

**Balog Imre**

**Some problems in discounted stochastic games**

**Doctoral School of Economics, Business and Informatics**

Supervisors:

**Ágoston Kolos Csaba Ph.D.**

**Pintér Miklós Ph.D.**

©Balog Imre

**Corvinus University of Budapest**

**Doctoral School of Economics, Business and Informatics**

**Some problems in discounted stochastic games**

Ph.D. Thesis

**Balog Imre**

Budapest, 2025



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The classification of stochastic games . . . . .	3
1.2	Literature review . . . . .	4
1.3	Reward functions and discounting methods . . . . .	7
1.4	Relevance of the research topic . . . . .	10
1.5	Research methods . . . . .	10
1.6	Research questions . . . . .	11
1.6.1	Finite stochastic games with generalised discounting . . . . .	12
1.6.2	Zero-sum stochastic games with separable discounting . . . . .	13
1.6.3	Discounted finitely additive Markov decision processes . . . . .	14
1.7	Structure of this thesis . . . . .	14
<b>2</b>	<b>Finite stochastic games with generalised discounting</b>	<b>16</b>
2.1	The game model . . . . .	17
2.2	Histories, plays and strategies . . . . .	20
2.3	Reward functions based on discounting . . . . .	22
2.4	Some notes on the research questions . . . . .	24
2.5	Continuous generalised games . . . . .	26
2.6	Results for finite stochastic games with generalised discounting . . . . .	29
2.7	Some comments on continuous generalised games . . . . .	34
2.7.1	Mixed extension . . . . .	34
2.7.2	The Big Match and the Paris Match . . . . .	37
<b>3</b>	<b>Zero-sum stochastic games with separable discounting</b>	<b>40</b>
3.1	Zero-sum countable stochastic games with separable discounting . . . . .	42
3.1.1	The game model . . . . .	42
3.1.2	Zero-sum countable stochastic games with separable discounting . . . . .	44

3.1.3	Supergames of zero-sum countable stochastic games with separable discounting . . . . .	49
3.1.4	Shapley operators for supergames . . . . .	55
3.1.5	Results for zero-sum countable stochastic games with separable discounting . . . . .	58
3.2	Zero-sum infinite stochastic games with separable discounting . . .	60
3.2.1	The game model . . . . .	60
3.2.2	Zero-sum infinite stochastic games with separable discounting . . . . .	64
3.2.3	Zero-sum Borel, Suslin, and Nowak stochastic games . . . . .	65
3.2.4	Results for zero-sum infinite stochastic games with separable discounting . . . . .	69
<b>4</b>	<b>Discounted finitely additive Markov decision processes</b>	<b>70</b>
4.1	The game model . . . . .	71
4.2	Discounted finitely additive MDPs . . . . .	74
4.3	Finitely additive MDPs with ripple discounting . . . . .	77
4.3.1	R-superprocesses . . . . .	77
4.3.2	Results for finitely additive MDPs with ripple discounting . . . . .	84
4.4	Finitely additive MDPs with separable discounting . . . . .	86
4.4.1	S-superprocesses . . . . .	86
4.4.2	Results for finitely additive MDPs with separable discounting . . . . .	90
<b>5</b>	<b>Conclusion</b>	<b>92</b>
5.1	Results . . . . .	93
<b>A</b>	<b>Mathematical background for Chapter 4</b>	<b>96</b>
A.1	Preliminaries . . . . .	96
A.2	The gambling problem . . . . .	97
A.3	The Dubins-Savage-integrals . . . . .	98
A.4	The Dubins-Savage-Sudderth-integral . . . . .	100
A.5	Application of the Dubins-Savage-integral in finitely additive Markov decision processes . . . . .	101

# List of Figures

1	Graphical representation of the finite stochastic game in Example 2.2. . . . .	19
2	Graphical representation of the two-person finite stochastic game with two states in Example 2.25. . . . .	33
3	Some examples of games that belong to the class of continuous generalised games, along with examples that do not do so. . . . .	34
4	Graphical representation of the zero-sum finite stochastic game with three states in Example 2.33. . . . .	38
5	Graphical representation of the two-person finite stochastic game with three states in Example 2.34. . . . .	38
6	Graphical representation of the zero-sum finite stochastic game with a single state in Example 3.7. . . . .	46
7	Graphical representation of the zero-sum finite stochastic game with a single state and arbitrary positive one-stage payoffs in Example 3.8. . . . .	48
8	Graphical representation of constructing a supergame. . . . .	51
9	Graphical representation of constructing an R-superprocess. . . . .	81
10	Graphical representation of the Markov decision process with three states in Example 4.19. . . . .	85
11	Graphical representation of constructing an S-superprocess. . . . .	88
12	Graphical representation of the Markov decision process with a single state in Example 4.27. . . . .	91

## List of Tables

1	Summary of the problems addressed and the approaches employed in each chapter. . . . .	12
2	Summary of our results for finite stochastic games with generalised discounting. . . . .	93
3	Summary of our results for zero-sum stochastic games with separable discounting. . . . .	94
4	Summary of our results for discounted finitely additive Markov decision processes. . . . .	95

# Acknowledgement

I would like to express my deepest gratitude to my supervisors, Miklós Pintér and Kolos Ágoston, for their invaluable guidance, generous investment of time, and unwavering support throughout my doctoral studies.

I also want to thank my opponents, Péter Bayer and Péter Vida, for thoroughly reviewing my thesis proposal. Their insightful feedback helped identify inaccuracies and provided valuable suggestions that significantly improved my work.

# Chapter 1

## Introduction

*"The game is afoot."*

---

SHERLOCK HOLMES  
in *The Adventure of the Abbey Grange*  
by Sir Arthur Conan Doyle

*Stochastic games* (also known as *Markov games*) is a powerful mathematical framework for analysing decision-making in dynamic environments. Since the theory of stochastic games is a broad branch of mathematics, we focus on *discrete-time* stochastic games with a *finite number of players*.

In this thesis, we define stochastic games by the following elements:

- (E1) *Players*: a nonempty finite set of decision makers interacting repeatedly over time.
- (E2) *States*: a nonempty set of situations in which the stochastic game can be. The number of states depends on the specific game model. To specify a given stochastic game, we must also specify the *initial state*, which serves as the starting point of the game.
- (E3) *Actions*: In each state, players have a nonempty set of available actions, respectively. The number of actions may vary across states and models.
- (E4) *Transition rule*: a mechanism that controls how the game moves from one state to another, based on the action chosen by the players.
- (E5) *Payoff functions*: each player has a mapping that assigns an immediate payoff (or cost) to each pair of state and action profiles.

We often use the term *state space* for the set of all states. Similarly, a player's *action space* means the set of all available actions for that player in each state of the stochastic game.

It is also important to highlight that this thesis frequently employs the following general principle (Parthasarathy and Babu, 2020; Solan, 2022). While specifying the *initial state* is necessary to define a stochastic game, it is often more practical to consider a family of stochastic games that differ only in their initial states.

There are various models of stochastic games, differing in their structure. In this thesis, we focus on the following framework. At the beginning of each stage, every player knows the full history of the game, including all past states and the actions taken by each player, and can observe the current state. Next, they choose their actions *simultaneously* and *independently*. The current state and the selected action profile jointly determine two outcomes: the immediate payoffs that each player receives and the probability distribution governing the transition to the next state. In the next state, everything starts over according to the previous rules. Depending on the specific structure of the stochastic game, this dynamic may evolve over either a *finite* or an *infinite time horizon*.

A *Markov decision process (MDP)* is a special stochastic game with a single decision-maker. Markov decision processes extend the concept of a *Markov chain*. While a Markov chain describes the probabilistic evolution of states over time, a Markov decision process generalises this framework by introducing the notions of *actions* and *payoffs* (see, e.g., Neyman (2003A)).

Stochastic games generalise Markov decision processes and *repeated games*, providing a unified framework for studying decision-making in dynamic multi-agent systems. Repeated games capture strategic interactions that unfold over time. Still, they assume a fixed game structure without state transitions - in other words, the players repeatedly engage in the same game across multiple stages. Thus, a repeated game is a stochastic game with a single state. In contrast, stochastic games use a state-dependent framework in which the environment evolves according to the current state and the players' past actions. This coupling between present decisions and future consequences adds significant complexity, as players must consider not only the immediate payoffs of their actions but also how those actions influence future states and the behaviour of other players.

## 1.1 The classification of stochastic games

This section presents a *selective classification* of stochastic games. The main purpose of this classification is to ensure consistent terminology throughout this thesis. It is important to clarify that our classification focuses on the aspects and classes most relevant to the scope of this thesis.

First, we consider *the number of players* in Element (E1). Suppose a stochastic game involves  $n$  players, where  $n$  is a positive integer. In that case, we refer to it as a stochastic game or, to emphasise the number of participants, as an  *$n$ -person stochastic game*. Such games are studied in Chapter 2. In contrast, Chapter 4 examines *Markov decision processes*. Between these two possibilities lies the class of *two-person stochastic games*, which form the central focus of Chapter 3.

As a point of interest, we note that stochastic games have also been analysed in settings with different fixed numbers of players (for example, Solan (1999) examined a three-person stochastic game). Moreover, stochastic games with infinitely many players have been investigated as well (see, for instance, Subir (2003), Doncel, Gast, and Gaujal (2016), or Sanjari and Yüksel (2021)). However, this line of research lies beyond the scope of the present thesis.

Next, we categorise stochastic games based on the *state and action spaces* in Elements (E2) and (E3). This thesis characterises a *finite stochastic game* by the following two conditions: (1) the state space is a nonempty finite set, and (2) in each state, every player has only finitely many available actions. Chapter 2 is devoted to studying ( $n$ -person) finite stochastic games.

In the first part of Chapter 3, we consider *two-person countable stochastic games*, where either the state space or the action spaces (or both) are countable. The second part of Chapter 3 extends the analysis to more general two-person stochastic games, which we shall, for simplicity, refer to as *stochastic games with a general state space*. Finally, Chapter 4 presents the most general framework, focusing only on the state and action spaces. Here, we study finitely additive Markov decision processes in which the state and action space are arbitrary nonempty sets.

As another classification point, we consider the dynamics of stochastic games. In addition to distinguishing between *discrete* and *continuous time*, it is also important to specify the duration of the stochastic game. A *finite-horizon stochastic game* ends after a fixed number of stages, while an *infinite-horizon stochastic game* never ends. In this thesis, it is important to mention that we restrict our attention to infinite-horizon stochastic games within a discrete-time framework.

This thesis examines only a specific class of stochastic games, namely those in

which players always act independently and simultaneously. We note that alternative models also exist, but these stochastic games are not discussed here (see, for example, *robust Markov decision processes* with a player in the role of *controller* and another in the role of *opponent* Jaśkiewicz and Nowak (2018, Section 3)). In addition, this thesis assumes that each player can observe the current state and is fully aware of all states and actions from the preceding stages of the stochastic game.

The transition rule in Element (E4) determines the dynamics of a stochastic game. The specific mathematical properties imposed on the transition rule result in different classes of models. In Chapters 2 and 3, we consider the standard *countable additivity framework*, which assumes that probability measures are  $\sigma$ -additive. Working with  $\sigma$ -additive measures is most commonly used in non-cooperative game theory. In contrast, Chapter 4 applies an alternative approach by relaxing this requirement. In this chapter, we consider the *finitely additive framework*, where probability measures are only required to be finitely additive. While it is still possible for finitely additive measures to be  $\sigma$ -additive in specific cases, this is not a standard assumption in Chapter 4.

Stochastic games can also be classified based on the characteristics of the one-stage payoff functions (see Element (E5)). A two-person stochastic game is called *zero-sum* if the sum of the one-stage payoffs is zero at every state and stage of play. A natural relaxation of the zero-sum stochastic games leads to *constant-sum stochastic games*, where the sum of the one-stage payoffs remains fixed across all states and stages. In contrast, *general-sum stochastic games* allow the sum of one-stage payoffs to vary depending on the game's current state and the chosen action profile.

Finally, a Markov decision process is said to be *positive* if the one-stage payoff function of Player 1 is non-negative for all states and actions, and *negative* if the one-stage payoff function is non-positive. Analogously, a two-person zero-sum stochastic game is called *positive* if Player 1's one-stage payoff function is non-negative across all states and actions.

## 1.2 Literature review

This section briefly summarises the literature on stochastic games. Our goal is not to provide a comprehensive review; instead, we focus on the most relevant works and highlight the key points significant to our discussion.

Modern game theory traces its origins to the introduction of mixed-strategy equilibria in finite zero-sum games (v. Neumann, 1928). In 1950, Nash (1950)

demonstrated that every finite  $n$ -player non-cooperative game possesses at least one Nash equilibrium in mixed strategies.

Stochastic games were introduced by Shapley (1953), who analysed zero-sum finite stochastic games. Shapley (1953) assumed that each play ends almost surely after finitely many stages, and proved that every zero-sum finite stochastic game with *exponential discounting* admits a *value*, with both players having optimal stationary strategies. This foundational result underpins the research in Chapter 3, where we address a related problem but replace *exponential discounting* with *separable discounting*.

Various models have been developed as generalisations of Shapley (1953)'s result on zero-sum stochastic games, of which we mention a few. For instance, Maitra and Parthasarathy (1970, 1971) studied zero-sum stochastic games with exponential discounting, assuming that the state and action spaces are compact metric spaces, and that both the one-stage payoff function and the transition rule satisfy suitable continuity conditions. Further results were obtained by Himmelberg, Parthasarathy, Raghavan and Van Vleck (1976); Couwenbergh (1980); Nowak (1985, 2003) on zero-sum stochastic games with exponential discounting, where the state space is a standard Borel space and the action spaces are compact metric spaces. The second part of Chapter 3 builds extensively on the work of Nowak (1984A), which treats zero-sum stochastic games with exponential discounting in an even more general framework.

In the study of zero-sum stochastic games, it is interesting to consider the subclass of positive zero-sum stochastic games. In this framework, players evaluate the so-called *total reward* rather than the discounted reward. Although several other reward functions can be defined, in Chapter 3 we restrict our analysis to the case of total reward.

The literature on positive zero-sum stochastic games with total reward is comprehensive (Frid, 1974; Maitra and Parthasarathy, 1971; Nowak, 1985, 2003). We highlight only two key results extensively utilised in Chapter 3. First, our solution for zero-sum countable stochastic games with separable discounting, presented in the first part of Chapter 3, is based on the work of Flesch, Predtetchinski and Sudderth (2018, 2020). Second, we rely on the findings of Nowak (1984B) for cases involving a general state space, which underpin the latter part of Chapter 3.

Chapter 2 investigates ( $n$ -person) finite stochastic games with *generalised discounting*. Our research extends the classical results of Fink (1964); Takahashi (1964); Sobel (1971), who established that every finite stochastic game with *exponential discounting* admits a stationary Nash equilibrium.

While Chapter 2 restricts our focus to finite stochastic games, it is worth noting the existence of broader results. For instance, Takahashi (1964) showed, using exponential discounting, that a stationary equilibrium exists in stochastic games with finite state spaces and compact action sets under suitable continuity conditions on the one-stage payoff functions and transition probabilities. More generally, for stochastic games with exponential discounting, arbitrary state spaces and compact action sets, the existence of an equilibrium has been established under suitable continuity assumptions on both the one-stage payoff functions and the transition rules (Solan, 1998; Mertens and Parthasarathy, 2003). In contrast, Levy and McLennan (2015) provided counterexamples of stochastic games with exponential discounting, general state and action spaces where no measurable equilibrium exists.

Markov decision processes, a special class of stochastic games, were first examined by Blackwell (1962). There are numerous results on Markov decision processes, depending on the assumptions applied on Elements (E2), (E3), (E4), and (E5) (see, for example, Puterman (1994); Feinberg and Shwartz (2002)).

Chapter 4 focuses on a discounted Markov decision process in the finitely additive framework. The starting point of our research is the results of Sudderth (2016, Section 3) on finitely additive Markov decision processes with *exponential discounting*. In Chapter 4, we examine both *ripple discounting* and *separable discounting*. We use the results of Sudderth (2016, Section 7) on negative finitely additive Markov decision processes to achieve our results.

Chapter 4 assumes the finitely additive framework, since, at the time of the research, results on *ripple discounting* could only be established for finitely additive Markov decision processes. The study of other stochastic games with ripple discounting, for instance, zero-sum games, could be an interesting direction for future research, but it is beyond the scope of this thesis.

There are several works for both supporting and challenging the use of *countably additive framework* and *finitely additive framework* (de Finetti, 2017; Savage, 1954). Bingham (2010) provided a comprehensive historical comparison of these two frameworks.

While not aiming for completeness, we mention some works that assume the finitely additive framework. Under certain conditions on the payoff functions, Marinacci (1997) showed that a normal form game admits a Nash equilibrium in the finitely additive framework. Flesch, Vermeulen and Zseleva (2021) extended this result by introducing the concept of the *legitimate equilibrium*. Similarly, Milchtaich (2023) proposed the notion of the *best-response equilibrium* and demon-

strated that, under suitable conditions, such an equilibrium exists in normal form games. In the case of stochastic games in the finitely additive framework, we only highlight the following result: Maitra and Sudderth (1998, Theorem 1.1, p. 258) showed that every zero-sum stochastic game with general state and action spaces admits a value if the reward function is a bounded Borel measurable function.

In this thesis, we focus on stochastic games that mainly use discounting. It is important to note that there exist other reward functions, as well (see, for instance, Puterman (1994); Feinberg and Shwartz (2002) for Markov decision processes and Parthasarathy and Babu (2020); Solan (2022) for stochastic games). Such reward functions not only expand the range of problems that stochastic games can practically address, but they also support theoretical research. A notable example is the *Big Match* introduced by Gillette (1957), which played a key role in developing the *uniform equilibrium concept*. Blackwell and Ferguson (1968) solved the problem of the Big Match, and later, Mertens and Neyman (1981) demonstrated that every two-person zero-sum stochastic game has a uniform value (for more detail, see, for example, Solan (2022, Chapter 9)).

### 1.3 Reward functions and discounting methods

In this section, we briefly present the discounting methods of this thesis through Example 1.1.

Players often face a dilemma between short-term gains and long-term benefits in stochastic games (Solan and Vieille, 2015). In the current stage, the one-stage payoffs may be tempting for the players. However, reaching these payoffs may lead to unfavourable outcomes in the future. Therefore, players must consider how their actions impact state transitions and influence future one-stage payoffs.

Reward functions are powerful for controlling this trade-off between short-term and long-term gains. In this thesis, we focus primarily on the discounted reward criteria, which assess the long-term values of one-stage payoffs by progressively reducing the significance of future payoffs. In cases of separable discounting, this principle may not necessarily hold (see, for example, Formula (SepDisc)).

In the following, we use the *Consumption–Saving Problem* (Samuelson, 1969) as a guiding example to introduce our discounting schemes.

*Example 1.1.* Consider a player, or decision-maker, with an initial wealth  $w_1 > 0$ . In the first stage, the player chooses a consumption level  $a_1 \in [0, w_1]$  and saves the amount  $w_1 - a_1 \geq 0$  for the next stage. From the consumption, the player receives the utility  $u(a_1)$ . At the end of the first stage, the player gets an income  $(y_1)$ , a non-negative random

variable.

At the beginning of the next stage, the player has the following wealth level:

$$w_2 = w_1 - a_1 + y_1.$$

In the second stage, the player selects an amount  $a_2 \in [0, w_2]$  to consume, saves the amount  $w_2 - a_2 \geq 0$ , and receives a new income  $y_2$ . This decision-making process persists over an infinite horizon.

The goal of the player is to maximise the expected value of the total exponential discounted utility given by

$$\sum_{t=1}^{\infty} \alpha^t u(a_t), \quad (\text{ExpDisc})$$

where  $\alpha \in [0, 1)$  is a fixed discount factor.

We refer to the discounting technique in Formula (ExpDisc) as *exponential discounting*. This method is a well-known discounting technique in stochastic games. For example, it appears in two-person finite zero-sum stochastic games (Shapley, 1953), finite stochastic games (Fink, 1964; Takahashi, 1964; Sobel, 1971), two-person stochastic games with general state space (Nowak, 1984A), and finitely additive Markov decision processes (Sudderth, 2016).

*Classical discounting* is one of the possible generalisations of the exponential discounting method. This discounting technique allows the discount rate to vary dynamically based on the current state and actions.

For instance, consider Example 1.1, and interpret the wealth level as a state. If the player applies classical discounting, the goal of the player becomes maximising the expected value of the *total classically discounted utility*:

$$\sum_{t=1}^{\infty} \left( \prod_{m=1}^t \beta(w_m, a_m) \right) u(a_t), \quad (\text{ClassDisc})$$

where  $\beta$  is a state- and action-dependent discount function such that  $\beta(w, a) \in [0, 1)$  for each pair of wealth level  $w$  and consumption level  $a$ .

This thesis considers three discounting methods that have not been extensively studied for stochastic games.

The first method assumes that the discount rate depends on the current state and may also vary over time. We refer to this method as *generalised discounting*.

For example, suppose that the player applies generalised discounting in Example 1.1. Then, the player aims to maximise the expected value of the *total generalised discounted utility*:

$$\sum_{t=1}^{\infty} \left( \prod_{m=1}^t \lambda(m, w_m) \right) u(a_t). \quad (\text{GenDisc})$$

In Formula (GenDisc),  $\lambda$  is a *generalised discount function* that depends on both time and state. Chapter 2 focuses on finite stochastic games with *generalised discounting*.

When the discount factor depends on time, state, and actions, we introduce two further discounting methods. *Ripple discounting* is a technique in which the past discount factors influence the current discount factor.

Specifically, in Example 1.1, if the player applies ripple discounting, then the player aims to maximise the expected value of the *total ripple discounted utility*:

$$\sum_{t=1}^{\infty} \left( \prod_{m=1}^t \Lambda(m, w_m, a_m) \right) u(a_t). \quad (\text{RippDisc})$$

In Formula (RippDisc),  $\Lambda$  is a *ripple discount function*. Chapter 4 focuses on finitely additive Markov decision processes with *ripple discounting*.

In contrast, when the current discount factor does not depend on the previous discount factors, we refer to this technique as *separable discounting*.

Returning to Example 1.1, with separable discounting, the player seeks to maximise the expected value of the *total separable discounted utility*:

$$\sum_{t=1}^{\infty} \delta(m, w_m, a_m) u(a_t). \quad (\text{SepDisc})$$

In Formula (SepDisc),  $\delta$  is a *separable discount function*. Both Chapters 3 and 4 consider the case of separable discounting.

We emphasise that these three discounting techniques mentioned above - *generalised discounting*, *ripple discounting*, and *separable discounting* - are introduced here only for illustrative purposes. In subsequent chapters, we provide their precise definitions within the context of the specific stochastic games.

In certain sections of this thesis, we focus on the *total reward* and the *long-run average reward*. For instance, in Example 1.1, the total reward is the expected value of the sum of the utilities obtained at each stage, that is, the expected value of

$$\sum_{t=1}^n u(a_t).$$

On the other hand, the *long-run average reward* is the expected value of

$$\limsup_{n \rightarrow \infty} \sum_{t=1}^n \frac{1}{n} u(a_t).$$

Once again, we emphasise that this chapter presents various reward functions solely for illustrative purposes. We provide the precise definitions of these reward functions in the subsequent chapters.

## 1.4 Relevance of the research topic

Stochastic games have a wide range of potential applications, such as competitions in weapons development (Winston, 1978), tax evasion (Raghavan, 2006), and conflicts over fishing resources (Levhari and Mirman, 1980). These applications represent just a few areas where such games can arise. Many articles and books discuss additional potential applications (see, for instance, White (1993); Puterman (1994); Filar and Vrieze (1997); Amir (2003); Solan and Vieille (2015); Chen (2019)).

In recent years, many articles have focused on discounted stochastic games with different discounting techniques. While this is not an exhaustive list, we highlight results on zero-sum games with general state space by Minjárez-Sosa (2015); González-Sánchez, Luque-Vásquez and Minjárez-Sosa (2019); Wu, Wang and Kong (2021); Wu, Tang and Medina (2022); Yu, Guo and Xia (2022), as well as results on Markov decision processes by (Wei and Guo, 2011; Ye and Guo, 2012; Wu and Guo, 2015). These studies primarily focused on various versions of the classical discounting technique.

The separable discounting differs from the generalised discounting and the classical discounting in that the current discount rate does not influence the future discount factors. We can think of the separable discounting as a specific weighting technique. Using weighting techniques is not a new approach in the theory of stochastic games (see, for example, Filar and Vrieze (1992); Altman, Feinberg and Shwartz (2000); Oliu-Barton (2021)).

This thesis focuses on the theory of discounted stochastic games, building upon the previously mentioned but non-exhaustive list. It is important to emphasise that this thesis aims to solve the theoretical problems.

## 1.5 Research methods

This thesis discusses discounted stochastic games using various mathematical branches based on the works of Aliprantis and Border (2006); Császár (1970); Kechris (1995); Laczkovich (1995); Srivastava (1998); Steen and Seebach (1978). Furthermore, each chapter of the thesis utilises specific game-theoretical materials.

Chapter 2 discusses finite stochastic games based on Parthasarathy and Babu (2020, Chapter 2), Solan (2022, Chapter 4) and Sorin (2003A). For continuous generalised games, we use the *Glicksberg Fixed Point Theorem* (Glicksberg, 1952).

Chapter 3 focuses on zero-sum stochastic games based on works of Parthas-

arathy and Babu (2020, Chapter 3) and Sorin (2003A,B). We use the work of Flesch et al. (2018, 2020) on positive zero-sum countable stochastic games. Additionally, Chapter 3 applies the mathematical preliminaries and the results of Nowak (1984A,B); Jaśkiewicz and Nowak (2018); Bertsekas and Shreve (1996) when investigating positive zero-sum stochastic games with a general state space. Finally, for the Shapley operators of one-shot games derived from positive zero-sum countable stochastic games, we also consider the works of Neyman (2003B); Sorin (2003C).

Chapter 4 presents the model of Sudderth (2016) on finitely additive Markov decision processes. Appendix A belongs to Chapter 3, and provides a brief overview of finitely additive measures (Dunford and Schwartz, 1957; Rao and Rao, 1983; Luxemburg, 1991) and the *gambling problem* (Dubins and Savage, 1965; Purves and Sudderth, 2010; Sudderth, 2016).

Finally, we use the following basic notations in this thesis:

- $\mathbb{N} = \{0, 1, 2, \dots\}$  denotes the non-negative integers.
- $\mathbb{T} = \{1, 2, 3, \dots\}$  denotes the set of all positive integers.
- $\mathbb{R}$  denotes the set of real numbers, and let

$$\overline{\mathbb{R}}_+ = \{x \in \mathbb{R} \mid x \geq 0\} \cup \{\infty\} \quad \text{and} \quad \overline{\mathbb{R}}_- = \{x \in \mathbb{R} \mid x \leq 0\} \cup \{-\infty\}.$$

- $|X|$  denotes the cardinality of the set  $X$ .
- $\mathcal{P}(X)$  denotes the power set of the set  $X$ .
- $\Delta(X, \mathcal{F})$  denotes the set of probability distributions on the measurable space  $(X, \mathcal{F})$ . If the  $\sigma$ -algebra is clear from the context - for example, if  $\mathcal{F} = \mathcal{P}(X)$  - we simply write  $\Delta(X)$  instead of  $\Delta(X, \mathcal{F})$ .
- $\text{Gam}(X)$  denotes the set of all gambles on  $X$ .

## 1.6 Research questions

This thesis investigates three problems in the theory of discounted stochastic games. For each problem, we formulate a structured hierarchy of research questions. We examine each problem in a separate chapter (see Table 1).

The main challenge in each chapter, as indicated in the first column of Table 1, arises from the combination of two elements: the *Game model* and the *Discounting*, which are listed in the second and third columns of the same table.

Chapter	Game model	Discounting	Approach
2	Finite stochastic games	Generalised	Continuous generalized games
3	Zero-sum stochastic games	Separable	Supergames
4	Finitely additive Markov decision processes	Ripple	Superprocesses
		Separable	

Table 1: Summary of the problems addressed and the approaches employed in each chapter.

The final column of Table 1 presents our approaches to examine each problem. While these methods are explained in detail in their respective chapters, it is important to highlight a key distinction here.

Chapters 3 and 4 utilise the concepts of *supergames* and *superprocesses*. These approaches extend the original stochastic game into a larger game by replacing the state space with a *position space*, which is often constructed by indexing states over time or representing them using the set of histories. This transformation is significant because it makes a strong connection between these games: the new game remains equivalent to the original one regarding our research goals.

In contrast, Chapter 2 takes a different approach by introducing a broader class of games that includes the original game as a special case.

With respect to Table 1, it must be emphasised that, although supergames (or superprocesses) are a well-established tool in stochastic game theory (see, for instance, Filar and Vrieze (1992); Judd, Yeltekin and Van Conklin (2006)), they remain relatively underutilised in the literature.

### 1.6.1 Finite stochastic games with generalised discounting

Chapter 2 explores *finite stochastic games with generalised discounting*. The starting point of our investigation is the fundamental result independently established by Fink (1964), Takahashi (1964), and Sobel (1971), which asserts that every finite stochastic game with exponential discounting admits a *stationary Nash equilibrium* (see Theorem 2.12).

In finite stochastic games, *generalised discounting* differs from *exponential discounting* in that the discount rate is not fixed but varies depending on the current stage and state. To the best of our knowledge, however, the analysis of finite stochastic games with *generalised discounting* has not yet been explored. Motivated

by this gap in the literature, we introduce our first research question:

(RQ1) *Does a finite stochastic game with generalised discounting admit a Nash equilibrium?*

The results by Fink (1964), Takahashi (1964), and Sobel (1971) not only state the existence of a Nash equilibrium in every finite stochastic game with exponential discounting, but also state that there exists at least one stationary strategy profile among the equilibrium strategy profiles. With this in mind, our second research question is as follows:

(RQ2) *Assuming the answer to Research Question (RQ1) is affirmative, is there a particular kind of equilibrium strategy profile?*

### 1.6.2 Zero-sum stochastic games with separable discounting

Chapter 3 examines *zero-sum stochastic games with separable discounting*. The existence of a value in zero-sum finite stochastic games with *exponential discounting* is a classical result by Shapley (1953). This result has been generalised in various directions, depending on the structural assumptions imposed on the game.

Our research focuses specifically on *zero-sum stochastic games with separable discounting*. To the best of our knowledge, several fundamental questions in this area remain open:

(RQ3) *Does a zero-sum stochastic game with separable discounting admit a value?*

If the answer to Research Question (RQ3) is affirmative, it naturally leads us to the next question:

(RQ4) *Assuming the answer to Research Question (RQ3) is affirmative, does there exist a 0-optimal strategy for each player?*

Finally, if we also receive a positive answer to Research Question (RQ4), then we consider the following research question:

(RQ5) *Assuming the answer to Research Question (RQ4) is affirmative, are there 0-optimal strategies that are either Markov or stationary for each player?*

### 1.6.3 Discounted finitely additive Markov decision processes

Chapter 4 investigates *finitely additive Markov decision processes with ripple discounting* and *separable discounting*. The starting point of our research is the following result by Sudderth (2016): in any finitely additive Markov decision process with *exponential discounting*, Player 1 always has an optimal stationary strategy (see Theorem 4.7).

When considering finitely additive Markov decision processes with *ripple discounting*, several fundamental questions remain open, to the best of our knowledge:

(RQ6) *Does a 0-optimal strategy exist for the player?*

(RQ7) *Assuming the answer to Research Question (RQ6) is affirmative, does a Markov or stationary 0-optimal strategy exist for the player?*

It is important to emphasise that the existence of an optimal reward is not a research question for finitely additive Markov decision processes with ripple or separable discounting, as it is automatically ensured by the definition of the discount function and the choice of game model.

Similarly, the following research questions arise in the context of *separable discounting*:

(RQ8) *Does a 0-optimal strategy exist for the player?*

(RQ9) *Assuming the answer to Research Question (RQ8) is affirmative, does a Markov or stationary 0-optimal strategy exist for the player?*

## 1.7 Structure of this thesis

Since the overall structure of the thesis has already been outlined in the previous sections, we provide only a summary here.

This chapter develops the basic intuition behind stochastic games and introduces the central research questions that guide our work.

Chapter 2 examines *finite stochastic games with generalised discounting* and addresses Research Questions (RQ1) and (RQ2).

Chapter 3 turns to *zero-sum stochastic games with ripple discounting*, focusing on Research Questions (RQ3), (RQ4), and (RQ5).

Chapter 4 investigates *finitely additive Markov decision processes with both ripple discounting and separable discounting*. It addresses Research Questions (RQ6) and

(RQ7) in the context of ripple discounting, and (RQ8) and (RQ9) in the context of separable discounting.

Chapter 5 provides a concise summary of the thesis and presents the results, reflecting on how they address the research questions.

# Chapter 2

## Finite stochastic games with generalised discounting

*“Allegro maestoso”*

---

W. A. MOZART: *Klavierkonzert Nr. 21 C-Dur, KV 467, 1. Satz*

This chapter focuses on *finite stochastic games with generalised discounting*. Our main goal is to address Research Questions (RQ1) and (RQ2).

We organise this chapter as follows. Section 2.1 introduces (*n*-person) *finite stochastic games* and describes their dynamics (Sorin, 2003A; Parthasarathy and Babu, 2020; Solan, 2022). It also offers graphical representations for the two-person finite stochastic games. Section 2.2 defines fundamental concepts such as *histories*, *plays*, and *behavioural strategies*. Section 2.3 presents *discounted reward functions*, focusing on *exponential* and *generalised discounting*. Section 2.4 discusses the concept of *Nash equilibrium* under the previously mentioned discounting techniques (Solan, 2022, Chapter 8). Section 2.5 introduces *continuous generalised games* based on the work of Glicksberg (1952), which lay the groundwork for addressing our two research questions in Section 2.6. Section 2.7 investigates additional properties of continuous generalised games and provides further theoretical insight for these games.

Chapter 2 is based on the work of Balog and Pintér (2025) and does not present any new results beyond those already reported therein (see Sections 2.3, 2.4, 2.5, 2.6, and 2.7).

## 2.1 The game model

This section introduces ( $n$ -person) finite stochastic games and describes their dynamics (Sorin, 2003A; Parthasarathy and Babu, 2020; Solan, 2022). The thesis primarily focuses on stochastic games presented in normal form, also known as strategic form. We conclude this section with a graphical representation of a two-person finite stochastic game, demonstrating that, in some cases, this graphical approach provides an equivalent way of specifying finite stochastic games compared to the normal form.

**Definition 2.1.** A ( $n$ -person) finite stochastic game is a tuple

$$\langle I, S, (A^i(s))_{s \in S}^{i \in I}, q, (r^i)^{i \in I} \rangle \quad (2.1)$$

which consists of the following components:

- (a) The nonempty finite set  $I$  denotes the set of players, with cardinality  $|I| = n$ .
- (b) The nonempty finite set  $S$  represents the set of states, also known as state space.
- (c) For each player  $i \in I$  and state  $s \in S$ ,  $A^i(s)$  denotes the nonempty finite set of actions available to player  $i \in I$  in state  $s \in S$ . Define the action space for player  $i \in I$  as

$$A^i = \bigcup_{s \in S} A^i(s).$$

Furthermore, define the set of action profiles in state  $s \in S$  as

$$A(s) = \prod_{i \in I} A^i(s).$$

- (d) The transition rule  $q: SA \rightarrow \Delta(S)$  determines how the  $n$ -person finite stochastic game evolves based on the current state and the players' actions. Here

$$SA = \{(s, a) \in S \times A: s \in S \text{ and } a \in A(s)\}$$

denotes the set of all action profiles across all states.

- (e) For each player  $i \in I$ , the function

$$r^i: SA \rightarrow \mathbb{R}$$

defines the one-stage payoff, assigning a real number to every pair  $(s, a) \in SA$ . The profile of these payoff functions for all players is denoted by

$$r = (r^i)^{i \in I},$$

and represents the one-stage payoff vector.

In line with the notations from Definition 2.1, we typically refer to finite stochastic games using the tuple

$$\langle I, S, (A^i)^{i \in I}, q, r \rangle$$

instead of the tuple presented in Formula (2.1).

It is important to highlight that Definition 2.1 does not present a single finite stochastic game but gives a family of specific finite stochastic games. The members of this family differ only in their initial states. To define a finite stochastic game precisely, we must specify its *initial state*. However, in many cases, analysing these games collectively proves to be advantageous (Solan, 2022).

In the next step, we explain how finite stochastic games progress over time. A finite stochastic game in Definition 2.1 starts at the initial state  $s_1 \in S$ , and the following happens at each stage  $t \in \mathbb{T}$ :

Step 1. The players observe the current state  $s_t \in S$ .

Step 2. Each player  $i \in I$  chooses an action  $a_t^i \in A^i(s_t)$  simultaneously and independently.

Step 3. The action profile  $(a_t^i)_{i \in I}$  is communicated to the players.

Step 4. The action profile  $(a_t^i)_{i \in I}$  induces the immediate one-stage payoff  $r^i(s_t, a_t)$  for each player  $i \in I$ .

Step 5. A new state  $s_{t+1} \in S$  is drawn according to the transition rule  $q(\cdot \mid s_t, a_t)$ , and the game proceeds at state  $s_{t+1}$ . Go back to Step 1.

We conclude this section by illustrating a two-player finite stochastic game. The goal of Example 2.2 is to show that in certain situations, we can represent stochastic games graphically in a way equivalent to using the normal form.

*Example 2.2.* Consider a finite stochastic game  $\langle I, S, (A^i)^{i \in I}, q, r \rangle$  where  $I = \{1, 2\}$  and  $S = \{s^1, s^2, s^3\}$ . The initial state of this two-person finite stochastic game is the state  $s^1$ .

The sets of available actions for player  $i \in \{1, 2\}$  in state  $s \in \{s^1, s^2, s^3\}$  are defined as follows:

$$A^1(s) = \begin{cases} \{\text{Up}, \text{Down}\} & \text{if } s = s^1, \\ \{\text{Black}, \text{White}\} & \text{if } s = s^2, \\ \{\text{East}\} & \text{if } s = s^3, \end{cases}$$

and

$$A^2(s) = \begin{cases} \{\text{Left}, \text{Right}\} & \text{if } s = s^1, \\ \{\text{North}\} & \text{if } s = s^2, \\ \{\text{Yellow}, \text{Green}\} & \text{if } s = s^3. \end{cases}$$

	Left (L)	Right (R)		North (N)
Up (U)	$-3, 2_{(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})}$	$4, -4_{(\frac{1}{7}, \frac{2}{7}, \frac{4}{7})}$		$-1, 2_{(0,0,1)}$
Down (D)	$2, -1_{(\frac{1}{3}, \frac{1}{6}, \frac{1}{2})}$	$-2, 2_{(\frac{1}{2}, 0, \frac{1}{2})}$		$2, -2_{(1,0,0)}$
	(a) state $s^1$			(b) state $s^2$

	Yellow (Y)	Green (G)
East (E)	$-1, 4_{(1,0,0)}$	$1, -3_{(0,1,0)}$
	(c) state $s^3$	

Figure 1: Graphical representation of the two-person finite stochastic game with three states in Example 2.2. Each entry in this figure displays the one-stage payoffs for both players, along with the transition probabilities indicated in the lower right subscript (enclosed in parentheses). The initial state of this two-person finite stochastic game is the state  $s^1$ .

In state  $s^2$ , Player 1 has the option to choose between the actions Black and White, while Player 2 can only select the North action. For simplicity, we refer to each action by its initial letter; for example, (L) stands for Left, and (U) stands for Up.

The transition rule  $q$  of this two-person finite stochastic game takes the following form:

$$\begin{aligned}
q(s^1 | s^1, U, L) &= \frac{1}{2}, & q(s^2 | s^1, U, L) &= \frac{1}{4}, & q(s^3 | s^1, U, L) &= \frac{1}{4}, \\
q(s^1 | s^1, U, R) &= \frac{1}{7}, & q(s^2 | s^1, U, R) &= \frac{2}{7}, & q(s^3 | s^1, U, R) &= \frac{4}{7}, \\
q(s^1 | s^1, D, L) &= \frac{1}{3}, & q(s^2 | s^1, D, L) &= \frac{1}{6}, & q(s^3 | s^1, D, L) &= \frac{1}{2}, \\
q(s^1 | s^1, D, R) &= \frac{1}{2}, & q(s^2 | s^1, D, R) &= 0, & q(s^3 | s^1, D, R) &= \frac{1}{2}, \\
q(s^1 | s^2, B, N) &= 0, & q(s^2 | s^2, B, N) &= 0, & q(s^3 | s^2, B, N) &= 1, \\
q(s^1 | s^2, W, N) &= 1, & q(s^2 | s^2, W, N) &= 0, & q(s^3 | s^2, W, N) &= 0, \\
q(s^1 | s^3, E, Y) &= 1, & q(s^2 | s^3, E, Y) &= 0, & q(s^3 | s^3, E, Y) &= 0, \\
q(s^1 | s^3, E, G) &= 0, & q(s^2 | s^3, E, G) &= 1, & q(s^3 | s^3, E, G) &= 0.
\end{aligned}$$

As a last element, we must provide the one-stage payoffs for the players, which are

defined as follows:

$$\begin{aligned}
r^1(s^1, U, L) &= -3, & r^1(s^1, U, R) &= 4, & r^1(s^1, D, L) &= 2, \\
r^1(s^1, D, R) &= -2, & r^1(s^2, B, N) &= -1, & r^1(s^2, W, N) &= 2, \\
r^1(s^3, E, Y) &= -1, & r^1(s^3, E, G) &= 1, & & \\
r^2(s^1, U, L) &= 2, & r^2(s^1, U, R) &= -4, & r^2(s^1, D, L) &= -1, \\
r^2(s^1, D, R) &= 2, & r^2(s^2, B, N) &= 2, & r^2(s^2, W, N) &= -2, \\
r^2(s^3, E, Y) &= 4, & r^2(s^3, E, G) &= -3. & & 
\end{aligned}$$

Example 2.2 considers a two-person finite stochastic game. Figure 1 shows this two-person finite stochastic game, with the row player identified as Player 1 and the column player as Player 2.

By the previous correspondence regarding the players, we depict the available actions for Player  $i \in \{1, 2\}$  at the state  $s \in \{s^1, s^2, s^3\}$  in Figure 1. Each entry in this figure displays the one-stage payoffs for both players, along with the transition probabilities indicated in the lower right subscript (enclosed in parentheses).

## 2.2 Histories, plays and strategies

This section introduces useful tools for analysing finite stochastic games (Solan, 2022). First, we define the *histories* and *plays*. Histories are important for introducing *strategies*, while plays are necessary for building the probabilistic framework of finite stochastic games. We conclude this section by presenting this probabilistic framework, which is crucial for discussing reward functions in Section 2.3.

We use histories to collect all the relevant information about the evolution of finite stochastic games. For every stage  $t \in \mathbb{T}$ , we define the collection of  $t$ -length *histories* as follows:

$$H_t = \begin{cases} S, & \text{if } t = 1 \\ (SA)^{t-1} \times S, & \text{if } t > 1. \end{cases}$$

Accordingly, a  $t$ -length history  $h_t = (s_1, a_1, \dots, s_{t-1}, a_{t-1}, s_t) \in H_t$  is a chronologically ordered series that gathers states from the first  $t \in \mathbb{T}$  stages as well as action profiles from the first  $t - 1$  stages.

If  $h = (s_1, a_1, s_2, \dots, s_{t-1}, a_{t-1}, s_t)$  is an arbitrary  $t$ -length history, then

- $\kappa(h)$  denotes the final state  $s_t$  of the history  $h$ ,
- $\text{len}(h)$  denotes the length of the history  $h$ .

In the case of Example 2.2, the history  $h = (s^1, (U, L), s^2, (B, N), s^3)$  is a 3-length history. This implies that  $\kappa(h) = s^3$  and  $\text{len}(h) = 3$ .

A *play* is a chronologically ordered infinite sequence of states and action profiles. The set of all histories ( $H$ ) and the set of all plays ( $H_\infty$ ) are denoted by:

$$H = \bigcup_{t \in \mathbb{T}} H_t \quad \text{and} \quad H_\infty = (SA)^\mathbb{T}.$$

A mixed action for player  $i \in I$  at state  $s \in S$  is a probability distribution over the set of available actions  $A^i(s)$ . Therefore, we denote the set of mixed actions for player  $i \in I$  at state  $s \in S$  by  $\Delta(A^i(s))$ .

In the next step, we introduce several classes of strategies. First, we present the so-called *behavioural strategy*, which specifies how a player chooses actions based on the given history.

**Definition 2.3.** A behavioural strategy for player  $i \in I$  is a map  $\sigma^i$  assigning a mixed action in  $\Delta(A^i(\kappa(h)))$  to each history  $h \in H$ .

We denote the collection of all behavioural strategies of player  $i \in I$  by  $\Sigma^i$ . Let  $\Sigma = \prod_{i \in I} \Sigma^i$  denote the collection of all behavioural strategy profiles.

Players may only consider a portion of the information from history, allowing for the introduction of special behavioural strategies.

**Definition 2.4.** A behavioural strategy of player  $i \in I$ , denoted by  $\sigma^i$ , is a Markov strategy if  $\sigma^i(h) = \sigma^i(h')$  holds for all histories  $h, h' \in H$  that satisfy these two criteria:

- both histories have the same length:  $\text{len}(h) = \text{len}(h')$ ;
- both histories reach the same final state:  $\kappa(h) = \kappa(h')$ .

Definition 2.4 states that a Markov strategy depends on the current stage and state.

**Definition 2.5.** A behavioural strategy of player  $i \in I$ , denoted by  $\sigma^i$ , is a stationary strategy if  $\sigma^i(h) = \sigma^i(h')$  holds for all histories  $h, h' \in H$  that satisfy the following criterium:

- both histories reach the same final state:  $\kappa(h) = \kappa(h')$ .

In other words, a stationary strategy depends on only the current state. According to Definitions 2.4 and 2.5, every stationary strategy is a Markov strategy.

A behavioural strategy  $\sigma^i$  is said to be *pure* if  $|\text{supp}(\sigma^i(h))| = 1$  for all histories  $h \in H$ . Similarly, we can define *pure Markov strategies* and *pure stationary strategies*.

We conclude this section by preparing the necessary components to present the reward functions in Section 2.3. Let

$$C(\hat{h}_t) = \left\{ h_\infty = (s_1, a_1, s_2, a_2, \dots, a_{t-1}, s_t, a_t, \dots) \in H_\infty \right. \\ \left. : s_1 = \hat{s}_1, a_1 = \hat{a}_1, s_2 = \hat{s}_2, a_2 = \hat{a}_2, \dots, a_{t-1} = \hat{a}_{t-1}, s_t = \hat{s}_t \right\} \quad (2.2)$$

for each history  $\hat{h}_t = (\hat{s}_1, \hat{a}_1, \hat{s}_2, \hat{a}_2, \dots, \hat{a}_{t-1}, \hat{s}_t) \in H_t$ . In other words, the cylindrical set  $C(\hat{h}_t)$  in Formula (2.2) contains all the plays which are compatible with the history  $\hat{h}_t$ .

For each stage  $t \in \mathbb{T}$ , the collection of all cylindrical sets forms an algebra  $\mathcal{H}_t$  over  $H_\infty$  (Solan, 2022, p. 8). Let  $\mathcal{H}$  denote the  $\sigma$ -algebra generated by the algebras  $(\mathcal{H}_t)_{t \in \mathbb{T}}$ . As a result,  $(H_\infty, \mathcal{H})$  is a measurable space.

By the Kolmogorov Extension Theorem (Aliprantis and Border, 2006, Theorem 15.26), every behavioural strategy profile  $\sigma = (\sigma^i)^{i \in I} \in \Sigma$  with the initial state  $s_1 \in S$  and the transition rule  $q$  determines a unique probability measure  $\mathbb{P}_{s_1}^\sigma$  on the measurable space  $(H_\infty, \mathcal{H})$  which defined as

$$\mathbb{P}_{s_1}^\sigma(C(\hat{h}_t)) = \chi_{\{s_1 = \hat{s}_1\}} \left( \prod_{\tau=1}^{t-1} \left( \prod_{i \in I} \sigma^i(\hat{a}_\tau^i | \hat{h}_\tau) \right) q(\hat{s}_{\tau+1} | \hat{s}_\tau, \hat{a}_\tau) \right)$$

for each history  $\hat{h}_t = (\hat{s}_1, \hat{a}_1, \hat{s}_2, \hat{a}_2, \dots, \hat{a}_{t-1}, \hat{s}_t) \in H_t$ . We denote the corresponding expectation operator with respect to the probability measure  $\mathbb{P}_{s_1}^\sigma$  by  $\mathbb{E}_{s_1}^\sigma$ .

## 2.3 Reward functions based on discounting

This section presents two discounting techniques for finite stochastic games: the *exponential discounting* and the *generalised discounting*. These methods are important for introducing the *exponential discounted reward* and the *generalised discounted reward* functions. We conclude the section by proving a key property of the generalised discounted reward functions.

In finite stochastic games, players receive one-stage payoffs at each stage, which raises the question of how players can compare the cash flow streams generated by different plays. One possible solution is to use *exponential discounting*.

**Definition 2.6.** Let  $\Gamma = \langle I, S, (A^i)^{i \in I}, q, r \rangle$  be a finite stochastic game starting from an initial state  $s \in S$ , and  $\alpha \in [0, 1)$  be a discount rate. The  $\alpha$ -exponential discounted reward for player  $i \in I$  under the behavioural strategy profile  $\sigma = (\sigma^i)^{i \in I} \in \Sigma$  is defined as

$$\gamma_\alpha^i(s, \sigma) = \mathbb{E}_s^\sigma \left[ \sum_{t=1}^{\infty} \alpha^{t-1} r^i(s_t, a_t) \right].$$

The  $\alpha$ -exponential discounted reward has several generalisations. In this chapter, we focus on *generalised discounting*, which is one potential direction for extending exponential discounting. In the case of generalised discounting, the discount rates vary based on the current stage and the current state of the finite stochastic game. The following step presents the *generalised discounted reward*.

**Definition 2.7.** Let  $\Gamma = \langle I, S, (A^i)^{i \in I}, q, r \rangle$  be a finite stochastic game. A function

$$\lambda: \mathbb{T} \times S \rightarrow [0, 1]$$

is a generalised discount function if there exists a number  $M_\lambda \in \mathbb{R}$  such that

$$\sum_{t=1}^{\infty} \left( \prod_{m=1}^t \lambda(m, s_m) \right) \leq M_\lambda \quad (2.3)$$

for every play  $h_\infty = (s_1, a_1, s_2, a_2, \dots) \in H_\infty$ .

Using the generalised discount function from Definition 2.7, we introduce the *generalised discounted reward* for finite stochastic games.

**Definition 2.8.** Let  $\Gamma = \langle I, S, (A^i)^{i \in I}, q, r \rangle$  be a finite stochastic game starting at an initial state  $s \in S$ , and let  $\lambda$  be a generalised discount function. The  $\lambda$ -generalised discounted reward for player  $i \in I$  under the behavioural strategy profile  $\sigma = (\sigma^i)^{i \in I} \in \Sigma$  is defined as

$$\gamma_\lambda^i(s, \sigma) = \mathbb{E}_s^\sigma \left[ \sum_{t=1}^{\infty} \left( \prod_{m=1}^t \lambda(m, s_m) \right) r^i(s_t, a_t) \right] = \int_{H_\infty} d_\lambda^i(h_\infty) d\mathbb{P}_{s, \sigma}, \quad (2.4)$$

where

$$d_\lambda^i(h_\infty) = \sum_{t=1}^{\infty} \left( \prod_{m=1}^t \lambda(m, s_m) \right) r^i(s_t, a_t) \quad (2.5)$$

for all plays  $h_\infty = (s_1, a_1, s_2, a_2, \dots) \in H_\infty$ .

We conclude this section by proving the following important property of the function  $d_\lambda^i$  in Formula (2.5):

**Lemma 2.9.** For each player  $i \in I$  and generalised discount function  $\lambda$ , the function  $d_\lambda^i$  in Formula (2.5) is continuous with respect to the product topology.

*Proof.* Let  $\varepsilon > 0$  be fixed and choose a bound  $K > 0$  such that

$$|r^i(s_t, a_t)| < K$$

for each stage  $t \in \mathbb{T}$ , state  $s_t \in S$  and action profile  $a_t \in A(s_t)$ .

We know that  $\lambda$  is a generalised discount function which meets the conditions outlined in Formula (2.3). This implies that there exists a stage  $t^* \in \mathbb{T}$  such that

$$\sum_{t=t^*+1}^{\infty} \prod_{m=1}^t \lambda(m, s_m) < \frac{\varepsilon}{2K}.$$

Now, consider the following two plays:  $h_{\infty}^x = (s_1^x, a_1^x, \dots, s_{t^*-1}^x, a_{t^*-1}^x, s_{t^*}^x, \dots)$  and  $h_{\infty}^y = (s_1^y, a_1^y, \dots, s_{t^*-1}^y, a_{t^*-1}^y, s_{t^*}^y, \dots)$  which coincide in the first  $t^* \in \mathbb{T}$  stages. In other words, there exists a history  $h_{t^*} = (s_1, a_1, \dots, a_{t^*-1}, s_{t^*}) \in H_{t^*}$  such that

$$h_{\infty}^x = (h_{t^*}, a_{t^*}^x, s_{t^*+1}^x, a_{t^*+1}^x, \dots) \quad \text{and} \quad h_{\infty}^y = (h_{t^*}, a_{t^*}^y, s_{t^*+1}^y, a_{t^*+1}^y, \dots).$$

In that case, we have

$$\begin{aligned} \left| d_{\lambda}^i(h_{\infty}^x) - d_{\lambda}^i(h_{\infty}^y) \right| &= \left| \sum_{t=1}^{\infty} \left( \prod_{m=1}^t \lambda(m, s_m) \right) r^i(s_t^x, a_t^x) \right. \\ &\quad \left. - \sum_{t=1}^{\infty} \left( \prod_{m=1}^t \lambda(m, s_m) \right) r^i(s_t^y, a_t^y) \right| \\ &= \left| \sum_{t=1}^{\infty} \left( \prod_{m=1}^t \lambda(m, s_m) \right) \left( r^i(s_t^x, a_t^x) - r^i(s_t^y, a_t^y) \right) \right| \\ &\leq \sum_{t=t^*+1}^{\infty} \left( \prod_{m=1}^t \lambda(m, s_m) \right) \left| \left( r^i(s_t^x, a_t^x) - r^i(s_t^y, a_t^y) \right) \right| < \varepsilon, \end{aligned}$$

meaning that  $d_{\lambda}^i$  in Formula (2.5) is a continuous function with respect to the product topology.  $\square$

The following observation immediately follows from Lemma 2.9.

**Corollary 2.10.** *For each player  $i \in I$  and discount rate  $\alpha \in [0, 1)$ , the function defined as*

$$d_{\alpha}^i(h_{\infty}) = \sum_{t=1}^{\infty} \alpha^{t-1} r^i(s_t, a_t)$$

*for all plays  $h_{\infty} = (s_1, a_1, s_2, a_2, \dots) \in H_{\infty}$  is continuous with respect to the product topology.*

## 2.4 Some notes on the research questions

This section briefly summarises the existence of the Nash equilibrium in finite stochastic games with *exponential discounting*. Next, we introduce the concept of the Nash equilibrium in finite stochastic games with *generalised discounting*. We conclude this section by discussing our Research Questions (RQ1) and (RQ2).

First, we introduce the Nash equilibrium in finite stochastic games with exponential discounting.

**Definition 2.11.** Let  $\Gamma = \langle I, S, (A^i)^{i \in I}, q, r \rangle$  be a finite stochastic game, and  $\alpha \in [0, 1)$  be a discount rate. A behavioural strategy profile  $\sigma_* = (\sigma_*^i)^{i \in I} \in \Sigma$  is a Nash equilibrium for the initial state  $s \in S$  if

$$\gamma_\alpha^i(s, \sigma_*) \geq \gamma_\alpha^i(s, (\sigma^i, \sigma_*^{-i}))$$

holds for each player  $i \in I$  and for all behavioural strategy  $\sigma^i \in \Sigma^i$ .

A behavioural strategy profile  $\sigma_*$  is a Nash equilibrium if it is a Nash equilibrium in each initial state.

In the next step, we present the following well-known result (Fink, 1964; Takahashi, 1964; Sobel, 1971):

**Theorem 2.12.** Any finite stochastic game with exponential discounting admits a stationary Nash equilibrium for each discount rate  $\alpha \in [0, 1)$ .

It is important to note that Theorem 2.12 states that there exists a stationary strategy profile, which is a Nash equilibrium. Remark 2.13 provides a brief overview of the proof for Theorem 2.12.

*Remark 2.13.* The proof of Theorem 2.12 consists of the following main steps (Solan, 2022, Section 8.3, p. 116-119):

- ED1) Let  $\Gamma$  be a finite stochastic game with exponential discounting, starting from the initial state  $s_1$ . We derive a one-shot game from  $\Gamma$  by stating that the game ends immediately after the first stage. In this structure, players take actions only in the first stage. To compensate for the abrupt termination of the game, we present players with a continuation payoff. This payoff depends on where the original finite stochastic game would have transitioned in the second stage.
- ED2) In this one-shot game, we introduce a special set-valued mapping with a fixed point according to Kakutani's Fixed Point Theorem (Kakutani, 1941). A fixed point in this specific set-valued mapping represents the one-shot game's equilibrium strategy and payoff profile.
- ED3) The construction implies that the equilibrium strategy profile of the one-shot game is stationary and that the equilibrium strategy profile of the original finite stochastic game inherits this characteristic.
- ED4) Finally, we demonstrate that the equilibrium payoff profile coincides with the payoff profile derived from the exponential discounted reward.

Now, we introduce the concept of Nash equilibrium for finite stochastic games with generalised discounting.

**Definition 2.14.** Let  $\Gamma = \langle I, S, (A^i)^{i \in I}, q, r \rangle$  be a finite stochastic game, and  $\lambda$  be a generalised discount function. A behavioural strategy profile  $\sigma_* = (\sigma_*^i)^{i \in I} \in \Sigma$  is a Nash equilibrium for the initial state  $s \in S$  if

$$\gamma_\lambda^i(s, \sigma_*) \geq \gamma_\lambda^i(s, (\sigma^i, \sigma_*^{-i}))$$

holds for each player  $i \in I$  and for all behavioural strategy  $\sigma^i \in \Sigma^i$ .

A behavioural strategy profile  $\sigma_*$  is a Nash equilibrium if it is a Nash equilibrium in each initial state.

As far as we know, no research has investigated the existence of a Nash equilibrium in finite stochastic games with generalised discounting. Therefore, it is natural to consider Research Questions (RQ1) and (RQ2).

It is important to note that we cannot apply the proof procedure sketched in Remark 2.13 because the discount rate varies over time in the case of generalised discounting. This observation implies that we cannot introduce a derived one-shot game in Step ED1) that would be equivalent to the original stochastic game in the spirit of Remark 2.13.

## 2.5 Continuous generalised games

This section outlines our approach to addressing Research Question (RQ1). It starts by presenting the notion of continuous generalised games and proceeds to establish the existence of a Nash equilibrium in such games, applying the Glicksberg Fixed Point Theorem (Glicksberg, 1952).

First, we introduce *continuous generalised games* as follows:

**Definition 2.15.** A continuous generalised game is a tuple

$$\langle I, P, (A_i)_{i \in I}, F, (f_i)_{i \in I} \rangle$$

which consists of the following components:

- (a)  $I$  is a nonempty finite set of players, with cardinality  $|I| = n$ .
- (b)  $P$  is a nonempty topological space of outcomes.
- (c) For each player  $i \in I$ ,  $A_i$  is a nonempty set of actions of player  $i \in I$ . Let  $A = \prod_{i \in I} A_i$  denote the set of action profiles.
- (d) The outcome mapping  $F: A \rightarrow P$  satisfies:

(OM1)  $F$  is affine,

(OM2)  $F(A)$  is a compact and convex set of  $P$ ,

(OM3) For every player  $i \in I$  and every truncated action profile  $a_{-i} \in A_{-i}$ , the set  $F(A_i \times \{a_{-i}\})$  is compact and convex.

(e) For each player  $i \in I$ , the payoff function  $f_i: P \rightarrow \mathbb{R}$  satisfies:

(PF1)  $f_i$  is continuous,

(PF2)  $f_i$  is concave on the set  $F(A_i \times \{a_{-i}\})$  for every  $a_{-i} \in A_{-i}$ .

In the next step, we introduce the concept of a *Nash equilibrium* (Nash, 1950, 1951) for continuous generalised games.

**Definition 2.16.** Let  $\langle I, P, (A_i)_{i \in I}, F, (f_i)_{i \in I} \rangle$  be a continuous generalised game. The strategy profile  $a^* \in A$  is a *Nash equilibrium* if

$$f_i \circ F(a^*) \geq f_i \circ F(a_i, a_{-i}^*)$$

holds for every player  $i \in I$  and every strategy  $a_i \in A_i$ .

Note that when  $P = A$ , i.e., when  $F = \text{id}$ , the Nash equilibrium in Definition 2.16 reduces to the well-known concept of a Nash equilibrium for normal form games.

We conclude this section by showing that every continuous generalised game admits a Nash equilibrium:

**Theorem 2.17.** *Every continuous generalised game has a Nash equilibrium.*

The proof of Theorem 2.17 proceeds in three steps. First, we establish key properties of the *individual best-response correspondence*. Next, we analyse the *joint best-response correspondence* similarly. Finally, we apply *Glicksberg Fixed Point Theorem* to demonstrate that the joint best-response correspondence admits a fixed point.

*Proof of Theorem 2.17.* Let  $\Gamma = \langle I, P, (A_i)_{i \in I}, F, (f_i)_{i \in I} \rangle$  be a continuous generalised game. For each player  $i \in I$ , we define the *individual best-response correspondence* as follows:

$$B_i(p) = \arg \max_{p' \in F(A_i \times \{a_{-i}\})} f_i(p') \quad (2.6)$$

for all  $p = F(a) \in F(A)$ .

As stated in Point (OM3) of Definition 2.15, the set  $F(A_i \times \{a_{-i}\})$  is compact for every player  $i \in I$  and each truncated action profile  $a_{-i} \in A_{-i}$ . Additionally,

Point (PF1) of the same definition states that the payoff function  $f_i$  is continuous for each player  $i \in I$ . These properties together ensure that the individual best-response correspondence  $B_i$  in Formula (2.6) is nonempty-valued for each player  $i \in I$  and for each  $p \in P$ .

Furthermore, Points (OM3) and (PF2) ensure that the set  $F(A_i \times \{a_{-i}\})$  is convex and that the payoff function is concave over this set for each  $i \in I$  and every  $a_{-i} \in A_{-i}$ . These conditions imply that the individual best-response correspondence  $B_i$  in Formula (2.6) is convex-valued for each  $i \in I$  and for each  $p \in P$ .

By Points (OM2) and (OM3), we know that the sets  $F(A)$  and  $F(A_i \times \{a_{-i}\})$  are compact for each player  $i \in I$  and every truncated action profile  $a_{-i} \in A_{-i}$ . Moreover, the payoff function  $f_i$  is continuous, as specified by Point (PF1) of Definition 2.15. Therefore, applying Berge's Maximum Theorem (Aliprantis and Border, 2006, Theorem 17.31, p. 570) ensures that the individual best-response correspondence  $B_i$ , given by Formula (2.6), is upper hemicontinuous for each player  $i \in I$ .

In the next step, we introduce the *joint best-response correspondence* defined by:

$$B(p) = F \left( \prod_{i \in I} (F^{-1}(B_i(p)))_i \right) \quad (2.7)$$

for every  $p \in F(A)$ .

It is important to note that the joint best-response correspondence  $B$  in Formula (2.7) is nonempty-valued. This follows from the fact that for any strategy profile  $a \in \prod_{i \in I} (F^{-1}(B_i(p)))_i$ , we have  $F(a) \in B(p)$  for all  $p \in F(A)$ . Moreover, the set  $\prod_{i \in I} (F^{-1}(B_i(p)))_i$  is convex for each  $p \in F(A)$ , since  $F$  is an affine function by Point (OM1) of Definition 2.15.

We have previously shown that each individual best-response correspondence  $B_i$  in Formula (2.6) is upper hemicontinuous. Therefore, the joint best-response correspondence  $B$  in Formula (2.7) is also upper hemicontinuous.

Finally, by *Glicksberg Fixed Point Theorem*, we conclude that the joint best-response correspondence  $B$  in Formula (2.7) has a fixed point. Let  $s^* \in F(A)$  denote such a fixed point. Then, there exists an action profile  $a^* \in A$  such that  $s^* = F(a^*)$ . It follows that  $a^*$  is a Nash equilibrium of the continuous generalised game  $\Gamma$ .  $\square$

We note that Theorem 2.17 extends the result of Glicksberg (1952, pp. 172–174). Additionally, we draw attention to the following point:

**Corollary 2.18.** *Let  $\langle I, P, (A_i)_{i \in I}, F, (f_i)_{i \in I} \rangle$  be a continuous generalised game. Suppose  $a \in A$  and  $\hat{a} \in A$  are Nash equilibria such that  $F(a) = F(\hat{a})$ . Then, for every  $\beta \in [0, 1]$ , the convex combination  $\beta a + (1 - \beta)\hat{a}$  is also a Nash equilibrium.*

## 2.6 Results for finite stochastic games with generalised discounting

This section addresses Research Questions (RQ1) and (RQ2), beginning with the former, as the answer to Research Question (RQ2) depends on the outcome of Research Question (RQ1).

To investigate Research Question (RQ1), we consider the pair  $(\Gamma_{\text{FSG}}, \lambda)$ , where  $\lambda$  denotes a generalised discount function, and  $\Gamma_{\text{FSG}} = \langle I, S, (A^i)_{i \in I}, q, r \rangle$  represents a family of finite stochastic games that differ only in their initial states.

As a preliminary observation for the family  $\Gamma_{\text{FSG}}$ , the Kolmogorov Extension Theorem (Aliprantis and Border, 2006, Theorem 15.26) ensures that for any initial state  $s \in S$ , transition rule  $q$  and behavioural strategy profile  $\sigma \in \Sigma$ , there exists a unique probability measure  $(s, \sigma)$  defined on the measurable space  $(H_\infty, \mathcal{B}(H_\infty))$ .

Next, we consider the following family of games

$$\Gamma_{\text{CGG}} = \langle I, P, (A_i)_{i \in I}, F, (f_i)_{i \in I} \rangle, \quad (2.8)$$

which is derived from the pair  $(\Gamma_{\text{FSG}}, \lambda)$ , and satisfies the following properties:

1. The set of players is the same in both games.
2. The set of outcomes is defined as

$$P = \Delta(H_\infty, \mathcal{B}(H_\infty)).$$

3. For each player  $i \in I$ , the set of actions is given by  $A_i = \Sigma^i$  where  $\Sigma^i$  denotes the the collection of all behavioural strategies of player  $i \in I$ .
4. The mapping  $F: \Sigma \rightarrow P$  assigns to each behavioural strategy profile  $\sigma \in \Sigma$  a probability measure  $F(\sigma) \in \Delta(H_\infty, \mathcal{B}(H_\infty))$ . This probability measure can be constructed by the Kolmogorov Extension Theorem (Aliprantis and Border, 2006, Theorem 15.26) based on the initial state  $s$  and the transition rule  $q$ .
5. For each player  $i \in I$ , the payoff function is the  $\lambda$ -generalised discounted reward as described in Formula (2.4). Formally, it is given by

$$f_i(\mathbb{P}_{s, \sigma}) = \int_{H_\infty} d_\lambda^i d\mathbb{P}_{s, \sigma}.$$

It is clear that each member in the family  $\Gamma_{\text{FSG}}$  corresponds uniquely to a member in the family  $\Gamma_{\text{CGG}}$  in Formula (2.8). This one-to-one relationship arises

because the mapping  $F$  and the payoff functions  $(f_i)_{i \in I}$  are fully determined by the selected finite stochastic game, its initial state, and the generalised discount function  $\lambda$ .

As an interesting observation, note the following: it is possible for two distinct strategy profiles  $\sigma, \sigma' \in \Sigma$ , where  $\sigma \neq \sigma'$ , to induce the same probability measure via the Kolmogorov Extension Theorem (Aliprantis and Border, 2006, Theorem 15.26). In other words,  $F(\sigma) = F(\sigma')$  can hold even though the behavioural strategies differ.

The following step demonstrates that  $\Gamma_{\text{CGG}}$  constitutes a family of continuous generalised games. To establish this, it suffices to show that each element of  $\Gamma_{\text{CGG}}$  meets the criteria specified in Definition 2.15.

First, we demonstrate that the mapping  $F$  of any game belonging to the family  $\Gamma_{\text{CGG}}$  in Formula (2.8) satisfies the conditions specified in Points(OM1) and (OM2).

**Proposition 2.19.** *For each member of  $\Gamma_{\text{CGG}}$  in Formula (2.8), the following properties hold:*

- (a) *The mapping  $F$  is affine.*
- (b) *The set  $F(\Sigma)$  is convex.*
- (c) *The set  $F(\Sigma)$  is weak\* compact.*

*Proof.* (a) It follows directly from the Kolmogorov Extension Theorem (Aliprantis and Border, 2006, Theorem 15.26).

(b) Since  $\Sigma$  is a convex set, and Point (a) establishes that  $F$  is an affine map, it follows that the set  $F(\Sigma)$  is also convex.

(c) Since the set of plays  $H_\infty$  forms a compact metrizable space, the set of probability measures  $\Delta(H_\infty, \mathcal{B}(H_\infty))$  is a weak\* compact metrizable set (Aliprantis and Border, 2006, Theorem 15.11). We note that the relevant dual pair in this context is  $(C(H_\infty), \text{cba}(H_\infty, \mathcal{B}(H_\infty)))$  (for more details, see Aliprantis and Border (2006, Chapter 15)).

Our goal is to show that the set

$$F(\Sigma) = \{(s, \sigma) : \sigma \in \Sigma\} \subset \Delta(H_\infty, \mathcal{B}(H_\infty))$$

is a weak\* closed set for any initial state  $s \in S$ . Since  $\Delta(H_\infty, \mathcal{B}(H_\infty))$  is weak\* compact, and any weak\* closed subset of a weak\* compact set is itself weak\* compact, it follows that  $F(\Sigma)$  is weak\* compact.

For each stage  $t \in \mathbb{T}$ , define

$$S_t = \{(s, \sigma)|_{H_t} : \sigma \in \Sigma \text{ and } H_t \subseteq H\}$$

as the set of probability measures consistent with the initial state  $s_1 \in S$  and the set of  $t$ -length histories  $H_t$ . Similarly, for every stage  $t \in \mathbb{T}$ , define

$$U_t = \{\nu \in \Delta(H_\infty, \mathcal{B}(H_\infty)) : \nu|_{(H_t, \mathcal{B}(H_t))} \in S_t\},$$

where  $\nu|_{(H_t, \mathcal{B}(H_t))}$  denotes the restriction of the probability measure  $\nu$  to the measurable space  $(H_t, \mathcal{B}(H_t))$ . In other words, each  $\nu \in U_t$  is a probability measure on  $\Delta(H_\infty, \mathcal{B}(H_\infty))$  whose restriction to  $t$ -length histories lies in  $S_t$ . Clearly,  $U_t \subseteq \Delta(H_\infty, \mathcal{B}(H_\infty))$  for every stage  $t \in \mathbb{T}$ .

Observe that the sets satisfy the nesting property  $U_{t+1} \subseteq U_t$  for all  $t \in \mathbb{T}$ . Moreover, each  $U_t$  is a weak\* closed subset of the weak\* compact set  $\Delta(H_\infty, \mathcal{B}(H_\infty))$  since it has a finite dimension.

Finally, since

$$F(\Sigma) = \bigcap_{t \in \mathbb{T}} U_t,$$

it follows that  $F(\Sigma) = \{(s, \sigma) : \sigma \in \Sigma\}$  is a weak\* closed set for every initial state  $s \in S$ . Being a weak\* closed subset of the weak\* compact set  $\Delta(H_\infty, \mathcal{B}(H_\infty))$ ,  $F(\Sigma)$  is, therefore, weak\* compact.  $\square$

Second, we prove that the mapping  $F$  of any game belonging to the family  $\Gamma_{\text{CGG}}$  in Formula (2.8) satisfies the properties outlined in Point (OM3).

**Proposition 2.20.** *For each member of  $\Gamma_{\text{CGG}}$  in Formula (2.8), the following properties hold: for every player  $i \in I$  and for every behavioural strategy profile  $\sigma \in \Sigma$ , the set  $F(\Sigma^i \times \{\sigma^{-i}\})$  is*

(a) *convex.*

(b) *compact.*

*Proof.* We only prove Point (a), since Point (b) can be shown using an argument similar to the one in the proof of Point (c) in Proposition 2.19.

It is clear that the set  $\Sigma^i \times \{\sigma^{-i}\}$  is convex for each player  $i \in I$  and for every truncated strategy profile  $\sigma^{-i} \in \Sigma^{-i}$ . Since  $F$  is an affine map by Point (a) of Proposition 2.19, we have that the set  $F(\Sigma^i \times \{\sigma^{-i}\})$  is convex for each player  $i \in I$  and for every truncated strategy profile  $\sigma^{-i} \in \Sigma^{-i}$ .  $\square$

Third, we prove that the payoff functions of any game belonging to the family  $\Gamma_{\text{CGG}}$  in Formula (2.8) meet the criteria stated in Points (PF1) and (PF2). We can establish this aim by observing the following property:

**Lemma 2.21.** *For each member of  $\Gamma_{\text{CGG}}$  in Formula (2.8), the following property holds: given any behavioural strategy profile  $\sigma \in \Sigma$  and any generalised discount function  $\lambda$ , the  $\lambda$ -generalised discounted reward in Definition 2.8 is affine.*

*Proof.* This result is a direct corollary of the definition of the  $\lambda$ -generalised discounted reward (see Definition 2.8).  $\square$

By applying Propositions 2.19 and 2.20, together with Lemma 2.21, we obtain the following result:

**Corollary 2.22.** *Every member of  $\Gamma_{\text{CGG}}$  in Formula (2.8) is a continuous generalised game.*

Based on the construction of the family  $\Gamma_{\text{CGG}}$  in Formula (2.8) and Corollary 2.22, we obtain the following result:

**Corollary 2.23.** *Every finite stochastic game with separable discounting is a continuous generalised game.*

We conclude this section by answering Research Questions (RQ1) and (RQ2). Theorem 2.17 and Corollary 2.23 imply the following conclusion:

**Theorem 2.24.** *Let  $\Gamma = \langle I, S, (A^i)^{i \in I}, q, r \rangle$  be a finite stochastic game with generalised discounting. Then, for any initial state  $s \in S$  and any generalised discount function  $\lambda$ ,  $\Gamma$  admits a  $\lambda$ -discounted Nash equilibrium.*

Theorem 2.24 gives a positive answer to Research Question (RQ1), that is, every finite stochastic game with generalised discounting has a discounted Nash equilibrium. Theorem 2.24 also implies that it makes sense to consider Research Question (RQ2).

*Example 2.25.* Consider the finite stochastic game  $\Gamma = \langle I, S, (A^i)^{i \in I}, q, r \rangle$  in Figure 2 where  $I = \{1, 2\}$  and  $S = \{s^1, s^2\}$ . The game  $\Gamma$  starts at the initial state  $s^1 \in S$ .

We assume that the players focus on the  $\lambda$ -generalised discounted reward. The generalised discount function  $\lambda$  is defined as follows:

$$\lambda(t, s_t) = \begin{cases} 0.9^t, & \text{if } s_t = s^1, \\ 1, & \text{if } (t, s_t) \in \{(2, s^2)\}, \\ 0 & \text{otherwise.} \end{cases}$$

	Left (L)	Right (R)
Up (U)	1, 1 <sub>(0,1)</sub>	0, 0 <sub>(1,0)</sub>
Down (D)	0, 0 <sub>(1,0)</sub>	3, 3 <sub>(1,0)</sub>

	No (N)
Yes (Y)	100, 100 <sub>(1,0)</sub>

(b)  $s^2$

(a)  $s^1$

Figure 2: Graphical representation of the two-person finite stochastic game with two states in Example 2.25. The initial state of this game is the state  $s^1$ .

Note that both players can maximise their  $\lambda$ -discounted rewards only if the play

$$(s^1, (U, L), s^2, (Y, N), s^1, (D, R), s^1, (D, R), \dots) \in H_\infty$$

happens with probability one.

For instance, this means that at the first stage, Player 1 chooses the action Up with probability one in the initial  $s^1$ . Following the transition to state  $s^2$ , Player 1 selects the action Yes with probability one. From that point forward, Player 1 consistently opts for the action Down with probability one at every subsequent stage. Let  $\hat{\sigma}^1$  denote this behavioural strategy.

Similarly, for Player 2, this means that at the first stage, Player 2 chooses the action Left with probability one in the initial state  $s^1$ . After transitioning to state  $s^2$ , Player 2 selects the action No with probability one. From then on, Player 2 consistently chooses the action Right with probability one at every stage. Let  $\hat{\sigma}^2$  denote this behavioural strategy.

It is clear that the behavioural strategy profile  $(\hat{\sigma}^1, \hat{\sigma}^2)$  in Example 2.25 constitutes a Nash equilibrium. However, it also follows that the two-person finite stochastic game in Example 2.25 does not admit a stationary Nash equilibrium; in other words, there is no equilibrium strategy profile consisting of only stationary strategies.

Example 2.25 leads to the following observation:

**Lemma 2.26.** *There exists a finite stochastic game with generalised discounting that does not have a stationary Nash equilibrium.*

In conclusion, every finite stochastic game with generalised discounting admits a Nash equilibrium. This result positively answers Research Question (RQ1).

In the case of Research Question (RQ2), our conclusion is the following: there exists a finite stochastic game with generalised discounting that does not admit a stationary Nash equilibrium, as demonstrated in Lemma 2.26.

It is important to emphasise that it remains an open question whether a Markov discounted Nash equilibrium exists for finite stochastic games with generalised discounting.

## 2.7 Some comments on continuous generalised games

Section 2.6 addressed Research Questions (RQ1) and (RQ2) by the results obtained for continuous generalised games. This section examines continuous generalised games in greater depth. We demonstrate that certain games, such as the *mixed extension* of games, belong to the class of continuous generalised games. We also present examples of games that do not belong to this class, including *Pick the Bigger Integer* (Wald, 1949), *Generalised Wald's Game*, the *Big Match* (Gillette, 1957), and the *Paris Match* (Sorin, 1986). The latter two are particularly significant, showing that even two-person finite stochastic games (with long-run average) do not belong to the class of continuous generalised games.

Figure 3 presents several examples of continuous generalised games, including finite stochastic games with generalised discounting. It is important to emphasise that finite stochastic games with exponential discounting also belong to this class. Additionally, when it exists, we consider the mixed extension of a normal form game as another example of a continuous generalised game (see Section 2.7.1).

### Continuous generalised games

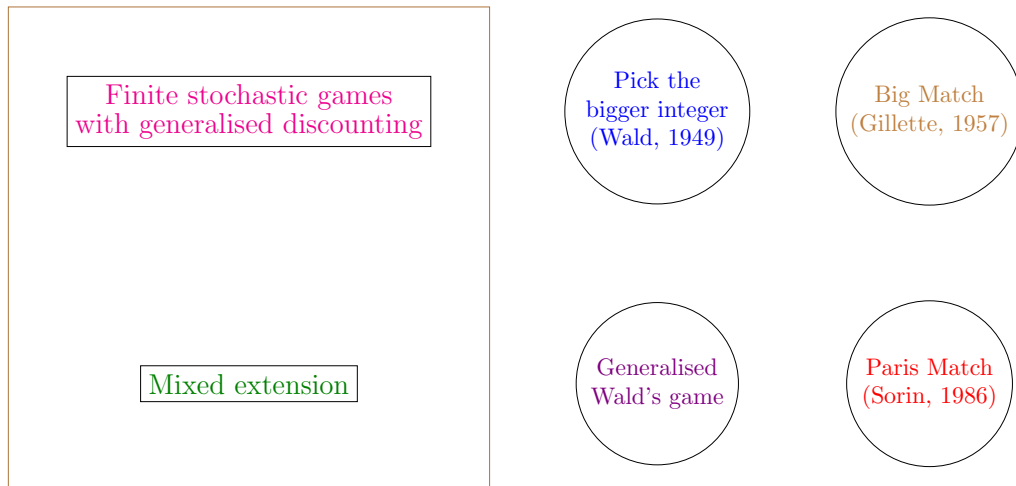


Figure 3: Some examples of games that belong to the class of continuous generalised games, along with examples that do not do so.

### 2.7.1 Mixed extension

First, we introduce the concept of the mixed extension of normal form games.

**Definition 2.27.** Let  $\Gamma = \langle I, (A_i)_{i \in I}, (f_i)_{i \in I} \rangle$  be a game which consists of the following components:

- (a)  $I$  represents the nonempty finite set of players.
- (b)  $A_i$  is the nonempty set of actions for player  $i \in I$ .
- (c)  $f_i: A \rightarrow \mathbb{R}$  is the payoff function for each player  $i \in I$  where

$$A = \prod_{i \in I} A_i$$

represents the set of action profiles.

Consider a game  $\hat{\Gamma} = \langle I, (\Delta(A_i))_{i \in I}, (u_i)_{i \in I} \rangle$  derived from the game  $\Gamma$ . Here, the payoff function  $u_i: \Delta(A_1) \times \cdots \times \Delta(A_n) \rightarrow \mathbb{R}$  for each player  $i \in I$  is defined as follows:

$$u_i(\mu_1, \dots, \mu_n) = \int_A f_i d(\mu_1 \times \cdots \times \mu_n).$$

If such a game  $\hat{\Gamma}$  exists, then we refer to  $\hat{\Gamma}$  as the mixed extension of the game  $\Gamma$ .

According to Definition 2.27, it follows that not every game has a mixed extension. To demonstrate this, we present two illustrative examples. The first, in Example 2.28, is the well-known game introduced by Wald (1949), commonly known as the *Pick the bigger integer game*.

*Example 2.28.* The Wald's game is the tuple

$$\langle I, (A_i)_{i \in I}, (f_i)_{i \in I} \rangle,$$

which consists of the following components:

- (a)  $I = \{1, 2\}$ , indicating that the game involves two players.
- (b) For each player  $i \in I$ , the action set is given by  $A_i = \mathbb{N}$ .
- (c) The payoff function for each player  $i \in I$  is defined as:

$$f_1(a_1, a_2) = -f_2(a_1, a_2) = \begin{cases} 1, & \text{if } a_1 \geq a_2, \\ -1, & \text{if } a_1 < a_2. \end{cases}$$

There are several variants of Wald's game. We focus on the following version:

*Example 2.29.* The generalised Wald's game is a tuple  $\langle I, (A_i)_{i \in I}, (f_i)_{i \in I} \rangle$  where

- $I = \{1, 2\}$ ,

- $A_i = ([0, 1], \tau_d)$ , that is, the unit interval endowed with the discrete topology for each player  $i \in I$ ,
- the payoff functions of the players are given by

$$f_1(a_1, a_2) = -f_2(a_1, a_2) = \begin{cases} 1 & \text{if } a_1 \geq a_2, \\ 0 & \text{if } a_1 < a_2. \end{cases}$$

**Lemma 2.30.** *The generalised Wald's game does not have a mixed extension.*

*Proof.* It is easy to see that  $f_1$  is the indicator function of the set  $\{(x, y) \in [0, 1] \times [0, 1] : x \geq y\}$ . However, this set does not belong to the  $\sigma$ -algebra generated by the cylindrical sets. Consequently,  $f_1$  is not integrable and  $\hat{f}_1$  is therefore not defined. As a result, the generalised Wald's game does not have a mixed extension.  $\square$

A mixed extension of a game does not necessarily exist. Proposition 2.31 provides sufficient conditions for the existence of a mixed extension. Furthermore, we demonstrate that a mixed extension is a continuous generalised game.

**Proposition 2.31.** *Let  $\Gamma_B = \langle I, (A_i)_{i \in I}, (f_i)_{i \in I} \rangle$  be the game in Definition 2.27, and assume that*

- (i)  $A_i$  is a compact metrizable topological space for each player  $i \in I$ ,
- (ii)  $f^i$  is a continuous function with respect to the product topology for each player  $i \in I$ .

Then,

- (a) the game  $\Gamma_B$  has a mixed extension denoted as  $\hat{\Gamma}_B$ .
- (b)  $\hat{\Gamma}_B$  is a continuous generalised game.

*Proof.* (a) According to Point (i) in Proposition 2.31,  $A_i$  is metrizable for each player  $i \in I$ . This implies that the probability measure  $\mu_i \in \Delta(A_i)$  is inner regular for each player  $i \in I$  (Aliprantis and Border, 2006, Theorem 12.5). Furthermore, we also assume in Point (i) in Proposition 2.31 that  $A_i$  is compact for each player  $i \in I$ . Since every closed subset of a compact set is compact,  $\mu_i \in \Delta(A_i)$  is a tight probability measure (for more details, see Aliprantis and Border (2006, Chapter 12)).

According to Point (ii) in Proposition 2.31,  $f^i$  is a continuous function with respect to the product topology for each player  $i \in I$ . It follows directly that  $f^i$  is integrable with respect to any probability measure defined on the Borel sets of the

product topology. Moreover,  $\mu = \times_{i \in I} \mu_i$  is a probability measure in the  $\sigma$ -algebra generated by cylindrical sets for any  $\mu_i \in \Delta(A_i)$ , where  $i \in I$ . It is straightforward to see that  $\mu$  is tight.

In Proposition 2.31, we assume that  $A_i$  is a compact metrizable space (in the weak\* topology), as stated in Point (i). Let  $A = \prod_{i \in I} A_i$  represent the product space. This product space is also a compact metrizable space. Moreover, any probability measure in  $\Delta(A)$  is tight. We know that  $\mu$  has a unique extension on the Borel sets of  $A$ .

In the rest of this proof,  $\mu$  represents an element from the set  $\Delta(A)$ . Based on the previous discussion, the expected payoff function  $\hat{f}_i$  in Definition 2.27 exists for each player  $i \in I$ . This implies that the mixed extension  $\hat{\Gamma}_B$  of the game  $\Gamma_B$  exists.

(b) Consider the dual pair  $(C(A), \text{cba}(A))$  where  $\text{cba}(A)$  denotes the space of bounded probability measures on  $\mathcal{B}(A)$ . Since  $A$  is a compact metrizable space, it follows that  $\Delta(A)$  is also a compact metrizable space in the weak\* topology (Aliprantis and Border, 2006, Theorem 15.11). We notice that the subbasis of the weak\* topology on the set  $\Delta(A)$  is given by

$$\left\{ \mu \in \Delta(A) : \left| \int_A f \, d\mu - \int_A f \, d\mu' \right| < \varepsilon \right\}$$

where  $f \in C(A)$ ,  $\mu' \in \Delta(A)$  and  $\varepsilon > 0$ . Since  $\prod_{i \in I} \Delta(A_i)$  is a weak\* closed subset of  $\Delta(A)$ , it follows that it is also weak\* compact.

It is also clear that the set  $\prod_{i \in I} \Delta(A_i)$  is convex. Additionally, as stated in Point (ii) of Proposition 2.31,  $f_i \in C(A)$  for each player  $i \in I$ . This indicates that the expected payoff function  $\hat{f}_i$  is continuous and linear, which implies that it is also concave.

Now, we have that  $\prod_{i \in I} \Delta(A_i)$  weak\* compact and  $\hat{f}_i$  is a linear continuous function for each player  $i \in I$ . Finally, let  $P = \prod_{i \in I} \Delta(A_i)$  and  $F = \text{id}$ . It is easy to see that  $F$  meets the conditions in Point (d) in Definition 2.15. This results in the mixed extension being a continuous generalised game.  $\square$

## 2.7.2 The Big Match and the Paris Match

This subsection considers the *Big Match* and the *Paris Match*. To study these games, we need the definition of the *long-run average reward*.

**Definition 2.32.** Let  $\Gamma = \langle I, S, (A^i)^{i \in I}, q, r \rangle$  be a finite stochastic game starting at an initial state  $s \in S$ . The long-run average reward for player  $i \in I$  under the behavioural

strategy profile  $\sigma = (\sigma^i)^{i \in I} \in \Sigma$  is defined as

$$\gamma_{is}^i(s, \sigma) = \mathbb{E}_s^\sigma \left[ \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n r^i(s_t, a_t) \right].$$

Example 2.33 revisits the *Big Match* (Gillette, 1957) (for further details, see Thuijsman (2003)).

*Example 2.33.* The *Big Match* (Gillette, 1957) is the zero-sum finite stochastic game in Figure 4 with the long-run average reward.

	Left (L)	Right (R)	
Up (U)	1, -1 <sub>(1,0,0)</sub>	0, 0 <sub>(1,0,0)</sub>	
Down (D)	0, 0 <sub>(0,1,0)</sub>	1, -1 <sub>(0,0,1)</sub>	
	(a) state $s^1$		
	Left (L)		Right (R)
Down (D)	0, 0 <sub>(0,1,0)</sub>		1, -1 <sub>(0,0,1)</sub>
	(b) state $s^2$		(c) state $s^3$

Figure 4: Graphical representation of the zero-sum finite stochastic game with three states in Example 2.33. The initial state of the game is  $s^1$ .

Example 2.34 revisits the *Paris Match* Sorin (1986), which is not zero-sum (for further details, see Thuijsman (2003)).

*Example 2.34.* The *Paris Match* Sorin (1986) is a two-person finite stochastic game in Figure 5 with the long-run average reward.

	Left (L)	Right (R)	
Up (U)	1, 0 <sub>(1,0,0)</sub>	0, 1 <sub>(1,0,0)</sub>	
Down (D)	0, 2 <sub>(0,1,0)</sub>	1, 0 <sub>(0,0,1)</sub>	
	(a) state $s^1$		
	Left (L)		Right (R)
Down (D)	0, 2 <sub>(0,1,0)</sub>		1, 0 <sub>(0,0,1)</sub>
	(b) state $s^2$		(c) state $s^3$

Figure 5: Graphical representation of the two-person finite stochastic game with three states in Example 2.34. The initial state of the game is  $s^1$ .

The following result is based on the Example of Flesch (1998, Example 11, p. 15):

**Lemma 2.35.** *The Big Match and the Paris Match do not belong to the class of continuous generalised games.*

*Proof.* We only present the case of the Big Match since the Paris Match follows similarly.

Suppose that Player 2 (the column player) follows the stationary strategy  $\sigma^2 = (1, 0)$  against the stationary strategy  $\sigma^1$  of Player 1 (the row player) in the first stage of the Big Match. In that case, the long-run average reward for Player 1 is

$$\gamma_{ls}^1(\sigma^1, \sigma^2) = \begin{cases} 1, & \text{if } \sigma^1 = (1, 0), \\ 0, & \text{if } \sigma^1 \in \{(1-x, x) : 0 < x \leq 1\}, \end{cases}$$

which is not a continuous function. □

# Chapter 3

## Zero-sum stochastic games with separable discounting

"Andante"

---

W. A. MOZART: *Klavierkonzert Nr. 21 C-Dur, KV 467, 2. Satz*

This chapter focuses on (*two-person*) *zero-sum stochastic games with separable discounting*. For simplicity, we refer to players as *Player 1* and *Player 2*, respectively. Our goal is to answer Research Question (RQ3), (RQ4) and (RQ5).

As a reminder, a two-person stochastic game is *zero-sum* if the sum of the one-stage payoffs is zero at every state and stage of play. The chapter also considers *positive zero-sum stochastic games*, where the one-stage payoff function (of Player 1) is non-negative in all states. For example, the zero-sum finite stochastic game in Figure 4 is positive.

To grasp the main objective of this chapter, consider the zero-sum finite stochastic game  $\langle I, S, (A^i)^{i \in I}, q, r \rangle$  in Figure 4, in which Player 1 seeks to maximise the expected value of the *separable discounted total payoff*, defined as

$$\sum_{t=1}^{\infty} \delta(t, s_t, a_t) r(s_t, a_t). \quad (3.1)$$

In Formula (3.1),  $\delta$  is a *separable discount function*. The subsequent sections provide a detailed formulation of the problem, including a rigorous definition of the discount function  $\delta$ , which ensures that the expected value of the *separable discounted total payoff* in Formula (3.1) is well-defined.

We use *supergames* to address Research Questions (RQ3), (RQ4), and (RQ5) (see Table 1). We derive the supergame from the original zero-sum stochastic

---

game with separable discounting by replacing its *state space* with the *position space*, the collection of all states indexed by time.

To illustrate all the technical difficulties that arise in addressing our research questions, we separate our investigation into two parts. The first part of this chapter focuses on zero-sum *countable* stochastic games (see Section 3.1), while the second turns to zero-sum stochastic games with a *general state space* (see Section 3.2).

The reason behind this separation is the following. In the case of zero-sum countable stochastic games, it is more straightforward and transparent to show that supergames do not necessarily admit a *value*. The reason is that a supergame is a zero-sum countable stochastic game with total reward. Naturally, this issue also arises in the second part of the chapter. It is important to highlight that we do not review the problem in detail again; instead, we specify models in which these issues do not arise.

We organise this chapter as follows. Section 3.1 explores zero-sum *countable* stochastic games with separable discounting. Subsection 3.1.1 introduces the *game model* of zero-sum stochastic games, while Subsection 3.1.2 presents the *separable discounting* and the related concepts for these games. Subsection 3.1.3 explains how we define *supergames* for zero-sum countable stochastic games with separable discounting. Subsection 3.1.4 introduces the *lower* and the *upper Shapley operators*. Finally, Subsection 3.1.5 presents our main results on zero-sum countable stochastic games with separable discounting.

Section 3.2 explores zero-sum *infinite* stochastic games with separable discounting. In the second part of the chapter, we discuss several results that are explained in greater detail in the first part. Subsection 3.2.1 presents the *game model* of zero-sum infinite stochastic games. Subsection 3.2.2 introduces the *separable discounting* for zero-sum infinite stochastic games. Since many models can be formulated as zero-sum infinite stochastic games, we restrict our attention in Subsection 3.2.3 to three specific types: the *zero-sum Borel stochastic game*, the *zero-sum Suslin stochastic game* and the *zero-sum Nowak stochastic game*. Subsection 3.2.4 presents our results for these specific stochastic games.

## 3.1 Zero-sum countable stochastic games with separable discounting

### 3.1.1 The game model

This subsection outlines the framework of *zero-sum countable stochastic games* and describes how they evolve over time. Additionally, it introduces several important concepts, such as *history* and *play*, and *behavioural strategies*. This subsection ends with a technical discussion to prepare for introducing reward functions (Parthasarathy and Babu, 2020; Flesch et al., 2018, 2020).

First, we define the class of zero-sum countable stochastic games. In these games, we usually omit the specification of the set of players, as it is evident from the context. Furthermore, we only need to specify the one-stage payoff function for Player 1, since the zero-sum nature of the game implies the one-stage payoffs for Player 2.

**Definition 3.1.** A zero-sum countable stochastic game is a tuple

$$\langle S, (A(s))_{s \in S}, (B(s))_{s \in S}, q, r \rangle,$$

which is played by Player 1 and Player 2 and consists of the following components:

- (a) The state space  $S$  is a nonempty and countable set of states.
- (b) For each state  $s \in S$ , the sets  $A(s)$  and  $B(s)$  denote the nonempty and countable sets of available action of Player 1 and Player 2 in state  $s \in S$ , respectively. We define the action spaces of Player 1 and Player 2 as

$$A = \bigcup_{s \in S} A(s) \quad \text{and} \quad B = \bigcup_{s \in S} B(s).$$

- (c) The transition rule  $q$  assigns a probability distribution  $q(\cdot \mid s, a, b)$  over the state space  $S$  to each triple  $(s, a, b) \in K$ , where

$$K = \{(s, a, b) \in S \times A \times B : s \in S, a \in A(s) \text{ and } b \in B(s)\}.$$

- (d) The one-stage payoff function  $r$  is a bounded real-valued function that maps each triple  $(s, a, b) \in K$  to a one-stage payoff  $r(s, a, b)$ .

It is important to emphasise that Definition 3.1 does not describe a single zero-sum countable stochastic game, but rather a family of such games. A complete

specification of any individual game within this family requires identifying its initial state.

Using the notation introduced in Definition 3.1, a zero-sum countable stochastic game  $\langle S, (A(s))_{s \in S}, (B(s))_{s \in S}, q, r \rangle$  can equivalently be expressed more concisely as  $\langle S, A, B, q, r \rangle$ .

This chapter frequently examines specific subclasses of zero-sum countable stochastic games. A zero-sum countable stochastic game is *positive* if its one-stage payoff function is non-negative. A *positive zero-sum finite stochastic game* is simply a zero-sum finite stochastic game with a non-negative one-stage payoff function.

Next, we describe the dynamics of zero-sum countable stochastic games. A zero-sum countable stochastic game starts at an initial state  $s_1 \in S$ , and in each stage  $t \in \mathbb{T}$ , the following events take place:

- Step 1. The players observe the current state  $s_t \in S$ .
- Step 2. Player 1 and Player 2 choose actions  $a_t \in A(s_t)$  and  $b_t \in B(s_t)$  simultaneously and independently.
- Step 3. The action profile  $(a_t, b_t)$  is communicated to the players.
- Step 4. The current state  $s_t$  and the action profile  $(a_t, b_t)$  induce the one-stage payoff  $r(s_t, a_t, b_t)$  for Player 1. The one-stage payoff for Player 2 is  $-r(s_t, a_t, b_t)$ .
- Step 5. A new state  $s_{t+1} \in S$  is drawn according to the transition rule  $q(\cdot \mid s_t, a_t, b_t)$ , and the game proceeds at state  $s_{t+1}$ . Go back to Step 1..

In analysing zero-sum countable stochastic games, gathering the information available at each stage is important. Given a zero-sum countable stochastic game in Definition 3.1, we define the set of histories of length  $t \in \mathbb{T}$  as follows:

$$H_t = \begin{cases} S, & \text{if } t = 1 \\ K^{t-1} \times S, & \text{if } t > 1. \end{cases}$$

Additionally, we define the set of all histories and the set of all plays as

$$H = \bigcup_{t \in \mathbb{T}} H_t \quad \text{and} \quad H_\infty = K^\mathbb{T}.$$

For histories, we introduce some additional notation. Given a history  $h \in H$ , let  $\kappa(h)$  denote the *final state* in the history, and  $\text{len}(h)$  denote the *length* of the history. For example, we have  $\kappa(h_t) = s_t$  and  $\text{len}(h_t) = t$  for the history  $h_t = (s_1, a_1, b_1, \dots, s_{t-1}, a_{t-1}, b_{t-1}, s_t) \in H_t$ .

We conclude this subsection by introducing several strategic concepts, followed by a technical supplement to aid in designing reward functions. Consider a zero-sum countable stochastic game  $\langle S, A, B, q, r \rangle$ . In this setting, a *mixed action* for Player 1 in state  $s \in S$  is a probability distribution over the set of available actions  $A(s)$ . We denote the set of such mixed actions for Player 1 by  $\Delta(A(s))$ . Similarly,  $\Delta(B(s))$  denotes the set of mixed actions for Player 2 at state  $s \in S$ .

**Definition 3.2.** Let  $\langle S, A, B, q, r \rangle$  be a zero-sum countable stochastic game.

A behavioural strategy for Player 1 is a mapping  $\pi$  that assigns to each history  $h \in H$  a mixed action in  $\Delta(A(\kappa(h)))$ .

The behavioural strategy  $\pi$  for Player 1 is called a Markov strategy if

$$\pi(h) = \pi(h') \quad \text{whenever} \quad \text{len}(h) = \text{len}(h') \text{ and } \kappa(h) = \kappa(h').$$

The behavioural strategy  $\pi$  for Player 1 is called a stationary strategy if

$$\pi(h) = \pi(h') \quad \text{whenever} \quad \kappa(h) = \kappa(h').$$

Strategies for Player 2 are defined analogously. Let  $\Pi$ ,  $\Pi_M$  and  $\Pi_S$  denote the sets of all behavioural strategies, Markov strategies, and stationary strategies for Player 1, respectively. Likewise, let  $\Theta$ ,  $\Theta_M$  and  $\Theta_S$  denote the corresponding sets of strategies for Player 2.

Finally, every strategy profile  $(\pi, \vartheta) \in \Pi \times \Theta$  with the initial state  $s \in S$  and the transition rule  $q$  determines a unique probability measure  $\mathbb{P}_s^{\pi, \vartheta}$  on  $(H_\infty, \mathcal{H})$ , by the Kolmogorov Extension Theorem (Aliprantis and Border, 2006, Theorem 15.26), where  $\mathcal{H}$  is the  $\sigma$ -algebra generated by the collection of cylindrical sets. We denote the corresponding expectation operator by  $\mathbb{E}_s^{\pi, \vartheta}$ .

### 3.1.2 Zero-sum countable stochastic games with separable discounting

This subsection introduces the concept of the *value* in zero-sum countable stochastic games with *separable discounting*. First, we give the notion of a *separable discount function*, which forms the basis for the corresponding idea of *separable discounted reward*. Following this, we formally define the *separable discounted value*. To emphasise the difficulty of separable discounting and how it differs fundamentally from the exponential discounting, we include two examples demonstrating how this alternative discounting method can lead to substantially different outcomes, even in zero-sum finite stochastic games.

First, we introduce the concept of separable discount functions within zero-sum stochastic games.

**Definition 3.3.** Let  $\langle S, A, B, q, r \rangle$  be a zero-sum countable stochastic game. A function

$$\delta: \mathbb{T} \times K \rightarrow [0, 1]$$

is a separable discount function if there exists a number  $M_\delta \in \mathbb{R}$  such that for every play  $h_\infty = (s_1, a_1, b_1, s_2, a_2, b_2, \dots) \in H_\infty$ , the following holds:

$$\sum_{t=1}^{\infty} \delta(t, s_t, a_t, b_t) \leq M_\delta. \quad (3.2)$$

Next, we introduce the concept of the *separable discounted reward* for zero-sum countable stochastic games.

**Definition 3.4.** Let  $\Gamma = \langle S, A, B, q, r \rangle$  be a zero-sum countable stochastic game starting from the initial state  $s \in S$ , and let  $\delta$  be a separable discount function. The  $\delta$ -separable discounted reward for Player 1 under the behavioural strategy profile  $(\pi, \vartheta) \in \Pi \times \Theta$  is defined as

$$\gamma_\delta(\pi, \vartheta)(s) = \mathbb{E}_s^{\pi, \vartheta} \left[ \sum_{t=1}^{\infty} \delta(t, s_t, a_t, b_t) r(s_t, a_t, b_t) \right]. \quad (3.3)$$

The *exponential discounted reward* in zero-sum countable stochastic games arises as a special case of the separable discounted reward. Specifically, consider the separable discount function  $\delta$  in Definition 3.4, and suppose it takes the following form for each stage  $t \in \mathbb{T}$ :

$$\delta(t, s_t, a_t, b_t) = f(t)g(s_t, a_t, b_t),$$

where  $f: \mathbb{T} \rightarrow [0, 1]$  and  $g: K \rightarrow [0, 1]$ . If we choose  $f(t) = \alpha^{t-1}$  and set  $g(s_t, a_t, b_t) = 1$  for all  $t \in \mathbb{T}$ ,  $s_t \in S$ ,  $a_t \in A(s_t)$  and  $b_t \in B(s_t)$ , then we recover the exponential discounted reward. Note that, in this case, Formula (3.2) holds for any constant  $M_\delta \geq \frac{1}{1-\alpha}$ .

In the next step, we define the *separable discounted value* for zero-sum countable stochastic games with separable discounting.

**Definition 3.5.** Let  $\Gamma = \langle S, A, B, q, r \rangle$  be a zero-sum countable stochastic game, and let  $\delta$  be a separable discount function. The lower and upper  $\delta$ -separable discounted values of  $\Gamma$  for the initial state  $s \in S$  are defined respectively as

$$\underline{v}_\delta(s) = \sup_{\pi \in \Pi} \inf_{\vartheta \in \Theta} \gamma_\delta(\pi, \vartheta)(s) \quad \text{and} \quad \bar{v}_\delta(s) = \inf_{\vartheta \in \Theta} \sup_{\pi \in \Pi} \gamma_\delta(\pi, \vartheta)(s).$$

If the lower  $\delta$ -separable discounted value  $\underline{v}_\delta(s)$  equals the upper  $\delta$ -separable discounted value  $\bar{v}_\delta(s)$  for a given initial state  $s \in S$ , then the game  $\Gamma$  is said to have a  $\delta$ -separable discounted value at the initial state  $s \in S$ , which is denoted by  $v_\delta(s)$ .

Moreover, if  $\Gamma$  has a separable discounted value for every initial state  $s \in S$ , we say that  $\Gamma$  has a  $\delta$ -separable discounted value.

In any zero-sum countable stochastic game  $\langle S, A, B, q, r \rangle$  with a separable discount function  $\delta$ , it is important to observe that the lower  $\delta$ -separable discounted value is always less than or equal to the upper  $\delta$ -separable discounted value. That is, for every initial state  $s \in S$ , the inequality  $v_\delta(s) \leq \bar{v}_\delta(s)$  holds.

Research Question (RQ3) primarily explores the condition under which the lower and upper separable discounted values always coincide in zero-sum countable stochastic games with separable discounting.

Research Question (RQ4) addresses the optimality of strategies, assuming that the answer to Research Question (RQ3) is affirmative. Consequently, the next step involves classifying strategies based on the separable discounted value.

**Definition 3.6.** Let  $\Gamma = \langle S, A, B, q, r \rangle$  be a zero-sum countable stochastic game which admits a  $\delta$ -separable discounted value  $v_\delta$ , where  $\delta$  denotes a separable discount function.

For any  $\varepsilon \geq 0$ , a behavioural strategy  $\pi \in \Pi$  for Player 1 is  $\varepsilon$ -optimal at the initial state  $s \in S$ , if

$$\gamma_\delta(\pi, \vartheta)(s) \geq v_\delta(s) - \varepsilon$$

for every behavioural strategy  $\vartheta \in \Theta$  for Player 2.

Likewise, for any  $\varepsilon \geq 0$ , a behavioural strategy  $\vartheta \in \Theta$  for Player 2 is  $\varepsilon$ -optimal at the initial state  $s \in S$ , if

$$\gamma_\delta(\pi, \vartheta)(s) \leq v_\delta(s) + \varepsilon$$

for every behavioural strategy  $\pi \in \Pi$  for Player 1.

A behavioural strategy is  $\varepsilon$ -optimal if it is  $\varepsilon$ -optimal for every initial state.

As is standard practice, we call a behavioural strategy *optimal* if it is 0-optimal in a zero-sum countable stochastic game.

We conclude this subsection by presenting two examples of zero-sum finite stochastic games with separable discounting. Examples 3.7 and 3.8 demonstrate the differences between separable discounting and exponential discounting. We mention that Example 3.7 provides a basis for Example 3.8.

	Left (L)	Right (R)
Up (U)	8	13
Down (D)	15	16

(a) state  $s^1$

Figure 6: Graphical representation of the zero-sum finite stochastic game with a single state in Example 3.7. Since the game is zero-sum, we present only the one-stage payoffs for Player 1. We omit the transition probabilities, as the game has only one state.

*Example 3.7.* Consider the zero-sum finite stochastic game  $\Gamma = \langle S, A, B, q, r \rangle$  in Figure 6, where the state space consists of a single element  $S = \{s^1\}$ . Both players have nonempty finite action space:  $A = A(s^1) = \{\text{Up}, \text{Down}\}$  and  $B = B(s^1) = \{\text{Left}, \text{Right}\}$ .

We assume that Players 1 and 2 focus on the  $\delta$ -separable discounted reward. The separable discount function  $\delta$  is defined as follows:

$$\delta(t, s_t, a_t, b_t) = \begin{cases} \frac{1}{2}, & \text{if } (t, s_t, a_t, b_t) \in \{(1, s^1, \text{U}, \text{L}), (1, s^1, \text{U}, \text{R})\}, \\ \frac{1}{3}, & \text{if } (t, s_t, a_t, b_t) \in \{(2, s^1, \text{D}, \text{L}), (2, s^1, \text{D}, \text{R})\}, \\ \frac{1}{4}, & \text{if } (t, s_t, a_t, b_t) \in \{(3, s^1, \text{U}, \text{L}), (3, s^1, \text{U}, \text{R})\}, \\ \frac{1}{5}, & \text{if } (t, s_t, a_t, b_t) \in \{(4, s^1, \text{D}, \text{L}), (4, s^1, \text{D}, \text{R})\}, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that the behavioural strategy  $\vartheta^* = (1, 0) \in \Theta$  is an optimal stationary strategy for Player 2, meaning Player 2 consistently chooses the Left action with probability one. Consequently, the  $\delta$ -separable discounted reward for Player 2 cannot exceed the value given by:

$$\frac{1}{2} \times (-8) + \frac{1}{3} \times (-15) + \frac{1}{4} \times (-8) + \frac{1}{5} \times (-15) = -4 - 5 - 2 - 3 = -14$$

regardless of the behavioural strategy chosen by Player 1 in response to  $\vartheta^*$ .

In contrast, for Player 1, any optimal behavioural strategy  $\pi^* = (\pi_1^*, \pi_2^*, \dots) \in \Pi$  can be expressed as:

$$\pi_t^* = \begin{cases} (1, 0), & \text{if } t \in \{1, 3\}, \\ (0, 1), & \text{if } t \in \{2, 4\}, \\ (x_t, 1 - x_t), & \text{if } t \geq 5, \end{cases}$$

where  $x_t \in [0, 1]$  for all  $t \geq 5$ . Under this behavioural strategy  $\pi^*$ , the  $\delta$ -separable discounted reward for Player 1 is guaranteed to be at least 14, irrespective of the behavioural strategy employed by Player 2 in response to  $\pi^*$ . It is clear that  $\pi^*$  is not a stationary strategy.

Example 3.7 highlights a key distinction from exponential discounting. In every zero-sum finite stochastic game with exponential discounting, Shapley (1953) showed that both players have optimal stationary strategies.

However, in the zero-sum finite stochastic game with separable discounting in Example 3.7, Player 1 lacks an optimal stationary strategy. Furthermore, neither player has an optimal stationary strategy in Example 3.8.

*Example 3.8.* Consider the modified version of the zero-sum finite stochastic game presented in Example 3.7, in Figure 7. In this variant, the one-stage payoffs  $c_1, c_2, c_3$ , and  $c_4$  are all positive real numbers.

	Left (L)	Right (R)
Up (U)	$c_1$	$c_2$
Down (D)	$c_3$	$c_4$

(a) state  $s^1$

Figure 7: Graphical representation of the zero-sum finite stochastic game with a single state and arbitrary positive one-stage payoffs in Example 3.8, where  $c_1, c_2, c_3$ , and  $c_4$  are all positive real numbers. Since the game is zero-sum, we present only the one-stage payoffs for Player 1. We omit the transition probabilities, as the game has only one state.

We assume that both Player 1 and Player 2 evaluate one-stage payoffs using the  $\delta$ -separable discounted reward criterion. The separable discount function  $\delta$  is defined as follows:

$$\delta(t, s_t, a_t, b_t) = \begin{cases} d_1, & \text{if } (t, s_t, a_t, b_t) \in \{(1, s^1, U, L)\} \\ d_2, & \text{if } (t, s_t, a_t, b_t) \in \{(2, s^1, D, R)\} \\ 0 & \text{otherwise,} \end{cases}$$

where  $d_1, d_2 \in (0, 1)$ .

For Player 1, any optimal behavioural strategy  $\pi^* = (\pi_1^*, \pi_2^*, \dots) \in \Pi$  is given by:

$$\pi_t^* = \begin{cases} (1, 0), & \text{if } t = 1, \\ (0, 1), & \text{if } t = 2, \\ (x_t, 1 - x_t), & \text{if } t \geq 3, \end{cases}$$

where  $x_t \in [0, 1]$  for all  $t \geq 3$ . This means Player 1 chooses action Up with probability one at the first stage, and action Down with probability one at the second stage. Consequently, such a behavioural strategy is not stationary, as it depends on the stage.

Likewise, Player 2 lacks an optimal stationary strategy, since any optimal behavioural strategy  $\vartheta^* = (\vartheta_1^*, \vartheta_2^*, \dots) \in \Theta$  is given by:

$$\vartheta_t^* = \begin{cases} (0, 1), & \text{if } t = 1, \\ (1, 0), & \text{if } t = 2, \\ (y_t, 1 - y_t), & \text{if } t \geq 3, \end{cases}$$

where  $y_t \in [0, 1]$  for all  $t \geq 3$ .

The main takeaway from Example 3.8 is that no player has an optimal stationary strategy, even in a zero-sum finite stochastic game with separable discounting.

### 3.1.3 Supergames of zero-sum countable stochastic games with separable discounting

This subsection explains how to derive a supergame from a zero-sum countable stochastic game with separable discounting. First, we introduce the supergames and comment on the key elements involved in their derivation. Next, we outline the temporal structure of the supergame. After that, we demonstrate several fundamental properties of the supergame and introduce the notion of *total reward* specific to these games. Building on the concept of *total reward*, we introduce the *lower total value*, *upper total value*, and the *total value* of the supergame.

Before defining supergames, we introduce the following notations. For a bounded function  $f: X \rightarrow \mathbb{R}$ , define

$$\text{glb}^*(f) := \min\{0, \text{glb}(f)\},$$

where  $\text{glb}(f)$  represents the greatest lower bound of  $f$ . Likewise,  $\text{lub}(f)$  denotes the least upper bound of the bounded function  $f$ .

**Definition 3.9.** Let  $\Gamma = \langle S, A, B, q, r \rangle$  be a zero-sum countable stochastic game starting at the initial state  $s_1 \in S$ , and let  $\delta$  denote a separable discount function.

A supergame  $\mathbb{S}(\Gamma, \delta)$  of  $\Gamma$  is a tuple

$$\langle X, C, D, p, f \rangle$$

which is played by the same players and consists of the following components:

- (a) The set  $X$ , referred to as the position space, is defined as follows:

$$X = S \times \mathbb{T} = \{(s, t) : s \in S \text{ and } t \in \mathbb{T}\}.$$

The supergame  $\mathbb{S}(\Gamma, \delta)$  starts at the initial position  $(s_1, 1) \in X$ .

- (b) The set  $C$  represents the action space for Player 1 and is defined as

$$C = \bigcup_{(s,t) \in X} C(s, t),$$

where for each position  $(s, t) \in X$ ,

$$C(s, t) = A(s)$$

denotes the set of available actions for Player 1 at the position  $(s, t) \in X$ .

(c) The set  $D$  represents the action space for Player 2 and is defined as

$$D = \bigcup_{(s,t) \in X} D(s,t),$$

where for each position  $(s,t) \in X$ ,

$$D(s,t) = B(s)$$

denotes the set of available actions for Player 2 at the position  $(s,t) \in X$ .

(d) The transition law  $p: Z \rightarrow \Delta(X)$  is defined by

$$p(Y \times \{t'\} \mid (s,t), c, d) = \begin{cases} q(Y \mid s, c, d), & \text{if } t' = t + 1, \\ 0, & \text{otherwise,} \end{cases} \quad (3.4)$$

for every  $Y \subseteq S$  where the set  $Z$  is given by

$$Z = \{(s,t,c,d) \in X \times C \times D : (s,t) \in X, c \in C(s,t) \text{ and } d \in D(s,t)\},$$

(e) The function  $f$  denotes the one-stage payoff function, and it is defined as follows:

$$f(s,t,c,d) = \delta(t,s,c,d) \left( r(s,c,d) - \text{glb}^*(r) \right) \quad (3.5)$$

for all  $(s,t) \in X$ .

In the supergame  $\mathbb{S}(\Gamma, \delta)$ , Player 1 aims to maximise the expected value of the total payoff, given by

$$\sum_{m=1}^{\infty} f(x_m, c_m, d_m), \quad (3.6)$$

whereas Player 2 aims to minimise this.

Before proceeding, we highlight several key observations regarding supergames. The primary motivation for introducing supergames is as follows. In establishing the existence of a value for a zero-sum countable stochastic game with exponential discounting, some steps are similar to those outlined in Remark 2.13, such as the use of a one-shot game. This step cannot be executed in separable discounting because the discount rate varies over time. However, this problem is addressed within the supergame framework by introducing the position space, which can be viewed as an expanded state space. The supergame is specifically designed to eliminate the dependence on time. As detailed in Definition 3.9, this modification is reflected in multiple aspects of the formal definition of the supergame.

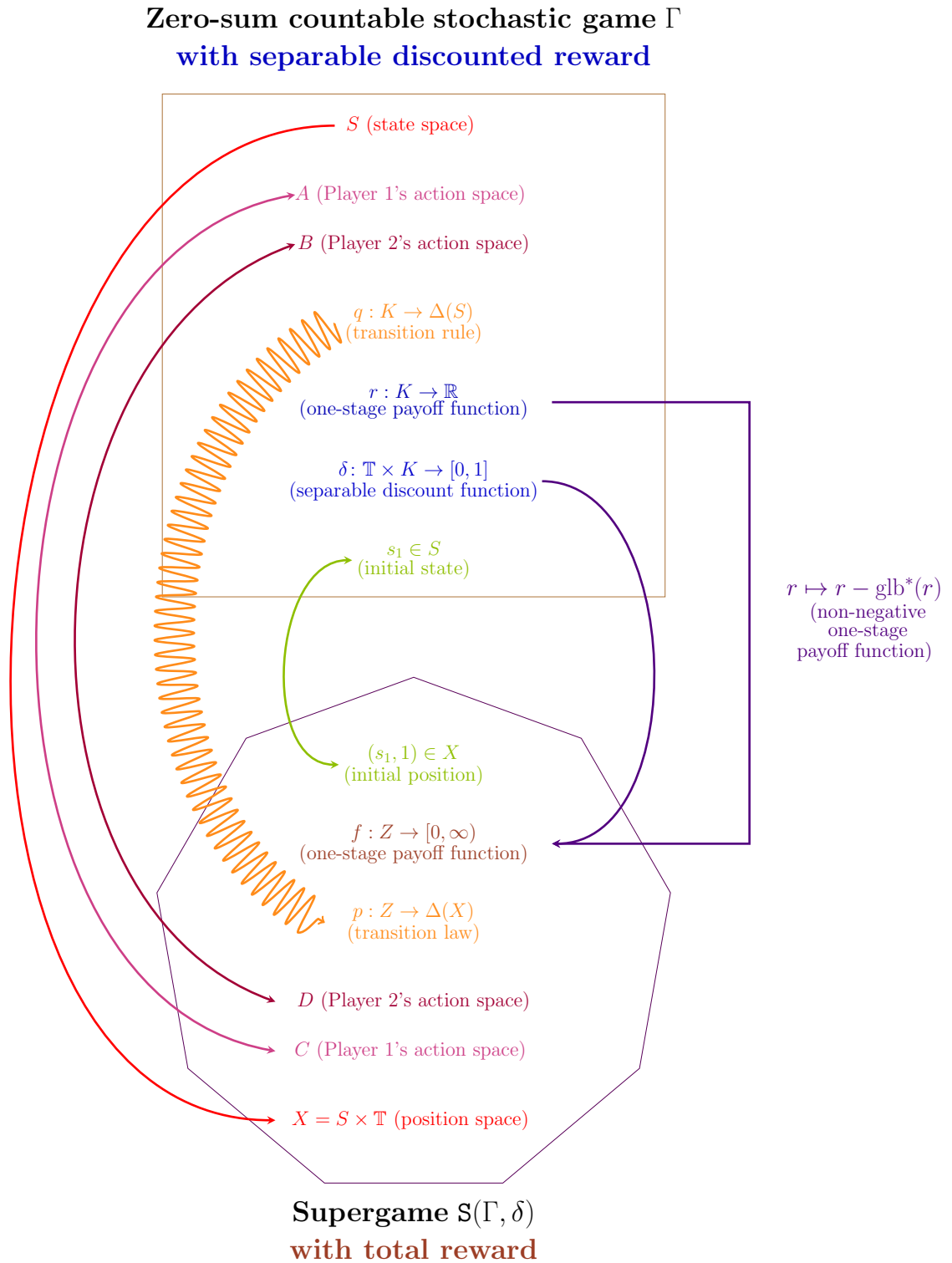


Figure 8: Graphical representation of the steps of constructing a supergame from a zero-sum countable stochastic game with separable discounting.

First, note that the zero-sum countable stochastic game  $\Gamma$  in Definition 3.9 has exactly one supergame  $S(\Gamma, \delta)$ . This uniqueness is a direct consequence of the fact that each element of  $S(\Gamma, \delta)$  is uniquely determined by the corresponding elements of the original zero-sum countable stochastic game  $\Gamma$  along with the separable discount function  $\delta$ .

Second, note that the position space  $X$  of the supergame  $S(\Gamma, \delta)$  is a (nonempty) countable set, even if the original zero-sum countable stochastic game  $\Gamma$  has a nonempty finite set of states.

Third, because we make a close and strong relationship between the original zero-sum countable stochastic game  $\Gamma$  and its corresponding supergame  $S(\Gamma, \delta)$ , the set of available actions for the players at any given position is clearly defined. For instance, in Formula (3.4), if the action  $c$  belongs to  $C(s, t)$ , it implies  $c \in A(s)$ , since set of available actions for Player 1 at position  $(s, t) \in X$  matches the set of available actions for Player 1 in state  $s \in S$ .

Figure 8 illustrates how the supergame is constructed. Both Definition 3.9 and Figure 8 implicitly rely on a foundational transformation. In particular, the original zero-sum countable stochastic game  $\Gamma = \langle S, A, B, q, r \rangle$  is converted into a *positive countable stochastic game*  $\hat{\Gamma} = \langle S, A, B, q, r^* \rangle$ , where  $r^* = r - \text{glb}^*(r)$ . Although this transformation is not explicitly described in Definition 3.9, it is present there via Formula (3.5).

It is important to note that if the original zero-sum countable stochastic game  $\Gamma$  possesses a separable discounted value, the transformation given by Formula (3.5) can potentially increase that value. However, this does not affect the optimality of the corresponding strategies. Furthermore, if  $\Gamma$  is already a positive countable stochastic game, then  $\text{glb}^*(r) = 0$ , and no modification of the one-stage payoff function  $r$  is required.

Finally, it is important to note that Player 1 seeks to maximise the expected value of the *separable discounted total payoff* in  $\Gamma$  and  $\hat{\Gamma}$ . By contrast, in  $S(\Gamma, \delta)$ , Player 1 strives to maximise the expected value of the *total payoff*. Despite this difference in their objectives, both approaches essentially tackle the same problem because of the relationship between  $\hat{\Gamma}$  and  $S(\Gamma, \delta)$  in their construction.

In the next step, we describe the dynamics of supergames. The supergame  $S(\Gamma, \delta)$  in Definition 3.9 starts from the initial position  $x_1 = (s_1, 1) \in X$ , and the following events take place at each stage  $n \in \mathbb{T}$ :

Step 1. The players observe the current position  $x_n \in X$ .

Step 2. Player 1 chooses an action  $c_n \in C(x_n)$  and Player 2 chooses an action

$d_n \in D(x_n)$  simultaneously and independently.

Step 3. The action profile  $(c_n, d_n)$  is communicated to the players.

Step 4. The current position  $x_n$  and the action profile  $(c_n, d_n)$  induce the one-stage payoff  $f(x_n, c_n, d_n)$  for Player 1. The one-stage payoff for Player 2 is  $-f(x_n, c_n, d_n)$ .

Step 5. A new position  $x_{n+1} \in X$  is drawn according to the transition law  $p(\cdot | x_n, c_n, d_n)$ , and the game proceeds at position  $x_{n+1}$ . Go back to Step 1..

Next, we highlight some properties of supergames. Lemma 3.10 is fundamental, as it situates supergames within the framework of two-person countable stochastic games.

**Lemma 3.10.** *For any zero-sum countable stochastic game  $\Gamma$  with the initial state  $s_1$  and separable discounting, the corresponding supergame  $\mathbb{S}(\Gamma, \delta)$  is a positive zero-sum countable stochastic game starting at the initial position  $(s_1, 1)$ .*

*Proof.* Consider the zero-sum countable stochastic game  $\Gamma = \langle S, A, B, q, r \rangle$  with an initial state  $s_1 \in S$  and a fixed separable discount function  $\delta$ .

The position space  $X$  of the associated supergame  $\mathbb{S}(\Gamma, \delta)$  is countable, as it is the Cartesian product of the nonempty countable sets  $S$  and  $\mathbb{T}$ .

Moreover, the one-stage payoff function  $f$  defined in Formula (3.5) is bounded. This property follows from the boundedness of the one-stage payoff function  $r$  and the separable discount function  $\delta$ , along with the fact that  $\text{glb}^*(r)$  is a real number.

Based on these observations, it is easy to see that  $\mathbb{S}(\Gamma, \delta)$  is a zero-sum countable stochastic game. It is also clear that  $f$  is a non-negative function.  $\square$

Building on Lemma 3.10, we define *histories*, *plays*, and *strategies* for supergames in the same way as in Subsection 3.1.1. For the supergame  $\mathbb{S}(\Gamma, \delta)$ , let  $\Phi$ ,  $\Phi_M$ , and  $\Phi_S$  denote the sets of all behavioural, all Markov, and all stationary strategies for Player 1, respectively. Correspondingly, for Player 2 in the same supergame  $\mathbb{S}(\Gamma, \delta)$ , we denote the sets of all behavioural, all Markov, and all stationary strategies by  $\Omega$ ,  $\Omega_M$ , and  $\Omega_S$ .

For any zero-sum countable stochastic game  $\Gamma$  with the initial state  $s_1$  and separable discounting, consider the associated supergame  $\mathbb{S}(\Gamma, \delta)$  starting at the initial position  $(s_1, 1) \in X$ . Given a strategy profile  $(\varphi, \omega) \in \Phi \times \Omega$  and the transition law  $p$ , there exists a unique probability measure  $\mathbb{P}_{x_1}^{\varphi, \omega}$  on the measurable space  $(H_\infty, \mathcal{H})$ . The corresponding expectation operator is denoted by  $\mathbb{E}_{x_1}^{\varphi, \omega}$ .

Next, we introduce the concept of *total reward* in the context of the supergame. This concept serves as the basis for defining the *lower total value*, the *upper total value*, and ultimately, the *total value*. To proceed, we begin with the following important observation:

*Remark 3.11.* For every play  $h_\infty \in H_\infty$ , the total payoff in Formula (3.6) is an element of  $\overline{\mathbb{R}}_+$ .

In the next step, we introduce the concept of total value in supergames.

**Definition 3.12.** Consider the zero-sum countable stochastic game  $\Gamma = \langle S, A, B, q, r \rangle$  with initial state  $s_1 \in S$  and a separable discount function  $\delta$ . In the associated supergame  $\mathbb{S}(\Gamma, \delta)$ , starting at the initial position  $x = (s_1, 1) \in X$ , the total reward for Player 1 under the behavioural strategy profile  $(\varphi, \omega) \in \Phi \times \Omega$  is defined as

$$\gamma(\varphi, \omega)(x) = \mathbb{E}_x^{\varphi, \omega} \left[ \sum_{m=1}^{\infty} f(x_m, c_m, d_m) \right]. \quad (3.7)$$

Theorem 3.13 is an immediate consequence of Definitions 3.3 and 3.9, together with Lemma 3.10. The property stated in Formula (3.8) plays a key role in positive zero-sum countable stochastic games and is commonly referred to as Strauch's property (Jaśkiewicz and Nowak, 2018, p. 239).

**Theorem 3.13.** Consider the zero-sum countable stochastic game  $\Gamma$  with an initial state  $s_1$  and a separable discount function  $\delta$ . The associated supergame  $\mathbb{S}(\Gamma, \delta)$  is a positive zero-sum countable stochastic game with the total reward criterion. Furthermore, it holds that

$$\sup_{\varphi, \omega} \gamma(\varphi, \omega)(x) \leq M_\delta \cdot \text{lub}(f) < \infty, \quad (3.8)$$

where  $x = (s_1, 1) \in X$ ,  $M_\delta \in \mathbb{R}$  is the bound for the separable discount function  $\delta$  in Formula (3.2), and  $\text{lub}(f)$  denotes the least upper bound of the bounded function  $f$ .

We conclude this subsection by introducing the *lower total value*, the *upper total value* and the *total value* for supergames.

**Definition 3.14.** Let  $\Gamma$  be a zero-sum countable stochastic game with an initial state  $s_1$  and a separable discount function  $\delta$ . Consider the associated supergame  $\mathbb{S}(\Gamma, \delta)$ , which starts at the initial position  $x = (s_1, 1) \in X$ . The lower total value of  $\mathbb{S}(\Gamma, \delta)$  is defined by

$$\underline{w}(x) = \sup_{\varphi \in \Phi} \inf_{\omega \in \Omega} \gamma(\varphi, \omega)(x). \quad (3.9)$$

Likewise, the upper total value of  $\mathbb{S}(\Gamma, \delta)$  is given by

$$\overline{w}(x) = \inf_{\omega \in \Omega} \sup_{\varphi \in \Phi} \gamma(\varphi, \omega)(x). \quad (3.10)$$

If the lower and upper total values are equal, i.e.,  $\underline{w}(x) = \overline{w}(x)$ , then the supergame  $S(\Gamma, \delta)$  is said to have a total value, denoted by  $w(x)$ .

The initial position of a supergame is always explicitly set by the initial state of the original countable zero-sum stochastic game from which it is derived. Accordingly, as shown in Definition 3.14, the lower total value, upper total value, and total value are all defined based on this initial point.

### 3.1.4 Shapley operators for supergames

This subsection presents some properties of supergames based on the work of Flesch et al. (2018, 2020) on positive zero-sum countable stochastic games with total reward (for more details on Shapley operators, see Neyman (2003B); Sorin (2003C)).

Let  $\Gamma$  be an arbitrary zero-sum countable stochastic game with the initial state  $s_1 \in S$  and the separable discount function  $\delta$ . Although it is currently unknown under what conditions the associated supergame  $S(\Gamma, \delta)$  possesses a total value, it follows that if such a total value exists for the supergame  $S(\Gamma, \delta)$ , then the original zero-sum countable stochastic game  $\Gamma$  admits a  $\delta$ -separable discounted value in the initial state  $s_1 \in S$ . Positive zero-sum countable stochastic games do not necessarily possess a total value. This fact also applies to supergames, a subclass of these games, as noted in Theorem 3.13.

First, we introduce *one-shot games*, which are derived from the supergame  $S(\Gamma, \delta)$ . Next, we define the *lower Shapley operator* and *upper Shapley operator* for these one-shot games. The subsection concludes with an exploration of key properties of these Shapley operators.

In this subsection, we adopt the following notations. By Remark 3.11, let  $\mathcal{L}_+(X)$  denote the space of all functions from the position space  $X$  to  $\overline{\mathbb{R}}_+$ . The lower total value, namely the function  $x \mapsto \underline{w}(x)$  in Formula (3.9), is an element of  $\mathcal{L}_+(X)$ . Similarly, the upper total value in Formula (3.10) also belongs to  $\mathcal{L}_+(X)$ .

First, we introduce the one-shot games. The central idea behind one-shot games is to consider an arbitrary zero-sum countable stochastic game  $\Gamma$  with an initial state  $s_1 \in S$  and a separable discount function  $\delta$ . From this setup, we construct the associated supergame  $S(\Gamma, \delta)$ , which starts at the initial position  $x = (s_1, 1) \in X$ . The one-shot game is then defined by stopping the supergame after its first stage. For this truncation, Player 1 receives a non-negative terminal payoff from Player 2, representing compensation for completing only the initial stage of the supergame. In the following step, we formalise this construction.

**Definition 3.15.** Let  $\Gamma = \langle S, A, B, q, r \rangle$  be a zero-sum countable stochastic game starting at the initial state  $s_1 \in S$ , and let  $\delta$  denote a separable discount function.

A one-shot game  $M(x, u)$  derived from the supergame  $S(\Gamma, \delta) = \langle X, C, D, p, f \rangle$  is a tuple

$$\langle x, C(x), D(x), u \rangle$$

which is played by the same players and consists of the following components:

- (a)  $x = (s_1, 1)$  is the initial position of the supergame  $S(\Gamma, \delta)$ .
- (b) The set  $C(x)$  represents the set of available actions for Player 1 at the initial position  $x \in X$ .
- (c) The set  $D(x)$  represents the set of available actions for Player 2 at the initial position  $x \in X$ .
- (d)  $u \in \mathcal{L}_+$  represents the continuation payoff function for Player 1.

The dynamics of the one-shot game  $M(x, u)$  in Definition 3.15 is as follows: at the end of the initial stage, Player 2 pays a non-negative payoff defined by

$$f(x, c, d) + \sum_{x' \in X} u(x')p(x' | x, c, d)$$

if the action profile  $(c, d) \in C(x) \times D(x)$  has been chosen. After this, the one-shot game  $M(x, u)$  ends.

It is important to emphasize that, while we explicitly define the zero-sum countable stochastic game  $\Gamma$  with its initial state  $s_1$  and the separable discount function  $\delta$ , which determines the associated supergame  $S(\Gamma, \delta)$ , Definition 3.15 introduces a family of one-shot games, where each member of this family is determined by a specific choice of the *continuation payoff function*  $u \in \mathcal{L}_+$ . One such condition is that one player's set of available actions at the initial state is (Flesch et al., 2020, p. 506).

In the following set, we introduce the notions of the *lower Shapley operator* and the *upper Shapley operator* for one-shot games. Given a one-shot game  $M(x, u)$ , the *lower Shapley operator*  $\mathcal{C}: \mathcal{L}_+ \rightarrow \mathcal{L}_+$  is defined by

$$(\mathcal{C}u)(x) = \sup_{\mu \in \Delta(C(x))} \inf_{\nu \in \Delta(D(x))} \sum_{\substack{c \in C(x) \\ d \in D(x)}} \left\{ f(x, c, d) + \sum_{x' \in X} u(x')p(x' | x, c, d) \right\} \mu(c)\nu(d). \quad (3.11)$$

In a similar way, we introduce the *upper Shapley operator*  $\mathcal{D}: \mathcal{L}_+ \rightarrow \mathcal{L}_+$ , defined as follows:

$$\begin{aligned}
 (\mathcal{D}u)(x) = & \inf_{\nu \in \Delta(D(x))} \sup_{\mu \in \Delta(C(x))} \sum_{\substack{c \in C(x) \\ d \in D(x)}} \\
 & \left\{ f(x, c, d) + \sum_{x' \in X} u(x') p(x' \mid x, c, d) \right\} \mu(c) \nu(d).
 \end{aligned} \tag{3.12}$$

It follows immediately that both the lower Shapley operator  $\mathcal{C}$  in Formula (3.11) and the upper Shapley operator  $\mathcal{D}$  in Formula (3.12) is monotonic: for any functions  $g_1, g_2 \in \mathcal{L}_+$  with  $g_1 \leq g_2$ , it follows that  $\mathcal{C}g_1 \leq \mathcal{C}g_2$  and  $\mathcal{D}g_1 \leq \mathcal{D}g_2$ .

**Definition 3.16.** *A function  $g_1 \in \mathcal{L}_+$  is a fixed point of the lower Shapley operator  $\mathcal{C}$  in Formula (3.11) if  $\mathcal{C}g_1 = g_1$ . Similarly, a function  $g_2 \in \mathcal{L}_+$  is a fixed point of the upper Shapley operator  $\mathcal{D}$  in Formula (3.12) if  $\mathcal{D}g_2 = g_2$ .*

It is well-known that if  $g \in \mathcal{L}_+$  is a fixed point of the lower Shapley operator  $\mathcal{C}$  in Formula (3.11), then for any real number  $c$  such that  $g + c \in \mathcal{L}_+$ , the function  $g + c$  is also a fixed point of the lower Shapley operator  $\mathcal{C}$ . An analogous statement holds for the upper Shapley operator  $\mathcal{D}$  in Formula (3.12).

We conclude this subsection by presenting some properties of the lower and upper Shapley operators. Before doing so, it is essential to introduce the following concepts. A function  $g_1 \in \mathcal{L}_+$  is called  *$\mathcal{C}$ -excessive* if  $\mathcal{C}g_1 \leq g_1$ . Similarly, a function  $g_2 \in \mathcal{L}_+$  is called  *$\mathcal{D}$ -excessive* if  $\mathcal{D}g_2 \leq g_2$ .

**Lemma 3.17.** *Let  $\Gamma = \langle S, A, B, q, r \rangle$  be a zero-sum countable stochastic game starting at the initial state  $s_1 \in S$ , and let  $\delta$  denote a separable discount function.*

*The lower total value  $\underline{w}(x)$  of the associated supergame  $\mathbb{S}(\Gamma, \delta)$  in Formula (3.9) has the following properties:*

- C1) *It is the least fixed point of the lower Shapley operator  $\mathcal{C}$  in Formula (3.11).*
- C2) *It is the least  $\mathcal{C}$ -excessive function.*

*Similarly, the upper total value  $\overline{w}(x)$  of  $\mathbb{S}(\Gamma, \delta)$  in Formula (3.10) has the following properties:*

- D1) *It is the least fixed point of the upper Shapley operator  $\mathcal{D}$  in Formula (3.12).*
- D2) *is the least  $\mathcal{D}$ -excessive function.*

Flesch et al. proved a generalised version of Theorem 3.17 for positive countable stochastic games with total reward (Flesch et al., 2020, Section 4, pp. 506–511). According to Theorem 3.13, Theorem 3.17 directly follows from these results.

### 3.1.5 Results for zero-sum countable stochastic games with separable discounting

This subsection presents our findings on zero-sum countable stochastic games with separable discounting.

**Theorem 3.18.** *Let  $\Gamma = \langle S, A, B, q, r \rangle$  be a zero-sum countable stochastic game with separable discounting.*

- 1) *If either  $A(s)$  or  $B(s)$  is finite for every state  $s \in S$ , then  $\Gamma$  admit a value.*
- 2) *If  $B(s)$  is finite for all states  $s \in S$ , then the following hold:*
  - *For every  $\varepsilon > 0$ , Player 1 has an  $\varepsilon$ -optimal Markov strategy.*
  - *Player 2 has an optimal Markov strategy.*

*Sketch of the proof.* 1) Let  $\Gamma = \langle S, A, B, q, r \rangle$  be a fixed zero-sum countable stochastic game starting at the initial state  $s_1 \in S$ , and let  $\delta$  denote a separable discount function. The associated supergame  $S(\Gamma, \delta) = \langle X, C, D, p, f \rangle$  thus begins at the initial position  $x = (s_1, 1) \in X$ . For any  $u \in \mathcal{L}_+$ , let  $M(x, u)$  denote the one-shot game derived from the supergame  $S(\Gamma, \delta)$ .

Flesch et al. (2020, Theorems 2 and 3, pp. 503–505) proved that if either  $C(x)$  or  $D(x)$  is a finite set, then the one-shot game  $M(x, u)$  possesses a value for every  $u \in \mathcal{L}_+$ . This result means that the lower Shapley operator  $\mathcal{C}$  in Formula (3.11) and the upper Shapley operator  $\mathcal{D}$  in Formula (3.12) coincide. To simplify notation, we define  $\mathcal{T} := \mathcal{C} = \mathcal{D}$  and refer to it as the *Shapley operator*.

By Theorem 3.17, both the lower total value  $\underline{w}(x)$  and the upper total value  $\overline{w}(x)$  are the least fixed point of the Shapley operator  $\mathcal{T}$ . Therefore, we have  $\underline{w}(x) = \overline{w}(x)$ , indicating that the supergame  $S(\Gamma, \delta)$  admits a total value. As a result, the zero-sum countable stochastic game  $\Gamma$  possesses a  $\delta$ -separable discounted reward, by the definition of the supergame  $S(\Gamma, \delta)$ .

2) For Player 2, following the argument presented in Point 2 of Theorem 12 in Flesch et al. (2020, p. 513), it can be shown that an optimal stationary strategy exists in the supergame  $S(\Gamma, \delta)$ . This strategy corresponds to an optimal Markov strategy in the original zero-sum countable stochastic game  $\Gamma$  with separable discounting since each position  $x = (s, t) \in X$  is uniquely defined by a pair of the state  $s \in S$  and the stage index  $t \in \mathbb{T}$ .

In the case of Player 1, we can follow the proof from Point 3 of Theorem 12 of Flesch et al. (2020) since the supergame  $S(\Gamma, \delta)$  has the Strauch's property (see Formula (3.8)). This implies that for every  $\varepsilon > 0$ , Player 1 has an  $\varepsilon$ -optimal stationary

strategy in the supergame  $S(\Gamma, \delta)$ . Consequently, Player 1 has a  $\varepsilon$ -optimal Markov strategy in the zero-sum countable stochastic game  $\Gamma$  with separable discounting for every  $\varepsilon > 0$ .  $\square$

In Point 2) of Theorem 3.18, the assumption that Player 2 has only finitely many available actions in each state is essential. By focusing on this condition, we can directly derive the following result from Theorem 3.18:

**Corollary 3.19.** *Let  $\Gamma$  be a zero-sum finite stochastic game with separable discounting. Then the following properties hold:*

- 1)  $\Gamma$  admits a value.
- 2) Player 1 has an optimal Markov strategy.
- 3) Player 2 has an optimal Markov strategy.

Finally, we present the following result, which is a direct consequence of Example 3.8. It is important to emphasise that Lemma 3.20 rules out the possibility of optimal stationary strategies in general. However, this does not contradict Theorem 3.18, as that theorem specifically addresses the optimality of Markov strategies.

**Lemma 3.20.** *There exists a zero-sum countable stochastic game with separable discounting for which no player has an optimal stationary strategy.*

*Proof.* See Examples 3.7 or 3.8.  $\square$

We conclude this subsection with a summary of our answers to Research Questions (RQ3), (RQ4), and (RQ5).

Regarding Research Question (RQ3), our results give a positive answer for zero-sum finite stochastic games, as demonstrated by Corollary 3.19. Similarly, for zero-sum countable stochastic games, the answer to Research Question (RQ3) is also positive, provided that the condition in Point 1) of Theorem 3.18 holds.

We provide the following answers to Research Questions (RQ4) and (RQ5). In every zero-sum finite stochastic game with separable discounting, both Player 1 and Player 2 have an optimal Markov strategy (see Corollary 3.19). In contrast, in every zero-sum countable stochastic game with separable discounting, Player 2 has an optimal Markov strategy, while Player 1 has an  $\varepsilon$ -optimal Markov strategy for any  $\varepsilon > 0$ , provided that a specific condition is satisfied (see Point 2) of Theorem 3.18).

## 3.2 Zero-sum infinite stochastic games with separable discounting

In the remainder of this chapter, we study *zero-sum infinite stochastic games* (or zero-sum stochastic games with a general state space) with *separable discounting*. Before we dive into this analysis, we introduce the necessary notations and provide the key definitions and remarks (for further details, see Laczkovich (1995); Kechrin (1995); Srivastava (1998); Bertsekas and Shreve (1996); Nowak (1984A,B, 2003); Jaśkiewicz and Nowak (2018)).

Let  $(S, \mathcal{S})$  be a measurable space, and let  $X$  denote a separable metric space equipped with the Borel  $\sigma$ -algebra  $\mathcal{B}(X)$ . We denote the set of all probability measures on  $(S, \mathcal{S})$  by  $\Delta(S)$ , and the set of all bounded measurable functions  $f: S \rightarrow \mathbb{R}$  by  $\mathbb{M}_b(S)$ .

Let  $C(X)$  denote the set of all continuous functions in  $\mathbb{M}_b(X)$ . Similarly,  $\overline{C}(X)$  denotes the set of all upper semicontinuous functions in  $\mathbb{M}_b(X)$ , and  $\underline{C}(X)$  denotes the set of all lower semicontinuous functions in  $\mathbb{M}_b(X)$ . Let  $\mathcal{K}(X)$  denote the family of compact subsets of  $X$ .

Given any family  $\mathcal{F} \subset \mathbb{M}_b(S)$ , we equip  $\Delta(S)$  with the  $\mathcal{F}$ -topology, defined as the coarsest topology in which every mapping  $\mu \mapsto \int f \, d\mu$  is continuous for all  $f \in \mathcal{F}$ . When  $\mathcal{F} = \mathbb{M}_b(S)$ , this topology is called the *MB-topology* on  $\Delta(S)$ .

A *Polish space* is a metric space that is both complete and separable. A *standard Borel space* is a nonempty Borel subset of a Polish space. An *analytic space* is any nonempty space that arises as the continuous image of a Polish space.

Let  $(X_1, \mathcal{F}_1)$  and  $(X_2, \mathcal{F}_2)$  be two measurable spaces.  $\mathbb{Q}(X_2 \mid X_1)$  denotes the set of all conditional probability measures on  $(X_2, \mathcal{F}_2)$  conditioned on  $(X_1, \mathcal{F}_1)$ .

A set-valued function  $F: S \rightrightarrows \mathcal{P}(X)$  is (*lower*) *measurable* if, for every open subset  $O$  of  $X$ , it holds that  $F^{-1}(O) = \{s \in S: F(s) \cap O \neq \emptyset\} \in \mathcal{S}$ .

### 3.2.1 The game model

This subsection provides a brief overview of the model of *zero-sum stochastic games* with a *general state space*, and presents the key related concepts (Nowak, 1984A, Section 2).

First, we formally define these stochastic games. Since our main aim is to analyse zero-sum stochastic games with a general state space under separable discounting, we introduce them within a general framework of zero-sum infinite stochastic games.

**Definition 3.21.** A zero-sum infinite stochastic game (or zero-sum infinite stochastic game with a general state space) is a tuple

$$\langle (S, \mathcal{S}), X, Y, A, B, q, r \rangle,$$

which is played by Player 1 and Player 2 and consists of the following components:

- (a) The pair  $(S, \mathcal{S})$  denotes a measurable space, where  $S$  is a nonempty state space.
- (b) The separable metric spaces  $X$  and  $Y$  are the action spaces for Player 1 and Player 2, respectively.
- (c) The measurable set-valued function  $A: S \rightrightarrows \mathcal{K}(X)$  assigns to each state  $s \in S$  a nonempty compact subset  $A(s)$  of  $X$ , representing the available actions for Player 1 in state  $s \in S$ .
- (d) The measurable set-valued function  $B: S \rightrightarrows \mathcal{K}(Y)$  assigns to each state  $s \in S$  a nonempty compact subset  $B(s)$  of  $Y$ , representing the available actions for Player 2 in state  $s \in S$ .
- (e) The function  $q \in \mathbb{Q}(S \mid S \times X \times Y)$  is the transition rule that governs the evolution of states based on the current state and chosen actions.
- (f) The function  $r \in \mathbb{M}_b(S \times X \times Y)$  is the one-stage payoff function for Player 1.

Like zero-sum countable stochastic games (see Definition 3.1), Definition 3.21 does not describe a single zero-sum infinite stochastic game but presents a family of specific zero-sum infinite stochastic games. The only difference between the games in this family is their initial state.

Next, we describe the dynamics of zero-sum infinite stochastic games. Each game starts from an initial state  $s_1 \in S$ . In every stage  $t \in \mathbb{T}$ , the following sequence of events takes place:

- Step 1. Both players observe the current state  $s_t \in S$ .
- Step 2. Player 1 and Player 2 simultaneously and independently select actions  $a_t \in A(s_t)$  and  $b_t \in B(s_t)$ , respectively.
- Step 3. The chosen action profile  $(a_t, b_t)$  is then disclosed to both players.
- Step 4. Given the current state  $s_t$  and the selected action profile  $(a_t, b_t)$ , Player 1 receives the one-stage payoff  $r(s_t, a_t, b_t)$ , while Player 2 receives the payoff  $-r(s_t, a_t, b_t)$ .

Step 5. The next state  $s_{t+1} \in S$  is drawn according to the transition rule  $q(\cdot | s_t, a_t, b_t)$ , and the game proceeds at state  $s_{t+1}$ . Go back to Step 1..

Consider an arbitrary zero-sum infinite stochastic game  $\langle (S, \mathcal{S}), X, Y, A, B, q, r \rangle$ . We define the set of *stories* of length  $t \in \mathbb{T}$  recursively as follows:

$$\mathbb{H}_t = \begin{cases} S, & \text{if } t = 1, \\ S \times X \times Y \times \mathbb{H}_{t-1}, & \text{if } t > 1, \end{cases}$$

and let

$$\mathbb{H} = (S \times X \times Y)^{\mathbb{T}}.$$

Similarly, the set of *histories* of length  $t \in \mathbb{T}$  is defined by:

$$H_t = \begin{cases} S, & \text{if } t = 1, \\ \text{Gr}(A \times B) \times H_{t-1}, & \text{if } t > 1, \end{cases}$$

where  $\text{Gr}(A \times B) = \{(s, a, b) : a \in A(s) \text{ and } b \in B(s)\}$ . Finally, the set of all *plays* is defined as:

$$H_\infty = \bigcap_{t=1}^{\infty} (H_t \times \mathbb{K}^{\mathbb{T}}),$$

where  $\mathbb{K} = X \times Y \times S$ .

*Remark 3.22.* For each stage  $t \in \mathbb{T}$ , the set  $H_t$  is a measurable subset of  $\mathbb{H}_t$ , where  $\mathbb{H}_t$  is equipped with the product  $\sigma$ -algebra. Similarly,  $H_\infty$  is a measurable subset of  $\mathbb{H}$ , also endowed with the product  $\sigma$ -algebra (Nowak, 1984A, p. 20).

This subsection concludes with an overview of key strategic concepts, followed by a technical appendix to support the introduction of reward functions. First, we introduce *behavioural strategies*. Notably, we define behavioural strategies using *stories* rather than *histories*, to avoid potential technical complications in *supergames* later on.

**Definition 3.23.** A sequence  $\pi = (\pi_1, \pi_2, \dots)$  is a behavioural strategy for Player 1 if it meets the following conditions for every stage  $t \in \mathbb{T}$  and every story  $h_t \in \mathbb{H}_t$ :

$$\pi_t \in \mathbb{Q}(X | \mathbb{H}_t) \quad \text{and} \quad \pi_t(A(\kappa(h_t)) | h_t) = 1,$$

where  $\kappa(h_t)$  denotes the final state of the story  $h_t$ .

We similarly define behavioural strategies for Player 2. Let  $\Pi$  and  $\Theta$  denote sets of all behavioural strategies for Player 1 and Player 2, respectively.

Next, we introduce Markov strategies. Under a Markov strategy, the player's choice at any stage depends exclusively on the current stage and the current state, without considering the story that led to that state.

**Definition 3.24.** A behavioural strategy  $\pi$  is called a Markov strategy for Player 1 if, for every stage  $t \in \mathbb{T}$ ,

$$\pi_t \in \mathbb{Q}(X | S).$$

Before exploring stationary strategies, we introduce several useful concepts: mixed action mappings and their selectors. Consider a zero-sum infinite stochastic game  $\langle (S, S), X, Y, A, B, q, r \rangle$ . For each state  $s \in S$ , let  $\Delta(A(s))$  and  $\Delta(B(s))$  denote the sets of *mixed actions* in state  $s \in S$  for Player 1 and Player 2, respectively. Additionally, we define *mixed action mappings* as follows:

$$\Delta_A(s) = \Delta(A(s)) \quad \text{and} \quad \Delta_B(s) = \Delta(B(s))$$

for each state  $s \in S$ .

*Remark 3.25.*  $\Delta_A$  and  $\Delta_B$  are set-valued functions with compact values, and they are also measurable (Nowak, 1984A, p. 19).

A selector of the mixed action mapping  $\Delta_A$  is a function  $f$  such that  $f(s) \in \Delta_A(s)$  for each state  $s \in S$ . Selectors for the mixed action mapping  $\Delta_B$  are defined similarly.

Next, we define stationary strategies.

**Definition 3.26.** A stationary strategy  $\pi$  for Player 1 is a Markov strategy that satisfies

$$\pi_t = f$$

for every stage  $t \in \mathbb{T}$ , where  $f$  is a measurable selector of the mixed action mapping  $\Delta_A$ .

Markov strategies and stationary strategies for Player 2 are defined analogously. Let  $\Pi_M$  and  $\Theta_M$  represent the sets of all Markov strategies for Player 1 and Player 2, respectively. Likewise, let  $\Pi_S$  and  $\Theta_S$  denote the sets of all stationary strategies for Player 1 and Player 2, respectively.

*Remark 3.27.* It is known that  $\Pi_S$  and  $\Theta_S$  are nonempty sets (Nowak, 1984A, Remark, p. 19).

Using Ionescu-Tulcea Theorem (Bertsekas and Shreve, 1996, Proposition 7.45), for every strategy profile  $(\pi, \vartheta) \in \Pi \times \Theta$ , given the initial state  $s \in S$  and the transition rule  $q$ , there exists a unique probability measure  $\mathbb{P}_s^{\pi, \vartheta}$  defined on the product  $\sigma$ -algebra of the infinite product  $\mathbb{K}^{\mathbb{T}}$ . Let  $\mathbb{E}_s^{\pi, \vartheta}$  denote the corresponding expectation operator.

### 3.2.2 Zero-sum infinite stochastic games with separable discounting

This subsection presents *separable discounting* for zero-sum infinite stochastic games. We define *separable discount functions*, followed by the introduction of *separable discounted reward* and *separable discounted value* for zero-sum infinite stochastic games. Finally, we classify strategies based on their optimality with respect to the separable discounted value. This material is very similar to that in Subsection 3.1.2, but here we develop it for zero-sum infinite stochastic games.

First, we define separable discount functions.

**Definition 3.28.** *Let  $\Gamma$  be a zero-sum infinite stochastic game from Definition 3.21. A function*

$$\delta: \mathbb{T} \times S \times X \times Y \rightarrow [0, 1]$$

*is called a separable discount function if there exists a number  $M_\delta \in \mathbb{R}$  such that for every play  $h_\infty = (s_1, a_1, b_1, s_2, a_2, b_2, \dots) \in H_\infty$ , the following holds:*

$$\sum_{t=1}^{\infty} \delta(t, s_t, a_t, b_t) \leq M_\delta. \quad (3.13)$$

Using separable discount functions, we introduce the *separable discounted reward*. This concept is analogous to the approach used in zero-sum countable stochastic games (see Definition 3.4).

**Definition 3.29.** *Let  $\Gamma$  be a zero-sum infinite stochastic, starting at the initial state  $s \in S$ , and let  $\delta$  be a separable discount function.*

*The  $\delta$ -separable discounted reward for Player 1 under the behavioural strategy profile  $(\pi, \vartheta) \in \Pi \times \Theta$  is defined as*

$$\gamma_\delta(\pi, \vartheta)(s) = \mathbb{E}_s^{\pi, \vartheta} \left[ \sum_{t=1}^{\infty} \delta(t, s_t, a_t, b_t) r(s_t, a_t, b_t) \right].$$

In the next step, we introduce the concepts of the *lower separable discounted value*, the *upper separable discounted value*, and the *separable discounted value* for zero-sum infinite stochastic games.

**Definition 3.30.** *Let  $\Gamma$  be a zero-sum infinite stochastic game, and let  $\delta$  be a separable discount function.*

*The lower and upper  $\delta$ -separable discounted values of  $\Gamma$  for the initial state  $s \in S$  are*

$$\underline{v}_\delta(s) = \sup_{\pi \in \Pi} \inf_{\vartheta \in \Theta} \gamma_\delta(\pi, \vartheta)(s) \quad \text{and} \quad \bar{v}_\delta(s) = \inf_{\vartheta \in \Theta} \sup_{\pi \in \Pi} \gamma_\delta(\pi, \vartheta)(s).$$

If the lower  $\delta$ -separable discounted value  $\underline{v}_\delta(s)$  equals the upper  $\delta$ -separable discounted value  $\bar{v}_\delta(s)$  for a given initial state  $s \in S$ , then  $\Gamma$  has a  $\delta$ -separable discounted value in the initial state  $s \in S$ , denoted by  $v_\delta(s)$ .

Furthermore, if  $\Gamma$  has a separable discounted value for every initial state  $s \in S$ , then  $\Gamma$  has a  $\delta$ -separable discounted value.

We conclude this subsection by classifying the behavioural strategies based on their optimality.

**Definition 3.31.** Let  $\Gamma$  be a zero-sum infinite stochastic game from Definition 3.21 which admits a  $\delta$ -separable discounted value  $v_\delta$ , where  $\delta$  denotes a separable discount function.

For any  $\varepsilon \geq 0$ , a behavioural strategy  $\pi \in \Pi$  for Player 1 is said to be  $\varepsilon$ -optimal at the initial state  $s \in S$ , if

$$\gamma_\delta(\pi, \vartheta)(s) \geq v_\delta(s) - \varepsilon$$

for every behavioural strategy  $\vartheta \in \Theta$  for Player 2.

Likewise, for any  $\varepsilon \geq 0$ , a behavioural strategy  $\vartheta \in \Theta$  for Player 2 is called  $\varepsilon$ -optimal at the initial state  $s \in S$ , if

$$\gamma_\delta(\pi, \vartheta)(s) \leq v_\delta(s) + \varepsilon$$

for every behavioural strategy  $\pi \in \Pi$  for Player 1.

A behavioural strategy is  $\varepsilon$ -optimal if it is  $\varepsilon$ -optimal for every initial state.

In the context of zero-sum infinite stochastic games, we use the term *optimal behavioural strategy* as a shorthand for a 0-optimal behavioural strategy.

### 3.2.3 Zero-sum Borel, Suslin, and Nowak stochastic games

This subsection introduces three distinct subclasses of zero-sum infinite stochastic games with separable discounting. These subclasses are based on the observation that supergames can be formulated for zero-sum infinite stochastic games with separable discounting, similar to the approach discussed in Subsection 3.1.3. We introduce these three special classes of games because their corresponding supergames belong to the class of positive zero-sum infinite stochastic games with total reward, which have already been extensively studied and for which established results are available.

First, we briefly summarise how to define the corresponding supergames for a zero-sum infinite stochastic game with separable discounting, along with the specific properties of this supergame. For simplicity, we avoid restating the entire formulation, as the concepts in Definition 3.9 and the elements shown in Figure 8 only require minor logical adjustments.

**Definition 3.32.** Let  $\Gamma$  be a zero-sum infinite stochastic starting at the initial state  $s_1 \in S$ , and let  $\delta$  be a separable discount function.

A supergame  $S(\Gamma, \delta)$  of  $\Gamma$  is a tuple

$$\langle (Z, \mathcal{Z}), X, Y, A^*, B^*, p, f \rangle,$$

which is played by the same players and consists of the following components:

(a)  $Z = S \times \mathbb{T}$  denotes the position space. Let  $\mathcal{Z} = \mathcal{S} \otimes \mathcal{B}(\mathbb{T})$ .

The supergame  $S(\Gamma, \delta)$  starts at the initial position  $(s_1, 1) \in Z$ .

(b) The separable metric spaces  $X$  and  $Y$  are the action spaces for Player 1 and Player 2, respectively.

(c) For each position  $(s, t) \in Z$ , let

$$A^*(s, t) = A(s) \quad \text{and} \quad B^*(s, t) = B(s).$$

(d) The transition law  $p \in \mathbb{Q}(Z \mid Z \times X \times Y)$  is defined by

$$p(E \times \{t'\} \mid (s, t), a, b) = \begin{cases} q(E \mid s, a, b), & \text{if } t' = t + 1, \\ 0, & \text{otherwise,} \end{cases}$$

for every  $E \in \mathcal{S}$ .

(e) The function  $f$  denotes the one-stage payoff function, and it is defined as follows:

$$f(s, t, a, b) = \delta(t, s, a, b) \left( r(s, a, b) - \text{glb}^*(r) \right)$$

for all  $(s, t) \in Z$ .

In the supergame  $S(\Gamma, \delta)$ , Player 1 aims to maximise the expected value of the total payoff, given by

$$\sum_{m=1}^{\infty} f(z_m, c_m, d_m),$$

whereas Player 2 aims to minimise this quantity.

First, Definition 3.32 implies that a supergame is a zero-sum infinite stochastic game. In a supergame, the one-stage payoff function is clearly non-negative. Moreover, every supergame has the Strauch's property, in a form analogous to Formula (3.8). Finally, following Definition 3.14, we can define the lower total value, upper total value, and total value for supergames (see, for example, Nowak (1984A,B)).

However, it is worth noting that supergames do not generally admit total value (as discussed earlier in Subsection 3.1.4). To ensure the existence of such a value, we focus on the results of Nowak (1984B) and introduce three classes of zero-sum infinite stochastic games whose associated supergames do admit a total value. This approach works because the results of Nowak (1984B) apply to positive zero-sum infinite stochastic games with total rewards, a class that includes supergames

As a reminder, we write  $\mathbb{M}_b(X)$  for the set of all bounded measurable functions  $f: X \rightarrow \mathbb{R}$ , where  $X$  is a separable metric space equipped with the Borel  $\sigma$ -algebra  $\mathcal{B}(X)$ . We denote the set of continuous functions in  $\mathbb{M}_b(X)$  by  $C(X)$ , and the sets of upper and lower semicontinuous functions in  $\mathbb{M}_b(X)$  by  $\overline{C}(X)$  and  $\underline{C}(X)$ , respectively.

Let  $S(\Gamma, \delta)$  denote the supergame of a zero-sum infinite stochastic game  $\Gamma$ , starting at the initial state  $s_1 \in S$ , where  $\delta$  represents a separable discount function. Consider the following assumptions (Nowak, 1984A, p. 22.) and (Nowak, 1984B, p. 149):

- A1) The players focus on  $\delta$ -separable discounted reward.
- A2)  $Z$ ,  $X$  and  $Y$  are standard Borel spaces.
- A3)  $X$  and  $Y$  are analytic spaces. Furthermore, the product  $\sigma$ -algebra  $\mathcal{S} \otimes \mathcal{B}(\mathbb{T})$  is closed under the Suslin operation (for more details, see Kechris (1995, Section 25C., pp. 198–201)).
- A4) The transition law  $p$  satisfies the following properties:

$$p(E \mid s, t, \cdot, b) \in C(A^*(s, t)) \quad \text{and} \quad p(E \mid s, t, a, \cdot) \in C(B^*(s, t))$$

for all  $E \in \mathcal{S}$  and for every  $(s, t, a, b) \in \text{Gr}(A^* \times B^*)$ , where

$$\text{Gr}(A^* \times B^*) = \{(s, t, a, b) : (s, t) \in Z, a \in A^*(s, t) \text{ and } b \in B^*(s, t)\}.$$

- A5) The transition law  $p$  satisfies the following properties:

$$p(E \mid s, t, \cdot, \cdot) \in C(A^*(s, t) \times B^*(s, t))$$

for all  $E \in \mathcal{S}$  and for every  $(s, t) \in Z$ .

- A6) The set

$$\left\{ p(\cdot \mid s, t, a, b) : (s, t, a, b) \in \text{Gr}(A^* \times B^*) \right\}$$

is relatively compact in the *MB-topology* in  $\Delta(Z)$  (see Section 3.2).

A7) The one-stage payoff function  $f$  satisfies the following properties:

$$f(s, t, \cdot, b) \in \overline{C}(A^*(s, t)) \quad \text{and} \quad f(s, t, a, \cdot) \in \underline{C}(B^*(s, t))$$

for all  $(s, t, a, b) \in \text{Gr}(A^* \times B^*)$ .

A8) The one-stage payoff function  $f$  satisfies the following properties:

$$f(s, t, \cdot, b) \in C(A^*(s, t)) \quad \text{and} \quad f(s, t, a, \cdot) \in C(B^*(s, t))$$

for all  $(s, t, a, b) \in \text{Gr}(A^* \times B^*)$ .

A9) The one-stage payoff function  $f$  satisfies the following properties:

$$f(s, t, \cdot, \cdot) \in C(A^*(s, t) \times B^*(s, t))$$

for every  $(s, t) \in Z$ .

First, we introduce the class of *zero-sum Borel stochastic games* based on the model of Nowak (1984A, Model (M3), p. 22).

**Definition 3.33.** Let  $S(\Gamma, \delta)$  be the supergame of a zero-sum infinite stochastic game  $\Gamma$  starting at the initial state  $s_1 \in S$ , where  $\delta$  denotes a separable discount function.

We say that  $\Gamma$  is a zero-sum Borel stochastic game if Assumption A1), A2), A4), A6) and A7) hold.

Second, we introduce the class of *zero-sum Suslin stochastic games*, following the model outlined by Nowak (1984A, Model (M2), p. 22).

**Definition 3.34.** Let  $S(\Gamma, \delta)$  be the supergame of a zero-sum infinite stochastic game  $\Gamma$  starting at the initial state  $s_1 \in S$ , where  $\delta$  denotes a separable discount function.

We say that  $\Gamma$  is a zero-sum Suslin stochastic game if Assumption A1), A3), A4), A6) and A8) hold.

Finally, we introduce the class of *zero-sum Nowak stochastic games*, following the model outlined by Nowak (1984A, Model (M1), p. 22).

**Definition 3.35.** Let  $S(\Gamma, \delta)$  be the supergame of a zero-sum infinite stochastic game  $\Gamma$  starting at the initial state  $s_1 \in S$ , where  $\delta$  denotes a separable discount function.

We say that  $\Gamma$  is a zero-sum Nowak stochastic game if Assumption A1), A4), A6) and A9) hold.

Theorem 3.36 directly follows from Definitions 3.33, 3.34 and 3.35; and the results of Nowak (1984B) on positive zero-sum infinite stochastic games with total reward (Nowak, 1984B, Theorems 7.1 and 7.2, p. 149-150).

**Theorem 3.36.** *Let  $\Gamma$  be a zero-sum Borel stochastic game, a zero-sum Suslin stochastic game, or a zero-sum Nowak stochastic game with separable discounting. Then, the following statements hold for the supergame  $\mathbb{S}(\Gamma, \delta)$ :*

- (a)  $\mathbb{S}(\Gamma, \delta)$  admit a total value.
- (b) For every  $\varepsilon > 0$ , Player 1 has an  $\varepsilon$ -optimal stationary strategy.
- (c) Player 2 has an optimal stationary strategy.

### 3.2.4 Results for zero-sum infinite stochastic games with separable discounting

This subsection presents our results on zero-sum infinite stochastic games with separable discounting. These results are a consequence of Theorem 3.36, which follows from the results of Nowak (1984A,B) on positive zero-sum infinite stochastic games with total reward.

We omit the proof of Theorem 3.37, as it follows the same reasoning as the proof of Theorem 3.18, with Theorem 3.36 playing a central role.

**Theorem 3.37.** *Let  $\Gamma$  be a zero-sum Borel stochastic game, a zero-sum Suslin stochastic game, or a zero-sum Nowak stochastic game with a separable discount function  $\delta$ . Then the following statements hold:*

- 1)  $\Gamma$  admit a value.
- 2) For every  $\varepsilon > 0$ , Player 1 has an  $\varepsilon$ -optimal Markov strategy.
- 3) Player 2 has an optimal Markov strategy.

We have the following results for our research questions. A zero-sum infinite stochastic game  $\Gamma$  with separable discounting admits a separable discounted value whenever  $\Gamma$  is a zero-sum Borel stochastic game, a zero-sum Suslin stochastic game, or a zero-sum Nowak stochastic game. This result positively answers the Research Question (RQ3).

In addressing Research Questions (RQ4) and (RQ5), we obtain the following result: Player 2 always has an optimal Markov strategy in any zero-sum infinite stochastic game  $\Gamma$  with separable discounting, provided that  $\Gamma$  is a zero-sum Borel, Suslin, or Nowak stochastic game. In contrast, Player 1, for every  $\varepsilon > 0$ , has an  $\varepsilon$ -optimal Markov strategy in any zero-sum infinite stochastic game  $\Gamma$  with separable discounting, provided that  $\Gamma$  is a zero-sum Borel, Suslin, or Nowak stochastic game.

# Chapter 4

## Discounted finitely additive Markov decision processes

*"Allegro vivace"*

---

W. A. MOZART: *Klavierkonzert Nr. 21 C-Dur, KV 467, 3. Satz*

This chapter explores discounted Markov decision processes (MDP) in the finitely additive framework (Sudderth, 2016). It is a departure from the countably additive framework in Chapters 2 and 3. In those chapters, all probability measures are assumed to be countably additive. Here, we concentrate on *finitely additive probability measures*. To maintain clarity, we use the term *charge* to denote any finitely additive probability measure, while acknowledging that such a set function may, in certain cases, also satisfy  $\sigma$ -additivity.

As a reminder, a *Markov decision process* is a stochastic game involving a single player, referred to as *Player 1*. This chapter also covers *negative Markov decision processes*, where the one-stage payoff function is non-positive in all states.

In this chapter, we study discounted finitely additive Markov decision processes through two discounting techniques: ripple discounting and separable discounting. The main emphasis is on ripple discounting, but we also include a brief discussion of separable discounting because the methods used to solve them are closely related. The chapter addresses Research Questions (RQ6) and (RQ7) in the context of ripple discounting, and Research Questions (RQ8) and (RQ9) in the context of separable discounting.

The chapter is organised as follows. Section 4.1 introduces the *game model* of finitely additive Markov decision processes and presents the different types of

*behavioural strategies*. Section 4.2 discusses discounting techniques for finitely additive Markov decision processes, including *exponential*, *separable* and *ripple discounting*. Section 4.3 focuses on finitely additive Markov decision processes with *ripple discounting*, while Section 4.4 examines the case of *separable discounting*. Each of these two sections concludes by addressing the corresponding research questions.

To keep the focus on finitely additive Markov decision processes, this chapter covers only the essential mathematical concepts. Additional technical details about the finitely additive framework can be found in Appendix A.

## 4.1 The game model

This section introduces finitely additive Markov decision processes and sets out the key theoretical concepts needed to analyse these games thoroughly. The framework we adopt follows the model presented in Sudderth (2016, Section 2).

Before delving into these games (or processes), we introduce some key terminology and notation. A *charge* is a finitely additive probability measure. When a charge  $\mu$  is defined on the  $\sigma$ -algebra of all subsets of a nonempty set  $X$ , that is, on  $\mathcal{P}(X)$ , it is called a *gamble*. The set of all gambles on  $X$  is denoted by  $\text{Gam}(X)$ . For further details, see Appendix A.

First, we introduce finitely additive Markov decision processes. It is important to emphasise that Definition 4.1 describes a family of such games (or processes), where each game differs only in its initial state.

**Definition 4.1.** A finitely additive Markov decision process is a tuple

$$\langle S, A, q, r \rangle$$

which consists of the following components:

- (a)  $S$  is a nonempty set called the state space.
- (b)  $A$  is a nonempty set called the action space.
- (c) The transition rule  $q$  assigns, to each state-action pair  $(s, a) \in S \times A$ , a gamble  $q(\cdot \mid s, a)$  on  $S$ .
- (d) The one-stage payoff function  $r$  is a bounded real-valued function that specifies the payoff  $r(s, a)$  associated with each pair  $(s, a) \in S \times A$ .

A finitely additive Markov decision process  $\langle S, A, q, r \rangle$  is *negative* if the one-stage payoff function  $r$  is non-positive.

Before moving forward, it is useful to emphasise a few important points related to Definition 4.1. Since the sets  $S$  and  $A$  are arbitrary but nonempty, this highlights a key benefit of employing the finitely additive framework introduced in this chapter. Additionally, we assume that both  $S$  and  $A$  are equipped with the discrete topology.

It is also important to emphasise that, for the sake of simplicity, we assume that all actions are available to the player in every state; that is,  $A(s) = A$  for all  $s \in S$ .

The finitely additive nature is clearly reflected in Definition 4.1, where the transition rule assigns a gamble to each state-action pair. To imagine the significance of working with gambles, consider the following property: for any  $\mu \in \text{Gam}(X)$ , the integral  $\int f \, d\mu$  is well-defined for every bounded function  $f: X \rightarrow \mathbb{R}$ , with  $X$  being a nonempty set. This property helps to avoid many integrability issues commonly encountered in countably additive Markov decision processes (Sudderth, 2016, p. 93).

In the next step, we explain how finitely additive Markov decision processes evolve. Such a game starts at an *initial state*  $s_1 \in S$ , after which the following sequence of events occurs at each stage  $t \in \mathbb{T}$ :

Step 1. Player 1 observes the current state  $s_t \in S$  and selects an action  $a_t \in A$ .

Step 2. Player 1 then receives the one-stage payoff  $r(s_t, a_t)$  based on the current state  $s_t$  and the chosen action  $a_t$ .

Step 3. The next state  $s_{t+1} \in S$  is drawn according to the transition probability  $q(\cdot \mid s_t, a_t)$ , and the game proceeds at state  $s_{t+1}$ . Go back to Step 1..

For each stage  $t \in \mathbb{T}$ , we define the collection of *t-length histories* as follows:

$$H_t = \begin{cases} S, & \text{if } t = 1 \\ (S \times A)^{t-1} \times S, & \text{if } t > 1. \end{cases}$$

A *play* is a possible trajectory in a finitely additive Markov decision process. We denote the collection of all histories ( $H$ ) and the collection of all plays ( $H_\infty$ ) by:

$$H = \bigcup_{t \in \mathbb{T}} H_t \quad \text{and} \quad H_\infty = (S \times A)^\mathbb{T}.$$

To conclude this section, we introduce various classes of strategies for finitely additive Markov decision processes. First, we introduce *behavioural strategies* as follows:

**Definition 4.2.** A sequence of mappings  $\sigma = (\sigma_1, \sigma_2, \dots)$  is called a behavioural strategy if, for each stage  $t \in \mathbb{T}$ ,

$$\sigma_t: S \times (A \times S)^{t-1} \mapsto \text{Gam}(A).$$

The strategy described in Definition 4.2 relies on the idea that, when Player 1 observes the history  $h_t = (s_1, a_1, \dots, s_t)$ , the action  $a \in A$  chosen by the strategy  $\sigma$  is drawn according to the probability distribution  $\sigma_t(h_t) \in \text{Gam}(A)$ . We denote the collection of all behavioural strategies by  $\Sigma$ .

A behavioural strategy  $\sigma$  is called *pure* if, for every history  $h \in H$ , the support of  $\sigma(h)$  contains exactly one action; that is,  $|\text{supp}(\sigma(h))| = 1$ .

Before introducing additional types of behavioural strategies, it is helpful to define the following concept:

**Definition 4.3.** A randomised selector is a mapping  $\varphi: S \rightarrow \text{Gam}(A)$ , that is, for each state  $s \in S$ , assigns a gamble  $\varphi(s) \in \text{Gam}(A)$ . On the other hand, a pure selector is a mapping that assigns a single action  $a \in A$  to every state  $s \in S$ .

Finally, we present the class of Markov and stationary strategies.

**Definition 4.4.** A behavioural strategy  $\sigma = (\sigma_1, \sigma_2, \dots)$  is called a (randomised) Markov strategy if there exists a sequence of randomised selectors  $(\varphi_1, \varphi_2, \varphi_3, \dots)$  such that, for every history  $h_t \in H$ ,

$$\sigma_t(h_t) = \varphi_t(s_t).$$

Furthermore, if there exists a sequence of pure selectors  $(f_1, f_2, f_3, \dots)$  such that for each history  $h_t \in H$ , the mapping  $\sigma_t(h_t)$  selects the single action  $\{f_t(s_t)\}$  with probability 1, then the behavioural strategy  $\sigma$  is called a pure Markov strategy.

**Definition 4.5.** A behavioural strategy  $\sigma = (\sigma_1, \sigma_2, \dots)$  is called a (randomised) stationary strategy if there exists a randomised selector  $\varphi$  such that, for every history  $h_t \in H$ ,

$$\sigma_t(h_t) = \varphi(s_t).$$

Furthermore, if there exists a pure selector  $f$  such that, for each history  $h_t \in H$ , the mapping  $\sigma_t(h_t)$  assigns probability one to the single action  $f(s_t)$ , then the behavioural strategy  $\sigma$  is called pure stationary.

For simplicity, we use the notation  $\varphi^\infty$  to represent a stationary strategy  $\sigma$ , where  $\varphi$  is the related randomised selector. Likewise, if  $\sigma$  is a pure stationary strategy with an associated pure selector  $f$ , we denote  $\sigma$  by  $f^\infty$ .

## 4.2 Discounted finitely additive MDPs

This section defines the *reward functions* for three discounting approaches: *exponential*, *separable*, and *ripple discounting*. It also introduces the *optimal rewards* for each discounting technique and classifies behavioural strategies based on their optimality under these discounting methods.

First, we introduce the concepts of *discounted reward* and *optimal reward*, and categorise behavioural strategies based on their optimality when exponential discounting is applied. Then, we present the result of Sudderth (2016, Theorem 3.1), which provides a benchmark for our investigation into separable and ripple discounting.

**Definition 4.6.** Let  $\langle S, A, q, r \rangle$  be a finitely additive Markov decision process with a discount rate  $\alpha \in [0, 1)$ .

For a given behavioural strategy  $\sigma \in \Sigma$  and an initial state  $s \in S$ , the  $\alpha$ -discounted reward (also called the  $\alpha$ -discounted value function) is defined as

$$\gamma_\alpha(\sigma)(s) = \mathbb{E}_s^\sigma \left[ \sum_{t=1}^{\infty} \alpha^{t-1} r(s_t, a_t) \right]. \quad (4.1)$$

The  $\alpha$ -optimal reward at an initial state  $s \in S$  is given by

$$v_\alpha(s) = \sup_{\sigma \in \Sigma} \gamma_\alpha(\sigma)(s).$$

Additionally, for any  $\varepsilon \geq 0$ , a behavioural strategy  $\sigma \in \Sigma$  is  $(\varepsilon, \alpha)$ -optimal if, for every initial state  $s \in S$ , it holds that

$$v_\alpha(s) \leq \gamma_\alpha(\sigma)(s) + \varepsilon.$$

We highlight three key points regarding Definition 4.6. First, it is easy to see that the formulation of the  $\alpha$ -discounted reward in Formula (4.1) guarantees the existence of an  $\alpha$ -optimal reward. Second, a  $(0, \alpha)$ -optimal behavioural strategy is simply called an  $\alpha$ -optimal behavioural strategy. Third, the *extended Dubins-Savage-integral* appears in Formula (4.1) (for more details, see Appendix A).

Before delving into the case of ripple and separable discounting, we present the well-known result for exponential discounting (Sudderth, 2016, Theorem 3.1), which serves as the starting point of our investigation.

**Theorem 4.7.** Let  $\langle S, A, q, r \rangle$  be a finitely additive Markov decision process with a discount rate  $\alpha \in [0, 1)$ . Then the following statements hold:

1. The  $\alpha$ -optimal reward function  $v_\alpha$  satisfies the Bellman equation for all  $s \in S$ , that is,

$$v_\alpha(s) = \sup_{a \in A} \left\{ r(s, a) + \alpha \int v_\alpha(s') q(s' | s, a) \right\}.$$

2. If the action space  $A$  is a finite set, then there exist an  $\alpha$ -optimal pure stationary strategy.
3. There always exists an  $\alpha$ -optimal stationary strategy.

In order to introduce ripple discounting for finitely additive Markov decision processes, we need the following definition:

**Definition 4.8.** Let  $\langle S, A, q, r \rangle$  be a finitely additive Markov decision process. A function

$$\Lambda: \mathbb{T} \times S \times A \rightarrow [0, 1]$$

is a ripple discount function if there exists a number  $M_\Lambda \in \mathbb{R}$  such that for every play  $h_\infty = (s_1, a_1, s_2, a_2, \dots) \in H_\infty$ , the following holds:

$$\sum_{t=1}^{\infty} \left( \prod_{m=1}^t \Lambda(m, s_m, a_m) \right) \leq M_\Lambda. \quad (4.2)$$

By employing the ripple discount function concept, we define the notions of discounted reward and optimal reward, and classify behavioural strategies according to their optimality under ripple discounting as follows:

**Definition 4.9.** Let  $\Gamma = \langle S, A, q, r \rangle$  be a finitely additive Markov decision process with a ripple discount function  $\Lambda$ .

For a given behavioural strategy  $\sigma \in \Sigma$  and an initial state  $s \in S$ , the  $\Lambda$ -ripple discounted reward is defined as

$$\gamma_\Lambda(\sigma)(s) = \mathbb{E}_s^\sigma \left[ \sum_{t=1}^{\infty} \left( \prod_{m=1}^t \Lambda(m, s_m, a_m) \right) r(s_t, a_t) \right]. \quad (4.3)$$

The  $\Lambda$ -optimal reward in an initial state  $s \in S$  is given by

$$v_\Lambda(s) = \sup_{\sigma \in \Sigma} \gamma_\Lambda(\sigma)(s).$$

Additionally, for any  $\varepsilon \geq 0$ , a behavioural strategy  $\sigma \in \Sigma$  is  $(\varepsilon, \Lambda)$ -optimal, if for every state  $s \in S$ , it holds that

$$v_\Lambda(s) \leq \gamma_\Lambda(\sigma)(s) + \varepsilon.$$

We notice that the definition of the ripple discount function  $\Lambda$  guarantees that the  $\Lambda$ -ripple discounted reward is well-defined in Formula (4.3). In other words, the extended Dubins-Savage-integral can be applied to the  $\Lambda$ -discounted sum of one-stage payoffs. For simplicity, a  $(0, \Lambda)$ -optimal behavioural strategy is simply called a  $\Lambda$ -optimal behavioural strategy.

Finally, we introduce the concept of *separable discounting* for finitely additive Markov decision processes. Before doing so, we first define the *separable discount function*.

**Definition 4.10.** *Let  $\langle S, A, q, r \rangle$  be a finitely additive Markov decision process. A function*

$$\delta: \mathbb{T} \times S \times A \rightarrow [0, 1]$$

*is a separable discount function if there exists a number  $M_\delta \in \mathbb{R}$  such that for every play  $h_\infty = (s_1, a_1, s_2, a_2, \dots) \in H_\infty$ , the following holds:*

$$\sum_{t=1}^{\infty} \delta(t, s_t, a_t) \leq M_\delta. \quad (4.4)$$

Building on the concept of separable discount functions, we define discounted reward and optimal reward, and classify behavioural strategies according to their optimality under separable discounting as follows:

**Definition 4.11.** *Let  $\Gamma = \langle S, A, q, r \rangle$  be a finitely additive Markov decision process with a separable discount function  $\delta$ .*

*For a given behavioural strategy  $\sigma \in \Sigma$  and an initial state  $s \in S$ , the  $\delta$ -separable discounted reward is defined by*

$$\gamma_\delta(\sigma)(s) = \mathbb{E}_s^\sigma \left[ \sum_{t=1}^{\infty} \delta(t, s_t, a_t) r(s_t, a_t) \right]. \quad (4.5)$$

*The  $\delta$ -optimal reward in an initial state  $s \in S$  is given by*

$$v_\delta(s) = \sup_{\sigma \in \Sigma} \gamma_\delta(\sigma)(s).$$

*Additionally, for any  $\varepsilon \geq 0$ , a behavioural strategy  $\sigma \in \Sigma$  is called  $(\varepsilon, \delta)$ -optimal if, for every  $s \in S$ , it holds that*

$$v_\delta(s) \leq \gamma_\delta(\sigma)(s) + \varepsilon.$$

The separable discount function  $\delta$  ensures that the  $\delta$ -separable discounted reward is well-defined in Formula (4.5). This means that the extended Dubins-Savage-integral (see Sections A.3 and A.5) can be used to evaluate the  $\delta$ -discounted sum of one-stage payoffs. For convenience, a  $(0, \delta)$ -optimal behavioural strategy is simply called  $\delta$ -optimal.

### 4.3 Finitely additive MDPs with ripple discounting

To analyse finitely additive Markov decision processes with *ripple discounting*, we utilise an approach similar to that described in Chapter 3. The key idea is to transform the original process into a new game, called the *superprocess*. The crucial step is to establish a strong correspondence between the two, since this link allows us to transfer the insights gained from the superprocess back to the original finitely additive Markov decision process with ripple discounting.

Since we also introduce superprocesses for finitely additive Markov decision processes with *separable discounting*, we distinguish between the two types of superprocesses to avoid confusion. Specifically, we refer to them as the R-superprocess (for ripple discounting) and the S-superprocess (for separable discounting).

#### 4.3.1 R-superprocesses

This subsection introduces *R-superprocesses* for finitely additive Markov decision processes with ripple discounting. It also analyses the connections between these classes of games and presents foundational results on R-superprocesses, based on the work of Sudderth (2016, Section 7).

Before formally defining *R-superprocesses*, we first set up the notation for a finitely additive Markov decision process  $\langle S, A, q, r \rangle$  with a ripple discount function  $\Lambda$ . For any history  $h_t = (s_1, a_2, s_2, \dots, a_{t-1}, s_t) \in H_t$ , we define

- $\text{len}(h_t)$ : the *length of the history*, given by  $\text{len}(h_t) = t$ .
- $\kappa(h_t)$ : the *final state in the history*, i.e.,  $\kappa(h_t) = s_t$ .
- $D_\Lambda(h_t)$ : the *predetermined discount factor*, defined as

$$D_\Lambda(h_t) = \begin{cases} 1, & \text{if } \text{len}(h_t) = 1, \\ \prod_{m=1}^{\text{len}(h_t)-1} \Lambda(m, s_m, a_m), & \text{if } \text{len}(h_t) \geq 2. \end{cases} \quad (4.6)$$

Additionally, for any bounded function  $f: X \rightarrow \mathbb{R}$ , we define

$$\text{lub}^*(f) = \max\{0, \text{lub}(f)\},$$

where  $\text{lub}(f)$  represents the least upper bound of  $f$ .

Next, we introduce the concept of an R-superprocess for finitely additive Markov decision processes under ripple discounting.

**Definition 4.12.** Let  $\Gamma = \langle S, A, q, r \rangle$  be a finitely additive Markov decision process with a ripple discount function  $\Lambda$ , starting at the initial state  $s_1 \in S$ .

An  $R$ -superprocess  $S(\Gamma, \Lambda)$  derived from  $\Gamma$  is a tuple

$$\langle Z, B, p, f \rangle$$

which is played by the same player and consists of the following components:

(a)  $Z$  is the set of positions (also known as position space), defined as

$$Z = \bigcup_{t \in \mathbb{T}} H_t,$$

where  $H_t$  denotes the set of all  $t$ -length histories in  $\Gamma$  for each  $t \in \mathbb{T}$ .

The  $R$ -superprocess  $S(\Gamma, \Lambda)$  starts from the initial position  $z = (s_1) \in Z$ , which is the unique history consisting solely of the initial state  $s_1$  of  $\Gamma$ .

(b) For each position  $z \in Z$ ,

$$B(z) = A(\kappa(z))$$

represents the set of available actions for Player 1 at position  $z \in Z$ . Let

$$B = \bigcup_{z \in Z} B(z),$$

which means that  $B$  is equal to  $A$ .

(c) The transition law  $p$  assigns, to each pair  $(z, b) \in Z \times B$ , a gamble  $p(\cdot \mid z, b)$  on  $Z$ , such that for every  $z, z' \in Z$  and  $b \in B(z)$ , the following condition holds:

$$\begin{aligned} p(\{\hat{z}: \hat{z} = (z, b, \kappa(z'))\} \mid z, b) &= q(\kappa(z') \mid \kappa(z), b), \\ p(\{\hat{z}: \text{len}(\hat{z}) \neq \text{len}(z) + 1\} \mid z, b) &= 0. \end{aligned} \quad (4.7)$$

(d) The one-stage payoff function  $f$  is defined as follows: for any position  $z = (s_1, a_1, \dots, a_{\text{len}(z)-1}, s_{\text{len}(z)}) \in Z$  and any action  $b \in B(z)$ , the function  $f$  is given by

$$f(z, b) = D_\Lambda(z) \Lambda(\text{len}(z), \kappa(z), b) \left( r(\kappa(z), b) - \text{lub}^*(r) \right) \quad (4.8)$$

where  $D_\Lambda$  denotes the predetermined discount factor in Formula (4.6).

In the  $R$ -superprocess  $S(\Gamma, \Lambda)$ , Player 1 seeks to maximise the expected value of the total payoff, which is defined as

$$\sum_{m=1}^{\infty} f(z_m, b_m). \quad (4.9)$$

For the remainder of this subsection, let  $\Gamma(s_1)$  denote a finitely additive Markov decision process with a ripple discount function  $\Lambda$ , starting at the initial state  $s_1 \in S$  as given in Definition 4.12.

First, we outline the dynamic of the R-superprocess  $\mathbb{S}(\Gamma(s_1), \Lambda)$ . It starts at the initial position  $z_1 = (s_1) \in Z$ , and at each stage  $n \in \mathbb{T}$ , the following events occur:

Step 1. Player 1 observes the current position  $z_n \in Z$  and selects an action  $b_n \in B(z_n)$ .

Step 2. Based on the current position  $z_n$  and the selected action  $b_n$ , Player 1 receives a one-stage payoff  $f(z_n, b_n)$ .

Step 3. A new position  $z_{n+1} \in Z$  is drawn according to the transition law  $p(\cdot \mid z_n, b_n)$ , and the R-superprocess proceeds at the position  $z_{n+1}$ . Go back to Step 1..

In the following, we present additional details regarding the construction of the R-superprocess  $\mathbb{S}(\Gamma(s_1), \Lambda)$ . Figure 9 depicts the connection between a finitely additive Markov decision process and the corresponding R-superprocess, as defined in Definition 4.12. From a technical viewpoint, Definition 4.12 consists of three primary components:

**Negative)**  $\Gamma(s_1)$  is transformed into a negative finitely additive Markov decision process  $\hat{\Gamma}(s_1) = \langle S, (A(s))_{s \in S}, q, r^* \rangle$  starting at the initial state  $s_1 \in S$ , where  $r^* = r - \text{lub}^*(r)$ . This transformation is implicit in Definition 4.12 and only becomes explicit in (4.8).

**Structural)** The position space  $Z$  of the R-superprocess  $\mathbb{S}(\Gamma(s_1), \Lambda)$  is considerably more complex than the state space  $S$  of the negative finitely additive Markov decision process  $\hat{\Gamma}(s_1)$ . However, the definitions of the elements of the R-superprocess  $\mathbb{S}(\Gamma(s_1), \Lambda)$  ensure that  $\Gamma(s_1)$ ,  $\hat{\Gamma}(s_1)$  and  $\mathbb{S}(\Gamma(s_1), \Lambda)$  fundamentally represent equivalent problems for Player 1. In particular, we mean:

- The R-superprocess  $\mathbb{S}(\Gamma(s_1), \Lambda)$  starts at the initial position  $z = (s_1) \in Z$  uniquely determined by the initial state  $s_1$  of  $\Gamma$  (see Point (a) in Definition 4.12).
- In each position of  $\mathbb{S}(\Gamma(s_1), \Lambda)$ , Player 1 has the same set of available actions as those in the corresponding state of  $\Gamma(s_1)$  (see Point (b) in Definition 4.12).

- The transition rule  $q$  of  $\Gamma(s_1)$  explicitly defines the essential parts of the transition law  $p$  throughout the condition in Formula (4.7).

**Objective)** Player 1 seeks to maximise the *ripple discounted total payoffs* in both games  $\Gamma(s_1)$  and  $\hat{\Gamma}(s_1)$ . In the R-superprocess  $S(\Gamma(s_1), \Lambda)$ , Player 1 aims to maximise the expected value of the *total payoff*. Despite these different goals, both tasks fundamentally address the same problem due to the definitions of  $\hat{\Gamma}(s_1)$  and  $S(\Gamma(s_1), \Lambda)$ .

It is important to note that if the finitely additive Markov decision process  $\Gamma(s_1)$  has a  $\Lambda$ -optimal reward at the initial state  $s_1 \in S$ , the transformation described in Component **Negative)** may decrease this value. This reduction occurs because the transformation involves subtracting the same constant,  $\text{lub}^*(r) \geq 0$ , from the one-stage payoff function  $r$  for every state-action pair  $(s, a) \in S \times A$ . Despite this potential decrease in the one-stage payoff function, the transformation does not affect the optimality of behavioural strategies. That is, if Player 1 has an  $\Lambda$ -optimal behavioural strategy in the original finitely additive Markov decision process  $\Gamma(s_1)$ , then that same behavioural strategy remains  $\Lambda$ -optimal in the negative finitely additive Markov decision process  $\hat{\Gamma}(s_1)$ , and vice versa. Moreover, suppose that  $\Gamma(s_1)$  is already a negative finitely additive Markov decision process. In that case, no adjustment to the one-stage payoff function is needed, and the transformation in Component **Negative)** has no effect.

We denote the class of R-superprocesses of  $\Gamma(s_1)$  with the ripple discount function  $\Lambda$  by  $\mathcal{R}(\Gamma(s_1), \Lambda)$ . A natural question to consider is the size of  $\mathcal{R}(\Gamma(s_1), \Lambda)$ . This question arises from the definition of the transition law  $p$  in Point (c) of Definition 4.12, as the conditions in (4.7) does not uniquely determine the gamble  $p(\cdot \mid z, b)$  on  $Z$  for each pair  $(z, b) \in Z \times B$ .

However, we do not explore this topic further in this subsection because any R-superprocess of  $\Gamma(s_1)$  with the ripple discount function  $\Lambda$  can, in the sense of Components **Structural)** and **Objective)**, be considered equivalent in addressing Research Questions (RQ6) and (RQ7).

To better understand the behaviour of the R-superprocesses of  $\Gamma(s_1)$  with the ripple discount function  $\Lambda$ , the following lemma provides a key observation.

**Lemma 4.13.** *Every R-superprocess from  $\mathcal{R}(\Gamma(s_1), \Lambda)$  is a negative finitely additive Markov decision process.*

*Proof.* Let  $S(\Gamma(s_1), \Lambda) \in \mathcal{R}(\Gamma(s_1), \Lambda)$  be a R-superprocess. By Point (a) of Definition 4.12, the position space  $Z$  of  $S(\Gamma(s_1), \Lambda)$  is a nonempty set. Similarly, the set of

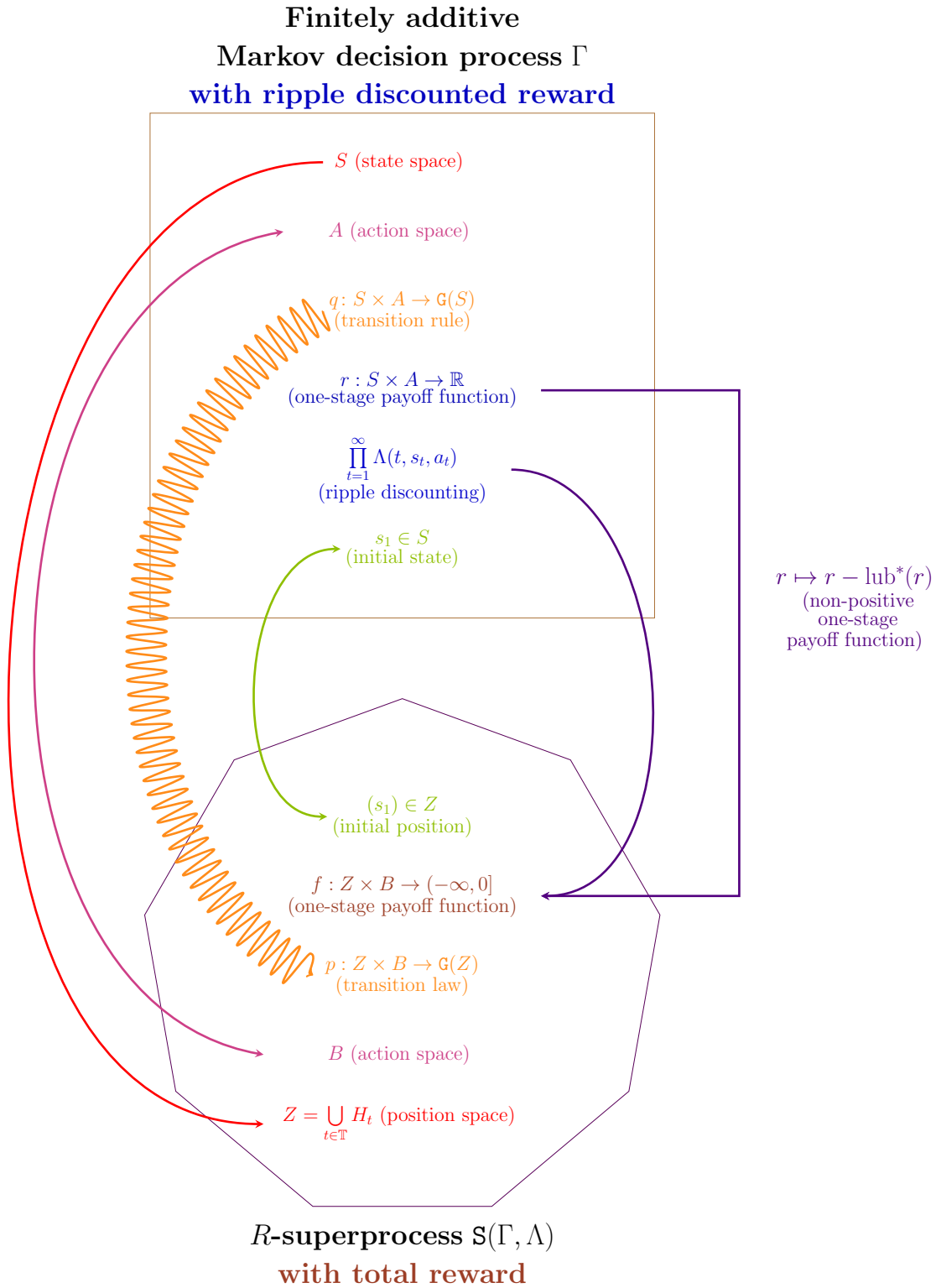


Figure 9: Graphical representation of the steps of constructing an  $R$ -superprocess from a finitely additive Markov decision process with ripple discounting.

available actions for Player 1 in position  $z \in Z$  is a nonempty set by Point (b) of Definition 4.12.

Moreover, the one-stage payoff function  $f$ , defined in Formula (4.8), is bounded. This property follows from the boundedness of both the one-stage payoff function  $r$  and the ripple discount function  $\Lambda$ , from the condition in Formula (4.2), and from the fact that  $\text{lub}^*(r)$  is a real number. Additionally, it is obvious that  $f$  is nonpositive.

Finally, the transition law  $p$  in Point (c) of Definition 4.12 automatically meets the required condition.  $\square$

By Lemma 4.13, we introduce histories, plays, and strategies for an R-superprocess in the same way as in Section 4.1. For the R-superprocess  $\mathcal{S}(\Gamma(s_1), \Lambda)$ , let  $\Theta$  denote the set of all behavioural strategies for Player 1.

The next step introduces the concept of the *total reward* for R-superprocesses. Before doing so, we note the following:

*Remark 4.14.* For R-superprocess from  $\mathcal{R}(\Gamma(s_1), \Lambda)$ , the total payoff in Formula (4.8) is an element of  $\overline{\mathbb{R}}_-$ .

**Definition 4.15.** For every R-superprocess from  $\mathcal{R}(\Gamma(s_1), \Lambda)$ , the total reward for Player 1 under the behavioural strategy  $\vartheta \in \Theta$  at the initial position  $z = (s_1) \in Z$  is

$$\gamma(\vartheta)(z) = \mathbb{E}_z^\vartheta \left[ \sum_{m=1}^{\infty} f(z_m, b_m) \right]. \quad (4.10)$$

The optimal total reward in the initial position  $z = (s_1) \in Z$  is

$$w(z) = \sup_{\vartheta \in \Theta} \gamma(\vartheta)(z). \quad (4.11)$$

Additionally, a behavioural strategy  $\vartheta \in \Theta$  is optimal if

$$w(z) = \gamma(\vartheta)(z).$$

It is important to note that the Dubins-Savage-Sudderth-integral appears in Formula (4.10) (for further details, see the Appendix A).

Theorem 4.16 directly follows from Definitions 4.8 and 4.12, as well as Lemma 4.13. The Strauch's property in Formula (4.12) is also crucial for negative finitely additive Markov decision processes.

**Theorem 4.16.** Every R-superprocess from  $\mathcal{R}(\Gamma(s_1), \Lambda)$  is a negative finitely additive Markov decision process with total reward. Moreover, for each R-superprocess from  $\mathcal{R}(\Gamma(s_1), \Lambda)$ ,

$$w(z) \geq M_\Lambda \cdot \text{glb}(f) > -\infty, \quad (4.12)$$

where  $M_\Lambda \in \mathbb{R}$  is the bound for the ripple discount function  $\Lambda$  in Formula (4.2) and  $\text{glb}(f)$  is the greatest lower bound of the one-stage payoff function  $f$ .

Theorem 4.16 is crucial because it allows us to directly utilise the results of Sudderth (2016) on negative finitely additive Markov decision processes with total reward in the context of R-superprocesses. We conclude this subsection by outlining these results.

By Remark 4.14, let  $\mathcal{L}_-(Z)$  represent the space of all functions from the position space  $Z$  to  $\overline{\mathbb{R}}_-$ . For every R-superprocess from  $\mathcal{R}(\Gamma(s_1), \Lambda)$ , we define the Bellman operator for  $u \in \mathcal{L}_-(Z)$  as

$$(Tu)(z) = \sup_{b \in B(z)} \left\{ f(z, b) + \int u(t)p(\mathrm{d}t \mid z, b) \right\}. \quad (4.13)$$

According to (4.13),  $(Tu)(z)$  captures the best that Player 1 can obtain when starting from the position  $z \in Z$ , by selecting an action  $b \in B(z)$  and receiving both the one-stage payoff  $f(z, b)$ , and accounting for the expected continuation value of  $u \in \mathcal{L}_-(Z)$  under the transition probabilities.

It is clear that the Bellman operator in Formula (4.13) does not have a unique fixed point. Specifically, if a function  $g \in \mathcal{L}_-(Z)$  is a fixed point of the Bellman operator  $T$ , then for any constant  $c$  such that  $g + c \in \mathcal{L}_-(Z)$ , the function  $g + c$  is also a fixed point.

By this observation, we introduce the following notion: a function  $u \in \mathcal{L}_-(Z)$  is called *deficient* if it satisfies  $Tu \geq u$  (Sudderth, 2016, p. 105). With this definition and by applying Theorem 4.16 and the results by (Sudderth, 2016, Theorems 7.1–7.3, pp. 103–105), we arrive at the following result:

**Theorem 4.17.** *For every R-superprocess from  $\mathcal{R}(\Gamma(s_1), \Lambda)$ , the following statements hold:*

(a) *The optimal total reward  $w$  in Formula (4.11) satisfies the Bellman equation:*

$$w(z) = (Tw)(z) = \sup_{b \in B(z)} \left\{ f(z, b) + \int w(z')p(\mathrm{d}z' \mid z, b) \right\}$$

*for all position  $z \in Z$ .*

(b) *The optimal total reward  $w$  is the greatest deficient function in  $\mathcal{L}_-(Z)$ .*

(c) *If the action space  $B$  is finite, then there exists an optimal pure stationary strategy.*

(d) *An optimal stationary strategy always exists.*

Similar statements of Theorem 4.17 were established by Sudderth (2016) for negative finitely additive Markov decision processes with total reward (Sudderth, 2016, Theorems 7.1–7.3). We omit the proof here, as it follows directly from those results. We mention that the presence of Strauch’s property in Formula (4.12) plays an important role for R-superprocesses.

### 4.3.2 Results for finitely additive MDPs with ripple discounting

This subsection presents our results on finitely additive Markov decision processes with ripple discounting and addresses Research Questions (RQ6) and (RQ7).

**Theorem 4.18.** *Let  $\Gamma = \langle S, A, q, r \rangle$  be a finitely additive Markov decision process with a ripple discount function  $\Lambda$ .*

- (a) *There always exists a  $\Lambda$ -optimal behavioural strategy.*
- (b) *If the action space  $A$  is finite, then there exists a  $\Lambda$ -optimal pure behavioural strategy.*

*Sketch of the proof.* (a) Let  $\Gamma(s_1) = \langle S, A, q, r \rangle$  be a finitely additive Markov decision process starting at the initial state  $s_1 \in S$ , and let  $\Lambda$  be a ripple discount function.

Referring to Subsection 4.3.1, particularly Definition 4.12 along with Components **Structural** and **Objective**, each R-superprocess from  $\mathcal{R}(\Gamma(s_1), \Lambda)$  essentially constitutes similar problem for Player 1 as the original game  $\Gamma(s_1)$  does.

However, Point (d) of Theorem 4.17 asserts that Player 1 has an optimal stationary strategy in every R-superprocess from  $\mathcal{R}(\Gamma(s_1), \Lambda)$ . Here, a stationary strategy in a R-superprocesses is the one that relies on only the current position, where a position encodes the history of the original game  $\Gamma$ .

Assume that  $\vartheta^*$  is an optimal stationary strategy in some R-superprocess from  $\mathcal{R}(\Gamma(s_1), \Lambda)$ . This stationary strategy  $\vartheta^*$  can be implemented by Player 1 in the original finitely additive Markov decision process  $\Gamma(s_1)$  with ripple discounting, specifically by employing the behavioural strategy  $\sigma^* \in \Sigma$ . Since  $\vartheta^*$  depends only on the current position, which is a history of  $\Gamma(s_1)$ , it follows that  $\sigma^*$  is a  $\Lambda$ -optimal behavioural strategy because it fully utilises all the relevant information available from the history.

- (b) The result can be established by a line of reasoning analogous to that in Point (a). □

Theorem 4.18 provides a positive answer to Research Question (RQ6), that is, for any finitely additive Markov decision process with ripple discounting, there exists an optimal behavioural strategy for Player 1.

Theorem 4.18 also answers Research Question (RQ7) as follows: there exists a  $\Lambda$ -optimal behavioural strategy. This immediately follows from the construction of the R-superprocesses. Now, consider the following example:

*Example 4.19.* Consider the (finitely additive) Markov decision process  $\Gamma = \langle S, A, q, r \rangle$  in Figure 10 where  $S = \{s^1, s^2, s^3\}$ .

Assume that Player 1 aims to maximise the  $\Lambda$ -ripple discounted reward, where the ripple discount function  $\Lambda$  is defined as:

$$\Lambda(t, s_t, a_t) = \begin{cases} \frac{1}{2}, & \text{if } (t, s_t, a_t) = (1, s^1, \text{U}), \\ \frac{2}{5}, & \text{if } (t, s_t, a_t) = (1, s^1, \text{D}), \\ 1, & \text{if } t = 2, \\ \frac{1}{3}, & \text{if } (t, s_t, a_t) = (3, s^3, \text{B}), \\ \frac{1}{2}, & \text{if } (t, s_t, a_t) = (3, s^3, \text{W}), \\ 0, & \text{otherwise.} \end{cases}$$

The game starts at the initial state  $s^1 \in S$ . No matter which action is taken in the first stage, the game transitions to state  $s^2$  in the second stage, and then moves to the absorbing state  $s^3$  in the third stage.

Now, suppose Player 1 reaches state  $s^3$  in the third stage. To maximise the  $\Lambda$ -ripple discounted reward, Player 1 needs to remember the action chosen in the first stage, since the predetermined discount factor in Formula (4.6) depends on that initial choice. For example, if the action **Up** was chosen at the outset, the predetermined discount factor is  $\frac{1}{2}$ . Alternatively, if **Down** was selected, the predetermined discount factor becomes  $\frac{2}{5}$ .

We note that any behavioural strategy is optimal for Player 1.

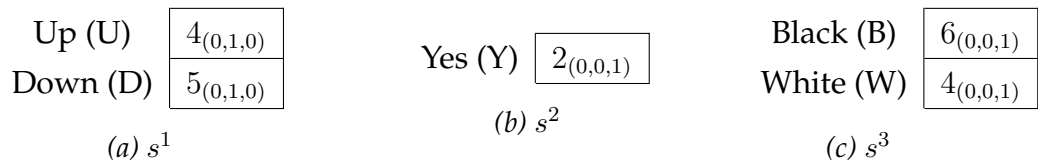


Figure 10: Graphical representation of the Markov decision process with three states in Example 4.19. The initial state of this Markov decision process is  $s^1$ .

## 4.4 Finitely additive MDPs with separable discounting

This section briefly discusses finitely additive Markov decision processes with *separable discounting*.

### 4.4.1 S-superprocesses

This subsection introduces the S-superprocesses and describes their dynamics. We do not repeat the material from Subsection 4.3.1, as the corresponding results for the S-superprocesses follow naturally from the earlier discussion. Therefore, in this section, we focus exclusively on the results that pertain to the S-superprocesses.

**Definition 4.20.** Let  $\Gamma = \langle S, A, q, r \rangle$  be a finitely additive Markov decision process with a separable discount function  $\delta$ , starting from the initial state  $s_1 \in S$ .

An S-superprocess  $\mathcal{S}(\Gamma, \delta)$  derived from  $\Gamma$  is a tuple

$$\langle X, B, p, f \rangle$$

which is played by the same player and consists of the following components:

- (a)  $X$  is the set of positions (also known as position space), defined as

$$X = S \times \mathbb{T} = \{(s, t) : s \in S \text{ and } t \in \mathbb{T}\}.$$

The S-superprocess  $\mathcal{S}(\Gamma, \delta)$  starts from the initial position  $(s_1, 1) \in X$ .

- (b) For each position  $(s, t) \in X$ ,

$$B(s, t) = A(s)$$

represents the set of available actions for Player 1 at position  $(s, t) \in X$ . Let

$$B = \bigcup_{(s,t) \in X} B(s, t),$$

which means that  $B$  is equal to  $A$ .

- (c) The transition law  $p$  assigns, to each pair  $((s, t), b) \in X \times B$ , a gamble  $p(\cdot \mid (s, t), b)$  on  $X$ , such that for every  $(s, t), (s', t') \in X$  and  $b \in B(s, t)$ , the following condition holds:

$$p((s', t') \mid (s, t), b) = \begin{cases} q(s' \mid s, b), & \text{if } t' = t + 1, \\ 0, & \text{otherwise.} \end{cases} \quad (4.14)$$

(d) The one-stage payoff function  $f$  is defined as follows:

$$f(s, t, b) = \delta(t, s, b) \left( r(s, b) - \text{lub}^*(r) \right) \quad (4.15)$$

for all  $(s, t) \in X$ .

In the S-superprocess  $\mathcal{S}(\Gamma, \delta)$ , Player 1 seeks to maximise the expected value of the total payoff, which is defined as

$$\sum_{m=1}^{\infty} f(x_m, b_m). \quad (4.16)$$

Figure 11 depicts the connection between a finitely additive Markov decision process and the corresponding S-superprocess, as defined in Definition 4.20.

For simplicity, let  $\Gamma(s_1)$  denote a finitely additive Markov decision process starting at the initial state  $s_1 \in S$  as given in Definition 4.20.

In the following step, we describe the dynamics of the S-superprocess  $\mathcal{S}(\Gamma(s_1), \delta)$ . The game begins at the initial position  $(s_1, 1) \in X$ , and at each stage  $n \in \mathbb{T}$ , the following events take place:

- Step 1. Player 1 observes the current position  $x_n \in X$  and selects an action  $b_n \in B(x_n)$ .
- Step 2. Based on the current position  $x_n$  and the selected action  $b_n$ , Player 1 receives a one-stage payoff  $f(x_n, b_n)$ .
- Step 3. A new position  $x_{n+1} \in X$  is drawn according to the transition law  $p(\cdot \mid x_n, b_n)$ , and the S-superprocess proceeds at the position  $x_{n+1}$ . Go back to Step 1..

We denote the class of S-superprocesses of  $\Gamma(s_1)$  with the separable discount function  $\delta$  by  $\mathcal{S}(\Gamma(s_1), \delta)$ .

The observations regarding R-superprocesses presented in Components **Negative**, **Structural**, and **Objective**, along with the necessary logical modifications, also apply directly to S-supergames.

Similar to Lemma 4.13, the proof of the following observation proceeds in the same manner.

**Lemma 4.21.** *Every S-superprocess from  $\mathcal{S}(\Gamma(s_1), \delta)$  is a negative finitely additive Markov decision process.*

The following observation is very important in the case of S-supergames:

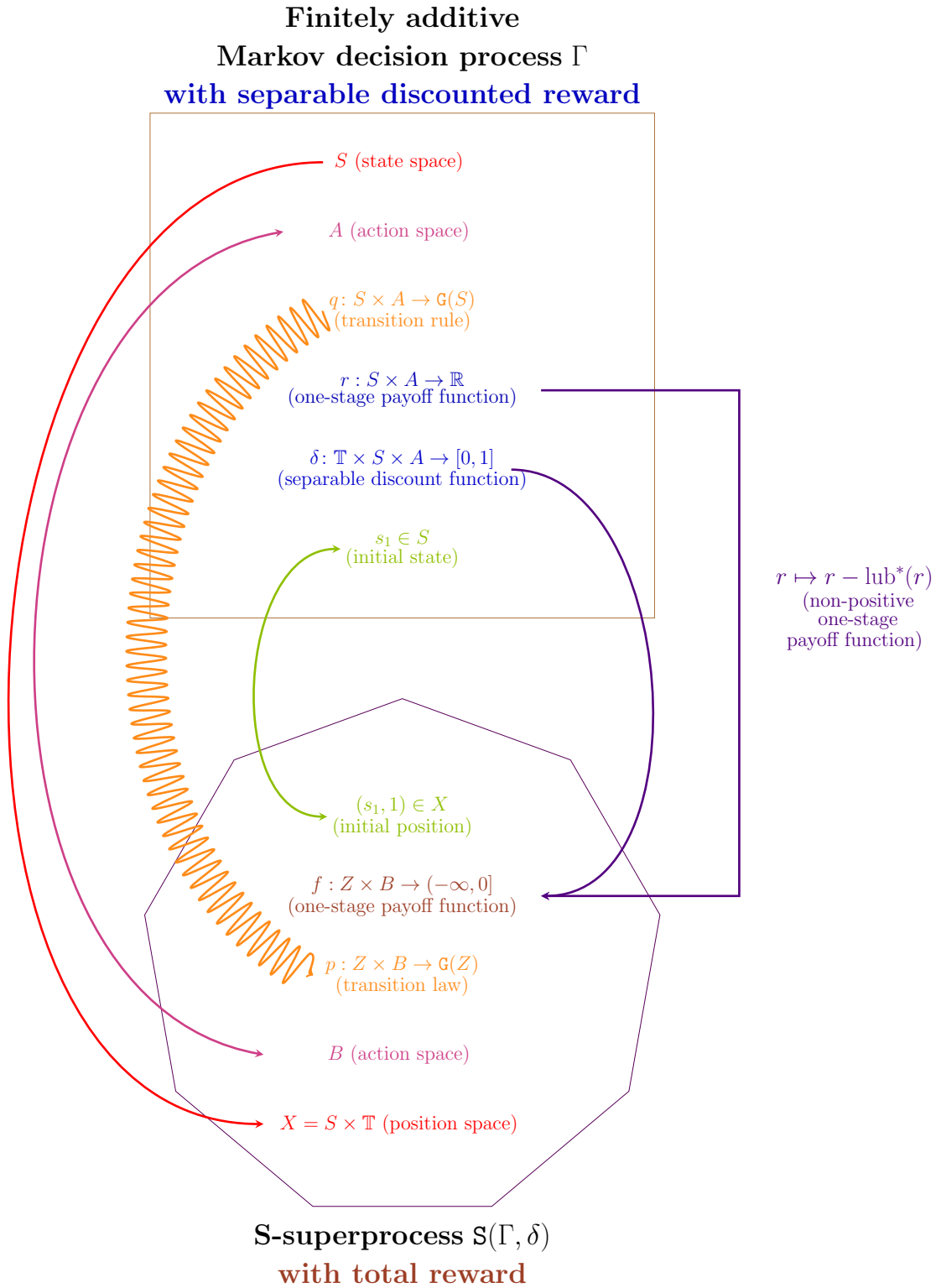


Figure 11: Graphical representation of the steps of constructing an S-superprocess from a finitely additive Markov decision process with separable discounting.

*Remark 4.22.* For every S-superprocess from  $\mathcal{S}(\Gamma(s_1), \delta)$ , the total payoff in Formula (4.15) is an element of  $\overline{\mathbb{R}}_-$ .

As with R-supergames, the total reward is also required in the case of S-supergames

**Definition 4.23.** For every S-superprocess from  $\mathcal{S}(\Gamma(s_1), \delta)$ , the total reward for Player 1 under the behavioural strategy  $\vartheta \in \Theta$  at the initial position  $x = (s_1, 1) \in X$  is defined as

$$\gamma(\vartheta)(x) = \mathbb{E}_x^{\vartheta} \left[ \sum_{m=1}^{\infty} f(x_m, b_m) \right]. \quad (4.17)$$

The optimal total reward in the initial position  $x = (s_1, 1) \in X$  is given by

$$w(x) = \sup_{\vartheta \in \Theta} \gamma(\vartheta)(x). \quad (4.18)$$

Additionally, a behavioural strategy  $\vartheta \in \Theta$  is called **optimal** if it holds that

$$w(x) = \gamma(\vartheta)(x).$$

Based on the arguments presented so far, Theorem 4.24 follows:

**Theorem 4.24.** Every S-superprocess from  $\mathcal{S}(\Gamma(s_1), \delta)$  is a negative finitely additive MDP with total reward. Moreover, for each S-superprocess from  $\mathcal{S}(\Gamma(s_1), \delta)$ ,

$$w(x) \geq M_{\delta} \cdot \text{glb}(f) > -\infty, \quad (4.19)$$

where  $M_{\delta} \in \mathbb{R}$  is bound for the separable discount function  $\delta$  in Formula (4.4) and  $\text{glb}(f)$  is the greatest lower bound of the one-stage payoff function  $f$ .

Theorem 4.25 yields results that parallel those obtained for R-supergames (see Theorem 4.17). The statements of Theorem 4.25 are direct corollaries from Sudderth's results (Sudderth, 2016, Theorems 7.1–7.3, pp. 103–105).

**Theorem 4.25.** For every S-superprocess from  $\mathcal{S}(\Gamma(s_1), \delta)$ , the following statements hold:

(a) The optimal total reward  $w$  in Formula (4.18) satisfies the Bellman equation:

$$w(z) = (Tw)(z) = \sup_{b \in B(z)} \left\{ f(z, b) + \int w(z') p(dz' | z, b) \right\}$$

for all position  $z \in Z$ .

(b) The optimal total reward  $w$  is the greatest deficient function in  $\mathcal{L}_-(Z)$ .

(c) If the action space  $B$  is finite, then there exists an optimal pure stationary strategy.

(d) There always exists an optimal stationary strategy.

### 4.4.2 Results for finitely additive MDPs with separable discounting

This subsection presents our results on finitely additive Markov decision processes with separable discounting and addresses Research Questions (RQ8) and (RQ9).

**Theorem 4.26.** *Let  $\Gamma = \langle S, A, q, r \rangle$  be a finitely additive Markov decision process with a separable discount function  $\delta$ .*

- (a) *There always exists a  $\delta$ -optimal Markov strategy.*
- (b) *If the action space  $A$  is finite, then there exists a  $\delta$ -optimal pure Markov strategy.*

*Sketch of the proof.* (a) Consider the finitely additive Markov decision process  $\Gamma(s_1) = \langle S, A, q, r \rangle$  starting at the initial state  $s_1 \in S$ , with a separable discount function  $\delta$ .

By following the argument used in the proof of Point (a) of Theorem 4.18, we can assume that there exists an optimal stationary strategy  $\vartheta^*$  in some S-superprocess from  $\mathcal{S}(\Gamma(s_1), \delta)$  from  $\mathcal{S}(\Gamma(s_1), \delta)$ .

This stationary strategy  $\vartheta^*$ , which depends only on the current position, that is, the combination of the stage index and the present state of  $\Gamma(s_1)$ , can be adopted by Player 1 in the original finitely additive Markov decision process with separable discounting. In particular, it can be implemented using a Markov strategy  $\sigma^* \in \Sigma_M$ . Consequently,  $\sigma^*$  is a  $\delta$ -optimal Markov strategy.

(b) The result can be established by a line of reasoning analogous to that in Point (a). □

Theorem 4.26 provides an affirmative answer to Research Question (RQ8), establishing that for any finitely additive Markov decision process with separable discounting, an optimal Markov strategy exists for Player 1.

Moreover, this result also addresses Research Question (RQ9): Player 1 has an optimal Markov strategy. Example 4.27 demonstrates that this result cannot be strengthened: there exists a finitely additive Markov decision process with separable discounting in which Player 1 does not have an optimal stationary strategy.

*Example 4.27.* Consider the (finitely additive) Markov decision process  $\Gamma = \langle S, A, q, r \rangle$  illustrated in Figure 12.

Up (U)	2
Down (D)	3

(a)  $s^1$

Figure 12: Graphical representation of the Markov decision process with a single state in Example 4.27. We omit the transition probabilities, as the game has only one state.

Suppose Player 1 aims to maximise the  $\delta$ -separable discounted reward, where the separable discount function  $\delta$  is defined as:

$$\delta(t, s_t, a_t) = \begin{cases} \frac{1}{2}, & \text{if } (t, s_t, a_t) = (1, s^1, \text{U}), \\ \frac{1}{3}, & \text{if } (t, s_t, a_t) = (2, s^1, \text{D}), \\ 0 & \text{otherwise.} \end{cases}$$

Any optimal strategy  $\sigma^* \in \Sigma$  for this problem can be written as:

$$\pi_t^* = \begin{cases} (1, 0), & \text{if } t = 1, \\ (0, 1), & \text{if } t = 2, \\ (x_t, 1 - x_t), & \text{if } t \geq 3, \end{cases}$$

where  $x_t \in [0, 1]$  for all  $t \geq 3$ . Clearly,  $\sigma^*$  is not a stationary strategy: in the first stage, Player 1 always chooses Up, while in the second stage, Player 1 always plays Down.

# Chapter 5

## Conclusion

*"Good luck, Parzival. And thanks. Thanks for playing my game."*

---

JAMES HALLIDAY  
in *Ready Player One*  
by Ernest Cline

This thesis examines three problems in the theory of discounted stochastic games, all motivated by the same fundamental question: what are the implications when the discount factor is time-dependent?

Chapter 1 provides a basic introduction to stochastic games and places them within the broader context of non-cooperative game theory (Parthasarathy and Babu, 2020; Solan and Vieille, 2015; Solan, 2022). After presenting a selective classification of stochastic games and a brief review of the relevant literature, we present three discounting methods: *generalised discounting*, *ripple discounting*, and *separable discounting*. The chapter concludes by formulating our research questions, which guide the investigations in the subsequent chapters.

Chapter 2 focuses on *n-person finite stochastic games* with *generalised discounting*. After establishing the framework for these games (Parthasarathy and Babu, 2020; Solan, 2022), we introduced the concept of generalised discounting (see, for instance, Definitions 2.8 or 2.14). To address Research Question (RQ1) and (RQ2), we introduce the concept of *continuous generalised games*, building on the work of (Glicksberg, 1952), and we demonstrate that every continuous generalised game possesses a Nash equilibrium (see Theorem 2.17). We applied this result to finite stochastic games (see Theorem 2.24). We concluded Chapter 2 with remarks on continuous generalised games.

Chapter 3 discusses *zero-sum stochastic games* with *separable discounting*. To ad-

dress Research Questions (RQ3), (RQ4) and (RQ5), we use the idea of supergames. To illustrate the challenges arising from the application of supergames, we divided our research into two parts: zero-sum *countable* stochastic games and zero-sum stochastic games with a *general state space*. For the case of zero-sum countable stochastic games, we provided a detailed discussion of the emerging problems and the methods by which these could be resolved based on the works of Flesch et al. (2018, 2020). For the case of zero-sum stochastic games with a general state space, we introduce three special classes based on the works of Nowak (1984A,B): zero-sum Borel, Suslin and Nowak stochastic games.

Chapter 4 examines discounted finitely additive Markov decision processes, focusing on both ripple and separable discounting. It addresses the relevant research questions and employs the concept of superprocesses. The chapter draws extensively on the work of Sudderth (2016) on negative finitely additive Markov decision processes with total reward.

## 5.1 Results

The thesis addresses our research questions as follows:

We have a positive answer for Research Question (RQ1): every finite stochastic game with generalised discounting admits a Nash equilibrium (Theorem 2.24). Concerning Research Question (RQ2), current results indicate that there exist finite stochastic games with generalised discounting that lack any equilibrium strategy profile consisting of only stationary strategies (Lemma 2.26). However, whether a Markov discounted Nash equilibrium exists in such games remains an open question. Table 2 comprehensively summarises these results.

	Behavioural strategy	Markov strategy	Stationary strategy
Existence of a Nash equilibrium	✓ (Theorem 2.24)	?	∅ (Example 2.25)

Table 2: Summary of our results for finite stochastic games with generalised discounting.

Regarding Research Question (RQ3), we summarise our results as follows. Every zero-sum *finite* stochastic game with separable discounting admits a value. The same holds for zero-sum *countable* stochastic games, provided that at least one player has only finitely many actions available in each state. Furthermore, every zero-sum *Borel*, *Suslin*, or *Nowak* stochastic game admits a value.

	Optimal Markov strategy	
	Player 1	Player 2
Zero-sum <i>finite</i> stochastic games	✓ (Corollary 3.19)	✓ (Corollary 3.19)
Zero-sum <i>countable</i> stochastic games	∅	∅
Zero-sum <i>Borel, Suslin or Nowak</i> stochastic games	∅	✓ (Theorem 3.37)

Table 3: Summary of our results for zero-sum stochastic games with separable discounting.

In addressing Research Questions (RQ4) and (RQ5), we find the following result for the players. Player 2 always has an optimal Markov strategy in zero-sum finite stochastic games with separable discounting (Corollary 3.19), as well as in zero-sum Borel, Suslin, or Nowak stochastic games (Theorem 3.37). This observation is not particularly surprising, as in the finite case, the set of available actions in each state is both non-empty and finite. In the latter three types of zero-sum infinite stochastic games, we assume that the sets of available actions are non-empty and compact in every state. For zero-sum countable stochastic games, Player 2 has an optimal Markov strategy provided that at least one player has only finitely many actions available in each state (Theorem 3.18).

Player 1 always has an optimal Markov strategy in zero-sum finite stochastic games with separable discounting (Corollary 3.19). In zero-sum Borel, Suslin, or Nowak stochastic games, Player 1 has an  $\varepsilon$ -optimal Markov strategy for every  $\varepsilon > 0$  (Theorem 3.37). For zero-sum countable stochastic games, Player 1 has an  $\varepsilon$ -optimal Markov strategy provided that at least one player has only finitely many actions available in each state (Theorem 3.18).

Table 3 comprehensively summarises these results. It is worth noting that in a zero-sum countable game with separable discounting, a value does not necessarily exist. Additionally, for Player 1, the conditions under which an optimal Markov strategy can be guaranteed in zero-sum Borel, Suslin, or Nowak stochastic games remain an open question. Furthermore, it is important to highlight that there exists a zero-sum countable stochastic game with separable dis-

counting in which no player has an optimal *stationary* strategy (Lemma 3.20).

	Ripple discounting	Separable discounting
Result	an optimal behaviour strategy always exists (Theorem 4.18)	an optimal Markov strategy always exists (Theorem 4.26)

Table 4: Summary of our results for discounted finitely additive Markov decision processes.

For Research Questions (RQ6) and (RQ7), we demonstrate that Player 1 always has an optimal behavioural strategy in a finitely additive Markov decision process with ripple discounting. Similarly, for Research Questions (RQ8) and (RQ9), we find that Player 1 always has an optimal Markov strategy in a finitely additive Markov decision process with separable discounting.

Table 4 summarises these results. It is important to highlight that there exists a (finitely additive) Markov decision process with separable discounting in which Player 1 do not have an optimal *stationary* strategy (Example 4.27). Similarly, there exists a (finitely additive) Markov decision process with ripple discounting in which Player 1 do not have an optimal *stationary* strategy.

# Appendix A

## Mathematical background for Chapter 4

*“Schnelle FüÙe, rascher Mut,  
schützt vor Feindes List und Wut.  
Fänden wir Tamino doch,  
sonst erwischen sie uns noch!”*

---

PAMINA UND PAPAGENO  
in *Die Zauberflöte*, KV 620, 16. Auftritt  
Text: E. Schikaneder, Musik: W. A. Mozart

### A.1 Preliminaries

This section reviews key results and concepts related to finitely additive measures (Dunford and Schwartz, 1957; Rao and Rao, 1983; Luxemburg, 1991). It is important to note that, unlike the usual treatment where only an algebra is assumed, we work with a  $\sigma$ -algebra in this section (Sudderth, 2016).

More precisely, let  $X$  be a nonempty set and let  $\mathcal{A}$  be a  $\sigma$ -algebra on  $X$ . A *finitely additive probability measure*, also known as a *charge*, on the measurable space  $(X, \mathcal{A})$ , is a set function  $\mu: \mathcal{A} \rightarrow [0, 1]$  that satisfies the following two properties:  $\mu(X) = 1$ , and  $\mu(A \cup B) = \mu(A) + \mu(B)$  for all disjoint sets  $A, B \in \mathcal{A}$ .

Let  $\mu$  be a finitely additive probability measure defined on the  $\sigma$ -algebra of all subsets of the  $X$ , that is, on  $\mathcal{P}(X)$ . In that case,  $\mu$  is called a *gamble* on  $X$ . We denote the class of all gambles on  $X$  by  $\text{Gam}(X)$ .

Let  $\mu$  be a charge defined on the measurable space  $(X, \mathcal{A})$ . The integral

$$\int f \, d\mu$$

is defined in the standard way when  $f$  is a  $\mathcal{A}$ -measurable simple function. We denote the space of all  $\mathcal{A}$ -measurable simple functions by  $S(X, \mathcal{A})$ .

For a bounded,  $\mathcal{A}$ -measurable function  $\varphi$ , the integral with respect to  $\mu$  is defined as

$$\int \varphi \, d\mu = \sup_{f \in S(X, \mathcal{A})} \left\{ \int f \, d\mu : f \leq \varphi \right\} = \inf_{f \in S(X, \mathcal{A})} \left\{ \int f \, d\mu : f \geq \varphi \right\}.$$

It is known that for every bounded,  $\mathcal{A}$ -measurable function  $\varphi$  and every charge  $\mu$  on the measurable space  $(X, \mathcal{A})$ , the functional  $\varphi \mapsto \int \varphi \, d\mu$  is linear. In addition, the integral of any constant function  $c$  satisfies  $\int c \, d\mu = c$ .

## A.2 The gambling problem

This section briefly introduces the gambling problem in a finitely additive framework (Dubins and Savage, 1965; Purves and Sudderth, 2010; Sudderth, 2016). Although the gambling problem and stochastic games are closely related, each has its own distinct terminology. For simplicity, we adopt the terminology of the gambling problem, meaning that some terms may differ from those typically used in stochastic games.

A *gambling problem* is a decision problem involving a single player, called the *gambler*. It is defined as follows: let  $F$  be a nonempty set of fortunes. There is a set-valued function  $\Gamma$  that assigns to each fortune  $x_0 \in F$  a nonempty set  $\Gamma(x_0) \subseteq \text{Gam}(F)$  of gambles. This set-valued function  $\Gamma$ , known as a *gambling house*, describes the environment in which the gambler plays.

The gambler starts with an *initial fortune*  $x = x_0 \in F$  and moves through a sequence of fortunes  $x_1, x_2, \dots$  as the gambling problem progresses step by step. Specifically, after observing the initial fortune  $x$ , the gambler picks a gamble  $\gamma$  from the set  $\Gamma(x)$  to determine the next fortune  $x_1$ . This fortune  $x_1$  is a random variable distributed according to  $\gamma$ . The gambler then selects another gamble from  $\Gamma(x_1)$  to determine the subsequent fortune  $x_2$ , and this procedure continues similarly at each stage.

An infinite sequence of fortunes is referred to as a *history*. Denote the set of all such histories by

$$H = F \times F \times F \times \dots$$

The sequence  $X_1, X_2, \dots$  represents the standard coordinate projections on  $H$ , where  $X_t(h) = h_t$  for all  $h \in H$  and  $t \in \mathbb{T}$ .

**Definition A.1.** A Dubins-Savage-strategy is defined as a sequence  $\pi = (\pi_0, \pi_1, \pi_2, \dots)$  where  $\pi_0 \in \text{Gam}(F)$ , and for each  $t \in \mathbb{T}$ , the function  $\pi_t$  maps from  $F^t$  to  $\text{Gam}(F)$ .

### A.3 The Dubins-Savage-integrals

This section introduces the Dubins-Savage-integral and its extended form, the extended Dubins-Savage-integral, as applied to gambling problems (Dubins and Savage, 1965; Purves and Sudderth, 2010; Sudderth, 2016).

Our first goal is to define the charge  $\mathbb{P}_\pi$  over the sequence of fortunes  $x_1, x_2, \dots$  generated by the Dubins-Savage-strategy  $\pi$ . To start, consider the following scenario: define the set

$$H_{AB} = A \times B \times F \times F \times F \times \dots \subseteq H,$$

where  $A$  and  $B$  are arbitrary subsets of  $F$ . Under this setup, the value of  $\mathbb{P}_\pi$  on such a set  $H_{AB}$  is defined as

$$\int_A \pi_1(x_1)(B) \pi_0(dx_1). \quad (\text{A.1})$$

Based on Formula (A.1), it is simpler to define the probability measure  $\mathbb{P}_\pi$  through its associated expectation operator  $\mathbb{E}_\pi$ . Let  $g: H \rightarrow \mathbb{R}$  be a bounded function. When  $g$  depends only on a finite number of coordinates, a natural way to define  $\mathbb{E}_\pi g$  is by using an iterated integral similar to that in Formula (A.1).

From the discussion above, it follows that  $\mathbb{E}_\pi$  can be extended to all bounded functions  $g: H \rightarrow \mathbb{R}$  while preserving the usual linearity properties. Nonetheless, there are many such extensions. By enforcing suitable regularity conditions, these multiple extensions can be reduced to a single one.

Alternatively, we follow the approach introduced by Dubins and Savage, which ensures a unique extension but only for a more limited class of functions that are closely connected to stop rules.

**Definition A.2.** A stop rule is a function  $t: H \rightarrow \mathbb{N}$  with the property that for each  $n \in \mathbb{N}$ , the event

$$\{h: t(h) = n\}$$

is completely determined by the first  $n$  coordinates of  $h$ .

We say that a function  $g: H \rightarrow \mathbb{R}$  is determined by the stopping time  $t$  if whenever  $t(h) = n$  and  $h' \in H$  matches  $h \in H$  on the first  $n$  coordinates, then  $g(h) = g(h')$ .

Finally, a function  $g: H \rightarrow \mathbb{R}$  is determined if there exists a stop rule  $t$  for which  $g$  is determined by  $t$ .

Let  $\mathcal{L}_{\text{BD}}$  denote the family of all bounded and determined functions  $g: H \rightarrow \mathbb{R}$ . We offer a concise summary of how the expectation operator  $\mathbb{E}_\pi$  can be defined on the linear space  $\mathcal{L}_{\text{BD}}$  based on Dubins and Savage (1965); Purves and Sudderth (2010); Sudderth (2016).

To precisely define the conditional expectation  $\mathbb{E}_\pi$  with respect to the charge  $\mathbb{P}_\pi$  given  $X_1 = x$ , we introduce the *conditional Dubins-Savage-strategy*  $\pi[p]$  as follows:

$$\begin{aligned}\pi[x]_0 &= \pi_1(x), \\ \pi[x]_t(x_2, \dots, x_{t+1}) &= \pi_{n+1}(x, x_2, \dots, x_{t+1})\end{aligned}$$

for all  $t \in \mathbb{T}$  and  $x, x_2, \dots, x_{t+1}$ .

With the conditional Dubins-Savage-strategy, we can express the expectation of a function  $g \in \mathcal{L}_{\text{BD}}$  under  $\pi$  as

$$\mathbb{E}_\pi g = \int \mathbb{E}_{\pi[x]}(gx) \pi_0(dx), \quad (\text{A.2})$$

where the  $x$ -section of  $g$  is given by

$$(gx)(h) = g(xh) = g(x, h_1, h_2, \dots)$$

for each history  $h = (h_1, h_2, \dots)$ .

Formula (A.2), together with the condition that  $\mathbb{E}_\pi c = c$  for constants  $c$  defines the *Dubins-Savage-integral* for every Dubins-Savage-strategies  $\pi$  and all functions  $g \in \mathcal{L}_{\text{BD}}$  (Dubins and Savage, 1965).

Let the set  $F$  be endowed with the discrete topology, and let  $H$  carry the product topology. (Dubins and Savage, 1965) showed that  $\mathcal{L}_{\text{BD}}$  contains the indicator functions of all clopen subsets of  $H$ . However, it is clear that  $\mathcal{L}_{\text{BD}}$  does not include all bounded Borel-measurable real-valued functions on  $H$  (Purves and Sudderth, 2010).

First, we consider the following set function

$$\mathbb{P}'_\pi(O) = \sup \{ \mathbb{P}_\pi(K) \mid K \subseteq O \text{ and } K \text{ is clopen} \}$$

for every open subset  $O \subseteq H$ . As shown by Dubins (1974), the set function  $\mathbb{P}'_\pi$  is finitely additive on the lattice of open sets and admits a unique extension to the

algebra generated by the open sets. We denote this extension by  $\mathbb{P}_\pi^*$ . It is worth noting that  $\mathbb{P}_\pi^*$  is not necessarily the only possible extension of  $\mathbb{P}_\pi$  to the open sets.

In the next step, we extend  $\mathbb{P}_\pi^*$  as described by Purves and Sudderth (1976). Define  $\mathcal{A}$  to be the collection of all sets  $A \subseteq H$  for which

$$\inf \{ \mathbb{P}_\pi^*(O) \mid A \subseteq O, O \text{ is open} \} = \sup \{ \mathbb{P}_\pi^*(C) \mid C \subseteq A, C \text{ is closed} \}. \quad (\text{A.3})$$

For any such set  $A \in \mathcal{A}$ , we denote this value in Formula (A.3) by  $\mathbb{P}_\pi^{**}(A)$ . Purves and Sudderth (1976) showed that  $\mathbb{P}_\pi^{**}$  is finitely additive on  $\mathcal{A}$  and  $\mathcal{A}$  contains the Borel  $\sigma$ -algebra. We use  $\mathbb{E}_\pi^{**}$  to denote the expectation operator corresponding to  $\mathbb{P}_\pi^{**}$ .

It is worth noting that the expectation  $\mathbb{E}_\pi^{**}g$  is well-defined for all bounded Borel measurable functions  $g: H \rightarrow \mathbb{R}$ . For the sake of simplicity, we henceforth refer to  $\mathbb{E}_\pi^{**}$  as the *extended Dubins-Savage-integral*.

## A.4 The Dubins-Savage-Sudderth-integral

The notion of the extended Dubins-Savage-integral fits naturally into the framework of finitely additive Markov decision processes, particularly when dealing with exponential, separable, or ripple discounting techniques. However, this approach proves inadequate for handling superprocesses - or more generally, the total reward case.

To address this shortcoming, we adopt the *Dubins-Savage-Sudderth-integral*, defined as follows (Sudderth, 2016, p. 106). Let  $\mathcal{L}_D^+$  denote the family of determined functions  $f: H \rightarrow [0, \infty)$ . Following the same reasoning used in the introduction of the Dubins-Savage-integral, we can define the *generalised Dubins-Savage-integral*  $\mathbb{E}_\pi^D$  for the family  $\mathcal{L}_D^+$  so that  $\mathbb{E}_\pi^D$  satisfies a condition analogous to Formula (A.2), and in particular, that  $\mathbb{E}_\pi^D c = c$  for any constant function  $c$ . It can be shown that the generalised Dubins-Savage-integral  $\mathbb{E}_\pi^D$  is uniquely defined for every Dubins-Savage-strategy  $\pi$  and for every function  $f \in \mathcal{L}_D^+$ .

For a function  $f: H \rightarrow \overline{\mathbb{R}}_+$ , the *Dubins-Savage-Sudderth-integral* is given by

$$\mathbb{E}_\pi^{\text{DSS}} f = \sup \{ \mathbb{E}_\pi^D w \mid w \leq f \text{ and } w \in \mathcal{L}_D^+ \}.$$

Additionally, the Dubins-Savage-Sudderth-integral of a function  $g: H \rightarrow \overline{\mathbb{R}}_-$  is given by

$$\mathbb{E}_\pi^{\text{DSS}} g = -(\mathbb{E}_\pi^{\text{DSS}}[-g]).$$

## A.5 Application of the Dubins-Savage-integral in finitely additive Markov decision processes

This section briefly overviews how the extended Dubins-Savage-integral can be applied in the context of finitely additive Markov decision processes, as discussed in Sudderth (2016).

Consider a finitely additive Markov decision process  $\Gamma = \langle S, A, q, r \rangle$  (see Definition 4.1) starting from the initial state  $s_1 \in S$ . For any subset  $B \subseteq A \times S$ , define the set

$$B_a = \{s \in S : (a, s) \in B\}.$$

A behavioural strategy  $\sigma \in \Sigma$  together with the initial state  $s_1 \in S$  of  $\Gamma$  define a Dubins-Savage-strategy  $\pi(s, \sigma)$  in the following way. For every pair  $(s_1, \nu) \in S \times \text{Gam}(A)$ , define the gamble  $\mu(s_1, \nu) \in \text{Gam}(A \times S)$  by

$$\mu(s_1, \nu)(B) = \int q(B_a \mid s_1, a) \nu(\mathrm{d}a)$$

for each subset  $B \subseteq A \times S$ .

Then, the Dubins-Savage-strategy  $\pi = (\pi_1, \pi_2, \dots)$  is defined as:

$$\begin{aligned} \pi_1 &= \mu(s_1, \sigma_1(s_1)), \\ \pi_t(x_1, x_2, \dots, x_{t-1}) &= \mu(s_t, \sigma_t(s_1, a_1, \dots, s_{t-1}, a_{t-1}, s_t)), \end{aligned}$$

where for  $t \geq 2$ , and  $(x_1, x_2, \dots, x_{t-1}) = ((a_1, s_2), (a_2, s_3), \dots, (a_{t-1}, s_t))$ .

Notice that  $\mathbb{E}_{s_1}^\sigma$  can be constructed as the *extended Dubins-Savage-integral* by following Section A.3.

# Bibliography

Aliprantis, C. D., and Border, K. C. (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Berlin, Heidelberg

<https://doi.org/10.1007/3-540-29587-9>

Altman, E., Feinberg, E. A. and Schwartz, A. (2000): *Weighted Discounted Stochastic Games with Perfect Information*. In: Filar, J.A., Gaitsgory, V., Mizukami, K. (eds) *Advances in Dynamic Games and Applications*. Annals of the International Society of Dynamic Games, vol 5. Birkhäuser, Boston, MA., Chapter 17, pp. 303–323.

[https://doi.org/10.1007/978-1-4612-1336-9\\_17](https://doi.org/10.1007/978-1-4612-1336-9_17)

Amir, R. (2003): *Stochastic Games in Economics and Related Fields: An Overview*. In: Neyman, A. and Sorin, S. (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 30, pp. 455–470.

[https://doi.org/10.1007/978-94-010-0189-2\\_30](https://doi.org/10.1007/978-94-010-0189-2_30)

**Balog, I. and Pintér, M. (to appear): *Folytonos általánosított játékok*. *Alkalmazott Matematikai Lapok***

Bertsekas, D. P. and Shreve, S. E. (1996): *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific Belmont, MA

Bingham, N. H. (2010): *Finite Additivity versus Countable Additivity*. *Electronic Journal for History of Probability and Statistics*, 6:1–35.

Blackwell, D. (1962): *Discrete Dynamic Programming*. *The Annals of Mathematical Statistics*, 33(2):719–726.

<https://doi.org/10.1214/aoms/1177704593>

Blackwell D., Ferguson T. S. (1968): *The Big Match*. *The Annals of Mathematical Statistics*, 39(1):159–163.

<https://doi.org/10.1214/aoms/1177698513>

- Chen, B. S. (2019): *Stochastic Game Strategies and their Applications*. CRC Press  
<https://doi.org/10.1201/9780429432941>
- Couwenbergh, H. A. M. (1980): *Stochastic games with metric state space*. *International Journal of Game Theory*, 9:25–36.  
<https://doi.org/10.1007/BF01784794>
- Császár, Á. (1970): *Bevezetés az általános topológiába*. Akadémiai Kiadó, Budapest
- Doncel, J., Gast, N. and Gaujal, B. (2016): *Are Mean-field Games the Limits of Finite Stochastic Games?*. *ACM SIGMETRICS Performance Evaluation Review*, 44(2):18–20.  
<https://doi.org/10.1145/3003977.3003984>
- Dubins, L. E. (1974): *On Lebesgue-like Extensions of Finitely Additive Measures*. *The Annals of Probability*. 2(3):456–463.  
<https://doi.org/10.1214/aop/1176996660>
- Dubins, L. E. and Savage, L. J. (1965): *How to Gamble if You Must: Inequalities for Stochastic Processes*. McGraw-Hill Book Company. New York.
- Dunford, N. and Schwartz, J. T. (1957): *Linear Operators, Part I, General Theory*. Interscience Publishers, New York
- Feinberg, A. E., Shwartz, A. (2002): *Handbook of Markov Decision Processes: Methods and Applications*. Springer New York, NY  
<https://doi.org/10.1007/978-1-4615-0805-2>
- Filar, J. A., Vrieze, O. J. (1992). *Weighted reward criteria in Competitive Markov Decision Processes*. *ZOR Zeitschrift für Operations Research Methods and Models of Operations Research*, 36:343–358.  
<https://doi.org/10.1007/BF01416234>
- Filar, J. and Vrieze, K. (1997): *Competitive Markov Decision Processes*. Springer New York, NY  
<https://doi.org/10.1007/978-1-4612-4054-9>
- de Finetti, B. (2017): *Theory of Probability: A Critical Introductory Treatment*. John Wiley & Sons Ltd.  
<https://doi.org/10.1002/9781119286387>

- Fink, A. M. (1964): *Equilibrium in a stochastic  $n$ -person game*. Journal of Science of the Hiroshima University, 28(1):89–93.  
<https://doi.org/10.32917/hmj/1206139508>
- Flesch, J. (1998): *Stochastic games with the average reward*. PhD Thesis, University of Maastricht.  
<https://doi.org/10.26481/dis.19981118jf>
- Flesch, J., Predtetchinski, A. and Sudderth, W. (2018): *Characterization and simplification of optimal strategies in positive stochastic games*. Journal of Applied Probability, 55(3):1–14.  
<https://doi.org/10.1017/jpr.2018.47>
- Flesch, J., Predtetchinski, A. and Sudderth, W. (2020): *Positive Zero-Sum Stochastic Games with Countable State and Action Spaces*. Applied Mathematics & Optimization, 82:499–516.  
<https://doi.org/10.1007/s00245-018-9536-3>
- Flesch, J., Vermeulen, D. and Zseleva, A. (2021): *Legitimate equilibrium*. International Journal of Game Theory, 50:787–800.  
<https://doi.org/10.1007/s00182-021-00768-y>
- Frid, E. B. (1974): *On Stochastic Games*. Theory of Probability & Its Applications, 18(2):389–393.  
<https://doi.org/10.1137/1118049>
- Gillette, D. (1957): *Stochastic games with zero stop probabilities*. Contributions to the Theory of Games, 3(39):179–187.  
<https://doi.org/10.1515/9781400882151-011>
- Glicksberg, I. L. (1952): *A Further Generalization of the Kakutani Fixed Point Theorem, with Application to Nash Equilibrium Points*. Proceedings of the American Mathematical Society, 3(1):170–174.  
<https://doi.org/10.2307/2032478>
- González-Sánchez, D., Luque-Vásquez, F. and Minjárez-Sosa, J.A. (2019): *Zero-Sum Markov Games with Random State-Actions-Dependent Discount Factors: Existence of Optimal Strategies*. Dynamic Games and Applications, 9(1):103–121.  
<https://doi.org/10.1007/s13235-018-0248-8>
- Himmelberg, C. J., Parthasarathy, T., Raghavan, T. E. S. and Van Vleck, F. S. (1976): *Existence of  $p$ -Equilibrium and Optimal Stationary Strategies in Stochastic Games*.

- Proceedings of the American Mathematical Society, 60(1):241–251.  
<https://doi.org/10.2307/2041151>
- Judd, K. L., Yeltekin, S. and Conklin, J. (2006): *Computing Supergame Equilibria*. *Econometrica*, 71(4):1239–1254.  
<https://doi.org/10.1111/1468-0262.t01-1-00445>
- Kakutani, S. (1941): *A generalization of Brouwer's fixed point theorem*. *Duke Mathematical Journal*, 8(3):457–459.  
<https://doi.org/10.1215/S0012-7094-41-00838-4>
- Kechris, A. S. (1995): *Classical Descriptive Set Theory*. Springer, New York  
<https://doi.org/10.1007/978-1-4612-4190-4>
- Jaśkiewicz, A. and Nowak, A. S. (2018): *Zero-Sum Stochastic Games*. In: Başar, T., Zaccour, G. (eds) *Handbook of Dynamic Game Theory*. Springer, Cham, Chapter 5, pp. 215–279.  
[https://doi.org/10.1007/978-3-319-44374-4\\_8](https://doi.org/10.1007/978-3-319-44374-4_8)
- Laczkovich, M. (1995): *Valós függvénytan*. ELTE Eötvös Kiadó
- Levhari, D. and Mirman, L. J. (1980): *The Great Fish War: An Example Using a Dynamic Cournot-Nash Solution*. *The Bell Journal of Economics*, 11(1):322–334.  
<https://doi.org/10.2307/3003416>
- Levy, Y. J. and McLennan, A. (2015): *Corrigendum to "Discounted Stochastic Games With No Stationary Nash Equilibrium: Two Examples"*. *Econometrica*, 83(3):1237–1252.  
<https://doi.org/10.3982/ECTA12183>
- Luxemburg, W. A. J. (1991): *Integration with Respect to Finitely Additive Measures*. In: *Positive Operators, Riesz Spaces, and Economics*. *Studies in Economic Theory*, vol 2. Springer, Berlin, Heidelberg, Chapter 6, pp. 109–150.  
[https://doi.org/10.1007/978-3-642-58199-1\\_6](https://doi.org/10.1007/978-3-642-58199-1_6)
- Maitra, A. and Parthasarathy, T. (1970): *On stochastic games*. *Journal of Optimization Theory and Applications*, 5:289–300.  
<https://doi.org/10.1007/BF00927915>
- Maitra, A. and Parthasarathy, T. (1971): *On stochastic games, II*. *Journal of Optimization Theory and Applications*, 8:154–160.  
<https://doi.org/10.1007/BF00928474>

- Maitra, A. and Sudderth, W. (1998): *Finitely additive stochastic games with Borel measurable payoffs*. *International Journal of Game Theory*, 27:257–267.  
<https://doi.org/10.1007/s001820050071>
- Marinacci, M. (1997): *Finitely additive and epsilon Nash equilibria*. *International Journal of Game Theory*, 26:315–333.  
<https://doi.org/10.1007/BF01263274>
- Milchtaich, I. (2023): *Best-response equilibrium: an equilibrium in finitely additive mixed strategies*. *International Journal of Game Theory*, 52:1317–1334.  
<https://doi.org/10.1007/s00182-023-00871-2>
- Mertens, J. F., Neyman, A. (1981): *Stochastic games*. *International Journal of Game Theory*, 10:53–66.  
<https://doi.org/10.1007/BF01769259>
- Mertens, J. F., Parthasarathy, T. (2003): *Equilibria for Discounted Stochastic Games*. In: Neyman, A. and Sorin, S. (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 10, pp. 131–172.  
[https://doi.org/10.1007/978-94-010-0189-2\\_10](https://doi.org/10.1007/978-94-010-0189-2_10)
- Minjárez-Sosa, J. A. (2015): *Markov control models with unknown random state-action-dependent discount factors*. *TOP*, 23(3):743–772.  
<https://doi.org/10.1007/s11750-015-0360-5>
- Nash, J. (1950): *Equilibrium Points in  $n$ -Person Games*. *Proceedings of the National Academy of Science*, 36(1):48–49.  
<https://doi.org/10.1073/pnas.36.1.48>
- Nash, J. (1951): *Non-Cooperative Games*. *Annals of Mathematics*, 54(2):286–295  
<https://doi.org/10.2307/1969529>
- v. Neumann, J. (1928): *Zur Theorie der Gesellschaftsspiele*. *Mathematische Annalen*, 100:295–320.  
<https://doi.org/10.1007/BF01448847>
- Neyman, A. (2003A): *From Markov Chains to Stochastic Games*. In: Neyman, A. and Sorin, S. (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 2, pp. 9–25.  
[https://doi.org/10.1007/978-94-010-0189-2\\_2](https://doi.org/10.1007/978-94-010-0189-2_2)

- Neyman, A. (2003B): *Stochastic Games and Nonexpansive Maps*. In: Neyman, A. and Sorin, S. (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 26, pp. 397–415.  
[https://doi.org/10.1007/978-94-010-0189-2\\_26](https://doi.org/10.1007/978-94-010-0189-2_26)
- Nowak, A. S. (1984A): *On zero-sum stochastic games with general state space I*. *Probability and Mathematical Statistics*, 4(1):13–32.
- Nowak A. S. (1984B): *On zero-sum stochastic games with general state space II*. *Probability and Mathematical Statistics*, 4(2):143–152.
- Nowak A. S. (1985): *Universally Measurable Strategies in Zero-Sum Stochastic Games*. *The Annals of Probability*, 13(1):269–287.  
<https://doi.org/10.1214/aop/1176993080>
- Nowak, A. S. (2003): *Zero-Sum Stochastic Games with Borel State Spaces*. In: Neyman, A. and Sorin, S. (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 7, pp. 77–91.  
[https://doi.org/10.1007/978-94-010-0189-2\\_7](https://doi.org/10.1007/978-94-010-0189-2_7)
- Oliu-Barton, M. (2021): *Weighted-average stochastic games with constant payoff*. *Operational Research*, 22(3):1675—1696.  
<https://doi.org/10.1007/s12351-021-00625-6>
- Parthasarathy, T. and Babu, S. (2020): *Stochastic Games and Related Concepts*. Springer Singapore  
<https://doi.org/10.1007/978-981-15-6577-9>
- Purves, R. A. and Sudderth, W. D. (1976): *Some Finitely Additive Probability*. *The Annals of Probability*, 4(2):259–276.  
<https://doi.org/10.1214/aop/1176996133>
- Purves, R. and Sudderth, W. (2010): *Big Vee: The story of a function, an algorithm, and three mathematical worlds*. *Sankhya*, 72:37–63.  
<https://doi.org/10.1007/s13171-010-0014-5>
- Puterman, M. L. (1994): *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc.  
<https://doi.org/10.1002/9780470316887>

- Raghavan, T. E. S. (2006): *A Stochastic Game Model of Tax Evasion*. In: Haurie, A., Muto, S., Petrosjan, L.A., Raghavan, T.E.S. (eds) *Advances in Dynamic Games. Annals of the International Society of Dynamic Games*, vol 8. Birkhäuser Boston. Chapter 21, pp. 397–420.  
[https://doi.org/10.1007/0-8176-4501-2\\_21](https://doi.org/10.1007/0-8176-4501-2_21)
- Rao, K. B. and Rao, M. B. (1983): *Theory of Charges: A Study of Finitely Additive Measures*. Academic Press, New York
- Samuelson, P. A. (1969): *Lifetime Portfolio Selection By Dynamic Stochastic Programming*. *The Review of Economics and Statistics*, 51(3):239–246.  
<https://doi.org/10.2307/1926559>
- Savage, L. J. (1954): *The foundations of statistics..* John Wiley & Sons  
<https://doi.org/10.1002/nav.3800010316>
- Sanjari, S. and Yüksel, S. (2021): *Optimal Solutions to Infinite-Player Stochastic Teams and Mean-Field Teams*. *IEEE Transactions on Automatic Control*, 66(3):1071–1086.  
<https://doi.org/10.1109/TAC.2020.2994899>
- Shapley, L. S. (1953): *Stochastic games*. *Proceedings of the National Academy of Sciences of the United States of America*, 39(10):1095–1100.  
<https://doi.org/10.1073/pnas.39.10.1095>
- Sobel, M. J. (1971): *Noncooperative Stochastic Games*. *The Annals of Mathematical Statistics*, 42(6): 1930–1935.  
<https://doi.org/10.1214/aoms/1177693059>
- Solan, E. (1998): *Discounted Stochastic Games*. *Mathematics of Operations Research*, 23(4):1010–1021.
- Solan, E. (1999): *Three-Player Absorbing Games*. *Mathematics of Operations Research*, 24(3):669–698.  
<https://doi.org/10.1287/moor.24.3.669>
- Solan, E. (2022): *A Course in Stochastic Game Theory*. Cambridge University Press  
<https://doi.org/10.1017/9781009029704>
- Solan, E. and Vieille, N. (2015): *Stochastic games*. *Proceedings of the National Academy of Sciences of the United States of America*, 112(45):13743–13746.  
<https://doi.org/10.1073/pnas.1513508112>

- Sorin, S. (1986): *Asymptotic properties of a non-zero sum stochastic game*. International Journal of Game Theory, 15:101–107.  
<https://doi.org/10.1007/BF01770978>
- Sorin, S. (2003A): *Classification and Basic Tools*. In: Neyman, A. and Sorin, S. (eds.), Stochastic Games and Applications, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 3, pp. 27–36.  
[https://doi.org/10.1007/978-94-010-0189-2\\_3](https://doi.org/10.1007/978-94-010-0189-2_3)
- Sorin, S. (2003B): *Discounted Stochastic Games: The Finite Case*. In: Neyman, A. and Sorin, S. (eds.), Stochastic Games and Applications, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 5, pp. 51–55.  
[https://doi.org/10.1007/978-94-010-0189-2\\_5](https://doi.org/10.1007/978-94-010-0189-2_5)
- Sorin, S. (2003C): *The Operator Approach to Zero-Sum Stochastic Games*. In: Neyman, A. and Sorin, S. (eds.), Stochastic Games and Applications, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 27, pp. 417–426.  
[https://doi.org/10.1007/978-94-010-0189-2\\_27](https://doi.org/10.1007/978-94-010-0189-2_27)
- Srivastava, S. M. (1998): *A Course on Borel Sets*. Springer New York, NY  
<https://doi.org/10.1007/b98956>
- Steen, L. A. and Seebach, J. A. (1978): *Counterexamples in Topology*. Springer New York, NY  
<https://doi.org/10.1007/978-1-4612-6290-9>
- Strauch, R. E. (1966). *Negative Dynamic Programming*. The Annals of Mathematical Statistics, 37(4):871–890.  
<https://doi.org/10.1214/aoms/1177699369>
- Subir, K. C. (2003): *Pure strategy Markov equilibrium in stochastic games with a continuum of players*. Journal of Mathematical Economics, 39(7):693–724.  
[https://doi.org/10.1016/S0304-4068\(03\)00041-7](https://doi.org/10.1016/S0304-4068(03)00041-7)
- Sudderth, W. D. (2016): *Finitely Additive Dynamic Programming*. Mathematics of Operations Research, 41(1):92–108.  
<https://doi.org/10.1287/moor.2015.0717>

- Takahashi, M. (1964): *Equilibrium points of stochastic non-cooperative n-person games*. Journal of Science of the Hiroshima University, 28(1):95–99.  
<https://doi.org/10.32917/hmj/1206139509>
- Thuijsman, F. (2003): *The Big Match and the Paris Match*. In: Neyman, A. and Sorin, S. (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 12, pp. 195–204.  
[https://doi.org/10.1007/978-94-010-0189-2\\_12](https://doi.org/10.1007/978-94-010-0189-2_12)
- Wald, A. (1949): *Statistical Decision Functions*. The Annals of Mathematical Statistics, 20(2):165–205.  
<https://doi.org/10.1214/aoms/1177730030>
- Wei, Q. and Guo, X. (2011): *Markov decision processes with state-dependent discount factors and unbounded rewards/costs*. Operations Research Letters 39(5):369–374.  
<https://doi.org/10.1016/j.orl.2011.06.014>
- White, D. J. (1993): *A Survey of Applications of Markov Decision Processes*. Journal of the Operational Research Society, 44(11):1073–1096.  
<https://doi.org/10.1057/jors.1993.181>
- Winston, W. (1978): *A Stochastic Game Model of a Weapons Development Competition*. SIAM Journal on Control and Optimization, 16(3):411–419.  
<https://doi.org/10.1137/0316026>
- Wu, X. and Guo, X. (2015): *First passage optimality and variance minimisation of Markov decision processes with varying discount factors*. Journal of Applied Probability, 52(2): 441–456.  
<https://doi.org/10.1239/jap/1437658608>
- Wu, X., Wang, Q. and Kong, Y. (2021): *Two-person zero-sum stochastic games with varying discount factors*. AIMS Mathematics, 6(10):11516–11529.  
<https://doi.org/10.3934/math.2021668>
- Wu, X., Tang, Y. and Medina, R. (2022): *Numerical Calculation of Optimal Policy Pairs in Zero-sum Stochastic Games with Varying Discount Factors*. Discrete Dynamics in Nature and Society, 2022(1):1–10.  
<https://doi.org/10.1155/2022/7474566>

- Ye, L. and Guo, X. (2012): *Continuous-Time Markov Decision Processes with State-Dependent Discount Factors*. *Acta Applicandae Mathematicae*, 121(1):5–27.  
<https://doi.org/10.1007/s10440-012-9669-3>
- Yu, Z., Guo, X. and Xia, L. (2022): *Zero-sum semi-Markov games with state-action-dependent discount factors*. *Discrete Event Dynamic Systems*, 32(4):545–571.  
<https://doi.org/10.1007/s10626-022-00366-4>