $\mathbf{A}(k)$ if k is controlled by Min. From (4.11) and (4.13) it follows that

$$\begin{aligned} F^\sharp_k(\zeta; r) \leqslant \sum_{i<k} F^{k-1}_i(\zeta; r)\, \mathbf{P}^{\sigma^k_\zeta,\tau}_k(S_{T^{[k]}_m+1} = i \mid T^{[k]}_m < \infty) \\ + \zeta\, \mathbf{P}^{\sigma^k_\zeta,\tau}_k(S_{T^{[k]}_m+1} = k \mid T^{[k]}_m < \infty) \\ + \sum_{i>k} r_i\, \mathbf{P}^{\sigma^k_\zeta,\tau}_k(S_{T^{[k]}_m+1} = i \mid T^{[k]}_m < \infty). \end{aligned} \tag{4.14}$$

For a play h, $h \in \{S_{T^{[k]}_m+1} = i\}$ for $i > k$, if and only if

— $T^{[k]}_m(h) < \infty$, i.e. h visits k at least m times,

— all states visited prior to $T^{[k]}_m$ are $\leqslant k$,

— $T^{[k]}_m + 1$ is the first moment when a stopping state $> k$ is visited and this state is i.

However, for the plays h satisfying these conditions the payoff $\varphi^{[k]}_{r^{-k}_\zeta}$ is equal to r_i. Thus

$$r_i = \mathbf{E}^{\sigma^k_\zeta,\tau}_k(\varphi^{[k]}_{r^{-k}_\zeta} \mid S_{T^{[k]}_m+1} = i), \quad \text{for } i > k. \tag{4.15}$$

As the second crucial observation let us note the following inequality:

$$\begin{aligned} F^{k-1}_i(\zeta; r) \leqslant\ \zeta\, \mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_{m+1} < \infty \mid S_{T^{[k]}_m+1} = i) \\ + \mathbf{E}^{\sigma^k_\zeta,\tau}_k(\varphi^{[k]}_{r^{-k}_\zeta} \mathbb{1}_{\{T^{[k]}_{m+1}=\infty\}} \mid S_{T^{[k]}_m+1} = i), \quad \text{for } i < k. \end{aligned} \tag{4.16}$$

The proof of (4.16), notationally somehow cumbersome, is postponed for a moment.

Using (4.15) and (4.16), we substitute r_i and $F_i^{k-1}(\zeta; r)$ in (4.14) and we obtain

$$F_k^\sharp(\zeta; r) \leqslant \zeta \sum_{i<k} \mathbf{P}_k^{\sigma_\zeta^k,\tau}(T_{m+1}^{[k]} < \infty \mid S_{T_m^{[k]}+1} = i)\mathbf{P}_k^{\sigma_\zeta^k,\tau}(S_{T_m^{[k]}+1} = i \mid T_m^{[k]} < \infty) \quad \text{(S1)}$$

$$+ \sum_{i<k} \mathbf{E}_k^{\sigma_\zeta^k,\tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mathbb{1}_{\{T_{m+1}^{[k]}=\infty\}} \mid S_{T_m^{[k]}+1} = i)\mathbf{P}_k^{\sigma_\zeta^k,\tau}(S_{T_m^{[k]}+1} = i \mid T_m^{[k]} < \infty) \quad \text{(S2)}$$

$$+ \zeta\, \mathbf{P}_k^{\sigma_\zeta^k,\tau}(S_{T_m^{[k]}+1} = k \mid T_m^{[k]} < \infty) \quad \text{(S3)}$$

$$+ \sum_{i>k} \mathbf{E}_k^{\sigma_\zeta^k,\tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mid S_{T_m^{[k]}+1} = i)\mathbf{P}_k^{\sigma_\zeta^k,\tau}(S_{T_m^{[k]}+1} = i \mid T_m^{[k]} < \infty). \quad \text{(S4)}$$

We shall show that

$$S2 + S4 = \mathbf{E}_k^{\sigma_\zeta^k,\tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mathbb{1}_{\{T_{m+1}^{[k]}=\infty\}} \mid T_m^{[k]} < \infty) \tag{4.17}$$

and

$$S1 + S3 = \zeta \cdot \mathbf{P}_k^{\sigma_\zeta^k,\tau}(T_{m+1}^{[k]} < \infty \mid T_m^{[k]} < \infty). \tag{4.18}$$

To prove (4.17) note that by Bayes' rule

$$\mathbf{E}_k^{\sigma_\zeta^k,\tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mathbb{1}_{\{T_{m+1}^{[k]}=\infty\}} \mid T_m^{[k]} < \infty) =$$
$$\sum_{i=1}^{n} \mathbf{E}_k^{\sigma_\zeta^k,\tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mathbb{1}_{\{T_{m+1}^{[k]}=\infty\}} \mid S_{T_m^{[k]}+1} = i, T_m^{[k]} < \infty)\mathbf{P}_k^{\sigma_\zeta^k,\tau}(S_{T_m^{[k]}+1} = i \mid T_m^{[k]} < \infty).$$

Note that the kth summand can be eliminated from the sum above because

$$\mathbf{E}_k^{\sigma_\zeta^k,\tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mathbb{1}_{\{T_{m+1}^{[k]}=\infty\}} \mid S_{T_m^{[k]}+1} = k, T_m^{[k]} < \infty) = 0. \tag{4.19}$$

Indeed $S_{T_m^{[k]}+1} = k$ means that the $(m+1)$th visit of the stopped state process to k takes place immediately after the mth visit, i.e. $T_{m+1}^{[k]} = T_m^{[k]} + 1 < \infty$, implying $\mathbf{E}_k^{\sigma_\zeta^k,\tau}(\mathbb{1}_{\{T_{m+1}^{[k]}=\infty\}} \mid S_{T_m^{[k]}+1} = k, T_m^{[k]} < \infty) = 0$ and (4.19) follows.

And finally, for $i > k$, $S_{T_m^{[k]}+1} = i$ means that at time $T_m^{[k]} + 1$ the stopped state process hits a stopping state thus $S_t^{[k]}$ will never return to k and therefore $\mathbf{P}_k^{\sigma_\zeta^k,\tau}(T_{m+1}^{[k]} = \infty \mid S_{T_m^{[k]}+1} = i) = 1$ implying

$$\mathbf{E}_k^{\sigma_\zeta^k,\tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mathbb{1}_{\{T_{m+1}^{[k]}=\infty\}} \mid S_{T_m^{[k]}+1} = i, T_m^{[k]} < \infty) =$$
$$\mathbf{E}_k^{\sigma_\zeta^k,\tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mid S_{T_m^{[k]}+1} = i, T_m^{[k]} < \infty), \quad \text{for } i > k.$$

This ends the proof of (4.17).

To prove (4.18), by Bayes' rule we obtain

$$\mathbf{P}_k^{\sigma_\zeta^k,\tau}(T_{m+1}^{[k]} < \infty \mid T_m^{[k]} < \infty) = \\ \sum_{i=1}^{n} \mathbf{P}_k^{\sigma_\zeta^k,\tau}(T_{m+1}^{[k]} < \infty \mid S_{T_m^{[k]}+1} = i, T_m^{[k]} < \infty)\mathbf{P}_k^{\sigma_\zeta^k,\tau}(S_{T_m^{[k]}+1} = i \mid T_m^{[k]} < \infty)$$

As we have already noted $S_{T_m^{[k]}+1} = k$ implies that $T_{m+1}^{[k]} = T_m^{[k]} + 1 < \infty$, i.e.

$$\mathbf{P}_k^{\sigma_\zeta^k,\tau}(T_{m+1}^{[k]} < \infty \mid S_{T_m^{[k]}+1} = k, T_m^{[k]} < \infty) = 1.$$

On the other hand, $S_{T_m^{[k]}+1} = i > k$ implies that $T_{m+1}^{[k]} = \infty$, i.e. $\mathbf{P}_k^{\sigma_\zeta^k,\tau}(T_{m+1}^{[k]} < \infty \mid S_{T_m^{[k]}+1} = i, T_m^{[k]} < \infty) = 0$, which terminates the proof of (4.18).

Now it suffices to notice that (4.17) and (4.18) imply (C1).

It remains to provide the missing proof of (4.16).

For all $t \geqslant 1$ we define the shift mapping,

$$\theta_t : H^\infty \to H^\infty.$$

which "forgets" all history prior to the moment t. Formally,

$$\text{for a } h = s_1, a_1, s_2, a_2, \ldots \in H^\infty, \quad \theta_t(h) = s_t, a_t, s_{t+1}, a_{t+1}, \ldots \ .$$

Consider the event

$$\{S_{T_m^{[k]}+1} = i < k\} \tag{4.20}$$

which consists of the plays that visit k for the mth time at the time $T_m^{[k]}$ and visit $i < k$ at the next time moment $T_m^{[k]} + 1$. Since $S_{T_m^{[k]}+1} = i < k$ implies $T_m^{[k]} < \infty$, for the plays belonging to (4.20) all states visited up to the moment $T_m^{[k]} + 1$ are $\leqslant k$.

Let us examine the following auxiliary game that is played under condition (4.20) and that starts at time $T_m^{[k]} + 1$ when the game visits i, $i < k$. We assume that the payoff applied in the auxiliary game to a play $h \in \{S_{T_m^{[k]}+1} = i < k\}$ is equal to

$$\varphi_{r_\zeta^{-k}}^{[k-1]}(\theta_{T_m^{[k]}+1}(h)),$$

i.e. after removing all history prior to the moment $T_m^{[k]}+1$ we apply to the remaining play the payoff $\varphi_{r_\zeta^{-k}}^{[k-1]}$. Suppose that in the residual game player Max plays according to σ_ζ^k while player Min continues to use the strategy τ.

We claim that

$$F_i^{k-1}(\zeta; r) \leqslant \mathbf{E}_k^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k-1]} \circ \theta_{T_m^{[k]}+1} \mid S_{T_m^{[k]}+1} = i), \quad \text{for } i < k, \tag{4.21}$$

i.e. the expected payoff obtained in the residual game (the right-hand side of (4.21)) is greater or equal to the value of the state i in the $\varphi_{r_\zeta^{-k}}^{[k-1]}$-game (which is $F_i^{k-1}(\zeta; r)$ by the induction hypothesis).

The strategy σ_ζ^k selects the same actions as σ_ζ^{k-1} for all states except k. But in the residual game it is irrelevant how player Max plays in k since for the plays that return to k the residual game is essentially over and player Max obtains the payoff ζ. Thus we can assume as well that in the residual game player Max select actions according to σ_ζ^{k-1}. But since σ_ζ^{k-1} is optimal for Max in the $\varphi_{r_\zeta^{-k}}^{[k-1]}$-game, this guarantees that in the residual game player Max obtains at least the value $F_i^{k-1}(\zeta; r)$ of the state i in the $\varphi_{r_\zeta^{-k}}^{[k-1]}$-game, i.e. (4.21) holds.

Now observe that for the plays $h \in \{T_{m+1}^{[k]} < \infty, S_{T_m^{[k]}+1} = i\}$ we have

$$\varphi_{r_\zeta^{-k}}^{[k-1]} \circ \theta_{T_m^{[k]}+1}(h) = \zeta$$

because k is stopping for the payoff $\varphi_{r_\zeta^{-k}}^{[k-1]}$ and ζ is the reward of k assigned by this payoff. Thus

$$\mathbf{E}_k^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k-1]} \circ \theta_{T_m^{[k]}+1} \mid T_{m+1}^{[k]} < \infty, S_{T_m^{[k]}+1} = i) = \zeta. \tag{4.22}$$

And finally, by Bayes' formula and using (4.21) and (4.22), we obtain

$$\begin{aligned}
&\mathbf{E}_k^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k-1]} \circ \theta_{T_m^{[k]}+1} \mid S_{T_m^{[k]}+1} = i) = \\
&\quad \mathbf{E}_k^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k-1]} \circ \theta_{T_m^{[k]}+1} \mid T_{m+1}^{[k]} < \infty, S_{T_m^{[k]}+1} = i) \cdot \mathbf{P}_k^{\sigma_\zeta^k, \tau}(T_{m+1}^{[k]} < \infty \mid S_{T_m^{[k]}+1} = i) \\
&\quad + \mathbf{E}_i^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k-1]} \circ \theta_{T_m^{[k]}+1} \mid T_{m+1}^{[k]} = \infty, S_{T_m^{[k]}+1} = i) \cdot \mathbf{P}_k^{\sigma_\zeta^k, \tau}(T_{m+1}^{[k]} = \infty \mid S_{T_m^{[k]}+1} = i) \\
&\qquad = \zeta\, \mathbf{P}_k^{\sigma_\zeta^k, \tau}(T_{m+1}^{[k]} < \infty \mid S_{T_m^{[k]}+1} = i) + \mathbf{E}_k^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mathbb{1}_{\{T_{m+1}^{[k]} = \infty\}} \mid S_{T_m^{[k]}+1} = i),
\end{aligned}$$

which terminates the proof of (4.16).

Proof of (C3):

Suppose that $\zeta < F_k^{\sharp}(\zeta; r)$.

Since $\varphi^{[k]}_{r^{-k}_\zeta} \leqslant 1$, from (C1) we obtain

$$\zeta \cdot \mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_{m+1} < \infty \mid T^{[k]}_m < \infty) + \mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_{m+1} = \infty \mid T^{[k]}_m < \infty) \geqslant F^\sharp_k(\zeta; r).$$

But $\mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_{m+1} = \infty \mid T^{[k]}_m < \infty) + \mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_{m+1} < \infty \mid T^{[k]}_m < \infty) = 1$, thus

$$\mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_{m+1} < \infty \mid T^{[k]}_m < \infty) \leqslant \frac{1 - F^\sharp_k(\zeta; r)}{1-\zeta} < 1.$$

Therefore

$$\begin{aligned}
\mathbf{P}^{\sigma^k_\zeta,\tau}_k(\forall m, T^{[k]}_m < \infty) &= \lim_{m\to\infty} \mathbf{P}^{\sigma^k_\zeta,\tau}_k(\forall i \leqslant m, T^{[k]}_i < \infty) \\
&= \lim_{m\to\infty} \mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_0 < \infty) \cdot \prod_{q=0}^{m-1} \mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_{q+1} < \infty \mid T^{[k]}_q < \infty) \\
&\leqslant \lim_{m\to\infty} \left(\frac{1 - F^\sharp_k(\zeta; r)}{1-\zeta}\right)^{m-1} \\
&= 0,
\end{aligned} \tag{4.23}$$

i.e. if player Max uses σ^k_ζ then almost surely k is visited only finitely many times.

Proof of (C2):
From (4.10) and (C1) it follows that

$$\begin{aligned}
F^\sharp_k(\zeta; r) \leqslant F^\sharp_k(\zeta; r) \cdot \mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_{m+1} < \infty \mid T^{[k]}_m < \infty) + \\
\mathbf{E}^{\sigma^k_\zeta,\tau}_k(\varphi^{[k]}_{r^{-k}_\zeta} \mathbb{1}_{\{T^{[k]}_{m+1} = \infty\}} \mid T^{[k]}_m < \infty)
\end{aligned}$$

which implies

$$F^\sharp_k(\zeta; r) \cdot \mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_{m+1} = \infty \mid T^{[k]}_m < \infty) \leqslant \mathbf{E}^{\sigma^k_\zeta,\tau}_k(\varphi^{[k]}_{r^{-k}_\zeta} \mathbb{1}_{\{T^{[k]}_{m+1} = \infty\}} \mid T^{[k]}_m < \infty).$$

Multiplying both sides by $\mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_m < \infty)$ we obtain

$$F^\sharp_k(\zeta; r) \cdot \mathbf{P}^{\sigma^k_\zeta,\tau}_k(T^{[k]}_{m+1} = \infty, T^{[k]}_m < \infty) \leqslant \mathbf{E}^{\sigma^k_\zeta,\tau}_k(\varphi^{[k]}_{r^{-k}_\zeta} \mathbb{1}_{\{T^{[k]}_{m+1} = \infty\}} \mathbb{1}_{\{T^{[k]}_m < \infty\}}). \tag{4.24}$$

Let us note that

$$\{\exists m, T_m^{[k]} = \infty\} = \{T_1^{[k]} = \infty\} \cup \bigcup_{i=1}^{\infty} \{T_{m+1}^{[k]} = \infty, T_m^{[k]} < \infty\},$$

where the events on the right-hand side are pairwise disjoint. Moreover,

$$\{S_1 = k\} \cap \{T_1^{[k]} = \infty\} = \varnothing$$

since if the game starts at k then $T_1^{[k]} = 1 < \infty$.

This implies that summing over m both sides of (4.24) we get

$$F_k^\sharp(\zeta; r) \cdot \mathbf{P}_k^{\sigma_\zeta^k, \tau}(\exists m, T_m^{[k]} = \infty) \leqslant \mathbf{E}_k^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mathbb{1}_{\{\exists m, T_m^{[k]} = \infty\}}). \tag{4.25}$$

Thus

$$\begin{aligned}\mathbf{E}_k^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k]}) &= \mathbf{E}_k^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mathbb{1}_{\{\exists m, T_m^{[k]} = \infty\}}) + \mathbf{E}_k^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k]} \mathbb{1}_{\{\forall m, T_m^{[k]} < \infty\}}) \\ &\geqslant F_k^\sharp(\zeta; r) \cdot \mathbf{P}_k^{\sigma_\zeta^k, \tau}(\exists m, T_m^{[k]} = \infty) + \zeta \cdot \mathbf{P}_k^{\sigma_\zeta^k, \tau}(\forall m, T_m^{[k]} < \infty),\end{aligned} \tag{4.26}$$

where the last inequality follows from (4.25) and from the fact that $\varphi_{r_\zeta^{-k}}^{[k]}(h) = \zeta$ for the plays such that $h \in \{\forall m, T_m^{[k]} < \infty\}$ (i.e. for the plays for which the stopping state process $S_i^{[k]}$ visits k infinitely often).

If $F_k^\sharp(\zeta; r) > \zeta$ then, by (C3), $\mathbf{P}_k^{\sigma_\zeta^k, \tau}(\forall m, T_m^{[k]} < \infty) = 0$ and thus $\mathbf{P}_k^{\sigma_\zeta^k, \tau}(\exists m, T_m^{[k]} = \infty) = 1$ and (4.26) implies (C2).

Similarly, if $F_k^\sharp(\zeta; r) = \zeta$ then (4.26) implies also

$$\mathbf{E}_k^{\sigma_\zeta^k, \tau}(\varphi_{r_\zeta^{-k}}^{[k]}) \geqslant \zeta = F_k^\sharp(\zeta; r).$$

This ends the proof of (C2). □

Lemma 4.12. *Assume that* ***(H.1)*** *and* ***(H.2)*** *are satisfied. Let* $\varnothing \neq D \subset [0,1]$ *and let* $\sigma_\star$ *be a pure memoryless strategy of* Max *optimal in the* $\varphi_{(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)}^{[k-1]}$*-game for all* $\xi \in D$*. Let* $w = \sup D$ *be the supremum of* D*. Then* $\sigma_\star$ *is optimal in the* $\varphi_{(r_1,\dots,r_{k-1},w,r_{k+1},\dots,r_n)}^{[k-1]}$*-game.*

Proof. By the assumptions of the lemma, for each state $i \in [k-1]$, each strategy τ of Min and each $\xi \in D$ we have

$$\mathbf{E}_i^{\sigma_\star, \tau}(\varphi_{(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)}^{[k-1]}) \geqslant \mathbf{Fix}_i^{k-1}(f)(r_1, \dots, r_{k-1}, \xi, r_{k+1}, \dots, r_n). \tag{4.27}$$

Since $\xi \leqslant w$ implies $\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)} \leqslant \varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w,r_{k+1},\dots,r_n)}$ we have

$$\mathbf{E}_i^{\sigma_\star,\tau}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)}) \leqslant \mathbf{E}_i^{\sigma_\star,\tau}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w,r_{k+1},\dots,r_n)}). \tag{4.28}$$

From (4.27) and (4.28)

$$\mathbf{E}_i^{\sigma_\star,\tau}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w,r_{k+1},\dots,r_n)}) \geqslant \mathbf{Fix}_i^{k-1}(f)(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n).$$

But $\mathbf{Fix}_i^{k-1}(f)(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)$ is a nonexpansive function of

$$(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)$$

and nonexpansive functions are also continuous which implies that if for some $a \in \mathbb{R}$,

$$a \geqslant \mathbf{Fix}_i^{k-1}(f)(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)$$

for all $\xi \in D$ then also

$$a \geqslant \mathbf{Fix}_i^{k-1}(f)(r_1,\dots,r_{k-1},w,r_{k+1},\dots,r_n)$$

for $w = \sup D$. In particular

$$\mathbf{E}_i^{\sigma_\star,\tau}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w,r_{k+1},\dots,r_n)}) \geqslant \mathbf{Fix}_i^{k-1}(f)(r_1,\dots,r_{k-1},w,r_{k+1},\dots,r_n).$$

□

In the following lemma we construct an optimal pure memoryless strategy for player Max in the $\varphi^{[k]}_r$-game.

Lemma 4.13. *Suppose that* ***(H.1)*** *and* ***(H.2)*** *are satisfied.*

Then for each reward vector $r \in [0,1]^n$ there exists a pure memoryless strategy $\sigma^k_\star$ for player Max *such that for each strategy τ of player* Min *and each state $i \in [k]$ we have*

$$\mathbf{Fix}_i^{k}(f)(r) \leqslant \mathbf{E}_i^{\sigma^k_\star,\tau}(\varphi^{[k]}_r). \tag{4.29}$$

Proof. Let us note

$$w := \mathbf{Fix}_k^{k}(f)(r).$$

As we did in the proof of Lemma 4.11, r_y^{-k} will be used to denote the reward vector $(r_1,\dots,r_{k-1},y,r_{k+1},\dots,r_n)$, for $y \in [0,1]$.

We first prove that there exists a pure memoryless strategy $\sigma^k_\star$ for player Max such that

(K.1) $\sigma^k_\star$ is optimal for Max in the $\varphi^{[k-1]}_{r_w^{-k}}$-game and

(K.2) $\mathbf{Fix}_k^k(f)(r) \leqslant \mathbf{E}_k^{\sigma_\star^k,\tau}(\varphi_r^{[k]})$.

As in the previous lemma we set

$$F_k^\sharp(\zeta;r) := f_k(F_1^{k-1}(\zeta;r),\ldots,F_{k-1}^{k-1}(\zeta;r),\zeta,r_{k+1},\ldots,r_n)$$

so that

$$\mathbf{Fix}_k^k(r) = \mu_{r_k}\zeta.F_k^\sharp(\zeta;r).$$

We examine three different cases.

Case 1: $r_k > \mathbf{Fix}_k^k(f)(r)$.
Since w is a fixed point of $F_k^\sharp$ we have

$$w = F_k^\sharp(w;r).$$

By Lemma 4.11 the last equality implies that player Max has a pure memoryless strategy $\sigma_\star^k$ which is optimal in the $\varphi_{r_w^{-k}}^{[k-1]}$-game and such that, for each strategy τ of player Min,

$$F_k^\sharp(w;r) \leqslant \mathbf{E}_k^{\sigma_\star^k,\tau}(\varphi_{r_w^{-k}}^{[k]}).$$

Now it suffices to note that $w < r_k$ implies that for all plays h,

$$\varphi_{r_w^{-k}}^{[k]}(h) \leqslant \varphi_r^{[k]}(h)$$

and therefore $\mathbf{E}_k^{\sigma_\star^k,\tau}(\varphi_{r_w^{-k}}^{[k]}) \leqslant \mathbf{E}_k^{\sigma_\star^k,\tau}(\varphi_r^{[k]})$ and we conclude that

$$\mathbf{Fix}_k^k(f)(r) \leqslant \mathbf{E}_k^{\sigma_\star^k,\tau}(\varphi_r^{[k]}).$$

Case 2: $r_k = \mathbf{Fix}_k^k(f)(r)$.
Immediately from Lemma 4.11 with $\zeta = r_k$.

Case 3: $r_k < \mathbf{Fix}_k^k(f)(r)$.
Since

$$\mathbf{Fix}_k^k(f)(r) = \mu_{r_k}\zeta.F_k^\sharp(\zeta;r) > r_k$$

by (3.1) applied to the mapping

$$\zeta \mapsto F_k^\sharp(\zeta;r)$$

$\mathbf{Fix}_k^k(f)(r)$ is in fact the least fixed point of this mapping. This implies that

$$F_k^\sharp(\xi;r) > \xi$$

for all ξ such that

$$r_k < \xi < \mathbf{Fix}_k^k(f)(r). \tag{4.30}$$

($\xi \mapsto F_k^\sharp(\xi;r)$ is strictly increasing for the arguments smaller than the least fixed point).

By Lemma 4.11 player Max has a pure memoryless strategy σ_ξ^k such that

(W.1) σ^k_ξ is optimal in the $\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)}$-game,

(W.2) $\mathbf{E}^{\sigma^k_\xi,\tau}_k(\varphi^{[k]}_{(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)}) \geqslant F^\sharp_k(\xi;r) > \xi$, and

(W.3) $\mathbf{P}^{\sigma^k_\xi,\tau}_k(S^{[k]}_i = k \text{ for infinitely many } i) = 0$ for all strategies τ of player Min.

Now it suffices to observe that the payoff mappings $\varphi^{[k]}_r$ and $\varphi^{[k]}_{(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)}$ differ only for the plays belonging to the set $\{S^{[k]}_i = k \text{ for infinitely may } i\}$ and this set has measure zero by **(W.3)**. Thus

$$\mathbf{E}^{\sigma^k_\xi,\tau}_k(\varphi^{[k]}_r) = \mathbf{E}^{\sigma^k_\xi,\tau}_k(\varphi^{[k]}_{(r_1,\dots,r_{k-1},\xi,r_{k+1},\dots,r_n)}).$$

Therefore, by **(W.2)**,

$$\mathbf{E}^{\sigma^k_\xi,\tau}_k(\varphi^{[k]}_r) \geqslant F^\sharp_k(\xi;r) > \xi. \tag{4.31}$$

For each pure memoryless strategy σ^k of player Max let

$$D(\sigma^k) = \{\xi \mid r_k < \xi < \mathbf{Fix}^k_k(f)(r) \text{ and } \sigma^k = \sigma^k_\xi\},$$

where, for each ξ, σ^k_ξ is a pure memoryless strategy for player Max satisfying (4.31) and **(W.1)**.

Since there is a finite number of pure memoryless strategies and each ξ such that $r_k < \xi < \mathbf{Fix}^k_k(f)(r)$ belongs to some $D(\sigma^k)$ there exists a pure memoryless strategy $\sigma^k_\star$ such that $\mathbf{Fix}^k_k(f)(r)$ is an accumulation point of $D(\sigma^k_\star)$. The elements of $D(\sigma^k_\star)$ are smaller than $\mathbf{Fix}^k_k(f)(r)$ thus, in fact, this accumulation point is the supremum of $D(\sigma^k_\star)$, i.e.

$$\mathbf{Fix}^k_k(f)(r) = \sup D(\sigma^k_\star).$$

Since, by (4.31), $\mathbf{E}^{\sigma^k_\star,\tau}_k(\varphi^{[k]}_r) > \xi$ for all $\xi \in D(\sigma^k_\star)$, we have also

$$\mathbf{E}^{\sigma^k_\star,\tau}_k(\varphi^{[k]}_r) \geqslant \sup D(\sigma^k_\star) = \mathbf{Fix}^k_k(f)(r).$$

Note also that, by Lemma 4.12, $\sigma^k_\star$ is optimal for player Max in the $\varphi^{[k-1]}_{r^{-k}_w}$-game.

This ends the proof of **(K.1)** and **(K.2)**.

To prove (4.29) for $i < k$ we proceed as follows.

By the induction hypothesis **(H.1)**, $\mathbf{Fix}^{k-1}_i(f)(r_1,\dots,r_{k-1},w,r_{k+1},\dots,r_n)$ is the value of state i in the $\varphi^{[k-1]}_{r^{-k}_w}$-game and, by **(K.1)**, $\sigma^k_\star$ is optimal in the same game, thus

$$\mathbf{Fix}^{k-1}_i(f)(r_1,\dots,r_{k-1},w,r_{k+1},\dots,r_n) \leqslant \mathbf{E}^{\sigma^k_\star,\tau}_i(\varphi^{[k-1]}_{r^{-k}_w}). \tag{4.32}$$

By Bayes' rule

$$\mathbf{E}_i^{\sigma_\star^k,\tau}(\varphi_{r_w^{-k}}^{[k-1]}) = \mathbf{E}_i^{\sigma_\star^k,\tau}(\varphi_{r_w^{-k}}^{[k-1]} \mid T_1^{[k]} < \infty)\mathbf{P}_i^{\sigma_\star^k,\tau}(T_1^{[k]} < \infty) + \mathbf{E}_i^{\sigma_\star^k,\tau}(\varphi_{r_w^{-k}}^{[k-1]} \mid T_1^{[k]} = \infty)\mathbf{P}_i^{\sigma_\star^k,\tau}(T_1^{[k]} = \infty) \tag{4.33}$$

where $T_1^{[k]}$ is as in Definition 4.10.

The plays satisfying $T_1^{[k]} = \infty$ never visit k thus for such plays it is irrelevant what is the reward of k and it is irrelevant if k is stopping or not, in particular we have

$$\mathbf{E}_i^{\sigma_\star^k,\tau}(\varphi_{r_w^{-k}}^{[k-1]} \mid T_1^{[k]} = \infty) = \mathbf{E}_i^{\sigma_\star^k,\tau}(\varphi_r^{[k]} \mid T_1^{[k]} = \infty). \tag{4.34}$$

The plays satisfying $T_1^{[k]} < \infty$ visit k and the states visited prior to the moment of the first visit to k are all $< k$. For such plays $\varphi_{r_w^{-k}}^{[k-1]}$ is equal to w implying

$$\mathbf{E}_i^{\sigma_\star^k,\tau}(\varphi_{r_w^{-k}}^{[k-1]} \mid T_1^{[k]} < \infty) = w. \tag{4.35}$$

On the other hand,

$$\mathbf{E}_i^{\sigma_\star^k,\tau}(\varphi_r^{[k]} \mid T_1^{[k]} < \infty) \geqslant w. \tag{4.36}$$

Indeed, we have $T_1^{[k]} < \infty$ for the plays that visit k and such that before the first visit to k all visited states were $< k$. For such plays the value of $\varphi_r^{[k]}$ does not depend on the history prior to the first visit to k. But by **(K.2)**, starting from k the strategy $\sigma_\star^k$ guarantees the expected payoff of at least $\mathbf{Fix}_k^k(f)(r) = w$ against any strategy of Min.

From (4.32),(4.33),(4.34),(4.35) and (4.36) we obtain

$$\mathbf{Fix}_i^{k-1}(f)(r_1, \dots, r_{k-1}, w, r_{k+1}, \dots, r_n) \leqslant \mathbf{E}_i^{\sigma_\star^k,\tau}(\varphi_r^{[k]} \mid T_1^{[k]} < \infty)\mathbf{P}_i^{\sigma_\star^k,\tau}(T_1^{[k]} < \infty) + \mathbf{E}_i^{\sigma_\star^k,\tau}(\varphi_r^{[k]} \mid T_1^{[k]} = \infty)\mathbf{P}_i^{\sigma_\star^k,\tau}(T_1^{[k]} = \infty) = \mathbf{E}_i^{\sigma_\star^k,\tau}(\varphi_r^{[k]}).$$

And now it remains to note that the definition of the nested nearest fixed point gives

$$\mathbf{Fix}_i^{k-1}(f)(r_1, \dots, r_{k-1}, w, r_{k+1}, \dots, r_n) = \mathbf{Fix}_i^{k-1}(f)(r_1, \dots, r_{k-1}, \mathbf{Fix}_k^k(f)(r), r_{k+1}, \dots, r_n) = \mathbf{Fix}_i^k(f)(r).$$

□

4.4.2 Dual games

In Section 3.3 we have defined the dual of the BMN mappings. In this section we define and examine the corresponding notion for the priority games.

Given an arena $\mathcal{A}$ the dual arena $\overline{\mathcal{A}}$ is defined in the following way:

- $\overline{\mathcal{A}}$ has the same states, actions and transition probabilities as $\mathcal{A}$,
- all states controlled by Max in $\mathcal{A}$ are controlled by Min in $\overline{\mathcal{A}}$,
- all states controlled by Min in $\mathcal{A}$ are controlled by Max in $\overline{\mathcal{A}}$.

From this definition it follows immediately that each strategy σ of player Max (respectively a strategy τ of Min) in $\mathcal{A}$ becomes a strategy of player Min (respectively Max) in $\overline{\mathcal{A}}$ and vice versa. Moreover, we have the equality of the corresponding induced probabilities,

$$\mathbf{P}_i^{\sigma,\tau}(\,\cdot\,;\mathcal{A}) = \mathbf{P}_i^{\tau,\sigma}(\,\cdot\,;\overline{\mathcal{A}}),$$

where the left-hand side denotes the probability induced on plays in $\mathcal{A}$ while the right-hand side denotes the probability on plays in $\overline{\mathcal{A}}$.

For each reward vector r, by $1-r$ we denote the reward vector $(1-r_1,\dots,1-r_n)$. Since for each play $h \in H^\infty$, $\varphi_r^{[k]}(h) = 1 - \varphi_{1-r}^{[k]}(h)$, we have the following equality concerning the expected payoffs for the (stopped) priority games played on $\mathcal{A}$ and $\overline{\mathcal{A}}$:

$$\mathbf{E}_i^{\sigma,\tau}(\varphi_r^{[k]};\mathcal{A}) = 1 - \mathbf{E}_i^{\tau,\sigma}(\varphi_{1-r}^{[k]};\overline{\mathcal{A}}). \tag{4.37}$$

This motivates the following definition.

Given a stopped priority game $(\mathcal{A},\varphi_r^{[k]})$ the *dual game* is the stopped priority game $(\overline{\mathcal{A}},\varphi_{1-r}^{[k]})$.

Note that a strategy σ is optimal for player Max in the game $(\mathcal{A},\varphi_r^{[k]})$ if and only if σ is optimal for player Min in the dual game $(\overline{\mathcal{A}},\varphi_{1-r}^{[k]})$.

A similar statement holds for strategies τ of Min. Therefore we have also the following equality for the game values:

$$\mathrm{val}_i(\mathcal{A},\varphi_r^{[k]}) = 1 - \mathrm{val}_i(\overline{\mathcal{A}},\varphi_{1-r}^{[k]}),$$

where $\mathrm{val}_i(\mathcal{A},\varphi_r^{[k]})$ is the value of state i in the original stopped priority game while $\mathrm{val}_i(\overline{\mathcal{A}},\varphi_{1-r}^{[k]})$ is the value of i in the dual game.

4.4.3 The duality of value mappings meets the duality of games

Recall the definition of a dual mapping given in Section 3.3, $\overline{f}(r) = 1 - f(1-r)$, where for $r = (r_1,\dots,r_n)$, $1-r = (1-r_1,\dots,1-r_n)$.

Lemma 4.14. *Let $f : [0,1]^n \to [0,1]^n$ be the value function of the one-step game, cf.* (4.2).

Then the dual mapping $\overline{f}$ is the value function of the one-step game played on the dual arena.

Proof. Let k be a state controlled by player Max in the dual arena $\overline{\mathcal{A}}$. Thus k is controlled by player Min in the original arena.

The value of state k for the one-step game played at k on the dual arena with reward vector r is

$$\max_{a \in \mathbf{A}(k)} \sum_i p(i|k,a) \cdot r_i = \max_{a \in \mathbf{A}(k)} (1 - \sum_i p(i|k,a) \cdot (1 - r_i)) =$$

$$1 - \min_{a \in \mathbf{A}(k)} \sum_i p(i|k,a) \cdot (1 - r_i) = 1 - f_k(1 - r) = \overline{f}_k(r).$$

Interchanging max and min we get the result when k is controlled by player Min in the dual arena. □

The duality leads directly to the following counterpart of Lemma 4.13.

Lemma 4.15. *Suppose that* ***(H.1)*** *and* ***(H.2)*** *are satisfied. For each reward vector $r \in [0,1]^n$ there exists a pure memoryless strategy $\tau_\star^k$ for player* Min *such that for each strategy σ of player* Max *and each $i \in [k]$ we have*

$$\mathbf{E}_i^{\sigma,\tau_\star^k}(\varphi_r^{[k]}) \leqslant \mathbf{Fix}_i^k(f)(r). \tag{4.38}$$

Proof. In the proof we will go back and forth between the priority game $(\mathcal{A}, \varphi_r^{[k]})$ and its dual $(\overline{\mathcal{A}}, \varphi_{1-r}^{[k]})$. To avoid ambiguity when we speak about the players then Max and Min are the maximizer and the minimizer in the original priority game while the maximizer and the minimizer in the dual game are named $\overline{\text{Max}}$ and $\overline{\text{Min}}$ respectively.

From Lemma 4.13 applied to the dual game we deduce that there exists a pure memoryless strategy $\tau_\star^k$ for player $\overline{\text{Max}}$ such that for each strategy σ of player $\overline{\text{Min}}$ and each state i,

$$\mathbf{E}_i^{\tau_\star^k,\sigma}(\varphi_{1-r}^{[k]}; \overline{\mathcal{A}}) \geqslant \mathbf{Fix}_i^k(\overline{f})(1 - r). \tag{4.39}$$

By Lemma 3.11,

$$\mathbf{Fix}_i^k(\overline{f})(1 - r) = 1 - \mathbf{Fix}_i^k(f)(r). \tag{4.40}$$

Using (4.37), (4.39) and (4.40) we obtain

$$\mathbf{E}_i^{\sigma,\tau_\star^k}(\varphi_r^{[k]}; \mathcal{A}) = 1 - \mathbf{E}_i^{\tau_\star^k,\sigma}(\varphi_{1-r}^{[k]}; \overline{\mathcal{A}}) \leqslant 1 - \mathbf{Fix}_i^k(\overline{f})(1 - r) = \mathbf{Fix}_i^k(f)(r).$$

□

Therefore we obtain finally:

Proof of Theorem 4.7. By Lemma 4.13 and Lemma 4.15. □

4.5 Remarks on priority games with infinite action or state sets

A turn-based stochastic priority game with an infinite number of actions may not have memoryless optimal strategies.

Let us consider the priority game player on the arena depicted on Figure 4.16. All states are controlled by player Max, $\mathbf{S} = \mathbf{S}^{\mathrm{Max}} = \{1, 2, 3\}$. State 2 is absorbing, state 3 has just one available action that leads to state 1 with probability 1.

State 1 has an infinite number of available actions $\mathbf{A}(1) = \{a_1, a_2, \ldots\}$ such that for all $i \geqslant 1$, $p(2|1, a_i) = \frac{1}{2^i}$ and $p(3|1, a_i) = 1 - \frac{1}{2^i}$. The reward vector is such that $r_1 = 0$, $r_2 = 0$ and $r_3 = 1$.

The value of state 1 is 1. But there does not exist a memoryless optimal strategy for player Max. In fact, for each memoryless strategy of Max the probability to reach state 2 is 1 which results in payoff 0. Moreover, player Max has no strategy (even with memory) securing the expected payoff 1. However, for each $\varepsilon > 0$, he has a strategy, which is not memoryless, securing for him the expected payoff of at least $1 - \varepsilon$. In fact, let $N \in \mathbb{N}$ be such that $1/2^{N-1} < \varepsilon$, and let be a strategy of player Max such that he plays action a_{N+i} if the game visited state 1 i times, then the probability to visit state 2 is $1/2^N + 1/2^{N+1} + 1/2^{N+2} + \ldots$ that converges to $1/2^{N-1} < \varepsilon$. Hence, the probability to visit state 3 infinitely often is $> 1 - \varepsilon$.

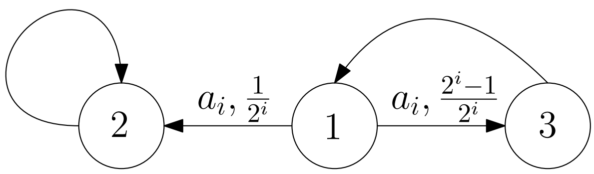

Figure 4.16 – Game with infinite set of actions where player Max does not have memoryless optimal strategy.

We can also consider priority games with an infinite number of states. To this end we first need to adapt the definition of the priority games to such a framework [3].

3. If the set of states is the set $\mathbb{N}$ of all natural numbers then $\limsup s_t$, where $s_1, s_2, \ldots$ is the infinite sequence of visited states can be equal to ∞, and the priority payoff of such a play is undefined if we try to apply the definition of Section 4.1.

A natural way to define a priority game with an infinite number of states is the one used for parity games.

Let $\mathbf{S}$ be an infinite set of states such that for each $s \in \mathbf{S}$ the set $\mathbf{A}(s)$ of actions available at s is finite. The game is played by two players, Max and Min, and each state is controlled by one of the players.

We assume that the arena is endowed with a *priority mapping*

$$\pi : \mathbf{S} \to \{1, \ldots, \ell\}$$

from states to a finite set of natural numbers.

The reward mapping

$$r : \{1, \ldots, \ell\} \to [0, 1]$$

maps priorities to the unit interval $[0, 1]$.

For each play $h = s_1, a_1, s_2, a_2 \ldots$, the priority payoff mapping is defined as

$$\varphi(h) = r_k, \quad \text{where } k = \limsup_t (\pi(s_t)).$$

Thus the payoff is the reward associated with the highest priority visited infinitely often.

Let us consider the priority stochastic game depicted on Figure 4.17. All states are controlled by player Max, $\mathbf{S} = \mathbf{S}^{\text{Max}} = \{s_d, s_w, s_1, s_2, \ldots\}$. The priorities are $\pi(s_d) = 0$, $\pi(s_w) = 1$ and, for all $i \geqslant 1$, $\pi(s_i) = 0$. The following rewards are assigned to the priorities: $r_0 = 0$ and $r_1 = 1$.

The game has the following actions: for all $i \geqslant 1$, $\mathbf{A}(s_i) = \{a, b\}$ and $p(s_d|s_i, a) = \frac{1}{2^i}$, $p(s_w|s_i, a) = 1 - \frac{1}{2^i}$ and $p(s_{i+1}|s_i, b) = 1$. State s_w has just a deterministic action a that moves to s_1 and state s_d is absorbing.

The value of the game for the initial state s_1 is 1. But

- for each memoryless strategy of player Max the expected payoff is 0 and
- player Max has no strategy securing the expected payoff 1 (but, as in the last game, for each $\varepsilon > 0$ player Max has a non-memoryless strategy σ securing for him the expected payoff of at least $1 - \varepsilon$). Strategy σ is built as follows: Let i be the times that state s_w was visited and let $N \in \mathbb{N}$ be such that $1/2^{N-1} < \varepsilon$. Then $\sigma(h)(a) = 1$ when $h = h' s_{N+i}$ and $\sigma(h)(b) = 1$ otherwise.

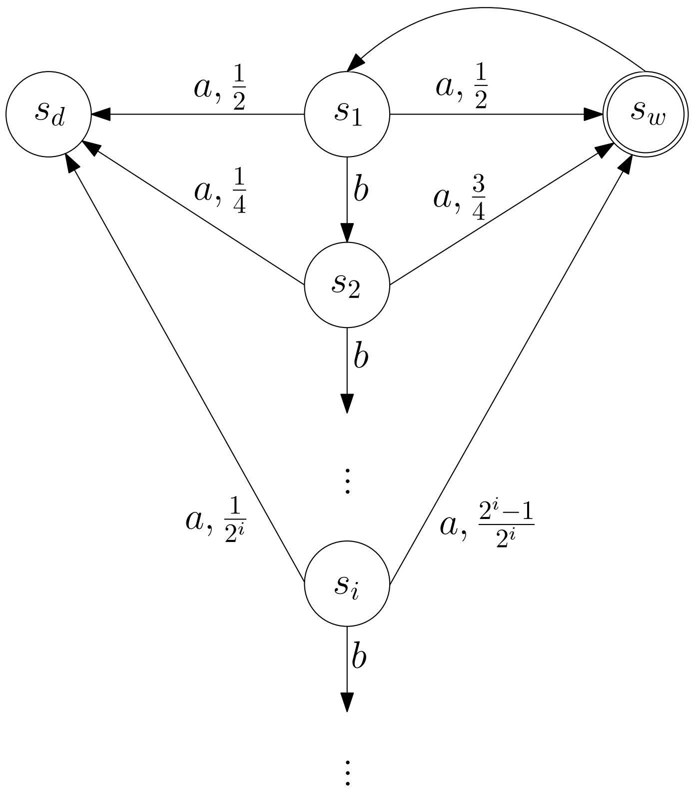

Figure 4.17 – Game with infinite set of states where player Max does not have optimal strategy.

Chapter 5

Concurrent stochastic priority games

In this chapter we study concurrent stochastic priority games.

Contrary to the turn-based stochastic games, in concurrent stochastic games, a given state is not controlled by any particular player. What happens instead is that the states are controlled jointly by both players. At each state both players choose actions independently and simultaneously and the probability to move to the next state depends on the actions chosen by both players.

The fact that the players choose actions simultaneously and independently at each stage has a significant impact on how the game is played. It turns out that in concurrent stochastic priority games, the players do not have optimal strategies, in general. However, they have ε-optimal strategies. But these strategies are neither pure nor memoryless [1].

The main result of this chapter is that the values of the concurrent stochastic priority games can be obtained as a nested nearest fixed point of appropriate monotone nonexpansive mapping. This result is analogous to the main result of the previous chapter. However, the proof is technically more involved, since we need to cope with the uncertainty due to the fact that the adversary player chooses actions independently and simultaneously at each state.

If the only possible rewards are 0 and 1, then the concurrent stochastic priority game is the same as the concurrent parity game examined by de Alfaro and Majumdar [dAM04]. These authors proved that the value of such game is given by a μ-calculus formula alternating the least and the greatest fixed points. Thus the result of this chapter is an extension of the result obtained in [dAM04], the only difference is that we replace greatest and least fixed points used in [dAM04] by the nearest fixed points.

1. See, for example, a game adapted by de Alfaro and Henzinger [dAH00] from [KS81] where both players do not have optimal strategies and for one of the players a ε-optimal strategy cannot be memoryless.

Our proof is however quite different. The proof of [dAM04] is not inductive. De Alfaro and Majumdar give a complete μ-calculus formula with all fixed points applied from the outset and show that this formula gives the values of all states in the concurrent parity game.

On the other hand, in our approach we provide a game interpretation of the nested fixed point formula where only some variables are bound by the fixed point while other variables are free. It turns out that such formula represents the values of the priority game where free variable correspond to absorbing states.

This approach makes our proof more structured than that of [dAM04].

Roughly speaking, we start with a trivial game where all states are absorbing. And next we transform the states, one by one, starting from the lowest priority state 1, next state 2, etc., from absorbing to nonabsorbing. We show by induction that, if f is the value mapping of the one-step game, then

$$\mu_{r_k} x_k. \dots \mu_{r_1} x_1. f(x_1, \dots, x_k, r_{k+1}, \dots, r_n), \tag{5.1}$$

where the free variables $x_{k+1}, \dots, x_n$ are evaluated to $r_{k+1}, \dots, r_n$, is the value vector of the priority game where the states $1, \dots, k$ are nonabsorbing while states $k + 1, \dots, n$ are transformed into absorbing states.

With this approach it suffices to show that solving the priority game where the states $1, \dots, k, k+1$ are nonabsorbing while the states $k+2, \dots, n$ are absorbing, corresponds to add the next r_{k+1}-nearest fixed point $\mu_{r_{k+1}} x_{k+1}$ to (5.1).

In this way we do not need to examine a fixed point formula where all n fixed points are applied at once. Instead, we just examine what happens if just one fixed point is added to (5.1).

The chapter is structured as follows. In Section 5.1 we define the *concurrent* stochastic priority games.

Section 5.2 defines and examines *one-step games*. These games are auxiliary matrix games played at each state. The crucial observation concerning the one-step game is that its value mapping f is monotone nonexpansive.

In Section 5.3, we define and examine the class of stopping concurrent priority games. In such games, all states greater[2] than a fixed state k are absorbing (or equivalently stopping). We prove by induction that (5.1) is the value vector of this game.

As a corollary we obtain that the values of concurrent priority games can be expressed as the nested nearest fixed points (without free variables).

2. greater in the priority order

5.1 Concurrent stochastic priority games

An arena for a two-player concurrent stochastic priority game is composed of a finite set of states $\mathbf{S} = [n] = \{1, 2, \ldots, n\} \subset \mathbb{N}$ (we assume without loss of generality that $\mathbf{S}$ is a subset of positive integers) and finite sets $\mathbf{A}$ and $\mathbf{B}$ of actions of players Max and Min. For each state i, $\mathbf{A}(i) \subseteq \mathbf{A}$ and $\mathbf{B}(i) \subseteq \mathbf{B}$ are the sets of actions that players Max and Min can play at s. We assume that $\mathbf{A}$ and $\mathbf{B}$ are disjoint and $(\mathbf{A}(i))_{i \in \mathbf{S}}$, $(\mathbf{B}(i))_{i \in \mathbf{S}}$ are partitions of $\mathbf{A}$ and $\mathbf{B}$.

For $i, j \in \mathbf{S}, a \in \mathbf{A}(i), b \in \mathbf{B}(i)$, $p(j|i, a, b)$ is the probability to move to j if players Max and Min execute respectively actions a and b at i.

An infinite game is played by players Max and Min. At each stage, given the current state i, the players choose simultaneously and independently actions $a \in \mathbf{A}(i)$ and $b \in \mathbf{B}(i)$ and the game moves to a new state j with probability $p(j|i, a, b)$. The couple (a, b) is called the *joint action*.

A finite history is a sequence $h = s_1, (a_1, b_1), s_2, (a_2, b_2), s_3 \ldots, s_t$ alternating states and joint actions and beginning and ending with a state. The length of h is the number of joint actions in h, in particular a history of length 0 consists of just one state and no actions. The set of finite histories is denoted H.

A strategy of player Max is a mapping $\sigma : H \to \Delta(\mathbf{A})$, where $\Delta(\mathbf{A})$ denotes the set of probability distributions over $\mathbf{A}$. We require that $\mathrm{supp}(\sigma(h)) \subseteq \mathbf{A}(i)$, where i is the last state of h and $\mathrm{supp}(\sigma(h)) := \{a \in \mathbf{A} \mid \sigma(h)(a) > 0\}$ is the support of the measure $\sigma(h)$.

A strategy σ is *memoryless* if $\sigma(h)$ depends only on the last state of h. Thus memoryless strategies of player Max can be identified with mappings from $\mathbf{S}$ to $\Delta(\mathbf{A})$ such that $\mathrm{supp}(\sigma(i)) \subseteq \mathbf{A}(i)$ for each $i \in \mathbf{S}$.

A strategy σ is *pure* if $\mathrm{supp}(\sigma(h))$ is a singleton for each h. Pure memoryless strategies of player Max are identified with mappings $\sigma : \mathbf{S} \to \mathbf{A}$ such that $\sigma(i) \in \mathbf{A}(i)$.

Strategies for player Min are defined in a similar way.

We write Σ and $\mathcal{T}$ to denote the sets of all strategies for player Max and Max respectively.

We use σ and τ (with subscripts or superscripts) to denote strategies of players Max and Min respectively.

An infinite history or a play is an infinite sequence
$h = s_1, (a_1, b_1), s_2, (a_2, b_2), s_3, (a_3, b_3), \ldots$ alternating states and joint actions. The set of infinite histories is denoted H^∞. For a finite history h, by h^+ we denote the cylinder generated by h consisting of all infinite histories with prefix h. We assume that H^∞ is endowed with the σ-algebra $\mathcal{B}(H^\infty)$ generated by the set of cylinders.

Strategies σ, τ of players Max and Min and the initial state i determine a probability measure $\mathbf{P}_i^{\sigma,\tau}$ on $(H^\infty, \mathcal{B}(H^\infty))$.

We define inductively $\mathbf{P}_i^{\sigma,\tau}$ for cylinders in the following way.

Let $h_0 = s_1$ be a finite history of length 0. Then

$$\mathbf{P}_i^{\sigma,\tau}(h_0^+) = \begin{cases} 0 & \text{if } i \neq s_1, \\ 1 & \text{if } i = s_1. \end{cases}$$

Let $h_{t-1} = s_1, (a_1, b_1), \ldots, s_{t-1}, (a_{t-1}, b_{t-1}), s_t$ and $h_t = h_{t-1}, (a_t, b_t), s_{t+1}$. Then

$$\mathbf{P}_i^{\sigma,\tau}(h_t^+) = \mathbf{P}_i^{\sigma,\tau}(h_{t-1}^+) \cdot \sigma(h_{t-1})(a_t) \cdot \tau(h_{t-1})(b_t) \cdot p(s_{t+1}|s_t, a_t, b_t).$$

Note that the set of cylinders is π-system (i.e. a family of sets closed under intersection) thus a probability defined on cylinders extends in a unique way to all sets of $\mathcal{B}(H^\infty)$.

The payoff mapping is a bounded Borel measurable mapping

$$\varphi : H^\infty \to \mathbb{R}.$$

The aim of player Max (player Min) is to maximize (resp. minimize) the expected payoff

$$\mathbf{E}_i^{\sigma,\tau}[\varphi] = \int_{H^\infty} \varphi(h) \mathbf{P}_i^{\sigma,\tau}(dh).$$

The game has value if for each state i there exists a real number v_i, the value of the game for the starting state i, such that

$$\inf_{\tau \in \mathcal{T}} \sup_{\sigma \in \Sigma} \mathbf{E}_i^{\sigma,\tau}[\varphi] = v_i = \sup_{\sigma \in \Sigma} \inf_{\tau \in \mathcal{T}} \mathbf{E}_i^{\sigma,\tau}[\varphi].$$

A strategy τ of player Min is ε-optimal, $\varepsilon \geqslant 0$, if for each state i and each strategy σ of player Max,

$$\sup_{\sigma \in \Sigma} \mathbf{E}_i^{\sigma,\tau}[\varphi] \leqslant v_i + \varepsilon.$$

Symmetrically, a strategy σ of player Max is ε-optimal if for each state i and each strategy τ of player Min,

$$\inf_{\tau \in \mathcal{T}} \mathbf{E}_i^{\sigma,\tau}[\varphi] \geqslant v_i - \varepsilon.$$

An ε-optimal strategy with $\varepsilon = 0$ is called *optimal*.

To define the concurrent stochastic priority game we endow the arena with the reward vector

$$r = (r_1, \ldots, r_n)$$

associating with each state i a reward $r_i \in \mathbb{R}$.

The priority payoff $\varphi_r(h)$ of an infinite history $h = s_1, (a_1, b_1), s_2, (a_2, b_2), s_3, \ldots$ is defined as

$$\varphi_r(h) = r_\ell, \quad \text{where } \ell = \limsup_t s_t. \tag{5.2}$$

Thus the payoff is equal to the reward of the greatest (in the usual integer order) state visited infinitely often.

The aim of player Max (player Min) is to maximize (resp. minimize) the expected priority payoff

$$\mathbf{E}_i^{\sigma,\tau}[\varphi_r] = \int_{H^\infty} \varphi_r(h) \mathbf{P}_i^{\sigma,\tau}(dh).$$

Concurrent priority games contain as special cases some other well known classes of games:

(i) If the reward mapping takes only values in $\{0, 1\}$ then we obtain the usual concurrent parity games [dAM04].

(ii) The second subclass of concurrent priority games is the class of Everett's recursive games [Eve57]. Everett's games are concurrent priority games having reward 0 for all nonabsorbing states [3].

Thus in Everett's games players receive the payoff 0 if the play remains forever in nonabsorbing states, otherwise, for plays ending in an absorbing state i, the payoff is equal to the reward r_i.

(iii) Everett's games contain as a subclass the class of reachability games. Reachability games are Everett's games such that all absorbing states have non-negative rewards [CdAH13, dAHK07].

(iv) The limsup games studied by Maitra and Sudderth [MS96] are the games with the payoff $\limsup_k r_{i_k}$, where $r_{i_1}, r_{i_2}, r_{i_3}, \ldots$ is the infinite sequence of rewards associated with the states visited at the stages $1, 2, 3, \ldots$ during the game. To see that limsup games are priority games it suffices to rename the states in such a way that $i < j$ implies $r_i \leqslant r_j$ for all states $i, j \in [n]$. If this condition is satisfied then the limsup payoff and the priority payoff are equal.

(v) The liminf games are the games with the payoff $\liminf_k r_{i_k}$, where $r_{i_1}, r_{i_2}, r_{i_3}, \ldots$ is the infinite sequence of rewards associated with the states visited at the stages $1, 2, 3, \ldots$ during the game.

Let us rename the states in such a way that, for all states $i, j \in [n]$, $i < j$ implies $r_i \geqslant r_j$. Then the liminf payoff is equal to the priority payoff, thus the liminf games constitute a subclass of priority games.

From the determinacy of Blackwell's games proved by Martin [Mar98] it follows that concurrent priority games have values, i.e. for each state i, $\sup_\sigma \inf_\tau \mathbf{E}_i^{\sigma,\tau}[\varphi_r] =$

3. A state i is absorbing if $p(i|i, a, b) = 1$ for all joint actions (a, b).

$\inf_\tau \sup_\sigma \mathbf{E}_i^{\sigma,\tau}[\varphi_r]$. (The Blackwell games do not have states but the result of Martin extends immediately to games with states as shown by Maitra and Sudderth [MS04].)

A proof of determinacy of concurrent stochastic parity games using fixed points was given by de Alfaro and Majumdar [dAM04]. For Everett's recursive games, Everett proved non only that such games have values but also that both players have ε-optimal memoryless strategies [Eve57]. For concurrent reachability games, player Min has an optimal memoryless strategy while player Max has, for each $\varepsilon > 0$, an ε-optimal memoryless strategy, [CdAH13].

Terminology: As in this chapter we deal only with concurrent stochastic priority games, always when we say a *priority game* it would mean concurrent stochastic priority games.

5.2 Concurrent one-step game

In this section we define an auxiliary one-step game. This simple game constitutes an essential ingredient in our solution to the general priority games.

Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ be a reward vector assigning to each state i the reward x_i.

A concurrent one-step game $\boldsymbol{M}(x)$ is the game played in the following way. If the game starts at a state k then players Max and Min choose independently and simultaneously actions $a \in \mathbf{A}(k)$ and $b \in \mathbf{B}(k)$. Suppose that upon execution of (a, b) the game moves to the next state m. This ends the game and player Max receives from player Min the payoff x_m.

A concurrent one-step game played at state k given the reward mapping x will be denoted $\boldsymbol{M}_k(x)$.

Note that $\boldsymbol{M}_k(x)$ can be seen as a matrix game where

$$\boldsymbol{M}_k(x)[a, b] := \sum_{m \in \mathbf{S}} x_m \cdot p(m|k, a, b)$$

is the (expected) payoff obtained by player Max from player Min when the players play actions a and b respectively.

The *value mapping of the one-step game* is the mapping $f = (f_1, \ldots, f_n)$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ such that, for each state $k \in [n]$,

$$f_k(x_1, \ldots, x_n) := \mathrm{val}(\boldsymbol{M}_k(x)), \tag{5.3}$$

where $\mathrm{val}(\boldsymbol{M}_k(x))$ is the value of the matrix game $\boldsymbol{M}_k(x)$, In other words, $f_k(x_1, \ldots, x_n)$ is the value of the concurrent one-step game played at state k seen as a function of the reward vector $x = (x_1, \ldots, x_n)$.

We will be interested in $f_k(x)$ seen as a function of the reward vector $x = (x_1, \ldots, x_n)$.

Since all entries in the matrix game $\boldsymbol{M}_k(x)$ belong to $\mathbb{R}$, $f_k(x) \in \mathbb{R}$, i.e. f_k is a mapping from $\mathbb{R}^n$ into $\mathbb{R}$.

Lemma 5.1. *The value mapping f of the one-step game defined in* (5.3) *is monotone and nonexpansive.*

Proof. It is easy to see that f is monotone and it is also straightforward that f is additively homogeneous, i.e, for all $x \in \mathbb{R}^n$,

$$f(x + \lambda \cdot e_n) = f(x) + \lambda \cdot e_n,$$

where $e_n = (1, \dots, 1) \in \mathbb{R}^n$ is the vector with 1 on all components. By Lemma 3.2 this implies that f is nonexpansive. □

In the sequel it will be convenient to assume that all rewards belong to the interval $[0, 1]$ rather than to $\mathbb{R}$. This can be achieved without loss of generality by a simple linear transformation, as we did in Section 4.2.

5.3 General concurrent stopping priority games

Concurrent stopping priority games generalize the priority games defined in Section 5.1 by allowing some states to be stopping. In particular if the number of stopping states is zero then we obtain concurrent priority games.

We solve concurrent priority stopping games by induction on the number of non-stopping states and we show that the value function can be expressed as the nearest fixed point of the value function (5.3) of the concurrent one-step game.

By S_t, A^1_t and A^2_t, $t = 1, 2, \dots$, we will denote stochastic processes such that S_t is the state visited at time t, A^1_t is the action executed by player Max at stage t and A^2_t is the action executed at stage t by player Min. i.e. for a play $h = s_1, (a_1, b_1), s_2, (a_2, b_2), s_3, \dots$, $S_t(h) = s_t$, $A^1_t(h) = a_t$ and $A^2_t(h) = b_t$.

For each state $k \in [n]$ we define the random variable

$$T_{>k} : H^\infty \to \mathbb{N} \cup \{\infty\}$$

such that

$$T_{>k} = \min\{t \mid S_t > k\}.$$

Thus $T_{>k}$ is the time of the first visit to a state greater than k.

We define a new stochastic process $S^{[k]}_t, t \in \mathbb{N}$, that we shall call the *stopped state process*:

$$S^{[k]}_t = \begin{cases} S_t & \text{if } T_{>k} \geqslant t, \\ S_q & \text{if } q = T_{>k} < t. \end{cases}$$

Thus if all previously visited states belong to $\{1,\ldots,k\}$ then $S_t^{[k]}$ is equal to the state visited at the current epoch t. However, if at some previous epoch a state $> k$ was visited then $S_t^{[k]}$ is the first such state. In other words, $S_t^{[k]}$ behaves as if the states $> k$ were absorbing, if $S_t^{[k]} > k$ then $S_q^{[k]} = S_t^{[k]}$ for all $q \geqslant t$.

For a given reward vector r and $k \in [n]$ we define the *stopping priority payoff* $\varphi_r^{[k]}$:

$$\varphi_r^{[k]} = r_\ell \quad \text{where } \ell = \limsup_t S_t^{[k]}.$$

The games with payoff $\varphi_r^{[k]}$ will be called *stopping priority games*. We will also speak about the $\varphi_r^{[k]}$-game to refer to the game with payoff $\varphi_r^{[k]}$. Similarly φ_r-game will stand for the usual priority game.

Note that once a state j greater than k is visited the game with payoff $\varphi_r^{[k]}$ is for all practical purposes over, independently of what can happen in the future the payoff is equal to the reward r_j of this state and the states visited after the moment $T_{>k}$ have no bearing on the payoff.

In the $\varphi_r^{[k]}$-game the states $[k]$ will be called *non-stopping* while the states $> k$, will be called *stopping*.

Note that since we have assumed that $\mathbf{S} = [n]$, i.e. n is the greatest state, we have $\varphi_r^{[n]} = \varphi_r$.

Note also that solving games starting in stopping states is trivial. If $i > k$ then for all plays h starting at i, $\varphi_r^{[k]}(h) = r_i$, thus $\mathbf{E}_i^{\sigma,\tau}(\varphi_r^{[k]}) = r_i$ for all strategies σ, τ, in particular the value of stopping state i, $i > k$, is r_i.

5.4 Constructing ε-optimal strategies

The rest of this section is devoted to the proof of the following main result characterizing the values of the stopping concurrent priority games by means of fixed points.

Theorem 5.2. *Let $f : [0,1]^n \to [0,1]^n$ be the value mapping of the concurrent one-step game defined in Section 5.2. For $0 \leqslant k \leqslant n$, let*

$$\mathbf{Fix}^k(f)$$

be the k-th nested fixed point of f, see Section 3.2. Then, for each reward vector r, for each initial state $i \in [n]$, the concurrent stopping priority $\varphi_r^{[k]}$-game starting at i has value $\mathbf{Fix}_i^k(f)(r)$.

Proof. For each $\varepsilon > 0$ we construct ε-optimal strategies for both players.

The proof is carried out by induction on k.

The case $k = 0$ is trivial since when all states are stopping then the value of each state is equal to its reward, i.e. the value of state i is $\mathbf{Fix}_i^0(f)(r) = r_i$.

Under the assumption that the theorem holds for $k-1$, i.e. $\mathbf{Fix}_i^{k-1}(f)(r)$ is the value of the non-stopping state $i \in [k-1]$ in the $\varphi_r^{[k-1]}$-game, we shall prove that $\mathbf{Fix}_i^k(f)(r)$ is the value of the non-stopping state $i \in [k]$ in the $\varphi_r^{[k]}$-game.

We will use the following notation:

$$w_k := \mathbf{Fix}_k^k(f)(r) = \mu_{r_k} x_k . f_k(F_1^{k-1}(x_k; r), \ldots, F_{k-1}^{k-1}(x_k; r), x_k, r_{k+1}, \ldots, r_n) \tag{5.4}$$

and

$$w_i := \mathbf{Fix}_i^k(f)(r) = F_i^{k-1}(w_k; r), \quad i \in [k-1], \tag{5.5}$$

where F_i^{k-1} are defined as in (3.3). Thus our aim is to prove that $(w_1, \ldots, w_{k-1}, w_k)$ are the values of the states $\{1, \ldots, k-1, k\}$ in the $\varphi_r^{[k]}$-game.

Since w_k is a fixed point of (5.4) we have

$$w_k = f_k(w_1, \ldots, w_{k-1}, w_k, r_{k+1}, \ldots, r_n). \tag{5.6}$$

Let T_m be the random time of the $m-$th visit to state k of the stopping state process $(S_t^{[k]})_{t \geqslant 1}$, i.e.

$$\begin{aligned} T_1 &= \min\{t \mid S_t^{[k]} = k\}, \\ T_m &= \min\{t \mid t > T_{m-1} \text{ and } S_t^{[k]} = k\} \quad \text{for } m > 1. \end{aligned} \tag{5.7}$$

Notice that T_m can be infinite if the number of visits of the stopping state process $S_t^{[k]}$ to the state k is smaller than m and $T_1 = 1$ if the game starts at k. Note that since T_m is defined w.r.t. the stopping state process $S_t^{[k]}$, $T_m < \infty$ implies that all states visited prior to the moment T_m are $\leqslant k$.

Let T be any random time, i.e. a mapping from plays to $\{1, 2, \ldots\} \cup \{\infty\}$ such that for each $m \in \{1, 2, \ldots\}$ the event $\{T = m\}$ belongs to the σ-algebra

$$\mathcal{F}_m = \sigma(S_1, (A_1^1, A_1^2), S_2, \ldots, S_m).$$

In other words, $\mathcal{F}_m$ is the σ algebra generated by the cylinders h_m^+, where h_m are histories of length m.

Intuitively that means that knowing the states and actions up to time m we can decide if $T = m$ or not.

Definition 5.3. *For a random time T, $\theta_T : H^\infty \to H^\infty$ will denote the shift mapping that maps plays to plays and is defined in the following way*

$$\theta_T(S_1, (A_1^1, A_1^2), S_2, \ldots) = S_T, (A_T^1, A_T^2), S_{T+1}, (A_{T+1}^1, A_{T+1}^2), S_{T+2}, (A_{T+2}^1, A_{T+2}^2), \ldots.$$

Thus the shift θ_T "forgets" all history prior to time T. Of course, θ_T is well defined only on plays such that $T < \infty$.

Below we use the shift θ_{T_m+1}, where T_m is the time of the mth visit to state k. This shift will be applied only to the plays with $T_m < \infty$.

5.4.1 $\varepsilon/2$-optimal strategy $\sigma_\star$ for player Max when $r_k < w_k$ and k is the starting state.

We assume that

$$r_k < w_k \tag{5.8}$$

and the aim is to construct a strategy $\sigma_\star$ for player Max satisfying

$$\mathbf{E}_k^{\sigma_\star,\tau}(\varphi_r^{[k]}) \geqslant w_k - \varepsilon/2 \tag{5.9}$$

for each strategy τ of Min.

Let

$$\eta \in (w_k - \varepsilon/2, w_k)$$

and define

$$\xi_i = F_i^{k-1}(\eta; r), \quad \forall i \in [k-1]. \tag{5.10}$$

By the induction hypothesis, ξ_i is the value of the $\varphi^{[k-1]}_{(r_1,\ldots,r_{k-1},\eta,r_{k+1},\ldots,r_n)}$-game starting at the state i.

Let us consider the concurrent one-step game $\boldsymbol{M}_k(\xi_1,\ldots,\xi_{k-1},\eta,r_{k+1},\ldots,r_n)$ played at state k. Then

$$\eta_\star := f_k(\xi_1,\ldots,\xi_{k-1},\eta,r_{k+1},\ldots,r_n) \tag{5.11}$$

is the value of this game.

By the properties of monotone nonexpansive mappings, (5.8) implies that w_k is in fact the least fixed point of the mapping

$$x_k \mapsto f_k(F_1^{k-1}(x_k;r),\ldots,F_{k-1}^{k-1}(x_k;r),x_k,r_{k+1},\ldots,r_n).$$

Thus $\eta < w_k$ implies that

$$\eta < f_k(\xi_1,\ldots,\xi_{k-1},\eta,r_{k+1},\ldots,r_n) = \eta_\star \leqslant w_k. \tag{5.12}$$

Fix δ such that

$$0 < \delta < \eta_\star - \eta. \tag{5.13}$$

We define the strategy $\sigma_\star$ of player Max in the following way:

- during the m-th visit to the state k, which takes place at time T_m, c.f. (5.7), player Max selects actions according to his optimal strategy in the concurrent one-step game $\boldsymbol{M}_k(\xi_1,\ldots,\xi_{k-1},\eta,r_{k+1},\ldots,r_n)$.
- during all stages j such that $T_m < j < T_{m+1}$, i.e. between the mth and $(m+1)$th visit to k, player Max plays according to his δ-optimal strategy for the $\varphi^{[k-1]}_{(r_1,\ldots,r_{k-1},\eta,r_{k+1},\ldots,r_n)}$-game.
  When he applies this strategy then we tacitly assume that after each visit to k player Max "forgets" all preceding history and he plays as if the game started afresh at the first state visited after the last visit to k.

From the optimality of $\sigma_\star$ in the concurrent one-step game $\boldsymbol{M}_k(\xi_1,\ldots,\xi_{k-1},\eta,r_{k+1},\ldots,r_n)$, we have

$$\begin{aligned}&\sum_{i<k}\xi_i\cdot\mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1}=i\mid T_m<\infty)\\&+\eta\cdot\mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1}=k\mid T_m<\infty)\\&+\sum_{i>k}r_i\cdot\mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1}=i\mid T_m<\infty)\\&\geqslant\eta_\star.\end{aligned}\tag{5.14}$$

Indeed, when player Max plays according to the strategy $\sigma_\star$ at the moment T_m then the current state is k and he plays using his optimal strategy in the concurrent one-step game $\boldsymbol{M}_k(\xi_1,\ldots,\xi_{k-1},\eta,r_{k+1},\ldots,r_n)$. Now it suffices to notice that the left-hand side of (5.14) is nothing else but the payoff that player Max obtains in the concurrent one-step game $\boldsymbol{M}_k(\xi_1,\ldots,\xi_{k-1},\eta,r_{k+1},\ldots,r_n)$ (because S_{T_m+1} is the state visited at the next time moment T_m+1). Since $\eta_\star$ is the value of this concurrent one-step game the inequality follows.

In the sequel we will note $\mathbb{1}_A$ the indicator of the event A, i.e. the mapping that is equal to 1 on A and to 0 on the complement of A.

Let us note the following equality:

$$\sum_{i>k}r_i\cdot\mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1}=i\mid T_m<\infty)=\mathbf{E}_k^{\sigma_\star,\tau}(\varphi_r^{[k]}\cdot\mathbb{1}_{\{S_{T_m+1}>k\}}\mid T_m<\infty).\tag{5.15}$$

Indeed, if a play belongs to the event $\{S_{T_m+1}=i,T_m<\infty\}$ for $i>k$ then $T_m<\infty$ means that at the moment T_m this play visits k and prior to T_m it never visited states $>k$ cf. (5.7), and at the next time moment T_m+1 such a play visits the stopping state $i>k$. But for such plays the payoff $\varphi_r^{[k]}$ is equal to r_i.

Consider now the event $\{S_{T_m+1}=i,T_m<\infty\}$, for $i<k$, see Figure 5.4.

This event consists of the plays such that

— the stopping state process $S_i^{[k]}$ visits k for the mth time at time T_m (this is guaranteed by $T_m<\infty$, cf.(5.7)) and

— at the next time moment T_m+1 the play visits the state $i<k$.

From the definition of $\sigma_\star$ it follows that starting from the time T_m+1 player Max plays using his δ-optimal strategy in the $\varphi^{[k-1]}_{(r_1,\ldots,r_{k-1},\eta,r_{k+1},\ldots,r_n)}$-game. Since, by the inductive hypothesis (5.10), the value of such a game for state i is ξ_i, we have

$$\mathbf{E}_k^{\sigma_\star,\tau}(\varphi^{[k-1]}_{(r_1,\ldots,r_{k-1},\eta,r_{k+1},\ldots,r_n)}\circ\theta_{T_m+1}\mid S_{T_m+1}=i,T_m<\infty)\geqslant\xi_i-\delta,\quad\text{for all } i<k,\tag{5.16}$$

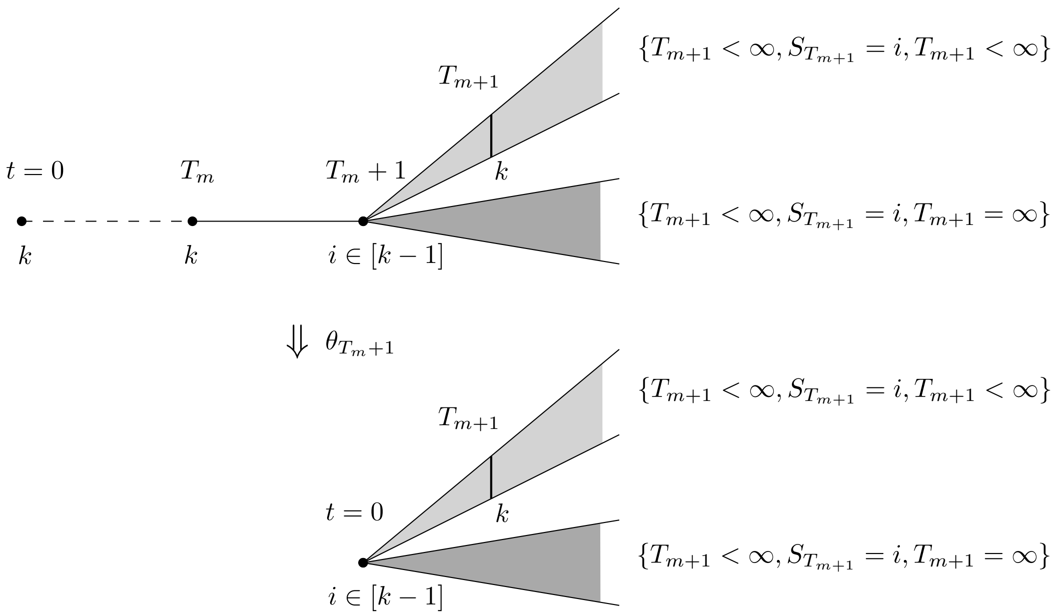

Figure 5.4 – The upper figure: The event $\{S_{T_m+1} = i, T_m < \infty\}$ consists of the plays that at time T_m visit state k for the mth time without ever visiting the states $> k$ before, and at time T_m+1 they visit state i, where $i < k$. These plays are partitioned into two sets. The set $\{T_{m+1} < \infty, S_{T_m+1} = i, T_m < \infty\}$ of plays that will visit k for the $(m+1)$th time and the set $\{T_{m+1} = \infty, S_{T_m+1} = i, T_m < \infty\}$ of the plays for which the mth visit in k was the last one. The lower figure : The shift mapping θ_{T_m+1} "forgets" all the history prior to the time $T_m + 1$.

where θ_{T_m+1} is the shift mapping that deletes all history prior to the time $T_m + 1$.

Using the fact that for all events A and B and each integrable mapping f we have $\mathbf{E}(f \mid A, B) \cdot P(A) = \mathbf{E}(f \cdot \mathbb{1}_{\{A\}} \mid B)$ we can rewrite (5.16) in the following form

$$\mathbf{E}_k^{\sigma_\star,\tau}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},\eta,r_{k+1},\dots,r_n)} \circ \theta_{T_m+1} \cdot \mathbb{1}_{\{S_{T_m+1}=i\}} \mid T_m < \infty) \geqslant \\ (\xi_i - \delta) \cdot \mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1} = i \mid T_m < \infty), \quad \text{for } i < k. \quad (5.17)$$

We shall prove that for $i < k$,

$$\mathbf{E}_k^{\sigma_\star,\tau}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},\eta,r_{k+1},\dots,r_n)} \circ \theta_{T_m+1} \cdot \mathbb{1}_{\{S_{T_m+1}=i\}} \mid T_m < \infty) = \\ \eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty, S_{T_m+1} = i \mid T_m < \infty) + \mathbf{E}_k^{\sigma_\star,\tau}(\varphi_r^{[k]} \cdot \mathbb{1}_{\{T_{m+1}=\infty\}} \cdot \mathbb{1}_{\{S_{T_m+1}=i\}} \mid T_m < \infty). \\ (5.18)$$

Indeed the left-hand side of (5.18) is the sum of

$$\mathbf{E}_k^{\sigma_\star,\tau}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},\eta,r_{k+1},\dots,r_n)} \circ \theta_{T_m+1} \cdot \mathbb{1}_{\{S_{T_m+1}=i\}} \cdot \mathbb{1}_{\{T_{m+1}=\infty\}} \mid T_m < \infty) \qquad (5.19)$$

and

$$\mathbf{E}_k^{\sigma_\star,\tau}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},\eta,r_{k+1},\dots,r_n)} \circ \theta_{T_m+1} \cdot \mathbb{1}_{\{S_{T_m+1}=i\}} \cdot \mathbb{1}_{\{T_{m+1}<\infty\}} \mid T_m < \infty). \tag{5.20}$$

Consider first (5.20). For plays h belonging to the event $\{T_{m+1} < \infty, S_{T_m+1} = i\}, i < k$, the shift θ_{T_m+1} removes all prefix history up to the time $T_m + 1$, see Figure 5.4. Since $T_{m+1} < \infty$ in the remaining suffix play $\theta_{T_m+1}(h)$ all visited states up to the next visit to k are $< k$. But for the plays that visit k at some moment and for which all states prior to this first visit to k are $< k$ the payoff $\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},\eta,r_{k+1},\dots,r_n)}$ is constant and equal to the reward η associated with k. Thus (5.20) is equal to

$$\eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty, S_{T_m+1} = i \mid T_m < \infty).$$

Let us examine now (5.19). The plays h belonging to the event $\{S_{T_m+1} = i, T_{m+1} = \infty, T_m < \infty\}$ have the following properties:

— at time T_m they visit k and all states visited prior to T_m are $\leqslant k$,

— at time $T_m + 1$, just after the mth visit to k, they visit the state i,

— since $T_{m+1} = \infty$ the suffix play $\theta_{T_m+1}(h)$ does not contain any occurrence of k (k is never visited for the $(m+1)$th time).

These properties assure that for such plays $\varphi^{[k]}_r(h) = \varphi^{[k]}_r(\theta_{T_m+1}(h))$. However, $\theta_{T_m+1}(h)$ has no occurrence of k, which implies for the resulting payoff it is irrelevant if k is stopping or not and what is the reward of k. Thus $\varphi^{[k]}_r(\theta_{T_m+1}(h)) = \varphi^{[k-1]}_{(r_1,\dots,r_{k-1},\eta,r_{k+1},\dots,r_n)}(\theta_{T_m+1}(h))$. This terminates the proof that (5.19) is equal to

$$\mathbf{E}_k^{\sigma_\star,\tau}(\varphi^{[k]}_r \cdot \mathbb{1}_{\{T_{m+1}=\infty\}} \cdot \mathbb{1}_{\{S_{T_m+1}=i\}} \mid T_m < \infty).$$

This concludes also the proof of (5.18).

From (5.17) and (5.18) we obtain

$$\begin{aligned}\eta\cdot\mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty, S_{T_m+1} = i \mid T_m < \infty)+\mathbf{E}_k^{\sigma_\star,\tau}(\varphi^{[k]}_r\cdot\mathbb{1}_{\{T_{m+1}=\infty\}}\cdot\mathbb{1}_{\{S_{T_m+1}=i\}} \mid T_m < \infty)\\ \geqslant (\xi_i - \delta)\cdot \mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1} = i \mid T_m < \infty).\end{aligned}$$

Summing both sides of this inequality for $i < k$ and rearranging the terms we obtain

$$\begin{aligned}\sum_{i<k} \xi_i \cdot \mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1} = i \mid T_m < \infty) &\leqslant \eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty, S_{T_m+1} < k \mid T_m < \infty)\\ &\quad + \mathbf{E}_k^{\sigma_\star,\tau}(\varphi^{[k]}_r \cdot \mathbb{1}_{\{T_{m+1}=\infty\}} \cdot \mathbb{1}_{\{S_{T_m+1}<k\}} \mid T_m < \infty)\\ &\quad + \delta \cdot \mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1} < k \mid T_m < \infty)\\ &\leqslant \eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty, S_{T_m+1} < k \mid T_m < \infty)\\ &\quad + \mathbf{E}_k^{\sigma_\star,\tau}(\varphi^{[k]}_r \cdot \mathbb{1}_{\{T_{m+1}=\infty\}} \cdot \mathbb{1}_{\{S_{T_m+1}<k\}} \mid T_m < \infty)\\ &\quad + \delta.\end{aligned}$$

The last inequality, (5.14) and (5.15) yield

$$\begin{aligned}\eta_\star \leqslant \eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}&(T_{m+1} < \infty, S_{T_m+1} < k \mid T_m < \infty) \\ &+\mathbf{E}_k^{\sigma_\star,\tau}(\varphi_r^{[k]} \cdot \mathbb{1}_{\{T_{m+1}=\infty\}} \cdot \mathbb{1}_{\{S_{T_m+1}<k\}} \mid T_m < \infty) \\ &+\delta \\ &+\eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1} = k \mid T_m < \infty) \\ &+\mathbf{E}_k^{\sigma_\star,\tau}(\varphi_r^{[k]} \cdot \mathbb{1}_{\{S_{T_m+1}>k\}} \mid T_m < \infty).\end{aligned} \tag{5.21}$$

Notice that

$$\begin{aligned}\mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty, S_{T_m+1} < k \mid T_m < \infty) + \mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1} = k \mid T_m < \infty) \\ = \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty \mid T_m < \infty)\end{aligned} \tag{5.22}$$

which allows to regroup the first and the fourth summand of right-hand side of (5.21). Indeed, $\{T_{m+1} < \infty, T_m < \infty\}$ is the union of three disjoint events, depending on whether the state visited at the next time moment $T_m + 1$ is $< k$, $= k$, or $> k$. But for the second of these events we have $\{T_{m+1} < \infty, T_m < \infty, S^{[k]}_{T_m+1} = k\} = \{T_m < \infty, S^{[k]}_{T_m+1} = k\}$ since $S^{[k]}_{T_m+1} = k$ implies that $T_{m+1} = T_m + 1 < \infty$.

And finally the third event $\{T_{m+1} < \infty, T_m < \infty, S^{[k]}_{T_m+1} > k\}$ is empty since $S^{[k]}_{T_m+1} > k$ means that at time $T_m + 1$ the game hits a stopping state thus the stopping state process will never return to k, therefore $T_{m+1} = \infty$. This terminates the proof of (5.22).

We can regroup also the second and the last summands of (5.21) since

$$\begin{aligned}\mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} = \infty, S_{T_m+1} < k \mid T_m < \infty) + \mathbf{P}_k^{\sigma_\star,\tau}(S_{T_m+1} > k \mid T_m < \infty) \\ = \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} = \infty \mid T_m < \infty)\end{aligned}$$

We obtain this again by presenting the event $\{T_{m+1} = \infty, T_m < \infty\}$ as the union of three disjoint events depending on the value of S_{T_m+1}. However, $S_{T_m+1} = k$ contradicts $T_{m+1} = \infty$ and $S_{T_m+1} > k$ implies $T_{m+1} = \infty$.

Using these observations we deduce from (5.21) that

$$\begin{aligned}\eta_\star \leqslant \eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}&(T_{m+1} < \infty \mid T_m < \infty) \\ &+ \mathbf{E}_k^{\sigma_\star,\tau}(\varphi_r^{[k]} \cdot \mathbb{1}_{\{T_{m+1}=\infty\}} \mid T_m < \infty) \\ &+ \delta.\end{aligned} \tag{5.23}$$

Since $\varphi_r^{[k]} \leqslant 1$, from (5.23) we obtain that

$$\eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty \mid T_m < \infty) + \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} = \infty \mid T_m < \infty) \geqslant \eta_\star - \delta.$$

But $\mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} = \infty \mid T_m < \infty) + \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty \mid T_m < \infty) = 1$ thus the last inequality yields

$$\mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty \mid T_m < \infty) \leqslant \frac{1+\delta-\eta_\star}{1-\eta} < \frac{1+(\eta_\star-\eta)-\eta_\star}{1-\eta} = 1.$$

Therefore

$$\begin{aligned}
\mathbf{P}_k^{\sigma_\star,\tau}(\forall m, T_m < \infty) &= \lim_{m\to\infty} \mathbf{P}_k^{\sigma_\star,\tau}(\forall i \leqslant m, T_i < \infty) \\
&= \lim_{m\to\infty} \mathbf{P}_k^{\sigma_\star,\tau}(T_0 < \infty) \cdot \prod_{q=0}^{m-1} \mathbf{P}_k^{\sigma_\star,\tau}(T_{q+1} < \infty \mid T_q < \infty) \\
&\leqslant \lim_{m\to\infty} \left(\frac{1-\eta_\star+\delta}{1-\eta}\right)^{m-1} \\
&= 0,
\end{aligned} \tag{5.24}$$

i.e. if player Max uses the strategy $\sigma_\star$ then with probability 1 the state k is visited only finitely many times.

Multiplying both sides of (5.23) by $\mathbf{P}_k^{\sigma_\star,\tau}(T_m < \infty)$, taking into account that $0 < \delta < \eta_\star - \eta$ and rearranging we get

$$\begin{aligned}
\mathbf{E}_k^{\sigma_\star,\tau}(\varphi_r^{[k]} \cdot \mathbb{1}_{\{T_{m+1}=\infty\}} \cdot \mathbb{1}_{\{T_m<\infty\}}) > \eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}(T_m < \infty) \\
- \eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} < \infty, T_m < \infty) \\
= \eta \cdot \mathbf{P}_k^{\sigma_\star,\tau}(T_{m+1} = \infty, T_m < \infty).
\end{aligned} \tag{5.25}$$

Since the events $\{T_{m+1} = \infty, T_m < \infty\}_{m\geqslant 0}$ and $\{\forall m, T_m < \infty\}$ form a partition of the sets of plays but the last event has probability 0, summing up both sides of (5.25) for all $m \geqslant 1$ we obtain

$$\mathbf{E}_k^{\sigma_\star,\tau}(\varphi_r^{[k]}) > \eta > w_k - \frac{\varepsilon}{2}$$

which terminates the proof of (5.9).

5.4.2 $\varepsilon/2$-optimal strategy $\tau_\star$ for player Min when $r_k \leqslant w_k$ and k is the starting state.

We assume that $r_k \leqslant w_k$ and $\varepsilon > 0$. The aim of this section is to construct a strategy $\tau_\star$ for player Min such that

$$\mathbf{E}_k^{\sigma,\tau_\star}(\varphi_r^{[k]}) \leqslant w_k + \varepsilon/2 \tag{5.26}$$

for each strategy σ of Max.

The strategy $\tau_\star$ of player Min is constructed in the following way.

(i) If the current state is k then player Min selects actions with probability given by his optimal strategy in the concurrent one-step game

$$\boldsymbol{M}_k(w_1, \ldots, w_{k-1}, w_k, r_{k+1}, \ldots, r_n).$$

Thus the strategy of player Min at k is "locally memoryless", the probability used to select actions to execute at k does not depend on the previous history.

(ii) During all stages j such that $T_m < j < T_{m+1}$ (between the mth and $(m+1)$th visit to state k) player Min plays using his $\varepsilon_m := \varepsilon/2^{m+1}$-optimal strategy in the $\varphi^{[k-1]}_{(r_1,\ldots,r_{k-1},w_k,r_{k+1},\ldots,r_n)}$-game [4]. In general the strategy played by Min between two visits to state k is not memoryless because ε_m changes at each visit to k. When player Min applies this strategy during all stages j, $T_m < j < T_{m+1}$, in the $\varphi^{[k]}_r$-game then we assume tacitly that starting from stage $T_m + 1$ player Min "forgets" all history preceding this stage and he plays this strategy as if the game started afresh at stage $T_m + 1$.

From the optimality of $\tau_\star$ in the concurrent one-step game $\boldsymbol{M}_k(w_1, \ldots, w_{k-1}, w_k, r_{k+1}, \ldots, r_n)$ we obtain

$$\begin{aligned}
&\sum_{j<k} w_j \cdot \mathbf{P}^{\sigma,\tau_\star}_k(S^{[k]}_{T_m+1} = j | T_m < \infty) \\
&+ w_k \cdot \mathbf{P}^{\sigma,\tau_\star}_k(S^{[k]}_{T_m+1} = k | T_m < \infty) \\
&+ \sum_{j>k} r_j \cdot \mathbf{P}^{\sigma,\tau_\star}_k(S^{[k]}_{T_m+1} = j | T_m < \infty) \\
&\leqslant w_k.
\end{aligned} \tag{5.27}$$

Indeed, at the time T_m the current visited state is k and player Min selects actions according to his optimal strategy in the concurrent one-step game $\boldsymbol{M}_k(w_1, \ldots, w_{k-1}, w_k, r_{k+1}, \ldots, r_n)$ and, by (5.6), the left-hand side of (5.27) gives the payoff in this concurrent one-step game while the right-hand side is the value of this game. Since he plays optimally the payoff cannot be greater than the value.

Let us consider the event

$$\{T_m < \infty, S_{T_m+1} = i\}, \quad \text{where } i < k. \tag{5.28}$$

This event, presented on the upper side of Figure 5.4, consists of plays h satisfying the following conditions:

(i) h visits k at least m times and prior to the m-th visit to k (which takes place at time T_m) the stopping states $\{k+1, \ldots, n\}$ were not visited, i.e. $S_t \in [k]$ for all $t < T_m$,

4. This strategy exists by the induction hypothesis.

(ii) at time T_m the game moves from k to i, i.e. $S_{T_m+1} = i$.

The definition of $\tau_\star$ says that starting from time $T_m + 1$, if the current state S_{T_m+1} is $< k$ and until the next visit to state k, player Min plays according to $\varepsilon/2^{m+1}$-optimal strategy in the $\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)}$-game. By (5.5), the value of the $\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)}$-game starting at state $i \in [k-1]$ is w_i.

Thus if we consider the game that, in some sense, restarts afresh at state i at time $T_m + 1$ and we apply to such residual game the payoff $\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)}$ and we assume that player Min plays $\tau_\star$ then the expected payoff will not be greater than $w_i + \varepsilon/2^{m+1}$, i.e.

$$\mathbf{E}^{\sigma,\tau_\star}_k(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)} \circ \theta_{T_m+1} \mid S_{T_m+1} = i, T_m < \infty) \leqslant w_i + \varepsilon/2^{m+1}. \quad (5.29)$$

where $f \circ g$ denotes the composition of mapping f and g.

Now let us note that (5.27) closely resembles (5.14) while (5.29) resembles (5.16). What is different but symmetric is that the first two formulas concern strategies $(\sigma_\star, \tau)$ and the last two $(\sigma, \tau_\star)$. Moreover, the inequalities are reversed. The following table resumes the correspondence between constants appearing in the formulas:

Eq. (5.14), (5.16)	Eq. (5.27), (5.29)
η	w_k
$\eta_\star$	w_k
ξ_i	w_i
δ	$-\varepsilon_m$

Thus exactly in the same way as we deduced (5.23) from (5.16) and (5.14) we can deduce from (5.27) and (5.29) the following formula analogous to (5.23) (just reverse the inequality and replace the constants as indicated above):

$$\begin{aligned} &w_k \cdot \mathbf{P}^{\sigma,\tau_\star}_k(T_{m+1} < \infty \mid T_m < \infty) \\ &+\mathbf{E}^{\sigma,\tau_\star}_k(\varphi^{[k]}_r \cdot \mathbb{1}_{\{T_{m+1}=\infty\}} \mid T_m < \infty) \\ &-\varepsilon_m \leqslant w_k. \end{aligned}$$

Rearranging the terms and multiplying by $\mathbf{P}^{\sigma,\tau_\star}_k(T_m < \infty)$ we obtain from this inequality that

$$\begin{aligned} \mathbf{E}^{\sigma,\tau_\star}_k(\varphi^{[k]}_r \cdot \mathbb{1}_{\{T_{m+1}=\infty\}} \cdot \mathbb{1}_{\{T_m<\infty\}}) &\leqslant w_k \cdot \mathbf{P}^{\sigma,\tau_\star}_k(T_{m+1} = \infty, T_m < \infty) + \frac{\varepsilon}{2^{m+1}} \cdot \mathbf{P}^{\sigma,\tau_\star}_k(T_m < \infty) \\ &\leqslant w_k \cdot \mathbf{P}^{\sigma,\tau_\star}_k(T_{m+1} = \infty, T_m < \infty) + \frac{\varepsilon}{2^{m+1}}. \end{aligned}$$

The events $\{T_{m+1} = \infty, T_m < \infty\}$ are pairwise disjoint and their union is equal to $\{\exists m, T_m = \infty\}$ thus summing over $m \geqslant 1$ both sides of the inequality we obtain

$$\mathbf{E}^{\sigma,\tau_\star}_k(\varphi^{[k]}_r \cdot \mathbb{1}_{\{\exists m, T_m=\infty\}}) \leqslant w_k \cdot \mathbf{P}^{\sigma,\tau_\star}_k(\exists m, T_m = \infty) + \varepsilon/2.$$

On the other hand, for all plays in $\{\forall m, T_m < \infty\}$ the state k is visited infinitely often thus $\varphi_r^{[k]}$ is equal to r_k.

Thus

$$\begin{aligned}\mathbf{E}_k^{\sigma,\tau_\star}(\varphi_r^{[k]}) &= \mathbf{E}_k^{\sigma,\tau_\star}(\varphi_r^{[k]} \cdot \mathbb{1}_{\{\exists m, T_m=\infty\}}) + \mathbf{E}_k^{\sigma,\tau_\star}(\varphi_r^{[k]} \cdot \mathbb{1}_{\{\forall m, T_m=\infty\}}) \\ &= \mathbf{E}_k^{\sigma,\tau_\star}(\varphi_r^{[k]} \cdot \mathbb{1}_{\{\exists m, T_m=\infty\}}) + r_k \cdot \mathbf{P}_k^{\sigma,\tau_\star}(\forall m, T_m < \infty) \\ &\leqslant w_k \cdot \mathbf{P}_k^{\sigma,\tau_\star}(\exists m, T_m = \infty) + r_k \cdot \mathbf{P}_k^{\sigma,\tau_\star}(\forall m, T_m < \infty) + \varepsilon/2 \\ &\leqslant w_k + \varepsilon/2.\end{aligned}$$

5.4.3 $\varepsilon/2$-optimal strategies for the other cases when the starting state is k

In Sections 5.4.1 and 5.4.2 we have constructed $\varepsilon/2$-optimal strategies for player Max when $w_k > r_k$ and for player Min when $w_k \geqslant r_k$ under the condition that $\mathbf{Fix}^{k-1}(f)(r)$ is the value vector of the $\varphi_r^{[k-1]}$-game.

But passing to the dual game, the last condition implies that $\mathbf{Fix}^{k-1}(\overline{f})(\overline{r})$ is the value vector in the dual stopping game with payoff $\varphi_{\overline{r}}^{[k-1]}$.

Therefore, proceeding exactly as in Section 5.4.1, we can construct a strategy $\tau^\star$ for player $\overline{\text{Max}}$ in the dual game with payoff $\varphi_{\overline{r}}^{[k]}$ such that

$$\mathbf{E}_k^{\tau^\star,\sigma}(\varphi_{\overline{r}}^{[k]}) \geqslant \overline{w}_k - \varepsilon/2 \tag{5.30}$$

for all strategies σ of player $\overline{\text{Min}}$ if

$$\overline{w}_k > \overline{r}_k. \tag{5.31}$$

By duality of games and fixed points, $\mathbf{E}_k^{\tau^\star,\sigma}(\varphi_{\overline{r}}^{[k]}) = 1 - \mathbf{E}_k^{\sigma,\tau^\star}(\varphi_r^{[k]})$, $\overline{w}_k = 1 - w_k$ and $\overline{r}_k = 1 - r_k$. Thus (5.30) is equivalent to $\mathbf{E}_k^{\sigma,\tau^\star}(\varphi_r^{[k]}) \leqslant w_k + \varepsilon/2$ and (5.31) is equivalent to $w_k < r_k$, i.e. we get a $\varepsilon/2$-optimal strategy of player Min in the $\varphi_r^{[k]}$-game if $w_k < r_k$.

In the similar way, applying the construction of Section 5.4.2 to the dual game and coming back to the original game we get a strategy $\sigma^\star$ for player Max such that $\mathbf{E}_k^{\sigma^\star,\tau}(\varphi_r^{[k]}) \geqslant w_k - \varepsilon/2$ if $w_k \leqslant r_k$.

5.4.4 ε-optimal strategies for the $\varphi_r^{[k]}$-game starting at states $< k$.

It remains to prove that

$$\mathbf{Fix}_i^k(f)(r) := F_i^{k-1}(w_k; r)$$

is the value of the $\varphi_r^{[k]}$-game starting in the state $i < k$. To this end we must construct strategies $\sigma_\sharp$ and $\tau_\sharp$ for player Max and Min, respectively, such that

$$\mathbf{E}_i^{\sigma,\tau_\sharp}(\varphi_r^{[k]}) \leqslant \mathbf{Fix}_i^k(f)(r) + \varepsilon \quad \text{and} \quad \mathbf{E}_i^{\sigma_\sharp,\tau}(\varphi_r^{[k]}) \geqslant \mathbf{Fix}_i^k(f)(r) - \varepsilon \tag{5.32}$$

for all strategies σ, τ. We define only the strategy $\tau_\sharp$ for player Min and prove the first equation of (5.32). The definition of $\sigma_\sharp$ and the proof of the right-hand side of (5.32) are symmetrical and are left to the reader.

Recall that T_1 was defined as the (random) time of the first visit of the stopped state process $S_t^{[k]}$ to the state k, cf. (5.7). Let $\tau_\star$ be the strategy of player Min defined at page 81 that satisfies (5.26), i.e $\tau_\star$ is an $\varepsilon/2$-optimal for player Min in the $\varphi_r^{[k]}$-game starting at the state k.

By the induction hypothesis, there exists an $\varepsilon/2$-optimal strategy α for player Min in the $\varphi^{[k-1]}_{(r_1,\ldots,r_{k-1},w_k,r_{k+1},\ldots,r_n)}$-game.

We define the strategy $\tau_\sharp$ for player Min by composing strategies α and $\tau_\star$ as follows:

$$\tau_\sharp(S_1,(A_1^1,A_1^2),\cdots,S_m) = \begin{cases} \alpha(S_1,(A_1^1,A_1^2),\cdots,S_m) & \text{if } T_1 > m, \\ \tau_\star(S_{T_1},(A_{T_1}^1,A_{T_1}^2),\cdots,S_m) & \text{if } T_1 \leqslant m. \end{cases}$$

Intuitively, $\tau_\sharp$ is the strategy such that player Min plays according to α until the first visit to k and starting from the moment of the first visit to k he switches to $\tau_\star$. Moreover, when he switches to $\tau_\star$ then he "forgets" all history prior to the moment T_1 and behaves as if the game have started afresh at k.

First we want to show that, for each strategy σ of player Max and for each state $i < k$,

$$\mathbf{E}_i^{\sigma,\tau_\sharp}(\varphi_r^{[k]} \mid T_1 < \infty) = \mathbf{E}_i^{\sigma,\tau_\sharp}(\varphi_r^{[k]} \circ \theta_{T_1} \mid T_1 < \infty) \leqslant w_k + \varepsilon/2$$

where θ_{T_1} is the shift operation, cf. Definition 5.3, and $w_k = \mathbf{Fix}_k^k(f)(r)$ is the value of k.

To justify the first equality let us notice that the plays with $T_1 < \infty$ do not visit the stopping states, i.e. the states $> k$, prior to T_1. Therefore the payoff $\varphi_r^{[k]}$ for such plays is not modified if we shift them by T_1.

The second inequality follows from the definition of $\tau_\sharp$. When the game hits state k at time T_1 player Min switches to strategy $\tau_\star$ and forgets the history prior to T_1. Since $\tau_\star$ is $\varepsilon/2$-optimal for player Min in the $\varphi_r^{[k]}$-game for plays starting at k, using this strategy limits the payoff to at most $w_k + \varepsilon/2$.

Now we examine the expected payoff for plays with $T_1 = \infty$. Such plays never visit k, therefore it is irrelevant for them if k is stopping or not like it is irrelevant what is the reward associated with k. Moreover, for such plays player Min plays according to strategy $\tau_\star$. For these reasons we have

$$\mathbf{E}_i^{\sigma,\tau_\sharp}(\varphi_r^{[k]} \mid T_1 = \infty) = \mathbf{E}_i^{\sigma,\tau_\star}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)} \mid T_1 = \infty). \tag{5.33}$$

From (5.33) we obtain

$$\begin{aligned}\mathbf{E}_i^{\sigma,\tau_\sharp}(\varphi_r^{[k]}) &= \mathbf{E}_i^{\sigma,\tau_\sharp}(\varphi_r^{[k]} \mid T_1 < \infty)\cdot \mathbf{P}_i^{\sigma,\tau_\sharp}(T_1 < \infty)\\ &\quad + \mathbf{E}_i^{\sigma,\tau_\sharp}(\varphi_r^{[k]} \mid T_1 = \infty)\cdot \mathbf{P}_i^{\sigma,\tau_\sharp}(T_1 = \infty)\\ &\leqslant (w_k + \varepsilon/2)\cdot \mathbf{P}_i^{\sigma,\tau_\sharp}(T_1 < \infty)\\ &\quad + \mathbf{E}_i^{\sigma,\tau_\star}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)} \mid T_1 = \infty)\cdot \mathbf{P}_i^{\sigma,\tau_\sharp}(T_1 = \infty).\end{aligned} \tag{5.34}$$

Since $\tau_\star$ is $\varepsilon/2$-optimal for player Min in the $\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)}$-game we have

$$\begin{aligned}F_i^{k-1}(w_k;r) + \varepsilon/2 &\geqslant \mathbf{E}_i^{\sigma,\tau_\star}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)})\\ &= \mathbf{E}_i^{\sigma,\tau_\star}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)} \mid T_1 < \infty)\cdot \mathbf{P}_i^{\sigma,\tau_\star}(T_1 < \infty)\\ &\quad + \mathbf{E}_i^{\sigma,\tau_\star}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)} \mid T_1 = \infty)\cdot \mathbf{P}_i^{\sigma,\tau_\star}(T_1 = \infty).\end{aligned}$$

Notice that plays with $T_1 < \infty$ have payoff w_k in the $\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)}$-game because k is stopping in this game and the reward of k is equal to w_k. Hence we can rewrite (5.35) as

$$\begin{aligned}F_i^{k-1}(w_k;r) + \varepsilon/2 &\geqslant w_k\cdot \mathbf{P}_i^{\sigma,\tau_\star}(T_1 < \infty)\\ &\quad + \mathbf{E}_i^{\sigma,\tau_\star}(\varphi^{k-1}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)} \mid T_1 = \infty)\cdot \mathbf{P}_i^{\sigma,\tau_\star}(T_1 = \infty).\end{aligned}$$

Thus

$$\begin{aligned}\mathbf{E}_i^{\sigma,\tau_\star}(\varphi^{[k-1]}_{(r_1,\dots,r_{k-1},w_k,r_{k+1},\dots,r_n)} &\mid T_1 = \infty)\cdot \mathbf{P}_i^{\sigma,\tau_\star}(T_1 = \infty)\\ &\leqslant F_i^{k-1}(w_k;r) + \varepsilon/2 - w_k\cdot \mathbf{P}_i^{\sigma,\tau_\star}(T_1 < \infty).\end{aligned} \tag{5.35}$$

From (5.34) and (5.35) and since $\mathbf{P}_i^{\sigma,\tau_\sharp}(T_1 < \infty) = \mathbf{P}_i^{\sigma,\tau_\star}(T_1 < \infty)$ we get

$$\begin{aligned}\mathbf{E}_i^{\sigma,\tau_\sharp}(\varphi_r^{[k]}) &\leqslant (w_k + \varepsilon/2)\cdot \mathbf{P}_i^{\sigma,\tau_\sharp}(T_1 < \infty) + F_i^{k-1}(w_k;r) + \varepsilon/2 - w_k\cdot \mathbf{P}_i^{\sigma,\tau_\star}(T_1 < \infty)\\ &= F_i^{k-1}(w_k;r) + \varepsilon/2 + (\varepsilon/2)\cdot \mathbf{P}_i^{\sigma,\tau_\sharp}(T_1 < \infty)\\ &\leqslant F_i^{k-1}(w_k;r) + \varepsilon\\ &= \mathbf{Fix}_i^k(f)(r) + \varepsilon\end{aligned}$$

which terminates the proof of the ε-optimality of $\tau_\sharp$.

5.4.5 Dual game

We have constructed a ε-optimal strategy for Max and Min for the game starting at k but the strategy for Max was constructed under the condition $r_k < w_k$ while the strategy for Min was constructed under the condition $r_k \leqslant w_k$.

How to obtain ε-optimal strategies for both players for two remaining cases ($r_k \geqslant w_k$ for Max and $r_k > w_k$ for Min) we use the natural duality of the nested fixed points and the games.

Let G be a priority game. The dual game $\overline{G}$ is obtained in the following way:

(Di) $\overline{G}$ has the same states, actions and transition probabilities as G,

(Dii) if $r = (r_1, \ldots, r_n)$ is the reward vector in G then $\overline{r} = (\overline{r}_1, \ldots, \overline{r}_n)$ is the reward vector in $\overline{G}$, where for $z \in [0, 1]$, $\overline{z} := 1 - z$,

(Diii) players Max and Min exchange the roles, in the dual game for each state $i \in \mathbf{S}$, $\mathbf{A}(i)$ are the actions of player Max while $\mathbf{B}(i)$ are the actions of player Min, moreover in the dual game player Max wants to minimize the priority payoff $\varphi_{\overline{r}}$ while Min wants to maximize the priority payoff $\varphi_{\overline{r}}$.

To avoid confusion, we write $\overline{\text{Max}}$ and $\overline{\text{Min}}$ to denote the players, respectively, maximizing and minimizing the priority payoff in the dual game.

A strategy σ is a strategy of player Max in G if and only if it is a strategy of player $\overline{\text{Min}}$ in the dual game $\overline{G}$. A symmetric property holds for strategies of player Min.

For each play h we have $\varphi_r(h) = 1 - \varphi_{\overline{r}}(h)$, thus $\mathbf{E}_i^{\sigma,\tau}(\varphi_r) = 1 - \mathbf{E}_i^{\tau,\sigma}(\varphi_{\overline{r}})$, where the left hand side is the expected payoff in G, while $\mathbf{E}_i^{\tau,\sigma}(\varphi_{\overline{r}})$ is the expected payoff in $\overline{G}$ when $\overline{\text{Max}}$ plays according to τ and $\overline{\text{Min}}$ plays according to σ.

This implies that $v_i = 1 - \overline{v}_i$, where v_i is the value of state i in G while $\overline{v}_i$ is the value of i in the $\overline{G}$. Moreover, a strategy is ε-optimal for player Max in G if and only if it is ε-optimal for player $\overline{\text{Min}}$ in $\overline{G}$. A symmetric property holds for strategies of player Min.

□

Chapter 6

Discussion and conclusions

In Chapter 4 we proved that in turn-based stochastic priority games both players have pure memoryless optimal strategies. Since the number of states and actions are finite, the number of possible pure memoryless strategies is also finite. Therefore, comparing game values obtained for all pairs of pure memoryless strategies (σ,τ), we can find pure memoryless optimal strategies. This method is highly inefficient.

The question whether there exists a more efficient way to find these pure memoryless optimal strategies for both players is open.

Concerning concurrrent priority games, in the future we hope to use the approach developed in Chapter 5 to find non-trivial classes of concurrent priority games where one or both players have ε-optimal memoryless strategies. In this direction let us mention the result of Secchi [Sec98] who proved that in concurrent limsup games [1] player Min has an ε-optimal memoryless strategy.

Another interesting problem is to find a method allowing to approximate the values of the concurrent priority games with a given accuracy.

1. A limsup game is a game with payoff equal to $\limsup_k r_{s_k}$, where $(s_k)_{k=1}^{\infty}$ are the visited states during the play.

Part II

Population questions

Chapter 7

Analysing population dynamics of Markov chains

In this chapter we analyse the simplest framework among discrete time stochastic finite state system: Markov chains. Contrary to what we did in the first part of the thesis, here we use another interpretation. Namely, the *population semantics*: it explains how a distribution over the states is transformed at each step. Let us consider the following example: let M be the transition stochastic matrix defined in (7.1). We can draw the Markov chain as showed in Figure 7.1: each arrow shows the probability to move from each state to another one. Assume that initially $1/2$ of the population are in state 2 and the other half is in state 3. Then if we want to know the proportion of the population in each state in the next step it suffices to multiply the matrix M by vector $(0, 1/2, 1/2)$ and we obtain $(0.45, 0.3, 0.25)$.

$$\begin{pmatrix} 0.1 & 0.7 & 0.2 \\ 0.4 & 0 & 0.6 \\ 0.5 & 0.3 & 0.2 \end{pmatrix} \tag{7.1}$$

With such semantics, properties considered are different than reachability, parity, etc. Instead, we want to know whether there exists a step at which the proportion of the distribution in a set *Goal* of states is higher than some threshold γ (population question). This is orthogonal to the question of bringing with high probability a pebble in a set of state, where the number of steps to bring the pebble is non uniform over all the runs (PCTL question) [BRS02] . The population question is much harder to verify than the PCTL question: it is actually not known whether this kind of question can be decided on Markov Chains ([AAOW15], as will be discussed in Section 7.1.2). In this chapter we approach this problem by studying the languages generated by Markov chains, whose regularity would entail the decidability of this question.

More precisely, in this chapter we study classes for which the language of tra-

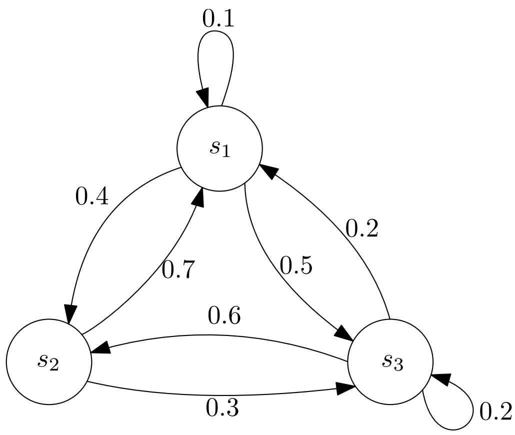

Figure 7.1 – Markov chain with three states.

jectories is (ω-)regular, allowing for the exact resolution of any regular question (e. g. checking any linear temporal logic with intervals ($LTL_{\mathcal{I}}$) formula as defined in [AAGT15], it means, a linear temporal logic in which an atomic proposition will assert that "the current probability of the node i lies in the interval d"). More precisely, we define the trajectory from a given initial distribution as an (infinite) word over the alphabet $\{A, B\}$. The $n-$th letter of a trajectory being A (for Above, respectively, B for Below) represents that after n steps the probability to be in *Goal* is greater than or equal to (respectively lesser than) the threshold γ. Further, we consider the language of MC as the set of trajectories (words) ranging over a (possibly infinite) set of initial distributions. Thus, we can answer questions such as: does there exist a trajectory from the set of initial distributions satisfying a regular property or do all trajectories satisfy it. We prove that the language of a MC with distinct real positive eigenvalues is regular.

7.1 Preliminaries and definitions

A distribution δ over Q is a function $\delta : Q \to [0,1]$ such that $\sum_{q \in Q} \delta(q) = 1$. Given $M \in |Q| \times |Q|$, the matrix associated with a MC, we denote by $M\delta$ the distribution given by $M\delta(q) = \sum_{q' \in Q} \delta(q')M(q', q)$ for all $q \in Q$. Notice that, considering δ and $M\delta$ as row-vectors, this corresponds to performing the matrix multiplication. That is, we consider M as a transformer of probabilities, as in [KVAK10, AAGT15]: $(M\delta)(q)$ represents exactly the probability to be in q after applying M once, knowing that the initial distribution is δ. Inductively, $(M^n\delta)(q)$ represents the probability to be in q after applying n times M, knowing that the initial distribution is δ.

For example, let $(\mathbf{S}, M)$ be the transition matrix of the Markov chain presented in Figure 7.1 and (7.1) with initial distribution given by

$$\delta_0 = \begin{pmatrix} 0 \\ 0.5 \\ 0.5 \end{pmatrix}.$$

Hence the initial probability to be in state 2 is $\frac{1}{2}$ and $\frac{1}{2}$ for state 3. The distribution at the next step is given by

$$M\delta = \begin{pmatrix} 0.45 \\ 0.3 \\ 0.25 \end{pmatrix}.$$

7.1.1 Motivation

As motivation, consider a population of yeast under osmotic stress [MTC+14]. The stress level of the population can be studied through a protein which can be marked (by a chemical reagent). For the sake of illustration, consider the following simplistic model of a Markov Chain M_{yeast} with the protein being in 3 different discrete states (namely the concentration of the protein being high (state 1), medium (state 2) and low (state 3)). The transition matrix, also denoted M_{yeast}, gives the proportion of yeast moving from one protein concentration level to another one, in one time step (say, 15 seconds).

$$M_{yeast} = \begin{pmatrix} 0.8 & 0.1 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.7 \end{pmatrix}$$

For instance, 20% of the yeast with high protein concentration will have low protein concentration at the next time step. The marker can be observed optically when the concentration of the protein is high. We know that the original proportion of yeast in state 1 is 1/3 (by counting the marked yeast population), but we are unsure of the mix between low and medium. The initial set of distributions is thus $\text{Init}_{yeast} = \{(1/3, x, 2/3 - x) \mid 0 \leqslant x \leqslant 2/3\}$. The language of M_{yeast} will tell us how the population evolves wrt the number of marked yeast being above or below the threshold $\gamma_{yeast} = 5/12$, depending on the initial distribution in Init_{yeast}. Now, suppose an experiment with yeasts reveals that there are at first less than 5/12 of *marked* yeast (i.e. with high concentration of proteins), then more than 5/12 of marked yeast, and eventually less than 5/12 of marked yeasts. That is, the trajectory is B for a while, then A for a while, then it stabilises at B, in other words, the trajectory is $B^n A^m B^\omega$ for some $n, m \geqslant 0$. Let us call this property as (P_{yeast}) (note that this is a regular property). We are interested in checking whether our simplistic model exhibits at least one trajectory with the property (P_{yeast}), and if yes, the range of initial values generating trajectories with this property.

Our method computes effectively the language of M_{yeast}, as M_{yeast} has positive real eigenvalues, answering the question whether there exists an initial trajectory s.t. property (P_{yeast}) holds.

7.1.2 Relation with the Skolem problem

Skolem problem can be formulated as follows: for an integer matrix M, does there exist n such that $M^n[s,t] = 0$? where $M^n[s,t] = e_s M^n e_t$ and e_i is a vector whose components consist of a one in the i-th position and 0 otherwise. On the other hand, the Markov reachability problem can be formulate as: given a stochastic matrix M with rational entries and a rational number r, does there exist n such that $M^n[s,t] = r$? Hence, the Markov reachability problem is a sub-case of Skolem, for the particular case where matrices are Markov chains. In [AAOW15] it is proved that Markov reachability problem is at least as hard as the Skolem problem, in particular, they show that the Skolem problem can be reduced to the Markov reachability problem in polynomial time.

We define three basic problems which have been studied extensively in different contexts. Given an initial distribution δ_0 and a MC $\mathcal{A}$ with Matrix M, target states *Goal* and threshold γ:

Existence problem: does there exist $n \in \mathbb{N}$ such that the probability to be in *Goal* after n iterations of M from δ_0 is γ (i.e., $\sum_{q \in Goal}(M^n \delta_0)(q) = \gamma$)?

Positivity problem: does there exist $n \in \mathbb{N}$ such that the probability to be in *Goal* after n iterations of M from δ_0 is at least γ (i.e., $\sum_{q \in Goal}(M^n \delta_0)(q) \geqslant \gamma$)?

Ultimate Positivity problem: does there exist $n \in \mathbb{N}$ s.t., for all $m \geqslant n$, the probability to be in *Goal* after m iterations of M from δ_0 is at least γ (i.e., $\sum_{q \in Goal}(M^m \delta_0)(q) \geqslant \gamma$)?

Note that all these problems are defined from a fix initial distribution δ_0. These problems for MCs are specific instances of problems over general recurrence sequences, that have been extensively studied [OW12, HHH06]. It turns out that the existence for the special MC case is as hard as the existence (Skolem) problem over general recurrence sequences as shown in [AAOW15].

Theorem 7.2. *[AAOW15, HHH06] For general MCs, the existence and positivity are as hard as the Skolem's problem.*

The positivity result comes from the interreducibility of Skolem's problem and the positivity problem for general recurrence sequences [HHH06]. The decidability of Skolem has been open for 40 years, and it has been shown that solving positivity, ultimate positivity or existence for general MCs even for a small number of states

(<50, depending on the problem considered) would entail major breakthroughs in diophantine approximations [OW14b].

7.1.3 Simple MCs

In order to obtain decidability, we will consider restrictions over the matrix M associated with the MC. The first restriction, fairly standard, is that M has distinct eigenvalues (they can be complex numbers too), which makes M diagonalizable.

Definition 7.3. *A stochastic matrix is* simple *if all its eigenvalues are distinct. A MC is* simple *if its associated transition matrix is.*

Some decidability results [OW14c, OW14a] have been proved in the case of distinct eigenvalues for variants of the Skolem, which implies the following for *simple* MCs:

Theorem 7.4. *For simple MCs, ultimate positivity is decidable [OW14c].*
For simple MCs with at most 9 states, positivity is decidable [OW14a].

We will consider the *simple* MC restriction. Notice that the decidability restrictions in Theorem 7.4 for these two closely related problems have led to two different papers [OW14a],[OW14c] in the same conference, using different techniques. As we want to answer in a uniform way any regular question (subsuming among others the above three problems and regular properties such as (P_{yeast})) for MCs of all sizes, we will later impose more restrictions. We start with the simple well-known observation that a simple MC has a unique stationary distribution.

Lemma 7.5. *Let M be a simple stochastic matrix. Then there exists a unique distribution δ_{stat} such that $M\delta_{stat} = \delta_{stat}$.*

Proof. We give a sketch of proof here. We will later get an analytical explanation of this result. We have $M\delta = \delta$ iff $(M - Id)\delta = 0$. As M is diagonalizable and 1 is a eigenvalue of M of multiplicity 1, we have $Ker(M - Id)$ is of dimension 1. The intersection of distributions and of $Ker(M - Id)$ is of dimension 0, that is, it is a single point. □

As usual with MCs, we consider the probability to be in the set of states *Goal* after n steps, that is $\sum_{q \in Goal}(M^n\delta)(q)$. We consider only one threshold γ, for simplicity. In fact, the case of multiple thresholds reduces to this case, since the behaviour is non-trivial for only one threshold, namely $\gamma_{stat} = \sum_{q \in Goal} \delta_{stat}(q)$, as Lemma 7.14 shows. Before to prove this Lemma we need some definitions.

7.1.4 Trajectories and ultimate periodicity

We want to know whether the n^{th} distribution $M^n\delta$ of the trajectory starting in distribution $\delta \in Init$ is above the hyperplane defined by $\sum_{q\in Goal} x_q = \gamma$, i.e., whether $\sum_{q\in Goal}[M^n\delta](q) \geqslant \gamma$. We will write $\rho_\delta(n) = A$ (Above) for $\sum_{q\in Goal}[M^n\delta](q) \geqslant \gamma$, and $\rho_\delta(n) = B$ (Below) else.

Definition 7.6. *The* trajectory $\rho_\delta = \rho_0\rho_1 \cdots \in \{A, B\}^\omega$ *from a distribution* δ *is the infinite word with* $\rho_n = \rho_\delta(n)$ *for all* $n \in \mathbb{N}$.

We write the eigenvalues of M as $p_0, \ldots, p_k$ with $||p_i|| \geqslant ||p_j||$ for all $i < j$. Notice that $k + 1 = |Q|$ the number of states (as the MC is simple). It is a standard result that all eigenvalues of Markov chains have modulus at most 1, and at least one eigenvalue is 1. We fix $p_0 = 1$. As shown in the next Lemma 7.7, we have, for some $a_i(\delta) \in \mathbb{C}$:

$$\rho_\delta(n) = A \text{ iff } \sum_{i=0}^{k} a_i(\delta)p_i^n \geqslant \gamma. \tag{7.2}$$

Lemma 7.7. *Given a matrix* M *with distinct eigenvalues* $(p_0, p_1, \ldots, p_k)$*, we have* $\rho_\delta(n) = A$ *iff* $\sum_{i=0}^{k} a_i(\delta)p_i^n \geqslant \gamma$ *for some constants* $a_i(\delta)_{i\leqslant k}$ *independent of* n.

Proof. As the eigenvalues are distinct the eigenvectors $(v_i)_{i\leqslant k}$ form a basis. Let $\delta = \alpha_i v_i$. By definition $\rho_\delta(n) = A$ iff $\sum_{q\in Goal} [M^n\delta](q) \geqslant \gamma$, then

$$\begin{aligned}
\gamma &\leqslant \sum_{q\in Goal} [M^n\delta](q) \\
&= \sum_{q\in Goal} \left(\sum_{i=0}^{k} \alpha_i M^n v_i\right) e_q \\
&= \sum_{q\in Goal} \left(\sum_{i=0}^{k} \alpha_i v_i p_i^n\right) e_q \\
&= \sum_{i=0}^{k} p_i^n \sum_{q\in Goal} \alpha_i v_i e_q,
\end{aligned}$$

with $e_q = (0, \ldots, 0, 1, 0, \ldots, 0)^t$ where 1 is it the q-th position. Now fixing

$$a_i(\delta) = \sum_{q\in Goal} \alpha_i v_i e_q, \tag{7.3}$$

we have $\rho_\delta(n) = A$ iff $\sum_{i=0}^{k} a_i(\delta) p_i^n \geqslant \gamma$.

□

In the following, we denote $u_\delta(n) = \sum_{i=0}^{k} a_i(\delta) p_i^n$ for all $n \in \mathbb{N}$, where $a_i(\delta)$ is defined in (7.3). If ρ_δ is (effectively) ultimately periodic (i.e, of the form uv^ω), every (omega) regular property, such as existence, positivity and ultimate positivity is decidable (and are in fact easy to check). Unfortunately, this is not always the case, even for small simple MCs.

Theorem 7.8. *[AAGT15] There exists an initial distribution δ_0 and simple MC $\mathcal{A}$ with 3 states, and coefficients and threshold in $\mathbb{Q}$, such that ρ_{δ_0} is not ultimately periodic.*

Proof Sketch. The MC is given by: *Goal* $= \{1\}$ is the first state, $\gamma = \frac{1}{3}$ and the associated matrix M_0 and initial distribution δ_0 are:

$$M_0 = \begin{pmatrix} 0.6 & 0.1 & 0.3 \\ 0.3 & 0.6 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{pmatrix} \text{ and } \delta_0 = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{2} \end{pmatrix}$$

The reason the trajectory is not ultimately periodic follows from the fact that the eigenvalues of M_0 are 1, $r_0 e^{i\theta_0}$ and $r_0 e^{-i\theta_0}$ with $r_0 = \sqrt{19}/10$ and $\theta_0 = cos^{-1}(4/\sqrt{19})$.

Figure 7.9 depicts the probability to be in state 1 (the solid line) and ρ_{δ_0} (the circles).

□

An easy way to obtain ultimately periodic trajectories is to restrict to eigenvalues v which are roots of real numbers, that is, there exists $n \in \mathbb{N}\backslash\{0\}$ with $v^n \in \mathbb{R}$.

Proposition 7.10. *Let $\mathcal{A}$ be a simple MC with eigenvalues $(p_i)_{i \leqslant m}$ all roots of real numbers. Then ρ_δ is ultimately periodic for all distributions δ. The (ultimate) period of ρ_δ can be chosen as any $m \in \mathbb{N}\backslash\{0\}$ such that p_i^m is a positive real number for all $i \leqslant m$.*

Proof. Let $m \in \mathbb{N}\backslash\{0\}$ such that $r_i = p_i^m$ is a positive real number for all i. Such an m exists. Indeed, let $n_i \in \mathbb{N}\backslash\{0\}$ such that $p_i^{n_i} \in \mathbb{R}$. Let ℓ be the lcm of $(n_i)_{i \leqslant k}$ and $m = 2\ell$. Hence every $r_i = p_i^m$ is a positive real number for all $i \leqslant k$.

Let δ a distribution. Taking (7.2), let $\rho(n) = \rho_\delta(mn)$ for all $n \in \mathbb{N}$. We have $\rho(n) = A$ iff $\sum_{i=0}^{k} a_i(\delta) r_i^n \geqslant 0$. We have $a_i(\delta) \in \mathbb{R}$ for all i.

For all $r \in \{r_i \mid i \leqslant k\}$, we denote I_r the set of indices i with $r = r_i$ (it is possible that several eigenvalues p_j are the roots of the same positive real r_i), and $a_r = \sum_{i \in I_r} a_i(\delta)$. Let r be the largest value in $\{r_i \mid i \leqslant k\}$ such that $a_r \neq 0$. Notice that if for all r, $a_r = 0$, then the trajectory is constant, equal to A^ω.

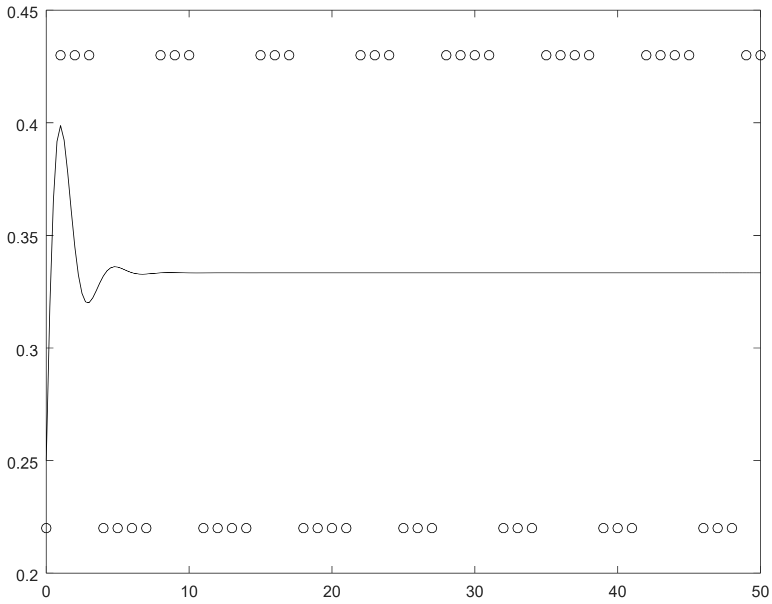

Figure 7.9 – The solid line represents $\sum_{i=0}^{k} a_i(\delta)p_i^n$ and the circles are above the graph if $\rho_{\delta_0}(n) = A$ and below if $\rho_{\delta_0}(n) = B$.

Obviously, $\rho(n)$ is asymptotically equivalent to $a_r r^n$ when n tends to infinity. That is, there exists N_δ such that for all $n \geqslant N_\delta$, $\rho(n)$ is of the sign of a_r. Now, consider initial distributions δ' in the finite set $\Delta = \{M^0\delta, \dots, M^{m-1}\delta\}$. Let N be the max over $N_{\delta'}$ for $\delta' \in \Delta$. We have that $\rho_\delta(mn+\ell) = \rho_{M^\ell\delta}(mn)$ for all $\ell \in \{0, \dots, m-1\}$. Let $u = \rho_\delta(0)\cdots\rho_\delta(mN-1)$ and $v = \rho_\delta(mN)\cdots\rho_\delta(m(N+1)-1)$. We have that $\rho_\delta = uv^\omega$, proving that ρ is ultimately periodic of (ultimate) period m. □

Now, for a finite state (Büchi) automaton $\mathcal{B}$ over the alphabet $\{A, B\}$, the membership problem, of whether a given single trajectory $\rho_\delta \in \mathcal{L}(\mathcal{B})$, is decidable. It is easy to obtain a (small) automaton $\mathcal{B}$ for each of the existence, positivity and ultimate positivity problem such that this problem is true iff $\rho_\delta \in \mathcal{L}(\mathcal{B})$. For instance, let us build a non-deterministic Büchi automaton for the ultimate positivity problem, let $\mathcal{B}$ be an automaton with two states $\{q_1, q_2\}$, acceptance condition $F = \{q_2\}$, initial state q_1 and non-deterministic transitions as depicted in Figure 7.11. It is easy to see that this automaton accepts words in which B occurs only finitely many times. We thus obtain the following proposition:

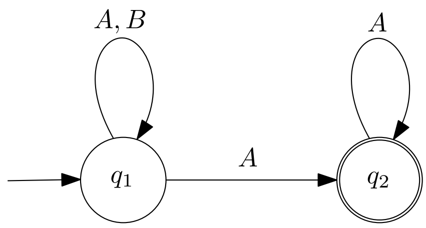

Figure 7.11 – Büchi automaton for the ultimate positivity problem.

Proposition 7.12. *Let $\mathcal{A}$ be a simple MC with eigenvalues all roots of real numbers. Let δ_0 be a distribution. Then the existence, positivity and ultimate positivity problems from initial distribution δ_0 are decidable.*

Proof. Let $\mathcal{A}$ be a simple MC with eigenvalues all roots of real numbers and let δ_0 be the initial distribution. Let ρ_δ be the symbolic trajectory defined in (7.2) and let $\mathcal{B}$ be the (Büchi) automaton such that positivity (or ultimate positivity) problem is true iff $\rho_\delta \in \mathcal{L}(\mathcal{B})$. As the membership problem is decidable, hence it suffices to decide if $\rho_\delta \in \mathcal{L}(\mathcal{B})$ to decide if positivity (or ultimate positivity) is true.

For the existence problem we have to modify the definition of ρ_δ, switching the inequality to an equality, i.e., $\rho_\delta(n) = A$ iff $\sum_{i=0}^{k} a_i(\delta)p_i^n = \gamma$, and to apply the same method of proof. □

Note that Propositions 7.10 and 7.12 hold even when the matrix associated with the MC is diagonalizable, but not necessarily simple.

7.2 Language of a MC

Using automata-based methods allows us to consider more complex problems, where the initial distribution is not fixed. We define the set Init of initial distributions as a convex polytope, that is the convex hull of a finite number of distributions.

Definition 7.13. *The language of a MC $\mathcal{A}$ wrt. the set of initial distributions* Init *is* $\mathcal{L}(\mathrm{Init}, \mathcal{A}) = \{\rho_\delta \mid \delta \in \mathrm{Init}\} \subseteq \{A, B\}^\omega$.

Note that A and B, and the language, depend on the threshold γ. As we assumed this threshold value to be fixed, the language only depends on $\mathcal{A}$ and Init. As $\mathcal{A}$ is often clear from the context, we will often write $\mathcal{L}(\mathrm{Init})$ instead of $\mathcal{L}(\mathrm{Init}, \mathcal{A})$. For the yeast example $M = M_{yeast}$, we have eigenvalues 1; 0.7; 0.6:

$$M \cdot \begin{pmatrix} 5/12 \\ 1/3 \\ 1/4 \end{pmatrix} = 1 \begin{pmatrix} 5/12 \\ 1/3 \\ 1/4 \end{pmatrix}; \quad M \cdot \begin{pmatrix} 5/12 \\ -5/12 \\ 0 \end{pmatrix} = 0.7 \begin{pmatrix} 5/12 \\ -5/12 \\ 0 \end{pmatrix}; \quad M \cdot \begin{pmatrix} 5/12 \\ 0 \\ -5/12 \end{pmatrix} = 0.6 \begin{pmatrix} 5/12 \\ 0 \\ -5/12 \end{pmatrix}$$

We can decompose two initial distributions $\delta_1, \delta_2 \in \text{Init}_{yeast}$ on the eigenvector basis:

$$\begin{pmatrix}1/3\\1/4\\5/12\end{pmatrix} = \begin{pmatrix}5/12\\1/3\\1/4\end{pmatrix} + \frac{1}{5}\begin{pmatrix}5/12\\-5/12\\0\end{pmatrix} - \frac{2}{5}\begin{pmatrix}5/12\\0\\-5/12\end{pmatrix}; \quad \begin{pmatrix}1/3\\1/3\\1/3\end{pmatrix} = \begin{pmatrix}5/12\\1/3\\1/4\end{pmatrix} - \frac{1}{5}\begin{pmatrix}5/12\\0\\-5/12\end{pmatrix}$$

Projecting on the first component, we have $\rho_{\delta_1}(n) = A$ iff $\frac{1}{12}0.7^n - \frac{1}{6}0.6^n \geqslant 0$, that is $\rho_{\delta_1} = B^4A^\omega$. Also, $\rho_{\delta_2}(n) = A$ iff $-\frac{1}{12}0.6^n \geqslant 0$, that is $\rho_{\delta_2} = B^\omega$. With the techniques developed in the following, we can prove more generally that, for all $n \in \mathbb{N}$, we can find an ϵ s.t., $\delta = (1/3 \quad 1/3 - \epsilon \quad 1/3 + \epsilon)^T$ has trajectory $\rho_\delta = B^nA^\omega$, and that $\mathcal{L}(\text{Init}_{yeast}) = B^*A^\omega \cup B^\omega$. Thus, property (P_{yeast}), from Introduction, does not hold for any initial distribution.

Now that we introduce the notions of language, we can prove the Lemma that we mentioned above.

Lemma 7.14. *For $\gamma \neq \gamma_{stat}$, we have $\mathcal{L}(\text{Init}, \mathcal{A})$ is regular.*

Proof. For all distributions δ, we have that $M^n\delta$ is converging (uniformly over all initial distributions) towards δ_{stat} as n tends to infinity. In fact, the proof of uniform convergence follows the following lines. In the case of *irreducible aperiodic* Markov Chains, it is well known that $M^n\delta$ converges uniformly towards a distribution δ_{stat} which does not depend upon the initial δ [LPW09]. For irreducible periodic Markov chain, $M^n\delta$ has the same property. Last, [AAGT15] lift this result to the general case (reducible chains) by a careful analysis.

Hence for all $\gamma \neq \sum_{q \in Goal} \delta_{stat}(q)$, there exists a N (independent of δ) such that either for all $n \geqslant N, \delta \in \text{Init}$, $M^n\delta$ will be strictly above γ, or for all $n \geqslant N, \delta \in \text{Init}$, $M^n\delta$ will be strictly below γ. This gives $\mathcal{L}(\text{Init}, \mathcal{A}) = S_1.A^\omega + S_2.B^\omega$ where S_1 and S_2 are finite sets of finite words of length $< N$. Hence $\mathcal{L}(\text{Init}, \mathcal{A})$ is regular. □

In general, if $\mathcal{L}(\text{Init}, \mathcal{A})$ is regular, then any regular question will be decidable. For instance, if $\mathcal{L}(\text{Init}, \mathcal{A})$ is regular, then it is decidable whether there exists $\delta_0 \in \text{Init}$ such that the existence problem is true for $\mathcal{A}, \delta_0$. One can also ask whether for a given convex polytope Q, some property (such as positivity) expressed e.g. with $LTL_\mathcal{I}$ [AAGT15] is true. Taking δ in the interior of Q, this corresponds to checking the robustness of the property around δ.

Clearly, simple PA $\mathcal{A}$ does not ensure the regularity of $\mathcal{L}(\text{Init}, \mathcal{A})$ because of Theorem 7.8 (by choosing $\text{Init} = \{\delta_0\}$ which is a convex polytope). Surprisingly, restricting eigenvalues to be *distinct* and *roots of real numbers* does not ensure regularity either [AGKV16]. In the following, we thus take a stronger restriction: we assume that the eigenvalues of M are *distinct* and *positive real numbers*. That is,

$p_0 = 1 > p_1 > \cdots > p_k \geqslant 0$ with $k+1 = |Q|$ the number of states. From Proposition 7.10, we obtain as corollary that for all δ_0, we have either $\rho_{\delta_0} = wA^\omega$ or $\rho_{\delta_0} = wB^\omega$ for w a finite word of $\{A, B\}^*$:

Corollary 7.15. *Let M be a simple (or just diagonalizable) stochastic matrix with positive real eigenvalues. Then every trajectory ρ_{δ_0} is ultimately constant.*

However, the language $\mathcal{L}(\text{Init}_{yeast}, M_{yeast})$ shows that $\mathcal{L}(\text{Init}, \mathcal{A})$ is not always of the simple form $\bigcup_{w \in W_A} wA^\omega \cup \bigcup_{w \in W_B} wB^\omega$, for W_A, W_B two finite sets of finite words over $\{A, B\}^*$. Nevertheless, in the next two sections, we succeed in proving the regularity of $\mathcal{L}(\text{Init}, \mathcal{A})$, which is our main result:

Theorem 7.16. *Let $\mathcal{A}$ be a MC with distinct positive real eigenvalues, and* Init *be a convex polytope of (initial) distributions. Then, $\mathcal{L}(\text{Init}, \mathcal{A})$ is effectively regular.*

7.2.1 Partition of the set Init of initial distributions

Recall that we write $u_\delta(n) := \sum_{i=0}^{k} a_i(\delta) p_i^n$, where $a_i(\delta)$ are given by Equation (7.2) from the previous section. Because the eigenvalues are real numbers, $a_i(\delta)$ is a real number for every i and δ. Notice that a_i is a linear function in δ, that is, $a_i(\alpha\delta_1 + \beta\delta_2) = \alpha a_i(\delta_1) + \beta a_i(\delta_2)$. The trajectory ρ_δ depends crucially on the sign of $a_0(\delta)$, and if $a_0(\delta) = 0$, on the sign of $a_1(\delta)$, etc. First, for all $i \leqslant k$, let $L_i = \{\delta \mid a_0(\delta) = \cdots = a_i(\delta) = 0\}$. This is a vector space (in $\mathbb{R}^k$), as for any $\nu_1, \nu_2 \in \mathbb{R}^k$, we have $\nu_1, \nu_2 \in L_i$ implies that any linear combination $\alpha\delta_1 + \beta\delta_2 \in L_i$ (since $a_i(\nu)$ is linear in ν, and the kernel of a linear function is a vector space).

We will divide the space of distributions into a finite set $\mathcal{H}$ of convex polytopes $H \in \mathcal{H}$ to keep the sign of each a_i constant on each polytope. Each $H \in \mathcal{H}$ satisfies that for all $e, f \in H$, for all $i \leqslant k$, we have $a_i(e), a_i(f)$ do not have different signs (either one is 0, or both are positive or both are negative). This can be done since $a_i(\nu)$ is continuous (as it is linear) and the set $\mathcal{H}$ is finite because for each $i \neq k$, sets $\{a_i(\delta) > 0, \forall\delta\}$ and $\{a_i(\delta) < 0, \forall\delta\}$ can be separated by an hyperplane in $\mathbb{R}^{k+1}$, so the space can be divided into at most 2^{k+2} parts. This is pictorially represented in the left of Figure 7.17. For instance, we divide Init_{yeast} into three polytopes: $\{(1/3, y, 2/3 - y) \mid y \leqslant 1/3\}$ and $\{(1/3, y, 2/3 - x) \mid 1/3 \leqslant y \leqslant 5/12\}$ and $\{(1/3, y, 2/3 - x) \mid y \geqslant 5/12\}$ as for $\delta = (1/3, 1/3, 1/3)$ we have $a_0(\delta) = 1$, $a_1(\delta) = 0$ (and $a_2(\delta) = -1/5$) and for $\delta = (1/3, 5/12, 1/4)$ we have $a_0(\delta) = 1$, $a_1(\delta) = -1/5$, $a_2(\delta) = 0$.

In general, we can assume that each of $H \in \mathcal{H}$ is the convex hull of $k+2$ points (else we divide further: this can be done as the space has dimension $k+1$). Consider the right part of Figure 7.17. Let Init be the convex hull of points e, f, g, h (in three dimensions) and $a_0(x) = 0$ and $a_2(x) > 0$ for all $x \in \{e, f, g, h, t\}$. Hence the sign of each trajectory ultimately depends upon $a_1(x)$. In the example, $a_1(g) = a_1(h) = 0$

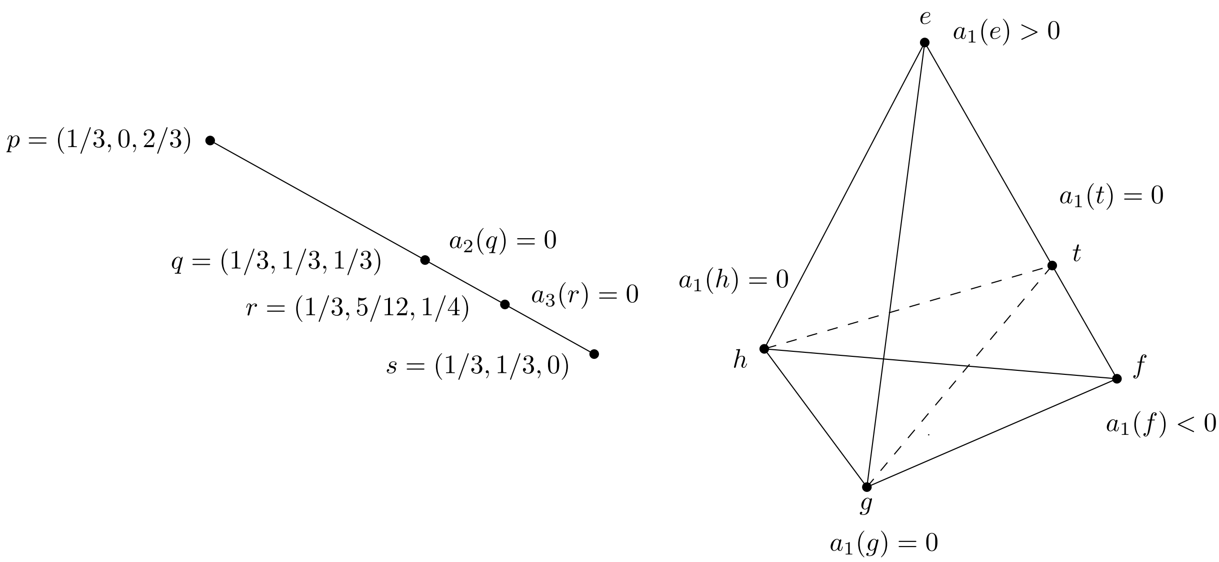

Figure 7.17 – Breaking into convex polytopes with constant signs

while $a_1(e) > 0 > a_1(f)$. Then there is a point t between e and f for which $a_1(t) = 0$ (in fact, $t = |a_1(f)|/(|a_1(e)|+|a_1(f)|)e+|a_1(e)|/(|a_1(e)|+|a_1(f)|)f$). We have $L_1 \cap \text{Init}$ is the convex hull of h, g, t. We break Init into two convex polytopes, the convex hull of h, g, t, e and the convex hull of h, g, t, f.

Let $H \in \mathcal{H}$. We let P be the finite set of (at most $k + 2$) extremities of H. In particular, H is the convex hull of P. Now it suffices to show that the language $\mathcal{L}(H)$ (taking H as the initial set of distributions) of each of these convex polytopes H is regular to prove that the language $\mathcal{L}(\text{Init}) = \bigcup_{H\in\mathcal{H}} \mathcal{L}(H)$ is regular.

7.2.2 High level description of the proof

The proof of the regularity of the language $\mathcal{L}(H)$ starting from the convex polytope H is performed as follows. We first prove that there exists a N_{max} such that the ultimate language (after N_{max} steps) of H is effectively regular using analytical techniques.

Definition 7.18. *Given N_{max}, the* ultimate language *from a convex polytope H is defined as* $\mathcal{L}_{ult}^{N_{max}}(H) = \{v \mid \exists w \in \{A, B\}^{N_{max}}, wv \in \mathcal{L}(H)\}$.

In the next section (Corollary 7.25), we show that this ultimate language $\mathcal{L}_{ult}^{N_{max}}(H)$ is regular, of the form $A^*B^* \cdots B^*A^\omega \cup A^*B^* \cdots A^*B^\omega$ with a bounded number of switches between A and B's. However, while for each prefix $w \in \{A, B\}^{N_{max}}$, the set H_w of initial distributions in H whose trajectory starts with w is a convex polytope; the language $\mathcal{L}(H_w)$ from H_w can be complex to represent. It is not in general $w\mathcal{L}_{ult}^{N_{max}}(H)$, but a strict subset.

In Section 7.4 (Lemma 7.28), we prove that the language $\mathcal{L}(H')$ associated with some carefully defined convex polytope $H' \subseteq H$ is a regular language, of the form $\bigcup_{w \in W} wA^iA^*B^* \cdots$
$B^*A^\omega \cup wA^iA^*B^* \cdots A^*B^\omega$ for a finite set W. Further, removing H' from H gives rise to a finite number of convex polytopes with a smaller number of "sign-changes", as formally defined in the next section. Hence we can apply the arguments inductively (requiring potentially to change the N_{max} considered). Finally, the union of these languages gives the desired regularity characterization for $\mathcal{L}(H)$.

7.3 Ultimate language

7.3.1 Limited number of switches.

We first show that the ultimate language $\mathcal{L}_{ult}^{N_{max}}(H)$ is included into $A^*B^*A^* \cdots A^*$ $B^\omega \cup A^*B^*A^* \cdots B^*A^\omega$ for some $N_{max} \in \mathbb{N}$, with a limited number of switches between A and B depending on properties of the set P of extremities of H.

We start by considering the generalisation of a sequence u_δ to a function over positive reals, and we will abuse the notation u_δ to denote both the sequence and the real function.

Definition 7.19. *A function of type $k \in \mathbb{N}$ is a function of the form $u : \mathbb{R}_{>0} \to \mathbb{R}$, with $u(x) = \sum_{j=0}^{k} \alpha_j p_j^x$, where $p_0 > \cdots > p_k > 0$.*

In Figure 7.20 function of type 2.

Now, let $u : \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ be a continuous function. We can associate with function u the (infinite) word $L(u) \in \{A, B\}^\omega, L(u) = (a_0a_1\ldots)$, where for all $n \in \mathbb{N}$, a_n is defined as $a_n = A$ if $u(n) \geqslant 0$ and $a_n = B$ otherwise. We have easily that $\rho_\delta = L(u_\delta)$. Knowing the zeros of u_δ and its sign before and after the zeros, defines uniquely the trajectory ρ_δ.

For example, let u be such that it has four zeros: $u(N - 0.04) = u(N + 10.3) = u(N + 20) = u(N + 35) = 0$ for some integer N. Assume that $u(0) < 0, u(N + 1) > 0, u(N + 11) < 0, u(N + 30) < 0$ and $u(N + 40) > 0$. Thus, by continuity of u, u is strictly negative on $[0, N - 1]$, strictly positive on $[N, N + 10]$, non-positive on $[N + 11, N + 34]$ and non-negative on $[N + 35, \infty)$. Thus the associated trajectory $\rho_\delta = B^N A^{11} B^{24} A^\omega$.

Hence, it is important to analyse the zeros of functions u_δ. If the number of zeros is bounded, then the number of alternations between A's and B's in any trajectory

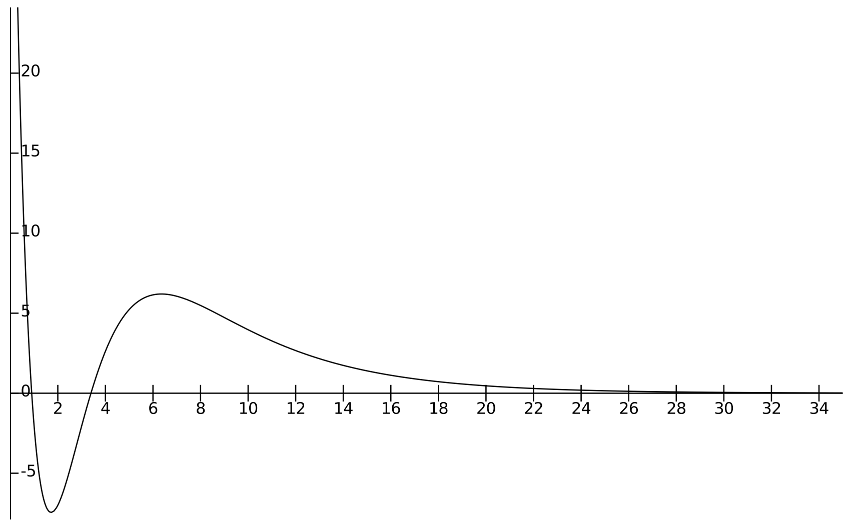

Figure 7.20 – Function of type 2 $f(x) = 40 \times 0.8^x - 380 \times 0.5^x + 390 \times 0.4^x (p_0 = 0.8, p_1 = 0.5, p_2 = 0.4)$

ρ_δ from $\delta \in H$ will be bounded. In fact, it is a standard result (which we do not use hence do not reprove here) that every type k function u has at most k zeros. We now show a more precise bound on the number of zeros. Namely, for the convex hull H' of a finite set P' of distributions in H, the number of alternations between A's and B's in H' is limited by the number of alternations of the sign of the dominant coefficients of the distributions in P'.

Let $z \in \mathbb{N}$. For $i \in \{0, \ldots, z\}$, let $u^i(x) := a_0^i p_0^x + a_1^i p_1^x + \cdots + a_k^i p_k^x$, with $p_0 > p_1 > p_2 > \ldots > p_k > 0$, representing the functions associated with the $z + 1$ extremities of H'. We denote $dom(u^i)$ the dominant coefficient of u^i, that is the smallest integer j with $a_j^i \neq 0$. We reorder $(u^i)_{i \in \{0,\ldots,z\}}$ such that $dom(u^i) \leqslant dom(u^{i+1})$ for all $i < z$. We denote $sign\_dom(u^i) \in \{+1, -1\}$ as the sign of $dom(u^i)$. We will assume, as for H, that for all i, i', j, a_j^i and $a_j^{i'}$ have the same sign, we can do this assumption as we show in Section 7.2.1. We let $Z(u^0, \cdots, u^z) = |\{i \leqslant z - 1 \mid sign\_dom(u^i) \neq sign\_dom(u^{i+1})\}|$. That is, $Z(u^0, \cdots, u^z)$ is the number of switches of sign between the dominant terms of u^i and u^{i+1}. We have $0 \leqslant Z(u^0, \cdots, u^z) \leqslant z$. Notice that as for $dom(u^i) = dom(u^j)$, we have $sign\_dom(u^i) = sign\_dom(u^j)$, $Z(u^0, \cdots, u^z)$ does not depend upon the choice in the ordering of $(u^i)_{i \in \{0,\ldots,z\}}$. We can now give a bound on the number of zeros of functions which are convex combinations of $u^0 \cdots u^z$.

Lemma 7.21. *Let $u^0 \cdots u^z$ be $z + 1$ type k functions. There exists a $N_{max} \in \mathbb{N}$ such that for all $\lambda_i \in [0, 1]$ with $\sum_i \lambda_i = 1$, denoting $u(x) = \sum_{i=0}^{z} \lambda_i u^i(x)$, $u(x)$ has at most $Z(u^0, \cdots, u^z)$ zeros after N_{max}. Further, if $u(x)$ has exactly $Z(u^0, \cdots, u^z)$*

zeros after N_{max}, then its sign changes exactly $Z(u^0, \cdots, u^z)$ times (that is, no zero is a local maximum/minimum).

In other words, we show that $u(x)$ behaves like a polynomial of degree $Z(u^0, \cdots, u^z)$ (as it has $Z(u^0, \cdots, u^z)$ dominating terms), although it has degree $k > Z(u^0, \cdots, u^z)$. To simplify notation, let $\ell(i) = \mathrm{dom}(u^i)$ for all i. We prove that the coefficients $a^i_j p^x_j$ for all $j > \ell(i)$ play a negligible role wrt. $a^i_{\ell(i)} p^x_{\ell(i)}$.

To do so, we use derivatives to study the sign of $u(x)$, which is a linear combination of $z + 1$ functions, u^i for all $0 \leqslant i \leqslant z$. Dividing $u(x)$ by a well chosen positive coefficient (of the form p^x) before differentiation allows us to obtain a linear combination of z functions. An induction allows us to conclude.

Proof. For all $r \in \mathbb{N}$, we introduce a small constant $\varepsilon(r) > 0$ depending on the number $(z - r)$ of functions considered. We start by defining $m(r, p_0, \ldots, p_k) > 0$, the min over all $0 \leqslant r \leqslant s \leqslant z$ and $0 \leqslant j \leqslant k$ with $j \neq \ell(r)$ of $|\frac{\log(\frac{p_{\ell(s)}}{p_{\ell(r)}})}{\log(\frac{p_j}{p_{\ell(r)}})}|$. The min exists and it is strictly positive because it is among a finite number of values, all strictly positive. We now define recursively $\varepsilon : \{0, \ldots, z\} \times \mathbb{R}^{z+1}_{>0} \to \mathbb{R}_{>0}$:

- $\varepsilon(z, p_0, \ldots, p_k) = \frac{1}{2k}$ and
- for all $0 \leqslant r < z$, $\varepsilon(r, p_0, \ldots, p_k) = \frac{m(r,p_0,\ldots,p_k)}{(1+3z)^2} \varepsilon(r + 1, p_0, \ldots, p_k)$.

It is now easy to show by induction that for all $q \notin \{p_0, \ldots, p_k\}$, for all r, $\varepsilon(r, \frac{p_0}{q}, \ldots, \frac{p_k}{q}) = \varepsilon(r, p_0, \ldots, p_k)$. We then define $\varepsilon(r) = \varepsilon(r, p_0, \ldots, p_k)$ for all $0 \leqslant r \leqslant z$. We can also show by induction that for all r, $\varepsilon(r) \leqslant \frac{1}{2k}$.

We will use the following technical lemma, which we prove later.

Lemma 7.22. *Let I be an interval of $\mathbb{R}_{\geqslant 0}$. Let $p_0 > \cdots > p_k > 0$ be positive reals. Let $v^i(x) := b^i_0 q^x_0 + b^i_1 q^x_1 + \cdots + b^i_k p^x_k$ be a function of type k for all $i \in \{r, \ldots, z\}$, $0 \leqslant r \leqslant z$, s.t.,*

- *for all $i \in \{r, \ldots, z\}$, all $j \neq \ell(i)$ and all $x \in I$, $|b^i_j p^x_j| \leqslant |\varepsilon(z, p_0, \cdots, p_k) b^i_{\ell(i)} p^x_{\ell(i)}|$ (if this holds, we say that $|b^i_j p^x_j|$ is* negligible *wrt $|b^i_{\ell(i)} p^x_{\ell(i)}|$ and call this the negligibility hypothesis)*

Then for all $\lambda_r \geqslant 0, \ldots, \lambda_z \geqslant 0$ with $\sum_{i=r}^z \lambda_i = 1$, the function $v : x \mapsto \sum_{i=r}^z \lambda_i v^i(x)$ has at most $Z(b^r_{\ell(r)} p^x_{\ell(r)}, \cdots, b^z_{\ell(z)} p^x_{\ell(z)})$ zeros in I.

Further, if $v(x)$ has exactly $Z(b^r_{\ell(r)} p^x_{\ell(r)}, \cdots, b^z_{\ell(z)} p^x_{\ell(z)})$ zeros in I, then its sign changes exactly $Z(b^r_{\ell(r)} p^x_{\ell(r)}, \cdots, b^z_{\ell(z)} p^x_{\ell(z)})$ times (that is, its zeros in I are not local maximum or minimum).

Notice that in Lemma 7.22, $\ell(i)$ is not necessarily the dominating factor for v^i. In fact, v^i is u^i plus some factors. If I is bounded, it can be the case that $|b^i_j| \gg |b^i_{\ell(i)}|$ with $j > \ell(i)$.

Assume Lemma 7.22 has been proved. We then apply Lemma 7.22 with $r = 0$, $v^i = u^i$ for all $i \leqslant z$ and $I = [N_{max}, \infty)$, with N_{max} chosen such that the negligibility hypothesis is verified, which is possible as $\ell(i)$ is the dominating factor of $u(i)$ for all i. This implies that u has $Z(b^r_{\ell(r)}p^x_{\ell(r)}, \cdots, b^z_{\ell(z)}p^x_{\ell(z)}) = Z(u^0, \ldots, u^z)$ many zeros, since these are the dominant coefficients of the u^i. Thus, we obtain the statement of Lemma 7.21: for all $\lambda_i \in [0,1]$ with $\sum_i \lambda_i = 1$, denoting $u(x) = \sum_{i=0}^{z} \lambda_i u^i(x)$, $u(x)$ has at most $Z(u^0, \cdots, u^z)$ zeros after N_{max}. Further, if $u(x)$ has exactly $Z(u^0, \cdots, u^z)$ zeros after N_{max}, then its sign changes exactly $Z(u^0, \cdots, u^z)$ times (that is, its zeros are not local maximum/minimums). This would complete the proof of Lemma 7.21.

□

It now remains to prove the technical Lemma 7.22, which we do by induction on r:

Proof of Lemma 7.22. For $r = z$, the lemma is trivial as one has a unique function $v^z(x) := b^z_0 q^x_0 + b^z_1 q^x_1 + \cdots + b^z_k p^x_k$. Let $\ell = \ell(z)$. For all $x \in I$, we have $\sum_{i \neq \ell} |b^z_i p^x_i| \leqslant k\varepsilon(z, p_0, \ldots, p_k)|b^z_\ell p^x_\ell| \leqslant k\frac{1}{2k}|b^z_\ell p^x_\ell| \leqslant \frac{1}{2}|b^z_\ell p^x_\ell|$. Hence the sign of $v^z(x)$ is the sign of b^z_ℓ for all $x \in I$. That is, v^z has no zero in I. The further statement is thus trivially verified in this case.

Let $0 \leqslant r \leqslant z$. Assume that the lemma is true for all instances with functions $(v^{r+1}, \ldots, v^z)$. Let us prove that the lemma is true for all instances with functions $(v^r, \ldots, v^z)$.

Let $v^i(x) := b^i_0 q^x_0 + b^i_1 q^x_1 + \cdots + b^i_k p^x_k$, for $i \in \{r, \ldots, z\}$ such that $1 \geqslant p_0 > p_1 > \ldots > p_k > 0$, $|b^i_j p^x_j| \leqslant |\varepsilon(r, p_0, \cdots, p_k) b^i_{\ell(i)} p^x_{\ell(i)}|$ for all $j \neq \ell(i)$ and $x \in I$. This hypothesis ensures that for all i, $(1 - k\varepsilon(r, p_0, \cdots, p_k))|b^i_{\ell(i)}|p^x_{\ell(i)} \leqslant |v^i(x)| \leqslant (1 + k\varepsilon(r, p_0, \cdots, p_k))|b^i_{\ell(i)}|p^x_{\ell(i)}$. As we have $\varepsilon(r, p_0, \ldots, p_k) \leqslant \frac{1}{2k}$ for all r, it gives

$$\frac{1}{2} b^i_{\ell(i)} p^x_{\ell(i)} \leqslant |v^i(x)| \leqslant \frac{3}{2} b^i_{\ell(i)} p^x_{\ell(i)} \tag{7.4}$$

Let $\lambda_1 \geqslant 0, \ldots, \lambda_z \geqslant 0$ with $\sum_{i \leqslant z} \lambda_i = 1$. Take the maximal $x_y \in I$ such that $v(x_y) = \sum_{r \leqslant i \leqslant z} \lambda_i v^i(x_y) = 0$ (if there is no such zero, then we are done). We can assume without loss of generality that $\ell(r) \neq \cdots \neq \ell(z)$, else it is easy to merge several u^i with the same $\ell(i)$ together (by replacing all u^i with the same $\ell(i)$ by the sum of all of them). We have $|\lambda_r v^r(x_y)| = |\sum_{i>r} \lambda_i v^i(x_y)|$ because x_y is a zero of v. Taking $s > r$ with $|\lambda_s v^s(x_y)|$ maximal, we have $|\sum_{i>r} \lambda_i v^i(x_y)| \leqslant z\lambda_s |v^s(x_y)|$. Thus $|\lambda_r v^r(x_y)| \leqslant z\lambda_s |v^s(x_y)|$.

We let $I' = I \cap [0, x_y]$. Using (7.4) for v^r and for v^s at $x_y \in I$, we have $\lambda_r |b^r_{\ell(r)}| p^{x_y}_{\ell(r)} \leqslant \lambda_s 3z |b^s_{\ell(s)}| p^{x_y}_{\ell(s)}$. Now, because $p_{\ell(r)} > p_{\ell(s)}$, we have for all $x \in I'$: $\lambda_r |b^r_{\ell(r)}| p^x_{\ell(r)} \leqslant \lambda_s 3z |b^s_{\ell(s)}| p^x_{\ell(s)}$. By applying the hypothesis of the negligibility, we thus get for all $x \in I'$ and all $j \neq \ell(r)$, $\lambda_r |b^r_j| p^x_j \leqslant \lambda_s 3z\varepsilon(r, p_0, \ldots, p_k)|b^s_{\ell(s)}| p^x_{\ell(s)}$. That is, the terms $\frac{\lambda_r}{\lambda_s} b^r_j q^x_j$, with $j \neq \ell(r)$ are small wrt $b^s_{\ell(s)} q^x_{\ell(s)}$ for $x \in I'$.

Let $q = p_{\ell(r)}$ and consider the function $v'(x) = \frac{v(x)}{q^x}$. Functions v' and v have the same zeros. We can derive v', which will cancel out every term using q^x: For all $r \leqslant i \leqslant z$, we define functions $f^i(x) := c_0^i(\frac{p_0}{q})^x + c_1^i(\frac{p_1}{q})^x + \cdots + c_k^i(\frac{p_k}{q})^x$ with:

— for $i \neq s$, f^i is the derivative of v^i, that is $c_j^i = \log(\frac{p_j}{q})b_j^i$ for $j \neq \ell(r)$, and $c_{\ell(r)}^i = 0$.

— $c_j^s = \log(\frac{p_j}{q})(b_j^s + \frac{\lambda_r}{\lambda_s}b_j^r)$ for $j \neq \ell(r)$, and $c_{\ell(r)}^s = 0$.

It is easy to check that $f(x) = \sum_{i=r+1}^{z} \lambda_i f^i(x)$ is the derivative of v'. We now prove the inequalities involving ε for $f^i(x)$ for all $x \in I'$. We do it for the most complex term, ie c_j^s with $j \neq \ell(s), \ell(r)$. We have

$$\begin{aligned}
\left|c_j^s(\frac{p_j}{q})^x\right| &= \left|\log(\frac{p_j}{q})\right| |(b_j^s + \frac{\lambda_r}{\lambda_s}b_j^r)|(\frac{p_j}{q})^x \\
&\leqslant \left|\log(\frac{p_j}{q})\right| \varepsilon(r, p_0, \ldots, p_k)(1+3z)|b_{\ell(s)}^s| \left(\frac{p_{\ell(s)}}{q}\right)^x \\
&\leqslant \left|\frac{\log(\frac{p_j}{q})}{\log(\frac{p_{\ell(s)}}{q})}\right| \varepsilon(r, p_0, \ldots, p_k)(1+3z)^2|c_{\ell(s)}^s| \left(\frac{p_{\ell(s)}}{q}\right)^x \\
&= \left|\frac{\log(\frac{p_j}{q})}{\log(\frac{p_{\ell(s)}}{q})}\right| m(r, p_0, \ldots, p_k)\varepsilon(r+1, p_0, \ldots, p_k)|c_{\ell(s)}^s| \left(\frac{p_{\ell(s)}}{q}\right)^x \\
&\leqslant \varepsilon(r+1, p_0, \ldots, p_k)|c_{\ell(s)}^s| \left(\frac{p_{\ell(s)}}{q}\right)^x
\end{aligned}$$

by definition of $m(r)$.

Recalling that $\varepsilon(r+1, \frac{p_0}{q}, \ldots, \frac{p_k}{q}) = \varepsilon(r+1, p_0, \ldots, p_k)$, we conclude $|c_j^s|(\frac{p_j}{q})^x \leqslant \varepsilon(r+1, \frac{p_0}{q}, \ldots, \frac{p_k}{q})|c_{\ell(s)}^s|(\frac{p_{\ell(s)}}{q})^x$ for all $x \in I'$, so we can apply the lemma to $f^{r+1}, \ldots, f^z$. Thus function f has at most $Z(c_{\ell(r+1)}^{r+1}p_{\ell(r+1)}^x, \ldots, c_{\ell(z)}^z p_{\ell(z)}^x)$ zeros in I'. It is easy to see that $c_{\ell(i)}^i$ has the opposite sign of $b_{\ell(i)}^i$ for all i, and thus we obtain $Z(c_{\ell(r+1)}^{r+1}p_{\ell(r+1)}^x, \ldots, c_{\ell(z)}^z p_{\ell(z)}^x) = Z(b_{\ell(r+1)}^{r+1}p_{\ell(r+1)}^x, \ldots, b_{\ell(z)}^z p_{\ell(z)}^x)$.

Now, consider v'. It has the same sign and zeros as v. Hence the last zero of v' in i is x_y. Because its derivative is f, v' (and thus v) has at most $1 + Z(b_{\ell(r+1)}^{r+1}p_{\ell(r+1)}^x, \ldots, b_{\ell(z)}^z p_{\ell(z)}^x)$ zeros in I'.

If $Z(b_{\ell(r)}^r p_{\ell(r)}^x, \cdots, b_{\ell(z)}^z p_{\ell(z)}^x) = 1 + Z(b_{\ell(r+1)}^{r+1}p_{\ell(r+1)}^x, \ldots, b_{\ell(z)}^z p_{\ell(z)}^x)$, (or if v has at most $Z(b_{\ell(r+1)}^{r+1}p_{\ell(r+1)}^x, \ldots, b_{\ell(z)}^z$ zeros), the induction proof is finished.

Else, we proceed by contradiction. It means that the sign of $b_{\ell(r)}^r$ and of $b_{\ell(r+1)}^{r+1}$ is the same. It also means that f has exactly $Z(b_{\ell(r+1)}^{r+1}p_{\ell(r+1)}^x, \ldots, b_{\ell(z)}^z p_{\ell(z)}^x)$ zeros and switches sign every time. Without loss of generality, assume that $b_{\ell(r+1)}^{r+1} > 0$. By induction, it is easy to see that the sign of $f(x_y)$ is the sign of $c^{r+1}_{\ell(r+1)}$, that is strictly negative.

In the same way, as $\ell(r)$ is the dominating factor of $v(x)$ in I, just after x_y (remember that $v(x_y) = v'(x_y) = 0$), the sign of v is $b^r_{\ell(r)} > 0$. This contradicts the continuity of v and the fact that $v(x_y) = 0$ and that its derivative is negative.

For the second statement, assume that v has exactly $\alpha := Z(b^r_{\ell(r)}p^x_{\ell(r)}, \dots, b^z_{\ell(z)}p^x_z)$ zeros in I. We know by the above that the derivative has exactly $\alpha - 1$ zeros $y_1, \dots, y_{\alpha-1}$ in I'. For all $i \in \{1, \alpha - 1\}$ there is one zero x_i of v between two consecutive zeros y_i, y_{i+1} of the derivative. Now, if by contradiction v does not change sign at one of its zeros, let say x_i, it means that $x_i = y_i$. In particular, it means that in $(y_i, y_{i+1}]$, there is no zero of v, which contradicts the fact that v has exactly α zeros in I'. It is also the case if the derivative is null at x_y. Last, v being continuous, it can not change sign after x_y as it has no zero other than x_y (by definition of x_y). □

Let $H \in \mathcal{H}$, and P its finite set of extremal points. We can apply Lemma 7.21 to $u^0, \dots, u^z$, the functions associated with the points of P (in decreasing order of dominating coefficient), and obtain a N_{max}. Now, since P is finite, the trajectories from P are ultimately constant, hence there exists N_y such that for all $i \leqslant y$, the trajectory of u^i is wA^ω or wB^ω for some $w \in \{A, B\}^{N_y}$. We define N_H to be the maximum of N_y and N_{max}. With this bound on the number of zeros, we deduce the following inclusion for the ultimate language $L^{N_H}_{ult}(H)$:

Corollary 7.23. *Let $y = Z(u^0, \dots, u^z)$. The ultimate language $L^{N_H}_{ult}(H) \subseteq C_1^* \cdots C_{y-1}^* C_y^\omega \cup C_1^* \cdots C_{y-1}^* C_{y-1}^\omega$ for $\{C_i, C_{i+1}\} = \{A, B\}$ for all $i < y$; and $C_y = A$ iff $sign\_dom(u^0)$ is positive.*

We can have 4 different sequences for $C_1^* \cdots C_{y-1}^* C_y^\omega$ with $\{C_i, C_{i+1}\} = \{A, B\}$, depending on the first and last letters C_1, C_y (or equivalently, C_y and parity of y which determines C_1).

The proof of our main result on regularity of $\mathcal{L}(H)$ will proceed by induction over the *switching-dimension* $Z(H)$ of H which we define as $Z(H) = Z(u^0, \dots, u^z)$. Notice that we could define the switching dimension for any convex set (not necessarily a polytope) whenever the sign of $a_i(\delta)$ does not change within the convex set. Finally, we also define $sign\_dom(H) = sign\_dom(u^0)$.

7.3.2 Characterization of the ultimate language.

We now show that the ultimate language of H is exactly $\mathcal{L}^{N_H}_{ult}(H) = A^*B^*A^* \cdots A^* B^\omega \cup A^*B^*A^* \cdots B^*A^\omega$, with at most $Z(H)$ switches of signs. We will state the associated technical Lemma 7.24 in the more general settings of "faces" as defined below, as it will be useful in the next section. Let P be the finite set of extremal points of a H. We call $(f^0, \dots, f^y) \subseteq P$ a *face* of H if $Z(v^0, \dots, v^y) = y = Z(H)$ for the functions $(v^0, \ldots, v^y)$ associated with the extremal points $(f^0, \ldots, f^y)$. Notice that denoting H' the convex hull of F, we can choose $N_{H'} = N_H$.

Lemma 7.24. *Given a face $(f^0, \ldots, f^y) \subseteq P$ of H with associated functions v^i, we have, for all $n_1, n_2, \ldots, n_y \in \mathbb{N}$ there exist $\lambda_i \in [0,1]$ with $\sum_i \lambda_i = 1$, such that denoting $\tilde{v}(x) = \sum_{i=1}^{y} \lambda_i v^i(x)$, $L(\tilde{v}) = wA^{n_1}B^{n_2} \ldots B^{n_y}A^{\omega}$ (assuming y is even) for some prefix $w \in \{A,B\}^{N_H}$.*

That is, for all $n_1, \ldots n_y$, one can find a prefix w of size N_H and a point δ in the convex hull of $e^1, \ldots, e^y$, such that $\rho_\delta = wA^{n_1}B^{n_2} \cdots B^{n_y}A^{\omega}$ (assuming the correct parity of y). Let H' be the convex hull of $f^0, \ldots, f^y$. As $(f^0, \ldots, f^y)$ is a face, $Z(H') = Z(H)$.

Proof. Let $N_{max} < n_1 < \cdots < n_y$ be integers. We define inductively $x_0 = N_{max} + 1/2$ and $x_j := x_{j-1} + n_j$ for all $1 \leqslant j \leqslant y$ if $n_j \neq 0$ and $x_j := x_{j-1} + \frac{1}{2y}$ if $n_j = 0$.

We build inductively a function $v_i^j(x)$, convex combination of $\{v^i, v^{i+1}, \ldots, v^{i+j}\}$, such that $v_i^j(x_k) = 0$ for all $k \in \{1, \ldots, j\}$. Further, if i is odd (resp. even), we have $v_i^j(x) > 0$ (resp. $v_i^j(x) < 0$) for all $x > x_j$. The initialization is trivial: we have that $\forall x > N_{max}, v^1(x)$ is positive, by choice of N_{max}. We let $v_i^0(x) = v^i$ for all i.

Induction step: Let $0 < j < y$. Assume that we have built $v_i^{j-1}(x)$ for all i. The first thing to remark is that for all i, any convex combination of $v_i^{j-1}(x)$ and $v_{i+1}^{j-1}(x)$ will have a zero at $x_1, \ldots, x_{j-1}$ as both terms are zero there. It remains to choose one which also have a zero at x_j. By induction, $\forall x > x_{j-1}, v_i^{j-1}(x)$ is positive (resp. negative) when i is odd (resp. even). Thus it exists $\lambda_i^j \in (0,1)$ such that $\lambda_i^j v_i^j(x_j) + (1-\lambda_i^j) v_{i+1}^j(x_j) = 0$. We thus define $v_i^j(x) = \lambda_i^j v_i^j(x) + (1-\lambda_i^j) v_{i+1}^j(x)$ and it has the required j zeros, after N_{max}. As it is a linear combination of $v_1 \cdots v_{i+j}$, it has exactly j zeros after N_{max} (by lemma 7.21), and thus, $\forall x > N_j, v_j^i(x)$ is positive (negative) if i is odd (even) (as it has no zero after x_j and we know its asymptotic behaviour).

Then v_1^y has $\{x_1, \ldots, x_y\}$ as zeros, and by lemma 7.21, it switches sign each time. Hence the language of v_y^1 is $wA^{n_1}B^{n_2} \ldots A^{\omega}$ (or $wB^{n_1}A^{n_2} \ldots A^{\omega}$ if y odd) for some prefix w of size $|w| = N_{max}$. □

Then, the ultimate language of H' (i.e., the language after prefixes of size N_H associated with y) contains $A^*B^* \ldots B^*A^{\omega}$ with y switches between A and B, which is the converse of Corollary 7.23. We can thus deduce the following about the ultimate language:

Corollary 7.25. *$L_{ult}^{N_H}(H) = L_{ult}^{N_H}(H') = C_1^* C_2^* \ldots C_y^* A^{\omega} \cup C_1^* C_2^* \ldots C_{y-1}^* B^{\omega}$ with $\{C_i, C_{i+1}\} = \{A, B\}$.*

Proof. We first prove the result for $L_{ult}^{N_H}(H')$. We can apply lemma 7.24 to H' and lemma 7.21 to H'. We obtain the first part of the union. Now, let $H'' \subseteq H'$ be the convex hull of $e^1, \cdots, e^y$ (that is excluding e^0). Each point δ in $H' \backslash H''$ has a trajectory which ends with A^ω, as $dom(u_\delta) = dom(v^1)$, and thus $sign\_dom(u_\delta) = sign\_dom(v^1)$ by construction of H (and $H' \subseteq H$). Thus the points with trajectory ending with B^ω are in H'', and applying lemma 7.21, we know that their ultimate trajectory has at most $y-1$ switches. Applying lemma 7.24 to H'', we obtain the second hand of the union. Now, $L_{ult}^{N_H}(H') \subseteq L_{ult}^{N_H}(H)$, and $L_{ult}^{N_H}(H) \subseteq C_1^* C_2^* \dots C_y^* A^\omega \cup C_1^* C_2^* \dots C_{y-1}^* B^\omega$ by Corollary 7.23. □

However, we cannot immediately conclude that $\mathcal{L}(H)$ is regular. Though N_H is finite, computable and there are a finite number of prefixes w of size N_H, we need to show that the subset of $\mathcal{L}_{ult}^{N_H}(H)$ appearing after a given $w \in \{A, B\}^{N_H}$ is (effectively) regular. This is what we do formally in the following section.

7.4 Regularity of the language

Let $\{e^0, \cdots, e^z\} = P$ the extremal points of H. Let u^p the function associated with each $e^p \in P$. We denote $y = Z(H) = Z((u^p)_{p \leqslant z})$. We will show the regularity of $\mathcal{L}(H)$ using an induction on $Z(H)$.

For $Z(H) = 0$, the regularity of $\mathcal{L}(H)$ is trivial as all the dominant coefficients have the same sign. Thus, by Corollary 7.23, the ultimate language is $\mathcal{L}_{ult}^{N_H}(H) = A^\omega$ and then the language is $\mathcal{L}(H) = \bigcup_{w \in W} wA^\omega$; or the ultimate language is $\mathcal{L}_{ult}^{N_H}(H) = B^\omega$ and the language is $\mathcal{L}(H) = \bigcup_{w \in W} wB^\omega$, for a finite set of $W \subseteq \{A, B\}^{N_H}$.

For $w \in \{A, B\}^{N_H}$, consider $H_w = \{\delta \in H \mid \rho_\delta = wv\}$, i.e., the language of words which begin with the prefix w. It is easy to see that $H_w \subseteq H$ is a polytope. Hence $Z(H_w) \leqslant Z(H)$. Observe that $\mathcal{L}(H) = \bigcup_{w \in \{A,B\}^{N_H}} \mathcal{L}(H_w)$. To show the regularity of $\mathcal{L}(H)$, we show the regularity of $\mathcal{L}(H_w)$ for each of the finitely many $w \in \{A, B\}^{N_H}$. For each $w \in \{A, B\}^{N_H}$, we have two cases: either $Z(H_w) < Z(H)$; then we apply the induction hypothesis and we are done. Or else, $Z(H_w) = Z(H) = y$. In this case, the sketch of proof is as follows:

- We show that there exists J such that for all $i \leqslant y$ and all $j \geqslant J$, we have a point h_j^i in H_w with trajectory $wC_1^j C_2 C_3 \cdots C_{i-1} C_i^\omega$. This is shown by applying lemma 7.24 to each face $(f^0, \dots, f^y)$ of H and then using convexity arguments and the fact that $Z(H_w) = Z(H)$.
- Subsequently, denoting H' the convex hull of $h_J^0 \cdots h_J^y$, we will deduce that $\mathcal{L}(H')$ is a regular language of the form $wC_1^J C_1^* C_2^* C_3^* \cdots C_{i-1}^* C_i^\omega$,
- Partitioning $H_w \backslash H'$ into a finite set of polytopes, we obtain polytopes of lower switching-dimensions, which have regular languages by induction.
- We conclude since the finite union of these regular languages is a regular language, namely $\mathcal{L}(H_w)$.

We now formalize the above proof sketch in a sequence of lemmas. For all faces F of H, applying Lemma 7.24 gives for all $j \in \mathbb{N}$, a point $g_j(F)$ of the convex hull of F with trajectory $w_j C_1^j C_2 C_3 \cdots C_y^\omega$, for some $w_j \in \{A,B\}^{N_H}$. We now prove that (g_j) converges towards f^y, the point of F with lowest dominant term.

Let $i \leqslant y = Z(H)$. A i-subface of H is a subset $F = (f^0, \ldots, f^i)$ of the set P of extremal points of H such that $Z(F) = i$.

Lemma 7.26. *For every $i \leqslant y$ and every i-subface $F_i = (f^0, \ldots, f^i)$ of H, $(g_j^i(F))_{j \in \mathbb{N}}$ converges towards f^i as j tends to infinity.*

Proof. For $i = 0$, the result is trivial. Let $0 < i \leqslant y$. By contradiction, assume that there exists a dimension d (as there is a finite number of dimensions) and an infinite set J of indices $j \in \mathbb{N}$ such that g_j^i is bounded away from f^i on dimension d. Let b be this bound. Let H' be the convex polytope made of points of the convex hull of F_i at distance at least b from f^i on dimension d (g^y is an extremal point of H, hence there is only one direction of being at distance at least b on dimension d). Applying lemma 7.21 to H', we obtain a bound $N_{H'}$ such that the number of switches after $N_{H'}$ (in general, $N_{H'} > N_H$) of any point of H' is at most $i-1$, as $Z(H') < Z(F_i) = i$. Now, as J is infinite, one can find a $j \in J$ with $j > N_{H'} + 1$. We have that the trajectory of $g_j^i \in H'$ is $w' C_1^j C_2 C_3 \cdots C_i^\omega$ for some $w' \in \{A,B\}^{N_H}$, which switches signs i times after $N_{H'}$, a contradiction. □

In the same way, for all $r < i$, we can prove that denoting $d_j^{i,r}$ the distance of g_j^i to the convex hull of $(f^0, \ldots, f^r)$, we have $d_j^{i,r+1}/d_j^{i,r}$ converges towards 0 as j tends to infinity. Let $D(e, f^0, \ldots, f^{r+1})$ be the distance from e to the convex hull of $(f^0, \ldots, f^{r+1})$ divided by the distance from e to the convex hull of $(f^0, \ldots, f^r)$. We thus want to show that $D(g_f^i, f^0, \ldots, f^{r+1})$ tends towards 0.

First, for $r = i-1$, this is trivial as $d_j^{i,r+1} = 0$ for all i,j. Else, for $r < i-1$, if it was not the case, there would exist a bound b and an infinite set J of indices with $d_j^{i,r+1}/d_j^{i,r} > b$ for all $j \in J$. Then as above, by considering H' the the convex polytope made of points e of the convex hull of F_i with $D(e, f^0, \ldots, f^{r+1}) > b$, we have $Z(H') < Z(F_i) = i$ and the same contradiction as above applies.

For all j, we consider $F(y,j)$ the convex hull of $\{g_j(F) \mid F \text{ is a face of } H\}$. Every point of $F(y,j)$ has trajectory $w' C_1^j C_2 C_3 \cdots C_y^\omega$ for some $w' \in \{A,B\}^{N_H}$. We then show by convexity that H_2 intersects $F(y,j)$, i.e., it has a point with trajectory $w' C_1^j C_2 C_3 \ldots C_y^\omega$.

Lemma 7.27. *Let a convex $H' \subseteq H$ and $w \in \{A,B\}^{N_H}$ with $Z(H'_w) = Z(H')$. There exists J s.t. for all $j > J$, $F(y,j) \cap Closure(H'_w) \neq \emptyset$.*

Proof. Let $y+1$ points $h^0, \ldots, h^y$ in $Closure(H'_w)$ such that $Z(h^0, \ldots, h^y) = y$. We choose J such that for all face $F = (f^0, \ldots, f^y)$ of H, for all $j > J$,

— $g_j^y(F)$ is closer to f^y than any h^i is from h^y, $i \neq y$.

— for all r and all $k > r$, $D(g_j^y(F), f^0, \ldots, f^r) < D(h^k, h^1, \ldots, h^r)$

Then we have that $Closure(H'_w)$ intersects the convex hull of $(g_j^i(F))_{F \text{ a face of } H}$.

As $g_j^y(F) \in F(y,j)$ for all j, F, we have In particular $F(y,j) \cap Closure(H'_w) \neq \varnothing$. □

Similarly, for all $i \leqslant y$ we can define a polytope $F(i,j)$. All the points in $F(i,j)$ have trajectory $w'C_1^jC_2C_3\cdots C_i^\omega$ for some $w' \in \{A,B\}^{N_H}$. We can find a J_i and a point $h_j^i \in H_w$ with trajectory $wC_1^jC_2C_3\cdots C_i^\omega$ for all $i \leqslant y$ and all $j > J_i$. Now, as the number of $i \leqslant y$ is bounded, one can find such a J uniform over all $i \leqslant y$ (by taking maximum over all i).

Consider $F(J)$ the convex hull of $F(0,J), \ldots, F(y,J)$. By convexity, all the points in $F(J)$ have their n-th letters of trajectory as C_1 for all $n \in [N_H + 1 \cdots N_H + J]$, since this is true for all points of $F(i,J)$. Hence, the language of $H_w \cap F(J)$ is included into $wC_1^JC_1^*C_2^*\cdots C_y^\omega \cup wC_1^JC_1^*C_2^*\cdots C_{y-1}^\omega$, because of the bound on the number of alternations after N_H of trajectories from points of H (Lemma 7.21). We show now that we have equality.

Lemma 7.28. *The language of the convex hull of* $\{h_J^0, \ldots, h_J^y\}$ *is exactly* $wC_1^JC_1^*C_2^*C_3^*\cdots C_{y-1}^*C_y^\omega \cup wC_1^JC_1^*C_2^*C_3^*\cdots C_{y-2}^*C_{y-1}^\omega$.

Hence the language of $H_w \cap F(J)$ is $wC_1^iC_1^*\cdots C_y^\omega \cup wC_1^iC_1^*\cdots C_{y-1}^\omega$.

Next, we prove Lemma 7.28 for which we first need an intermediate lemma describing the exact language of the convex hull of two points of H_w. In the following, we will abuse notation of a point to also define the function associated with its trajectory: $g(n) \geqslant 0$ iff the n-th letter of the trajectory starting from g is an A.

Lemma 7.29. *Let* $e_0 \cdots e_y$ *be points of* H_w *with* $Z(e_0, \ldots, e_y) = Z(H_w)$. *Assume that the trajectory of* $e = e_k$ *is* $wC_1^{i_1}C_2^{i_2}\cdots C_{k-1}^{i_{k-1}}C_k^\omega$ *with* $i_j > 0$ *and* $\{C_j, C_{j+1}\} = \{A,B\}$ *for all* $j < k$. *Assume also that the trajectory of* $f = e_{k-1}$ *is* $wC_1^{i_1}C_2^{i_2}\cdots C_{k-2}^{i_{k-2}}C_{k-1}^\omega$. *Let* $i' > i_{k-1}$. *Then there is a point* g *on the segment* (e,f) *with* $g(N_{max} + \sum_{j=1}^{k-2} i_j + i' + 1/2) = 0$.

Notice that any g on (e,f) has at least $k-2$ zeros, one in each $(N_{max} + i_1 + \cdots + i_j, N_{max} + i_1 + \cdots + i_{j+1})$. The g we will build thus have trajectory $wC_1^{i_1}C_2^{i_2}\cdots C_{k-1}^{i'}C_k^\omega$. Hence, the language of $[e,f)$ is

$$wC_1^{i_1}C_2^{i_2}\cdots C_{k-1}^{i_{k-1}}C_{k-1}^*C_k^\omega$$

Proof. Let $i > N$. Let g define a point on (e,f) to be specified later. For $a \in \{e,f,g\}$, we define u_a as the function associated to the point a. Let $x := |w| + i_1 + i_2 + \ldots + i_{z-3} + i + 1/2$. We have $u_e(x) > 0$ and $u_f(x) < 0$ (in the unlikely case where $u_f(x) = 0$

with this x, i.e., $u_f(x) = 0$ implies the letter is B and the derivative of u_f is null in x, we just take $x + 1/4$. Because of the maximal number of zeros of u_f, $u_f(x + 1/4) \neq 0$ if $u_f(x) = 0$). So there exists $\lambda \in (0, 1)$ such that $\lambda u_e(x) + (1-\lambda)u_f(x) = 0$. Let g be the point $\lambda e + (1 - \lambda)f$ on segment (e, f), and u_g its associated function. We have $u_g = \lambda u_e + (1-\lambda)u_f$ by linearity. Further, as $g = \lambda e+(1-\lambda)f$ and both e and f have prefix $wA^{i_1}B^{i_2}A^{i_3} \cdots A^{i_{z-3}}$, then g has also prefix $wA^{i_1}B^{i_2}A^{i_3} \cdots A^{i_{z-3}}$. It means that u_g changes sign between $|w|+i_1-1$ and $|w|+i_1$, ..., between $|w|+i_1+i_2+\ldots+i_{z-3}-1$ and $|w|+i_1+i_2+\ldots+i_{z-3}$. In particular, u_g has a zero in every of these $z-2$ intervals. Thus u_g has $z - 1$ zeros. By lemma 7.21, it switches signs exactly at these zeros, and never elsewhere in $[N_{max}, +\infty)$. Thus the trajectory of g is $wA^{i_1}B^{i_2}A^{i_3} \cdots A^{i_{z-2}}B^iA^\omega$. Further, as g is on the segment $[e, f]$, both $e, f \in H_w$ and H_w is convex, then $g \in H_w$. □

We can now finish the proof of lemma 7.28.

Lemma 7.28. *Let $e_0 \cdots e_y$ be points of H_w with $Z(e_0, \ldots, e_y) = Z(H_w)$. Let $J \in \mathbb{N}$. Assume that the trajectory of e_i is $wC_1^JC_2C_3 \cdots C_i^\omega$ with $\{C_j, C_{j+1}\} = \{A, B\}$ for all $j < i$ (that is e_i has the maximum number of alternance in its subspace). Then the language of the convex hull of $\{e_0, \ldots, e_y\}$ is exactly $wC_1^iC_1^*C_2^*C_3^* \cdots C_{y-1}^*C_y^\omega \cup wC_1^iC_1^*C_2^*C_3^* \cdots C_{y-2}^*C_{y-1}^\omega$.*

Proof. We first consider the case $wC_1^iC_1^*C_2^*C_3^* \cdots C_{k-1}^*C_k^\omega$. Then, we consider the other case of $wC_1^iC_1^*C_2^*C_3^* \cdots C_{k-2}^*C_{k-1}^\omega$ in a second step.

Let x be a point in the interior of the convex hull of $e_1 \cdots e_z$. Then the trajectory of x is wC_1^iu for some infinite word u as all the point $e_1 \cdots e_z$ are of this type and by linearity of M^i for all i. Now, by lemma 7.21, the number of alternation after w is at most $z - 1$, hence the trajectory of x is of the form $wC_1^{i+i_1}C_2^{i_2}C_3^{i_3} \cdots C_{k-1}^{i_{k-1}} \cdots C_k^\omega$ with $i_j \in \mathbb{N}$ for all j. We will show that every of these trajectories is reached for a point in the convex hull of $e_1 \cdots e_z$.

Let $(i_j)_{j \leqslant k}$ be a family of integers. At first, we assume that $i_j \neq 0$ for all j. For all $j \in \{1, \ldots, z - 1\}$ let $x_j := N_{max} + i + j$. Also, for all $j \in \{1, \ldots, z - 1\}$, we define $y_j := N_{max} + i + i_1 + \ldots + i_j + 1/2$.

We will prove that there exists a point f in the interior of the convex hull of $e_1, \cdots, e_z$ such that $f(y_j) = 0$ for all $j \in \{1, \ldots, z-1\}$. Then Lemma 7.21 will imply that the language of f is $wC_1^{i+i_1}C_2^{i_2}C_3^{i_3} \cdots C_{k-1}^{i_{z-1}} \cdots C_z^\omega$.

We build f by induction. Applying lemma 7.29 for all $j \in \{1, \ldots, z-2\}$ to e_j, e_{j+1}, we obtain a point e_j^1 in (e_j, e_{j+1}) such that $e_j^1(y_{z-1}) = 0$. As e_j^1 is in (e_j, e_{j+1}), by linearity, the prefix of its trajectory is $wC_1^iC_2 \cdots C_{j-1}C_j$ (and it ends up with C_{j+1}^ω), which implies that it has additionally $j - 1$ zeros in $(N_{max} + i, N_{max} + i + j + 1)$, with $N_{max} + i + j + 1 \leqslant y_{z-1}$.

Thus, the sign of $e_j^1(x)$ is constant in $x \in [x_{j-1}+1, y_{z-1})$, depending on the parity of j. In particular, $y_{z-2} \in [x_{j-1} + 1, y_{z-1})$ for all $j \leqslant z - 2$.

We now consider points $(e_j^2)_{j\leqslant z-3}$ in the convex hull of $(e_j^1)_{j\leqslant z-2}$. Thus any of these points have $e_j^2(y_{z-1}) = 0$ by linearity. Let $j \in \{1, \dots, z-3\}$. We chose e_j^2 in the segment (e_j^1, e_{j+1}^1) such that $e_j^2(y_{z-2}) = 0$. It is possible as the sign of $e_j^1(y_{z-2}) > 0$ and the sign of $e_{j+1}^1(y_{z-2}) < 0$ (or vice versa, depending on the parity of j). We have that e_j^2 has $j+1$ zeros: y_{z-1}, y_{z-2} and one zero in every of $[x_k, x_{k+1})$ for all $k < j$.

By induction, we get $f := e_1^{z-1}$ such that $f(y_i) = 0$ for $1 \leqslant i \leqslant z-1$ and it switches sign between each zeros, hence its trajectory is $wC_1^{i+i_1}C_2^{i_2}\cdots C_{z-1}^{i_{z-1}}C_z^{\omega}$. Hence the case for $i_j > 0$ for all j is solved.

Consider now the case where some $i_j = 0$. First, if $i_1 = 0$, then the above procedure works. Now, for $i_j = 0$ for $j \neq 1$, it means that the desired trajectory is $wC_1^{i+i_1}C_2^{i_2}\cdots C_{j-1}^{i_{j-1}}\,C_{j+1}^{i_{j+1}}\cdots C_{z-1}^{i_{z-1}}C_z^{\omega} = wC_1^{i+i_1}C_2^{i_2}\cdots C_{j-2}^{i_{j-2}}C_{j-1}^{i_{j-1}+i_{j+1}}C_{j+2}^{i_{j+2}}\cdots C_{z-1}^{i_{z-1}}C_z^{\omega}$ as $C_{j-1} = C_{j+1}$, hence with 2 less switches. It suffices to start with the above procedure, but with $z' = z-2$ and points $e_1 \cdots e_{z'} = e_{z-2}$. For instance, take e_1, e_2. Their trajectories are respectively wC_1^{ω} and $wC_1^{i}C_2^{\omega}$. Applying lemma 7.29, we get the existence of a point f_1 in the convex hull of e_1, e_2 with a zero in $y_1 = N_{max}+i+i_1+1/2$. Its trajectory is $wC_1^{i+i_1}C_2^{\omega}$.

Last, for the case of $wC_1^{i}C_1^{*}C_2^{*}C_3^{*}\cdots C_{k-2}^{*}C_{k-1}^{\omega}$, it suffices to proceed in the same way in the convex hull of $(e_0, \dots, e_{y-1})$. □

Next, we note that the set $H_w\backslash F(J)$ may not be convex. However, one can partition $H_w\backslash F(J)$ into a finite number of convex polytopes. Now, let G be a convex polytope in $H_w\backslash F(J)$. We want to show that $Z(G) < Z(H_w) = Z(H) = y$. Indeed, else, one could apply Lemma 7.27 to $G_w = G$ and for some J' obtain $F(i,j)\cap G \neq \varnothing$ for any $j > J'$, which contradicts G being a convex set in $H_w\backslash F(J)$.

Hence one can compute the language of every G inductively, and each of them is regular. Finally, this leads to the regularity of $\mathcal{L}(H_w)$ by finite union, and to the regularity of $\mathcal{L}(H)$, and again by finite union to the regularity of $\mathcal{L}(\text{Init})$. This concludes our proof of the main regularity result, i.e., Theorem 7.16.

7.5 Discussion and conclusions

In this chapter, we have shown the following, summed up in table 7.1: if the eigenvalues of the transition (row-stochastic) matrix associated with the MC are distinct roots of real numbers, then any trajectory from a given initial distribution is ultimately periodic. This is tight, in the sense that, there are examples of trajectories which are not ultimately periodic even for MCs with 3 states [AAGT15, Tur68] (with some eigenvalue not root of any real number). Further, the eigenvalues are distinct positive real numbers, then the language generated by a MC starting from a convex polytope of initial distributions is *effectively regular*. Surprisingly, this result is also tight: there exist MCs with eigenvalues being distinct roots of real numbers (starting from a convex initial set) which generate a non-regular language.

Theorem 7.28. *[AGKV16] There exists a MC $\mathcal{A}_1$ with eigenvalues which are roots of real values and 7 states such that $\mathcal{L}(\text{Init}, \mathcal{A}_1)$ is not regular.*

We proved that if the eigenvalues of the transition matrix associated with the Markov chain are all distinct positive real numbers and we know these values, then the language, for any convex polytope of initial distributions, is effectively regular. We proved that by building its language of trajectories.

Notice that in general, the eigenvalues of a Markov chain can only be approximated. However, in case these eigenvalues are rational, then one can use the rational root theorem (see, for example, [Lan13]) in order to find them explicitly. This also provides a test whether all the eigenvalues are rational, and if yes, whether they are all positive numbers.

Hence, if the Markov chain of the reduction from a Skolem problem to a Markov reachability problem have distinct positive real eigenvalues and they are known or its eigenvalues are distinct positive rational values, then we can decide the original Skolem problem.

Though Markov Chains are a simple formalism, there are still many basic problems, whose decidability is open and thought to be very hard. Indeed, it is surprising yet significant that even after assuming strong hypotheses, their behaviours cannot be described easily.

Property of eigenvalues of MC	Regular language	Ultimately periodic trajectories
Distinct, positive real numbers	✓ (Thm.7.16)	✓ (from below)
Distinct, roots of real numbers	× [AGKV16]	✓ (Prop.7.10)
Distinct	× (from above)	× ([AAGT15], Thm.3)

Table 7.1 – A summary of the results in this chapter.

Besides imposing strong restriction as positive eigenvalues, another way to tackle the problem is to approximate it, asking whether for all ϵ there exists a number of steps n_ϵ after which the probability to be in *Goal* is at least $\gamma - \epsilon$. The decidability and precise complexity of this problem has been explored in [CKV14]. A more general approximation scheme, valid for much more general questions which can be expressed in some LTL logic, has also been tackled by generating a regular language of *approximated* behaviors [AAGT15], where the authors define a notion of an ϵ-approximation of a disitribution ξ, such that ξ_ϵ is an ϵ-approximation of ξ iff

ξ_ϵ and ξ are in the same class until some n_ϵ, that depends on ϵ, and after that both distributions are in the same set of *final classes*, a set of configurations where the configurations cycle in the steady state phase.

We now explain the relationship between checking population questions on MC and MDP and checking reachability for stochastic systems with imperfect information. In some sense, checking population questions is harder than checking reachability for systems with full observation (as this is decidable), but it is simpler than reachability with imperfect information.

Hence, finding strategies ensuring quantified reachability in MDPs with imperfect information (that is in POMDP, i.e. partially observable MDPs) is harder than solving population problems for MDPs (because population questions on MDPs corresponds to the particular imperfect information case of PAs, that there is no information). In turn, this is harder than the case where the (PO)MDP is unary (that is it a Markov chain that there is no choice of action), and in this case quantified reachability in unary POMDPs and population questions on Markov Chains is the same problem.

A *Probabilistic Automaton* (PA) can be defined as a MDP such that all actions are available in each state and the player do not know in which of these states he is.

Unary PAs [CKV14, Tur68], have an alphabet with a single letter. That is, there is a unique strategy, and the model is essentially a Markov chain.

Population questions on MDP, with uniform strategy per time point correspond to reachability in PA. Assume that there exists a number n of steps such that there is at least γ of the population in Goal after n steps of Markov Chain. Then playing n steps of the associated unary PA, there is probability at least γ to reach *Goal*. Reciprocally, a wining strategy of a unary PA translates to a number of steps after which at least γ of the population is in *Goal*. Hence reachability for unary PA is open (Skolem complete).

For PAs, the problem of whether there is a strategy to reach *Goal* with probability at least a threshold γ (also called a cut-point) is already undecidable [Ber74]. Even approximating this probability has been shown undecidable in PAs [MHC03]. In fact, deciding whether there exists a sequence of strategies with probability arbitrarily close to $\gamma = 1$ is already undecidable [GO10], and only very restricted subclasses are known to ensure decidability [FGO12, CT12].

List of Figures

4.2 A two-player stochastic arena . . . 37
4.4 Transition probabilities in $\mathcal{A}$ with strategies σ and τ. . . . 41
4.9 Game with states $\mathbf{S} = \{1, 2, 3\}$ and actions $\mathbf{A}$as defined above. . . . 47
4.16 Game with infinite set of actions where player Max does not have memoryless optimal strategy. . . . 63
4.17 Game with infinite set of states where player Max does not have optimal strategy. . . . 65

5.4 The upper figure: The event $\{S_{T_m+1} = i, T_m < \infty\}$ consists of the plays that at time T_m visit state k for the mth time without ever visiting the states $> k$ before, and at time $T_m + 1$ they visit state i, where $i < k$. These plays are partitioned into two sets. The set $\{T_{m+1} < \infty, S_{T_m+1} = i, T_m < \infty\}$ of plays that will visit k for the $(m + 1)$th time and the set $\{T_{m+1} = \infty, S_{T_m+1} = i, T_m < \infty\}$ of the plays for which the mth visit in k was the last one. The lower figure : The shift mapping θ_{T_m+1} "forgets" all the history prior to the time $T_m + 1$. . . . 78

7.1 Markov chain with three states. . . . 94
7.9 The solid line represents $\sum_{i=0}^{k} a_i(\delta)p_i^n$ and the circles are above the graph if $\rho_{\delta_0}(n) = A$ and below if $\rho_{\delta_0}(n) = B$. . . . 100
7.11 Büchi automaton for the ultimate positivity problem. . . . 101
7.17 Breaking into convex polytopes with constant signs . . . 104
7.20 Function of type 2 $f(x) = 40 \times 0.8^x - 380 \times 0.5^x + 390 \times 0.4^x (p_0 = 0.8, p_1 = 0.5, p_2 = 0.4)$ . . . 106

Bibliography

[AAGT15] M. Agrawal, S. Akshay, B. Genest, and P. S. Thiagarajan. Approximate verification of the symbolic dynamics of Markov chains. *Journal of the ACM*, 62(1):2:1–2:34, 2015.

[AAOW15] S. Akshay, T. Antonopoulos, J. Ouaknine, and J. Worrell. Reachability problems for Markov chains. *Information Processing Letters*, 115(2):155–158, 2015.

[AGKV16] S. Akshay, B. Genest, B. Karelović, and N. Vyas. On Regularity of Unary Probabilistic Automata. In *33rd Symposium on Theoretical Aspects of Computer Science*, volume 47 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 8:1–8:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016.

[AN01] A. Arnold and D. Niwiński. *Rudiments of μ-calculus*, volume 146 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, 2001.

[Ber74] A. Bertoni. The solution of problems relative to probabilistic automata in the frame of the formal languages theory. In *GI Jahrestagung*, pages 107–112, 1974.

[BK76] T. Bewley and E. Kohlberg. The asymptotic theory of stochastic games. *Mathematics of Operations Research*, 1:197–208, 1976.

[BRS02] D. Beauquier, A. Rabinovich, and A. Slissenko. A logic of probability with decidable model-checking. In *International Workshop on Computer Science Logic*, volume 2471, pages 306–321. Springer, 2002.

[CdAH13] K. Chatterjee, L. de Alfaro, and T. A. Henzinger. Strategy improvement for concurrent reachability and turn-based stochastic safety games. *Journal of Computer and System Sciences*, 79(5):640–657, 2013.

[CJH04] K. Chatterejee, M. Jurdziński, and T.A. Henzinger. Quantitative stochastic parity games. In *Proceedings of the 15th Annual Symposium on Discrete Algorithms*, pages 114–123, 2004.

[CJK+16] C. S. Calude, S. Jain, B. Khoussainov, W. Li, and F. Stephan. Deciding parity games in quasipolynomial time, Technical report, CDMTCS, 2016.

[CKV+11] R. Chadha, V. Korthikanti, M. Vishwanathan, G. Agha, and Y. Kwon. Model checking MDPs with a unique compact invariant set of distributions. In *2011 Eighth International Conference on Quantitative Evaluation of SysTems*, pages 121–130, 2011.

[CKV14] R. Chadha, D. Kini, and M. Viswanathan. Decidable problems for unary PFAs. In *International Conference on Quantitative Evaluation of Systems*, volume 8657, pages 329–344. Springer, 2014.

[Con92] A. Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203–224, 1992.

[CT80] M. G. Crandall and L. Tartar. Some relations between nonexpansive and order preserving mappings. *Proceedings of the American Mathematical Society*, 78(3):385–390, 1980.

[CT12] K. Chatterjee and M. Tracol. Decidable problems for probabilistic automata on infinite words. In *Logic in Computer Science (LICS), 2012 27th Annual IEEE Symposium on*, pages 185–194. IEEE, 2012.

[dA97] L. de Alfaro. *Formal Verification of Probabilistic Systems*. PhD thesis, Stanford University, 1997.

[dAH00] L. de Alfaro and T.A. Henzinger. Concurrent omega-regular games. In *Logic in Computer Science. Proceedings. 15th Annual IEEE Symposium on*, pages 141–154. IEEE, 2000.

[dAHK07] L. de Alfaro, T.A. Henzinger, and O. Kupferman. Concurrent reachability games. *Theoretical Computer Science*, 386:188–217, 2007.

[dAM04] L. de Alfaro and R. Majumdar. Quantitative solution to omega-regular games. *Journal of Computer and System Sciences*, 68:374–397, 2004.

[EJ91] E.A. Emerson and C. Jutla. Tree automata, μ-calculus and determinacy. In *Foundations of Computer Science. Proceedings., 32nd Annual Symposium on*, pages 368–377. IEEE, 1991.

[Eve57] H. Everett. Recursive games. *Contributions to the Theory of Games*, 3(39):47–78, 1957.

[FGO12] N. Fijalkow, H. Gimbert, and Y. Ouahlad. Deciding the value 1 problem for probabilistic leaktight automata. In *Logic in Computer Science (LICS), 27th Annual IEEE Symposium on*, volume 11, pages 295–304. IEEE, 2012.

[FV97] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.

[GH82] Y. Gurevich and L. Harrington. Trees, automata and games. In *Proceedings of the fourteenth annual ACM symposium on Theory of computing*, pages 60–65. ACM, 1982.

[GO10] H. Gimbert and Y. Oualhadj. Probabilistic automata on finite words: Decidable and undecidable problems. In *International Colloquium on Automata, Languages, and Programming*, volume 6199, pages 527–538. Springer, 2010.

[GTW02] E. Grädel, W. Thomas, and T. Wilke, editors. *Automata, Logics, and Infinite Games*, volume 2500 of *LNCS*. Springer, 2002.

[HHH06] V. Halava, T. Harju, and M. Hirvensalo. Positivity of second order linear recurrent sequences. *Discrete Applied Mathematics*, 154(3), 2006.

[HK66] A. J. Hoffman and R. M. Karp. On nonterminating stochastic games. *Management Science*, 12(5):359–370, 1966.

[JFM15] S. Zamir J-F. Mertens, S. Sorin. *Repeated games.* Cambridge University Press, 2015.

[KS81] P. R. Kumar and T. H. Shiau. Existence of value and randomized strategies in zero-sum discrete-time stochastic dynamic games. *SIAM Journal on Control and Optimization*, 19(5):617–634, 1981.

[KVAK10] V. A. Korthikanti, M. Viswanathan, G. Agha, and Y. Kwon. Reasoning about MDPs as transformers of probability distributions. In *Quantitative Evaluation of Systems, Seventh International Conference on the*, pages 199–208. IEEE, 2010.

[KZ15] B. Karelović and W. Zielonka. Nearest fixed points and concurrent priority games. In *International Symposium on Fundamentals of Computation Theory*, volume 9210, pages 381–393. Springer, 2015.

[Lan13] S. Lang. *Undergraduate analysis.* Springer Science & Business Media, 2013.

[LPW09] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times.* American Mathematical Soc., 2009.

[Mar98] D. A. Martin. The determinacy of Blackwell games. *Journal of Symbolic Logic*, 63(4):1565–1581, 1998.

[MHC03] O. Madani, S. Hanks, and A. Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artificial Intelligence*, 147(1-2):5–34, 2003.

[MM02] A. K. McIver and C. C. Morgan. Games, probability, and the quantitative μ-calculus q$M\mu$. In *International Conference on Logic for Programming Artificial Intelligence and Reasoning*, volume 2514, pages 292–310. Springer, 2002.

[MS96] A. P. Maitra and W. D. Sudderth. *Discrete Gambling and Stochastic Games*. Springer, 1996.

[MS04] A. Maitra and W. Sudderth. Stochastic games with Borel payoffs. In A. Neyman and S. Sorin, editors, *Stochastic Games and Applications*, volume 570 of *NATO Science Series C, Mathematical and Physical Sciences*, pages 367–373. Kluwer Academic Publishers, 2004.

[MTC+14] L. Maruthi, I. Tkachev, A. Carta, E. Cinquemani, P. Hersen, G. Batt., and A. Abate. Towards real-time control of gene expression at the single cell level: a stochastic control approach. In *International Conference on Computational Methods in Systems Biology*, volume 8859, pages 155–172. Springer, 2014.

[Ney03] A. Neyman. Stochastic games and nonexpansive maps. In A. Neyman and S. Sorin, editors, *Stochastic Games and Applications*, volume 570 of *NATO Science Series C, Mathematical and Physical Sciences*, pages 397–415. Kluwer Academic Publishers, 2003.

[NS04] A. Neyman and S. Sorin, editors. *Stochastic Games and Applications*, volume 570 of *NATO Science Series C, Mathematical and Physical Sciences*. Kluwer Academic Publishers, 2004.

[OW12] J. Ouaknine and J. Worrell. Decision problems for linear recurrence sequences. In *International Workshop on Reachability Problems*, volume 7550, pages 21–28. Springer, 2012.

[OW14a] J. Ouaknine and J. Worrell. On the Positivity Problem for simple linear recurrence sequences. In *International Colloquium on Automata, Languages, and Programming*, volume 8573, pages 318–329. Springer, 2014.

[OW14b] J. Ouaknine and J. Worrell. Positivity problems for low-order linear recurrence sequences. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 366–379. Society for Industrial and Applied Mathematics, 2014.

[OW14c] J. Ouaknine and J. Worrell. Ultimate Positivity is decidable for simple linear recurrence sequences. In *International Colloquium on Automata, Languages, and Programming*, volume 8573, pages 330–341. Springer, 2014.

[Paz71] A. Paz. *Introduction to Probabilistic Automata*. Academic Press, 1971.

[Sec98] P. Secchi. On the existence of good stationary strategies for nonleavable stochastic games. *International Journal of Game Theory*, 27(1):61–81, 1998.

[Sha53] L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953.

[Sor02] S. Sorin. *A First Course on Zero-Sum Repeated Games*, volume 37. Springer, 2002.

[Tar55] A. Tarski. A lattice-theoretical fixpoint theroem and its aplications. *Pacific journal of Mathematics*, 5:285–309, 1955.

[Tur68] P. Turakainen. On stochastic languages. *Information and Control*, 12(4):304–313, 1968.

[Wal02] I. Walukiewicz. Monadic second-order logic on tree-like structures. *Theoretical Computer Science*, 275(1-2):311–346, 2002.

[Zie98] W. Zielonka. Infinite games on finitely coloured graphs with applications to automata on infinite trees. *Theoretical Computer Science*, 200:135–183, 1998.