Université de Mons
Faculté des Sciences
Département de Mathématique

F.R.S.-FNRS

The Many Faces of Strategy Complexity

James C. A. Main

A dissertation submitted in fulfilment of the requirements of the degree of *Docteur en Sciences* of the Université de Mons

Jury

Nathalie Bertrand Université de Rennes, INRIA, CNRS, IRISA	Reviewer
Thomas Brihaye UMONS – Université de Mons	Secretary
Véronique Bruyère UMONS – Université de Mons	President
Nathanaël Fijalkow CNRS, LaBRI & Université de Bordeaux	Reviewer
Antonín Kučera Masaryk University	Reviewer
Mickaël Randour F.R.S.-FNRS & UMONS – Université de Mons	Supervisor

Abstract

A *reactive system* is a system that continuously interacts with its (uncontrollable) environment. Controllers for reactive systems are notoriously difficult to design, due to the possibly infinite behaviours that the environment may exhibit. This motivates the need for approaches to *automatically design* controllers. *Reactive synthesis* allows one to obtain a correct-by-construction controller automatically from a formal specification. The synthesis problem can be solved by means of a *game-theoretic approach*: we model the interaction of the system and the environment as a game and compute well-performing strategies of the system in this game. A *strategy* of the system player in such a game is the formal counterpart of a *controller* of the system.

A central question is to understand *how complex* strategies must be to enforce specifications. A classical representation of a strategy is via a *Mealy machine*, i.e., a finite automaton with outputs along its transitions. This model is used to define a classical measure of strategy complexity: the size of the smallest Mealy machine inducing it. This is known as the *memory* of the strategy. We explore *different visions* of strategy complexity: starting from this classical model, moving on to randomisation and finally to alternative representations.

First, we consider strategy complexity in the *memory framework* in multi-player turn-based games played on deterministic graphs. We consider multi-player games with (variants of) *reachability* objectives, and focus on *Nash equilibria*, a classical solution concept in multi-player games. We study the sufficient amount of memory to design Nash equilibria in which a given set of players win. We obtain that the memory needed in games with reachability

objectives for such Nash equilibria depends only on the number of players, and that finite memory suffices if all players aim to visit their targets infinitely often rather than only once.

Second, we consider *randomisation* in strategies. Randomisation is useful to balance different goals or to hide one's intentions from others. Randomisation in strategies can be integrated into decision making in different ways. With *mixed strategies*, one tosses a coin at the start of a play to select a deterministic strategy (possibly among infinitely many), and follows this strategy for the entire play. With *behavioural strategies*, one tosses a coin at each step to select an action. Kuhn's theorem, a seminal result in game theory, asserts the equivalence of these two models of randomisation in a broad class of games, called games with perfect recall. We investigate an *analogue of Kuhn's theorem* for finite-memory strategies: we classify the different variants of randomised strategies based on stochastic Mealy machines with respect to their expressiveness and obtain a hierarchy of randomised finite-memory strategies.

As all models of randomisation do not share the same expressiveness, it yields *another measure of strategy complexity*. This measure is not directly related to memory requirements: there can be a trade-off between memory and randomisation requirements in general. We thus investigate *randomisation requirements* in a setting in which randomisation is required: *Markov decision processes* (MDPs) *with multiple objectives*. An MDP is a one-player game where the environment is fully stochastic. Each strategy in an MDP with multiple objectives yields a vector of expected payoffs: we investigate the structure of the set of such expectation vectors under all strategies. We obtain that in this setting, under wide-ranging assumptions, a *limited form of randomisation suffices*.

Finally, we study an *alternative representation* of strategies in a class of infinite-state MDPs. We study *one-counter MDPs*: finite MDPs augmented with a counter that can be decremented, incremented, or left unchanged on each transition. In this setting, strategies with no memory need not admit a finite representation. We consider a natural class of counter-based strategies that admit *finite representations* based on partitions of counter values into *intervals*. For two reachability-based objectives, we provide PSPACE algorithms to solve the problem of checking whether a strategy enforces the objective

with high enough probability and to solve the problem of determining whether there exists a well-performing strategy whose representation satisfies constraints either its structure.

Our results highlight the *multi-dimensional nature* of strategy complexity. We explore several of these dimensions with the goal of providing building blocks for an extensive framework of strategy complexity.

Résumé

Un *système réactif* est un système qui maintient une interaction continue avec son environnement incontrôlable. Les contrôleurs pour des systèmes réactifs sont particulièrement difficiles à concevoir, au vu du nombre potentiellement infini de comportements que l'environnement peut adopter. Cette difficulté motive le besoin d'approches pour *automatiquement concevoir* des contrôleurs pour des systèmes réactifs. La *synthèse réactive* est une approche automatique qui permet d'obtenir un bon contrôleur à partir d'une spécification formelle. Le problème de synthèse peut être résolu au moyen d'une approche basée sur la *théorie des jeux* : l'interaction du système et de son environnement est modélisée par le biais d'un jeu et on y calcule des stratégies performantes du joueur système. Une *stratégie* de ce joueur correspond à un modèle formel d'un *contrôleur* du système.

Une question centrale est de comprendre *à quel point les stratégies doivent être complexes* pour satisfaire des spécifications. Une façon classique de représenter une stratégie peut se faire par le biais d'une machine de Mealy, c'est-à-dire un automate fini avec des sorties sur ses transitions. Ce modèle permet également de définir une mesure classique de complexité pour les stratégies : la *mémoire* de la stratégie, quantifiée par la taille de la plus petite machine de Mealy induisant la stratégie. Dans ce manuscrit, on considère *différentes visions* de complexité des stratégies : le modèle classique de mémoire, l'aléatoire dans la prise de décision et, finalement, les représentations alternatives de stratégies.

Tout d'abord, on étudie la complexité des stratégies par le biais de la mémoire dans des jeux multi-joueurs joués sur des graphes déterministes. On considère les équilibres de Nash dans des jeux avec des variantes d'objectifs

d'accessibilité. On étudie la quantité de mémoire suffisante pour construire un équilibre de Nash dans lequel un sous-ensemble donné de joueurs gagne. On montre que la mémoire suffisante dans les jeux d'accessibilité pour de tels équilibres de Nash ne dépend que du nombre de joueurs, et qu'il suffit d'avoir de la mémoire finie si l'objectif de tous les joueurs est d'atteindre leur cible infiniment souvent plutôt qu'une seule fois.

Ensuite, on s'intéresse à l'aléatoire dans les stratégies. Intégrer de l'aléatoire dans ses décisions permet, par exemple, de cacher ses intentions à ses adversaires, et cette intégration peut se faire de diverses manières. D'une part, avec une *stratégie mixte*, on tire au sort une stratégie déterministe au début d'une partie (parmi un ensemble potentiellement infini) que l'on suit pour l'intégralité de la partie. D'autre part, avec une *stratégie comportementale*, on tire au sort une action à chaque étape de la partie. Un théorème célèbre de Kuhn affirme que ces deux classes de stratégies aléatoires sont équivalentes dans une grande classe de jeux. On propose un *analogue au théorème de Kuhn* pour les stratégies à mémoire finie : on fournit une hiérarchie des différentes variantes des stratégies aléatoires basées sur des machines de Mealy selon leur expressivité.

Étant donné que tous les modèles d'aléatoire ne sont pas équivalents, le type d'aléatoire constitue une *autre mesure de complexité des stratégies.* Cette mesure n'est pas directement correlée à la mémoire : il peut y avoir un compromis entre la mémoire et l'aléatoire en général. On étudie les *besoins d'aléatoire* dans un contexte où il est nécessaire : les *processus de décision de Markov (PDM) avec plusieurs objectifs.* Un PDM est un jeu à un joueur où l'environnement est entièrement stochastique. Chaque stratégie dans un PDM à plusieurs objectifs fournit un vecteur de gains espérés : on étudie la structure de l'ensemble de ces espérances pour toutes les stratégies. On conclut que dans ce contexte, sous des hypothèses peu restrictives, *une forme restreinte d'aléatoire suffit.*

Finalement, on décrit une *représentation alternative* de stratégies dans une classe de PDM à espace d'états infini. On étudie les PDM à un compteur : des PDM finis avec un compteur qui peut être incrémenté, décrémenté ou laissé tel quel sur chaque transition. Dans ce contexte, même les stratégies sans mémoire n'admettent pas toujours de représentations finies. On considère des stratégies qui peuvent être représentées par une partition constituée d'intervalles *finiment représentable* de l'ensemble des valeurs de compteurs. Pour deux objectifs dérivés de l'accessibilité, on propose des algorithmes en espace polynomial pour résoudre un problème de *vérification* (une stratégie donnée satisfait-elle l'objectif avec une probabilité suffisante ?) et deux problèmes de *réalisabilité* étant donné des contraintes sur la structure de la stratégie (existe-t-il une stratégie suffisamment bonne respectant les contraintes imposées ?).

Les résultats de cette thèse soulignent la *nature multi-dimensionnelle* de la complexité des stratégies. On explore différentes facettes de la complexité des stratégies dans le but de contribuer à la conception d'un cadre formel extensif pour la complexité des stratégies.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Mickaël Randour. Words cannot express how thankful I feel for all that I have learned from you, both as a researcher and as a person. All of the conversations we have shared and all of the advice I have received from you is invaluable to me. You introduced me to this field and supported me as your student since right before the start of my master's: your support over the course of these many years is something I can never thank you enough for.

I would also like to sincerely thank all of the other members of my jury: Nathalie Bertrand, Thomas Brihaye, Véronique Bruyère, Nathanaël Fijalkow and Antonín Kučera. I am happy that you have accepted to review my work.

I am also grateful to all of my co-authors in addition to Mickaël: Michal Ajdarów, Thomas Brihaye, Aline Goeminne, Petr Novotný and Jeremy Sproston. I have learned a lot from our collaborations. This manuscript is a crystallisation of the results and the experience I have accumulated through our shared work.

I feel very lucky for the environment I have had the chance to learn and work in these past years. I extend my thanks to all of the staff in the Departments of Mathematics and Computer Science in UMONS for all that you have taught me as teachers and all of the nice interactions we have had as colleagues in later years. I am particularly grateful to Véronique and Thomas for all that you have taught me, all of your advice, our numerous conversations and your overall kindness.

I am also happy to have shared my office(s) with wonderful people. Thank you to Aline, Pierre, Clément, Charly, Gaëtan, Chloé, Nicolas, Florent, Luca, Christophe, Sougata, Sanjana and Pikachu for your greatness, the (very welcome) breaks and conversations, and the occasional board game nights.

I am very thankful to my long-time friends. They only deserve this one sentence because they are not that good. I say this in jest; your friendship is invaluable to me. I would like to thank Gab, Jules, Lev, Line, Manu and Stef for all of the wonderful moments we have shared together in the many years we have known each other. Thank you for supporting me for all these years in spite of my occasional absences.

Last but not least, I am most grateful to my family. I thank my mother and father for their unwavering support, my sister for her colourful and fun personality, and my aunt, uncle and grandmother for all the tea and biscuits they have provided me with throughout the years.

Contents

Chapter 1

Introduction

1.1 Context

An era of ever-present computation. In the modern day, computer systems are deeply integrated in many aspects of society and life, e.g., in entertainment, transportation, healthcare and communication, and we constantly interact with them. As with most things in life, it is most desirable that these computer systems operate correctly. This is all the more true for contexts in which *safety is critical*, e.g., a software issue in a car susceptible to cause a crash should be corrected before the car goes on the market. This motivates the need for *reliable mechanisms* to ascertain the *correctness* of systems.

Detecting errors. A classical technique to detect faults in computer systems is through *testing*. While the usefulness of testing cannot be contested, it is not a foolproof method: tests can only demonstrate the presence of bugs, not guarantee their absence. In particular, when dealing with *reactive systems* [HP85], i.e., systems that constantly interact with an uncontrollable environment through inputs and outputs, exhaustive testing is unrealistic or impossible. This is due to the different behaviours that can be adopted by the environment, of which there can be infinitely many.

Formal methods offer another avenue to check the correctness of systems. *Model checking* is a technique that automatically checks whether a formal model of a system satisfies a specification formulated in some formalism such as *linear temporal logic* [Pnu77]. In other words, model checking provides *mathematical* *guarantees* on the behaviour of a (model of a) system. The guarantees obtained through model checking are with respect to the input model: an *accurate model* (with respect to the specification) is necessary to apply this technique in practice. Model checking emerged in the eighties in independent works by Clarke and Emerson [CE81] and Queille and Sifakis [QS82]. Similarly to logic (e.g., [Büc62]), *automata* play a major role in model checking [VW86]. We refer the reader to the books [BK08, CHVB18] for extensive presentations of model checking.

Automatic design. Model checking requires a model of the system to be verified. In some cases, it is desirable to start from the specification and *automatically design* a system satisfying the specification. This corresponds to the *synthesis problem*. This problem was introduced by Church [Chu57]: he asked whether there exists an algorithm to synthesise a logical circuit from a specification in monadic second order logic. This problem was solved by Rabin [Rab69], and Büchi and Landweber [BL69] in the late sixties.

A variant of the synthesis problem, of particular relevance for reactive systems, is the *controller synthesis problem*. Instead of building a system directly from a formula, the goal is to automatically design a *controller* for an incomplete (reactive) system that enforces the required specification. This variant of the synthesis problem was first studied by Ramadge and Wonham [RW89] for discrete event systems. The (controller) synthesis problem can be tackled by framing the interaction of the system and its environment as a *game* and exploiting models from *game theory*.

Game theory. Game theory is a mathematical field studying models of strategic interactions between agents called *players*. The roots of modern game theory come from Morgenstern and von Neumann's book *"Theory of Games and Economic Behavior"* [vM44]. We also refer the reader to the seminal book on game theory of Obsorne and Rubinstein [OR94], which incorporates many advances made since the inception of the field. In a game, players are *rational*: they aim to maximise their utility, are aware of the alternatives available to them, make decisions based only on the facts at their disposal, are capable to determine the best decisions and will (selfishly) select these alternatives.

The goal of controller synthesis is to design a controller that enforces a specification *regardless of the behaviour of the environment*. Therefore, the controller synthesis problem amounts to studying two-player games in which the system player and environment player are adversaries. Such games are called *zero-sum games*: their goals are the opposite of one another. The goal in the analysis of a zero-sum game is to find a decision-making plan, called a *strategy*, for each player that yields a good outcome regardless of the decisions of the other player.

Models in which players are not competing also exist: these are called *non-zero-sum games*, in which there can be more than two players. Multi-player non-zero-sum games are also of interest for controller design. For instance, if the goal is to control a system consisting of several components, each with its own specification, it may prove too restrictive to assume that all components are adversarial to one another. In multi-player non-zero-sum games, *Nash equilibria* [Nas50] are a classical formalisation of rational behaviour. Intuitively, a Nash equilibrium is a contract between the players, described by one strategy per player, such that none can benefit by unilaterally breaking it.

1.2 A myriad of game models for synthesis

Game-theoretic approaches for synthesis utilise *games played on graphs* [GTW02, BCJ18, FBB+23]. There exist many variants of this model. We first describe one of the simplest models: (non-terminating) two-player games played in a turn-based fashion on a finite graph. We then comment on the extensions that are considered in this manuscript below.

Basic model. We consider a finite directed graph whose vertices, called *states* are partitioned between the two players and whose edges are labelled by *actions*. We start by placing a pebble on an initial state. In each round, the player in control of the current state chooses an action labelling an outgoing edge of the state, and moves the pebble along the edge. This interaction goes on forever and yields an infinite *play*.

The players are allowed to make use of all of the information of the ongoing play in their decision making. This yields the following formalisation of a *strategy*: a function that assigns, to each finite play prefix, an action to be chosen after this prefix. For controller synthesis, strategies of the system player constitute *formal blueprints* for the sought controller for the system.

Extending the model. The structure on which a game is played, like the partitioned graph in the above description, is called an *arena*. The above arena model is *turn-based*, *deterministic*, *finite* and the player are *perfectly informed*. It can be used to model games that respect all of these conditions, such as chess. We can generalise this arena model in several directions, and we can combine these various generalisations.

First, we can incorporate randomness in the transitions of the arena: in each round, after an action is selected, the next state is selected by a distribution that depends *only* on the current state and chosen action. This yields *stochastic game arenas* (e.g., [Con92]). A stochastic game arena can be used to model games in which dice are used, e.g., backgammon.

A special case of particular interest is that of one-player stochastic arenas. Such arenas are known as *Markov decision processes* (MDPs). MDPs are a classical framework for decision making in uncertain environments, which are notably used not only in the fields of formal methods (e.g., [BK08, FBB$^+$23]), but also in reinforcement learning (e.g., [SB18]).

Second, we can extend the model to have the players make their decisions *concurrently*, i.e., the players make their decisions simultaneously and without communicating, in each round (e.g., [dAH00, dAHK07]). A simple example of a concurrent game is rock paper scissors. We call arenas *concurrent* if the players choose their actions simultaneously and *turn-based* otherwise. We remark that one-shot concurrent games are one of the foundational models of game theory [Bor21, von28].

Third, the model described above assumes that the players are *perfectly informed* when making decisions. This assumption is not realistic for some applications, e.g., when dealing with systems with imprecise or unreliable sensors. Games with *imperfect information* (e.g., [OR94, CD12b, BGG17]) can be used to model such situations. In such games, players perceive observations that may correspond to several states at once, and must make their decisions based on these observations. For instance, card games such as poker, in which one only perceives their own hand and not that of the other players, is a game of imperfect information.

Fourth, we can also consider arenas with infinitely many states or infinitely many actions. This can model games in which there can be infinitely many configurations, such as Monopoly (the bank has unlimited funds). Another example is the class of *one-counter MDPs* [BBE$^+$10], which are finitely-presented MDPs with a countable state space. As we study one-counter MDPs in the latter part of this manuscript, we postpone an explanation of this model to the sequel.

1.3 Strategy complexity

Regardless of the arena model, strategies are defined in a similar way and constitute the formal counterpart of controllers in game-based approaches to synthesis. Therefore, in practice, *the simpler the strategy, the better*. On the one hand, small controllers are preferable, e.g., for deployment on resource-constrained embedded systems. On the other hand, it is preferable to have a controller that is understandable to one that is opaque. This motivates a key question: *what makes a strategy complex?*

Memory. A classical measure of the complexity of a strategy is the size of its *memory*, which quantifies the information that the player has to retain to execute the strategy. More precisely, a *finite-memory strategy* is a strategy that can be represented by a *Mealy machine* [Mea55], i.e., a finite automaton with outputs along its edges. The amount of memory of a strategy is the size of the smallest Mealy machine that encodes it. According to this measure, strategies with less memory are preferable and the simplest strategies are *memoryless strategies*, i.e., strategies that disregard the past and make decisions based only on the latest observation.

For many classical specifications in simpler arenas, memoryless strategies suffice. For the sake of illustration, let us consider *reachability objectives*: a reachability objective requires that we visit a target set of states in the arena. Reachability objectives are central in synthesis (see [BGMR23] and references therein). In two-player turn-based zero-sum games on infinite deterministic arenas, memoryless strategies suffice to force a visit to the target whenever possible [GTW02]. Similarly, in two-player turn-based zero-sum games on *finite* stochastic arenas (in MDPs in particular), memoryless strategies suffice to maximise the worst-case probability of visiting the target [Con92]. In countable MDPs, while there need not exist an optimal strategy, memoryless strategies are as powerful as general strategies to maximise reachability probabilities [Orn69, KMS$^+$20].

Of course, some specifications require strategies with memory to be enforced. In contrast to the above, infinite memory may be required to play (almost) optimally for a reachability objective in countable stochastic arenas [KMST24]. Another example consists of *conjunctions of reachability objectives* (e.g. [FH13]) for which the goal is to visit several targets. In this case, memory is necessary already in finite one-player deterministic arenas.

Randomised decision making. Regardless of memory, the definition of a strategy we have used up to now is not well-suited to concurrent and imperfect information settings. This can be observed already with rock paper scissors: regardless of the chosen action, the worst-case outcome is a loss. This highlights a need for richer strategies. In this case, we can do better with *randomised strategies*.

Strategies with randomisation may prove necessary when balancing multiple objectives (e.g., [EKVY08, RRS17, DKQR20]), in concurrent games (e.g., [dAHK07]) and in contexts of partial information (e.g., [CD12b, BGG17]). Whether a strategy is randomised or not can be seen as another aspect of its complexity.

Strategies that deterministically assign actions to each play history are called *pure*. There exist two classical definitions of randomised strategies. On the one hand, *mixed strategies* are distributions over pure strategies. When playing according to a mixed strategy, a pure strategy is drawn at the beginning and is followed throughout the play. On the other hand, *behavioural strategies* assign distributions over available actions to each history. When following a behavioural strategy, actions are drawn randomly at each step according to the distributions it provides.

In the most general settings, the classes of mixed and behavioural strategies are not equivalent in terms of the outcomes they can generate [OR94, Chap. 11]. *Kuhn's theorem* [Kuh53, Aum64] asserts their equivalence in the *perfect recall* setting, i.e., if players never forget their prior knowledge and can observe their own actions. We will see that in our setting, even with imperfect recall, all behavioural strategies are equivalent to some mixed strategy (cf. Theorem 2.47).

Mealy machines can be augmented with randomisation to obtain *finite-memory randomised strategies.* The way that randomisation is integrated in Mealy machines can have an impact in many respects. In several works, stochastic Mealy machines are either defined with stochastic outputs and deterministic updates (i.e., automaton transitions) or stochastic outputs and updates. Both definitions encompass *memoryless randomised (behavioural) strategies.*

The chosen definition of stochastic Mealy machines can impact several aspects. First, more general models can allow one to obtain smaller winning strategies in games. For instance, for almost-surely winning strategies in turn-based stochastic Muller games, while pure finite-memory strategies suffice to win almost surely, smaller Mealy machines can be obtained by allowing stochastic outputs [Cha07] and even smaller Mealy machines can be obtained by allowing both stochastic outputs and updates [Hor09]. Second, some behaviours that can be achieved with the stochastic update model cannot be obtained with deterministic updates [dAHK07, CDH10]: there is a *gap in expressiveness* between the two models. Finally, in spite of the previous two points, it is not necessarily desirable to default to the most expressive model, as model checking them is undecidable in general [GO10].

We highlight two consequences of the above. On the one hand, in spite of Kuhn's theorem, not all classes of randomised strategies, including those that are Mealy machine-based, are equally expressive, powerful or concise. We thus can distinguish different classes of randomised strategies, and study *randomisation requirements* similarly to memory requirements. On the other hand, there can be a trade-off between memory requirements and randomisation requirements. In a nutshell, there are several contributing factors to *strategy complexity*, i.e., it should be seen as a multi-dimensional measure.

Representing strategies. Pure memoryless strategies are the simplest strategies with respect to the complexity measures described above. We can argue, however, that all such strategies are not equally simple. For instance, a constant strategy is much more simple than a strategy that assigns a different action to each state. This indicates that *memory* and *randomisation* do not fully characterise the complexity of strategies.

In practice, when designing controllers for systems with limited resources (e.g., embedded systems), these controllers must be have a compact representation. Already for memoryless strategies, an explicit representation as a table assigning (distributions over) actions to each state can contain a lot of redundancy, seeing as state spaces are often large. This motivates a need for *efficient representations of strategies*. Similarly, for *infinite arenas*, pure memoryless strategies need not admit a finite representation, and thus cannot be implemented in practice. For such settings, we need *tailored models* to deal with infinite state spaces.

The *size of representations* of strategies thus provides another measure of (part of) the complexity of a strategy. For instance, [Gel14] presents a model based on Turing machines, and defines three ad hoc complexity measures: the size of the machine and the time and space complexity of the computations made throughout a play. The need for small representations is also a core motivation of a series of works on *decision tree representations* of memoryless strategies [BCC$^+$15, BCKT18, JKW23], which exploit the structure of the state space to obtain small controllers.

1.4 Contributions

One of our main goals is to refine our understanding of *strategy complexity*. To tackle this wide-ranging question, we consider *various factors* that contribute to the complexity of strategies. In this thesis, we focus on three different aspects of strategy complexity. First, we consider *memory* requirements measured via *Mealy machines*, due to its well-established relevance. We then move on to *randomisation*, which allows for a richer classes of strategies, and is necessary in some instances. We study both the *expressiveness* of randomised strategies and *randomisation requirements* for a given class of MDPs. Finally, we study *alternative representations* of strategies (with respect to Mealy machines): alternative representations can provide insight into the *structure* underlying the decision-making rules of a strategy and yield more *compact representations* than Mealy machines. Each of these three directions give information regarding a *facet of strategy complexity* and highlight the *multi-dimensional* nature of strategy complexity.

In the remainder of this section, we provide a brief description of our main contributions. We refer the reader to Chapter 3 for an extended description of each contribution and its context.

Memory for Nash equilibria. We first focus on a classical measure of strategy complexity: *memory.* We consider a class of non-zero-sum games played on turn-based deterministic arenas, in which all players have a goal of the same type. We focus on variants of reachability specifications. We consider games where all players have a *reachability objective*, games where all players have a *Büchi objective* and games where all players have a *shortest-path cost function.* While reachability objectives are satisfied after visiting a target set once, Büchi objectives require visiting a target infinitely often. Shortest-path cost functions model a quantitative version of reachability: each transition in the arena is assigned a *non-negative integer weight*, and the shortest-path cost assigns the sum of weights to the first occurrence of a target to each play, or positive infinity if no target is visited. The goal of the players is to minimise their cost. In our setting, we assume that the weights are the same for all players.

Our main goal is to understand how much memory is sufficient to implement a *Nash equilibrium* (NE). In the games we study, there can exist several NEs from a given initial state. Furthermore, NEs in which all players lose can coexist with NEs in which all players win. NEs of the latter kind are preferable to those of the former type. For instance, when modelling different system components as players, it is preferable that as many component specifications as possible are satisfied. For this reason, we study *how much memory is sufficient* to obtain a preferable NE (i.e., with which all players lose no utility) from a given NE. In other words, we study upper bounds on the sufficient amount of memory to implement a *constrained Nash equilibrium.* This contribution is based on the single-author paper [Mai24].

We focus on *move-independent Mealy machines*, i.e., a model of strategy representation for which memory updates depend only on the states seen along the play (this definition is used, e.g., in [CRR14, CHVB18, BBGT21]). For multi-player reachability and shortest-path games, we obtain memory upper bounds that are quadratic in the number of players and are *independent of the arena*. For multi-player Büchi games, we obtain that finite memory suffices, and that arena-independent bounds cannot be obtained.

Randomisation and expressiveness. We move on to another aspect of strategy complexity: *randomisation*. First, we focus on the *expressiveness* of models of randomised strategies based on Mealy machine. This contribution is based on joint work with Mickaël Randour [MR24].

Kuhn's theorem provide a sufficient condition yielding the equivalence of mixed and behavioural strategies. A natural question, inspired by this result, is to understand the expressiveness of different *variants of stochastic Mealy machines* as representations of strategies with respect to each other. We study this question in *finite multi-player concurrent arenas with perfect recall*, and in the more general settings that do not assume that the arena is finite, that there is perfect recall or that either assumption holds (although we restrict ourselves to countable arenas in all cases).

We provide a full taxonomy of classes of randomised finite-memory strategies in terms of expressiveness, i.e., in terms of the distributions over plays that can arise when using strategies from a given class. More precisely, we use the same criterion that is used to compare mixed and behavioural strategies in Kuhn's theorem: *outcome equivalence*. Two strategies of a player in an arena are *outcome-equivalent* if they induce the same distributions over plays no matter the strategy of the other players. In particular, this criterion is agnostic to whatever specification we consider: two outcome-equivalent strategies will perform equally well for all goals.

Randomisation complexity. We study *randomisation requirements* in a setting in which randomisation is necessary. We consider *multi-objective MDPs*, i.e., MDPs with *multi-dimensional payoff functions*. A payoff function assigns a numerical value to each play. In multi-objective MDPs, the goal is typically to *achieve a vector* (from an initial state), i.e., find a strategy whose expected payoff vector is greater than the given vector (for the component-wise order). Achieving vectors requires *randomisation* in general. We study the structure of sets of expected payoff vectors in multi-objective MDPs and its impact on randomisation requirements. These results are based on joint work with Mickaël Randour [MR25].

First, we study the relationship between the set of expected payoff vectors obtained through pure strategies and the set of all expected payoffs in *countable MDPs*. We focus on *universally unambiguously integrable payoffs*, i.e., payoffs whose expectation is well-defined for all strategies. We show that any expected payoff vector can be approximated with a *convex combination* of pure expected payoffs. We obtain finer results for *universally integrable payoffs*, i.e., payoff whose expectation is finite for all strategies: all expected payoff vectors are convex combinations of pure expected payoffs. For both cases, it follows that mixed strategies with a *finite support* (i.e., that randomise over a finite set) often suffice to achieve vectors.

While unrelated to randomisation requirements, we also provide sufficient conditions on *continuous payoff functions* in *finite MDPs* that guarantee that the set of expected payoffs is closed.

Finite representations of strategies. Finally, we investigate finite representations of memoryless strategies in one-counter MDPs. *One-counter MDPs* [BBE$^+$10] are finite MDPs augmented with a counter that can be incremented (by one), decremented (by one) or left unchanged on each transition. An OC-MDP induces a possibly infinite MDP over a set of *configurations* given by states of the underlying MDP and counter values. In this induced MDP, any play that reaches counter value zero is interrupted; this event is called *termination*. We consider two variants of the model: *unbounded OC-MDPs*, where counter values can grow arbitrarily large, and *bounded OC-MDPs*, in which plays are interrupted when a fixed counter upper bound is reached.

The counter in OC-MDPs can, e.g., model resource consumption along plays [BBE$^+$10], or serve as an abstraction of unbounded data types and structures [BKK11]. It can also model the passage of time: OC-MDPs generalise *finite-horizon MDPs*, in which a bound is imposed on the number of steps (see, e.g., [BKN+19]). OC-MDPs can be also seen as an extension of *one-counter Markov chains* with non-determinism. One-counter Markov chains are equivalent to (discrete-time) quasi-birth-death processes [EWY10], a model studied in queuing theory.

We consider two objectives in OC-MDPs: *state-reachability*, i.e., reaching a target set of states, and *selective termination*, i.e., reaching a target set of states with counter value zero, thus generalising termination. The synthesis problem for the latter is not known to be decidable and is connected to major open problems in number theory [PB24, OW14]. Furthermore, even memoryless strategies in OC-MDPs might be impossible to build in practice due to the possibly infinite configuration space. To overcome these obstacles, we introduce two classes of *concisely represented* strategies based on a (possibly infinite) partition of counter values in intervals. We collectively refer to these strategies as *interval strategies*.

For both classes of strategies, and both objectives, we study the verification problem and two synthesis problems. On the one hand, the verification problem asks whether a given strategy ensures a high enough probability for the objective. On the other hand, our synthesis problems asks for *structurally-constrained* strategies. For the first problem, we fix the interval partition of the strategy as an input. For the other problem, we give parameters constraining the representation of the interval partition. We develop a generic approach based on a compression of the induced countable MDP that yields decidability in all cases, with all complexities within PSPACE. These contributions originate from joint work with Michal Ajdarów, Petr Novotný and Mickaël Randour [AMNR25].

1.5 Outline

This manuscript is divided into six parts. At the end of this manuscript, the reader can find an index of technical terms and a table of notations. We illustrate the overall structure of the manuscript in Figure 1.1.

Part I introduces the background and the contributions of this thesis. In Chapter 2, we introduce basic mathematical notation, background and all game models related to our contributions. Appendix A complements this chapter:

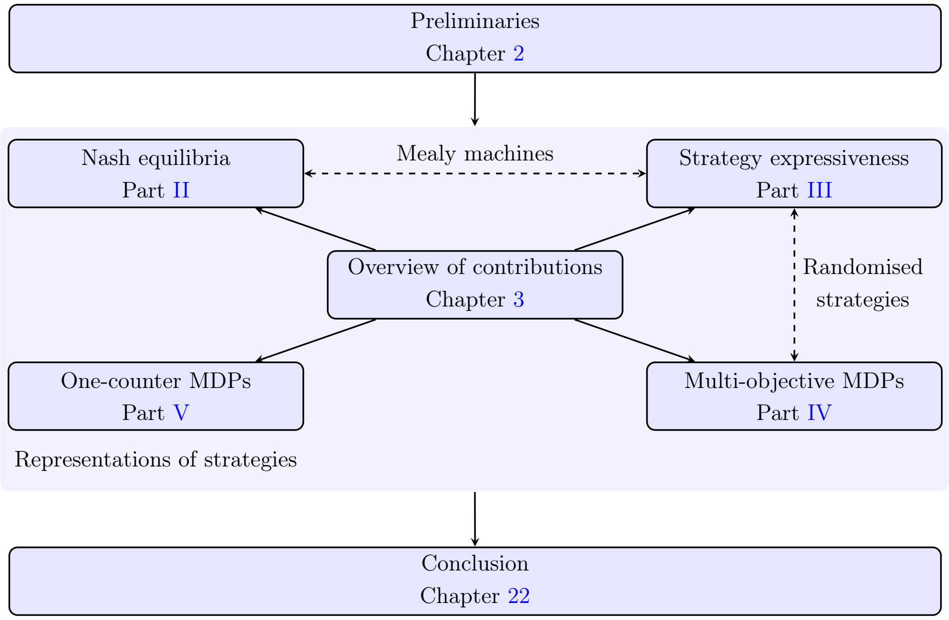


Figure 1.1: Overview of the structure of this thesis.

it includes the proofs of the results of Chapter 2 and additional background. Chapter 3 provides an extended high-level presentation of each contribution outlined above. Reading up to Chapter 3 should give the reader a global overview of the results of this thesis.

Parts II–V are dedicated to the contributions highlighted in the previous section. Chapters 4, 8, 12 and 16 are the first of their respective part and serve as local introductions. They complement Chapter 3 by providing a summary-like outline of their part. We provide a brief description of the content of each part, and refer to these chapters for an extended presentation.

Part II presents our results regarding memory requirements for constrained Nash equilibria. In Chapter 5, we discuss constrained Nash equilibria and the model of Mealy machines we use. Chapter 6 provides an overview of existing results, and adaptations, when necessary, on zero-sum and non-zero-sum games on graphs that we use to construct NEs. Finally, we present our results regarding memory in Chapter 7.

We study the expressiveness of randomised strategies in Part III. Chapter 9 provides a discussion of the definition of outcome-equivalence and a proof of

Kuhn's theorem. Chapters 10 and 11 respectively showcase inclusions and non-inclusions between classes of finite-memory strategies.

We focus on multi-objective MDPs in Part IV. We introduce multi-objective-specific notation in Chapter 13, along with some essential integration-related results for the next chapter. Chapter 14 presents our results relating sets of pure expected payoffs with sets of all expected payoffs. We study continuous payoffs in multi-objective MDPs in Chapter 15.

Part V is dedicated to interval strategies in OC-MDPs. We define interval strategies, discuss some basic properties and formalise our decision problems in Chapter 17. We present our compression idea in Chapter 18. Chapter 19 and 20 provide algorithms for verification and synthesis respectively based on the compression approach. Finally, we complement the upper bounds obtained through these algorithms with lower bounds in Chapter 21.

We conclude in Part VI, which consists of a single chapter, Chapter 22.

1.6 Publication history

Contributions in this thesis originate from four published articles [MR22, MR24, Mai24, AMNR25], where the journal paper [MR24] extends the conference paper [MR22], and one technical report [MR25]. We briefly comment on the other publications of the author as they have contributed to shaping the current vision of the author.

We first discuss work on *timed automata and games*. A timed automaton [AD94] is a finite automaton augmented with real-valued variables called clocks that increase at the same rate. Clocks model the passage of time. A timed game is played on a timed automaton by two players. We follow the game model of [dAFH$^+$03]: in each round, the two players concurrently select a delay and an action, and a transition is taken following the move with least delay. In joint work with Mickaël Randour and Jeremy Sproston [MRS21, MRS22], we study *window parity objectives* in timed automata and games. These objectives are a variant of the classical parity objective with timing constraints; see also [BHR16, BDOR20] for the study of window parity objectives in discrete-time models. We provide verification algorithms for timed automata and synthesis algorithms in timed games for window parity objectives with matching lower and upper complexity bounds.

Second, we mention an invited contribution co-authored with Thomas Brihaye, Aline Goeminne and Mickaël Randour [BGMR23]. This work is a (non-exhaustive) survey on (variants of) reachability games on graphs. We explore two notions of complexity: the computational probability of solving games and the strategy complexity required to play optimally.

1.7 Related work

We only discuss related works connected to several parts of the manuscript. Additional references can be found in the additional context described in Chapter 3, and some related work is discussed at the end of the introductory chapters of each part.

We refer the reader to [BK08] for a general introduction to model checking and Markov decision processes, to [BCJ18] for a presentation of reactive synthesis and games and to [FBB$^+$23] as a general reference on games on graphs.

We first discuss memory requirements. Many works that provide synthesis algorithms also study how much memory is necessary and how much memory is sufficient to enforce the specification (which can be winning in a zero-sum game or achieving a vector in a multi-objective MDP); examples of such endeavours can found, e.g., in [FH13, CD12a, CRR14, RRS17, BGHM17]. Understanding memory bounds can also yield complexity-theoretic results. For instance, the existence of memoryless winning strategies for both players in zero-sum parity [EJ88] and mean-payoff [EM79] games on deterministic turn-based finite arenas can be used to show that the complexity of determining the winner in such games is in $\mathsf{NP} \cap \mathsf{co}\text{-}\mathsf{NP}$.

Memory requirements are sensitive to the way that strategies are defined. As explained previously, whether we allow randomisation or not can impact memory requirements. The power and conciseness of strategies also depends on the information they are allowed to register to update their memory. For instance, if objectives and payoffs are defined via sequences of *colours* labelling transitions (as in, e.g., [GZ05, BLO$^+$22, BRV23]) and only colours may be used in memory updates, the amount of memory to win in a game can be greater than with general strategies [Koz24].

By colouring transitions, we can define objectives (i.e., sets of winning plays), payoffs and, more generally, preference relations over plays *independently of the arena*. This formalism can be used to characterise the *power of finite-memory strategies* for all games with the same winning condition. There are several works endeavouring to understand when finite memory is sufficient; all of the (non-exhaustive) works we mention below are in the *turn-based perfect information setting*. Gimbert and Zielonka, and Colcombet and Niwiński provide characterisations of winning conditions for which optimal memoryless strategies exist for both players in zero-sum games on finite [GZ05] and infinite [CN06] arenas respectively. Bouyer et al. provide characterisations for games in which *arena-independent finite-memory strategies* (i.e., the same colour-based update scheme can be used to win in all arenas) suffice for both players in finite deterministic arenas [BLO+22], finite stochastic arenas [BORV23] and infinite deterministic arenas [BRV23]. There also exist *sufficient conditions* on winning conditions that ensure the existence of memoryless strategies: see [GK23] for zero-sum games on finite stochastic arenas and [Gim07] for finite MDPs. In the same vein, Le Roux and Pauly provide conditions on games on finite deterministic arenas such that finite-memory NEs exist in non-zero-sum games whenever certain conditions on the corresponding zero-sum games hold [LP18].

The last direction we discuss is related to the representation of strategies. We mention a few strategy representations. First, we highlight the previously mentioned model of Gelderie based on *Turing machines* [Gel14]. Turing machine models have also been used to quantify the reasoning ability of players by means of computational complexity classes [DJ23]. We have also mentioned *decision tree* representations of memoryless strategies [BCC+15, BCKT18, JKW23]; recently this approach was extended to strategies with memory by combining Mealy machines with decision trees [ACKK24]. In reinforcement learning, *neural networks* are used to compute and represent strategies [SB18]. For strategies with memory, *recurrent neural networks* can be used (e.g., [KFG19, CJW+19, CJT20]). Finally, we mention [SFM24], which studies *programmatic policies*, i.e., strategies represented by programs.

Funding

The author of this thesis is a Research Fellow of the Fonds de la Recherche Scientifique – FNRS (F.R.S.-FNRS) and member of the TRAIL institute. He is supervised by Mickaël Randour (F.R.S.-FNRS & Université de Mons), an F.R.S.-FNRS Research Associate and member of the TRAIL institute. The author and Mickaël Randour were supported by the F.R.S.-FNRS under Grants n° F.4520.18 (MIS ManySynth) and n° T.0188.23 (PDR ControlleRS). Petr Novotný, co-author of the contributions presented in Part V, has been supported by the Czech Science Foundation grant no. GA23-06963S.

Part I:

Games, Markov decision processes and strategies

Chapter 2

Preliminaries

This chapter introduces the background of this thesis and the notation used in the subsequent chapters. Section 2.1 presents the general mathematical notation and recalls some geometric results. Sections 2.2 to 2.8 introduce the models of games considered in this manuscript and related relevant notions, such as strategies, payoff functions and objectives. We defer some additional material, most notably basic topology definitions and the proofs of some results stated in this chapter to Appendix A (Page 399).

Contents

2.1 Mathematical background

2.1.1 Sets, functions and words

We write $\mathbb{N}$, $\mathbb{Q}$ and $\mathbb{R}$ for the sets of non-negative integers, rational numbers and real numbers respectively. We denote the extended real line by $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. We let $\mathbb{N}_{>0} = \mathbb{N} \setminus \{0\}$ denote the set of positive integers, and let $\bar{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$ and $\bar{\mathbb{N}}_{>0} = \mathbb{N}_{>0} \cup \{+\infty\}$. Given $n, n' \in \bar{\mathbb{N}}$, we let $[\![n, n']\!]$ denote the set $\{k \in \mathbb{N} \mid n \leq k \leq n'\}$ of natural numbers ranging between n and n', and if $n = 0$, we shorten the notation to $[\![n']\!]$. In particular, with this notation, we have $[\![\infty]\!] = \mathbb{N}$.

Let $A' \subseteq A$ and $B' \subseteq B$ be sets and $f\colon A \to B$ be a function. We let $\mathbb{1}_{A'}\colon A \to \{0, 1\}$ denote the indicator function of A'. We let $\mathsf{Im}(f) = f(A) = \{f(a) \mid a \in A\}$ denote the image of f. We let $f^{-1}(B') = \{a \in A \mid f(a) \in B'\}$ denote the inverse image of B' by f. For any $b \in B$, we write $f^{-1}(b)$ instead of $f^{-1}(\{b\})$ to lighten notation.

The cardinality of A is denoted by $|A|$. We let A^*, A^+ and A^ω respectively denote the set of finite, non-empty finite and infinite words over A. We write ε for the empty word. A subset $\mathcal{L} \subseteq A^*$ is *prefix-free* if no word of $\mathcal{L}$ is a strict prefix of another word of $\mathcal{L}$.

2.1.2 Probability theory

Let A be a countable set. We write $\mathcal{D}(A)$ for the set of probability distributions over A, i.e., the set of functions $\mu\colon A \to [0,1]$ such that $\sum_{a\in A} \mu(a) = 1$. The support of a distribution $\mu \in \mathcal{D}(A)$ is $\mathsf{supp}(\mu) = \{a \in A \mid \mu(a) > 0\}$. A Dirac distribution is a distribution $\mu \in \mathcal{D}(A)$ such that $|\mathsf{supp}(\mu)| = 1$, i.e., there exists $a \in A$ such that $\mu(a) = 1$.

Given a set B and a σ-algebra $\mathcal{F}$ over B, we denote by $\mathcal{D}(B, \mathcal{F})$ the set of probability distributions over the measurable space $(B, \mathcal{F})$. Let $\mu \in \mathcal{D}(B, \mathcal{F})$ and $f\colon B \to \bar{\mathbb{R}}$ be a measurable function. We say that f is μ-integrable if it is integrable with respect to μ, i.e., if $\int_B |f| \mathrm{d}\mu \in \mathbb{R}$. We extend the Lebesgue integral to non-positive functions in the following way: if f is non-positive, we let $\int_B f \mathrm{d}\mu = -\int_B -f \mathrm{d}\mu$. If f is non-negative, non-positive or μ-integrable, we say that $\int_B f \mathrm{d}\mu$ is the μ-integral of f.

2.1.3 Topology notation

We only provide some notation in this section. We recall some classical definitions and results in Appendix A.1, including the definitions of the product topology, continuity, compactness and the usual topology of $\bar{\mathbb{R}}$.

Let $(X, \mathcal{T})$ be a Hausdorff topological space. For all $D \subseteq X$, we let $\mathsf{cl}(D)$ and $\mathsf{int}(D)$ denote the closure and interior of D. The boundary of $D \subseteq X$ is the set $\mathsf{bd}(D) = \mathsf{cl}(D) \setminus \mathsf{int}(D)$.

2.1.4 Vectors and geometry

Vector spaces

Vectors are written in boldface to distinguish them from scalars. Let $d \in \mathbb{N}_{>0}$. We let $\mathbf{0}_d$ and $\mathbf{1}_d \in \mathbb{R}^d$ respectively be the vectors of $\mathbb{R}^d$ where all components are zero and one respectively. We omit the dimension subscript whenever there is no ambiguity on the dimension of the space.

Given $\mathbf{v} = (v_j)_{1\leq j\leq d}, \mathbf{w} = (w_j)_{1\leq j\leq d} \in \mathbb{R}^d$, we let $\langle \mathbf{v}, \mathbf{w}\rangle = \sum_{j=1}^{d} v_j w_j$ denote the scalar product of $\mathbf{v}$ and $\mathbf{w}$. We let $\|\cdot\|_2$ denote the Euclidean norm on $\mathbb{R}^d$, defined by $\|\mathbf{v}\|_2 = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$ for all $\mathbf{v} \in \mathbb{R}^d$.

Given a linear map $L\colon \mathbb{R}^d \to \mathbb{R}^{d'}$ (where $d' \in \mathbb{N}_{>0}$), we let $\mathsf{ker}(L)$ denote the kernel of L, i.e., the set $\{\mathbf{v} \in \mathbb{R}^d \mid L(\mathbf{v}) = \mathbf{0}_{d'}\}$. A *linear form* is a linear map whose co-domain is $\mathbb{R}$. We denote linear forms by x^*, y^*, ...

The affine span of a set $D \subseteq \mathbb{R}^d$, which we denote by $\mathsf{aff}(D)$, is the smallest affine set (i.e., translation of a vector subspace of $\mathbb{R}^d$) in which D is included. Subsets of $\mathbb{R}^d$ whose affine span is not $\mathbb{R}^d$ have empty interior: strict affine subspaces of $\mathbb{R}^d$ cannot contain any ball with positive radius (the affine span of any such ball is $\mathbb{R}^d$). Instead of considering the interior of such sets, we consider their relative interior. The *relative interior* of a set $D \subseteq \mathbb{R}^d$, denoted by $\mathsf{ri}(D)$, is the interior of D as a subset of $\mathsf{aff}(D)$ (with the induced topology).

Let $D \subseteq \mathbb{R}^d$. The interior of D is a subset of $\mathsf{ri}(D)$ by definition. Furthermore, $\mathsf{int}(D)$ and $\mathsf{ri}(D)$ coincide if and only if $\mathsf{aff}(D) = \mathbb{R}^d$ or $\mathsf{ri}(D) = \emptyset$. Otherwise, these sets differ. For instance, the segment $[\mathbf{0}_2, \mathbf{1}_2] \subseteq \mathbb{R}^2$ has empty interior. However, we have $\mathsf{ri}([\mathbf{0}_2, \mathbf{1}_2]) =]\mathbf{0}_2, \mathbf{1}_2[$ (because $\mathsf{aff}([\mathbf{0}_2, \mathbf{1}_2])$ is the line of equation $x = y$).

Ordering vectors

We consider two order relations on $\bar{\mathbb{R}}^d$: the component-wise order and the lexicographic order. Let $\mathbf{q} = (q_j)_{1\leq j\leq d}$ and $\mathbf{p} = (p_j)_{1\leq j\leq d} \in \bar{\mathbb{R}}^d$. For the component-wise order, we write $\mathbf{q} \leq \mathbf{p}$ if and only if $q_j \leq p_j$ for all $1 \leq j \leq d$. For the lexicographic ordering over $\mathbb{R}^d$, we write $\mathbf{q} \leq_{\mathsf{lex}} \mathbf{p}$ if and only if $\mathbf{q} = \mathbf{p}$ or $q_j \leq p_j$ where $j = \min\{j' \leq d \mid q_{j'} \neq p_{j'}\}$. We write $\mathbf{q} <_{\mathsf{lex}} \mathbf{p}$ if $\mathbf{q} \leq_{\mathsf{lex}} \mathbf{p}$ and $\mathbf{q} \neq \mathbf{p}$. We recall that the component-wise order is partial, whereas the lexicographic order is a total order.

Let $D \subseteq \bar{\mathbb{R}}^d$. We say that $\mathbf{q} \in D$ is a *Pareto-optimal* element of D if it is maximal for the component-wise order, i.e., if there does not exist $\mathbf{p} \in D$ such that $\mathbf{q} \leq \mathbf{p}$ and $\mathbf{q} \neq \mathbf{p}$. We say that D is *downward-closed* if for all $\mathbf{q} \in D$ and $\mathbf{p} \in \bar{\mathbb{R}}^d$, $\mathbf{p} \leq \mathbf{q}$ implies $\mathbf{p} \in D$. We let $\mathsf{down}(D)$ denote the *downward closure* of D, which is defined as the smallest (with respect to set inclusion) downward-closed set in which D is included. A set and its downward closure have the same set of Pareto-optimal elements.

Convexity

A *convex combination* of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^d$ is a linear combination $\sum_{m=1}^{n} \alpha_m \cdot \mathbf{v}_m$ such that $\alpha_1, \ldots, \alpha_n \in [0, 1]$ and $\sum_{m=1}^{n} \alpha_m = 1$. We refer to a sequence of coefficients $\alpha_1, \ldots, \alpha_n \in [0, 1]$ such that $\sum_{m=1}^{n} \alpha_m = 1$ as *convex combination coefficients*. Given $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d$, we let $[\mathbf{v}, \mathbf{w}] = \{\alpha \cdot \mathbf{v} + (1 - \alpha)\mathbf{w} \mid \alpha \in [0, 1]\}$ denote the (closed) segment from $\mathbf{v}$ to $\mathbf{w}$; it is the set of convex combinations of $\mathbf{v}$ and $\mathbf{w}$. Open and half-open segments are defined analogously.

Let $D \subseteq \mathbb{R}^d$. The *convex hull* of D, denoted by $\mathsf{conv}(D)$, is the set of all convex combinations of elements of D. The set D is *convex* if for all $\mathbf{v}, \mathbf{w} \in D$, $[\mathbf{v}, \mathbf{w}] \subseteq D$, or, equivalently, if $D = \mathsf{conv}(D)$. If D is convex, we say that $\mathbf{q} \in D$ is an *extreme point* of D if $\mathbf{q} \notin \mathsf{conv}(D \setminus \{\mathbf{q}\})$, i.e., if $\mathbf{q}$ is not a convex combination of elements of D other than $\mathbf{q}$ and we let $\mathsf{extr}(D)$ denote the set of extreme points of D. Extreme points generalise the notion of vertices of polytopes.

The definition of a convex combination does not bound the number of involved vectors. However, in $\mathbb{R}^d$, it is sufficient to only consider convex combinations involving no more than $d + 1$ vectors. This is formalised by the following theorem.

Theorem 2.1 (Carathéodory's theorem for convex hulls [Roc70, Thm. 17.1])**.** *Let $D \subseteq \mathbb{R}^d$ and $\mathbf{q} \in \mathsf{conv}(D)$. There exists $D' \subseteq D$ such that $|D'| \leq d + 1$ and $\mathbf{q} \in \mathsf{conv}(D')$.*

Carathéodory's theorem can be used to show that the convex hull of a compact subset D of $\mathbb{R}^d$ is itself compact. It follows from the theorem that a sequence of elements in $\mathsf{conv}(D)$ can be described by $d + 1$ sequences of vectors in the compact set D and $d + 1$ sequences of convex combination coefficients in the compact interval $[0, 1]$. We can use the compactness of all of these sets to extract a convergent sequence from any sequence in $\mathsf{conv}(D)$. We provide a formal argument in Appendix A.3.

Lemma 2.2. *Let $d \in \mathbb{N}_{>0}$. Let $D \subseteq \mathbb{R}^d$. If D is compact, then $\mathsf{conv}(D)$ is also compact.*

Convexity does not generalise to $\bar{\mathbb{R}}^d$. Most notably, convex combinations of vectors of $\bar{\mathbb{R}}^d$ (defined in the same way as above) may be ill-defined. Although we consider convex combinations of elements of $\bar{\mathbb{R}}^d$ in Part IV, these are always guaranteed to be well-defined: we will not consider convex combinations where $+\infty$ and $-\infty$ both occur on a dimension in two vectors of the convex combination.

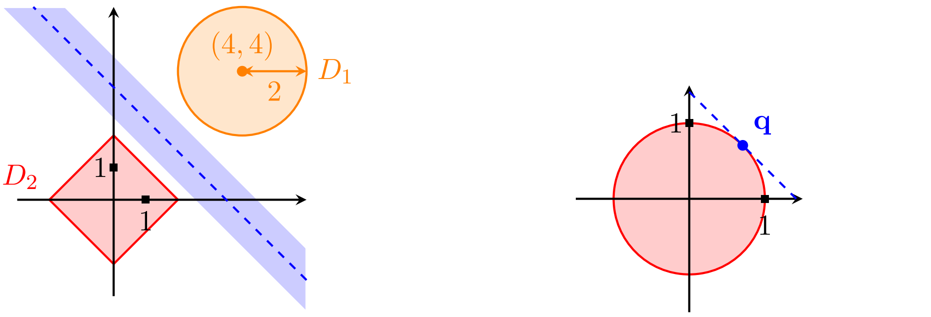
Hyperplane separation

A *hyperplane* H of $\mathbb{R}^d$ is a set of the form $\{\mathbf{v} \in \mathbb{R}^d \mid x^*(\mathbf{v}) = \alpha\}$ for some non-zero linear form x^* and $\alpha \in \mathbb{R}$. Let D_1 and $D_2 \subseteq \mathbb{R}^d$. The sets D_1 and D_2 are *strongly separated* by a hyperplane if there exists a non-zero linear form x^* such that $\inf_{\mathbf{q} \in D_1} x^*(\mathbf{q}) > \sup_{\mathbf{p} \in D_2} x^*(\mathbf{p})$. A convex set $D \subseteq \mathbb{R}^d$ is *supported* by a hyperplane at $\mathbf{q} \in D$ if there exists a non-zero linear form x^* such that, for all $\mathbf{p} \in D$, $x^*(\mathbf{p}) \leq x^*(\mathbf{q})$; a *supporting hyperplane* in this case is $H = \{\mathbf{v} \in \mathbb{R}^d \mid x^*(\mathbf{v}) = x^*(\mathbf{q})\}$. We provide an illustration of the notions of separating and supporting hyperplanes in Figure 2.1.

We recall a variant of the hyperplane separation theorem and the supporting hyperplane theorem. We first outline a sufficient condition such that two disjoint convex sets can be strongly separated.

Theorem 2.3 (Hyperplane separation theorem [Roc70, Cor. 11.4.2])**.** *Let D_1 and D_2 be two convex subsets of $\mathbb{R}^d$. If $\mathsf{cl}(D_1) \cap \mathsf{cl}(D_2) = \emptyset$ and D_1 or D_2 is bounded, then there exists a hyperplane strongly separating D_1 and D_2.*

Figure 2.1a illustrates a setup in which we can apply the theorem. Theorem 2.3 can be applied whenever D_1 is a singleton set $\{\mathbf{q}\}$ (it is bounded) and $\mathbf{q} \notin \mathsf{cl}(D_2)$ to separate $\mathbf{q}$ from $\mathsf{cl}(D_2)$. The next theorem provides a sufficient condition for the existence of a supporting hyperplane at a given point of a convex set.



(a) The dashed blue line ($x + y = \frac{7}{2}$) strongly separates D_1 and D_2: they lie on a different side of the line and there is a strip around the line that is disjoint from D_1 and D_2.

(b) The dashed blue line ($x + y = \sqrt{2}$) supports the unit ball for the Euclidean norm in $\mathbb{R}^2$ at $\mathbf{q} = \frac{\sqrt{2}}{2}\mathbf{1}$.

Figure 2.1: Illustration of separating and supporting hyperplanes.

Theorem 2.4 (Supporting hyperplane theorem [Roc70, Thm. 11.6])**.** *Let* $D \subseteq \mathbb{R}^d$ *be convex and* $\mathbf{q} \in D$*. If* $\mathbf{q} \notin \mathsf{ri}(D)$*, then there exists a hyperplane* H *supporting* D *at* $\mathbf{q}$ *such that* $D \not\subseteq H$*.*

2.2 Arenas and Markov decision processes

All game models that we study in the sequel can be formalised as a special case of multi-player concurrent stochastic games played on graphs.

When considering an n-player game (where $n \in \mathbb{N}_{>0}$), we denote player i by $\mathcal{P}_i$ for all $i \in [\![1, n]\!]$. At the start of a play, a pebble is placed on some initial state (i.e., a vertex of the graph). In each round, all players simultaneously select an action available in the current state and the next state is chosen randomly following a distribution depending only on the current state and the actions chosen by the players. The game proceeds for an infinite number of rounds, yielding an infinite play.

The formal structures on which such games are played are called *arenas*.

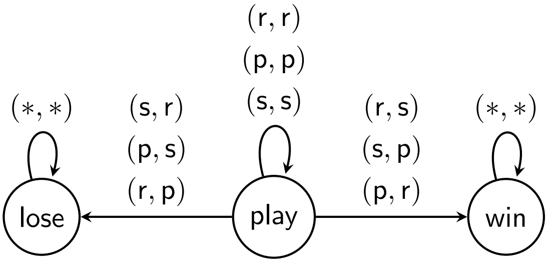


Figure 2.2: A concurrent arena modelling rock-paper-scissors. The self-loops of states win and lose can be taken with all pairs of actions.

Definition 2.5 (Multi-player concurrent stochastic arena)**.** Let $n \in \mathbb{N}_{>0}$. An n-player *(perfect-information) concurrent stochastic arena*, or simply an *arena*, is a tuple $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$ where S is a non-empty countable set of states, $A^{(i)}$ is a countable set of actions for each $i \in [\![1, n]\!]$ and $\delta \colon S \times \prod_{i\in[\![1,n]\!]} A^{(i)} \to \mathcal{D}(S)$ is a (partial) probabilistic transition function. The arena $\mathcal{A}$ is *finite* if S is finite and $A^{(i)}$ is finite for all $i \in [\![1, n]\!]$.

For two-player arenas, we slightly change the notation and denote them by tuples $(S, A^{(1)}, A^{(2)}, \delta)$. A one-player arena is called a *Markov decision process* (MDP), and we denote MDPs by $\mathcal{M} = (S, A, \delta)$.

We fix an n-player arena $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$. We let $\bar{A} = \prod_{i\in[\![1,n]\!]} A^{(i)}$. Elements of $\bar{A}$ are called *action profiles.* We denote action profiles with a bar to emphasise that they are tuples of actions. Given $\bar{a} \in \bar{A}$, we adopt the convention that $\bar{a}$ is given by the tuple $(a^{(1)}, \dots, a^{(n)})$.

For any state $s \in S$, we let $\bar{A}(s) = \{\bar{a} \in \bar{A} \mid \delta(s, \bar{a}) \text{ is defined}\}$ and require that there exist subsets $A^{(i)}(s)$ of $A^{(i)}$ for all $i \in [\![1, n]\!]$ such that $\bar{A}(s) = \prod_{i\in[\![1,n]\!]} A^{(i)}(s)$. In other words, the actions available to a player in a state are not constrained by the choices of the others. We assume without loss of generality that for all $s \in S$, $\bar{A}(s)$ is non-empty, i.e., there are no deadlocks in the arena.

We now present a simple two-player arena and an MDP to illustrate the definition of an arena.

Example 2.1 (Rock paper scissors)**.** Rock paper scissors is a two-player game where two players simultaneously choose an action among rock, paper and

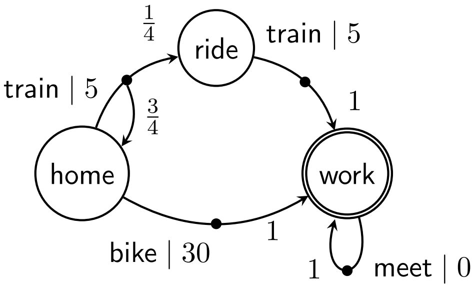


Figure 2.3: An MDP. The numerical weights next to actions represent the time taken by each action.

scissors. In this game, rock beats scissors, scissors beats paper and paper beats rock; the player whose action beats the action of the other wins. This game can be modelled by the two-player arena depicted in Figure 2.2, in which it is assumed that players replay whenever a tie occurs. Formally, this arena is given by the tuple $(S, A^{(1)}, A^{(2)}, \delta)$ where $S = \{\mathsf{play}, \mathsf{win}, \mathsf{lose}\}$, $A^{(1)} = A^{(2)} = \{\mathsf{r}, \mathsf{p}, \mathsf{s}\}$ where the actions r, p, s respectively represent rock, paper and scissors, and the (deterministic) transition function δ is as depicted in the figure, e.g., we have $\delta(\mathsf{play}, \mathsf{r}, \mathsf{p})(\mathsf{lose}) = \delta(\mathsf{play}, \mathsf{s}, \mathsf{p})(\mathsf{win}) = 1$. $\lhd$

Example 2.2. Figure 2.3 depicts an MDP $\mathcal{M}$ representing a situation where a person can choose between taking their bicycle or the train to reach work. Whether a delay occurs is modelled by the stochastic transition between states home and ride. In this example, a numerical weight is assigned to each state-action pair: weights represent the time taken by each action, and model the fact that taking the train to work takes less time than taking the bicycle.

Formally, we have $\mathcal{M} = (S, A, \delta)$ where $S = \{\mathsf{home}, \mathsf{ride}, \mathsf{work}\}$, $A = \{\mathsf{train}, \mathsf{bike}, \mathsf{meet}\}$ and the transition function δ is as depicted on the illustration, e.g., $\delta(\mathsf{home}, \mathsf{train})(\mathsf{ride}) = 1 - \delta(\mathsf{home}, \mathsf{train})(\mathsf{home}) = \frac{1}{4}$. $\lhd$

A *play* of $\mathcal{A}$ is an infinite sequence $s_0\bar{a}_0s_1\ldots \in (S\bar{A})^\omega$ such that for all $\ell \in \mathbb{N}$, $\delta(s_\ell, \bar{a}_\ell)(s_{\ell+1}) > 0$. A *history* is a finite prefix of a play ending in a state. Given a play $\pi = s_0\bar{a}_0s_1\bar{a}_1\ldots$ and $\ell \in \mathbb{N}$, we write $\pi_{\leq \ell}$ for the prefix history $s_0\bar{a}_0\ldots\bar{a}_{\ell-1}s_\ell$ and $\pi_{\geq \ell}$ for the suffix play $s_\ell\bar{a}_\ell s_{\ell+1}\ldots$, and use the same notation for prefixes and suffixes of histories. For any history $h = s_0\bar{a}_0\ldots\bar{a}_{k-1}s_k$, we let $\mathsf{first}(h) = s_0$ and $\mathsf{last}(h) = s_k$. Similarly, for a play π, we denote its first state by $\mathsf{first}(\pi)$. We write $\mathsf{Plays}(\mathcal{A})$ to denote the set of plays of $\mathcal{A}$, $\mathsf{Hist}(\mathcal{G})$ to denote the set of histories of $\mathcal{A}$. Given some initial state $s_{\mathsf{init}} \in S$, we write $\mathsf{Hist}(\mathcal{A}, s_{\mathsf{init}})$ for the set of histories starting in state s_{init}.

Let $h = s_0\bar{a}_0 s_1 \dots \bar{a}_{\ell-1} s_\ell$ and $h' = s_\ell \bar{a}_{\ell+1} s_{\ell+1} \dots \bar{a}_{r-1} s_r$ be two histories such that $\mathsf{last}(h) = \mathsf{first}(h')$. We let $h \cdot h' = s_0\bar{a}_0 s_1 \dots \bar{a}_{\ell-1} s_\ell \bar{a}_{\ell+1} s_{\ell+1} \dots \bar{a}_{r-1} s_r$ denote the concatenation of h and h' without repeating state s_ℓ. We abusively call $h \cdot h'$ the *concatenation* of h and h'. The concatenation $h \cdot \pi$ of a history h and a play π such that $\mathsf{last}(h) = \mathsf{first}(\pi)$ is defined similarly.

Concurrent stochastic multi-player arenas subsume several models that have been studied in their own right. First, there are the previously mentioned special cases of finite arenas, MDPs and two-player arenas. Second, there is the class of turn-based arenas. An arena is turn-based if at each round, only one player can influence the next transition.

Definition 2.6. The arena $\mathcal{A} = (S, (A^{(i)})_{i \in [\![1,n]\!]}, \delta)$ is *turn-based* if for all states $s \in S$, there exists $i^\star \in [\![1,n]\!]$ such that, for all $i \in [\![1,n]\!] \setminus \{i^\star\}$, $|A^{(i)}(s)| = 1$; we say that $\mathcal{P}_{i^\star}$ controls s.

Turn-based arenas are traditionally described by a partition of the state space into states controlled by the different players. We use this presentation when dealing with turn-based arenas. Formally, if $\mathcal{A} = (S, (A^{(i)})_{i \in [\![1,n]\!]}, \delta)$ is turn-based, we present it as a tuple $((S_i)_{i \in [\![1,n]\!]}, A, \delta')$ where $(S_i)_{i \in [\![1,n]\!]}$ is a partition of S, $A = \bigcup_{i \in [\![1,n]\!]} A^{(i)}$ is the set of all actions and, for all $i \in [\![1,n]\!]$, $s \in S_i$ and $\bar{a} \in \bar{A}(s)$, $\delta'(s, a^{(i)}) = \delta(s, \bar{a})$.

Assume that $\mathcal{A}$ is turn-based. In this case, we view the plays of $\mathcal{A}$ as elements of $(SA)^\omega$ instead of elements of $(S\bar{A})^\omega$. Similarly, histories are seen as elements of $(SA)^*S$. Intuitively, we omit the information related to players who have no choice. Definitions presented in the sequel for concurrent arenas can be adapted to the turn-based setting in this way. We let, for all $i \in [\![1,n]\!]$, $\mathsf{Hist}_i(\mathcal{A}) = \mathsf{Hist}(\mathcal{A}) \cap (SA)^*S_i$ denote the set of histories ending in a state controlled by $\mathcal{P}_i$.

A class of arenas of particular interest consists of the arenas with transitions that are not subject to randomness.

Definition 2.7. An arena $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$ is *deterministic* if for all $s \in S$ and $\bar{a} \in \bar{A}(s)$, $\delta(s, \bar{a})$ is a Dirac distribution.

When dealing with a deterministic arena, we view the transition function δ as a function $\delta\colon S \times \bar{A} \to S$. For instance, the arena of Example 2.1 is deterministic, unlike the MDP of Example 2.2. A deterministic MDP is a graph whose edges are labelled by actions.

Markov chains are a class of stochastic processes. We view Markov chains as a special case of arenas: a Markov chain can be seen as an arena where all players only have one action. Due to the absence of action choices, we omit actions from Markov chains, and use the following definition.

Definition 2.8. A (discrete-time) *Markov chain* is a tuple $\mathcal{C} = (S, \delta)$ where S is a countable set of states and $\delta\colon S \to \mathcal{D}(S)$ is a probabilistic transition function.

Let $\delta\colon S \to \mathcal{D}(S)$ be a Markov chain. Plays and histories of $\mathcal{C}$ are defined similarly to those of an arena, except we omit actions. More precisely, a *play* of $\mathcal{C}$ is a sequence $s_0s_1s_2\ldots \in S^\omega$ such that, for all $\ell \in \mathbb{N}$, $\delta(s_\ell)(s_{\ell+1}) > 0$. A *history* of $\mathcal{C}$ is a finite prefix of a play of $\mathcal{C}$. We use the notation $\mathsf{Plays}(\mathcal{C})$ and $\mathsf{Hist}(\mathcal{C})$ for the set of histories and plays of $\mathcal{C}$, like for arenas.

2.3 Topology on the set of plays

We present the usual topology on the set of plays of an arena. Probability distributions over the set of plays of an arena (when the non-determinism is resolved) or of a Markov chain are defined over the Borel σ-algebra for this topology, i.e., the σ-algebra generated by open sets of plays.

Let $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$ be an n-player arena. We endow $\mathsf{Plays}(\mathcal{A})$ with a metrisable topology as follows. First, we equip $(S\bar{A})^\omega$ with the product topology, where S and $\bar{A}$ are both equipped with the discrete topology. It follows that $(S\bar{A})^\omega$ is a metrisable topological space (as a countable product of metrisable spaces). The topology of $\mathsf{Plays}(\mathcal{A})$ is the topology induced on $\mathsf{Plays}(\mathcal{A})$ by the topology of $(S\bar{A})^\omega$.

A base of the topology of $\mathsf{Plays}(\mathcal{A})$ is the set of cylinder sets (in the sense of general topology, see Appendix A.1). We define, for any $h \in \mathsf{Hist}(\mathcal{A})$, the *cylinder* of h as the set $\mathsf{Cyl}_{\mathcal{A}}(h) = \{\pi \in \mathsf{Plays}(\mathcal{A}) \mid h \text{ is a prefix of } \pi\}$, consisting of plays that extend h. For any set of histories $\mathcal{H} \subseteq \mathsf{Hist}(\mathcal{A})$, we write $\mathsf{Cyl}_{\mathcal{A}}(\mathcal{H}) = \bigcup_{h \in \mathcal{H}} \mathsf{Cyl}_{\mathcal{A}}(h)$. Two history cylinders intersect if and only if one of the histories is a prefix of another. In particular, for a prefix-free $\mathcal{H} \subseteq \mathsf{Hist}(\mathcal{A})$, the union defining $\mathsf{Cyl}_{\mathcal{A}}(\mathcal{H})$ is disjoint. We drop the subscript $\mathcal{A}$ from the notation of cylinders when the arena is clear for the context.

It can be shown that the set of cylinders of histories are also a base of the topology of $\mathsf{Plays}(\mathcal{A})$. We provide a proof in Appendix A.4.

Lemma 2.9. *The set* $\{\mathsf{Cyl}(h) \mid h \in \mathsf{Hist}(\mathcal{A})\}$ *of history cylinders is a base of the topology of* $\mathsf{Plays}(\mathcal{A})$.

Since $\mathsf{Hist}(\mathcal{A})$ is countably infinite, Lemma 2.9 implies that all open subsets of $\mathsf{Plays}(\mathcal{A})$ are countable unions of history cylinders. Therefore, the topology of $\mathsf{Plays}(\mathcal{A})$ is a subset of the σ-algebra generated by history cylinders. In particular, the σ-algebra generated by history cylinders is the Borel σ-algebra (for the standard topology) of $\mathsf{Plays}(\mathcal{A})$.

As mentioned above, the topology of $\mathsf{Plays}(\mathcal{A})$ is metrisable. It is induced, e.g., by the metric $\mathsf{dist}_{\mathsf{play}}$ over $\mathsf{Plays}(\mathcal{A})$ defined by, for all $\pi, \pi' \in \mathsf{Plays}(\mathcal{M})$, $\mathsf{dist}_{\mathsf{play}}(\pi, \pi') = 2^{-r}$, where $r = 0$ if $\mathsf{first}(\pi) \neq \mathsf{first}(\pi')$, and, otherwise, $r = \sup\{\ell \in \mathbb{N}_{>0} \mid \pi_{\leq \ell-1} = \pi'_{\leq \ell-1}\}$. This distance can be derived from the discrete metric and a standard argument to prove that countable products of metric spaces are metrisable. Open balls with a positive radius for $\mathsf{dist}_{\mathsf{play}}$ are history cylinders; it follows from Lemma 2.9 that $\mathsf{dist}_{\mathsf{play}}$ induces the usual topology of $\mathsf{Plays}(\mathcal{A})$.

Finite topological spaces are compact. Therefore, if $\mathcal{A}$ is finite, $(S\bar{A})^\omega$ is a compact space: any product of compact topological spaces is compact (see Theorem A.3 for countable products). We show in Appendix A.4 that $\mathsf{Plays}(\mathcal{A})$ is a closed subset of $(S\bar{A})^\omega$, and is therefore also a compact space whenever $\mathcal{A}$ is finite.

Lemma 2.10. *The set* $\mathsf{Plays}(\mathcal{A})$ *is a closed subset of* $(S\bar{A})^\omega$. *In particular,* $\mathsf{Plays}(\mathcal{A})$ *is a compact space whenever* $\mathcal{A}$ *is finite.*

2.4 Strategies

2.4.1 Definition

A strategy is a function that describes how a player should play. In the perfect information setting, players observe the entirety of the play history when making decisions. Arenas with imperfect information, in which players must make their choices based on observations that could represent several states rather than the states themselves, are introduced in Section 2.7.

In general, a strategy can use the whole past of the current play in its decision making, i.e., a strategy can use an unbounded amount of memory. Furthermore, players need not act in a deterministic fashion: they can use randomisation to select an action. Formally, strategies are defined as follows.

Definition 2.11. A *(behavioural) strategy* of $\mathcal{P}_i$ is a function $\sigma_i \colon \mathsf{Hist}(\mathcal{A}) \to \mathcal{D}(A^{(i)})$ such that for all histories $h \in \mathsf{Hist}(\mathcal{A})$, $\mathsf{supp}(\sigma_i(h)) \subseteq A^{(i)}(\mathsf{last}(h))$.

In other words, a strategy assigns, to any history, a distribution over the actions available to $\mathcal{P}_i$ in the last state of the history. In the turn-based setting, we view strategies of $\mathcal{P}_i$ as functions over the set of histories ending in a state controlled by $\mathcal{P}_i$ (i.e., $\mathsf{Hist}_i(\mathcal{A})$).

A strategy that only uses information on the current state of the play is called memoryless: a strategy σ_i of $\mathcal{P}_i$ is *memoryless* if for all histories $h, h' \in \mathsf{Hist}(\mathcal{A})$, $\mathsf{last}(h) = \mathsf{last}(h')$ implies $\sigma_i(h) = \sigma_i(h')$. Memoryless strategies can be viewed as functions $S \to \mathcal{D}(A^{(i)})$. A strategy is called *pure* if it does not use randomisation. A pure strategy of $\mathcal{P}_i$ can be viewed as a function $\mathsf{Hist}(\mathcal{A}) \to A^{(i)}$. Strategies that are both memoryless and pure can be viewed as functions $S \to A^{(i)}$.

We write $\Sigma^i(\mathcal{A})$ for the set of all (behavioural) strategies of $\mathcal{P}_i$ in $\mathcal{A}$ and $\Sigma^i_{\mathsf{pure}}(\mathcal{A})$ for the set of pure strategies of $\mathcal{P}_i$ in $\mathcal{A}$.

A strategy profile is a tuple made of one strategy per player.

Definition 2.12. A *strategy profile* is a tuple $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$ of strategies where σ_i is a strategy of $\mathcal{P}_i$ for all $i \in [\![1, n]\!]$.

For all $i \in [\![1, n]\!]$, to highlight the strategy of $\mathcal{P}_i$ in a strategy profile $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$, we write $\sigma = (\sigma_i, \sigma_{-i})$; we (abusively) say that σ_{-i} is a strategy profile of the players other than $\mathcal{P}_i$.

We say that a strategy profile is *pure* (resp. *memoryless*) if all strategies in the profile are pure (resp. memoryless).

We change our terminology and notation slightly whenever $\mathcal{A}$ is an MDP, i.e., when there is only a single player. Instead of referring to strategies of the unique player, we instead refer to strategies of the MDP. We denote strategies without any reference to a player, e.g., we write σ instead of σ_1. For sets of strategies, we drop the exponent from the notations $\Sigma^1(\mathcal{A})$ and $\Sigma^1_{\mathsf{pure}}(\mathcal{A})$, and write $\Sigma(\mathcal{A})$ and $\Sigma_{\mathsf{pure}}(\mathcal{A})$ respectively.

2.4.2 Outcomes and probabilities over plays

Let $i \in [\![1, n]\!]$, σ_i be a strategy of $\mathcal{P}_i$ and σ be a strategy profile. A play or play prefix $s_0\bar{a}_0s_1\dots$ is *consistent* with σ_i if for all action indices ℓ, it holds that $\sigma_i(s_0\bar{a}_0\dots s_\ell)(a_\ell^{(i)}) > 0$.[1] A play or play prefix is consistent with σ if it is consistent with all strategies in σ. A play that is consistent with σ_i (respectively σ) is called an *outcome* of σ_i (respectively σ).

Let $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$ be a strategy profile and s_{init} be an initial state. The strategy profile σ induces a Markov chain over the set of histories of $\mathcal{A}$ starting in s, which yields a distribution over plays of $\mathcal{A}$. We do not formalise this Markov chain, and instead directly define the distribution over Borel subsets of $\mathsf{Plays}(\mathcal{A})$. This presentation allows us to avoid having to formalise distributions over plays of Markov chains first.

We write $\mathcal{F}_\mathcal{A}$ for the Borel σ-algebra of $\mathsf{Plays}(\mathcal{A})$. We define the probability measure $\mathbb{P}^\sigma_{\mathcal{A},s_{\mathsf{init}}} \in \mathcal{D}(\mathsf{Plays}(\mathcal{A}), \mathcal{F}_\mathcal{A})$ induced by following σ from s_{init} in $\mathcal{A}$ as follows. For any history $h = s_0\bar{a}_0\dots s_r \in \mathsf{Hist}(\mathcal{A}, s_{\mathsf{init}})$, we define

$$\mathbb{P}^\sigma_{\mathcal{A},s_{\mathsf{init}}}(\mathsf{Cyl}_\mathcal{A}(h)) = \prod_{\ell=0}^{r-1} \left(\delta(s_\ell, \bar{a}_\ell)(s_{\ell+1}) \cdot \prod_{i=1}^{n} \sigma_i(s_0\bar{a}_0\dots s_\ell)(a_\ell^{(i)}) \right).$$

For any history $h \in \mathsf{Hist}(\mathcal{A}) \setminus \mathsf{Hist}(\mathcal{A}, s_{\mathsf{init}})$, we set $\mathbb{P}^\sigma_{s_{\mathsf{init}}}(\mathsf{Cyl}_\mathcal{A}(h)) = 0$. By the Ionescu-Tulcea extension theorem [Kal21, Thm. 8.24], the measure described

[1] We use the terminology of consistency not only for plays and histories, but also for prefixes of plays that end with an action profile.

above can be extended in a unique fashion to $(\mathsf{Plays}(\mathcal{A}), \mathcal{F}_{\mathcal{A}})$. Whenever $\mathcal{A}$ is clear from the context, we drop it from the notation, i.e., we write $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}$ instead of $\mathbb{P}^{\sigma}_{\mathcal{A},s_{\mathsf{init}}}$.

Remark 2.13 (Probability of inconsistent plays). Let $i \in [\![1, n]\!]$ and σ_i be a strategy of $\mathcal{P}_i$. For any history $h \in \mathsf{Hist}(\mathcal{A})$, if h is not consistent with σ_i, then $\mathbb{P}^{\sigma_i,\sigma_{-i}}_{\mathsf{first}(h)}(\mathsf{Cyl}\,(h)) = 0$ for all strategy profiles σ_{-i} of the players other than $\mathcal{P}_i$. Since the set of plays that is inconsistent with σ_i can be written as the union of the cylinders of histories that are inconsistent with σ_i, it follows that the probability of this set when $\mathcal{P}_i$ follows σ_i is zero no matter the initial state and strategy profile of the other players. ◁

For a Markov chain $\mathcal{C}$ with state space S, the cylinder of a history is defined similarly for arenas, and so is the distribution over plays of $\mathcal{C}$. As there are no action choices in $\mathcal{C}$, we omit the strategy from the notation above: we write $\mathbb{P}_{\mathcal{C},s_{\mathsf{init}}}$ to highlight $\mathcal{C}$, and, if $\mathcal{C}$ is clear from the context, we write $\mathbb{P}_{s_{\mathsf{init}}}$.

If $\mathcal{A}$ is deterministic and σ_i is pure for all $i \in [\![1, n]\!]$, then for all states s, there is a single play consistent with σ starting in s; we denote this unique play by $\mathsf{Out}_{\mathcal{A}}(\sigma, s)$.

When dealing with memoryless strategy profiles, we can seen the Markov chains induced by the strategy profile as a Markov chain over S. We use this vision of induced Markov chains in Part V to study the probability of some events in large and countable Markov decision processes. For this reason (and to lighten the notation), we only provide a formal decision for Markov decision processes.

Definition 2.14. Let $\mathcal{M} = (S, A, \delta)$ be an MDP and σ be a memoryless strategy of $\mathcal{M}$. We define the *Markov chain induced by σ on $\mathcal{M}$* as the Markov chain (S, δ') such that, for all s, $s' \in S$, $\delta'(s)(s') = \sum_{a \in A(s)} \sigma(s)(a) \cdot \delta(s, a)(s')$.

While this definition does abstract actions away from plays and histories, it preserves the probability of reaching one state from another in the MDP when following the strategy.

2.4.3 Mixed strategies

Randomised strategies in the sense of Definition 2.11 are called *behavioural strategies*. Intuitively, a behavioural strategy casts a die at each step of the play to select an action. Another way of integrating randomisation in decision making is through *mixing*: a *mixed strategy* is a distribution over pure strategies. When playing with a mixed strategy, a pure strategy is chosen at the beginning of the play and is followed for the whole play. With respect to the previous intuition, in this case, we cast a die once at the start of a play then no longer use randomisation.

To formally define mixed strategies, we must introduce a σ-algebra over the set of pure strategies of $\mathcal{A}$. The set of pure strategies $\Sigma^i_{\mathsf{pure}}(\mathcal{A})$ of $\mathcal{P}_i$ in $\mathcal{A}$ can be written as $\prod_{h\in\mathsf{Hist}(\mathcal{A})} A^{(i)}(\mathsf{last}(h))$. We let $\mathcal{F}_{\Sigma^i_{\mathsf{pure}}(\mathcal{A})}$ denote the σ-algebra generated by sets of the form

$$\left\{\sigma_i \in \Sigma^i_{\mathsf{pure}}(\mathcal{A}) \mid \sigma_i(h) = \tau_i(h) \text{ for all } h \in \mathcal{H}\right\}$$

where $\tau_i \in \Sigma^i_{\mathsf{pure}}(\mathcal{A})$ and $\mathcal{H}$ is a finite set of histories. Such sets are cylinders of $\prod_{h\in\mathsf{Hist}(\mathcal{A})} A^{(i)}(\mathsf{last}(h))$ in the sense of the product topology, when we endow the subsets of $A^{(i)}$ in the previous product with the discrete topology. We formalise mixed strategies as follows.

Definition 2.15. A *mixed strategy* of $\mathcal{P}_i$ in $\mathcal{A}$ is a probability distribution $\mu_i \in \mathcal{D}(\Sigma^i_{\mathsf{pure}}(\mathcal{A}), \mathcal{F}_{\Sigma^i_{\mathsf{pure}}(\mathcal{A})})$. A *mixed strategy profile* is a tuple of the form $(\mu_i)_{i\in[\![1,n]\!]}$ where μ_i is a mixed strategy of $\mathcal{P}_i$ for all $i \in [\![1,n]\!]$.

We assume that pure strategies are a special case of mixed strategies by identifying pure strategies with the corresponding Dirac mixed strategy.

Let $\mu = (\mu_i)_{i\in[\![1,n]\!]}$ be a mixed strategy profile and $s_{\mathsf{init}} \in S$ be an initial state. The distribution $\mathbb{P}^{\mu}_{\mathcal{A},s_{\mathsf{init}}}$ over $\mathsf{Plays}(\mathcal{A})$ induced by μ from s_{init} is defined, for all histories $h = s_0\bar{a}_0 s_1 \ldots \bar{a}_{r-1}s_r \in \mathsf{Hist}(\mathcal{A})$, by $\mathbb{P}^{\mu}_{\mathcal{A},s_{\mathsf{init}}}(\mathsf{Cyl}(h)) = 0$ if $s_0 \neq s_{\mathsf{init}}$, and, otherwise, is defined by

$$\mathbb{P}^{\mu}_{\mathcal{A},s_{\mathsf{init}}}(\mathsf{Cyl}(h)) = \left(\prod_{i\in[\![1,n]\!]} \mu_i(\Sigma^i_h)\right) \cdot \left(\prod_{\ell=0}^{r-1} \delta(s_\ell, \bar{a}_\ell)(s_{\ell+1})\right)$$

where $\Sigma^i_h = \{\sigma_i \in \Sigma^i_{\mathsf{pure}}(\mathcal{A}) \mid h \text{ is consistent with } \sigma_i\}$ for all $i \in [\![1, n]\!]$. The Ionescu-Tulcea extension theorem ensures that the partially-defined measure above can be extended in a unique fashion to $(\mathsf{Plays}(\mathcal{A}), \mathcal{F}_{\mathcal{A}})$.

Alternatively, the distribution induced by a profile of mixed strategies can be written as an integral over all pure strategies profiles for the product distribution. We require the following technical property to formally present this definition. We defer its proof, based on induction on the Borel hierarchy, to Appendix A.5.

Lemma 2.16. *Let $\Omega \subseteq \mathsf{Plays}(\mathcal{A})$ be measurable and let $s \in S$. The function $P_\Omega \colon \prod_{i=1}^n \Sigma^i_{\mathsf{pure}}(\mathcal{A}) \to [0,1] : \sigma \to \mathbb{P}^\sigma_s(\Omega)$ is measurable.*

The following lemma provides an equivalent definition of the distribution induced by a mixed strategy. A proof is also provided in Appendix A.5.

Lemma 2.17. *Let $\mu = (\mu_i)_{i\in[\![1,n]\!]}$ be a mixed strategy profile and $s_{\mathsf{init}} \in S$ be an initial state. Let $\mu_1 \times \cdots \times \mu_n$ denote the (unique) product measure over $\prod_{i=1}^n \Sigma^i_{\mathsf{pure}}(\mathcal{A})$ obtained from $\mu_1, \cdots, \mu_n$. For all measurable $\Omega \subseteq \mathsf{Plays}(\mathcal{A})$, we have*

$$\mathbb{P}^\mu_{\mathcal{A},s_{\mathsf{init}}}(\Omega) = \int_{\sigma\in\prod_{i=1}^n \Sigma^i_{\mathsf{pure}}(\mathcal{A})} \mathbb{P}^\sigma_{\mathcal{A},s_{\mathsf{init}}}(\Omega)\mathrm{d}(\mu_1 \times \cdots \times \mu_n)(\sigma).$$

In the *perfect recall* setting, i.e., when players can remember all of their past information and the actions they have chosen, behavioural and mixed strategies share the same expressive power. This result is known as Kuhn's theorem [Aum64]. Perfect information and perfect recall are not equivalent: perfect information is a special case of perfect recall, where players are fully informed. In Section 2.7, we define arenas with imperfect information and formalise perfect recall. We also formally state Kuhn's theorem in that section.

2.4.4 Finite-memory strategies

A strategy is said to be *finite-memory* if it can be encoded by a Mealy machine, i.e., an automaton with outputs along its edges. We can include randomisation in the initialisation, outputs and updates (i.e., transitions) of the Mealy machine.

This yields the following definition.

Definition 2.18. Let $i \in [\![1, n]\!]$. A *(stochastic) Mealy machine* of $\mathcal{P}_i$ is a tuple $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$, where M is a finite set of memory states, $\mu_{\mathsf{init}} \in \mathcal{D}(M)$ is an initial distribution, $\mathsf{nxt}_{\mathfrak{M}} \colon M \times S \to \mathcal{D}(A^{(i)})$ is a (stochastic) next-move function and $\mathsf{up}_{\mathfrak{M}} \colon M \times S \times A^{(i)} \to \mathcal{D}(M)$ is a (stochastic) update function.

Let $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be a Mealy machine of $\mathcal{P}_i$. If its initial distribution μ_{init} is a Dirac distribution for some $m_{\mathsf{init}} \in M$, we write $\mathfrak{M}$ as $(M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$. We say that $\mathfrak{M}$ has a *deterministic update* (resp. *next-move*) function if the image of $\mathsf{up}_{\mathfrak{M}}$ (resp. $\mathsf{nxt}_{\mathfrak{M}}$) only contains Dirac distributions. We assume that deterministic update (resp. next-move) functions are of the type $M \times S \times \bar{A} \to M$ (resp. $M \times S \to A^{(i)}$). A Mealy machine is *deterministic* if its initial distribution is a Dirac distribution and its update and next-move functions are deterministic.

We describe how $\mathfrak{M}$ works. Let $s_0 \in S$. At the start of a play, an initial memory state m_0 is selected randomly following μ_{init}. Then, at each step ℓ of the play, an action profile $\bar{a} \in \bar{A}(s_\ell)$ is sampled where $a^{(i)}_\ell$ of $\mathcal{P}_i$ is chosen following the distribution $\mathsf{nxt}_{\mathfrak{M}}(m_\ell, s_\ell)$ and the actions of the other players are independently chosen according to their respective strategies. The memory state $m_{\ell+1}$ is then randomly updated following the distribution $\mathsf{up}_{\mathfrak{M}}(m_\ell, s_\ell, \bar{a}_\ell)$ and the arena state $s_{\ell+1}$ is chosen following the distribution $\delta(s_\ell, \bar{a}_\ell)$, with these two choices being made independently.

We now explain how to derive a strategy from a Mealy machine. When in a memory state $m \in M$ and arena state $s \in S$, the probability of an action $a^{(i)} \in A^{(i)}(s)$ being chosen is given by $\mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)})$. Therefore, the probability of choosing the action $a^{(i)} \in A^{(i)}$ after some history $h = ws$ (where $w \in (S\bar{A})^*$ and $s = \mathsf{last}(h)$) is given by the sum, for each memory state $m \in M$, of the probability that m was reached after w has taken place (i.e., after $\mathfrak{M}$ processes w), multiplied by $\mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)})$. Therefore, to provide a formal definition of the strategy induced by $\mathfrak{M}$, we require a description of the distribution over memory states of $\mathfrak{M}$ after elements of $(S\bar{A})^*$ take place (under the strategy induced by $\mathfrak{M}$).

We provide an inductive definition of the distribution over memory states of $\mathfrak{M}$ after some element of $(S\bar{A})^*$ has taken place. This inductive formula can be derived by analysing the Markov chain obtained when fixing a Mealy machine of $\mathcal{P}_i$ and strategies of the other players. We defer the derivation of the inductive formula, which relies on conditional probabilities, to Appendix A.6.

The distribution μ_ε over memory states after the empty word ε (i.e., nothing) has taken place is by definition μ_{init}. Assume inductively that we know the distribution μ_w for $w = s_0\bar{a}_0 \dots s_{\ell-1}\bar{a}_{\ell-1}$. We explain how to derive $\mu_{ws_\ell\bar{a}_\ell}$ from μ_w for any state $s_\ell \in \mathsf{supp}(\delta(s_{\ell-1}, \bar{a}_{\ell-1}))$ and for any pair of actions $\bar{a}_\ell \in \bar{A}(s_\ell)$.

In general, the choice of an action by $\mathcal{P}_i$ conditions what the predecessor memory states could be. First, we note that if $\mathsf{nxt}_{\mathfrak{M}}(m', s_\ell)(a_\ell^{(i)}) = 0$ holds for all memory states $m' \in \mathsf{supp}(\mu_w)$, then the action $a_\ell^{(i)}$ is actually never chosen. We leave this case undefined (the related conditional probabilities are ill-defined) and assume that $a_\ell^{(i)} \in \mathsf{supp}(\mathsf{nxt}_{\mathfrak{M}}(m', s_\ell))$ for some $m' \in \mathsf{supp}(\mu_w)$. The equation for $\mu_{ws_\ell\bar{a}_\ell}$ uses the likelihood of being in a memory state knowing that the action $a_\ell^{(i)}$ was chosen, and not μ_w directly. We have, for any memory state $m \in M$,

$$\mu_{ws_\ell\bar{a}_\ell}(m) = \frac{\sum_{m' \in M} \mu_w(m') \cdot \mathsf{up}_{\mathfrak{M}}(m', s_\ell, \bar{a}_\ell)(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s_\ell)(a_\ell^{(i)})}{\sum_{m' \in M} \mu_w(m') \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s_\ell)(a_\ell^{(i)})}. \tag{2.1}$$

This quotient is not well-defined whenever $\mathsf{nxt}_{\mathfrak{M}}(m', s_\ell)(a_\ell^{(i)}) = 0$ holds for all $m' \in \mathsf{supp}(\mu_w)$, further justifying the distinction above.

Using these distributions, we formally define the (partial) strategy $\sigma_i^{\mathfrak{M}}$ induced by the Mealy machine $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ as the strategy $\sigma_i^{\mathfrak{M}} \colon \mathsf{Hist}(\mathcal{A}) \to \mathcal{D}(A^{(i)})$ such that for all histories $h = ws$, for all actions $a^{(i)} \in A^{(i)}(s)$,

$$\sigma_i^{\mathfrak{M}}(h)(a^{(i)}) = \sum_{m \in M} \mu_w(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)}).$$

This strategy is only partially defined because distributions μ_w are not defined for all $w \in (S\bar{A})^*$. Due to the inductive definition of μ_w, all histories for which $\sigma_i^{\mathfrak{M}}$ is undefined are of the form $h\bar{a}h'$ such that $\sigma_i^{\mathfrak{M}}$ is defined for h and $\sigma_i^{\mathfrak{M}}(h)(a^{(i)}) = 0$. In other words, $\sigma_i^{\mathfrak{M}}$ is only undefined over histories with a prefix that is inconsistent with $\sigma_i^{\mathfrak{M}}$. Therefore, no matter how the partial definition of $\sigma_i^{\mathfrak{M}}$ given above is extended, it does not influence the induced probability distribution over plays involving this strategy. We define finite-memory strategies as strategies whose behaviour can be induced by using a Mealy machine.

Definition 2.19. A strategy σ_i of $\mathcal{P}_i$ is a *finite-memory strategy* if there exists a Mealy machine $\mathfrak{M}$ of $\mathcal{P}_i$ such that σ_i agrees with $\sigma_i^{\mathfrak{M}}$ over the domain of latter.

We say that a strategy profile is a *finite-memory strategy profile* if all strategies within are finite-memory.

Let $\mathfrak{M}$ be a Mealy machine of $\mathcal{P}_i$ with a deterministic initialisation and deterministic updates. In this case, the distribution over memory states of $\mathfrak{M}$ after a history prefix takes place is a Dirac distribution. This state can be determined by iterating memory updates from the initial memory state.

Definition 2.20. Let $\mathfrak{M} = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be a Mealy machine of $\mathcal{P}_i$ with a deterministic initialisation and a deterministic update function $\mathsf{up}_{\mathfrak{M}} \colon M \times S \times \bar{A} \to M$. We define the *iterated memory update function* $\widehat{\mathsf{up}_{\mathfrak{M}}} \colon (S\bar{A})^* \to M$ by induction, by letting, $\widehat{\mathsf{up}_{\mathfrak{M}}}(\varepsilon) = m_{\mathsf{init}}$ and, for any $ws\bar{a} \in (S\bar{A})^+$, $\widehat{\mathsf{up}_{\mathfrak{M}}}(ws\bar{a}) = \mathsf{up}_{\mathfrak{M}}(\widehat{\mathsf{up}_{\mathfrak{M}}}(w), s, \bar{a})$.

2.5 Specifications

2.5.1 Objectives and payoffs

An arena describes the interaction of the players without specifying their goals. These goals can be modelled in several ways depending on the considered specification. We consider two ways of formalising the goals of players. First, the goal of a player can be specified through a set of good plays, which we call an objective.

Definition 2.21. An *objective* is a measurable set of plays.

We say that a play $\pi \in \mathsf{Plays}(\mathcal{A})$ satisfies an objective $\Omega \subseteq \mathsf{Plays}(\mathcal{A})$ if $\pi \in \Omega$. Intuitively, the goal of a player is to have their objective be satisfied with high probability.

Second, we consider specifications modelled by assigning a numerical value to plays.

Definition 2.22. A *payoff function* (or payoff for short) is a measurable function $f\colon \mathsf{Plays}(\mathcal{M}) \to \bar{\mathbb{R}}$.

In general, players strive to obtain a high expected payoff. When a payoff function models a cost to be minimised, we highlight this by referring to the function as a *cost function*.

Remark 2.23. A payoff function can be used to model the specification given by an objective: in a probability space, the expectation of the indicator of an event is the probability of the event. Therefore, to an objective Ω, we associate the payoff $\mathbb{1}_\Omega$. Similarly, to an objective Ω, we associate the cost function $\mathbb{1}_{\mathsf{Plays}(\mathcal{A})\setminus\Omega}$. It follows that all definitions for payoffs and cost functions directly extend to objectives. ◁

We define a game as an arena along with payoffs (or costs) for each player.

Definition 2.24. A *game* (over $\mathcal{A}$) is a pair $\mathcal{G} = (\mathcal{A}, (f_i)_{i\in[\![1,n]\!]})$ where f_i is the payoff (or cost) function of $\mathcal{P}_i$ for all $i \in [\![1, n]\!]$.

Given a game $\mathcal{G} = (\mathcal{A}, (\mathbb{1}_{\Omega_i})_{i\in[\![1,n]\!]})$ where the payoffs are indicators of objectives $\Omega_1, \ldots, \Omega_n$, we abuse notation and write $\mathcal{G} = (\mathcal{A}, (\Omega_i)_{i\in[\![1,n]\!]})$ instead.

2.5.2 Expected payoffs

Let $f\colon \mathsf{Plays}(\mathcal{A}) \to \bar{\mathbb{R}}$ be a payoff function. Let σ be a strategy profile and $s \in S$ be a state of $\mathcal{A}$. The $\mathbb{P}^\sigma_s$-integral of f is only formally defined whenever f is non-negative, non-positive or $\mathbb{P}^\sigma_s$-integrable. If f is such a payoff, we let $\mathbb{E}^\sigma_s(f) = \int_{\pi\in\mathsf{Plays}(\mathcal{M})} f(\pi)\mathrm{d}\mathbb{P}^\sigma_s(\pi)$; $\mathbb{E}^\sigma_s(f)$ is the expected payoff of the strategy profile σ from s (for f). We also generalise the notion of expected payoff to a broader class of payoff functions as follows.

Definition 2.25. Let $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$ denote the non-negative and non-positive parts of f. We say that f has an *unambiguous* $\mathbb{P}_s^\sigma$*-integral* if $\mathbb{E}_s^\sigma(f^+) \in \mathbb{R}$ or $\mathbb{E}_s^\sigma(f^-) \in \mathbb{R}$. If f has an unambiguous $\mathbb{P}_s^\sigma$-integral, we abuse notation and let $\mathbb{E}_s^\sigma(f) = \mathbb{E}_s^\sigma(f^+) - \mathbb{E}_s^\sigma(f^-)$.

We reserve the notation $\mathbb{E}$ for the expectation of payoffs, and use integrals for other probability spaces.

In Part IV, we study MDPs with multiple payoff functions (for the unique player) and provide general results for this setting. In this context, we limit ourselves to payoffs for which the expected payoff is unambiguously defined under all strategies from all initial states.

Definition 2.26. The payoff f is *universally unambiguously integrable* if f has an unambiguous $\mathbb{P}_s^\sigma$-integral for all strategy profiles σ and all $s \in S$.

In particular, all non-negative and non-positive payoffs are universally unambiguously integrable. We will see that payoffs that are integrable no matter the strategy and initial state are of particular interest.

Definition 2.27. The payoff f is *universally integrable* if it is $\mathbb{P}_s^\sigma$-integrable, i.e., if $\mathbb{E}_s^\sigma(|f|) \in \mathbb{R}$, for all strategy profiles σ and all $s \in S$.

All bounded functions are universally integrable. In particular, the indicator of any objective falls into this category.

2.5.3 Continuous payoffs

Let $f\colon \mathsf{Plays}(\mathcal{A}) \to \bar{\mathbb{R}}$ be a payoff. We say that f is *continuous* at a play π to mean that it is continuous at π with respect to the usual topologies of $\mathsf{Plays}(\mathcal{A})$ and $\bar{\mathbb{R}}$. Since all open subsets of $\mathsf{Plays}(\mathcal{A})$ are unions of history cylinders, continuity at a play can be characterised as follows.

Definition 2.28. Let $f\colon \mathsf{Plays}(\mathcal{M}) \to \bar{\mathbb{R}}$ be a payoff and let $\pi \in \mathsf{Plays}(\mathcal{M})$.

- If $f(\pi) \in \mathbb{R}$, then f is continuous at π if and only if for all $\varepsilon > 0$, there exists $\ell \in \mathbb{N}$ such that for all $\pi' \in \mathsf{Cyl}\,(\pi_{\leq \ell})$, $|f(\pi) - f(\pi')| < \varepsilon$.

- If $f(\pi) = +\infty$ (resp. $-\infty$), then f is continuous at π if and only if for all $M \in \mathbb{R}$, there exists $\ell \in \mathbb{N}$ such that for all plays $\pi' \in \mathsf{Cyl}\,(\pi_{\leq \ell})$, $f(\pi') \geq M$ (resp. $f(\pi') \leq -M$).

The payoff f is *continuous* if it is continuous at all plays.

When $\mathcal{A}$ is finite, $\mathsf{Plays}(\mathcal{A})$ is compact. Therefore, continuous real-valued payoffs over finite arenas are *uniformly continuous*, which is a stronger form of continuity. In general, uniformly continuous payoffs can be defined as follows.

Definition 2.29. Let $f\colon \mathsf{Plays}(\mathcal{A}) \to \mathbb{R}$ be a real-valued payoff. The payoff f is *uniformly continuous* if and only if for all $\varepsilon > 0$, there exists $\ell \in \mathbb{N}$ such that for all plays π, $\pi' \in \mathsf{Plays}(\mathcal{M})$, $\pi_{\leq \ell} = \pi'_{\leq \ell}$ implies that $|f(\pi) - f(\pi')| < \varepsilon$.

We provide some examples of continuous payoffs in Appendix A.9.

2.5.4 Some classical objectives

We now present some classical objectives. A core objective in verification and synthesis is the reachability objective. A reachability objective requires that a set of target states be visited.

Definition 2.30. Let $T \subseteq S$ be a target (i.e., a set of target states). The *reachability objective* $\mathsf{Reach}(T)$ is the set $\{s_0\bar{a}_0 s_1 \bar{a}_1 \ldots \in \mathsf{Plays}(\mathcal{A}) \mid \exists \ell \in \mathbb{N},\, s_\ell \in T\}$. If $T = \{t\}$, we write $\mathsf{Reach}(t)$ instead of $\mathsf{Reach}(\{t\})$.

For instance, in the model of rock paper scissors of Example 2.1, the goal of winning for $\mathcal{P}_1$ can be modelled by the reachability objective $\mathsf{Reach}(\mathsf{win})$.

The complement of a reachability objective is called a safety objective: it requires that a set of unsafe states never be visited.

Definition 2.31. Let $U \subseteq S$ be a set of unsafe states. The *safety objective* $\mathsf{Safe}(U)$ is the set $\mathsf{Plays}(\mathcal{A}) \setminus \mathsf{Reach}(U)$. If $U = \{t\}$, we write $\mathsf{Safe}(t)$ instead of $\mathsf{Safe}(\{t\})$.

The reachability objective requires visiting a target once. Büchi objectives model the requirement of visiting a target infinitely often.

Definition 2.32. Let $T \subseteq S$ be a target. The *Büchi objective* $\mathsf{Büchi}(T)$ is the set $\{s_0\bar{a}_0s_1\bar{a}_1\ldots \in \mathsf{Plays}(\mathcal{A}) \mid \forall\, \ell \in \mathbb{N}, \exists\, r \geq \ell,\, s_r \in T\}$. If $T = \{t\}$, we write $\mathsf{Büchi}(t)$ instead of $\mathsf{Büchi}(\{t\})$.

A Büchi objective can be used, e.g., to model the requirement of having to win infinitely often in a variant of the rock paper scissors arena of Example 2.1 in which players replay after each round even when there is not a draw.

Finally, the complement of a Büchi objective is the co-Büchi objective. It requires avoiding a set of unsafe states from some point on.

Definition 2.33. Let $U \subseteq S$ be a set of unsafe states. The *co-Büchi objective* $\mathsf{coBüchi}(U)$ is the set $\mathsf{Plays}(\mathcal{A}) \setminus \mathsf{Büchi}(U)$. If $U = \{t\}$, we write $\mathsf{coBüchi}(t)$ instead of $\mathsf{coBüchi}(\{t\})$.

2.5.5 Some classical payoff functions

We consider payoff functions that are defined from numerical weights that are assigned to transitions of $\mathcal{A}$. Formally, a weight function is a function $w\colon S \times \bar{A} \to \mathbb{R}$. We fix a weight function w.

A *discounted-sum payoff* is defined as an accumulated sum of weights multiplied by powers of a discount factor in $[0, 1[$.

Definition 2.34 (Discounted-sum payoff)**.** Let $\lambda \in [0, 1[$ be a discount factor. We let $\mathsf{DSum}_w^\lambda\colon \mathsf{Plays}(\mathcal{A}) \to \mathbb{R}$ be the payoff function defined by $\mathsf{DSum}_w^\lambda(\pi) = \sum_{\ell=0}^{\infty} \lambda^\ell w(s_\ell, \bar{a}_\ell)$ for all plays $\pi = s_0\bar{a}_0s_1\bar{a}_1\ldots \in \mathsf{Plays}(\mathcal{A})$.

When w is bounded in absolute value (this holds by default in finite arenas), any discounted-sum payoff built from w is well-defined and bounded. This implies that discounted-sum payoffs are universally integrable whenever w is bounded. We can also show that discounted-sum are uniformly continuous when w is bounded (see Lemma A.12).

A *total-reward* payoff corresponds to the accumulated weights along a play with no discounting.

Definition 2.35 (Total reward payoff)**.** We let $\mathsf{TRew}_w\colon \mathsf{Plays}(\mathcal{A}) \to \bar{\mathbb{R}}$ be the payoff function defined by $\liminf_{r\to\infty} \sum_{\ell=0}^{r} w(s_\ell, \bar{a}_\ell)$ for all plays $\pi = s_0\bar{a}_0s_1\bar{a}_1\ldots \in \mathsf{Plays}(\mathcal{A})$.

We use a limit-inferior to ensure that this payoff is well-defined for all plays, as the series of weights along a play need not converge.

A *shortest-path* cost function can be seen as a quantitative variant of reachability and is defined as the accumulated sum of weights up to the first visit of the target set. We refer to this function as a cost function, as it is often used to model the goal of minimising the time or cost to reach a target.

Definition 2.36 (Shortest-path cost)**.** Let $T \subseteq S$ be a target. We let $\mathsf{SPath}_w^T\colon \mathsf{Plays}(\mathcal{A}) \to \bar{\mathbb{R}}$ be the payoff function defined by, for all plays $\pi = s_0\bar{a}_0s_1\bar{a}_1\ldots$, $\mathsf{SPath}_w^T(\pi) = +\infty$ if $\pi \notin \mathsf{Reach}(T)$ and, otherwise, $\mathsf{SPath}_w^T(\pi) = \sum_{\ell=0}^{r-1} w(s_\ell, \bar{a}_\ell)$ where $r = \min\{\ell \in \mathbb{N} \mid s_\ell \in T\}$.

We attribute an infinite cost to plays that do not visit the target to give them the largest possible penalty. A shortest-path cost function is continuous whenever there exists a positive lower bound on weights (see Lemma A.13).

Remark 2.37. If $\mathcal{A}$ is turn-based, we consider weight functions to be of the form $w\colon S \times A \to \mathbb{R}$ (where A denotes the set of all actions of all players). The payoffs defined above can be directly adapted to accommodate this change. $\lhd$

2.6 Solution concepts

We consider two types of games: *two-player zero-sum games* and *multi-player games*. In a two-player zero-sum game (Section 2.6.1), the two players compete for opposite goals. In multi-player games (Section 2.6.2), each player has their own goal they aim to optimise, and their respective goals need not be opposite to one another.

2.6.1 Zero-sum games

Assume that $n = 2$. Intuitively, a game on $\mathcal{A}$ is zero-sum if the goal of $\mathcal{P}_1$ is to maximise a payoff and the goal of $\mathcal{P}_2$ is to minimise the same payoff. We formalise this as follows.

Definition 2.38. A two-player game $\mathcal{G} = (\mathcal{A}, (f_1, f_2))$ is a *zero-sum game* if $f_2 = -f_1$.

We drop the payoff of $\mathcal{P}_2$ from the notation of zero-sum games, i.e., we write $\mathcal{G} = (\mathcal{A}, f)$ instead of $(\mathcal{A}, (f, -f))$. We fix a (universally unambiguously integrable) payoff f and the zero-sum game $\mathcal{G} = (\mathcal{A}, f)$ for the following definitions.

We present definitions with a maximisation point of view, i.e., we assume that the goal of $\mathcal{P}_1$ is to maximise the expectation of f. If the goal of $\mathcal{P}_1$ is to minimise the expectation of f, i.e., if f is a cost function, it suffices to exchange the roles of the two players in the following.

Definition 2.39 (Value)**.** Let $s_{\mathsf{init}} \in S$ be an initial state. If

$$\sup_{\sigma_1 \in \Sigma^1(\mathcal{A})} \inf_{\sigma_2 \in \Sigma^2(\mathcal{A})} \mathbb{E}^{\sigma_1, \sigma_2}_{s_{\mathsf{init}}}(f) = \inf_{\sigma_2 \in \Sigma^2(\mathcal{A})} \sup_{\sigma_1 \in \Sigma^1(\mathcal{A})} \mathbb{E}^{\sigma_1, \sigma_2}_{s_{\mathsf{init}}}(f), \tag{2.2}$$

we refer to the above as the *value* of s_{init} in $\mathcal{G}$ and denote it by $\mathsf{Val}_{\mathcal{G}}(s_{\mathsf{init}})$. A game is *determined* if the value is well-defined in all states.

A strategy σ_1 of $\mathcal{P}_1$ *ensures* $\theta \in \bar{\mathbb{R}}$ from s_{init} if for all strategies σ_2 of $\mathcal{P}_2$, $\mathbb{E}^{\sigma_1, \sigma_2}_{s_{\mathsf{init}}}(f) \geq \theta$. Symmetrically, a strategy σ_2 of $\mathcal{P}_2$ *ensures* $\theta \in \bar{\mathbb{R}}$ from s_{init} if for all strategies σ_1 of $\mathcal{P}_1$, $\mathbb{E}^{\sigma_1, \sigma_2}_{s_{\mathsf{init}}}(f) \leq \theta$. A strategy σ_i of $\mathcal{P}_i$ is *optimal* from s_{init} if it ensures $\mathsf{Val}_{\mathcal{G}}(s_{\mathsf{init}})$ from s_{init}, and σ_i is *uniformly optimal* if it ensures $\mathsf{Val}_{\mathcal{G}}(s)$ from s for all $s \in S$.

Example 2.3 (Optimal strategies in rock paper scissors)**.** In rock paper scissors, if one player plays uniformly at random, they can ensure a probability of $\frac{1}{2}$ of winning. We show this by reasoning on the arena $\mathcal{A}$ modelling rock paper scissors presented in Example 2.1, and the zero-sum reachability game $\mathcal{G} = (\mathcal{A}, \mathsf{Reach}(\mathsf{win}))$. In $\mathcal{G}$, the goal of $\mathcal{P}_1$ is to maximise their probability of

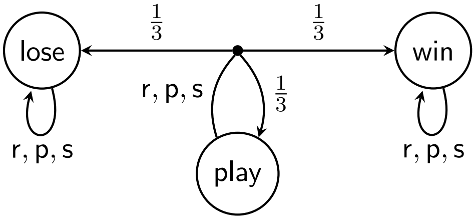


Figure 2.4: The MDP obtained from the rock paper scissors arena of Example 2.1 by fixing a uniform randomised strategy for one of the players.

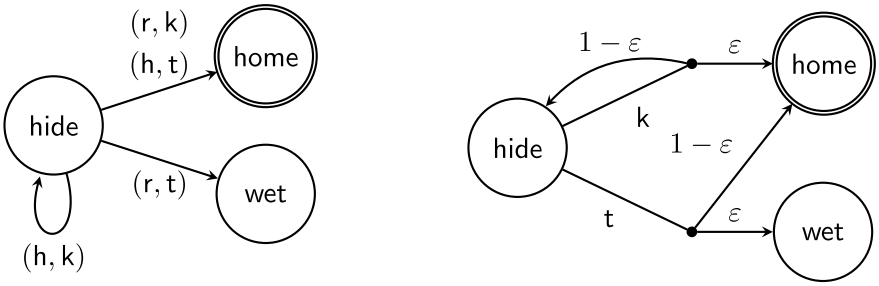
winning, whereas the goal of $\mathcal{P}_2$ is to prevent $\mathcal{P}_1$ from winning. Let σ_1 and σ_2 be the memoryless strategies of $\mathcal{A}$ that select an action uniformly at random in state play.

We show that these strategies ensure $\frac{1}{2}$ for both players. Fix $i \in \{1, 2\}$. Figure 2.4 illustrates the MDP induced on $\mathcal{A}$ when fixing the strategy of $\mathcal{P}_i$ to be σ_i. It can be constructed by noting that, regardless of the choice of $\mathcal{P}_{3-i}$ (i.e., the other player) in play, $\mathcal{P}_{3-i}$ has a uniform probability of winning, losing or having a draw against σ_i. This MDP can be seen as a Markov chain: the choices of its player do not influence the probability of reachability objectives. Therefore, no matter the chosen strategy in this MDP, states win and lose will be reached with probability $\frac{1}{2}$. Since these two states are unreachable from one another, it follows that σ_1 and σ_2 ensure $\frac{1}{2}$ from play in $\mathcal{G}$. Furthermore, because both players have strategies ensuring the same threshold, we obtain that $\mathsf{Val}_{\mathcal{G}}(\mathsf{play}) = \frac{1}{2}$ and that σ_1 and σ_2 are optimal from play. $\lhd$

If $f = \mathbb{1}_\Omega$ for some objective Ω, we use specialised terminology. First, a strategy σ_1 of $\mathcal{P}_1$ is *(surely) winning* from s_{init} if all outcomes π of σ_1, π satisfies Ω. Symmetrically, a strategy σ_2 of $\mathcal{P}_2$ is *winning* from s_{init} if all of its outcomes π do not satisfy Ω. Analogously to uniformly optimal strategies, for $i \in \{1, 2\}$, we define *uniformly winning* strategies of $\mathcal{P}_i$ as strategies that are winning from each state from which $\mathcal{P}_i$ has a winning strategy.

A strategy σ_1 of $\mathcal{P}_1$ is *almost-surely winning* from s_{init} if, for all strategies σ_2 of $\mathcal{P}_2$, $\mathbb{P}^{\sigma_1,\sigma_2}_{s_{\mathsf{init}}}(\Omega) = 1$, i.e., if σ_1 ensures 1 for the payoff $\mathbb{1}_\Omega$. A strategy σ_1 of $\mathcal{P}_1$ is *positively winning* from s_{init} if, for all strategies σ_2 of $\mathcal{P}_2$, $\mathbb{P}^{\sigma_1,\sigma_2}_{s_{\mathsf{init}}}(\Omega) > 0$, i.e., if Ω is satisfied with positive probability no matter the strategy of $\mathcal{P}_2$.

The value of a state in $\mathcal{G}$, intuitively, represents the greatest expected payoff



(a) The arena of the snowball game of [dAHK07]. State home is the target of $\mathcal{P}_1$. We omit self-loops on states home and wet to lighten the figure.

(b) The MDP induced on the snowball game by having $\mathcal{P}_1$ select action r with probability $\varepsilon \in [0, 1]$ regardless of the history.

Figure 2.5: A concurrent reachability game. The actions r, h, t and k respectively represent the actions run, hide, throw and keep.

that $\mathcal{P}_1$ can ensure and the lowest expected payoff that $\mathcal{P}_2$ can ensure. Let $s_{\mathsf{init}} \in S$. If $\mathsf{Val}_{\mathcal{G}}(s_{\mathsf{init}}) \in \mathbb{R}$, due to the supremum in the definition of the value, $\mathcal{P}_1$ has strategies that can ensure $\mathsf{Val}_{\mathcal{G}}(s) - \varepsilon$ for all $\varepsilon > 0$, i.e., $\mathcal{P}_1$ can ensure thresholds arbitrarily close to the value. If $\mathsf{Val}_{\mathcal{G}}(s_{\mathsf{init}}) = +\infty$, then $\mathcal{P}_1$ has strategies ensuring M for all $M \in \mathbb{N}$. However, optimal strategies need not necessarily exist, even if the value does.

Example 2.4 (Non-existence of optimal strategies)**.** We present the snowball game [dAHK07], a concurrent reachability game in which $\mathcal{P}_1$ has no optimal strategy from a given state in spite of the value of this state being 1.

In the snowball game, $\mathcal{P}_1$ wants to return home without being hit by $\mathcal{P}_2$ who has a single snowball. The arena $\mathcal{A}$ depicted in Figure 2.5a models the interaction of the two players. At each step of a play, $\mathcal{P}_1$ can either remain in hiding or run back home, whereas $\mathcal{P}_2$ can either keep their single snowball or throw it. If $\mathcal{P}_1$ runs when $\mathcal{P}_2$ throws their snowball, then $\mathcal{P}_1$ loses. If $\mathcal{P}_1$ remains hidden and $\mathcal{P}_2$ keeps their snowball, then the play continues for an additional step. In any other case, $\mathcal{P}_1$ reaches home without being hit by a snowball. The objective of $\mathcal{P}_1$ is a reachability objective: we consider the zero-sum reachability game $\mathcal{G} = (\mathcal{A}, \mathsf{Reach}(\mathsf{home}))$.

First, we claim that $\mathsf{Val}_{\mathcal{G}}(\mathsf{hide}) = 1$. Let $\varepsilon \in]0, 1[$. Let σ_1^{ε} be the memoryless strategy of $\mathcal{P}_1$ such that $\sigma_1^\varepsilon(\mathsf{hide})(\mathsf{r}) = \varepsilon$. We use the MDP of Figure 2.5b obtained by fixing σ_1 in $\mathcal{A}$ to conclude that the strategy σ_1^ε ensures $1 - \varepsilon$ from hide. In MDPs, there exist optimal strategies for safety objectives that are pure and memoryless (e.g., [BK08]), and thus the smallest probability of visiting home in the MDP of Figure 2.5b is $1 - \varepsilon$. Since $\mathcal{P}_1$ can ensure $1 - \varepsilon$ from hide for all $\varepsilon \in]0, 1[$, it follows that $\mathsf{Val}_{\mathcal{G}}(\mathsf{hide}) = 1$.

However, $\mathcal{P}_1$ does not have an optimal strategy. It suffices to check for a memoryless optimal strategy: if there exists an almost-surely winning strategy from a state in a zero-sum concurrent reachability game, then there exists a memoryless almost-surely winning strategy from this state [dAHK07]. However, by examining the MDP of Figure 2.5b, we can see that all choices of ε allow $\mathcal{P}_2$ to prevent a visit to home with positive probability. ◁

2.6.2 Multi-player games

We lift the assumption that $n = 2$ of the previous section and fix (universally unambiguously integrable) payoffs f_1, …, f_n and the game $\mathcal{G} = (\mathcal{A}, (f_1, \dots, f_n))$.

In a two-player zero-sum game, the two players compete with one another for opposite objective. Due to this, we perform a worst-case analysis; this is reflected, e.g., in the definition of the value (Definition 2.39). In a non-zero-sum context, the different players each have their own goals that do not necessarily conflict with one another. We thus use a different solution concept for such games.

We consider *Nash equilibria* (NEs). Intuitively, an NE from an initial state is a strategy profile that can be seen as a contract between the players such that none of the players have an incentive to unilaterally deviate from the agreement. We define NEs for cost functions. We use this definition as in Part II, we mainly study NEs in games with shortest-path cost functions and with reachability and Büchi objectives.

Definition 2.40. Let $s_{\mathsf{init}} \in S$. Assume that, for all $i \in [\![1, n]\!]$, f_i is a cost function (i.e., $\mathcal{P}_i$ wants to minimise it). A *Nash equilibrium* (NE) from s_{init} in $\mathcal{G}$ is a strategy profile $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$ such that, for all $i \in [\![1, n]\!]$ and all strategies

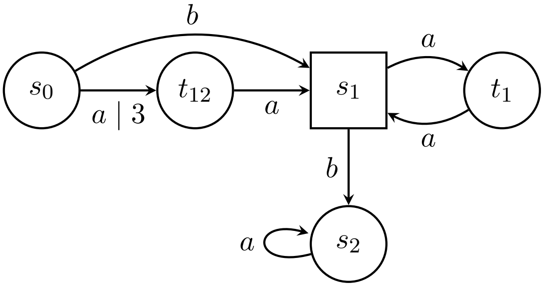


Figure 2.6: A turn-based arena. Circles and squares respectively denote $\mathcal{P}_1$ and $\mathcal{P}_2$ states. Unspecified weights are 1 and are omitted to lighten the figure.

$\tau_i \in \Sigma^i(\mathcal{A})$,

$$\mathbb{E}^{\sigma}_{s_{\mathsf{init}}}(f_i) \leq \mathbb{E}^{\tau_i,\sigma_{-i}}_{s_{\mathsf{init}}}(f_i).$$

Let $s_{\mathsf{init}} \in S$. Given a strategy profile $\sigma = (\sigma_i)_{i\in[\![1,n]\!]}$, $i \in [\![1,n]\!]$ and a strategy $\tau_i \in \Sigma^i(\mathcal{A})$, we say that τ_i is a *profitable deviation* (with respect to σ from s_{init}) if $\mathbb{E}^{\sigma}_{s_{\mathsf{init}}}(f_i) > \mathbb{E}^{\tau_i,\sigma_{-i}}_{s_{\mathsf{init}}}(f_i)$. Therefore, a strategy profile is an NE if and only if no player has a profitable deviation.

The *(expected) cost profile* of an NE σ from $s_{\mathsf{init}} \in S$ is $(\mathbb{E}^{\sigma}_{s_{\mathsf{init}}}(f_i))_{i\in[\![1,n]\!]}$. In general, there may exist several NEs in a game with incomparable cost profiles with respect to the component-wise ordering on $\bar{\mathbb{R}}^n$.

Example 2.5. We consider the turn-based deterministic arena $\mathcal{A}$ depicted in Figure 2.6. We let w denote the weight function that is equal to 1 for all state-action pairs other than (s_0, a) and such that $w(s_0, a) = 3$ (as indicated in the figure). We let $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}^{T_1}_w, \mathsf{SPath}^{T_2}_w))$ where $T_1 = \{t_{12}, t_1\}$ and $T_2 = \{t_{12}\}$.

The memoryless strategy profile (σ_1, σ_2) with $\sigma_1(s_0) = a$ and $\sigma_2(s_1) = b$ is an NE from s_0 with cost profile $(3, 3)$. On the one hand, if $\mathcal{P}_1$ assigns positive probability to action b in s_0, then T_1 would not be visited with positive probability by definition of σ_2. It follows that their expected payoff, if they deviate from σ, would be infinite, and thus $\mathcal{P}_1$ does not have a profitable deviation. On the other hand, $\mathcal{P}_2$ cannot improve their cost by deviating because their target T_2 is visited before s_1 is reached.

Another pure NE from s_0 is the memoryless strategy profile (σ'_1, σ'_2) such that $\sigma'_1(s_0) = s_1$ and $\sigma'_2(s_1) = t_1$. The cost profile of this NE is $(2, +\infty)$, which is incomparable with $(3, 3)$. $\lhd$

When studying pure Nash equilibria in games on deterministic arenas, it is sufficient to only consider pure deviations when checking the existence of a profitable deviation. Fix an initial state s_{init}. Intuitively, if $\mathcal{P}_i$ has a (randomised) profitable deviation τ_i with respect to a pure strategy profile $\sigma = (\sigma_i, \sigma_{-i})$ from s_{init}, then the set of plays with a smaller cost for $\mathcal{P}_i$ than $\mathsf{Out}_{\mathcal{A}}(\sigma, s_{\mathsf{init}})$ has positive probability under (τ_i, σ_{-i}) from s_{init}. Any pure strategy that follows along a play of this set is a profitable deviation of $\mathcal{P}_i$. The above idea is formalised in the proof presented in Appendix A.7 of the following result.

Lemma 2.41. *Assume that $\mathcal{A}$ is deterministic and that, for all $i \in [\![1, n]\!]$, f_i is a cost function. Let $s_{\mathsf{init}} \in S$ and $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$ be a pure strategy profile. Let $i \in [\![1, n]\!]$ and write $\sigma = (\sigma_i, \sigma_{-i})$. The following statements are equivalent:*

(i) $\mathcal{P}_i$ has a profitable deviation with respect to σ from s_{init};

(ii) there exists a play π from s_{init} consistent with σ_{-i} such that $f_i(\pi) < f_i(\mathsf{Out}_{\mathcal{A}}(\sigma, s_{\mathsf{init}}))$.

(iii) $\mathcal{P}_i$ has a pure profitable deviation with respect to σ from s_{init};

In particular, σ is an NE from s_{init} if and only if no player has a pure profitable deviation.

2.7 Imperfect information

Up to now, we have considered arenas with perfect information, i.e., in which players are fully informed throughout the play. We now introduce arenas with imperfect information, in which players perceive observations rather than the states and actions of the play directly. Arenas with perfect information are a special case of such arenas, in which observations are exactly the states and actions.

2.7.1 Definition

In the imperfect information setting, the players are not fully informed of the current state of the play and the actions that are used along the play. Instead, they perceive an *observation* for each state and action, and this observation may be shared between different states and actions, making them indistinguishable. These observations are not shared between the players; each player perceives the ongoing play differently. We formalise this model as follows.

Definition 2.42. Let $n \in \mathbb{N}_{>0}$. An *n-player arena with imperfect information* is defined as a tuple $\mathfrak{P} = (\mathcal{A}, (\mathcal{Z}_i, \mathsf{Obs}_i)_{i \in [\![1,n]\!]})$ where $\mathcal{A} = (S, (A^{(i)})_{i \in [\![1,n]\!]}, \delta)$ is an n-player arena (with perfect information), and for $i \in [\![1,n]\!]$, $\mathcal{Z}_i$ is a countable set of observations of $\mathcal{P}_i$ and $\mathsf{Obs}_i \colon S \cup \bigcup_{i' \in [\![1,n]\!]} A^{(i')} \to \mathcal{Z}_i$ is the observation function of $\mathcal{P}_i$. We require that for all $i \in [\![1,n]\!]$ and all $s, s' \in S$, $\mathsf{Obs}_i(s) = \mathsf{Obs}_i(s')$ implies $A^{(i)}(s) = A^{(i)}(s')$, i.e., in two states that are indistinguishable for $\mathcal{P}_i$, the same actions are available to $\mathcal{P}_i$.

A one-player arena with imperfect observation is called a *partially observable Markov decision process* (POMDP). We fix $\mathfrak{P} = (\mathcal{A}, (\mathcal{Z}_i, \mathsf{Obs}_i)_{i \in [\![1,n]\!]})$ where $\mathcal{A} = (S, (A^{(i)})_{i \in [\![1,n]\!]}, \delta)$ for the remainder of the section. We say that $\mathfrak{P}$ is finite if $\mathcal{A}$ is finite.

Plays and histories of $\mathfrak{P}$ are respectively defined as plays and histories of $\mathcal{A}$. We reuse the notations $\mathsf{Plays}(\mathfrak{P})$ and $\mathsf{Hist}(\mathfrak{P})$ for the sets of plays of $\mathfrak{P}$ and histories of $\mathfrak{P}$ respectively. We extend the observation functions to action profiles and to histories as follows. Let $i \in [\![1,n]\!]$. For all $\bar{a} = (a^{(i')})_{i' \in [\![1,n]\!]} \in \bar{A}$, we let $\mathsf{Obs}_i(\bar{a}) = (\mathsf{Obs}_i(a^{(i')}))_{i' \in [\![1,n]\!]}$. For all histories $h = s_0\bar{a}_0 \dots s_r$ of $\mathfrak{P}$, we let $\mathsf{Obs}_i(h) = \mathsf{Obs}_i(s_0)\mathsf{Obs}_i(\bar{a}_0) \dots \mathsf{Obs}_i(s_r)$. We say that two histories h, $h' \in \mathsf{Hist}(\mathfrak{P})$ are *indistinguishable* for $\mathcal{P}_i$ if $\mathsf{Obs}_i(h) = \mathsf{Obs}_i(h')$.

In $\mathfrak{P}$, players select actions based on the sequence of observations they have perceived up to the point of decision. Formally, we define strategies of $\mathfrak{P}$ as strategies of $\mathcal{A}$ that agree on histories that share the same observation. This definition avoids having to redefine notions such as consistency and distributions induced by plays.

Definition 2.43. Let $i \in [\![1, n]\!]$. A pure (resp. behavioural) strategy σ_i of $\mathcal{P}_i$ in $\mathcal{A}$ is a *pure (resp. behavioural) observation-based strategy* in $\mathfrak{P}$ if for all histories $h, h' \in \mathsf{Hist}(\mathfrak{P})$, $\mathsf{Obs}_i(h) = \mathsf{Obs}_i(h')$ implies that $\sigma_i(h) = \sigma_i(h')$. A *mixed observation-based strategy* of $\mathcal{P}_i$ in $\mathfrak{P}$ is a mixed strategy of $\mathcal{P}_i$ in $\mathcal{A}$ that assigns a probability of zero to the set of pure strategies that are not observation-based.

Remark 2.44 (Measurability of the set of observation-based strategies)*.* The definition of a mixed observation-based strategy assumes that the set of strategies that are observation-based is in the σ-algebra we consider on the set of pure strategies of a player. It suffices to show that the complement of this set is measurable.

Let $i \in [\![1, n]\!]$. A pure strategy σ_i of $\mathcal{P}_i$ is not observation-based if there are two histories h and h' that are indistinguishable for $\mathcal{P}_i$ such that $\sigma_i(h) = a^{(i)}$ and $\sigma_i(h') = b^{(i)}$ for distinct $a^{(i)}$, $b^{(i)} \in A^{(i)}(\mathsf{last}(h))$. For any pair of indistinguishable histories (h, h') and pair of distinct actions $(a^{(i)}, b^{(i)}) \in A^{(i)}(\mathsf{last}(h)) \times A^{(i)}(\mathsf{last}(h))$, the set of strategies that assign $a^{(i)}$ to h and $b^{(i)}$ to h' is one of the sets we have used to generate our σ-algebra over $\Sigma^i_{\mathsf{pure}}(\mathcal{A})$). Since the set of non-observation-based strategies can be written as the countable union of these sets, we obtain that the set of observation-based strategies and its complement are both measurable. $\lhd$

Pure and behavioural observation-based strategies of $\mathcal{P}_i$ in $\mathfrak{P}$ can be seen as a functions $\mathsf{Obs}_i(\mathsf{Hist}(\mathfrak{P})) \to A^{(i)}$ and $\mathsf{Obs}_i(\mathsf{Hist}(\mathfrak{P})) \to \mathcal{D}(A^{(i)})$ respectively. We refer to (behavioural) strategies of the arena $\mathcal{A}$ with perfect information as *history-based strategies* to distinguish them from observation-based strategies.

2.7.2 Perfect recall and Kuhn's theorem

We now define perfect recall. Perfect recall in games in extensive form with imperfect information is defined by two properties: players never forget their previous knowledge and can infer their previous action choices from the information at their disposal. We refer the reader to [OR94, Chap. 11] for a definition of extensive form games with imperfect information and perfect recall in that context.

In our setting, players make their decisions based on the sequence of observations of the current history, i.e., the first property requiring that players never forget their past knowledge is built into our model. Therefore, in $\mathfrak{P}$, a player has perfect recall if they can distinguish their own actions from one another.

Definition 2.45. For all $i \in [\![1, n]\!]$, $\mathcal{P}_i$ has *perfect recall* in $\mathfrak{P}$ if $A^{(i)} \subseteq \mathcal{Z}_i$ and for all $a^{(i)} \in A^{(i)}$ and $x \in S \cup \bigcup_{i' \in [\![1,n]\!]} A^{(i')}$, $\mathsf{Obs}_i(x) = a^{(i)}$ if and only if $x = a^{(i)}$.

Kuhn's theorem asserts the equivalence of mixed and behavioural strategies for players who have perfect recall. Whether two strategies are equal is not a satisfactory measure of equivalence of strategies. On the one hand, equality does not allow us to compare mixed and behavioural strategies, as they are different objects syntactically. On the other hand, two behavioural strategies may yield the same outcomes despite being different: the actions suggested by a strategy in an inconsistent history can be changed without affecting the distributions induced by the strategy. Therefore, instead of using the equality of strategies as a measure of equivalence, we consider some weaker notion of equivalence, referred to as *outcome equivalence*.

Definition 2.46 (Outcome equivalence)**.** Two randomised strategies σ_i and τ_i of $\mathcal{P}_i$ in $\mathcal{A}$ are *outcome-equivalent* if for all *pure* strategy profiles σ_{-i} of the players other than $\mathcal{P}_i$ and for all initial states s_{init}, the probability distributions $\mathbb{P}^{\sigma_1, \sigma_{-i}}_{s_{\mathsf{init}}}$ and $\mathbb{P}^{\tau_1, \sigma_{-i}}_{s_{\mathsf{init}}}$ coincide.

In the above definition, we only quantify over *pure strategies* of the others players for syntactic reasons: we have only defined distributions induced by strategy profiles where all strategies are mixed or all strategies are behavioural. We discuss this definition in Chapter 9.1. In particular, we will see that the outcome-equivalence of two mixed (resp. behavioural) strategies in the sense of Definition 2.46 implies that these strategies induce the same distributions with all mixed (resp. behavioural) strategies of the other players.

With the definition of outcome-equivalence, we can now state Kuhn's theorem formally. We provide a proof of Kuhn's theorem in Chapter 9.2.

Theorem 2.47 (Kuhn's theorem [Kuh53, Aum64])**.** *Let $i \in [\![1, n]\!]$. For every behavioural observation-based strategy σ_i of $\mathcal{P}_i$ in $\mathfrak{P}$, there exists an outcome-equivalent mixed strategy μ_i. If $\mathcal{P}_i$ has perfect recall, then for every mixed observation-based strategy μ_i of $\mathcal{P}_i$ in $\mathfrak{P}$, there exists an outcome-equivalent behavioural strategy σ_i.*

We remark that to derive a mixed strategy from a behavioural strategy, perfect recall is not necessary. This is due to the fact that a part of the definition of perfect recall holds automatically in our setting: players never forget their prior knowledge as they make their decisions based on histories of increasing length. In contrast to this, we can show that perfect recall is required to derive a behavioural strategy from a mixed strategy (see Example 9.1).

2.7.3 Observation-based Mealy machines

We now discuss finite-memory strategies in a context of imperfect information. We define a finite-memory strategy in $\mathfrak{P}$ as a strategy induced by a (stochastic) Mealy machine of $\mathcal{A}$ that the updates and next-move functions of which agree on inputs that share the same observation. We call such Mealy machines *observation-based.*

Definition 2.48. Let $i \in [\![1, n]\!]$. Let $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be a Mealy machine of $\mathcal{P}_i$. We say that $\mathfrak{M}$ is *observation-based* if for all $m \in M$, $s, t \in S$ and $\bar{a}, \bar{b} \in \bar{A}$, if $\mathsf{Obs}_i(s) = \mathsf{Obs}_i(t)$ and $\mathsf{Obs}_i(\bar{a}) = \mathsf{Obs}_i(\bar{b})$, then $\mathsf{up}_{\mathfrak{M}}(m, s, \bar{a}) = \mathsf{up}_{\mathfrak{M}}(m, t, \bar{b})$ and $\mathsf{nxt}_{\mathfrak{M}}(m, s) = \mathsf{nxt}_{\mathfrak{M}}(m, t)$.

An observation-based Mealy machine $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ of $\mathcal{P}_i$ can be seen as a tuple $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ where its update and next-move functions are of the form $\mathsf{up}_{\mathfrak{M}} \colon M \times \mathcal{Z}_i^3 \to \mathcal{D}(M)$ and $\mathsf{nxt}_{\mathfrak{M}} \colon M \times \mathcal{Z}_i \to \mathcal{D}(A^{(i)})$ respectively.

An observation-based Mealy machine need not induce an observation-based behavioural strategy (see Example 9.1). We provide two sufficient conditions that ensure that strategies induced by observation-based Mealy machines are behavioural observation-based strategies in Chapter 9.3. If the owner of the Mealy machine has perfect recall, then the induced strategy is an observation-based behavioural strategy (Lemma 9.8). Otherwise, if the Mealy machine has a single initial state and deterministic updates, then the strategy it induces is an observation-based behavioural strategy. (Lemma 9.9).

2.8 One-counter Markov decision processes

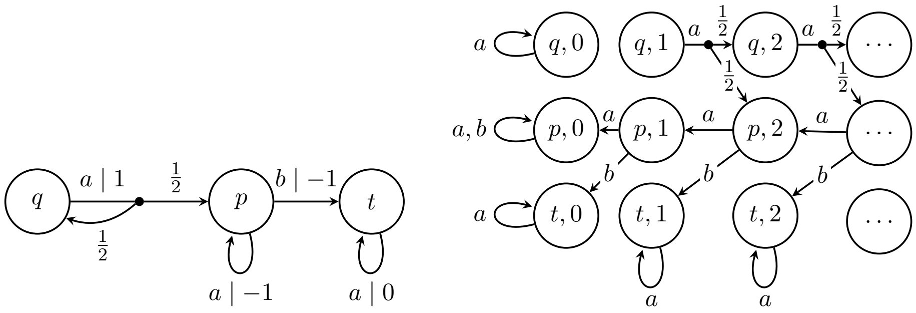
In Part V, we study one-counter MDPs, which are finite MDPs that induce countable-state MDPs. One-counter MDPs (OC-MDPs) extend MDPs with a counter that can be incremented, decremented or left unchanged on each transition.

Definition 2.49. A *one-counter MDP* is a tuple $\mathcal{Q} = (Q, A, \delta, w)$ where (Q, A, δ) is a finite MDP and $w\colon S \times A \to \{-1, 0, 1\}$ is a (partial) weight function that assigns an integer weight from $\{-1, 0, 1\}$ to state-action pairs.

For all $q \in Q$ and $a \in A$, we require that $w(q, a)$ be defined whenever $a \in A(q)$, i.e., all transitions are labelled by some weight. A *configuration* of $\mathcal{Q}$ is a pair (q, k) where $q \in Q$ and $k \in \mathbb{N}$. In the sequel, by plays, histories, strategies etc. of $\mathcal{Q}$, we refer to the corresponding notion with respect to the MDP (Q, A, δ) underlying $\mathcal{Q}$.

Remark 2.50. The weight function of an OC-MDP is not subject to randomisation in our definition, i.e., the weight of any transition is determined by the outgoing state and chosen action. In particular, counter values can be inferred from histories and be taken in account to make decisions in $\mathcal{Q}$. ◁

Let $\mathcal{Q} = (Q, A, \delta, w)$ be an OC-MDP. The OC-MDP $\mathcal{Q}$ induces an MDP over the infinite countable space of configurations. Transitions in this induced MDP are defined using δ for the probability of updating the state and w for the deterministic change of counter value. We interrupt any play whenever a configuration with counter value 0 is reached. Intuitively, such configurations can be seen as situations in which we have run out of an energy resource. We may also impose an upper bound on the counter value, and interrupt any plays that reach this counter upper bound. We refer to OC-MDPs with a finite upper bound on counter values as *bounded OC-MDPs* and OC-MDPs with no upper bounds as *unbounded OC-MDP*. We provide a unified definition for both



(a) An OC-MDP $\mathcal{Q}$. Weights are indicated next to actions.

(b) The MDP induced by the OC-MDP of Figure 2.7a.

Figure 2.7: An OC-MDP and the MDP over configurations it induces.

semantics.

Definition 2.51. Let $\mathcal{Q} = (Q, A, \delta, w)$ be an OC-MDP and let $B \in \bar{\mathbb{N}}_{>0}$ be a counter upper bound. We define the MDP $\mathcal{M}^{\leq B}(\mathcal{Q}) = (Q \times [\![B]\!], A, \delta^{\leq B})$ where $\delta^{\leq B}$ is defined, for all configurations $s = (q, k) \in Q \times [\![B]\!]$, actions $a \in A(q)$ and states $p \in Q$, by $\delta^{\leq B}(s, a)(p, k + w(q, a)) = \delta(q, a)(p)$ if $k \notin \{0, B\}$, and $\delta^{\leq B}(s, a)(s) = 1$ otherwise.

The state space of $\mathcal{M}^{\leq B}(\mathcal{Q})$ is finite if and only if the counter upper bound B is finite.

We briefly illustrate the semantics of a model via a simple illustration.

Example 2.6. We illustrate an OC-MDP $\mathcal{Q}$ in Figure 2.7a. A fragment of the MDP $\mathcal{M}^{\leq B}(\mathcal{Q})$ over configurations induced by $\mathcal{Q}$, for some $B \geq 3$, is depicted in Figure 2.7b. $\lhd$

We also introduce one-counter Markov chains. In one-counter Markov chains, we authorise stochastic counter updates (to simplify our presentation), i.e., counter updates are integrated in the transition function. This contrasts with OC-MDPs where deterministic counter updates are used to allow strategies to observe counter updates. We only require the unbounded semantics in this case.

Definition 2.52. A *one-counter Markov chain* is a tuple $\mathcal{R} = (Q, \delta)$ where Q is a finite set of states and $\delta\colon Q \to \mathcal{D}(Q \times \{-1, 0, 1\})$ is a probabilistic transition and counter update function. The one-counter Markov chain $\mathcal{R}$ induces a Markov chain $\mathcal{C}^{\leq\infty}(\mathcal{R}) = (Q \times \mathbb{N}, \delta^{\leq\infty})$ such that for any configuration $s = (q, k) \in Q \times \mathbb{N}$, any $p \in Q$ and any $u \in \{-1, 0, 1\}$, we have $\delta^{\leq\infty}(s)((p, k + u)) = \delta(q)(p, u)$ if $k \neq 0$ and $\delta^{\leq\infty}(s)(s) = 1$ otherwise.

In Part V, we focus on two reachability-based objectives in OC-MDPs. We recall that reachability objectives are central in synthesis (see [BGMR23]). On the one hand, we study state-reachability, which requires reaching a target regardless of the counter value.

Definition 2.53. Let $T \subseteq Q$ be a set of target states. The *state-reachability objective* for T in $\mathcal{M}^{\leq B}(\mathcal{Q})$ is defined as $\mathsf{Reach}(T \times [\![B]\!])$. We abusively denote this objective as $\mathsf{Reach}(T)$.

The second objective we consider is called selective termination: it requires reaching a counter value of zero in a target state.

Definition 2.54. Let $T \subseteq Q$ be a set of target states. The *selective termination objective*, denoted by $\mathsf{Term}(T)$, for T in $\mathcal{M}^{\leq B}(\mathcal{Q})$ is defined as $\mathsf{Reach}(T \times \{0\})$.

The selective termination objective generalises the *termination objective*, which requires reaching counter value zero.

2.9 Complexity theory

We assume that the reader is familiar with complexity theory and the traditional time and space complexity classes (in particular NP, co-NP and PSPACE). We refer the reader to the books of Sipser [Sip96] and Papadimitriou [Pap94] for complexity-theoretic background. In Part V, we present complexity bounds derived from the Blum-Shub-Smale model of computation and the decidability of the theory of the reals. We summarise some relevant properties for our purposes in the following.

2.9.1 The Blum-Shub-Smale model of computation

Some of our complexity bounds rely on a model of computation introduced by Blum, Shub and Smale [BSS89]. A Blum-Shub-Smale (BSS) machine, intuitively, is a random-access-memory machine with registers storing real numbers. Arithmetic computations on the content of registers are in constant time in this model.

The class of decision problem that can be solved in polynomial time in the BSS model coincides with the class of decision problems that can be solved, in the Turing model, in polynomial time with a PosSLP oracle [ABKM09]. The PosSLP problem asks, given a division-free straight-line program (intuitively, an arithmetic circuit), whether its output is positive. The PosSLP problem lies in the counting hierarchy and can be solved in polynomial space [ABKM09].

2.9.2 Theory of the reals

The theory of the reals refers to the set of sentences in the signature of ordered fields (i.e., fully quantified first-order logical formulae) that hold in $\mathbb{R}$. The problem of deciding whether a sentence is in the theory of the reals is decidable; if the number of quantifier blocks is fixed, this can be done in PSPACE [BPR06, Rmk. 13.10]. Furthermore, the problem of deciding the validity of an existential (resp. universal) formula is NP-hard (resp. co-NP-hard) [BPR06, Rmk. 13.9]. Our complexity bounds refer to the complexity classes ETR (existential theory of the reals) and co-ETR, which contain the problems that can be reduced in polynomial time to checking the membership of an *existential sentence* and *universal sentence* in the theory of the reals respectively.

Chapter 3

Contribution overview

In this chapter, we provide an overview of the precise problems we tackle and of the results presented in the later parts. The contributions of this thesis are structured into four parts. Each part is related to the concept of *strategy complexity* in games.

In Part II, we focus on the *classical Mealy machine* model and present upper bounds on the memory that is sufficient to construct (constrained) Nash equilibria in games with reachability or Büchi objectives and in games with shortest-path costs. In Part III, we revisit Kuhn's theorem in the finite-memory setting: we investigate how variations of randomised Mealy machines compare to one another in terms of expressiveness. We study the structure of expected payoff sets in Markov decision processes with multiple payoffs in Part IV and conclude that restricted randomisation suffices in this setting in many cases. Finally, in Part V, we study one-counter Markov decision processes, which induce countable MDPs. We study decision problems for interval strategies, a class of strategies that admit finite interval-based representations.

Contents

3.1 Memory for Nash equilibria

We motivate and summarise the results presented in Part II. These results are based on the single author paper [Mai24].

3.1.1 Context

Strategy complexity. In the context of synthesis via game theory, strategies are the formal counterpart of controllers. Therefore, *simpler strategies* are preferable in general whenever they exist. The complexity of a strategy is often measured by the *amount of memory* it requires (e.g., [FH13, CD12a, CRR14, RRS17, BGHM17]), which is formalised by the number of states of the smallest Mealy machine that induces it.

For a given class of games, this leads to two natural questions: "*how much memory is sufficient to enforce the specification?*" and "*how much memory is necessary to enforce the specification?*" in the studied games. The first question amounts to determining an upper bound on the amount of memory sufficient to win whenever possible (which may be parameterised by some property of the specification or of the arena). The second question asks for a lower bound witnessing how much memory could be needed in the worst case. Both the upper and lower bounds are sensitive to the considered class of strategies (e.g., randomised or not) and how Mealy machines are formalised (e.g., where we introduce randomisation) – see, e.g., [CdH04, Cha07, Hor09, CRR14].

We investigate the first question for pure Nash equilibria in multi-player deterministic turn-based games with respect to *move-independent* Mealy machines, i.e., Mealy machines whose updates do not depend on the actions occurring along the play. This definition of a Mealy machine is natural in games where arenas are described by graphs with unlabelled edges, as the sequence of states contains all of the information regarding the play (it is used, e.g., in [CRR14, CHVB18, BBGT21]). This choice has an impact on the obtained memory bounds (see Chapter 5.3), it can also be seen as imposing a restricted form of *imperfect information*: the players can only observe states and not actions throughout the play.

Nash equilibria. We study *Nash equilibria* [Nas50] in multi-player non-zero-sum games on infinite *deterministic turn-based arenas*. Recall that an NE from an initial state is a strategy profile such that no player has an incentive to unilaterally deviate from their strategy (Definition 2.40). We focus on *pure* NEs in games with reachability-related objectives, as reachability objectives are central in synthesis [BGMR23]. More precisely, we consider games with reachability and Büchi objectives, and with shortest-path cost functions all built on a single non-negative integer weight function (i.e., the weights are the same for all players).

It is known that NEs exist from all states in the games we consider here. For games with reachability and Büchi objectives, this follows from the existence result of [Umm06] for games where all players have ω-regular objectives. In games with shortest-path cost functions on finite arenas, the existence of NE follows from results for games with continuous cost functions if all weights are positive [FL83], and from [BDS13] for the case of non-negative weights. We build on the ideas of [BDS13] to extend their existence result to shortest-path games with non-negative integer weights on infinite arenas (Chapter 6.3).

Although NEs are guaranteed to exist, several incomparable NEs may co-exist within a game (Example 2.5). In practice, NEs where more players win are preferable. For instance, when modelling different components of a system as players with their own objectives, it is desirable that as many component specifications as possible be satisfied. This is the core motivation of the *constrained NE existence problem* (Definition 5.1): does there exist an NE whose cost profile is bounded from above by some input vector? We are interested in bounding *how much memory is sufficient* for solutions to the constrained NE existence problem when using *move-independent Mealy* *machines*. In general, memory is necessary for such solutions (see Examples 5.1 and 5.2).

A related question, that we do not address, is to quantify how much memory is sufficient for any NE. To the best of our knowledge, whether memoryless NEs always exist in the games we consider is not known.

Nash equilibria and punishment. A useful technique to construct NEs in the games we consider is through *punishment*. Given a play, all players agree to follow the play, and if any player deviates from the play, then all of the others band together to sabotage the deviating player. If the initial play is an NE outcome, then the strategy profile resulting from this construction is an NE from the first state of the play. This punishing mechanism is based on the proof of the folk theorem for NEs in repeated games [Fri71, OR94], which describes the set of NE payoff profiles in these games.

The punishment mechanism can be used to obtain memory upper bounds for solutions to the constrained NE existence problem that depend on the size of the arena (e.g., [BDS13, Umm08, BBGT21]). The main ideas of the classical argument are as follows. First, one shows that there exist plays resulting from NEs with a *finite representation*, e.g., a lasso. We then construct a (move-independent) Mealy machine whose states are given by the finite representation of the play and additional memory states to punish any deviating players. If some player is inconsistent with the play, the other players switch to a (finite-memory) punishing strategy to sabotage the deviating player; this enforces the stability of the equilibrium. The size of the Mealy machine depends on that of the finite description of the play, and thus *depends on the size of the arena*. Furthermore, this approach does not translate to infinite arenas, e.g., if the considered NE outcome is a simple play, it cannot be encoded in a (finite) Mealy machine.

3.1.2 Contributions

Summary. The contributions presented in Part II are twofold. First, we present constructions of *move-independent Mealy machines* for solutions to the constrained NE existence problem for reachability games (Theorem 7.7), shortest-path games with a single non-negative integer weight function (The-

Arena size	Finite	Infinite
Reachability	n^2 (Thm. 7.7)	
Shortest-path	$n^2 + 2n$ (Thm. 7.9)	
Büchi	$\lvert S\rvert + n^2 + n$ (Lem. 7.14)	Finite-memory (Thm. 7.13)

Table 3.1: Table of memory upper bounds for solutions to the constrained pure NE existence problem in n-player turn-based deterministic games. The set S denotes the state space of the arena. Bounds are given with respect to the move-independent Mealy machine model.

orem 7.9) and Büchi games (Theorem 7.13) that apply to arbitrary arenas, bypassing the finite-arena requirement of existing approaches. In other words, for these three types of games, we show that from any NE, we can derive another NE where all strategies are finite-memory and such that the same players accomplish their objective, without increasing their cost for shortest-path games.

Second, for reachability and shortest-path games, we provide memory bounds that are *independent* of the size of the arena which are quadratic in the number of players. For Büchi games, we show that finite memory suffices in finite and infinite arenas, and provide an explicit (arena-dependent) memory bound for finite arenas. We also argue that arena-independent memory bounds cannot be obtained in Büchi games (Example 7.6): we provide a family of two-player games played on finite arenas where NEs with an outcome in which the second player wins require a memory of size linear in the size of the arena. Table 3.1 summarises our memory bounds in finite and infinite arenas.

We briefly comment on the main elements that are used to obtain our results. We focus on reachability and shortest-path games, and only provide limited intuition for Büchi games (for which we have more limited results).

Punishing strategies. We build on a variation of the punishment mechanism, and therefore we require *punishing strategies* with limited memory. We use memoryless punishing strategies. Punishing strategies are obtained via zero-sum reachability, Büchi or shortest-path games: to punish a player $\mathcal{P}_i$, the other players band together and try to optimise the opposite of the cost or objective of $\mathcal{P}_i$.

In reachability and Büchi zero-sum games on deterministic turn-based arenas, the two players have memoryless uniformly optimal strategies [Maz01, EJ88]. In zero-sum shortest-path games, the adversary (whose goal is to maximise the shortest-path cost) does not necessarily have a uniformly optimal strategy in infinite arenas. Nonetheless, we can show that the adversary has a memoryless strategy that can punish the others enough to successfully implement the punishment mechanism regardless of the point of deviation (Theorem 6.5).

Simplifying Nash equilibria outcomes. We derive move-independent Mealy machines implementing an NE from well-shaped NE outcomes. We obtain these well-shaped outcomes by *simplifying* outcomes of other NEs. This simplification process does not increase the cost of any player for shortest-path games and leaves the set of winners unchanged for reachability and Büchi games.

We provide an intuition of the simplification process for NE outcomes in shortest-path and reachability games in Figure 3.1. Intuitively, we first break up the play into *segments* connecting the first occurrences of each visited target. We then transform each finite segment into a simple history such that it is not possible to reach the end of one of these histories faster (with respect to the weight function) by rearranging the states.

We use a similar simplification process for NE outcomes in Büchi games: we obtain either a play that can be decomposed into an ultimately periodic sequence of simple histories or a play that can be decomposed into a sequence of segments such that no state occurs in two distinct odd-indexed (resp. even-indexed) segments.

To show that the simplified outcomes are indeed NE outcomes, we rely on *characterisations* of NE outcomes: sufficient and necessary conditions that ensure that a play is the outcome of an NE (Theorems 6.8 and 6.9).

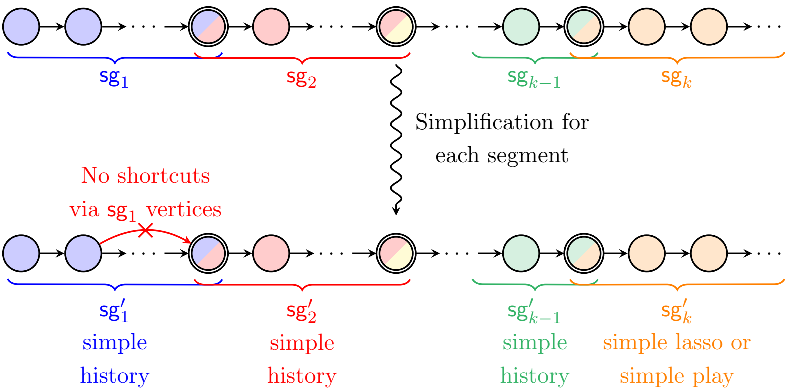

sg_1
sg_2
sg_{k-1}
sg_k
Simplification for each segment
No shortcuts via sg_1 vertices
sg'_1 simple history
sg'_2 simple history
sg'_{k-1} simple history
sg'_k simple lasso or simple play

Figure 3.1: Simplification process for an NE outcome in a multi-player shortest-path game. Doubly circles states denote the first occurrence of a target state for each player in the play.

Relaxing the punishment mechanism. Implementing the punishment mechanism with a move-independent Mealy machine for a given outcome requires a complete description of the intended outcome. In particular, we cannot obtain finite-memory strategies in infinite arenas through this approach. We propose a relaxation of the punishment mechanism: players only punish some specific deviations that can be considered as severe.

In reachability and shortest-path games, players keep track of two pieces of information: the current segment of the intended (simplified) outcome and the last player to have moved. It follows that players cannot react to in-segment deviations; however, none of these deviations can be profitable due to the absence of shortcuts. If a state outside of the current segment is reached, then the last player to have moved must have deviated: this deviation is deemed to be severe and is punished.

In Büchi games, the situation is slightly different: some in-segment deviations may be profitable, e.g., if a target of a player whose objective is not satisfied occurs in the segment and they can loop back to it. Intuitively, this phenomenon prevents us from obtaining arena-independent upper bounds on the size of move-independent Mealy machines implementing a solution to the constrained NE existence problem. To circumvent the issue, we use a two-phase approach: punish all deviations until there are no more targets of losing players in the remaining segments, then operate like in the above games.

3.2 Revisiting Kuhn's theorem with finite-memory assumptions

We motivate and summarise the results presented in Part III. These results are based on a collaboration with Mickaël Randour [MR24]. In the previous part, we have focused on pure strategies. We now move on to randomised strategies.

3.2.1 Context

Randomness in strategies. Strategies may require *randomisation* in concurrent games (e.g. [CD12b, dAHK07] and Example 2.3), games with imperfect information (e.g., [BGG17]) or in multi-objective settings (e.g., [EKVY08, RRS17, DKQR20], see also Part IV). There are different ways of implementing randomisation in strategies. On the one hand, a *mixed strategy* (Definition 2.15) randomly selects a pure strategy at the start of the play, and commits to it for the whole play. On the other hand, a *behavioural strategy* (Definition 2.11) selects an action randomly in each step.

Kuhn's theorem. In full generality, the classes of mixed and behavioural strategies are incomparable (e.g., [CDH10] or [OR94, Chap. 11]). Nonetheless, Kuhn's theorem [Aum64] (Theorem 2.47) asserts their equivalence under a mild hypothesis: if a player has *perfect recall*, then all of their mixed strategies admit an outcome-equivalent behavioural strategy and vice-versa. Intuitively, two strategies are outcome-equivalent (Definition 2.46) if they generate the same distributions over plays regardless of the decisions of the other players. We remark that, in our model of arenas (with imperfect information), mixed strategies are no less expressive than behavioural strategies even without perfect recall.

Finite-memory strategies. For reactive synthesis, infinite-memory strategies, along with randomised ones relying on infinite supports, are undesirable for implementation. We study finite-memory strategies described by stochastic Mealy machines (Definition 2.18). Randomisation can be implemented in these

Mealy machines in different ways: the *initialisation*, *outputs* or *transitions* can be randomised or deterministic respectively (see, e.g., [CDH10]). The equivalence stated in Kuhn's theorem motivates study of *the expressive power of different variants of stochastic Mealy machines.*

Kuhn's theorem and finite memory. The techniques underlying Kuhn's theorem cannot be extended directly to the finite-memory setting. Kuhn's theorem crucially relies on two properties. First, mixed strategies can be distributions over an *infinite* set of pure strategies. Second, strategies can use *infinite memory.* For instance, consider a memoryless behavioural strategy that flips a coin in each round to choose one of two actions. Such a strategy generates infinitely many sequences of actions, therefore any equivalent mixed strategy needs the ability to randomise between infinitely many pure strategies. Moreover, infinitely many of these sequences require infinite memory to be generated due to their non-regularity.

The previous example illustrates that it may not be possible to emulate a memoryless behavioural strategy by mixing finitely many pure finite-memory strategies, in spite of the idea of mixing pure finite-memory strategies being the natural finite-memory counterpart of mixed strategies. This highlights a dependency of expressive power depending on the randomisation power allowed in stochastic Mealy machines. In the sequel, we classify different classes of stochastic Mealy machines depending on their expressive power.

3.2.2 Contributions

Classifying Mealy machines. We classify finite-memory strategies following the type of stochastic Mealy machines that can induce them. We introduce a concise notation for each class: we use three-letter acronyms of the form XYZ with $\mathrm{X}, \mathrm{Y}, \mathrm{Z} \in \{\mathrm{D}, \mathrm{R}\}$, where X, Y and Z respectively refer to the *initialisation*, *outputs* and *updates* of the Mealy machines, with D and R respectively denoting deterministic and randomised components. For instance, we will write RRD to denote the class of Mealy machines that have randomised initialisation and outputs, but deterministic updates. We also apply this terminology to finite-memory strategies: we will say that a finite-memory strategy is in the class XYZ — i.e., it is an XYZ strategy — if it is induced by an XYZ Mealy machine.

We briefly comment on the appearance of some of these classes in the literature. Strategies in the class DRD have been referred to as *behavioural* finite-memory strategies in [CDH10]. This name comes from the randomised outputs, reminiscent of behavioural strategies that output a distribution over actions after a history. In some works, finite-memory randomised strategies are defined as DRD strategies (e.g., [Cha07, BFRR17]). Adding randomisation in outputs constitutes a natural approach to extend deterministic Mealy machines to encode randomised strategies; this way, the information maintained by the player is not subject to any randomness.

Similarly, RDD strategies have been referred to as *mixed* finite-memory strategies [CDH10]. The general definition of a mixed strategy is a distribution over pure strategies: under a mixed strategy, a player randomly selects a pure strategy at the start of a play and plays according to it for the whole play. RDD strategies are similar in the way that the random initialisation can be viewed as randomly selecting some DDD strategy (i.e., a pure finite-memory strategy) among a *finite* selection of such strategies. It follows that RDD are a special case of finite-support mixed strategies.

The elements of RRR, the broadest class of finite-memory strategies, have been referred to as general finite-memory strategies [CDH10] and stochastic-update finite-memory strategies [BBC$^+$14a, CKK17]. The latter name highlights the random nature of updates and insists on the difference with models that rely on deterministic updates, that are common in the literature.

Settings. We study four classes of *multi-player concurrent stochastic arenas*, where each class is described following (i) whether perfect recall is assumed or not and (ii) whether the arenas are finite or countable. These classes encompass two-player turn-based (deterministic) arenas with perfect information and (partially observable) Markov decision processes as particular subcases.

For each class of arenas, we compare strategy classes on the basis of outcome equivalence (Definition 2.46). We say that a class of strategies Σ_1 is *no less expressive* than a class Σ_2 if for all strategies in Σ_2, we can find an outcome-equivalent strategy in Σ_1; we abbreviate this by saying that Σ_2 is included in Σ_1. Through this comparison criterion, we establish a *Kuhn-like taxonomy* of the classes of finite-memory strategies obtained by varying which Mealy machine components (initialisation, outputs and updates) are randomised.

We illustrate this taxonomy through lattices highlighting the inclusion relationships. To separate classes of strategies, we provide separation examples on an MDP with one state and two actions, or a variation thereof if the separation result does not hold in finite arenas with perfect information. This MDP is arguably the simplest setting in which we can distinguish strategy classes.

In the remainder of this section, we comment on our results for finite arenas with perfect recall and for countable arenas with no assumption regarding perfect recall. We provide the lattices for the other classes of arenas we consider in Chapter 8.

Finite arenas with perfect recall. We first consider finite arenas with perfect recall. Our results are illustrated in the lattice illustrated in Figure 3.2. In the figure, a line between two strategy classes represents the strict inclusion of the lower class in the above class.

Unsurprisingly DDD strategies, i.e., pure finite-memory strategies, are the least expressive. For instance, in deterministic MDPs, DDD strategies can only induce a single outcome from each state unlike the other classes.

Several inclusions follow directly from some classes having more randomisation power than others: a deterministic component can be emulated using Dirac distributions. This argument yields, e.g., the inclusion of DRD in RRD. We obtain three inclusions that do not follow from such an argument: RDD $\subsetneq$ DRD, RRR $\subsetneq$ DRR and RRR $\subsetneq$ RDR. We briefly discuss each of these.

First, we obtain that RDD strategies can be emulated by DRD strategies, i.e., finite-memory *mixed strategies* are less expressive than finite-memory *behavioural strategies* (Theorem 10.2). Intuitively, our construction yields a DRD Mealy machine that keeps track of all of the strategies mixed by the RDD one, and we use randomised outputs to postpone the randomised initialisation: whenever two of the strategies that are mixed disagree, we randomly choose actions and discard the inconsistent strategies. The inclusion of RDD in DRD is strict. For instance, in a deterministic MDP, all RDD strategies have finitely many outcomes, whereas the DRD strategy that chooses actions uniformly at

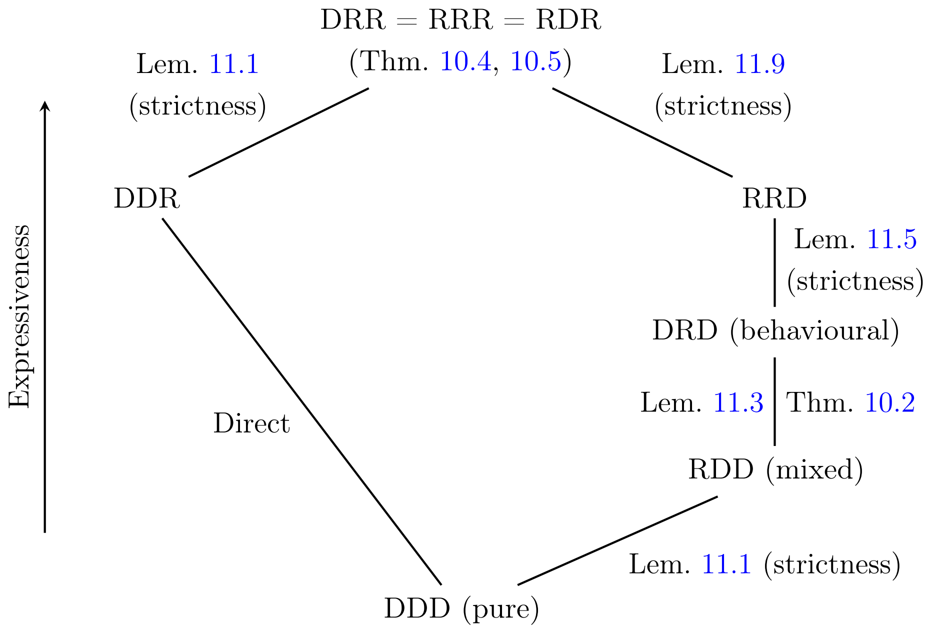


Figure 3.2: Lattice of finite-memory strategy classes in terms of expressive power in *finite multi-player arenas with perfect recall.* Each line in the figure indicates that the class above is strictly more expressive than the class below.

random in each step has infinitely many.

Second, we prove that RRR strategies can be emulated by DRR strategies, i.e., we can remove randomised initialisation from general Mealy machines without losing out on expressiveness (Theorem 10.4). The main idea is to add a fresh initial state to the Mealy machine. On the one hand, with randomised outputs, we can emulate the choices of the original RRR strategy in the first step of the game. On the other hand, with well-chosen randomised updates, we can arrange for the distribution over memory states after the first memory update to coincide in the DRR Mealy machine and the original RRR one.

The two inclusions sketched above exchange a randomised initialisation for another form of randomisation by using completely different constructions. We note that there is no uniform argument through which we can transform an RXY strategy into a DRY strategy in general; this is witnessed by the strict inclusion of DRD in RRD.

Finally, we show that RRR strategies can be emulated by RDR strategies, i.e., randomised outputs do not provide any additional expressive power in general Mealy machines (Theorem 10.5). In this case, the main idea is to use randomised initialisation and randomised outputs to preemptively draw actions for each state in each step.

We remark that we cannot remove both the randomised initialisation and randomised outputs from RRR strategies: this yields the DDR class. DDR strategies are less expressive that RRR strategies because they cannot make a random decision in the first step of a game. However, DDR is not a subset of RRD because, as suggested by the results above, stochastic updates enable essentially all behaviours that can be expressed by RRR strategies (from the second step of a play on). This explains why the class DDR is in its own branch in the lattice.

Countable arenas with no assumption on recall. We now consider our most general setting: countable arenas, possibly with imperfect recall. We illustrate the relevant lattice in Figure 3.3.

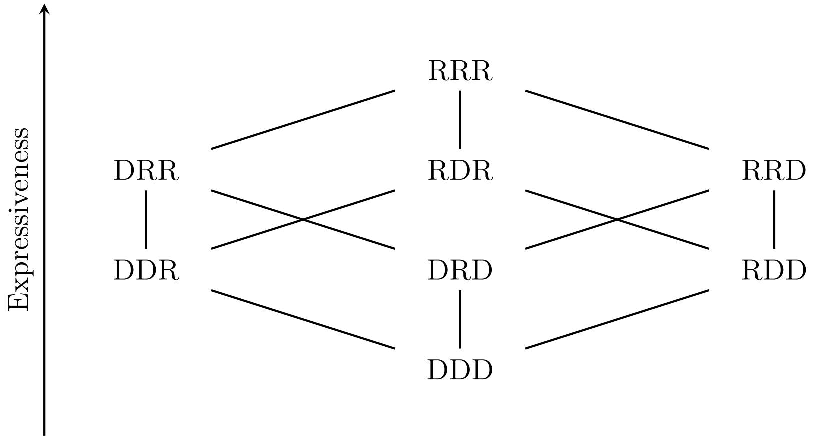


Figure 3.3: Lattice of finite-memory strategy classes in terms of expressive power in *general multi-player arenas*.

The only inclusions that hold in this general setting are due to some classes having more randomisation power than others. All separation results that hold in finite arenas with perfect recall extend to this setting. We briefly highlight two additional non-inclusions that result from broadening our setting.

First, we observe that RDD is not included in DRR (in the previous setting, this inclusion followed from RDD $\subseteq$ DRD $\subseteq$ DRR). This can be shown by a POMDP with one state and two indistinguishable actions, thus with imperfect recall. An RDD strategy that mixes the two constant strategies of this POMDP does not admit a DRR equivalent, as a DRR strategy has no means to determine the first action occurring in the game.

Second, we obtain that DRD is no longer included in RDR (which followed previously from RRR = RDR). This can be shown via an MDP with one action and infinitely many actions: a memoryless DRD strategy can play all actions with positive probability, but an RDR strategy can only use as many actions as there are memory states.

Each of the above examples exploits either imperfect recall or an infinite arena, but not both. In particular, they yield separations in the two settings on which we do not comment in this section.

In a nutshell. We provide a full picture of the expressiveness of randomised finite-memory strategies in variants of the classical Mealy machine model. Through our use of outcome-equivalence to compare strategy classes, we obtain a taxonomy that is agnostic to the choice of payoffs and objectives, and the way these are defined, e.g., over sequences of states or via colours labelling transitions.

In Chapter 11, we complement our separation results on finite arenas with game instances from the literature for which strategies of some class suffice and others do not to enforce a specification. For instance, we note that RDD strategies suffice in multi-objective MDPs with reachability objectives but not DDD strategies [EKVY08], and that DRD strategies suffice to win almost-surely in a zero-sum concurrent reachability game, but RDD strategies do not [dAHK07]. Therefore, in a sense, our taxonomy of strategy classes also extends to the framework in which strategies are compared on the basis of their performance with respect to specifications.

3.3 The structure of payoff sets in multi-objective Markov decision processes

We motivate and summarise the results presented in Part IV. The results presented in this section are based on joint work with Mickaël Randour [MR25].

3.3.1 Context

A multi-dimensional vision of strategy complexity. The *randomisation power* of a strategy is *another factor* that contributes to its complexity, in addition to the memory of the strategy. Randomisation is distinct from *memory* from the standpoint of strategy complexity: for some specifications, randomisation can be traded-off for memory, already in the one-dimensional case [CdH04, Hor09, CRR14, MPR20]. Furthermore, as highlighted by the classification of randomised finite-memory strategies presented in Section 3.2, *not all randomised strategies are created equal* even in the perfect information setting.

For memory requirements, we are often interested in the sufficient and necessary amounts of memory to enforce a specification. For randomisation, we can ask the similar question *"what is the simplest form of randomisation that is sufficient to enforce a given specification?"* whenever randomisation is necessary. In the following, we study randomisation requirements in *multi-objective Markov decision processes*, a setting in which randomisation is necessary.

We note that randomisation requirements can be studied without taking memory in account, i.e., we can identify other natural subclasses of randomised strategies besides those of the previous section. We will be interested in *finite-support mixed strategies*, i.e., mixed strategies that randomise over finitely many pure strategies. Such strategies can thus only use a limited form of randomisation.

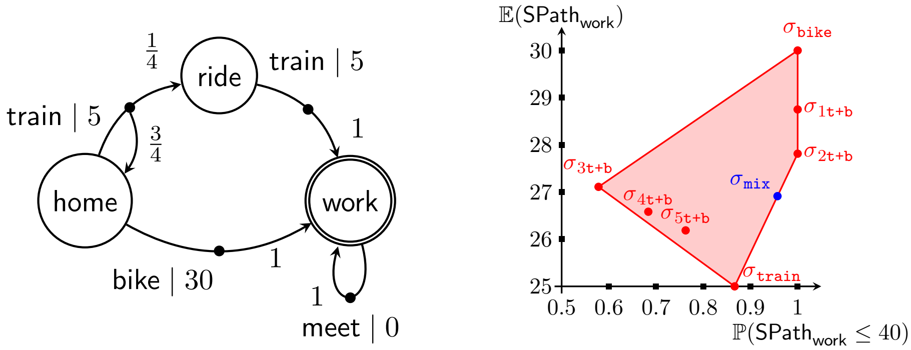
Multi-objective Markov decision processes. We consider Markov decision processes with multi-dimensional payoff functions, i.e., *multi-objective Markov decision processes*. Multi-dimensional payoff functions can be used to model specifications imposing several simultaneous constraints on a system, e.g., constraints on the response time and energy consumption of a system. In this setting, some expected payoff vectors may be incomparable; an analysis of trade-offs between the different dimensions may be necessary.

The goal is generally to determine, given a vector, whether it is *achievable*, i.e., whether there exists a strategy whose expected payoff from an initial state is greater than or equal to the vector in all components (e.g., [EKVY08, CFW13, RRS17]). A related problem is to compute or approximate the *Pareto curve* of the set of expected payoffs, i.e., the expected payoffs that are *Pareto-optimal* (e.g., [FKP12, CKK17, QK21]). Intuitively, an expected payoff is Pareto-optimal if there is no strategy whose expected payoff is as good on all dimensions and strictly better on one dimension. Alternatively, one can look for strategies with expected payoffs that are optimal for the *lexicographic* order over vectors (e.g., [HPS$^+$21, CKM$^+$23, BCM$^+$23]). For instance, in a two-dimensional setting, this equates to finding strategies that maximise the expected payoff on the second dimension *among* the strategies that maximise the expected payoff on the first dimension.

In general, strategies with both *memory and randomisation* may be necessary in multi-objective MDPs to achieve some vectors (see, e.g., [RRS17, DKQR20, BGMR23] and the example below). Our focus is on *randomisation requirements* in (countable) multi-objective MDPs: we study whether *limited randomisation power* suffices to achieve vectors.

A simple example. For the sake of illustration, let us consider the MDP depicted in Figure 3.4a. This MDP models a situation where a person wants to go to work. They must choose between riding their bicycle or taking the train. However, the train may be delayed with high probability due to an ongoing strike. The goal of the commuter is twofold: maximise the likelihood of reaching work within 40 time units (to reach work on time) and do so as fast as possible on average. We model the two goals with a shortest-path cost function.

We illustrate the set of expected payoffs for this situation in Figure 3.4b. We label some expected payoffs with their corresponding strategies. On the one hand, σ_{train} and σ_{bike} denote the pure strategies that always choose the action in the subscript. On the other hand, we denote by $\sigma_{\ell\mathsf{t}+\mathsf{b}}$ the strategy that attempts to take the train ℓ times before choosing the bicycle.



(a) Transitions are labelled by actions and probabilities. Weights next to actions represent the time taken by the action. The target is doubly circled.

(b) A set of expected payoffs. The probability of reaching work before 40 time units is given on the first dimension. The other dimension represents the expected time to reach work.

Figure 3.4: An MDP with two payoffs and its associated set of expected payoffs.

This simple example highlights the need for randomisation and memory in multi-objective MDPs. When limited to pure strategies, for instance, it is not possible to reach work on time with probability exceeding 90% while guaranteeing an expected commute time lower than 27. While the figure does not feature the expected payoffs of all pure strategies, we observe that any pure strategy whose payoff is not represented will reach work on time with a smaller probability than σ_{train}, and thus will not be satisfactory. However, as suggested by the point labelled by σ_{mix} on the illustration, these constraints can be satisfied by mixing σ_{train} and $\sigma_{\mathsf{2t+b}}$. In fact, here, all expected payoffs can be obtained by *finite-support mixed strategies*. We explain below how this property generalises.

We remark that, in general, expected payoff sets need not be convex polytopes like in the previous example. We depict a set of expected payoffs that is not a polytope in Figure 3.5; this set has infinitely many extreme points. This figure is taken from a multi-objective MDP with two discounted-sum payoffs that is presented in Chapter 14.1.

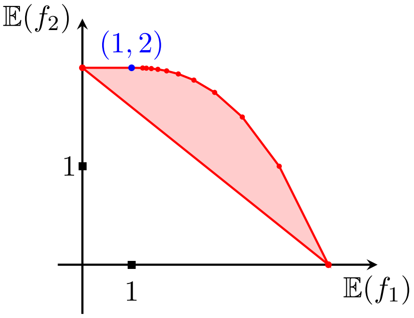


Figure 3.5: An expected payoff set that is not a polytope; taken from the example of Chapter 14.1 of a multi-objective MDP with two discounted-sum payoffs.

3.3.2 Contributions

Philosophy. Our goal is to provide *general results*, i.e., that apply to a broad class of payoffs. We only consider *universally unambiguously integrable* payoffs (Definition 2.26), i.e., payoffs that have a well-defined (possibly infinite) expectation no matter the strategy. This constitutes a natural requirement when aiming to reason about expectations. We obtain finer results for *universally integrable* payoffs (Definition 2.27), i.e., payoffs whose expectation is *finite* under all strategies.

Overview. Our contributions are twofold. On the one hand, we study the structure of sets of expected payoffs in *countable multi-objective MDPs*, focusing on the relationship between what can be obtained with pure and with randomised strategies. Through these relationships, we obtain results regarding randomisation requirements in multi-objective MDPs. On the other hand, we investigate sufficient conditions that ensure that any achievable vector is dominated by a *Pareto-optimal payoff*. We show that this holds for a subclass of *continuous* payoffs in *finite MDPs*. We remark that Pareto-optimal payoffs need not exist in general; already in the one-dimensional case, optimal strategies need not exist.

Payoff type	Univ. int.	Univ. unamb. int.
Lexico. opt.	Pure strat. (Thm. 14.1)	Inf. support (Ex. 14.1)
Achieving a vect.	Fin.-support mixed strat. (Thm. 14.4)	Inf. support (Ex. 14.4)
Approx. a vect.	Fin.-support mixed strat. (Thm. 14.4)	Fin.-support mixed strat. (Thm. 14.7)

Table 3.2: Summary of randomisation requirements in countable multi-objective MDPs for lexicographic optimisation, achieving a vector and approximating an expected payoff vector. Univ. int. and univ. unamb. int. respectively stand for universally integrable and universally unambiguously integrable. For each case, either pure strategies or finite-support mixed strategies suffice, or finite-support mixed strategies do not suffice.

The structure of expected payoff sets. We relate the set of expected payoffs of pure strategies and the set of expected payoffs of general strategies in *countable multi-objective MDPs*. We summarise the results described in the following in Table 3.2.

First, we prove that for universally integrable (multi-dimensional) payoffs, for all strategies, there exists a *pure strategy* with a greater or equal expected payoff in the lexicographic sense (Theorem 14.1). In other words, *randomisation is not necessary* for lexicographic optimisation of universally integrable payoffs.

This first result serves as a building block to one of our main results: any expected payoff vector of a universally integrable (multi-dimensional) payoff is a *convex combination* of expected payoffs of pure strategies (Theorem 14.4). From a strategic perspective, this means that in multi-objective MDPs, *finite-support mixed strategies* suffice to exactly obtain any (Pareto-optimal) expected payoff vector. As a corollary, we obtain that any extreme point of a set of expected payoffs of a universally integrable payoff can be obtained by a pure strategy (Corollary 14.5). These results generalise known properties for classical combinations of objectives on finite MDPs for which the set of expected payoffs is a convex polytope (e.g., combinations of ω-regular specifications [EKVY08] and universally integrable total-reward payoffs [FKP12]), and for which the set of expected payoff vectors need not be a polytope (e.g., discounted-sum payoffs [CFW13] – see Figure 3.5).

Although none of the previous properties generalise to the whole class of universally unambiguously integrable payoffs (even in finite MDPs, see Examples 14.1 and 14.4), we prove that, for such payoffs, convex combinations of pure strategies can be used to *approximate* any expected payoff (Theorem 14.7).

In both cases, we can bound the number of strategies that are mixed depending on the number of dimensions d. Depending on the setting, we can match or approximate expected payoffs by mixing no more than $d+1$ strategies (Theorem 14.8).

Our results highlight the role of randomisation in strategies for multi-objective MDPs: it is useful *only* to balance the payoffs on different dimensions in many cases.

To establish these results, we mainly reason on *mixed strategies* rather than behavioural strategies. Mixed strategies provide a crucial link between expected payoffs of pure and randomised strategies: the expected payoff of a mixed strategy is an integral with respect to the mixed strategy of the expected payoffs under all pure strategies (Lemma 13.4). Kuhn's theorem allows us to extend our results to behavioural strategies.

We comment on the applicability of the above results. The class of *universally integrable* payoffs is large: it contains most classical payoffs. Indeed, all *bounded* payoffs are de facto in this class. For example, all indicators of objectives are in it. This means that all settings where one considers the *probabilities* of sets of plays (i.e., either an inherently qualitative objective or one arising from fixing a threshold for a quantitative payoff) are in it, for *any* definition of such sets of plays. Classical examples include ω-regular objectives [EKVY08], window objectives [BDOR20] or percentiles queries [RRS17]. Bounded payoffs also encompass discounted-sum [Sha53] and mean-payoff [BBC$^+$14b, CKK17] functions. Heterogeneous combinations, such as, e.g., combinations of energy and mean-payoff [BHRR19] also fit under this umbrella.

Payoffs that are not universally unambiguously integrable fall out of scope of our results. It makes sense to exclude such payoffs: their expectation need not be well-defined. Such out-of-scope payoffs include some instances of total reward payoff or shortest-path cost with both negative and positive weights, e.g., when there are plays with a payoff of positive infinity and negative infinity with non-zero probability under some strategy. When only non-negative weights (or only non-positive weights) are used, a classical restriction [FKP12, RRS17], these two types of payoffs are universally unambiguously integrable (but not necessarily universally integrable), and thus are covered by our results.

We now briefly discuss memory requirements in multi-objective MDPs with respect to mixing. For some universally integrable payoffs in finite MDPs, the sets of expected payoffs are polytopes whose extreme points can be obtained via pure finite-memory strategies (e.g., ω-regular objectives [EKVY08] or mean-payoff [BBC$^+$14b]). It follows that for these objectives, mixing over *finitely many pure finite-memory strategies* is sufficient to fulfil any achievability requirement. In other words, one of the least expressive models of randomised finite-memory strategies (with respect to our classification in Part III), i.e., RDD strategies, suffices. Furthermore, the blow-up in memory from this mixing argument is small: it suffices to mix, at most, one more strategy than the number of payoffs.

Continuous payoffs. We provide a sufficient condition on payoffs to guarantee that the set of expected payoffs is closed in *finite MDPs*. We show that this is the case for continuous universally *square integrable* payoffs, i.e., continuous payoffs that are universally integrable when squared (Theorem 15.8). Universally square integrable payoffs are a special case of universally integrable payoffs (due to the Cauchy-Schwarz inequality, see, e.g., [Dur19, Thm. 1.5.2.]).

The class of continuous universally square-integrable payoffs includes real-valued continuous payoffs (in particular *discounted-sum payoffs*) and *universally integrable shortest-path* costs based on a *positive* weight function (see Chapter 15.4). It follows that for combinations of these payoffs, any achievable vector can be bounded from above by a Pareto-optimal expected payoff. We remark that expected payoff sets for continuous universally square integrable payoffs need not be polytopes: this is witnessed by the example illustrated in Figure 3.5, which is based on discounted-sum payoffs.

To prove that expected payoffs sets are closed for continuous universally square integrable payoffs in finite MDPs, we introduce a notion of convergence of behavioural strategies (Chapter 15.1) and show that the function mapping a strategy to the expectation under this strategy is continuous (Theorem 15.7). Our approach depends on the compactness of the set of plays, and therefore does not translate to countable MDPs.

3.4 Counter-based strategies in one-counter Markov decision processes

We motivate and summarise the results presented in Part V. The results presented in this section are based on joint work with Michal Ajdarów, Petr Novotný and Mickaël Randour [AMNR25].

3.4.1 Context

Strategies and their representations. In reactive synthesis via games, the goal is to automatically construct a strategy representing the sought controller. Traditionally, synthesised strategies are finite-memory and are represented by Mealy machines. A special subclass of particular interest is that of *memoryless strategies*.

Memoryless strategies are functions assigning (distributions over) actions to each state. Therefore, in *infinite arenas*, even these strategies, which can be seen as the simplest strategies from the viewpoint of memory, need not admit a finite representation. The contribution presented in this part focuses on small (and particularly, finite) counter-based representations of memoryless strategies in a fundamental class of infinite-state MDPs: *one-counter MDPs*.

One-counter MDPs. *One-counter MDPs* [BBE+10] (see Definition 2.49) are finite MDPs augmented with a counter that can be incremented (by one), decremented (by one) or left unchanged on each transition. Considering such counter updates is not restrictive for modelling: any integer counter update can be obtained with several transitions. However, this impacts the complexity of decision problems.

An OC-MDP induces a possibly infinite MDP over a set of *configurations* given by states of the underlying MDP and counter values (Definition 2.51). In this induced MDP, we interrupt plays that reach counter value zero; this event is called *termination*. We consider two variants of the model: *unbounded OC-MDPs*, where counter values can grow arbitrarily large, and *bounded OC-MDPs*, in which plays are interrupted when a fixed counter upper bound is reached. OC-MDPs are small representations of large MDPs: unbounded OC-MDPs have infinitely many configurations and bounded OC-MDPs have exponentially many configurations with respect to the binary encoding of the counter upper bound.

Termination is the canonical objective in OC-MDPs [BBE+10]. Also relevant is the more general *selective termination* objective (Definition 2.54), which requires terminating in a target set of states. In this work, we study both the selective termination objective and the *state-reachability* objective (Definition 2.53), which requires visiting a target set of states regardless of the counter value.

Synthesis in OC-MDPs. Optimal strategies need not exist in unbounded OC-MDPs for these objectives [BBEK13]. The general synthesis problem in unbounded OC-MDPs for selective termination is not known to be decidable, and it is at least as hard as the positivity problem for linear recurrence sequences [PB24], whose decidability would yield a major breakthrough in number theory [OW14]. Optimal strategies exist in bounded OC-MDPs: the induced MDP is finite and we consider reachability objectives. However, constructing optimal strategies is already EXPTIME-hard for reachability in finite-horizon MDPs [BKN+19], i.e., OC-MDPs in which all weights are negative (and thus the number of steps in the play is bounded by the initial counter value).

We propose to tame the inherent complexity of analysing OC-MDPs by restricting our analysis to a class of succinctly representable (yet natural and expressive) strategies called *interval strategies.*

3.4.2 Contributions

Interval strategies. We introduce *interval* strategies (Definition 17.2): an interval strategy is based on some (finite or infinite but finitely-representable) partitioning of $\mathbb{N}$ into intervals, and the strategy's decision depends on the current state and on the interval containing the current counter value. More precisely, we focus on two classes of these strategies. On the one hand, in bounded and unbounded OC-MDPs, we consider *open-ended interval strategies* (OEISs) for which the underlying partitioning is finite. On the other hand, in unbounded OC-MDPs, we also consider *cyclic interval strategies* (CISs): strategies for which there exists a (positive integer) period such that, for any two counter values that differ by the period, we take the same decisions.

We collectively refer to OEISs and CISs as *interval strategies*. While interval strategies are not sufficient to play optimally in unbounded OC-MDPs [BKSV08] (see also Examples 17.2 and 17.3), they can be used to approximate the supremum probability for the objectives we consider (Lemma 17.5).

We can show that OEISs in bounded OC-MDPs and CISs in unbounded OC-MDPs can be exponentially more concise than equivalent Mealy machines, and OEISs in unbounded OC-MDPs can even represent infinite-memory strategies (Chapter 17.2).

Decision problems. For selective termination and state-reachability, we consider three decision problems. First is the *interval strategy verification problem* (Definition 17.6): it asks whether the probability of the objective from an initial state under the given strategy is greater or equal to a given threshold. The other two problems are realisability problems for structurally-constrained interval strategies. On the one hand, the *fixed-interval realisability* problem (Definitions 17.7 and 17.8) asks, given an interval partition, whether there is an interval strategy built on this partition that ensures the objective with a probability greater than a given threshold. Intuitively, in this case, the system designer specifies the desired structure of the controller and it remains to specify the action choices for each interval. On the other hand, the *parameterised realisability* problem for interval strategies (Definitions 17.9 and 17.10), asks whether there exists a well-performing strategy built on a partition of size no more than a parameter d such that no finite interval is larger than a second parameter n. We consider two variants of the realisability problems: one for checking the existence of a suitable pure strategy and another for randomised strategies. Randomisation allows for better performance when imposing structural constraints on strategies (see Example 17.4), but pure strategies are however often preferred for synthesis [DKQR20], as randomness is undesirable for certain applications (e.g., in the medical field).

Compressed Markov chains. Analysing the performance of a memoryless strategy in an MDP amounts to studying the Markov chain it induces on the MDP (Definition 2.14). Our results rely on the analysis of a *compressed Markov chain* derived from the (potentially infinite) Markov chain induced by an interval strategy (Chapter 18). We remove certain configurations and aggregate several transitions into one. This compressed Markov chain preserves the probability of selective termination and of hitting counter upper bounds (Theorem 18.4). However, its transition probabilities may require exponential-size representations or even be irrational. To represent these probabilities, we characterise them as the least solutions of quadratic systems of equations (Theorems 18.6 and 18.9); these characterisations are respectively derived from and inspired by those of [KEM06] for termination probabilities in probabilistic pushdown automata. Compressed Markov chains are finite for OEISs and are induced by a one-counter Markov chain for CISs (Section 18.5).

Complexity results. We summarise our complexity results in Table 3.3. The crux of our algorithmic results is the aforementioned compression. For verification, we reduce the problem to checking the validity of a universal formula in the theory of the reals, by exploiting our characterisation of transition probabilities in compressed Markov chains. This induces a PSPACE upper bound. For bounded OC-MDPs, we can do better: verification can be solved in polynomial time in the unit-cost arithmetic RAM model of computation of Blum, Shub and Smale [BSS89] (described in Chapter 2.9), by computing transition and reachability probabilities of the compressed Markov chain. This yields a $\mathsf{P}^{\mathsf{PosSLP}}$ complexity in the Turing model (see [ABKM09]).

Both realisability variants exploit the verification approach through the theory of the reals. For fixed-interval realisability for pure strategies, we exploit non-determinism to select good strategies and then verify them with the above. In the randomised case, in essence, we build on the verification formulae and existentially quantify over the probabilities of actions under the sought strategy. Finally, for parameterised realisability, we build on our algorithms for the fixed-interval case by first non-deterministically building an appropriate partition.

We also provide complexity lower bounds. We show that all of our considered

Semantics	Bounded		Unbounded			
Strategy type	*Open-ended*		*Open-ended*		*Cyclic*	
Verification	$\mathsf{P}^{\mathsf{PosSLP}}$ Thm. 19.1		co-ETR Thm. 19.5		co-ETR Thm. 19.9	
	sqrt-sum-hard Thm. 21.7		sqrt-sum-hard [EWY10]			
Realisability (both variants)	*Pure* $\mathsf{NP}^{\mathsf{PosSLP}}$ Thm. 20.2	*Random* $\mathsf{NP}^{\mathsf{ETR}}$ Thm. 20.4	*Pure* $\mathsf{NP}^{\mathsf{ETR}}$ Thm. 20.5	*Random* PSPACE Thm. 20.7	*Pure* $\mathsf{NP}^{\mathsf{ETR}}$ Thm. 20.8	*Random* PSPACE Thm. 20.10
	NP-hard (termination, Thm. 21.12) and sqrt-sum-hard (cf. verification)					

Table 3.3: Complexity bound summary for our problems. All bounds are below PSPACE. Square-root-sum-hardness results are derived from instances of the form $\sum \sqrt{x_i} \geq y$.

problems are hard for the square-root-sum problem (Definition 21.1), a problem that is not known to be solvable in polynomial time but that is solvable in polynomial time in the Blum-Shub-Smale model [Tiw92]. We also prove NP-hardness for our realisability problems for selective termination, already when checking the existence of good single-interval strategies. We provide a reduction from the problem of deciding whether a finite directed graph contains a Hamiltonian cycle (Definition 21.11).

Impact. Our results provide a natural class of strategies for which realisability is decidable (whereas the general case remains open and known to be difficult [PB24]), and with arguably low complexity (for synthesis). Furthermore, the class of interval strategies is of practical interest thanks to their concise representation and their inherently understandable structure (in contrast to the corresponding Mealy machine representation).

Part II:

Memory, a classical measure of strategy complexity

Chapter 4

Introduction

In this part, we present the results described in Chapter 3.1, originating from the single-author paper [Mai24]. We study Nash equilibria in multi-player games on turn-based deterministic arenas with reachability objectives, shortest-path costs and Büchi objectives. We investigate how much memory is sufficient to obtain a constrained pure Nash equilibrium from an initial state, i.e., a Nash equilibrium whose cost profile is bounded from above (with respect to the component-wise order) by a given vector, when considering Mealy machines whose updates disregard actions.

We refer the reader to Chapter 3.1 for an extended presentation of the context. We divide this part into three chapters. We summarise their contents below. We comment on related work at the end of this chapter.

Memory for constrained Nash equilibria. Chapter 5 presents the constrained Nash equilibrium existence problem and move-independent Mealy machines. Throughout this part, we focus on move-independent Mealy machines as our means of representing strategies and quantifying memory.

In an n-player game, the *constrained NE existence problem* asks, given an initial state and an n-dimensional vector, whether there exists an NE from the initial state whose cost profile is bounded from above by the given vector (Definition 5.1). Such an NE is called a *solution* to the constrained NE existence problem. Memory is necessary in general for solutions to the constrained NE existence problem: it is useful if there are several targets to be visited (Example 5.1) and to threaten other players to prevent them from having profitable deviations (Example 5.2).

We then consider *move-independent strategies and Mealy machines.* A strategy is move-independent if its decisions depend only on the sequence of game states occurring throughout the play. Similarly, a deterministic Mealy machine is move-independent if its memory updates depend only on states and not on actions. When considering deterministic turn-based arenas described by directed graphs (i.e., there is at most one transition from one state to another), all strategies are move-independent – this presentation of arenas is used, e.g., in [GTW02, BCJ18].

In finite arenas, all finite-memory move-independent strategies are induced by move-independent Mealy machines (Lemma 5.5). However, this is no longer the case in infinite arenas: *move-dependent Mealy machines* can represent pure strategies that cannot be represented through (finite) move-independent Mealy machines (Example 5.3). In particular, upper bounds on the sufficient memory for solutions to the constrained NE existence problem obtained through general (action-aware) Mealy machines do not yield upper bounds when using move-independent Mealy machines. For this reason, we directly construct move-independent Mealy machines in our analysis.

Punishing strategies and characterisations of NE outcomes. In Chapter 6, we provide an overview of several technical results that we use to construct Nash equilibria. We present results for zero-sum games on deterministic turn-based arenas and characterisations of outcomes of pure Nash equilibria.

We build NEs through a variant of the punishing mechanism: the players follow along a given outcome and, if some player deviates from the intended outcome, the other players join together to sabotage the deviating player. This sabotage is executed through *punishing strategies* obtained via *coalition games* (Definition 6.7). A coalition game is a zero-sum game derived from a multiplayer game in which one player $\mathcal{P}_i$ plays against the coalition of the other players, who aim to maximise the cost of $\mathcal{P}_i$. We recall classical results on zero-sum reachability and Büchi games on deterministic turn-based arenas from which we obtain memoryless punishing strategies (see Theorems 6.1 and 6.2). We also show that pure memoryless punishing strategies exist in shortest-path games (Theorem 6.5), despite the fact that, in zero-sum shortest path games, optimal strategies need not exist for the second player who aims to maximise the shortest-path cost (Example 6.1).

To implement the punishing mechanism, we require an NE outcome. Furthermore, this outcome must be well-structured to result from a finite-memory strategy profile. In Chapter 7, we describe methods to appropriately simplify NE outcomes while improving their cost profile. To guarantee that this simplification approach yields NE outcomes, we use *characterisations of NE outcomes* based mainly on values in zero-sum coalition games (Theorems 6.8 and 6.9). Intuitively, these characterisations state that a play is an NE outcome if the values of states along the play do not suggest the existence of a profitable deviation.

We close the chapter by discussing the fact that pure NEs exist from all states in the games we consider. This follows from known results in all cases besides shortest-path games on infinite arenas. We recall these results and we provide an existence proof in multi-player shortest-path games with non-negative integer weights that applies to finite and infinite arenas (see Theorem 6.11). Our argument is based on our characterisation of NE outcomes in these games.

Memory bounds for Nash equilibria. We provide our main results regarding memory requirements for constrained Nash equilibria in Chapter 7. On the one hand, we show that there exist *arena-independent* upper bounds on the sufficient amount of move-independent memory for constrained NEs in reachability games (Theorem 7.7) and in shortest-path games (Theorem 7.9) that hold in particular in *infinite arenas*. On the other hand, we show that for Büchi games, if there exists a solution to an instance of the constrained pure NE existence problem, then there exists one where the strategies are induced by move-independent Mealy machines (Theorem 7.13).

For all of the aforementioned games, we rely on a relaxation of the punishment mechanism: we consider NE outcomes and construct strategies such that players only punish certain deviations and disregard some others. To obtain finite-memory strategies via this mechanism, we build on well-structured outcomes; see Lemmas 7.3 and 7.4 for shortest-path and reachability games and Lemmas 7.11 and 7.12 for Büchi games.

Related work. We discuss three directions related to multi-player non-zero sum games. The first direction is related to the constrained equilibrium existence problem as a *decision problem* (e.g., [BMR14, BBMU15]). On finite arenas, deciding the existence of a constrained NE is NP-complete for reachability and (non-negative weighted) shortest-path games [BBGT21] and is in P for Büchi objectives [Umm08]. This problem has also been studied for other types of equilibria, e.g., for subgame perfect equilibria in reachability and (non-negative weighted) shortest-path games [BBG$^+$20] and in mean-payoff games [BvdBR23, BRvdB22] and for secure equilibria in weighted games for the supremum, infimum, limit superior, limit inferior and mean payoffs [BMR14].

Second, the construction of our finite-memory NEs rely on *characterisations* of plays resulting from NEs. Their purpose is to ensure that the punishment mechanism can be used to guarantee the stability of an equilibrium. In general, these characterisations can be useful from an algorithmic perspective; deciding the existence of a constrained NE boils down to finding a play that satisfies the characterisation. Characterisations appear in the literature for NEs [Umm08, UW11, BBMU15], but also for other types of equilibria, e.g., subgame perfect equilibria [BBG$^+$20] and secure equilibria [BMR14].

Chapter 5

The role of memory in constrained Nash equilibria

In general, even if Nash equilibria are guaranteed to exist, there need not be guarantees on their cost profile. In practice however, we are more interested in Nash equilibria in which the players have a low cost, e.g., if players model different (independent) components of a system to be controlled. This motivates the *constrained (pure) Nash equilibrium existence problem*, which asks, in an multi-player game, whether there exists a pure NE from a given initial state whose cost profile is no more than a given vector. Such an NE is called a solution to the constrained existence problem.

In the subsequent chapters, we endeavour to provide upper bounds on the amount of memory that is sufficient for these solutions when considering *move-independent Mealy machines*, i.e., Mealy machines whose memory updates only depend on states (and thus disregard actions). Strategies induced by such Mealy machines are called *move-independent*: their decisions do not take in account the actions taken throughout the play.

The main purpose of this chapter is to formalise the constrained existence problem, due to its role as a motivation for this work, and move-independent Mealy machines, as our chosen strategy representation. We also provide examples witnessing the need for memory in solutions to the constrained NE existence problem.

In Section 5.1, we formalise the constrained NE problem. We then prove the necessity of memory in Section 5.2 and highlight two roles that memory plays. Finally, we discuss move-independent strategies and move-independent (deterministic) Mealy machines in Section 5.3. In addition to formalising these notions, we investigate the impact of removing actions from Mealy machine updates on memory size requirements.

We fix an n-player deterministic turn-based arena $\mathcal{A} = ((S_i)_{i\in[\![1,n]\!]}, A, \delta)$ for the remainder of the chapter.

Contents

5.1 Constrained equilibrium existence problem

For all types of games we consider in this section, deciding whether a pure NE exists from a given initial state is trivial: there exists an NE from any initial state. We discuss NE existence results in Chapter 6.3. Existence results from the literature do not directly apply to games with shortest-path costs on infinite arenas; we provide an explicit existence proof instead (Theorem 6.11).

Although NEs are guaranteed to exist, several incomparable NEs may co-exist within a game, as we have seen in Example 2.5, i.e., this existence result does not provide any guarantees on the quality of the cost profile. Therefore, a natural question is to ask, given an initial state, whether there exists an NE from this initial state in which the costs of the players are good enough. This problem is formalised as follows.

Definition 5.1 (Constrained NE existence problem)**.** Let $\mathcal{G} = (\mathcal{A}, (f_i)_{i\in[\![1,n]\!]})$ be a game where f_i is a cost function for all $i \in [\![1, n]\!]$. The *constrained (pure) NE existence problem* asks, given an initial state $s_{\mathsf{init}} \in S$ and $\mathbf{q} \in \bar{\mathbb{R}}^n$, whether there exists a pure NE σ from s_{init} such that $(f_i(\mathsf{Out}_{\mathcal{A}}(\sigma, s_{\mathsf{init}})))_{i\in[\![1,n]\!]} \leq \mathbf{q}$. Such an NE is called a *solution* to the constrained NE existence problem.

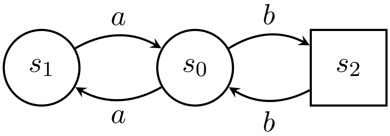


Figure 5.1: A two-player deterministic turn-based arena.

In games where the goals of the players are given by objectives, the constrained NE existence problem asks whether there exists an NE the outcome of which is winning for a given subset of players. Our goal is to *bound the sufficient amount of memory* for solutions to the constrained NE existence problem.

5.2 The necessity of memory

Memory may be required for solutions to the constrained Nash equilibrium existence problem. In this section, we provide two examples that highlight this need for memory. We use these examples to explain the roles of memory for NEs in games with reachability objectives; this discussion is inspired by an invited contribution co-authored with Thomas Brihaye, Aline Goeminne and Mickaël Randour [BGMR23]. While we only focus on reachability objectives, the following also applies to games with shortest-path cost functions and games with Büchi objectives, due to their similarity.

Intuitively, memory serves *two roles*. On the one hand, memory is useful to *satisfy several objectives*, i.e., for reachability, to visit several targets that do not all lie on a single simple history. On the other hand, memory may be useful to *prevent profitable deviations* of other players. We illustrate these needs via two examples.

We first provide an example highlighting the need for memory to visit several targets.

Example 5.1. We let $\mathcal{A}$ be the arena depicted in Figure 5.1. We consider the reachability game $\mathcal{G} = (\mathcal{A}, (\mathsf{Reach}(s_1), \mathsf{Reach}(s_2))$, i.e., the target of $\mathcal{P}_i$ is the state s_i for $i \in \{1, 2\}$. We show that there exists a solution to the constrained NE existence problem instance asking for a pure NE from s_0 such that the objectives of both players are satisfied in its outcome, and that any solution to this problem instance requires memory.

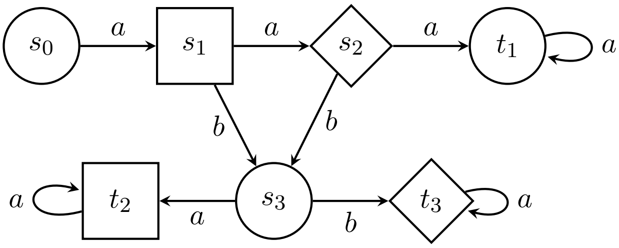


Figure 5.2: A three-player deterministic turn-based arena. Circles, squares and diamonds are resp. $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$ states.

We let σ_2 denote the pure memoryless strategy of $\mathcal{P}_2$ in $\mathcal{A}$ such that $\sigma_2(s_2) = b$. This is the only strategy of $\mathcal{P}_2$ in $\mathcal{A}$. We consider a pure strategy σ_1 of $\mathcal{P}_1$ that alternates between actions a and b after each visit to s_0. It follows that the outcome of σ_1 and σ_2 from s_0 satisfies the objectives of the two players, i.e., the strategy profile (σ_1, σ_2) is an NE from s_0 such that both players win in its outcome.

However, there are no pure memoryless NE from s_0 with an outcome that is winning for both players: the only two outcomes from s_0 of pure memoryless strategy profiles in $\mathcal{A}$ are $(s_0as_1a)^\omega$ and $(s_0bs_2b)^\omega$. In other words, memory is needed to obtain an NE from s_0 that visits both s_1 and s_2. ◁

We now showcase the second application of memory: to prevent profitable deviations. In this case, we use memory to implement a *punishment mechanism*: the players band together to sabotage any player who strays from the intended outcome of the NE (see also: the description of this mechanism in Chapter 3.1).

Example 5.2. We consider the three-player reachability game $\mathcal{G}$ on the arena depicted in Figure 5.2 where the objective of $\mathcal{P}_i$ is $\mathsf{Reach}(t_i)$ for $i \in [\![1,3]\!]$. We claim that memory is necessary to obtain a pure NE from s_0 such that t_1 is visited in its outcome.

Let $\sigma = (\sigma_i)_{i\in[\![1,3]\!]}$ be a pure memoryless strategy profile such that t_1 is visited in $\mathsf{Out}_{\mathcal{A}}(\sigma, s_0)$, i.e., $\sigma(s_0) = \sigma(s_1) = \sigma(s_2) = a$. We claim that σ is not an NE from s_0. If $\sigma(s_3) = a$, then $\mathcal{P}_2$ has a profitable deviation by choosing b in s_1 instead of a. Similarly, if $\sigma(s_3) = b$, then $\mathcal{P}_3$ has a profitable deviation by choosing b in s_2 instead of a. Therefore, σ is not an NE from s_0.

We can construct a pure NE $\sigma = (\sigma_i)_{i\in[\![1,3]\!]}$ from s_0 such that t_1 is visited as follows. We let σ_2 and σ_3 be the pure memoryless strategies of $\mathcal{P}_2$ and $\mathcal{P}_3$ that select action a in all states. We define σ_1 to be the strategy that chooses b in s_3 if s_2 has not been visited, and chooses a otherwise. We check that σ is an NE from s_0 as follows: if $\mathcal{P}_2$ deviates and uses action b in s_1, then $\mathcal{P}_1$ reacts to this deviation by avoiding t_2, and similarly, if $\mathcal{P}_3$ deviates and uses action b in s_2, then $\mathcal{P}_1$ reacts to this deviation by avoiding t_3.

We have shown that there exists an NE from s_0 in $\mathcal{G}$ such that the target of $\mathcal{P}_1$ is visited, and that any such NE requires memory. $\lhd$

The two previous examples illustrate that memory is necessary for solutions to the constrained NE existence problem.

5.3 Move-independent strategies

Upper bounds on the memory sufficient to win in zero-sum games or to obtain NEs in multi-player games are sensitive to the choice of strategy model. Intuitively, more general models of strategies yield smaller bounds. For instance, in certain settings, better memory bounds can be obtained by considering randomised strategies instead of pure strategies (e.g., [CdH04, CHP08, Hor09, CRR14, MPR20]). On the other hand, establishing memory bounds for more restrictive models implies memory bounds for more general models.

In this section, we introduce move-independent strategies and Mealy machines. We then show that all finite-memory move-independent strategies can be implemented with move-independent Mealy machines in finite arenas: we provide a translation from a general Mealy machine to a move-independent one that involves encoding the previous visited state in the memory. Finally, we provide an example illustrating that any translation from Mealy machines implementing move-independent strategies to move-independent Mealy machines requires a blow-up relative to the number of states of the arena. In particular, in infinite arenas, strategies implemented by move-independent Mealy machines are *a strict subset* of move-independent finite-memory strategies.

Definitions are given in Section 5.3.1. We show that, in finite arenas, move-independent pure strategies are finite-memory strategies if and only if they are induced by a move-independent Mealy machine in Section 5.3.2. Finally, we show that a blow-up proportional to the size of the memory state space is necessary to construct move-independent Mealy machines from general ones in Section 5.3.3.

5.3.1 Definitions

A strategy is move-independent if its decisions are agnostic to the actions chosen throughout histories. This can be formalised as follows.

Definition 5.2. Let $i \in [\![1, n]\!]$. A strategy σ_i of $\mathcal{P}_i$ in $\mathcal{A}$ is *move-independent* if for any two histories $h = s_0 a_0 s_1 \dots a_{r-1} s_r$ and $h = s_0 b_0 s_1 \dots b_{r-1} s_r$ that traverse the same sequence of states, we have $\sigma_i(h) = \sigma_i(h')$.

Similarly, we say that a Mealy machine is move-independent if its updates disregard the chosen actions. Formally, a deterministic move-independent Mealy machine is defined as follows.

Definition 5.3. Let $i \in [\![1, n]\!]$. A deterministic Mealy machine $\mathfrak{M} = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ of $\mathcal{P}_i$ is *move-independent* if for all $m \in M$, $s \in S$ and $a, b \in A(s)$, $\mathsf{up}_{\mathfrak{M}}(m, s, a) = \mathsf{up}_{\mathfrak{M}}(m, s, b)$. If $\mathfrak{M}$ is move-independent, we view its update function as a function $M \times S \to M$.

As the name suggests, a deterministic move-independent Mealy machine induces a move-independent strategy.

Remark 5.4 (Stochastic Mealy machines)*.* The above definition is not satisfactory for stochastic Mealy machines: for a stochastic Mealy machine, even if the update function disregards the chosen action, the strategy it induces may not be move-independent. For instance, consider a one-state deterministic MDP with two actions and the mixed strategy that randomises over the two pure constant strategies of the MDP. This strategy is not move-independent: its decisions depend on the choice of action in the first round. However, this strategy can be implemented by a Mealy machine with two states, a randomised initialisation and updates that leave memory states unchanged. In particular, these updates are agnostic to actions. This motivates the restriction of the above definition to deterministic Mealy machines. ◁

We close the section by commenting on the use of move-independent Mealy machines in the literature. Some authors define Mealy machines as move-independent Mealy machines (e.g., [CRR14, CHVB18, BBGT21]), i.e., with updates that depend only on states and not actions. This definition is also used in the paper [Mai24] from which the results presented in Chapter 7 originate. In these works, (turn-based) arenas are presented as finite directed graphs where vertices are partitioned among the different players and transitions are described by the edge relation of the graph. In such arenas, all strategies are move-independent. Furthermore, as these works are concerned with finite arenas, a pure strategy is finite-memory if and only if it is induced by a move-independent Mealy machine. However, this choice of model can influence any memory requirements obtained for winning strategies or equilibrium profiles.

In our case, restricting our attention to move-independent strategies can also be seen as imposing a specific type of imperfect information on the decision making of the players: they can only observe the state of the world and not the actions that are taken by the others. This models the situation in which actions are internal to the players, and only their effect on the state of the arena can be observed.

5.3.2 Move-independent Mealy machines in finite arenas

We now assume that $\mathcal{A}$ is finite. Our goal is to show that any pure finite-memory strategy that is move-independent is induced by a move-independent Mealy machine. We consider a (move-dependent) deterministic Mealy machine $\mathfrak{M}$ that induces a move-independent strategy and derive a move-independent Mealy machine $\mathfrak{N}$ from $\mathfrak{M}$ as follows. The state space of $\mathfrak{N}$ is obtained by augmenting memory states of $\mathfrak{M}$ with the state visited in the previous step of the play. We use this additional information to mimic the memory updates of $\mathfrak{M}$ in $\mathfrak{N}$ one time step later. We formalise this idea in the following proof.

Lemma 5.5. *Assume that the state space of $\mathcal{A}$ is finite. Let $i \in [\![1, n]\!]$. For all deterministic Mealy machines $\mathfrak{M} = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ of $\mathcal{P}_i$ that induce move-independent strategies, there exists a move-independent Mealy machine inducing an outcome-equivalent strategy with $|M| \cdot |S| + 1$ memory states.*

Proof. For all $s \in S \setminus S_i$ and $t \in S$, we let $a_{s,t}$ denote a *fixed* action such that $\delta(s, a_{s,t}) = t$ if such an action exists. We use these actions in the definition of the update and next-move functions of our move-independent Mealy machines, to avoid having to observe actions.

Let $\mathfrak{M} = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be a deterministic Mealy machine inducing a strategy of $\mathcal{P}_i$. We define an outcome-equivalent move-independent Mealy machine $\mathfrak{N} = (N, \top, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ as follows. First, we let $N = (M \times S) \cup \{\top\}$ where $\top$ is a new initial memory state. For all $m \in M$ and all $s, t \in S$, we define $\mathsf{up}_{\mathfrak{N}}(\top, t) = (m_{\mathsf{init}}, t)$ and

$$\mathsf{up}_{\mathfrak{N}}((m, s), t) = \begin{cases} (\mathsf{up}_{\mathfrak{M}}(m, s, \mathsf{nxt}_{\mathfrak{M}}(m, s)), t) & \text{if } s \in S_i \\ (\mathsf{up}_{\mathfrak{M}}(m, s, a_{s,t}), t) & \text{otherwise,} \end{cases}$$

and, if $t \in S_i$, we define $\mathsf{nxt}_{\mathfrak{N}}(\top, t) = \mathsf{nxt}_{\mathfrak{M}}(m_{\mathsf{init}}, t)$ and

$$\mathsf{nxt}_{\mathfrak{N}}((m, s), t) = \begin{cases} \mathsf{nxt}_{\mathfrak{M}}(\mathsf{up}_{\mathfrak{M}}(m, s, \mathsf{nxt}_{\mathfrak{M}}(m, s)), t) & \text{if } s \in S_i \\ \mathsf{nxt}_{\mathfrak{M}}(\mathsf{up}_{\mathfrak{M}}(m, s, a_{s,t}), t) & \text{otherwise.} \end{cases}$$

We remark that $|N| = |M| \cdot |S| + 1$.

We now show that $\mathfrak{M}$ and $\mathfrak{N}$ induce outcome-equivalent strategies. Two strategies are outcome-equivalent if and only if they agree over the histories consistent with the two strategies (see the outcome-equivalence criterion of Lemma 9.1 for a formal argument); we establish outcome-equivalence through this property. Let σ_i and τ_i denote the strategies induced by $\mathfrak{M}$ and $\mathfrak{N}$ respectively.

First, we establish the following property by induction: for any history prefix $w = w'sa \in (SA)^+$ consistent with σ_i, there exists a history $u's \in \mathsf{Hist}(\mathcal{A})$ consistent with σ_i and sharing the same sequence of states as $w's$ such that $\widehat{\mathsf{up}_{\mathfrak{N}}}(w) = (\widehat{\mathsf{up}_{\mathfrak{M}}}(u'), s)$. We prove this by induction on the number of actions in the history prefixes. For the base case, we consider a history prefix $w = sa$ and let u' be the empty word. By definition of $\mathsf{up}_{\mathfrak{N}}$ (and its iterated version), we have $\widehat{\mathsf{up}_{\mathfrak{N}}}(w) = (m_{\mathsf{init}}, s)$. This ends the proof of the base case.

We now let $w = w''tbsa \in (SA)^+$ be a history prefix consistent with σ_i. We let $w' = w''tb$. We assume by induction that there exists a history $u''t$ consistent with σ_i and sharing the same sequence of states as $w''t$ such that

$\widehat{\mathsf{up}_{\mathfrak{N}}}(w') = (\widehat{\mathsf{up}_{\mathfrak{M}}}(u''), t)$. We first define a suitable history prefix u'. If $t \in S_i$, we let $u' = u''tb$. Otherwise, we let $u' = u''ta_{t,s}$. In both cases, we obtain that $u's$ shares the same sequence of states as $w's$ and is consistent with σ_i: for the first case, it follows from the consistency of w with σ_i and move-independence of σ_i.

To end the induction argument, it remains to show that $\widehat{\mathsf{up}_{\mathfrak{N}}}(w) = (\widehat{\mathsf{up}_{\mathfrak{M}}}(u'), s)$. By definition and the induction hypothesis, we have

$$\widehat{\mathsf{up}_{\mathfrak{N}}}(w) = \mathsf{up}_{\mathfrak{N}}(\widehat{\mathsf{up}_{\mathfrak{N}}}(w'), s) = \mathsf{up}_{\mathfrak{N}}\left((\widehat{\mathsf{up}_{\mathfrak{M}}}(u''), t), s\right).$$

We distinguish two cases, depending on whether $\mathcal{P}_i$ controls t. First, assume that $t \in S_i$. By consistency of w with σ_i, we obtain that $\mathsf{nxt}_{\mathfrak{M}}(\widehat{\mathsf{up}_{\mathfrak{M}}}(w''), t) = b$. It follows from the above and the definition of $\mathsf{up}_{\mathfrak{N}}$ that

$$\widehat{\mathsf{up}_{\mathfrak{N}}}(w) = \left(\mathsf{up}_{\mathfrak{M}}\left(\widehat{\mathsf{up}_{\mathfrak{M}}}(u''), t, b\right), s\right) = \left(\widehat{\mathsf{up}_{\mathfrak{M}}}(u'), s\right).$$

This ends the first case. We now assume that $t \notin S_i$. It follows from the above, the definition of $\mathsf{up}_{\mathfrak{N}}$ and of u' that

$$\widehat{\mathsf{up}_{\mathfrak{N}}}(w) = \left(\mathsf{up}_{\mathfrak{M}}\left(\widehat{\mathsf{up}_{\mathfrak{M}}}(u''), t, a_{t,s}\right), s\right) = \left(\widehat{\mathsf{up}_{\mathfrak{M}}}(u'), s\right).$$

This ends the inductive argument.

We use the property proven above to show that σ_i and τ_i agree over the set of histories consistent with σ_i. For histories consisting of a single state s (all of which are consistent with σ_i), the result is direct by definition of $\mathsf{nxt}_{\mathfrak{N}}$. Let $h = w'tbs \in \mathsf{Hist}_i(\mathcal{A})$ be a history consistent with σ_i in which at least two states occur. We let $w = w'tb$. We let $u't \in \mathsf{Hist}_i(\mathcal{A})$ be a history consistent with σ_i sharing the same sequence of states as $w't$ such that $\widehat{\mathsf{up}_{\mathfrak{N}}}(w) = (\widehat{\mathsf{up}_{\mathfrak{M}}}(u'), t)$, given by the above property. We distinguish two cases depending on the ownership of t (due to the definition of $\mathsf{nxt}_{\mathfrak{N}}$).

First, we assume that $t \in S_i$. By consistency of h with σ_i and move-independence of σ_i, we obtain that

$$b = \sigma_i(w't) = \sigma_i(u't) = \mathsf{nxt}_{\mathfrak{M}}\left(\widehat{\mathsf{up}_{\mathfrak{M}}}(u'), t\right).$$

We let $u = u'tb$; us and h share the same sequence of states. Therefore, by move-independence of σ_i, we have

$$\sigma_i(h) = \sigma_i(us) = \mathsf{nxt}_{\mathfrak{M}}(\widehat{\mathsf{up}_{\mathfrak{M}}}(u), s).$$

We now apply our assumption on $\widehat{\mathsf{up}_{\mathfrak{N}}}(w)$ and $b = \mathsf{nxt}_{\mathfrak{M}}(\widehat{\mathsf{up}_{\mathfrak{M}}}(u'), t)$, together with the definition of $\mathsf{nxt}_{\mathfrak{N}}$ to obtain

$$\tau_i(h) = \mathsf{nxt}_{\mathfrak{M}}\left(\mathsf{up}_{\mathfrak{M}}\left(\widehat{\mathsf{up}_{\mathfrak{M}}}(u'), t, b\right), s\right) = \mathsf{nxt}_{\mathfrak{M}}(\widehat{\mathsf{up}_{\mathfrak{M}}}(u), s) = \sigma_i(h).$$

This ends the proof of the case $t \in S_i$.

We now assume that $t \notin S_i$. In this case, we let $u = u'ta_{t,s}$. Like above, move-independence of σ_i implies that

$$\sigma_i(h) = \sigma_i(us) = \mathsf{nxt}_{\mathfrak{M}}(\widehat{\mathsf{up}_{\mathfrak{M}}}(u), s).$$

We now apply our assumption on $\widehat{\mathsf{up}_{\mathfrak{N}}}(w)$ with the definition of $\mathsf{nxt}_{\mathfrak{N}}$ to obtain

$$\tau_i(h) = \mathsf{nxt}_{\mathfrak{M}}\left(\mathsf{up}_{\mathfrak{M}}\left(\widehat{\mathsf{up}_{\mathfrak{M}}}(u'), t, a_{t,s}\right), s\right) = \mathsf{nxt}_{\mathfrak{M}}(\widehat{\mathsf{up}_{\mathfrak{M}}}(u), s) = \sigma_i(h).$$

We have shown the outcome-equivalence of σ_i and τ_i. □

We comment on the increase in size of the memory state space that follows from the construction of Lemma 5.5 in the following section.

5.3.3 The cost of move independence

Lemma 5.5 requires that the state space of $\mathcal{A}$ be finite. The construction presented in its proof yields a move-independent Mealy machine whose memory state space depends on the size of $|S|$. We show through the following example that such a dependency cannot be avoided in general. Through the same example, we show that the result of Lemma 5.5 cannot be adapted to infinite arenas: there exists a move-independent finite-memory strategy in an infinite arena that cannot be encoded in a (finite) move-independent Mealy machine.

Example 5.3. We define a two-player turn-based arena for each non-empty subset of the natural numbers. For each of these arenas, we define a pure finite-memory strategy of $\mathcal{P}_1$ in this arena and show that if there exists a move-independent Mealy machine inducing it, then it requires a number of memory states equal to the size of the considered subset of $\mathbb{N}$.

Let $I \subseteq \mathbb{N}$ be non-empty. We define a two-player turn-based arena $\mathcal{A}_I = (\{t\} \times I, (\{s\} \times I) \cup \{s_{\mathsf{init}}, s_{\neq}, s_{=}\}, I, \delta)$ where the transition function δ is defined,

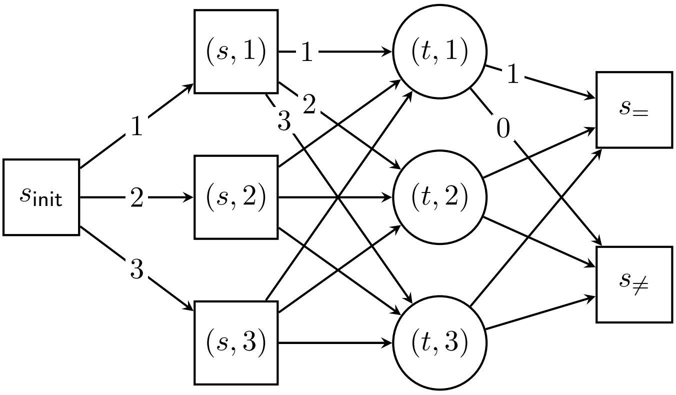


Figure 5.3: The arena $\mathcal{A}_{[\![1,3]\!]}$ of Example 5.3. Action labels for transitions from $(s,2)$, $(s,3)$, $(t,2)$ and $(t,3)$ are omitted to lighten the figure. The strategy of $\mathcal{P}_1$ that moves to $s_=$ if and only if the counter value in the previously visited states coincide cannot be represented with a move-independent Mealy machine with less than three states.

for all $k, a \in I$, by $\delta(s_{\mathsf{init}}, a) = (s, a)$, $\delta((s,k), a) = (t, a)$, $\delta((t,k), 1) = s_=$, $\delta((t,k), 0) = s_{\neq}$ and $s_=$ and $s_{\neq}$ are absorbing. We depict $\mathcal{A}_{[\![1,3]\!]}$ in Figure 5.3. We remark that the size of the state space of $\mathcal{A}_I$ is proportional to the size of I: if I is finite, we have $2 \cdot |I| + 3$ states in the arena.

We consider the strategy σ_1 of $\mathcal{P}_1$ that moves from (t,k) to $s_=$ if and only if the play starts in (t,k) or if (s,k) was visited previously. We provide a two-state (move-dependent) Mealy machine implementing this strategy and then show that there is no move-independent Mealy machine with strictly less than $|I|$ states that induces it.

We define $\mathfrak{M} = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ as follows. We let $M = \{0,1\}$ and $m_{\mathsf{init}} = 1$. For any $m \in M$ and $a, k \in I$ such that $a \neq k$, we let $\mathsf{up}_{\mathfrak{M}}(m, (s,k), a) = 0$. In other cases, the update function leaves the memory state unchanged. Finally, for all $m \in M$ and $k \in I$, we let $\mathsf{nxt}_{\mathfrak{M}}(m, (s,k)) = m$. It is easy to see that $\mathfrak{M}$ implements the strategy described previously.

We now show that any move-independent Mealy machine with less than $|I|$ states cannot induce σ_1. For any such Mealy machine, because there are $|I|$ actions but fewer memory states, there exist two different history prefixes $s_{\mathsf{init}}a(s,a)c$ and $s_{\mathsf{init}}b(s,b)c$ that lead to the same memory state for all actions $c \in I$ (this last property follows from the Mealy machine being move-independent). By choosing $c = a$, we obtain that the strategy induced by the move-independent Mealy machine agrees on the histories $s_{\mathsf{init}}a(s,a)a(t,a)$ $s_{\mathsf{init}}b(s,b)a(t,a)$, and thus differs from σ_1. $\lhd$

Chapter 6

Punishing strategies and characterisations of Nash equilibrium outcomes

In Chapter 7, we build Nash equilibria via an adaptation of the classical punishment mechanism. Intuitively, the punishment mechanism functions as follows: if some player deviates from the intended outcome, the other players coordinate as a coalition to prevent the player from having a profitable deviation. The strategy of the coalition used to sabotage the deviating player is called a *punishing strategy*. To obtain finite-memory Nash equilibria through the punishment mechanism, we need *simple punishing strategies* and *well-structured Nash equilibrium outcomes*.

We present results on strategies in zero-sum games in Section 6.1, that imply the existence of simple punishing strategies. We then provide characterisations of Nash equilibrium outcomes in Section 6.2; we use them in Chapter 7 to construct well-structured Nash equilibrium outcomes. Although unrelated to Chapter 7, we close this section by discussing existence results for Nash equilibria in Section 6.3.

Contents

6.1 Zero-sum games

We present an overview of relevant results for two-player zero-sum games where $\mathcal{P}_1$ has a reachability, Büchi objective, or a shortest-path cost function. The main takeaway of this section is that we can always find pure memoryless punishing strategies in the three classes of games we consider.

In Section 6.1.1, we recall classical results on reachability and Büchi games. In Section 6.1.2, we show that memoryless punishing strategies exist in shortest-path games.

We fix a two-player turn-based deterministic arena $\mathcal{A} = (S_1, S_2, A, \delta)$ and a target $T \subseteq S$ for the remainder of this section.

6.1.1 Reachability and Büchi games

In a two-player zero-sum game where $\mathcal{P}_1$'s goal is expressed by an objective, we call *winning region* the set of states from which $\mathcal{P}_1$ has a (surely) winning strategy.

We first discuss zero-sum reachability games. Let $\mathcal{G} = (\mathcal{A}, \mathsf{Reach}(T))$ denote the zero-sum reachability game on $\mathcal{A}$ where the target of $\mathcal{P}_1$ is T. A well-known result is that zero-sum reachability games on turn-based deterministic arenas enjoy *memoryless determinacy*: they are determined and for both players, there exist *pure memoryless uniformly winning strategies*. In other words, in $\mathcal{G}$, there exist pure memoryless strategies σ_1 and σ_2 such that, for all $s \in S$,

- if $\mathcal{P}_1$ has a winning strategy from s, then all outcomes of σ_1 from s are in $\mathsf{Reach}(T)$;
- otherwise, all outcomes of σ_2 from s are in $\mathsf{Safe}(T) = \mathsf{Plays}(\mathcal{A}) \setminus \mathsf{Reach}(T)$.

The classical proof of memoryless determinacy of reachability games (see, e.g., [Maz01, Prop. 2.18]) relies on a characterisation of the winning region of $\mathcal{P}_1$ in $\mathcal{G}$ as the least fixed point of the *controllable predecessor operator* when starting from T. Intuitively, this operator adds takes a set of states X, and adds to it the states from which $\mathcal{P}_1$ can enforce a visit to X in a single step. A by-product of this characterisation is that all strategies (even with memory) of $\mathcal{P}_2$ that select actions that do not enter the winning region of $\mathcal{P}_1$ whenever possible are uniformly winning strategies of $\mathcal{P}_2$. We use this property to establish the effectiveness of our punishing strategies. We summarise the properties of interest for reachability games in the following theorem.

Theorem 6.1. *Zero-sum reachability games on turn-based deterministic arenas are determined and both players have pure memoryless uniformly winning strategies in reachability games. Any pure strategy of $\mathcal{P}_2$ that only selects actions that do not lead to the winning region of $\mathcal{P}_1$ whenever it can be avoided are uniformly winning strategies of $\mathcal{P}_2$.*

We now move on to zero-sum Büchi games. Let $\mathcal{G} = (\mathcal{A}, \mathsf{Büchi}(T))$ denote the zero-sum Büchi game on $\mathcal{A}$ where the target of $\mathcal{P}_2$ is T. Like reachability games, Büchi games also enjoy memoryless determinacy. This can be seen as a corollary of the memoryless determinacy of parity games [EJ88], a class of objectives subsuming Büchi and co-Büchi objectives.

Theorem 6.2. *Büchi games on deterministic arenas are determined and both players have pure memoryless uniformly winning strategies.*

6.1.2 Shortest-path games

We now study zero-sum shortest-path games on $\mathcal{A}$ in which weights are non-negative integers. Let $w\colon S \times A \to \mathbb{N}$ be a weight function and $\mathcal{G} = (\mathcal{A}, \mathsf{SPath}_w^T)$ be a zero-sum shortest-path game on $\mathcal{A}$. Recall that the goal of $\mathcal{P}_1$ is to *minimise* the shortest-path cost function.

Shortest-path games are determined (see, e.g., [BGHM17] for finite arenas). We provide a direct argument using the determinacy of games with open objectives [GS53]: in a game on a deterministic arena with an open objective, from all initial states, one of the players has a pure (surely) winning strategy. An objective is open if it can be written as a union of cylinder sets. A by-product of our argument is that $\mathcal{P}_1$ has a pure optimal strategy from any state.

Lemma 6.3. *The game $\mathcal{G}$ is determined, $\mathcal{P}_1$ has a pure optimal strategy from all states and $\mathcal{P}_2$ has a pure optimal strategy from all states with a finite value.*

Proof. Let $\theta \in \mathbb{R}$. We let $\{\mathsf{SPath}_w^T \leq \theta\} = \{\pi \in \mathsf{Plays}(\mathcal{A}) \mid \mathsf{SPath}_w^T(\pi) \leq \theta\}$. The objective $\{\mathsf{SPath}_w^T \leq \theta\}$ is open: it is the union of the cylinder of histories of weight no more than θ ending in a state of T. Therefore, in the zero-sum game $\mathcal{G}_\theta = (\mathcal{A}, \{\mathsf{SPath}_w^T \leq \theta\})$, for all $s \in S$, either $\mathcal{P}_1$ or $\mathcal{P}_2$ has a pure surely winning strategy from s [GS53].

Let $s \in S$. For all $\theta \in \mathbb{R}$, if $\mathcal{P}_1$ wins from s in $\mathcal{G}_\theta$, then $\mathcal{P}_1$ can ensure (a cost of at most) θ in $\mathcal{G}$, whereas if $\mathcal{P}_2$ wins from s in $\mathcal{G}_\theta$, then $\mathcal{P}_2$ can ensure (a cost of at least) θ in $\mathcal{G}$.

First, assume that for all $\theta \in \mathbb{N}$, $\mathcal{P}_2$ wins from s in $\mathcal{G}_\theta$. We conclude that $\mathsf{Val}_{\mathcal{G}}(s) = +\infty$. All strategies of $\mathcal{P}_1$ are optimal (because all plays have a cost at most $+\infty$), and thus $\mathcal{P}_1$ has a pure optimal strategy from s.

Assume now that there exists some $\theta \in \mathbb{N}$ such that $\mathcal{P}_1$ wins in $\mathcal{G}_\theta$ from s and let $\theta^\star$ denote the minimum of all such $\theta \in \mathbb{N}$. We claim that $\mathsf{Val}_{\mathcal{G}}(s) = \theta^\star$. First, we observe that $\mathcal{P}_2$ wins in $\mathcal{G}_{\theta^\star - \frac{1}{2}}$ from s and thus $\mathcal{P}_2$ has a pure strategy ensuring a cost of at least $\theta^\star - \frac{1}{2}$ from s in $\mathcal{G}$. Indeed, if $\mathcal{P}_1$ has a winning strategy in $\mathcal{G}_{\theta^\star - \frac{1}{2}}$, then $\mathcal{P}_1$ wins in $\mathcal{G}_{\theta^\star - 1}$ because $\mathsf{SPath}_w^T \colon \mathsf{Plays}(\mathcal{A}) \to \bar{\mathbb{N}}$. For the same reason, it follows that $\mathcal{P}_2$ can ensure a cost of at least $\theta^\star$ from s in $\mathcal{G}$. We conclude that $\mathsf{Val}_{\mathcal{G}}(s) = \theta^\star$. Any pure winning strategy of $\mathcal{P}_1$ from s in $\mathcal{G}_{\mathsf{Val}_{\mathcal{G}}(s)}$ is optimal in $\mathcal{G}$ from s and, similarly, any pure winning strategy of $\mathcal{P}_2$ from s in $\mathcal{G}_{\mathsf{Val}_{\mathcal{G}}(s) - \frac{1}{2}}$ is optimal in $\mathcal{G}$ from s. □

We now refine the result on optimal strategies of $\mathcal{P}_1$ proven above: $\mathcal{P}_1$ has a memoryless uniformly optimal strategy (even if $\mathcal{A}$ is infinite). To prove this, we follow the following steps. First, we argue the existence of optimal strategies for $\mathcal{P}_1$ regardless of memory and uniformity. Second, we construct a shortest-path game by removing transitions from $\mathcal{A}$ without introducing deadlocks. We then show that values coincide in this new game and the original game. Finally, we establish that memoryless uniformly winning reachability strategies of the new game are optimal in both this new game and the original shortest-path game.

Theorem 6.4. *In $\mathcal{G}$, $\mathcal{P}_1$ has a pure uniformly optimal memoryless strategy that is uniformly winning in the reachability game* $(\mathcal{A}, \mathsf{Reach}(T))$.

Proof. By Lemma 6.3, $\mathsf{Val}_{\mathcal{G}}(s)$ is well-defined for all $s \in S$.

First, we show that for all $s \in S_1 \setminus T$, there exists $a \in A(s)$ such that $\mathsf{Val}_{\mathcal{G}}(s) = \mathsf{Val}_{\mathcal{G}}(\delta(s, a)) + w(s, a)$. Let $s \in S_1 \setminus T$. If $\mathcal{P}_1$ uses $a \in A$ in s, $\mathcal{P}_1$ can ensure $\mathsf{Val}_{\mathcal{G}}(\delta(s, a)) + w(s, a)$ at best. It follows that $\mathsf{Val}_{\mathcal{G}}(s) = \min\{\mathsf{Val}_{\mathcal{G}}(\delta(s, a)) + w(s, a)) \mid a \in A(s)\}$ (this minimum is well-defined because $\bar{\mathbb{N}}$ is well-ordered). This implies the claim.

Second, we claim that for all $s \in S_2 \setminus T$ and all $a \in A(s)$, we have $\mathsf{Val}_{\mathcal{G}}(s) \geq \mathsf{Val}_{\mathcal{G}}(\delta(s, a)) + w(s, a)$. Let $s \in S_2 \setminus T$ and $a \in A(s)$. Let $s' = \delta(s, a)$. We first assume that $\mathsf{Val}_{\mathcal{G}}(s')$ is finite. Then, $\mathcal{P}_2$ has an optimal strategy from s' by Lemma 6.3. It follows that $\mathcal{P}_2$ can ensure $\mathsf{Val}_{\mathcal{G}}(s') + w(s, a)$ from s by playing action a in s and playing optimally from s', which implies the desired inequality. Assume now that $\mathsf{Val}_{\mathcal{G}}(s')$ is infinite. Then for all $\theta \in \mathbb{N}$, $\mathcal{P}_2$ has a strategy ensuring θ from s'. We conclude, similarly to the previous case, that $\mathsf{Val}_{\mathcal{G}}(s)$ is infinite and therefore satisfies the inequality.

Third, we remove transitions of $\mathcal{A}$ to derive a game $\mathcal{G}'$ in which values are unchanged with respect to $\mathcal{G}$. Intuitively, we remove transitions from states of $\mathcal{P}_1$ that cannot be used by an optimal strategy. Let $\mathcal{A}' = (S_1, S_2, A, \delta')$ denote the arena where δ' is the restriction of δ over the union of $S_2 \times A$ and

$$\{(s, a) \in S_1 \times A \mid a \in A(s) \text{ and } \mathsf{Val}_{\mathcal{G}}(s) = \mathsf{Val}_{\mathcal{G}}(\delta(s, a)) + w(s, a)\}\,.$$

By the first point above, $\mathcal{A}'$ does not have any deadlocks.

We let $\mathcal{G}' = (\mathcal{A}', \mathsf{SPath}_w^T)$. We claim that

(i) for all $s \in S$, $\mathsf{Val}_{\mathcal{G}}(s) = \mathsf{Val}_{\mathcal{G}'}(s)$ and

(ii) the winning regions in the reachability games $(\mathcal{A}, \mathsf{Reach}(T))$ and $(\mathcal{A}', \mathsf{Reach}(T))$ coincide.

For (i), we observe that for all $s \in S$, $\mathsf{Val}_{\mathcal{G}'}(s) \geq \mathsf{Val}_{\mathcal{G}}(s)$ by construction of $\mathcal{A}'$: $\mathcal{P}_1$ has fewer strategies than in $\mathcal{A}$, but no actions of $\mathcal{P}_2$ have been removed. In particular, if $\mathsf{Val}_{\mathcal{G}}(s) = +\infty$, then $\mathsf{Val}_{\mathcal{G}'}(s) = +\infty$. We show the other inequality of (i) by induction on $\mathsf{Val}_{\mathcal{G}}(s)$ for states of finite value. For the base case, let $s \in S$ such that $\mathsf{Val}_{\mathcal{G}}(s) = 0$. In this case, an optimal strategy of $\mathcal{P}_1$ from s (which exists by Lemma 6.3) surely reaches T by only using zero-weight transitions and that only traversing states with zero-value in $\mathcal{G}$ until T regardless of the choices of $\mathcal{P}_2$. It follows that this same strategy can be used to ensure a cost of 0 in $\mathcal{G}'$. This ends the argument for the base case.

We now assume by induction that for all $\beta \leq \theta$ and all $s \in S$, if $\mathsf{Val}_{\mathcal{G}}(s) = \beta$, then $\mathsf{Val}_{\mathcal{G}'}(s) = \beta$. Let $s \in S$ such that $\mathsf{Val}_{\mathcal{G}}(s) = \theta + 1$ and let us show that $\mathsf{Val}_{\mathcal{G}'}(s) = \theta + 1$. To this end, we construct an optimal strategy from s in $\mathcal{G}'$ as follows.

Fix a pure strategy σ_1 of $\mathcal{P}_1$ in $\mathcal{A}$ that is optimal from s. We consider the strategy σ_1' of $\mathcal{P}_1$ in $\mathcal{A}'$ that plays consistently with σ_1 until a state s' with $\mathsf{Val}_{\mathcal{G}}(s') < \mathsf{Val}_{\mathcal{G}}(s)$ is reached, and then plays accordingly to an optimal strategy from s' in $\mathcal{G}'$. Under the assumption that σ_1' is well-defined, it ensures $\mathsf{Val}_{\mathcal{G}}(s)$ from s in $\mathcal{G}'$ by construction.

It remains to show that σ_1' only uses actions that are available in $\mathcal{A}'$. We prove this by contradiction. Assume that there exists a history $h \in \mathsf{Hist}(\mathcal{A})$ consistent with σ_1' starting in s such that $\sigma_1'(h)$ is an action that is not enabled in $\mathsf{last}(h)$ in $\mathcal{A}'$. By choosing h of minimal length, we obtain that $h \in \mathsf{Hist}(\mathcal{A}')$. Since σ_1' switches to a strategy of $\mathcal{A}'$ once a state with value strictly less than $\mathsf{Val}_{\mathcal{G}}(s)$ is reached, all states of h have value equal to $\mathsf{Val}_{\mathcal{G}}(s)$ and all transitions in h have weight zero. Furthermore, h is consistent with σ_1 and $\sigma_1'(h) = \sigma_1(h)$. By definition of transitions in $\mathcal{A}'$, we have

$$w(\mathsf{last}(h), \sigma_1(h)) + \mathsf{Val}_{\mathcal{G}}(\delta(\mathsf{last}(h), \sigma_1(h))) > \mathsf{Val}_{\mathcal{G}}(\mathsf{last}(h)) = \mathsf{Val}_{\mathcal{G}}(s).$$

We now extend h to construct an outcome of σ_1 with a cost exceeding $\mathsf{Val}_{\mathcal{G}}(s)$, which contradicts the optimality of σ_1 from s. We extend h by choosing the actions of $\mathcal{P}_2$ according to strategies that ensure some threshold from $\mathsf{last}(h)$. If $\mathsf{Val}_{\mathcal{G}}(\delta(\mathsf{last}(h), \sigma_1(h))) \in \mathbb{N}$, we extend h by relying on a strategy of $\mathcal{P}_2$ that is optimal from $\delta(\mathsf{last}(h), \sigma_1(h))$ (which exists by (see Lemma 6.3). Otherwise, if $\mathsf{Val}_{\mathcal{G}}(\delta(\mathsf{last}(h), \sigma_1(h)))$ is infinite, we extend h by using a strategy of $\mathcal{P}_2$ that ensures a cost strictly greater than $\mathsf{Val}_{\mathcal{G}}(s)$ from $\delta(\mathsf{last}(h), \sigma_1(h))$. Through this scheme, we obtain an outcome of σ_1 from s with cost greater than $\mathsf{Val}_{\mathcal{G}}(s)$, which shows that σ_1 is not optimal from s. This ends the proof of (i).

We now prove that (ii) holds. Clearly any state that is winning in

$(\mathcal{A}', \mathsf{Reach}(T))$ also is in $(\mathcal{A}, \mathsf{Reach}(T))$. Conversely, fix a state s that is winning for $\mathcal{P}_1$ in $(\mathcal{A}, \mathsf{Reach}(T))$. If its value is finite, the claim follows from (i). Therefore, assume that $\mathsf{Val}_{\mathcal{G}}(s) = +\infty$. Let σ_1 be a winning strategy of $\mathcal{P}_1$ from s in $(\mathcal{A}, \mathsf{Reach}(T))$. Its behaviours in states of infinite value need not be restricted to obtain a strategy of $\mathcal{A}'$, as the outgoing edges from $\mathcal{P}_1$ states of infinite value are the same in $\mathcal{A}$ and $\mathcal{A}'$. Furthermore, all outcomes of σ_1 eventually reach a state of finite value. By changing σ_1 so it conforms to a strategy optimal in $\mathcal{G}'$ from the earliest such visited state, we obtain a strategy σ_1' that is winning from s in $(\mathcal{A}', \mathsf{Reach}(T))$. This ends the proof of (ii).

Finally, we prove the claim of the theorem. Let σ_1 be a memoryless uniformly winning reachability strategy in the reachability game $(\mathcal{A}', \mathsf{Reach}(T))$. It follows from (ii) that σ_1 is also a uniformly winning reachability strategy in $(\mathcal{A}, \mathsf{Reach}(T))$. It remains to show that σ_1 is optimal from all states with finite value. Let $s_0 \in S$ such that $\mathsf{Val}_{\mathcal{G}}(s_0)$ is finite. Let $\pi = s_0 a_0 s_1 \dots$ be consistent with σ_1. We prove that $\mathsf{SPath}_w^T(\pi) \leq \mathsf{Val}_{\mathcal{G}}(s_0)$. Let $r = \min\{\ell \in \mathbb{N} \mid s_\ell \in T\}$ (it exists by (ii) and the fact that $\mathsf{Val}_{\mathcal{G}}(s_0)$ is finite). We have, by choice of E' and the third claim above, that

$$\begin{aligned}
\mathsf{SPath}_w^T(\pi) &= w(\pi_{\leq r}) \\
&= \sum_{\ell=0}^{r-1} w(s_\ell, a_\ell) \\
&\leq \sum_{\ell=0}^{r-1} \mathsf{Val}_{\mathcal{G}}(s_\ell) - \mathsf{Val}_{\mathcal{G}}(s_{\ell+1}) \\
&= \mathsf{Val}_{\mathcal{G}}(s_0),
\end{aligned}$$

because $\mathsf{Val}_{\mathcal{G}}(s_r) = 0$. This shows that σ_1 is optimal from s and ends the proof. □

On the other hand, we can show that $\mathcal{P}_2$ does not necessarily have a pure optimal strategy from states with an infinite value in an infinitely branching shortest-path game. Furthermore, although there do exist pure optimal strategies of $\mathcal{P}_2$ from states with a finite value, there need not exist memoryless pure strategies of $\mathcal{P}_2$ that are optimal from all states with finite value. In other words, intuitively, it is not possible to have a strategy that is uniformly optimal

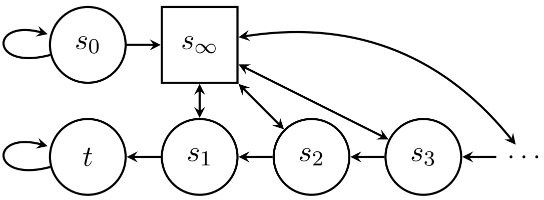


Figure 6.1: An infinite-state turn-based deterministic arena. The action labelling an edge (s, s') is the outgoing state s'; we omit actions from the figure to lighten it. Circles and squares respectively denote states of $\mathcal{P}_1$ and $\mathcal{P}_2$.

over all finite-value states.

Example 6.1. We consider the two-player countable-state arena $\mathcal{A}$ depicted in Figure 6.1 and the two-player zero-sum game $\mathcal{G} = (\mathcal{A}, \mathsf{SPath}_1^{\{t\}})$ where 1 denotes the constant weight function assigning 1 to all pairs in $S \times A$.

Let $\theta \in \mathbb{N}_{>0}$. We have $\mathsf{Val}_{\mathcal{G}}(s_\theta) = \theta$. On the one hand, $\mathcal{P}_1$ can ensure a cost of no more than θ from s_θ by moving leftward in the illustration. On the other hand, $\mathcal{P}_2$ can ensure a cost of at least θ from s_θ with the memoryless strategy that moves from s_∞ to s_θ. It follows that this same memoryless strategy of $\mathcal{P}_2$ ensures $\theta + 1$ from s_∞. We conclude that $\mathsf{Val}_{\mathcal{G}}(s_\infty) = +\infty$.

However, $\mathcal{P}_2$ does not have an optimal pure strategy from s_∞. Consider a pure strategy σ_2 of $\mathcal{P}_2$. If σ_2 moves from s_∞ to s_θ in the first round, then $\mathcal{P}_2$ cannot ensure a cost higher than $\theta + 1$ from s_∞. Therefore, $\mathcal{P}_2$ does not have a pure optimal strategy from s_∞.

We now show that $\mathcal{P}_2$ does not have a pure memoryless strategy that is optimal from all finite-value state in $\mathcal{G}$. Consider the pure memoryless strategy σ_2 of $\mathcal{P}_2$ such that $\sigma_2(s_\infty) = s_\theta$ for some $\theta \in \mathbb{N}_{>0}$. This strategy ensures, at best, a cost of $\theta + 2$ from the state $s_{\theta+3}$; if $\mathcal{P}_1$ moves from $s_{\theta+3}$ to s_∞, then moves leftwards from s_θ, the cost of the resulting outcome is $\theta + 2 < \mathsf{Val}_{\mathcal{G}}(s_{\theta+3})$. Therefore, there is no memoryless strategy of $\mathcal{P}_2$ in this game that ensures, from all finite-value states, their value. $\lhd$

To implement the punishment mechanism to construct finite-memory NEs, we need punishing strategies that are effective regardless of the state from which the punishment starts. Example 6.1 shows that we cannot do this with (uniform) optimal strategies. We establish a weaker, albeit sufficient property of $\mathcal{G}$: there exists a family of $\mathcal{P}_2$ memoryless strategies $(\sigma_2^\theta)_{\theta\in\mathbb{N}}$ such that, for all $\theta \in \mathbb{N}$, σ_2^θ is winning from any state in the winning region $W_2(\mathsf{Safe}(T))$ of $\mathcal{P}_2$ in the reachability game $(\mathcal{A}, \mathsf{Reach}(T))$ and ensures the minimum of θ and the value of the state from any other state. Intuitively, the parameter θ quantifies by how much $\mathcal{P}_1$ should be sabotaged (uniformly).

Let $\theta \in \mathbb{N}$. The construction of σ_2^θ can be sketched as follows. On $W_2(\mathsf{Safe}(T))$, we let σ_2^θ coincide with a uniformly winning memoryless strategy of $\mathcal{P}_2$ in $(\mathcal{A}, \mathsf{Reach}(T))$. Outside of $W_2(\mathsf{Safe}(T))$, σ_2^θ selects successors such that the sum of the edge weight and the value of the successor state is maximum if there is one such maximum, and otherwise, selects a successor such that this sum is at least θ (which exists because all such sums are in $\bar{\mathbb{N}}$). This definition of σ_2^θ yields the desired properties.

Theorem 6.5. *Let $s \in W_2(\mathsf{Safe}(T))$ denote the winning region of $\mathcal{P}_2$ in the reachability game $(\mathcal{A}, \mathsf{Reach}(T))$. For all $\theta \in \mathbb{N}$, there exists a memoryless strategy σ_2^θ of $\mathcal{P}_2$ such that, for all $s \in S$:*

(i) σ_2^θ is winning from s for $\mathcal{P}_2$ in $(\mathcal{A}, \mathsf{Reach}(T))$ if $s \in W_2(\mathsf{Safe}(T))$ and

(ii) σ_2^θ ensures a cost of at least $\min\{\mathsf{Val}_{\mathcal{G}}(s), \theta\}$ from s.

Proof. In the following, we extend w to histories by letting $w(h) = \sum_{\ell=1}^{r-1} w(s_\ell, a_\ell)$ for all $h = s_0a_0s_1 \dots a_{r-1}s_r \in \mathsf{Hist}(\mathcal{A})$.

Let $\sigma_2^{\mathsf{Safe}(T)}$ be a memoryless uniformly winning strategy of $\mathcal{P}_2$ in $(\mathcal{A}, \mathsf{Reach}(T))$ (cf. Theorem 6.1). For $s \in S_2$, we let $\sigma_2^\theta(s) = \sigma_2^{\mathsf{Safe}(T)}(s)$ if $s \in W_2(\mathsf{Safe}(T))$, otherwise, if $\max_{a\in A(s)} w(s, a) + \mathsf{Val}_{\mathcal{G}}(\delta(s, a))$ is defined, we let $\sigma_2^\theta(s)$ be an action achieving this maximum, and, otherwise, we let $\sigma_2^\theta(s) = a$ where $a \in A(s)$ is such that $w(s, a) + \mathsf{Val}_{\mathcal{G}}(\delta(s, a)) \geq \theta$.

We prove that σ_2^θ satisfies the claimed properties. First, we observe that any play starting in $W_2(\mathsf{Safe}(T))$ consistent with $\sigma_2^{\mathsf{Safe}(T)}$ never leaves $W_2(\mathsf{Safe}(T))$. Property (i) follows. To establish (ii), we show the following property: for any history $h = s_0a_0s_1a_1 \dots s_r$ that is consistent with σ_2^θ such that for all $\ell < r$, $s_\ell \notin T$, it holds that $w(h) + \min\{\mathsf{Val}_{\mathcal{G}}(s_r), \theta\} \geq \min\{\mathsf{Val}_{\mathcal{G}}(s_0), \theta\}$.

We proceed by induction on the length of histories. For a history of the form $h = s_0$, the property is immediate. We now consider a suitable history

$h = h = s_0a_0s_1a_1 \dots s_r$ and assume the property holds for $h' = h_{\leq r-1}$ by induction (note that h' is of the suitable form as well). We discuss two cases depending on whether $\mathsf{Val}_{\mathcal{G}}(s_{r-1})$ is finite and split each case depending on whom controls $\mathsf{last}(h') = s_{r-1}$.

We first assume that $\mathsf{Val}_{\mathcal{G}}(s_{r-1})$ is finite. Both players have pure optimal strategies from any state s with finite value (Lemma 6.3). We observe that

$$\begin{aligned} w(h) + \min\{\mathsf{Val}_{\mathcal{G}}(s_r), \theta\} &= w(h') + w(s_{r-1}, a_{r-1}) + \min\{\mathsf{Val}_{\mathcal{G}}(s_r), \theta\} \\ &\geq w(h') + \min\{\mathsf{Val}_{\mathcal{G}}(s_r) + w(s_{r-1}, a_{r-1}), \theta\}. \end{aligned}$$

To conclude by induction, it suffices to show that $\mathsf{Val}_{\mathcal{G}}(s_r) + w(s_{r-1}, a_{r-1}) \geq \mathsf{Val}_{\mathcal{G}}(s_{r-1})$. If $s_{r-1} \in S_1$, $\mathcal{P}_1$ can ensure a cost of at most $\mathsf{Val}_{\mathcal{G}}(s_r) + w(s_{r-1}, a_{r-1})$ from s_{r-1} by playing action a_{r-1} in s_{r-1} and then playing optimally from s_r, yielding the desired inequality. Assume now that $s_{r-1} \in S_2$. For all $a \in A(s_{r-1})$, it holds that if $\mathcal{P}_2$ can ensure $\beta \in \mathbb{N}$ from $\delta(s_{r-1}, a)$, then $\mathcal{P}_2$ can ensure $\beta + w(\delta(s_{r-1}, a))$ from s_{r-1}. It follows that

$$\mathsf{Val}_{\mathcal{G}}(s_{r-1}) = \sup_{a \in A(s_{r-1})} \mathsf{Val}_{\mathcal{G}}(\delta(s_{r-1}, a)) + w(s_{r-1}, a).$$

Because $\mathsf{Val}_{\mathcal{G}}(s_{r-1})$ is finite and values are in $\bar{\mathbb{N}}$, we obtain that $\sigma_2^\theta(s_{r-1})$ is an action witnessing that the above supremum is a maximum, and obtain that $\mathsf{Val}_{\mathcal{G}}(s_{r-1}) = \mathsf{Val}_{\mathcal{G}}(\delta(s_{r-1}, \sigma_2^\theta(s_{r-1}))) + w(s_{r-1}, \sigma_2^\theta(s_{r-1}))$. This ends the proof of this case.

Assume now that $\mathsf{Val}_{\mathcal{G}}(s_{r-1}) = +\infty$. We first consider the case $s_{r-1} \in S_1$. Then all successors of s_{r-1} have an infinite value, otherwise $\mathcal{P}_1$ could ensure a finite cost from s_{r-1} by moving to a successor with finite value and playing optimally from there. We must therefore show that $w(h) + \theta \geq \min\{\mathsf{Val}_{\mathcal{G}}(s_0), \theta\}$. This follows directly from the induction hypothesis $w(h') + \theta \geq \min\{\mathsf{Val}_{\mathcal{G}}(s_0), \theta\}$ and the inequality $w(h) \geq w(h')$.

Next, we assume that $s_{r-1} \in S_2$. It follows from $\mathsf{Val}_{\mathcal{G}}(s_{r-1}) = +\infty$ and the definition of σ_2^θ that $\mathsf{Val}_{\mathcal{G}}(s_r) + w(s_{\ell-1}, a_{\ell-1}) \geq \theta \geq \min\{\mathsf{Val}_{\mathcal{G}}(s_0), \theta\}$. The desired inequality follows from $w(h) \geq w(s_{\ell-1}, a_{\ell-1})$, ending the induction proof.

Let $s \in S \setminus W_2(\mathsf{Safe}(T))$. We now use the previous property to conclude that σ_2^θ ensures $\min\{\mathsf{Val}_{\mathcal{G}}(s), \theta\}$ from s. Let π be a play consistent with σ_2^θ

starting in s. If π does not visit T, then $\mathsf{SPath}_w^T(\pi) = +\infty$. Otherwise, let h be the prefix of π up to the first state in T included. Then, we have $\mathsf{SPath}_w^T(\pi) = w(h) \geq \min\{\mathsf{Val}_{\mathcal{G}}(s), \theta\}$ by the previous property. This shows that σ_2^θ ensures $\min\{\mathsf{Val}_{\mathcal{G}}(s), \theta\}$ from s, ending the proof. □

Remark 6.6 (Optimal strategies for $\mathcal{P}_2$). The proof above suggests that if $A(s)$ is a finite set for all $s \in S_2$, then $\mathcal{P}_2$ has a memoryless uniformly optimal strategy for the SPath_w^T cost function. We formalise this below.

Assume that $A(s)$ is a finite set for all $s \in S_2$. In this case, the definition of σ_2^θ is independent of θ. Let $\sigma_2 = \sigma_2^0$ and let us show that σ_2 is optimal from all states.

Let $s \in S$. It follows from the proof above that σ_2 is optimal from all states with finite value and all states in $W_2(\mathsf{Safe}(T))$. We therefore assume that $\mathsf{Val}_{\mathcal{G}}(s) = +\infty$. It suffices to show that $s \in W_2(\mathsf{Safe}(T))$. We proceed by contradiction and assume that s is in the winning region of $\mathcal{P}_1$ in the reachability game $(\mathcal{A}, \mathsf{Reach}(T))$. We argue that $\mathsf{Val}_{\mathcal{G}}(s)$ is finite. Fix a strategy σ_1 that is winning from s for $\mathcal{P}_1$ in the reachability game $(\mathcal{A}, \mathsf{Reach}(T))$. All plays starting in s_0 that are consistent with σ_1 eventually reach T. The set of their prefixes up to the first occurrence of a state of T is a finitely branching tree: branching occurs only when $\mathcal{P}_2$ selects an action. If s has an infinite value, i.e., there are histories in the tree with arbitrarily large weight, then the tree must be infinite. By König's lemma [Kön27], there must be an infinite branch in this tree, i.e., a play consistent with σ_1 that does not visit T. This contradicts the assumption that σ_1 is winning from s. Therefore, $\mathsf{Val}_{\mathcal{G}}(s)$ must be finite. This is a contradiction with $\mathsf{Val}_{\mathcal{G}}(s) = +\infty$, which yields the desired result. ◁

6.2 Characterising Nash equilibria outcomes

We provide characterisations of plays that are outcomes of Nash equilibria in reachability, Büchi and shortest-path games. These characterisations refer to zero-sum games with the same objective or cost function. Intuitively, a play is an NE outcome if and only if the cost incurred by a player from a state of the play is no more than the value of said state in the zero-sum game where the state owner plays against the others. In other words, there is a profitable deviation if and only if there is a profitable deviation when the other players are adversaries.

We fix a turn-based deterministic arena $\mathcal{A} = ((S_i)_{i\in[\![1,n]\!]}, A, \delta)$ and targets $T_1, \dots, T_i \subseteq S$ for the remainder of this section.

In the following, we consider values in so-called *coalition games*. For all $i \in [\![1, n]\!]$, we let $\mathcal{A}_i$ denote the two-player arena $(S_i, S \setminus S_i, A, \delta)$ where all players other than $\mathcal{P}_i$ are grouped in a coalition.

Definition 6.7. Given a game $\mathcal{G} = (\mathcal{A}, (f_i)_{i\in[\![1,n]\!]})$ and $i \in [\![1, n]\!]$, we let $\mathcal{G}_i = (\mathcal{A}_i, f_i)$ be the zero-sum game where all players coordinate against $\mathcal{P}_i$ as a coalition. We call $\mathcal{G}_i$ a *coalition game* (against $\mathcal{P}_i$).

We provide a characterisation for NE outcomes based on values in coalition games for reachability and Büchi games in Section 6.2.1 and a characterisation for shortest-path games in Section 6.2.2.

6.2.1 Reachability and Büchi games

We present a characterisation of NE outcomes in games where all players have either a reachability objective or a Büchi objective. In this section, we allow games where players may have objectives of different types, to provide a uniform characterisation for the NE outcomes of the considered objectives.

We consider the game $\mathcal{G} = (\mathcal{A}, (\Omega_i)_{i\in[\![1,n]\!]})$ where, for all $i \in [\![1, n]\!]$, we have $\Omega_i \in \{\mathsf{Reach}(T_i), \mathsf{Büchi}(T_i)\}$. Let $W_i(\Omega_i)$ be the winning region of the first player of the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \Omega_i)$, in which $\mathcal{P}_i$ is opposed to the other players. Let $\pi = s_0a_0s_1a_1 \ldots \in \mathsf{Plays}(\mathcal{A})$ be a play. Then π is an outcome of an NE from s_0 if and only if, for all $i \in [\![1, n]\!]$ such that $\pi \notin \Omega_i$, all states in π are in the complement of $W_i(\Omega_i)$.

On the one hand, if there is $i \in [\![1, n]\!]$ such that the objective of $\mathcal{P}_i$ is not satisfied and $W_i(\Omega_i)$ is visited along π, then $\mathcal{P}_i$ has a profitable deviation by switching to a (memoryless uniform) winning strategy in $\mathcal{G}_i$ once $W_i(\Omega_i)$ is reached. Conversely, one constructs a Nash equilibrium as follows. The players follow the play π, and, if $\mathcal{P}_i$ deviates from π, then all other players conform to a (memoryless uniformly) winning strategy for the second player in $\mathcal{G}_i$ for the complement of Ω_i (i.e., $\mathsf{Safe}(T_i)$ if Ωi is a reachability objective or $\mathsf{coBüchi}(T_i)$

if Ω_i is a Büchi objective). This ensures no player has a profitable deviation.

We formally state and prove this characterisation below. The following characterisation is similar to the characterisation of NE outcomes in finite arenas of [CFGR16], where it is assumed that all players have objectives of the same type. We provide a formal proof below for the sake of completeness. In the following argument, we exploit the *prefix-independence* of the Büchi objective: adding or removing a prefix to a play does not change whether the Büchi objective is satisfied or not.

Theorem 6.8. *Let $\mathcal{G} = (\mathcal{A}, (\Omega_i)_{i\in[\![1,n]\!]})$ be a game where, for all $i \in [\![1, n]\!]$, $\Omega_i \in \{\mathsf{Reach}(T_i), \mathsf{Büchi}(T_i)\}$. Let $W_i(\Omega_i)$ denote the winning region of the first player of the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \Omega_i)$. Let $\pi = s_0a_0s_1a_1\dots$ be a play. Then π is the outcome of an NE from s_0 if and only if, for all $i \in [\![1, n]\!]$ such that $\pi \notin \Omega_i$, $s_\ell \notin W_i(\Omega_i)$ for all $\ell \in \mathbb{N}$.*

Proof. First, assume that there exists some $i \in [\![1, n]\!]$ such that $\pi \notin \Omega_i$ and there exists some $\ell \in \mathbb{N}$ such that $s_\ell \in W_i(\Omega_i)$. We claim that for all strategy profiles σ such that $\pi = \mathsf{Out}_{\mathcal{A}}(\sigma, s_0)$, $\mathcal{P}_i$ has a profitable deviation with respect to σ from s_0 (i.e., π is not the outcome of an NE). We consider a pure strategy τ_i of $\mathcal{P}_i$ that agrees with σ_i on strict prefixes of $\pi_{\leq\ell}$ and otherwise agrees with a memoryless uniformly winning strategy of the coalition game $\mathcal{G}_i$. The play $\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_0)$ is the concatenation of $\pi_{\leq\ell}$ and a play π' starting in s_ℓ that is consistent with τ_i. It follows from $s \in W_i(\Omega_i)$ that $\pi' \in \Omega_i$. Because Ω_i is either a reachability objective or is prefix-independent, it follows that $\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_0) \in \Omega_i$. We have shown that τ_i is a profitable deviation with respect to σ from s_0.

We now prove the converse. Assume that for all $i \in [\![1, n]\!]$ such that $\pi \notin \Omega_i$, $s_\ell \notin W_i(\Omega_i)$ for all $\ell \in \mathbb{N}$. We formalise the NE suggested prior to the proof. For all $i \in [\![1, n]\!]$, we fix a memoryless uniformly winning strategy τ_{-i} of the second player in the coalition game $\mathcal{G}_i$. Let $i \in [\![1, n]\!]$. We define σ_i as follows. We let σ_i be arbitrary over $\mathsf{Hist}(\mathcal{A}) \setminus \mathsf{Hist}(\mathcal{A}, s_0)$ (we are only concerned with plays starting in s_0). Let $h \in \mathsf{Hist}(\mathcal{A}, s_0)$. If $h = \pi_{\leq\ell}$ for some $\ell \in \mathbb{N}$ and $\mathsf{last}(h) \in S_i$, we let $\sigma_i(h) = a_\ell$. We now assume that h is not a prefix of π. Let h' be the longest common prefix of π and h; h' is necessarily a history because transitions are deterministic. Let $i' \in [\![1, n]\!]$ such that $\mathsf{last}(h') \in S_{i'}$. If $i' = i$, we let $\sigma_i(h')$ be arbitrary; this is the case in which $\mathcal{P}_i$ has deviated from π. Otherwise, we let $\sigma_i(h) = \tau_{-i}(\mathsf{last}(h))$.

We let $\sigma = (\sigma_i)_{i\in[\![1,n]\!]}$. It is easy to see that $\mathsf{Out}_{\mathcal{A}}(\sigma, s_0) = \pi$. It remains to establish that σ is an NE from s_0. Let $i \in [\![1, n]\!]$. If $\pi \in \Omega_i$, i.e., if the objective of $\mathcal{P}_i$ is satisfied, then $\mathcal{P}_i$ has no profitable deviation. We now assume that $\pi \notin \Omega_i$. Let τ_i be a strategy of $\mathcal{P}_i$. We must show that $\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_0)$ does not satisfy Ω_i. If $\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_0) = \pi$, there is nothing to show. We assume the contrary. It follows that $\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_0)$ can be written as the concatenation of a prefix of π and of a play π' consistent with τ_{-i} by definition of σ. Since all states of π are outside of $W_i(\Omega_i)$, π' is not in Ω_i. If Ω_i is a reachability objective, there are no visits to T_i in $\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_0)$ because there are none in π nor in π'. Otherwise, we obtain that $\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_0) \notin \Omega_i$ by prefix-independence.

We have thus shown that $\mathcal{P}_i$ does not have a profitable deviation, i.e., σ is an NE from s_0 in $\mathcal{G}$. □

6.2.2 Shortest-path games

We now provide a characterisation for NE outcomes in shortest-path games. We state this result for games in which each player has their own weight function. For all $i \in [\![1, n]\!]$, let $w_i \colon E \to \mathbb{N}$ be a weight function for $\mathcal{P}_i$. We consider the shortest-path game $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}_{w_i}^{T_i})_{i\in[\![1,n]\!]})$. For any $s \in S$, we denote by $\mathsf{Val}_{\mathcal{G}}^{i}(s)$ the value of s in the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \mathsf{SPath}_{w_i}^{T_i})$. As in the previous section, we let $W_i(\mathsf{Reach}(T_i))$ denote the winning region of the first player of the coalition reachability game $(\mathcal{A}_i, \mathsf{Reach}(T_i))$.

Theorem 6.8 asserts that the value (i.e., whether a player wins) is sufficient to characterise NE outcomes in reachability games. A natural generalisation of this characterisation in shortest-path games would be to impose, for each player, a constraint on the values of all suffixes of the play up to a target of the player (or for all states of the play if no target appears), stating that the cost of the suffix is preferable to the value of its first state. This matches an existing characterisation in finite arenas [BBGT21, Thm. 15]. However, we argue that this is not sufficient in infinite arenas with the following example.

Example 6.2. Let us consider the arena depicted in Figure 6.1 (Page 112) and let $T_1 = \{t\}$ and $T_2 = \{s_0\}$. It follows from $\mathsf{Val}^1_{\mathcal{G}}(s_\infty) = +\infty$ that $\mathsf{Val}^1_{\mathcal{G}}(s_0) = +\infty$ (the former equality is shown in Example 6.1). Therefore, the cost of all suffixes of the play s_0^ω for $\mathcal{P}_1$ matches the value of their first state s_0. However, for any strategy profile resulting in s_0^ω from s_0, $\mathcal{P}_1$ has a profitable deviation in moving to s_∞ and using a reachability strategy to ensure a finite cost. $\lhd$

A value-based characterisation fails because of states $s \in W_i(\mathsf{Reach}(T_i))$ such that $\mathsf{Val}^i_{\mathcal{G}}(s)$ is infinite. Despite the infinite value of such states, $\mathcal{P}_i$ has a strategy such that their cost is finite no matter the behaviour of the others. For this reason, to characterise NE outcomes, we impose additional conditions on players whose targets are not visited that are related to coalition reachability games.

We show, using a similar approach to the proof of Theorem 6.8, that a play in a shortest-path game is the outcome of an NE if and only if it is an outcome of an NE for the reachability game $(\mathcal{A}, (\mathsf{Reach}(T_i))_{i\in[\![1,n]\!]})$ such that, for players who do see their targets, the values in $(\mathcal{A}_i, \mathsf{SPath}^{T_i}_{w_i})$ suggest they do not have a profitable deviation, in a sense we formalise below.

Theorem 6.9. *Let* $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}^{T_i}_{w_i})_{i\in[\![1,n]\!]})$. *Let* $\pi = s_0a_0s_1\ldots \in \mathsf{Plays}(\mathcal{A})$. *Then* π *is an outcome of an NE from* s_0 *in* $\mathcal{G}$ *if and only*

(i) for all $i \in [\![1,n]\!]$ *such that* $\pi \notin \mathsf{Reach}(T_i)$ *and for all* $\ell \in \mathbb{N}$, *we have* $s_\ell \notin W_i(\mathsf{Reach}(T_i))$ *and*

(ii) for all $i \in [\![1,n]\!]$ *such that* $\pi \in \mathsf{Reach}(T_i)$ *and all* $\ell \leq r_i$, *it holds that* $\mathsf{SPath}^{T_i}_{w_i}(\pi_{\geq \ell}) \leq \mathsf{Val}^i_{\mathcal{G}}(s_\ell)$ *where* $r_i = \min\{r \in \mathbb{N} \mid s_r \in T_i\}$.

Proof. For all $i \in [\![1,n]\!]$, we let $\mathcal{G}_i = (\mathcal{A}_i, \mathsf{SPath}^{T_i}_{w_i})$ denote the coalition game against $\mathcal{P}_i$. Recall that we denote values in $\mathcal{G}_i$ by $\mathsf{Val}^i_{\mathcal{G}}$.

We first prove that if (i) or (ii) does not hold, then π cannot be the outcome of an NE. Let $\sigma = (\sigma_i)_{i\in[\![1,n]\!]}$ such that $\pi = \mathsf{Out}_{\mathcal{A}}(\sigma, s_0)$. We show that some player has a profitable deviation with respect to σ from s_0.

First, assume that (i) does not hold. Let $i \in [\![1,n]\!]$ such that $\pi \notin \mathsf{Reach}(T_i)$

and $\ell \in \mathbb{N}$ such that $s_\ell \in W_1(\mathsf{Reach}(T_i))$. Consider a strategy τ_i of $\mathcal{P}_i$ such that $\pi_{\leq \ell}$ is consistent with τ_i and τ_i agrees with a uniformly winning memoryless strategy in the zero-sum reachability game $(\mathcal{A}_i, \mathsf{Reach}(T_i))$ for all histories that are not a prefix of $\pi_{\leq \ell}$. We obtain that $\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_0)$ is the concatenation of $\pi_{\leq \ell}$ and a play starting in s_ℓ that is in $\mathsf{Reach}(T_i)$. This shows that τ_i is a profitable deviation, and therefore σ is not an NE from s_0.

We now assume that (ii) does not hold. Let $i \in [\![1, n]\!]$ such that $\pi \in \mathsf{Reach}(T_i)$, $r_i = \min\{r \in \mathbb{N} \mid s_r \in T_i\}$ and $\ell \leq r_i$ such that $\mathsf{SPath}^{T_i}_{w_i}(\pi_{\geq \ell}) > \mathsf{Val}^i_{\mathcal{G}}(s_\ell)$. Similarly to above, we consider a strategy τ_i of $\mathcal{P}_i$ such that $\pi_{\leq \ell}$ is consistent with τ_i and τ_i agrees with a memoryless uniform optimal strategy in the zero-sum shortest-path game $(\mathcal{A}_i, \mathsf{SPath}^{T_i}_{w_i})$ for all histories that are not a prefix of $\pi_{\leq \ell}$. In this case, we obtain that $\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_0)$ is the concatenation of $\pi_{\leq \ell}$ and a play π' starting in s_ℓ such that $\mathsf{SPath}^{T_i}_{w_i}(\pi') \leq \mathsf{Val}^i_{\mathcal{G}}(s_\ell)$. This implies that $\mathsf{SPath}^{T_i}_{w_i}(\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_0)) < \mathsf{SPath}^{T_i}_{w_i}(\pi)$. We have shown that τ_i is a profitable deviation with respect to σ, ending the proof of the first implication.

We now show the converse implication. Let $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$ be a strategy profile such that all players follow π, and if $\mathcal{P}_i$ deviates from π, the coalition consisting of the other players switches to a winning strategy in the reachability game $(\mathcal{A}_i, \mathsf{Reach}(T_i))$ if $\pi \notin \mathsf{Reach}(T_i)$ and otherwise the coalition switches to a strategy that ensures $\min\{\mathsf{Val}^i_{\mathcal{G}}(s_\ell), \mathsf{SPath}^{T_i}_{w_i}(\pi_{\geq \ell}) + 1\}$ (we specify a minimum to ensure that the threshold to be ensured is finite) from s_ℓ if the deviation occurs in s_ℓ. It is easy to see that no player has a profitable deviation with respect to σ from in s_0 thanks to (i) and (ii); one can use a straightforward adaptation of the arguments of the proof of Theorem 6.9 to show this. □

Remark 6.10. We complement the above proof by arguing that [BBGT21, Thm. 15] holds in a class of arenas more general than finite arenas. Informally, this characterisation states that a play is the outcome of an NE if and only if condition (ii) of Theorem 6.9 holds for all players (the minimum in the condition is replaced by an infinimum to be well-defined for all players).

This characterisation only fails when there are states $s \in W_i(\mathsf{Reach}(T_i))$ with $\mathsf{Val}^i_{\mathcal{G}}(s) = +\infty$. However, such states do not exist if there are finitely many enabled actions in $\mathcal{P}_2$ states (refer to Remark 6.6). Therefore, the finite-arena characterisation of [BBGT21] extends to finitely-branching arenas. ◁

6.3 Existence of Nash equilibria

It is known that Nash equilibria exist from all states in games where all players have a reachability or a Büchi objective [Umm06], and in games on finite arenas with shortest-path cost functions built on non-negative weights [BDS13]. In this section, we prove the existence of Nash equilibria in games where all players have a shortest-path cost function on arbitrary arenas by building on the approach of [BDS13]. We remark that the argument given below can also be adapted to prove the existence of Nash equilibria in games with reachability and Büchi objectives.

We fix a turn-based deterministic arena $\mathcal{A} = ((S_i)_{i\in[\![1,n]\!]}, A, \delta)$, and, for all $i \in [\![1, n]\!]$, a weight function $w_i \colon S \times A \to \mathbb{N}$ and a target $T_i \subseteq S$ for the remainder of this section. We let $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}^{T_i}_{w_i})_{i\in[\![1,n]\!]})$.

In [BDS13], the authors construct a Nash equilibrium from any state as follows: the players follow uniformly optimal memoryless strategies from their coalition game (in which they are opposed to the other players), and, whenever someone plays otherwise, the other players switch (and commit) to (memoryless uniformly optimal) punishing strategies. In finite arenas, this construction yields a finite-memory Nash equilibrium, as the resulting outcome is a lasso (i.e., we eventually keep repeating the same simple cycle).

Theorem 6.4 guarantees the existence of memoryless uniformly optimal strategies of $\mathcal{P}_1$ in two-player zero-sum shortest-path games. Although memoryless uniformly optimal strategies need not exist for the adversary in a two-player zero-sum shortest-path game (Example 6.1), the strategies provided by Theorem 6.5 suffice to implement the punishing mechanism described above. To avoid redundancy with the proof of Theorem 6.9, we use its characterisation instead of formalising the NE suggested above.

Theorem 6.11. *Let $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}^{T_i}_{w_i})_{i\in[\![1,n]\!]})$. There exists an NE in $\mathcal{G}$ from any initial state.*

Proof. Let $s_0 \in S$ be an initial state. For all $i \in [\![1, n]\!]$, let σ_i be a memoryless uniformly optimal strategy in the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \mathsf{SPath}^{T_i}_{w_i})$ that is uniformly winning in $(\mathcal{A}_i, \mathsf{Reach}(T_i))$, the existence of which follows from

Theorem 6.4. We define $\sigma = (\sigma_i)_{i\in[\![1,n]\!]}$. We argue that $\pi = \mathsf{Out}_{\mathcal{A}}(\sigma, s_0)$ is the outcome of an NE by Theorem 6.9. Establishing this implies the existence of an NE from s_0.

The first condition of Theorem 6.9 follows from the strategies σ_i being memoryless uniformly winning strategies in the reachability game $(\mathcal{A}_i, \mathsf{Reach}(T_i))$. The second condition follows from the uniform optimality of the strategies σ_i in $\mathcal{G}_i$: it ensures $\mathsf{Val}_{\mathcal{G}}(s)$ from all $s \in S$. Therefore, the inequality in the second condition must hold in all relevant cases. □

Chapter 7

Memory requirements for constrained Nash equilibria

This chapter presents our upper bounds on the sufficient amount of memory for solutions to the constrained pure Nash equilibrium existence problem when using move-independent Mealy machines in games on turn-based deterministic arenas. We only consider *pure strategies* for the remainder of the chapter.

Section 7.1 briefly introduces some terminology and an abuse of notation used to lighten proofs. We establish, in Section 7.2, that from any NE in a reachability or shortest-path game, we can derive an NE from the same state given by move-independent Mealy machines of size *quadratic in the number of players* whose outcome has a less or equal cost profile than the outcome of the original NE. In particular, we obtain an upper bound on the size of these Mealy machines that is *independent of the arena*. We then consider Büchi games in Section 7.3, in which we show that from any NE in a Büchi game, we can obtain an NE from the same initial state given by move-independent Mealy machines such that the same players win in the outcomes of the two NEs.

We fix a turn-based deterministic arena $\mathcal{A} = ((S_i)_{i\in[\![1,n]\!]}, A, \delta)$, target sets $T_1, \dots, T_n \subseteq S$ and a weight function $w\colon E \to \mathbb{N}$ for the whole chapter.

Contents

7.1 Terminology and notation

Segments of plays. Throughout this section and the next section, by a *segment* of $\pi \in \mathsf{Plays}(\mathcal{A})$, we mean either an infix h of π (i.e., we can write $\pi = h' \cdot h \cdot \pi'$ for some $h' \in \mathsf{Hist}(\mathcal{A})$ and $\pi' \in \mathsf{Plays}(\mathcal{A})$) or suffix $\pi_{\geq \ell}$ of π. We denote segments by sg to avoid distinguishing finite and infinite segments of plays.

A history is *simple* if no state occurs twice within this history. A *lasso* is a play of the form $h \cdot w^\omega$ where $h \in \mathsf{Hist}(\mathcal{A})$ and $w = s_0 a_1 \dots s_r a_r \in (SA)^+$ is such that ws_0 is a cycle of $\mathcal{A}$. A lasso is *simple* if it can be written as $h \cdot w^\omega$ in such a way that no state occurs twice in $h \cdot w$. A *simple segment* is either a simple history, a simple play or a simple lasso.

Weight of a history. As in the proof of Theorem 6.5, we extend w to histories by letting $w(h) = \sum_{\ell=1}^{r-1} w(s_\ell, a_\ell)$ for all $h = s_0 a_0 s_1 \dots a_{r-1} s_r \in \mathsf{Hist}(\mathcal{A})$.

Iterated update function. To prove that the finite-memory strategies we introduce are NEs, we reason on the memory states reached after a given history. In the remainder of the chapter, we focus on *move-independent Mealy machines* (Definition 5.3). As the updates of these Mealy machines depend only on the sequence of states along a history, we make the following abuse of notation with respect to the iterated memory update function (defined over $(SA)^*$ – see Definition 2.20) of a move-independent Mealy machine.

Let $i \in [\![1, n]\!]$ and $\mathfrak{M} = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be a move-independent Mealy machine of $\mathcal{P}_i$ in $\mathcal{A}$. Let $h \in \mathsf{Hist}(\mathcal{A})$ be a history. We let $\widehat{\mathsf{up}_{\mathfrak{M}}}(h) = \widehat{\mathsf{up}_{\mathfrak{M}}}(ha)$ for some $a \in A(\mathsf{last}(h))$. This definition is independent of the chosen action due to the move-independence of $\mathfrak{M}$. Formally, $\widehat{\mathsf{up}_{\mathfrak{M}}}(h)$ is the memory state reached after h has taken place. This abuse of notation enables us to avoid introducing actions when reasoning on the memory state reached after a given history.

7.2 Reachability and shortest-path games

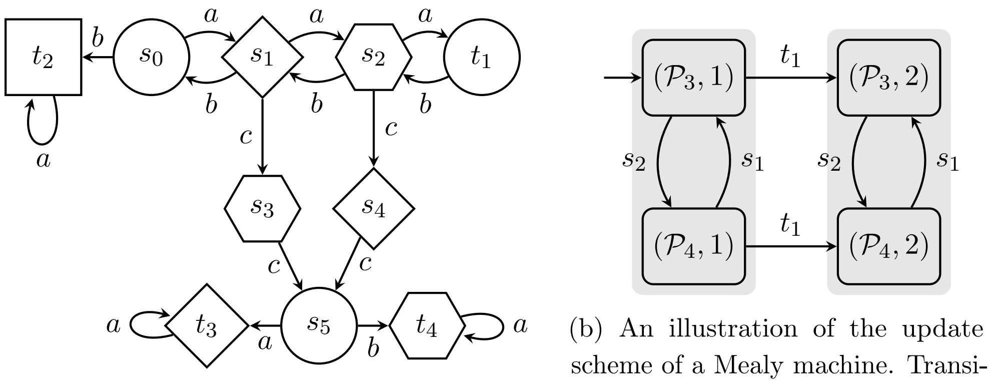
We first study multi-player reachability and shortest-path games on $\mathcal{A}$. The strategies forming our finite-memory NEs behave differently than those described in the NE outcome characterisations of Theorem 6.8 and Theorem 6.9. While the strategies used to establish these characterisation implement a strict *punishment mechanism*, we provide strategies that do not resort to punishing player who deviate from the intended outcome. Instead, if a deviation occurs, the players may attempt to keep following a suffix of the equilibrium's original outcome so long as the deviation does not appear to prevent it.

We first illustrate this idea with examples in Section 7.2.1. We then describe well-structured NE outcomes from which we design our finite-memory NEs in Section 7.2.2. Section 7.2.3 then provides partially-defined move-independent Mealy machines derived from NE outcomes Finally, we extend these Mealy machines to construct NEs in reachability games in Section 7.2.4 and in shortest-path games in Section 7.2.5.

7.2.1 Illustrating finite-memory Nash equilibria

We illustrate the core ideas behind our move-independent Mealy machines implementing NEs on simple examples. We provide an example in a reachability game and in a shortest-path game, to illustrate the slight differences that arise in these two contexts. We open with a reachability game.

Example 7.1. We consider the reachability game $\mathcal{G}$ on the arena depicted in Figure 7.1a where the objective of $\mathcal{P}_i$ is $\mathsf{Reach}(t_i)$ for $i \in [\![1, 4]\!]$. We present a



(a) An arena. Circles, squares, diamonds and hexagons are resp. $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$, $\mathcal{P}_4$ states.

(b) An illustration of the update scheme of a Mealy machine. Transitions that do not change the memory state are omitted.

Figure 7.1: A reachability game and a representation of a move-independent Mealy machine update scheme suitable for an NE from s_0.

finite-memory move-independent pure NE with outcome

$$\pi = s_0 a s_1 a s_2 a t_1 b s_2 b s_1 b s_0 b (t_2 a)^\omega$$

to illustrate the idea behind the upcoming construction. To aid readability, we note that the sequence of states underlying π is $s_0 s_1 s_2 t_1 s_2 s_1 s_0 t_2^\omega$.

First, observe that π can be seen as the combination of the simple history $\mathsf{sg}_1 = s_0 a s_1 a s_2 a t_1$ and the simple lasso $\mathsf{sg}_2 = t_1 b s_2 b s_1 b s_0 b (t_2 a)^\omega$. The simple history sg_1 connects the initial state to the first visited target, and the simple lasso sg_2 connects the first target to the second and contains the suffix of the play. Therefore, if we were not concerned with the stability of the equilibrium, the outcome π could be obtained by using a finite-memory strategy profile where all strategies are defined by a Mealy machine with state space $[\![1, 2]\!]$. Intuitively, these strategies would follow sg_1 while remaining in their first memory state 1, then, when t_1 is visited, they would update their memory state to 2 and follow sg_2.

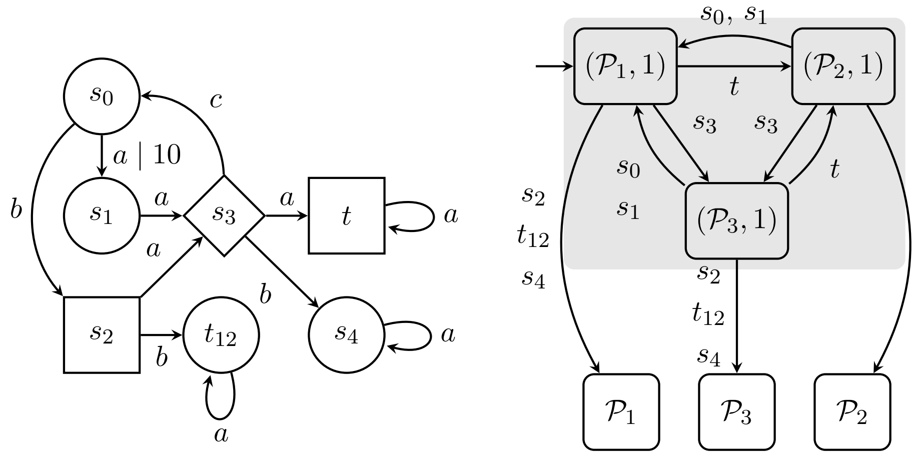
We build on these simple Mealy machines with two states. We include additional information in each memory state. We depict a suitable Mealy machine state space and update scheme in Figure 7.1b. We only label transitions of the Mealy machines with states instead of state-action pairs as we consider move-independent strategies. The rectangles grouping together states $(\mathcal{P}_3, j)$ and $(\mathcal{P}_4, j)$ represent the memory state j of the simpler Mealy machine, for $j \in [\![1, 2]\!]$. Intuitively, the additional information encodes the last player to act among the players whose objective is not satisfied in π. More precisely, an update is performed from the memory state $(\mathcal{P}_i, j)$ only if the state fed to the Mealy machine appears in sg_j for $j \in [\![1, 2]\!]$.

By construction, if $\mathcal{P}_i$ (among $\mathcal{P}_3$ and $\mathcal{P}_4$) deviates and exits the set of states of sg_j when in a memory state of the form $(\cdot, j)$, then the memory updates to $(\mathcal{P}_i, j)$ and does not change until the play returns to some state of sg_j (which is not possible here due to the structure of the arena, but may be in general). For instance, assume that $\mathcal{P}_3$ moves from s_1 to s_3 (with action c). after the history $h = \mathsf{sg}_1 b s_2 b s_1$. Then the Mealy machine state after h is $(\mathcal{P}_3, 2)$ and no longer changes from there on.

It remains to explain how the next-move function of the Mealy machine should be defined to ensure an NE. Essentially, for a state of the form $(\mathcal{P}_i, j)$ and states in sg_j, we assign actions as in the simpler two-state Mealy machine described previously. On the other hand, for a state of the form $(\mathcal{P}_i, j)$ and a state not in sg_j, we use a memoryless punishing strategy against $\mathcal{P}_i$. In this particular case, we need only specify what $\mathcal{P}_1$ should do in s_5. Naturally, in memory state $(\mathcal{P}_i, j)$, $\mathcal{P}_1$ should move to the target of the other player. It is essential to halt memory updates for states s_3 and s_4 to ensure that the correct player is punished.

We close this example with comments on the structure of the Mealy machine. Assume the memory state is of the form $(\mathcal{P}_i, j)$. If a deviation occurs and leads to a state of sg_j other than the intended one, then the other players will continue trying to progress along sg_j and do not specifically try punishing the deviating player. Similarly, if after a deviation leaving the set of states of sg_j (from which point the memory is no longer updated until this set is rejoined), a state of sg_j is visited again, then the players resume trying to progress along this history and memory updates resume. In other words, these finite-memory strategies do not pay attention to all deviations and do not have dedicated memory that commit to punishing deviating players for the remainder of a play after a deviation. $\lhd$

We now give an example in game with shortest-path cost functions. The



(a) A weighted arena. Transition weights are indicated next to actions. Unlabelled transitions have a weight of 1.

(b) A Mealy machine update scheme. Transitions that do not change the memory state are omitted.

Figure 7.2: A shortest-path game and a representation of a move-independent Mealy machine update scheme suitable for some NE from s_0.

Mealy machines we propose for this case are slightly larger: it may be necessary to commit to a punishing strategy if the set of states of the segment that the players want to progress along is left. This requires additional memory states. Our example illustrates that it may be necessary to punish deviations from players whose targets are visited, as they could possibly improve their cost otherwise.

Example 7.2. Let $\mathcal{A}$ and w respectively denote the arena and weight function depicted in Figure 7.2a. We consider the game $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}_w^{T_i})_{i \in [\![1,3]\!]})$ where the targets of $\mathcal{P}_1$ and $\mathcal{P}_2$ are $T_1 = T_2 = \{t, t_{12}\}$ and the target of $\mathcal{P}_3$ is $T_3 = \{t\}$. We argue that a finite-memory NE with outcome $\pi = s_0 a s_1 a s_3 (at)^\omega$ from s_0 cannot be obtained by adapting the construction of Example 7.1. We provide an alternative construction that builds on the same ideas.

The play π is a simple lasso, much like the second part of the play in the previous example. First, let us assume a Mealy machine similar to that of Example 7.1, i.e., such that it tries to progress along π whenever it is in one of its states. The update scheme of such a Mealy machine would be obtained by removing the transitions to states of the form $\mathcal{P}_i$ from Figure 7.2b (replacing them by self-loops).

If $\mathcal{P}_3$ uses a strategy based on such a Mealy machine, then $\mathcal{P}_1$ has a profitable deviation from s_0. Indeed, if $\mathcal{P}_1$ moves from s_0 to s_2 with action b, then either $\mathcal{P}_1$ incurs a cost of 2 if $\mathcal{P}_2$ uses action b in s_2 (i.e., moves to t_{12}) or a cost of 3 if $\mathcal{P}_2$ uses action a (i.e., moves to s_3) as $\mathcal{P}_3$ would then use a and move to t by definition of the Mealy machine. To circumvent this issue, if $\mathcal{P}_i$ exits the set of states of π, we update the memory to the punishment state $\mathcal{P}_i$. This results in the update scheme depicted in Figure 7.2b. Next-move functions to obtain an NE can be defined as follows, in addition to the expected behaviour to obtain π: for $\mathcal{P}_2$, $\mathsf{nxt}_{\mathfrak{M}_2}((\mathcal{P}_1, 1), s_2) = \mathsf{nxt}_{\mathfrak{M}_2}(\mathcal{P}_1, s_2) = s_3$ and for $\mathcal{P}_3$, $\mathsf{nxt}_{\mathfrak{M}_3}(\mathcal{P}_1, s_3) = s_4$.

Similarly to the previous example, players do not explicitly react to deviations that move to states of π; if $\mathcal{P}_3$ deviates after reaching s_3 and moves back to s_0, the memory of the other players does not update to state $\mathcal{P}_3$. Intuitively, there is no need to switch to a punishing strategy for $\mathcal{P}_3$ as going back to the start of the intended outcome is more costly than conforming to it, preventing the existence of a profitable deviation. ◁

Remark 7.1. Example 7.2 differs slightly from the general construction below. According to the general construction, we should decompose π into two parts: a history $s_0as_1as_3at$ from the initial state to the first target and the suffix $(ta)^\omega$ of the play after all targets are visited. Furthermore, we can argue that such a split is sometimes necessary (see Example 7.3). ◁

7.2.2 Simple Nash equilibria outcomes

A common trait of the NE outcomes of Examples 7.1 and 7.2 is that they are derived from NE outcomes that can be written as a concatenation of simple segments. We construct our move-independent finite-memory NEs from such outcomes. In this section, we show that given an NE outcome from an initial state s, we can find another NE outcome whose form is suitable to generalise the ideas underlying Examples 7.1 and 7.2.

We formulate the results of this section for shortest-path games. They apply to reachability games through the following observation: a pure NE of $(\mathcal{A}, (\mathsf{Reach}(T_i))_{i\in[\![1,n]\!]})$ from a state $s \in S$ is an NE of the shortest-path game $(\mathcal{A}, (\mathsf{SPath}_0^{T_i})_{i\in[\![1,n]\!]})$ where all weights are zero. Therefore, we fix $\mathcal{G} =$

$(\mathcal{A}, (\mathsf{SPath}_w^{T_i})_{i \in [\![1,n]\!]})$ for the remainder of the section.

First, we introduce segment decompositions of plays.

Definition 7.2. Let $\pi \in \mathsf{Plays}(\mathcal{A})$. A *segment decomposition* of π is a (possibly infinite) sequence $\mathcal{S} = (\mathsf{sg}_j)_{j=1}^{k}$ of segments of π such that π is the concatenation $\mathsf{sg}_1 \cdot \mathsf{sg}_2 \cdot \ldots$ of the segments in $\mathcal{S}$. The segment decomposition $\mathcal{S}$ is *finite* if $k \in \mathbb{N}_{>0}$ and is *simple* if all segments in $\mathcal{S}$ are simple.

In the following, we assume that among the histories of a decomposition, there are none of the form $h = s$, i.e., there are no trivial segments.

The goal of this section is to prove that given an NE outcome of $\mathcal{G}$, we can find an NE outcome of $\mathcal{G}$ with the same initial state, a preferable cost profile and a finite simple segment decomposition (with some additional technical properties).

To concisely formulate our results, we introduce some notation. For any $\pi = s_0 a_0 s_1 \ldots \in \mathsf{Plays}(\mathcal{A})$, we let $\mathsf{VisPl}^{\mathcal{G}}(\pi) = \{i \in [\![1,n]\!] \mid \pi \in \mathsf{Reach}(T_i)\}$ denote the set of players whose targets are visited in π and $\mathsf{VisPos}^{\mathcal{G}}(\pi) = \{\min\{\ell \in \mathbb{N} \mid s_\ell \in T_i\} \mid i \in \mathsf{VisPl}^{\mathcal{G}}(\pi)\}$ be the set of indices of π at which the target of a player is visited for the first time.

We consider two types of NE outcomes in reachability games. First, we consider NE outcomes such that all players who see their target have the initial state of the outcome in it. This generalises the case in which no players see their target. From these outcomes, we can directly derive an NE outcome that is a simple lasso or simple play.

Lemma 7.3. *Let $\pi' \in \mathsf{Plays}(\mathcal{A})$ be the outcome of an NE from $s_0 \in S$ in game $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}_w^{T_i})_{i \in [\![1,n]\!]})$ such that $\mathsf{VisPos}^{\mathcal{G}}(\pi') \subseteq \{0\}$. There exists an NE outcome $\pi \in \mathsf{Plays}(\mathcal{A})$ from s_0 with the same cost profile as π' that is a simple lasso or a simple play and such that $\mathsf{VisPos}^{\mathcal{G}}(\pi) \subseteq \{0\}$. In particular, π has the simple segment decomposition (π).*

Proof. We first observe that if π' is a simple play, the result follows immediately. Therefore, we assume that π' is not a simple play. This implies that there is a simple lasso $\pi \in \mathsf{Plays}(\mathcal{A})$ starting from s_0 that only uses states that occur in π'. It follows that $\mathsf{VisPos}^{\mathcal{G}}(\pi) \subseteq \{0\}$. By Theorem 6.9, π is an NE outcome; condition (i) of the characterisation follows from it holding for π' and condition (ii) holds because $\mathsf{VisPos}^{\mathcal{G}}(\pi) \subseteq \{0\}$. □

We now consider NE outcomes such that some player sees their target later than in the initial state. From an NE outcome π', we derive an NE outcome π (from $\mathsf{first}(\pi')$) with a finite simple decomposition such that the simple histories of this decomposition end in the first occurring elements of the visited target sets along π'. The idea is to first decompose π' into segments connecting these target elements. We then replace the histories of this decomposition with simple histories and change the last segment so the concatenation of the last two segments is a simple lasso or simple play. We impose an additional condition on the simple histories in the decomposition of the resulting play π, to ensure that, when building a finite-memory NE from π, no player can obtain profitable deviation by skipping ahead in a segment.

Lemma 7.4. *Let π' be the outcome of an NE from $s_0 \in S$ in $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}_w^{T_i})_{i\in[\![1,n]\!]})$. Assume that $|\mathsf{VisPos}^{\mathcal{G}}(\pi') \setminus \{0\}| = k > 0$. There exists an NE outcome π from s_0 in $\mathcal{G}$ with $\mathsf{VisPos}^{\mathcal{G}}(\pi) \setminus \{0\} = \{\ell_1 < \ldots < \ell_k\}$ that admits a simple segment decomposition $(\mathsf{sg}_1, \ldots, \mathsf{sg}_{k+1})$ such that*

(i) $(\mathsf{sg}_1, \ldots, \mathsf{sg}_k \cdot \mathsf{sg}_{k+1})$ is also a simple decomposition of π;

(ii) for all $j \in [\![1, k]\!]$, $\mathsf{sg}_1 \cdot \ldots \cdot \mathsf{sg}_j = \pi_{\leq \ell_j}$;

(iii) for all $j \in [\![1, k]\!]$, $w(\mathsf{sg}_j)$ is minimum among all histories that share their first and last state with sg_j and traverse a subset of the states occurring in sg_j; and

(iv) for all $i \in [\![1, n]\!]$, $\mathsf{SPath}_w^{T_i}(\pi) \leq \mathsf{SPath}_w^{T_i}(\pi')$.

Proof. We define π by describing the simple decomposition $\mathcal{S} = (\mathsf{sg}_1, \ldots, \mathsf{sg}_{k+1})$. Let $\ell'_1 < \ldots < \ell'_k$ be the elements of $\mathsf{VisPos}^{\mathcal{G}}(\pi') \setminus \{0\}$ and $\ell'_0 = 0$. For $j \in [\![1, k]\!]$, we let sg'_j denote the segment of π between positions ℓ'_{j-1} and ℓ'_j. We let sg_j be a simple history that shares its first and last state with sg'_j and traverses a subset of the states occurring in sg'_j, with minimal weight among all such histories (actions that do not occur in sg'_j may be used in sg_j). It remains to define the segment sg_{k+1}. We let sg_{k+1} be $\pi'_{\geq \ell_k}$ if $\mathsf{sg}_k \cdot \pi'_{\geq \ell_k}$ is a simple play, and otherwise we let sg_{k+1} be any play starting in $\mathsf{last}(\mathsf{sg}_k)$ such that $\mathsf{sg}_k \cdot \mathsf{sg}_{k+1}$ is a simple lasso in which only states of π occur. It follows from this choice of sg_{k+1} that $\pi = \mathsf{sg}_1 \cdot \ldots \cdot \mathsf{sg}_{k+1}$ satisfies condition (i).

We now argue that the play π is an NE outcome satisfying conditions (ii)-(iv). Let $\pi = s_0 a_0 s_1 \ldots$, $\ell_1 < \ldots < \ell_k$ be the elements of $\mathsf{VisPos}^{\mathcal{G}}(\pi) \setminus \{0\}$ and $\ell_0 = 0$. To argue that π is an NE outcome, we rely on the characterisation in Theorem 6.9. Because π' is an NE outcome and all states occurring in π occur in π', it follows the first condition of the characterisation of Theorem 6.9 holds for π.

For the second condition of the characterisation, we fix $i \in \mathsf{VisPl}^{\mathcal{G}}(\pi)$ and $j_i \leq k$ such that $\ell_{j_i} = \min\{\ell \in \mathbb{N} \mid s_\ell \in T_i\}$. We show that for all $\ell \leq \ell_{j_i}$, we have $\mathsf{SPath}_w^{T_i}(\pi_{\geq \ell}) \leq \mathsf{Val}_{\mathcal{G}}^i(s_\ell)$ where $\mathsf{Val}_{\mathcal{G}}^i(s_\ell)$ is the value of s_ℓ in the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \mathsf{SPath}_w^{T_i})$.

Let $\ell \leq \ell_{j_i}$ and $j \leq j_i$ such that $\ell_j \leq \ell < \ell_{j+1}$. By construction, there is an occurrence of s_ℓ in the segment sg'_j of π'. We consider a suffix $\pi'_{\geq \ell'}$ of π' starting from an occurrence of s_ℓ in sg'_j. The desired inequality follows from the relations $\mathsf{SPath}_w^{T_i}(\pi_{\geq \ell}) \leq \mathsf{SPath}_w^{T_i}(\pi'_{\geq \ell'}) \leq \mathsf{Val}_{\mathcal{G}}^i(s_\ell)$.

We prove that the first inequality holds by contradiction. Assume that $\mathsf{SPath}_w^{T_i}(\pi_{\geq \ell}) > \mathsf{SPath}_w^{T_i}(\pi'_{\geq \ell'})$. It must be the case that either a suffix of sg'_j starting in s_ℓ must have weight strictly less than the suffix $s_\ell a_\ell \ldots s_{\ell_j}$ of sg_j, or that $w(\mathsf{sg}_{j'}) > w(\mathsf{sg}'_{j'})$ for some $j < j' \leq j_i$. Both possibilities contradict the choice of the elements of $\mathcal{S}$, therefore we have $\mathsf{SPath}_w^{T_i}(\pi_{\geq \ell}) \leq \mathsf{SPath}_w^{T_i}(\pi'_{\geq \ell'})$. The second inequality holds by Theorem 6.9 as π' is an NE outcome. We remark (for condition (iv)) that in the special case $\ell = 0$, the first inequality implies that $\mathsf{SPath}_w^{T_i}(\pi) \leq \mathsf{SPath}_w^{T_i}(\pi')$ as we can choose $\ell' = 0$. We have shown that π is an NE outcome.

We now show that conditions (ii)-(iv) hold. Condition (ii) follows immediately by construction. Let $j \in [\![1, k]\!]$. The minimum in condition (iii) for j is attained by some simple history. By construction, it must be realised by sg_j. This implies that condition (iii) holds. For condition (iv), due to the above, we need only consider players who do not see their target. For these players,

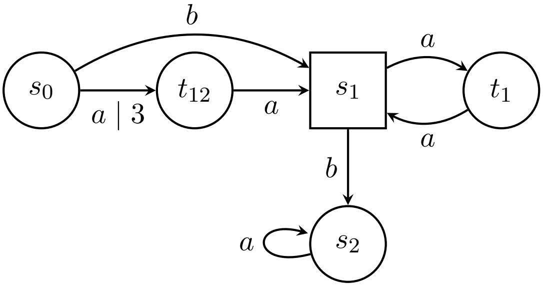


Figure 7.3: The turn-based arena of Figure 2.6. Circles and squares respectively denote $\mathcal{P}_1$ and $\mathcal{P}_2$ states. Unspecified weights are 1 and are omitted to lighten the figure.

the condition follows from the equality $\mathsf{VisPl}^{\mathcal{G}}(\pi) = \mathsf{VisPl}^{\mathcal{G}}(\pi')$ implying that players have an infinite cost in π if and only if they have an infinite cost in π'. This concludes the proof that conditions (ii)-(iv) are satisfied by π. □

We provide further comments on the statement of Lemma 7.4. Due to condition (i) on the outcome, we could consider decompositions with one less element. However, working with a decomposition where these segments are merged may prevent us from ensuring the stability of an NE with strategies that only punish players who exit the current segment of the decomposition (as in Example 7.1) during the play. Intuitively, some player could have an incentive to move to sg_{k+1} before reaching the last state of sg_k. We illustrate one such situation in the following example.

Example 7.3. We revisit the game used in Example 2.5. We recall the relevant arena $\mathcal{A}$ and weight function w in Figure 7.3. We consider the game $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}_w^{T_1}, \mathsf{SPath}_w^{T_2}))$ where the target of $\mathcal{P}_1$ is $T_1 = \{t_1, t_{12}\}$ and the target of $\mathcal{P}_2$ is $T_2 = \{t_{12}\}$. The play $\pi = s_0 a t_{12} (a s_1 a t_1 a)^\omega$ is a simple lasso that is an NE outcome by Theorem 6.9. We claim that there are no NEs where $\mathcal{P}_2$ selects action a in s_1, i.e., where $\mathcal{P}_2$ plays consistently with the simple decomposition $\mathcal{S} = (\pi)$ when in its unique segment.

Let σ_2 be a strategy of $\mathcal{P}_2$ such that $\sigma_2(h) = a$ for all $h \in \mathsf{Hist}(\mathcal{A})$ such that all states in h occur in π and $\mathsf{last}(h) = s_1$. The history $h = s_0 b s_1$ satisfies these last two properties, and thus $\sigma_2(h) = a$. If $\mathcal{P}_1$ moves from s_0 to s_1 while $\mathcal{P}_2$

follows σ_2, $\mathcal{P}_1$ can obtain a cost of 2 rather than 3 (3 being the cost of π for both players). This shows that there are no NEs where $\mathcal{P}_2$ uses σ_2. ◁

Condition (i) on decompositions in Lemma 7.4 is relevant for reachability games. The issue highlighted by Example 7.3 is specific to the shortest-path setting: players whose targets are visited do not have profitable deviations in reachability games. Merging these last two segments provides us with a smaller decomposition, which in turn yields smaller memory bounds for move-independent finite-memory NEs in reachability games. Intuitively, in this qualitative setting, there is no need to distinguish the last two segments of the decomposition given by the lemma.

7.2.3 Decomposition-based finite-memory strategies

Lemma 7.3 and Lemma 7.4 imply that we can improve the cost profile of any NE outcome by considering NE outcomes that admit a (well-structured) simple segment decomposition. We now endeavour to construct move-independent finite-memory NEs from such outcomes. In this section, we introduce *strategies based on a simple segment decomposition*. We then provide partially-defined Mealy machines that induce strategies based on simple segment decompositions. We build on these Mealy machines in the following section to obtain our arena-independent memory bounds for (move-independent) pure NEs in reachability and shortest-path games. We fix a play π that admits a simple decomposition $\mathcal{S} = (\mathsf{sg}_1, \dots, \mathsf{sg}_k)$ for the remainder of the section.

In the NEs of Examples 7.1 and 7.2, not all deviations were punished: players would try to continue along the segment of the intended outcome being built as long as it is not left. For instance, in Example 7.2, if $\mathcal{P}_3$ deviates, resulting in the history $s_0 a s_1 a s_3 c s_0$, the other players do not try to prevent $\mathcal{P}_3$ from reaching a target (despite it being possible from s_0). Instead, they attempt to follow the moves suggested by the segment $s_0 a s_1 a s_3 t$, i.e., they maintain their initial behaviour. In a sense, it is because this history is *coherent* with the considered simple decomposition, i.e., whenever the players try to complete a segment, the set of states of this segment is never left.

Formally, we say that a history is coherent with $\mathcal{S}$ if there is some $j \in [\![1, k]\!]$ such that it is j-coherent with $\mathcal{S}$. We define j-coherence inductively as follows. The base case of the induction is the history s_0; it is 1-coherent with $\mathcal{S}$. We now consider a j-coherent history h, and let $a \in A(\mathsf{last}(h))$ and $s = \delta(\mathsf{last}(h), a)$. If $j < k$ and $s = \mathsf{last}(\mathsf{sg}_j)$, then has is $(j+1)$-coherent with $\mathcal{S}$. Otherwise, if s occurs in sg_k, then has is j-coherent. In any other case, has is not coherent with $\mathcal{S}$.

We now define strategies that, given a coherent history, attempt to complete the segment in progress. First, we define the action that players should use after a coherent history. Given a history h that is j-coherent, we define the *next action* of h with respect to $\mathcal{S}$ as the action that follows $\mathsf{last}(h)$ in sg_j. We prove that this action is well-defined below.

Lemma 7.5. *Let h be a history that is coherent with the simple decomposition $\mathcal{S} = (\mathsf{sg}_1, \ldots, \mathsf{sg}_k)$. The next action of h with respect to $\mathcal{S}$ is well-defined.*

Proof. We assume that h is j-coherent. We establish existence and uniqueness of this action. Uniqueness follows from the simplicity of the decomposition. Existence is clear if $j = k$: sg_k does not have a final state.

We therefore assume that $j < k$ and establish the existence of a next action by contradiction. Assume there is no suitable action. This implies that $\mathsf{last}(h) = \mathsf{last}(\mathsf{sg}_j)$. By simplicity and the absence of trivial histories in a decomposition, we have $\mathsf{first}(h_j) \neq \mathsf{last}(h_j)$. Therefore, there must be a prefix of h that is j-coherent by definition of j-coherence. We obtain that h should either be $(j+1)$-coherent or not coherent, a contradiction. □

We say that a pure strategy of $\mathcal{P}_i$ is *based on* $\mathcal{S}$ if to any history $h \in \mathsf{Hist}_i(\mathcal{A})$ that is coherent with $\mathcal{S}$, it assigns the next action of h with respect to $\mathcal{S}$. Any strategy profile $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$ such that σ_i is based on $\mathcal{S}$ for all $i \in [\![1, n]\!]$ is such that $\mathsf{Out}_{\mathcal{A}}(\sigma, s_0) = \pi$.

We now formalise partially-defined Mealy machines for all players that induce strategies that are based on $\mathcal{S}$. These Mealy machines serve as the basis for the finite-memory NEs described in the next sections. These Mealy machines share the same memory state space, initial memory state and memory update function.

The memory state space is made of pairs of the form $(\mathcal{P}_i, j)$ where $i \in [\![1, n]\!]$ and $j \in [\![1, k]\!]$. We do not consider all such pairs, e.g., it is not necessary in

Example 7.1. Therefore, we parameterise our construction by a non-empty set of players $I \subseteq [\![1, n]\!]$. We consider the memory state space $M^{I,\mathcal{S}} = \{\mathcal{P}_i \mid i \in I\} \times [\![1, k]\!]$. The initial state $m_{\mathsf{init}}^{I,\mathcal{S}}$ is any state of the form $(\mathcal{P}_i, 1) \in M^{I,\mathcal{S}}$.

The update function $\mathsf{up}_{\mathfrak{M}}^{I,\mathcal{S}}$ behaves similarly to Figure 7.1b. It keeps track of the last player in I to have moved and the current segment. Formally, for any $(\mathcal{P}_i, j) \in M^{I,\mathcal{S}}$ and state s occurring in sg_j, we let $\mathsf{up}_{\mathfrak{M}}^{I,\mathcal{S}}((\mathcal{P}_i, j), s) = (\mathcal{P}_{i'}, j')$ where (i) i' is such that $s \in S_{i'}$ if $s \in \bigcup_{i'' \in I} S_{i''}$ and otherwise $i' = i$, and (ii) $j' = j + 1$ if $j < k$ and $s = \mathsf{last}(\mathsf{sg}_j)$ and $j' = j$ otherwise. Updates from $(\mathcal{P}_i, j)$ for a state that does not appear in sg_j are left undefined.

The next-move function $\mathsf{nxt}_{\mathfrak{M}_i}^{I,\mathcal{S}}$ of $\mathcal{P}_i$ proposes the action following the input state in the current segment. Formally, given a memory state $(\mathcal{P}_{i'}, j) \in M^{I,\mathcal{S}}$ and a state $s \in S_i$ that occurs in sg_j, we let $\mathsf{nxt}_{\mathfrak{M}_i}^{I,\mathcal{S}}((\mathcal{P}_{i'}, j), s)$ be the action occurring after s in sg_{j+1} if $j < k$ and $s = \mathsf{last}(\mathsf{sg}_j)$, and otherwise we let it be the action occurring after s in sg_j. Like updates, the next-move function is left undefined in memory states $(\mathcal{P}_i, j)$ for a state that does not appear in sg_j.

We now prove that any finite-memory strategy induced by a Mealy machine that extends the partially defined Mealy machine $(M^{I,\mathcal{S}}, m_{\mathsf{init}}^{I,\mathcal{S}}, \mathsf{up}_{\mathfrak{M}}^{I,\mathcal{S}}, \mathsf{nxt}_{\mathfrak{M}_i}^{I,\mathcal{S}})$ is based on $\mathcal{S}$. To this end, we establish that if a history h is j-coherent with $\mathcal{S}$, then the memory state after the Mealy machine processes h is of the form $(\mathcal{P}_i, j)$.

Lemma 7.6. *Let $\mathfrak{M}_i = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}_i}, \mathsf{up}_{\mathfrak{M}_i})$ be a move-independent Mealy machine of $\mathcal{P}_i$ such that $M^{I,\mathcal{S}} \subseteq M$, $m_{\mathsf{init}}^{I,\mathcal{S}} = m_{\mathsf{init}}$, $\mathsf{up}_{\mathfrak{M}_i}$ and $\mathsf{nxt}_{\mathfrak{M}_i}$ coincide with $\mathsf{up}_{\mathfrak{M}}^{I,\mathcal{S}}$ and $\mathsf{nxt}_{\mathfrak{M}_i}^{I,\mathcal{S}}$ respectively on the domain of the latter functions. The strategy σ_i induced by $\mathfrak{M}_i$ is based on $\mathcal{S}$ and for all $h \in \mathsf{Hist}(\mathcal{A})$, if h is j-coherent with $\mathcal{S}$, then $\widehat{\mathsf{up}_{\mathfrak{M}_i}}(h) = (\mathcal{P}_{i'}, j)$ for some $i' \in I$.*

Proof. We first show the second claim of the lemma. We proceed by induction on the number of states in h. The only coherent history with a single state is s_0. The assumption that $\mathcal{S}$ contains no trivial segments ensures that $s_0 \neq \mathsf{last}(\mathsf{sg}_1)$, thus we have that $\mathsf{up}_{\mathfrak{M}_i}(m_{\mathsf{init}}, s_0)$ is of the form $(\mathcal{P}_{i'}, 1)$.

We now consider a j-coherent history h and assume by induction that $\widehat{\mathsf{up}_{\mathfrak{M}_i}}(h) = (\mathcal{P}_{i'}, j)$. Let $a \in A(\mathsf{last}(h))$ such that has is coherent with $\mathcal{S}$ where $s = \delta(\mathsf{last}(h), a)$. It follows that s occurs in sg_j. Therefore, $\widehat{\mathsf{up}_{\mathfrak{M}_i}}(has) =$

$\mathsf{up}_{\mathfrak{M}_i}((\mathcal{P}_{i'}, j), s) = \mathsf{up}_{\mathfrak{M}}^{I,\mathcal{S}}((\mathcal{P}_{i'}, j), s)$. We distinguish two cases. If $j < k$ and $s = \mathsf{last}(\mathsf{sg}_j)$, then has is $(j+1)$-coherent and by definition of $\mathsf{up}_{\mathfrak{M}}^{I,\mathcal{S}}$, $\mathsf{up}_{\mathfrak{M}}^{I,\mathcal{S}}((\mathcal{P}_{i'}, j), s)$ is of the form $(\mathcal{P}_{i''}, j+1)$. Otherwise, has is j-coherent and by definition of $\mathsf{up}_{\mathfrak{M}}^{I,\mathcal{S}}$, $\mathsf{up}_{\mathfrak{M}}^{I,\mathcal{S}}((\mathcal{P}_{i'}, j), s)$ is of the form $(\mathcal{P}_{i''}, j)$.

It remains to argue that σ_i is based on $\mathcal{S}$. Let $h \in \mathsf{Hist}_i(\mathcal{A})$ be a history coherent with $\mathcal{S}$. If h contains only one state, then by coherence $h = s_0$. The definition of $\mathsf{nxt}_{\mathfrak{M}_i}^{I,\mathcal{S}}$ ensures that $\sigma_i(s_0)$ is the next action of the history s_0 with respect to $\mathcal{S}$. If h contains more than one state, let $h = h'as$ and assume that h' is j-coherent. By the previous point, it holds that $\widehat{\mathsf{up}_{\mathfrak{M}_i}}(h') = (\mathcal{P}_{i'}, j)$ for some $i' \in I$. Therefore, $\sigma_i(h) = \mathsf{nxt}_{\mathfrak{M}_i}^{I,\mathcal{S}}((\mathcal{P}_{i'}, j), s)$. It follows from the definition of $\mathsf{nxt}_{\mathfrak{M}_i}^{I,\mathcal{S}}$ that σ_i maps h to its next action with respect to $\mathcal{S}$. □

7.2.4 Nash equilibria in reachability games

We consider the reachability game $\mathcal{G} = (\mathcal{A}, (\mathsf{Reach}(T_i))_{i\in[\![1,n]\!]})$. We generalise the construction illustrated in Example 7.1. We construct finite-memory NEs by extending the partially-defined Mealy machines of Section 7.2.3. We build on the NE outcomes with simple decompositions provided by Lemma 7.3 and Lemma 7.4.

The general idea of the construction is to use the state space $M^{I,\mathcal{S}}$ where $I \subseteq [\![1, n]\!]$ is the set of players who do not see their targets if it is non-empty, or a single arbitrary player if all players see their target. Let $i' \in I$ and $j \in [\![1, k]\!]$. We extend $\mathsf{up}_{\mathfrak{M}}^{I,\mathcal{S}}$ so that the memory state is unchanged when performing an update in a memory state $(\mathcal{P}_{i'}, j)$ by reading a game state that does not occur in sg_j. In this same situation, the next-move function $\mathsf{nxt}_{\mathfrak{M}_i}^{I,\mathcal{S}}$ is extended to assign moves from a memoryless uniformly winning strategy of the second player in the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \mathsf{Reach}(T_i))$ (which exists by Theorem 6.1). The equilibrium's stability is a consequence of the NE outcome characterisation in Theorem 6.8 and the description of winning strategies of the second player in two-player zero-sum reachability games of Theorem 6.1: if the target of $\mathcal{P}_i$ is not visited in the intended outcome, all states along this play are not winning for the first player of the coalition game $\mathcal{G}_i$. We formalise the explanation above in the proof of the following theorem.

Theorem 7.7. *Let $\mathcal{G} = (\mathcal{A}, (\mathsf{Reach}(T_i))_{i\in[\![1,n]\!]})$ be a reachability game. Let σ' be a pure NE from a state s_0. There exists a pure finite-memory NE σ from s_0 such that $\mathsf{VisPl}^{\mathcal{G}}(\mathsf{Out}_{\mathcal{A}}(\sigma, s_0)) = \mathsf{VisPl}^{\mathcal{G}}(\mathsf{Out}_{\mathcal{A}}(\sigma', s_0))$ where each strategy of σ is induced by a move-independent Mealy machine of size at most n^2. More precisely, we can bound the size of these Mealy machines by*

$$\max\{1, n - |\mathsf{VisPl}^{\mathcal{G}}(\mathsf{Out}_{\mathcal{A}}(\sigma', s_0))|\} \cdot \max\{1, |\mathsf{VisPos}^{\mathcal{G}}(\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)) \setminus \{0\}|\}.$$

Proof. Let $k = \max\{1, |\mathsf{VisPos}^{\mathcal{G}}(\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)) \setminus \{0\}|\}$. By Lemma 7.3 or Lemma 7.4 (condition (i) on the outcome, there exists an NE outcome π from s_0 that admits a simple segment decomposition $\mathcal{S} = (\mathsf{sg}_1, \ldots, \mathsf{sg}_k)$ and such that $\mathsf{VisPl}^{\mathcal{G}}(\pi) = \mathsf{VisPl}^{\mathcal{G}}(\mathsf{Out}_{\mathcal{A}}(\sigma', s_0))$.

Let $I = [\![1, n]\!] \setminus \mathsf{VisPl}^{\mathcal{G}}(\pi)$ if it is not empty and otherwise let $I = \{1\}$. For $i \in I$, let τ_{-i} denote a memoryless strategy of the second player in the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \mathsf{Reach}(T_i))$ that is uniformly winning on their winning region (Theorem 6.1). We let $W_{-i}(\mathsf{Safe}(T_i))$ denote this winning region.

We formally extend the Mealy machines of Section 7.2.3. Let $i \in [\![1, n]\!]$. We consider the Mealy machine $\mathfrak{M}_i = (M^{I,\mathcal{S}}, m^{I,\mathcal{S}}_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}_i}, \mathsf{up}_{\mathfrak{M}})$ where $\mathsf{nxt}_{\mathfrak{M}_i}$ and $\mathsf{up}_{\mathfrak{M}}$ respectively extend $\mathsf{nxt}^{I,\mathcal{S}}_{\mathfrak{M}_i}$ and $\mathsf{up}^{I,\mathcal{S}}_{\mathfrak{M}}$ as follows. Let $(\mathcal{P}_{i'}, j) \in M^{I,\mathcal{S}}$ and $s \in S$ that does not occur in sg_j. We let $\mathsf{up}_{\mathfrak{M}}((\mathcal{P}_{i'}, j), s) = (\mathcal{P}_{i'}, j)$ and, if $s \in S_i$, we let $\mathsf{nxt}_{\mathfrak{M}_i}((\mathcal{P}_{i'}, j), s) = \tau_{-i'}(s)$ if $i' \neq i$ and otherwise we let $\mathsf{nxt}_{\mathfrak{M}_i}((\mathcal{P}_{i'}, j), s)$ be arbitrary. We let σ_i denote the strategy induced by $\mathfrak{M}_i$. By definition of $M^{I,\mathcal{S}}$, we have the asserted bounds on the memory size of σ_i. It follows from Lemma 7.6 that the outcome of $\sigma = (\sigma_i)_{i\in[\![1,n]\!]}$ from s_0 is π.

We now prove that σ is an NE from s_0. It suffices to show that for all $i \notin \mathsf{VisPl}^{\mathcal{G}}(\pi)$, $\mathcal{P}_i$ does not have a profitable deviation. Fix $i \notin \mathsf{VisPl}^{\mathcal{G}}(\pi)$. It suffices to show that all histories starting in s_0 that are consistent with the strategy profile $\sigma_{-i} = (\sigma_{i'})_{i'\neq i}$ only traverse states of $W_{-i}(\mathsf{Safe}(T_i))$. We prove this by induction on the length of histories. To aid in this proof, we show in parallel that for all *has* $\in \mathsf{Hist}(\mathcal{A}, s_0)$ consistent with σ_{-i}, if $(\mathcal{P}_{i'}, j) = \widehat{\mathsf{up}_{\mathfrak{M}}}(h)$ and s does not occur in sg_j, then $i' = i$.

We first remark that, by our characterisation of NE outcomes in reachability games (Theorem 6.8), all states occurring in π are in $W_{-i}(\mathsf{Safe}(T_i))$. The base case of the induction is the history s_0. Both claims hold without issue. We now assume that the claim holds for a history has starting in s_0 consistent with σ_{-i} by induction, and show they hold for $hasbt$, assumed consistent with σ_{-i}. Let $(\mathcal{P}_{i'}, j) = \widehat{\mathsf{up}_{\mathfrak{M}}}(m^{I,\mathcal{S}}_{\mathsf{init}}, h)$.

We first argue that $t \in W_{-i}(\mathsf{Safe}(T_i))$. Assume that $s \in S_i$. The induction hypothesis implies that $s \in W_{-i}(\mathsf{Safe}(T_i))$. If $t \notin W_{-i}(\mathsf{Safe}(T_i))$, then by determinacy of zero-sum reachability games (Theorem 6.1), $\mathcal{P}_1$ could force a visit to T_i from t and therefore could also win from s, which is a contradiction. We now assume that $s \notin S_i$. We have $b = \sigma_{i''}(has) = \mathsf{nxt}_{\mathfrak{M}_{i''}}((\mathcal{P}_{i'}, j), s)$ for some $i'' \neq i$. We consider two cases. If s occurs in sg_j, then $t = \delta(s, b)$ occurs in π by definition of $\mathsf{nxt}_{\mathfrak{M}_{i''}}$. This implies that $t \in W_{-i}(\mathsf{Safe}(T_i))$. Otherwise, by the induction hypothesis, we have $i' = i$, which implies $b = \tau_{-i}(s)$, and thus $t \in W_{-i}(\mathsf{Safe}(T_i))$ (otherwise, τ_{-i} would not be winning in $\mathcal{G}_i$).

We now move on to the second half of the induction. Let $(\mathcal{P}_{i''}, j') = \widehat{\mathsf{up}_{\mathfrak{M}}}(has)$. By definition of $\mathsf{up}_{\mathfrak{M}}$, we have $j' = j + 1$ if $j < k$ and $s = \mathsf{last}(\mathsf{sg}_j)$ and $j' = j$ otherwise. It follows that s occurs in sg_j if and only if it occurs in $\mathsf{sg}_{j'}$. Assume that t does not occur in $\mathsf{sg}_{j'}$. We consider two cases. First, assume that s occurs in $\mathsf{sg}_{j'}$. We must have $s \in S_i$ by definition of σ_{-i}. The definition of $\mathsf{up}^{I,\mathcal{S}}_{\mathfrak{M}}$ ensures that $i'' = i$. Second, assume that s does not occur in $\mathsf{sg}_{j'}$. On the one hand, we have $i' = i$ by the induction hypothesis. On the other hand, the definition of $\mathsf{up}_{\mathfrak{M}}$ implies that $(\mathcal{P}_{i''}, j') = (\mathcal{P}_{i'}, j)$, implying $i'' = i$. This ends the proof by induction.

We have shown that players whose targets are not visited in $\pi = \mathsf{Out}_{\mathcal{A}}(\sigma, s_0)$ do not have a profitable deviation. This shows that σ is an NE from s_0. □

Theorem 7.7 provides a memory bound that is *linear in the number of players* when at most one player does not see their target, and when at most one player sees their target.

Corollary 7.8. *If there exists a pure NE from s_0 such that at most one player sees (resp. does not see) their target in its outcome, then there is a pure finite-memory NE from s_0 such that the same targets are visited in its outcome and all strategies are induced by a move-independent Mealy machine of size at most n.*

7.2.5 Nash equilibria in shortest-path games

We now consider the shortest-path game $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}_w^{T_i})_{i\in[\![1,n]\!]})$. We extend the partially-defined Mealy machines described in Section 7.2.3 to generalise the strategies provided in Example 7.2. Once again, we build on NE outcomes with simple decompositions provided by Lemma 7.3 and Lemma 7.4.

First, let us comment on the use of a different construction than in games with reachability objectives. Example 7.2 highlights that it may be necessary to commit to punishing strategies when the current segment of the decomposition of the intended outcome is left. Therefore, it is not sufficient to simply extend the construction used for Theorem 7.7 (i.e., freezing memory updates outside of the current segment) to monitor and punish players whose targets are visited.

We modify the construction of Theorem 7.7 as follows. We change the approach in such a way that players commit to punishing any player who exits the current segment of the intended outcome. When the current segment is left, if the memory state is of the form $(\mathcal{P}_i, j)$, the memory switches to a newly introduced memory state $\mathcal{P}_i$ that is never left. This switch can only occur if $\mathcal{P}_i$ deviates. The next-move function, for this memory state, assigns moves from a punishing strategy obtained from the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \mathsf{SPath}_w^{T_i})$ by Theorem 6.5, chosen to hinder $\mathcal{P}_i$ enough to ensure that in case of a deviation, the cost of $\mathcal{P}_i$ is at least that of the original outcome.

The conditions imposed on outcomes of Lemma 7.4 (notably condition (iii)) and the characterisation of Theorem 6.9 imply the correctness of this construction. Condition (iii) of Lemma 7.4 ensures that a player cannot reach their target with a lesser cost by changing the order in which states of the segment are visited, whereas the characterisation of Theorem 6.9 guarantees that the punishing strategies sabotage deviating players sufficiently.

Theorem 7.9. *Let $\mathcal{G} = (\mathcal{A}, (\mathsf{SPath}_w^{T_i})_{i\in[\![1,n]\!]})$ be a shortest-path game. Let σ' be a pure NE from a state s_0. There exists a pure finite-memory NE σ from s_0 such that $\mathsf{VisPl}^{\mathcal{G}}(\mathsf{Out}_{\mathcal{A}}(\sigma, s_0)) = \mathsf{VisPl}^{\mathcal{G}}(\mathsf{Out}_{\mathcal{A}}(\sigma', s_0))$ and, for all $i \in [\![1, n]\!]$, $\mathsf{SPath}_w^{T_i}(\mathsf{Out}_{\mathcal{A}}(\sigma, s_0)) \leq \mathsf{SPath}_w^{T_i}(\mathsf{Out}_{\mathcal{A}}(\sigma', s_0))$ where each strategy of σ is induced by a move-independent Mealy machine of size of at most $n^2 + 2n$.*

More precisely, we can bound the size of these Mealy machines by

$$n \cdot (|\mathsf{VisPos}^{\mathcal{G}}(\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)) \setminus \{0\}| + 2).$$

Proof. Let $k = |\mathsf{VisPos}^{\mathcal{G}}(\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)) \setminus \{0\}|$. By Lemma 7.3 and Lemma 7.4, there exists an NE outcome π from s_0 which admits a simple segment decomposition $\mathcal{S} = (\mathsf{sg}_1, \ldots, \mathsf{sg}_{k+1})$ satisfying conditions (ii)-(iv) of Lemma 7.4 (these conditions hold trivially if Lemma 7.3 is applicable).

If $\mathsf{VisPl}^{\mathcal{G}}(\pi)$ is non-empty, we let $\theta = \max\{\mathsf{SPath}_w^{T_i}(\pi) \mid i \in \mathsf{VisPl}^{\mathcal{G}}(\pi)\}$, and, otherwise we let $\theta = 1$. For $i \in [\![1, n]\!]$, let τ_{-i} denote a memoryless strategy of the second player in the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \mathsf{SPath}_w^{T_i})$ such that τ_{-i} is uniformly winning on their winning region $W_{-i}(\mathsf{Safe}(T_i))$ in the reachability game $(\mathcal{A}_i, \mathsf{Reach}(T_i))$ and such that τ_{-i} ensures a cost of at least $\min\{\mathsf{Val}_{\mathcal{G}}^i(s), \theta\}$ from any $s \in S$, where $\mathsf{Val}_{\mathcal{G}}^i(s)$ denotes the value of s in $\mathcal{G}_i$. The existence of these strategies is guaranteed by Theorem 6.5.

We extend the Mealy machine of Section 7.2.3. We work with $I = [\![1, n]\!]$ in the following and drop I from the notation to lighten it. Let $i \in [\![1, n]\!]$. We consider the Mealy machine $\mathfrak{M}_i = (M, m_{\mathsf{init}}^{\mathcal{S}}, \mathsf{up}_{\mathfrak{M}}, \mathsf{nxt}_{\mathfrak{M}_i})$. We let $M = M^{\mathcal{S}} \cup \{\mathcal{P}_i \mid i \in [\![1, n]\!]\}$. The functions $\mathsf{up}_{\mathfrak{M}}$ and $\mathsf{nxt}_{\mathfrak{M}_i}$ extend $\mathsf{up}_{\mathfrak{M}}^{\mathcal{S}}$ and $\mathsf{nxt}_{\mathfrak{M}_i}^{\mathcal{S}}$ as follows. For all $(\mathcal{P}_{i'}, j) \in M^{\mathcal{S}}$ and $s \in S$ that does not occur in sg_j, we let $\mathsf{up}_{\mathfrak{M}}((\mathcal{P}_{i'}, j), s) = \mathcal{P}_{i'}$ and, if $s \in S_i$, we let $\mathsf{nxt}_{\mathfrak{M}_i}((\mathcal{P}_{i'}, j), s) = \tau_{-i'}(s)$ if $i' \neq i$ and $\mathsf{nxt}_{\mathfrak{M}_i}((\mathcal{P}_i, j), s)$ is left arbitrary. For all $i' \in [\![1, n]\!]$ and $s \in S$, we let $\mathsf{up}_{\mathfrak{M}}(\mathcal{P}_{i'}, s) = \mathcal{P}_{i'}$ and, if $s \in S_i$, we let $\mathsf{nxt}_{\mathfrak{M}_i}(\mathcal{P}_{i'}, s) = \tau_{-i'}(s)$ if $i' \neq i$ and $\mathsf{nxt}_{\mathfrak{M}_i}(\mathcal{P}_i, s)$ is left arbitrary.

We let σ_i denote the strategy induced by $\mathfrak{M}_i$. We have $|M| = n \cdot (k + 2)$, therefore the memory size of σ_i satisfies the announced bounds. Furthermore, it follows from Lemma 7.6 that the outcome of $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$ from s_0 is π and σ_i is based on $\mathcal{S}$ for all $i \in [\![1, n]\!]$.

We now argue that σ is an NE from s_0. Let $i \in [\![1, n]\!]$. Let π' be a play starting in s_0 consistent with $\sigma_{-i} = (\sigma_{i'})_{i' \neq i}$. To end the proof, it suffices to show that we have $\mathsf{SPath}_w^{T_i}(\pi') \geq \mathsf{SPath}_w^{T_i}(\pi)$.

We first show the following claim. If some prefix of π' is not coherent with $\mathcal{S}$, then there exists $\ell \in \mathbb{N}$ such that $\pi'_{\leq \ell}$ is the longest prefix of π' coherent with $\mathcal{S}$ and $\pi'_{\geq \ell}$ is a play that is consistent with τ_{-i}. Assume that some prefix of π' is not coherent with $\mathcal{S}$. Let $\ell \in \mathbb{N}$ such that $\pi'_{\leq \ell}$ is the longest prefix of π' coherent with $\mathcal{S}$, and assume that it is j-coherent. As the strategies of σ_{-i} are based on $\mathcal{S}$, we must have $\mathsf{first}(\pi'_{\geq \ell}) \in S_i$. Lemma 7.6 and the definition of $\mathsf{up}_{\mathfrak{M}}$ ensure that $\widehat{\mathsf{up}_{\mathfrak{M}}}(\pi'_{\leq \ell}) = (\mathcal{P}_i, j)$. It follows from $\pi'_{\leq \ell+1}$ being inconsistent with $\mathcal{S}$ that its last state does not occur in sg_j. The definitions of $\mathsf{up}_{\mathfrak{M}}$ and $\mathsf{nxt}_{\mathfrak{M}_{i'}}$ for $i' \neq i$ combined with the above ensure that $\pi'_{\geq \ell}$ is consistent with τ_{-i}.

We now show that $\mathsf{SPath}^{T_i}_w(\pi') \geq \mathsf{SPath}^{T_i}_w(\pi)$. We first assume that $i \notin \mathsf{VisPl}^{\mathcal{G}}(\pi)$. We establish that $\pi' \notin \mathsf{Reach}(T_i)$. By the characterisation of NE outcomes of Theorem 6.9, all states occurring in π belong to $W_{-i}(\mathsf{Safe}(T_i))$. Therefore, if all prefixes of π' are coherent with $\mathcal{S}$, as all states of π' occur in π, it holds that T_i is not visited in π'. Otherwise, let $\ell \in \mathbb{N}$ such that $\pi'_{\leq \ell}$ is the longest prefix of π' that is coherent with $\mathcal{S}$ and $\pi'_{\geq \ell}$ is consistent with τ_{-i}. No states of T_i occur in $\pi'_{\leq \ell}$ by coherence with $\mathcal{S}$. It follows from the coherence of $\pi'_{\leq \ell}$ with $\mathcal{S}$ that $\mathsf{first}(\pi'_{\geq \ell}) = \mathsf{last}(\pi'_{\leq \ell})$ occurs in π. We obtain that $\mathsf{first}(\pi'_{\geq \ell}) \in W_{-i}(\mathsf{Safe}(T_i))$, therefore $\pi_{\geq \ell} \notin \mathsf{Reach}(T_i)$. We have shown that for all $i \notin \mathsf{VisPl}^{\mathcal{G}}(\pi)$, we have $\mathsf{SPath}^{T_i}_w(\pi') = \mathsf{SPath}^{T_i}_w(\pi) = +\infty$.

We now assume that $i \in \mathsf{VisPl}^{\mathcal{G}}(\pi)$. The desired inequality is immediate if T_i is not visited in π'. Similarly, it holds directly if $s_0 \in T_i$. We therefore assume that we are in neither of the previous two cases. We write the shortest prefix of π' ending in T_i (the weight of which is $\mathsf{SPath}^{T_i}_w(\pi')$) as a combination $h \cdot h'$ where h is its longest prefix that is coherent with $\mathcal{S}$. We note that h' is consistent with τ_{-i} because h is a prefix of the longest prefix of π' coherent with $\mathcal{S}$: if h is a strict prefix, then h' is a trivial (i.e., one-state) history, and otherwise it follows from the above.

We provide lower bounds on the weights of h and h'. Assume that h is j-coherent. By definition of coherence, we can write h as a history combination $h_1 \cdot \ldots \cdot h_j$ where, for all $j' < j$, $h_{j'}$ shares its first and last states with $\mathsf{sg}_{j'}$ and contains only states of $\mathsf{sg}_{j'}$, and h_j shares its first state with sg_j and contains only states of sg_j. Let sg'_j be the prefix of sg_j up to $\mathsf{last}(h_j)$.

On the one hand, we have $\sum_{j'<j} w(h_{j'}) \geq \sum_{j'<j} w(\mathsf{sg}_{j'})$ and $w(h_j) \geq w(\mathsf{sg}'_j)$. This follows from π satisfying property (iii) of Lemma 7.4 (for h_j, having $w(h_j) < w(\mathsf{sg}'_j)$ would contradict property (iii) with respect to sg_j). On the other hand, by choice of τ_{-i}, we obtain that $w(h') \geq \min\{\mathsf{Val}^i_{\mathcal{G}}(\mathsf{first}(h')), \theta\}$. From the characterisation of Theorem 6.9, we obtain $\mathsf{Val}^i_{\mathcal{G}}(\mathsf{first}(h')) \geq \mathsf{SPath}^{T_i}_w(\pi) -$

$w(\mathsf{sg}_1 \cdot \ldots \cdot \mathsf{sg}_{j-1} \cdot \mathsf{sg}'_j)$. It follows from $\mathsf{SPath}_w^{T_i}(\pi) \geq \theta$ that $w(h') \geq \mathsf{SPath}_w^{T_i}(\pi) - w(\mathsf{sg}_1 \cdot \ldots \cdot \mathsf{sg}_{j-1} \cdot \mathsf{sg}'_j)$. By combining all of the above inequalities, we obtain $\mathsf{SPath}_w^{T_i}(\pi') \geq \mathsf{SPath}_w^{T_i}(\pi)$, ending the proof. □

In this case, Theorem 7.9 provides the memory bound $2n$ if no players visit their target. However, the construction of Theorem 7.7 applies to such NEs in shortest path games. We obtain the following result.

Corollary 7.10. *If there exists a pure NE from s_0 such that no players see their target in its outcome, then there is a pure move-independent finite-memory NE from s_0 such that no players see their target in its outcome such that all strategies are induced by move-independent Mealy machines of size at most n.*

7.3 Büchi games

We now prove that for any NE from a given initial state in a Büchi game on $\mathcal{A}$, we can find a move-independent finite-memory NE from the same initial state where the same objectives are satisfied. We provide several examples in Section 7.3.1. We illustrate that decomposition-based strategies are no longer sufficient for Büchi objectives, and that we cannot obtain arena-independent memory bounds when restricted to move-independent strategies. We build on the techniques of Section 7.2.3 to provide move-independent finite-memory NEs in Section 7.3.3.

Throughout this section, for the sake of simplicity, we extend the definition of simple segment to also include *simple cycles*, i.e., cycles of $\mathcal{A}$ in which only all states occur only once besides the first state, which occurs exactly twice.

7.3.1 Limitations of decomposition-based strategies

For reachability and shortest-path games, we relied on simple segment decompositions between consecutive targets along an NE outcome to obtain finite-memory NEs. Our strategies based on these decompositions do not explicitly punish players who deviate without leaving the current segment. Intuitively, this can pose an issue in a Büchi game as a losing player may have an incentive to loop within a segment that contains one of their targets.

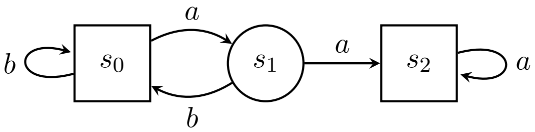


Figure 7.4: An arena where a direct decomposition-based approach fails to obtain an NE.

Example 7.4. Consider the game on the arena depicted in Figure 7.4 where the objectives of $\mathcal{P}_1$ and $\mathcal{P}_2$ are $\mathsf{Büchi}(\{s_1\})$ and $\mathsf{Büchi}(\{s_2\})$ respectively. The play $s_0as_1(as_2)^\omega$ is the outcome of an NE by Theorem 6.8. To mimic the construction underlying Theorems 7.7 and 7.9, we would consider a finite-memory strategy based on the decomposition $\mathcal{S} = (s_0as_1as_2, (s_2a)^\omega)$. However, if $\mathcal{P}_2$ uses a strategy based on $\mathcal{S}$, the memoryless strategy σ_1 of $\mathcal{P}_1$ such that $\sigma_1(s_1) = b$ would be a profitable deviation of $\mathcal{P}_1$: if $\mathcal{P}_1$ uses this strategy, we obtain the outcome $(s_0as_1b)^\omega$, as $\mathcal{P}_2$ would not punish the deviation of $\mathcal{P}_1$. ◁

In the previous example, the issue with the proposed decomposition lies with the occurrence of a target of $\mathcal{P}_1$, whose objective is not satisfied in the intended outcome, within some segment of the decomposition. To circumvent this issue in the next section, we construct strategies that follow two phases. In their first phase, these strategies punish any deviations from the intended outcome. For their second phase, we adapt the strategies of Section 7.2.3. To ensure that no profitable deviations may exist, we start the second phase at a point of the intended outcome from which no more targets of losing players occur.

The following example illustrates that the punishing mechanism used for finite-memory NEs in reachability games does not suffice. In other words, players must commit to punishing strategies once some player exits the current segment in the second phase mentioned above.

Example 7.5. Consider the game on the arena depicted in Figure 7.5 where the objectives of $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ are $\mathsf{Büchi}(\{s_1\})$ and $\mathsf{Büchi}(\{s_2, s_4\})$ and $\mathsf{Büchi}(\{s_4\})$ respectively. The play $\pi = (s_0as_1as_3a)^\omega$ is the outcome of an NE by Theorem 6.8. Consider a $\mathcal{P}_1$ strategy based on the decomposition (π) that uses the punishment mechanism we introduced for reachability games. Then the behaviour of $\mathcal{P}_1$ does not change if $\mathcal{P}_2$ moves from s_0 to s_2 instead of s_1: $\mathcal{P}_1$

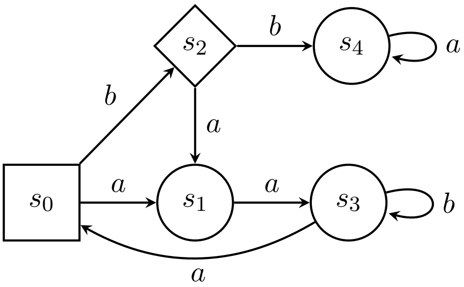


Figure 7.5: An arena where there exists an NE such that players should commit to punishing strategies once a segment of a given decomposition of its outcome is left. The diamond is a $\mathcal{P}_3$ state.

would move from s_1 to s_3 and then to s_0. It follows that $\mathcal{P}_2$ would have a profitable deviation no matter the strategy of $\mathcal{P}_3$.

To obtain an NE where all players use strategies based on the decomposition (π), $\mathcal{P}_1$, must commit to a punishing strategy for $\mathcal{P}_2$ if s_2 is visited. For $\mathcal{P}_2$ and $\mathcal{P}_3$ we consider the memoryless strategies σ_2 and σ_3 such that $\sigma_2(s_0) = \sigma_3(s_2) = a$. It is easy to check that this is an NE.

We remark that there is no NE from s_0 using the construction we had used in reachability games in this game such that the objective of $\mathcal{P}_1$ is satisfied in the outcome of the NE. $\lhd$

In the two-phase approach described above, players must precisely enforce a specific segment of a play during the first phase of the strategy (profile). This results in move-independent finite-memory NEs with a size that is dependent on the arena. We provide an example that proves that arena-independent memory bounds cannot be obtained in general for move-independent solutions to the constrained existence NE problem in Büchi games. We illustrate it with a generalisation of the game of Example 7.4.

Example 7.6 (Arena-dependence of memory size for move-independent NEs in Büchi games)**.** Our argument is based on a family of two-player turn-based deterministic arenas $(\mathcal{A}^p)_{p\geq 1}$. We fix $p \in \mathbb{N}_{>0}$ for the remainder of the example. We define $\mathcal{A}^p = ((S_1^p, S_2^p), A^p, \delta^p)$ as follows. We let $S_1^p = \{t_1, \ldots, t_p\}$, $S_2^p = \{s_1, \ldots, s_{p+1}\}$ and $A = \{a_=, a_+\} \cup \{a_q \mid q \in [\![1, p+1]\!]\}$. The transition function $\delta\colon S \times A \to S$ is defined by the three following rules. Let $q \in [\![1, p]\!]$. From s_q,

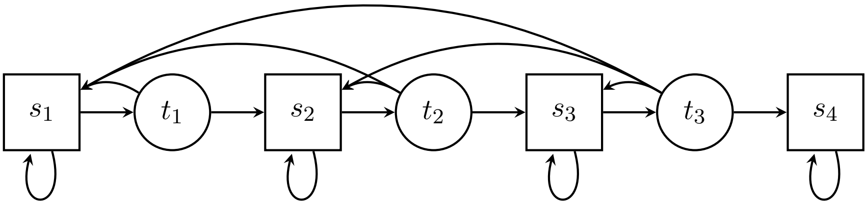


Figure 7.6: The graph underlying the arena $\mathcal{A}^3$ of Example 7.6, in which $\mathcal{P}_2$ needs a memory of size at least 3 in any NE from s_1 in which $\mathcal{P}_2$ wins. Circles and squares are resp. $\mathcal{P}_1$ and $\mathcal{P}_2$ vertices.

there is a self-loop labelled by $a_=$ and a transition to t_q labelled by a_+. There also is an $a_=$-labelled self-loop in s_{p+1}. Finally, for all $q' \in [\![1, q+1]\!]$, we have $\delta^p(t_q, a_{q'}) = s_{q'}$. We illustrate $\mathcal{A}^3$ (without action labels) in Figure 7.6, and remark that $\mathcal{A}^1$ matches the arena of Figure 7.4 up to a renaming of the states and actions.

We now define a game on $\mathcal{A}^p$. We define $T_1^p = S_1^p$ and $T_2^p = \{s_{p+1}\}$, and consider the Büchi game $\mathcal{G}^p = (\mathcal{A}^p, (\mathsf{Büchi}(T_1^p), \mathsf{Büchi}(T_2^p)))$. We prove that, in $\mathcal{G}^p$,

(i) there exists a pure move-independent finite-memory NE $\sigma = (\sigma_1, \sigma_2)$ from s_1 such that $\mathsf{Out}_{\mathcal{A}}(\sigma, s_1) \in \mathsf{Büchi}(T_2^p)$, σ_1 is memoryless and σ_2 is induced by a Mealy machine with at most $p+1$ states and

(ii) for all pure move-independent strategy profiles $\sigma = (\sigma_1, \sigma_2)$ such that which σ_2 is induced by a Mealy machine with at most p memory states, σ is not an NE from s_1 such that $\mathsf{Out}_{\mathcal{A}}(\sigma, s_1) \in \mathsf{Büchi}(T_2^p)$.

It follows from these two points that arena-dependent memory is necessary for move-independent constrained NEs in Büchi games.

We start with claim (i). For $\mathcal{P}_1$, we consider the memoryless strategy σ_1 such that for all $q \leq p$, $\sigma_1(t_q) = a_{q+1}$. For $\mathcal{P}_2$, we consider the strategy σ_2 induced by the Mealy machine $\mathfrak{M} = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ defined as follows. We let $M = [\![1, p+1]\!]$ and $m_{\mathsf{init}} = 1$.

Memory updates follow three rules. First, the memory is not updated in s_{p+1} and in $\mathcal{P}_1$ states. Formally, for all $q \in M$ and $q' \in [\![1, p]\!]$, we let

$\mathsf{up}_{\mathfrak{M}}(q, s_{p+1}) = \mathsf{up}_{\mathfrak{M}}(q, t_{q'}) = q$. Second, if the index of the current $\mathcal{P}_2$ game state does not match the memory state, we do not update the memory, i.e., for all $q, q' \in [\![1, p]\!]$ such that $q \neq q'$, we let $\mathsf{up}_{\mathfrak{M}}(q, s_{q'}) = q$. Finally, for all $q \in [\![1, p]\!]$, if in memory state q and game state s_q, we increment the memory state, i.e., we let $\mathsf{up}_{\mathfrak{M}}(q, s_q) = q + 1$.

For the next-move function, we take the self-loop if the memory state is not the index of the current state, and otherwise take the other action. Formally, for all $q \in M$ and $q' \in [\![1, p]\!]$, w e let $\mathsf{nxt}_{\mathfrak{M}}(q, s_{q'}) = a_+$ if $q = q'$ and $\mathsf{nxt}_{\mathfrak{M}}(q, s_{q'}) = a_=$ otherwise. Intuitively, σ_2 moves rightward (with respect to the depiction of the arena in Figure 7.6) at each step so long as $\mathcal{P}_1$ does so, and stops progressing as soon as $\mathcal{P}_1$ moves to the left.

To lighten notation, in the remainder of this example, we omit actions from plays and histories of $\mathcal{A}^p$: the sequence of states of a play uniquely determines the sequence of actions due to the structure of $\mathcal{A}^p$.

The outcome of σ_1 and σ_2 from s_1 is $s_1 t_1 \dots s_p t_p s_{p+1}^\omega$, which is winning for $\mathcal{P}_2$. To argue that (σ_1, σ_2) is an NE, it suffices to argue that $\mathcal{P}_1$ does not have a profitable deviation. Let τ_1 be an arbitrary strategy of $\mathcal{P}_1$. We consider two cases. First, assume that for all $q \in [\![1, p]\!]$, for $h_q = s_1 t_1 \dots s_q t_q$, we have $\tau_1(h_q) = a_{q+1}$. In this case, the outcome from s_1 of (τ_1, σ_2) matches that of (σ_1, σ_2), hence τ_1 is not a profitable deviation of $\mathcal{P}_1$. Second, assume that for some $q \in [\![1, p]\!]$, we have $\tau_1(h_q) = a_{q'}$ for some $q' < q + 1$. We consider the smallest such q. It follows from a straightforward induction that $\widehat{\mathsf{up}_{\mathfrak{M}}}(h_q) = q + 1$. By definition of $\mathfrak{M}$, we obtain that the outcome of τ_1 and σ_2 from s_1 is the play $h_q s_{q'}^\omega$, hence τ_1 is not a profitable deviation in this case either. This shows that (σ_1, σ_2) is an NE from s_1.

We now prove claim (ii). We let $\sigma = (\sigma_1, \sigma_2)$ be a strategy profile such that its outcome from s_1 is winning for $\mathcal{P}_2$ and such that σ_2 is given by a Mealy machine $\mathfrak{M} = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ with memory size at most p. We establish that $\mathcal{P}_1$ has a profitable deviation if the initial vertex is s_1.

We first observe that all vertices of $\mathcal{A}^p$ occur in $\mathsf{Out}_{\mathcal{A}}(\sigma, s_1)$; the arena is such that reaching s_{p+1} from s_1 implies so. For all $q \in [\![1, p]\!]$, we let h_q be the shortest prefix of $\mathsf{Out}_{\mathcal{A}}(\sigma, s_1)$ that ends in t_q, and let h_0 denote the empty word. It follows from our assumption on the memory size of $\mathfrak{M}$ that there exist $q < q' \in [\![1, p]\!] \cup \{0\}$ such that $\widehat{\mathsf{up}_{\mathfrak{M}}}(h_q) = \widehat{\mathsf{up}_{\mathfrak{M}}}(h_{q'})$. Let $h = v_0 \dots v_r$ be the non-empty history such that $h_{q'} = h_q h$. We show that $h_q h^\omega$ is consistent with σ_2, which implies the existence of a profitable deviation for $\mathcal{P}_1$.

The history $h_q h$ is a prefix of $\mathsf{Out}_{\mathcal{A}}(\sigma, s_1)$, thus it is consistent with σ_2. We continue by induction. Assume that $h_q h^z v_0 \dots v_{\ell+1}$ (with $\ell + 1 \leq r$) is consistent with σ_2 and that $\widehat{\mathsf{up}_{\mathfrak{M}}}(h_q h^z v_0 \dots v_\ell) = \widehat{\mathsf{up}_{\mathfrak{M}}}(h_q v_0 \dots v_\ell)$. We remark that these properties hold for the case $z = 0$ and $\ell = r - 1$, which we consider to be the base case. We show that same property holds for $\ell + 1$. In the following, we abusively let v_{r+1} denote v_0 to avoid treating the case $\ell + 1 = r$ separately.

We have $\widehat{\mathsf{up}_{\mathfrak{M}}}(h_q h^z v_0 \dots v_\ell v_{\ell+1}) = \widehat{\mathsf{up}_{\mathfrak{M}}}(h_q v_0 \dots v_\ell v_{\ell+1})$ directly from the induction hypothesis and the definition of $\widehat{\mathsf{up}_{\mathfrak{M}}}$. We consider two cases for the consistency. First, we assume that $v_{\ell+1} \in S_1^p$. The consistency of the relevant history follows directly by induction as $\mathcal{P}_2$ does not select the last transition. Second, we assume that $v_{\ell+1} \in S_2^p$. The induction hypothesis implies that $h_q h^z v_0 \dots v_{\ell+1} v_{\ell+2}$ is consistent with σ_2 if and only if the action labelling the transition from $v_{\ell+1}$ to $v_{\ell+2}$ is $\mathsf{nxt}_{\mathfrak{M}}(\widehat{\mathsf{up}_{\mathfrak{M}}}(h_q h^z v_0 \dots v_\ell), v_{\ell+1})$. By consistency of $h_q h$ with σ_2, the action $\mathsf{nxt}_{\mathfrak{M}}(\widehat{\mathsf{up}_{\mathfrak{M}}}(h_q v_0 \dots v_\ell), v_{\ell+1})$ is the sought action. By combining this with the induction hypothesis on memory updates, we conclude the required consistency claim.

We have shown that in the game $\mathcal{G}^p$, any move-independent NE from s_1 with an outcome that is winning for $\mathcal{P}_2$ requires $\mathcal{P}_2$ to have a memory size of at least p, which is roughly half of the number of vertices in the game. This ends our illustration that arena-dependent memory is required for general constrained move-independent NEs in Büchi games. $\lhd$

7.3.2 Simple Nash equilibria outcomes

In a Büchi game, there need not exist NE outcomes with a finite simple decomposition as we have shown in games with reachability objectives and shortest-path costs. We consider two ways of simplifying NE outcomes in Büchi games depending on the form of these outcomes.

First, we consider NE outcomes such that some state occurs infinitely often in it. The following example illustrates that we cannot transform these outcomes into plays with a finite simple decomposition in general.

Example 7.7. In the two-player arena $\mathcal{A}$ depicted in Figure 7.7, if we consider the game in which the objectives of $\mathcal{P}_1$ and $\mathcal{P}_2$ are respectively $\mathsf{Büchi}(\{s_1\})$ and

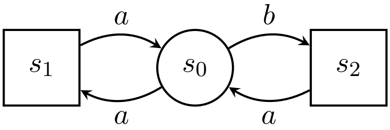


Figure 7.7: A two-player arena where there is no outcome from s_0 with a finite simple decomposition satisfying the objectives $\mathsf{Büchi}(\{s_1\})$ and $\mathsf{Büchi}(\{s_2\})$.

$\mathsf{Büchi}(\{s_2\})$, it is easy to see that the play $\pi = (s_0as_1as_0bs_2a)^\omega$ is the outcome of a pure NE from s_0 as both players satisfy their objective.

We observe that there is no play of $\mathcal{A}$ with a finite simple decomposition that visits both s_1 and s_2 infinitely often; simple lassos of $\mathcal{A}$ can only loop in one of s_1 or s_2. We remark that π admits the ultimately periodic simple decomposition $(s_0as_1, s_1as_0bs_2, s_2as_0as_1, s_1as_0bs_2, s_2as_0as_1, \ldots)$. We will build move-independent NEs on plays that admit such decompositions. ◁

We now prove that from any NE outcome of $\mathcal{G}$ in which a state occurs infinitely often, we can derive an NE outcome from the same initial state that admits an ultimately periodic simple decomposition. We choose the first segment of this decomposition to correspond to the first phase of the two-phase approach mentioned in the previous section: no targets of losing players must appear outside of this first segment.

The idea of the following proof is as follows. There is $\ell \in \mathbb{N}$ such that no targets of losing players occur in $\pi_{\geq \ell}$ by definition of the Büchi objective. Furthermore, due to the presence of an infinitely occurring state in π, for all $i \in [\![1, n]\!]$ such that $\pi \in \mathsf{Büchi}(T_i)$, there is $t_i \in T_i$ such that all these states are connected by simple histories or simple cycles that traverse only states in $\pi_{\geq \ell}$. Our desired decomposition is obtained by letting its first segment be a simple history starting in s_0 up to some t_i and then selecting the other segments to be the simple segments mentioned above such that all of the relevant targets appears in their concatenation. Through this approach, we obtain an NE outcome with an ultimately periodic decomposition $\mathcal{S} = (\mathsf{sg}_j)_{j \in \mathbb{N}}$ of period no more than n.

Lemma 7.11. *Let π' be the outcome of an NE from $s_0 \in S$ in $\mathcal{G}$ such that some state occurs infinitely often in π' and let $k = |\{i \in [\![1, n]\!] \mid \pi' \in \mathsf{Büchi}(T_i)\}|$. There exists an NE outcome π from s_0 in $\mathcal{G}$ such that, for all $i \in [\![1, n]\!]$, $\pi \in \mathsf{Büchi}(T_i)$ if and only if $\pi' \in \mathsf{Büchi}(T_i)$, and π admits an infinite simple segment decomposition $(\mathsf{sg}_0, \mathsf{sg}_1, \dots)$ such that*

(i) for all $j \geq 1$ and all $i \in [\![1, n]\!]$, if $\pi \notin \mathsf{Büchi}(T_i)$, then no state of T_i occurs in sg_j and

(ii) for all $j \geq 1$, $\mathsf{sg}_j = \mathsf{sg}_{j+\max\{k,1\}}$.

Proof. We assume without loss of generality that $k \geq 1$, i.e., that at least one Büchi objective of $\mathcal{G}$ is satisfied. If this is not the case, we can add a new player controlling no states whose target is S to enforce this assumption. For convenience of notation, we assume that π satisfies $\mathsf{Büchi}(T_1)$, …, $\mathsf{Büchi}(T_k)$.

We fix $\ell \in \mathbb{N}$ such that for all $i > k$, no states of T_i occur in $\pi_{\geq \ell}$. For all $i \in [\![k]\!]$, we fix $t_i \in T_i$ which appears in $\pi_{\geq \ell}$. We define the decomposition $\mathcal{S}$ as follows. We let sg_0 be a simple history from s_0 to t_1 using only vertices from π (we tolerate the history s_0 if $s_0 = t_1$). For all $1 \leq j < k$, we let sg_j be a simple history or cycle from t_j to t_{j+1}. Finally, we let sg_k be a simple segment from t_k to t_1. Other segments are defined so condition (ii) holds. By construction, the π generated from this decomposition satisfies condition (i). Theorem 6.8 ensures that π is the outcome of an NE. □

When dealing with a play in an infinite arena, there need not be a state occurring infinitely often within the play. This is the case, e.g., in arenas devoid of cycles. We provide an example illustrating a game in which an NE in which all players win cannot have an outcome with a finite simple decomposition, nor with an ultimately periodic decomposition.

Example 7.8. We consider the two-player arena $\mathcal{A}$ (where all states are controlled by $\mathcal{P}_1$) depicted in Figure 7.8, and the game $\mathcal{G} = (\mathcal{A}, (\mathsf{Büchi}(\{t_\ell \mid \ell \in 2\mathbb{N}\}), \mathsf{Büchi}(\{t_\ell \mid \ell \in 2\mathbb{N} + 1\})))$. On the one hand, we observe that the play π obtained by playing the sequence of actions $(b^2a)^\omega$ is the outcome of an NE from s_0 in $\mathcal{G}$ as both objectives are satisfied.

We now show that no play of $\mathcal{A}$ that admits a finite or ultimately periodic

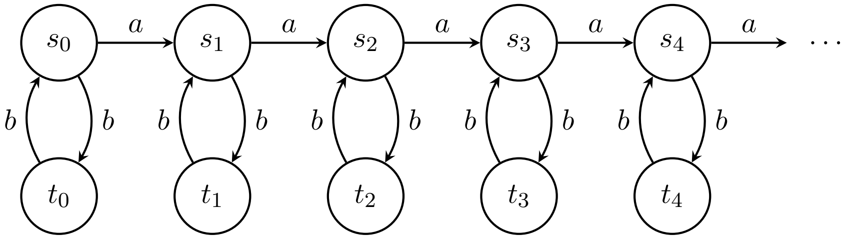


Figure 7.8: A two-player arena where there is no outcome from s_0 with a finite simple decomposition satisfying the objectives $\mathsf{Büchi}(\{t_\ell \mid \ell \in 2\mathbb{N}\})$ and $\mathsf{Büchi}(\{t_\ell \mid \ell \in 2\mathbb{N}+1\})$.

simple decomposition satisfies both Büchi objectives of $\mathcal{G}$. A play admitting an ultimately periodic simple decomposition is a lasso. Similarly, a play admitting a finite simple decomposition has a simple play or simple lasso as a suffix. In both cases, it is not possible to visit both targets infinitely often due to the structure of $\mathcal{A}$.

The play π admits the infinite simple decomposition $\mathcal{S} = (\mathsf{sg}_j)_{j\in\mathbb{N}}$ of π defined by $\mathsf{sg}_0 = s_0bt_0$ and, for all $j \in \mathbb{N}_{>0}$, $\mathsf{sg}_j = t_{j-1}bs_{j-1}as_jbt_j$, consisting of simple histories connecting consecutive target visits along π. An important property of $\mathcal{S}$ that we will use in the sequel to construct finite-memory NE is that no two segments of $\mathcal{S}$ with an index of the same parity have states in common. $\lhd$

We now explain how to derive NE outcomes with decompositions similar to that of the previous example from arbitrary NE outcomes. Let $\pi \in \mathsf{Plays}(\mathcal{A})$ be the outcome of an NE from $s_0 \in S$ in $\mathcal{G}$ such that no state occurs infinitely often in π. We sketch how to derive an NE outcome from π that has an infinite simple segment decomposition $\mathcal{S} = (\mathsf{sg}_j)_{j\in\mathbb{N}}$ such that for all odd (resp. even) $j \neq j'$, no state of sg_j occurs in $\mathsf{sg}_{j'}$ and, for all $j \geq 1$, no targets of players whose objective is not satisfied by π occurs in sg_j.

Once again, we aim to use sg_0 of $\mathcal{S}$ to implement the first phase of our two-phase mechanism. We obtain sg_0 similarly to the case of outcomes with states occurring infinitely often: there exists a position $\ell \in \mathbb{N}$ such that no states of T_i occur in $\pi_{\geq \ell}$ for all $i \in [\![1, n]\!]$ such that $\pi \notin \mathsf{Büchi}(T_i)$. We let $\ell_0 \geq \ell$ be such that $s_{\ell_0} \in T_i$ for some $i \in [\![1, n]\!]$ such that $\pi \in \mathsf{Büchi}(T_i)$, and choose sg_0 to be a simple history that shares its first and last states with $\pi_{\leq \ell_0}$

and that uses only states occurring in this prefix.

The other segments are constructed by induction. We explain how sg_1 is defined from ℓ_0 to illustrate the idea of the general construction. As no states appear infinitely often in π, there exists some position $\ell_1 > \ell_0$ such that no state of $\pi_{\leq \ell_0}$ occurs in $\pi_{\geq \ell_1}$ and $s_{\ell_1} \in T_i$ for some $i \in [\![1, n]\!]$ such that $\pi \in \mathsf{Büchi}(T_i)$. We let sg_1 be a simple history that starts in $\mathsf{last}(\mathsf{sg}_0)$, ends in $\mathsf{last}(\pi_{\leq \ell_1})$ and uses only states that occur in the segment of π between positions ℓ_0 and ℓ_1. If we construct sg_2 similarly from ℓ_1 (by induction), then it shares no states with sg_0 by choice of ℓ_1. Proceeding with this inductive construction while ensuring that all targets that are visited infinitely often in π occur infinitely often in the decomposition being constructed, we obtain the desired decomposition. Furthermore, the play described by this decomposition is an NE outcome the characterisation of NE outcomes of Theorem 6.8. We formalise the sketch above in the following proof.

Lemma 7.12. *Let π' be the outcome of an NE from $s_0 \in S$ in $\mathcal{G}$ such that no state occurs infinitely often in π'. There exists an NE outcome π from s_0 such that, for all $i \in [\![1, n]\!]$, $\pi \in \mathsf{Büchi}(T_i)$ if and only if $\pi \in \mathsf{Büchi}(T_i)$, and π admits an infinite simple segment decomposition $\mathcal{S} = (\mathsf{sg}_j)_{j \in \mathbb{N}}$ such that*

(i) for all $j \geq 1$ and all $i \in [\![1, n]\!]$, if $\pi \notin \mathsf{Büchi}(T_i)$, then no state of T_i occurs in sg_j and

(ii) for all $j \neq j'$, sg_j and $\mathsf{sg}_{j'}$ have no states in common if j and j' have the same parity.

Proof. We let $\pi' = s_0 a_0 s_1 \ldots$ We assume without loss of generality that there exists $i \in [\![1, n]\!]$ such that $\pi \in \mathsf{Büchi}(T_i)$ (this can be enforced by adding a player for whom all states are targets if necessary). For convenience of notation, we assume that $\{i \in [\![1, n]\!] \mid \pi \in \mathsf{Büchi}(T_i)\} = [\![1, k]\!]$ where $k = |\{i \in [\![1, n]\!] \mid \pi \in \mathsf{Büchi}(T_i)\}| \geq 1$. We define the sought outcome π via an infinite segment decomposition.

We let $\ell \in \mathbb{N}$ such that for all $i > k$, no states of T_i occur in $\pi_{\geq \ell}$. There exists some position $\ell_0 \geq \ell$ such that $s_{\ell_0} \in T_1$. We let sg_0 be a simple history from s_0 to s_{ℓ_0} that only traverses states occurring in $\pi'_{\leq \ell_0}$.

We now assume that segments $\mathsf{sg}_0, \ldots, \mathsf{sg}_j$ and positions $\ell_0, \ldots, \ell_j$ are defined. We assume by induction that (a) for all $j' \leq j$, $\mathsf{last}(\mathsf{sg}_{j'}) \in T_{j' \bmod k+1}$, (b) for all $j' \leq j$, $\mathsf{sg}_{j'}$ contains only states occurring in $\pi'_{\leq \ell_{j'}}$, (c) for all $j' < j$, no states of $\pi'_{\leq \ell_{j'}}$ occur in $\pi'_{\geq \ell_{j'+1}}$, and (d) for all $j'' \leq j' - 2$, $\mathsf{sg}_{j'}$ and $\mathsf{sg}_{j''}$ have no states in common.

We define ℓ_{j+1} and sg_{j+1} as follows. There exists some $\ell \in \mathbb{N}$ such that no state of $\pi'_{\leq \ell_j}$ occurs in $\pi'_{\geq \ell}$. We choose $\ell_{j+1} > \ell$ such that $s_{\ell_{j+1}} \in T_{(j'+1) \bmod k+1}$. We let sg_{j+1} be a simple history from s_{ℓ_j} to $s_{\ell_{j+1}}$ that uses only states occurring in the segment $s_{\ell_j} a_{\ell_j} \ldots s_{\ell_{j+1}}$ of π'.

We argue that the induction hypothesis is preserved by this choice. Properties (a), (b) and (c) hold by definition. We show that (d) holds, i.e., that for all $j' \leq j-1$, sg_{j+1} and $\mathsf{sg}_{j'}$ have no states in common. Let $j' \leq j-1$. It holds that the states occurring in sg_{j+1} all appear in $\pi'_{\geq \ell_j}$. By induction, all states of $\mathsf{sg}_{j'}$ occur in $\pi'_{\leq \ell_{j'}}$, and none of these states occur in $\pi'_{\geq \ell_j}$. This ends the inductive construction.

It remains to prove that the play $\pi = \mathsf{sg}_0 \cdot \mathsf{sg}_1 \cdot \ldots$ is the outcome of an NE. This is immediate by Theorem 6.8 as all states appearing in π appear in π' and, by construction, π and π' satisfy the same Büchi objectives of $\mathcal{G}$. □

7.3.3 Finite-memory Nash equilibria

In this section, we prove the following theorem by considering the two cases distinguished in the previous section.

Theorem 7.13. *Let $\mathcal{G} = (\mathcal{A}, (\mathsf{Büchi}(T_i))_{i \in [\![1,n]\!]})$ be a Büchi game. Let σ' be a pure NE from a state $s_0 \in S$. There exists a finite-memory pure NE σ from s_0 such that all strategies of σ are induced by move-independent Mealy machines and, for all $i \in [\![1,n]\!]$, $\mathsf{Out}_{\mathcal{A}}(\sigma, s_0) \in \mathsf{Büchi}(T_i)$ if and only if $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0) \in \mathsf{Büchi}(T_i)$.*

We build finite-memory move-independent NEs from NE outcomes and decompositions obtained through Lemma 7.11 and Lemma 7.12. In both cases, we construct strategies following two phases from these decompositions. In the first phase, players react and punish any deviation from the sequence of states of the first segment of the decomposition (i.e., if different actions are used but they lead to the same successor, no punishment is used). This can be achieved by adding all states of the segment in memory. A deviation can be detected by checking whether the current state of the play is the current state in the memory until the first segment is completed. In the second phase, the players follow a strategy based on the suffix of the decomposition that excludes the first segment, implemented similarly to the partially-defined strategies of Section 7.2.3 that we had used for games with reachability objectives and shortest-path costs. In practice, the constructions for the second phase differ slightly for each type of outcome.

First, we consider an NE σ' from s_0 in $\mathcal{G}$ such that a state occurs infinitely often in $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)$. By Lemma 7.11, we can derive an NE outcome π from s_0 such that π and $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)$ satisfy the same objectives of $\mathcal{G}$, and π admits a simple segment decomposition $\mathcal{S} = (\mathsf{sg}_0, \mathsf{sg}_1, \ldots, \mathsf{sg}_k, \mathsf{sg}_1, \ldots, \mathsf{sg}_k, \ldots)$ where no targets of losing players occur in segments $\mathsf{sg}_1, \ldots, \mathsf{sg}_k$ of $\mathcal{S}$.

We only describe the second phase of the two-phase approach described above. For the second phase, we switch to a strategy that is based on the periodic decomposition $\mathcal{S}' = (\mathsf{sg}_1, \mathsf{sg}_2, \ldots, \ldots)$ of period k. We slightly adapt the construction presented in Section 7.2.3 to loop back to the memory states for the first segment at the end of sg_k. More precisely, when reading $\mathsf{last}(\mathsf{sg}_k)$ in memory states of the form $(\mathcal{P}_i, k)$, we update the memory to an appropriate memory state of the form $(\mathcal{P}_{i'}, 1)$.

By completing the above behaviour with switches to memoryless punishing strategies (Theorem 6.2) if sg_0 is not accurately simulated or if a player exits the current segment, we obtain a finite-memory NE. The stability of the NE follows from the characterisation of Theorem 6.8 for deviations that induce the use of punishing strategies and the property that no targets of losing players occur in segments sg_j, $j \geq 1$ for other (in-segment) deviations. We formally present these finite-memory strategies in the following proof.

Lemma 7.14. *Let σ' be a pure NE from a state s_0 such that some state occurs infinitely often in $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)$. There exists a finite-memory pure NE σ from s_0 such that all strategies of σ are induced by move-independent Mealy machines and, for all $i \in [\![1, n]\!]$, $\pi \in \mathsf{Büchi}(T_i)$ if and only if $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0) \in \mathsf{Büchi}(T_i)$.*

If $\mathcal{A}$ is finite, move-independent Mealy machines of size at most $|S| + n^2 + n$ suffice to implement the strategies of σ.

Proof. Let π be an NE outcome obtained via Lemma 7.11 from $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)$. We let $k = \max\{1, |\{i \in [\![1, n]\!] \mid \pi \in \mathsf{Büchi}(T_i)\}|\}$ and $(\mathsf{sg}_0, \mathsf{sg}_1, \ldots)$ be the decomposition provided by the lemma. We argue that π is the outcome of a finite-memory NE. Before constructing the Mealy machines, we first introduce some notation. Let $I \subseteq [\![1, n]\!]$ be $\{i \in [\![1, n]\!] \mid \pi \notin \mathsf{Büchi}(T_i)\}$ if this set is not empty, or $\{1\}$ otherwise. For all $i \in I$, we let τ_{-i} be a memoryless uniformly winning strategy for the second player of the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \mathsf{Büchi}(T_i))$ (it exists by Theorem 6.2) and let $W_{-i}(\mathsf{coBüchi}(T_i))$ denote the winning region of this player in $\mathcal{G}_i$. We write $\mathsf{sg}_0 = s_0 a_0 \ldots s_r$. Finally, we define $S_I = \bigcup_{i \in I} S_i$.

For each $i \in [\![1, n]\!]$, we define a Mealy machine $\mathfrak{M}_i = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}_i}, \mathsf{up}_{\mathfrak{M}})$ as follows. We define M as the union of sets

$$\{s_\ell \in S \mid 0 \leq \ell \leq r\} \cup (\{\mathcal{P}_i \mid i \in I\} \times [\![1, k]\!]) \cup \{\mathcal{P}_i \mid i \in I\}$$

and $m_{\mathsf{init}} = s_0$. The memory states that are game states correspond to the first phase in the sketch above, and the others to the second phase. We note that the memory bounds claimed for finite arenas follow from sg_0 being simple.

The update function $\mathsf{up}_{\mathfrak{M}}$ is defined as follows. For all $\ell < r$, we let $\mathsf{up}_{\mathfrak{M}}(s_\ell, s_\ell) = s_{\ell+1}$. We let $\mathsf{up}_{\mathfrak{M}}(s_r, s_r) = (\mathcal{P}_i, 1)$ where $i \in I$ is such that $\mathcal{P}_i$ controls s_r if $s_r \in S_I$ and i is arbitrary otherwise. For all $\ell \leq r$ and $s \neq s_\ell$, if $\ell \geq 1$ and there exists $i \in I$ such that $s_{\ell-1} \in S_i$ (i.e., $s \in S_I$), we let $\mathsf{up}_{\mathfrak{M}}(s_\ell, s) = \mathcal{P}_i$, and otherwise the update is arbitrary. Intuitively, in the first phase, the strategy checks that the current state matches the one it should be while following sg_0 and switches to a special punishment state if a deviation is detected. For all states of the form $(\mathcal{P}_i, j) \in M$ and all $s \in S$ occurring in sg_j, we let $\mathsf{up}_{\mathfrak{M}}((\mathcal{P}_i, j), s) = (\mathcal{P}_{i'}, j')$ where (a) $\mathcal{P}_{i'}$ is the player controlling s if $s \in S_I$ and otherwise, $i' = i$, and (b) $j' = j$ if $s \neq \mathsf{last}(\mathsf{sg}_j)$, and otherwise, if $s = \mathsf{last}(\mathsf{sg}_j)$, we set $j' = j + 1$ if $j < k$ and $j' = 1$ otherwise. For all states of the form $(\mathcal{P}_i, j) \in M$ and all $s \in S$ that do not occur in sg_j, we let $\mathsf{up}_{\mathfrak{M}}((\mathcal{P}_i, j), s) = \mathcal{P}_i$. Finally, for all $i \in I$ and $s \in S$, we let $\mathsf{up}_{\mathfrak{M}}(\mathcal{P}_i, s) = \mathcal{P}_i$.

Let $i \in I$. We now define $\mathsf{nxt}_{\mathfrak{M}_i}$. Let $s \in S_i$. We first consider memory states of the form s_ℓ. Fix $\ell \leq r$. If $s = s_\ell$, we let $\mathsf{nxt}_{\mathfrak{M}_i}(s_\ell, s) = a_\ell$ if $\ell \neq r$ and let $\mathsf{nxt}_{\mathfrak{M}_i}(s_r, s_r)$ be the first action of sg_1 (i.e., the only action that follows s_r in sg_1). Assume that $s \neq s_\ell$. If $\ell \geq 1$ and $s_{\ell-1} \in S_I$, we let $i' \in I$ such that $s \in S_{i'}$ and set $\mathsf{nxt}_{\mathfrak{M}_i}(s_\ell, s) = \tau_{-i'}(s)$ if $i' \neq i$. We let $\mathsf{nxt}_{\mathfrak{M}_i}(s_\ell, s)$ be arbitrary in all other cases.

We now deal with memory states of the form $(\mathcal{P}_{i'}, j)$. Fix $i' \in I$ and $j \in [\![1, k]\!]$. If s occurs in sg_j and $s \neq \mathsf{last}(\mathsf{sg}_j)$, we let $\mathsf{nxt}_{\mathfrak{M}_i}((\mathcal{P}_{i'}, j), s)$ be the action following s in sg_j. If $s = \mathsf{last}(\mathsf{sg}_j)$ and $j < k$ (resp. $j = k$), we let $\mathsf{nxt}_{\mathfrak{M}_i}((\mathcal{P}_{i'}, j), s)$ be the first action of sg_{j+1} (resp. sg_1). If s does not occur in sg_j, we let $\mathsf{nxt}_{\mathfrak{M}_i}((\mathcal{P}_{i'}, j), s) = \tau_{-i'}(s)$. Finally, we let $\mathsf{nxt}_{\mathfrak{M}_i}(\mathcal{P}_{i'}, s) = \tau_{-i'}(s)$ if $i' \neq i$, and otherwise we let it be arbitrary.

We let σ_i be the strategy induced by $\mathfrak{M}_i$. It can be shown by induction that the outcome of $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$ is π. We omit the proof here; it is very close to the argument for coherence appearing in the proof of Lemma 7.6.

It remains to show that σ is an NE from s_0. It suffices to show that for all $i \in [\![1, n]\!]$, if $\pi \notin \mathsf{Büchi}(T_i)$, then $\mathcal{P}_i$ does not have a profitable deviation. We fix one such $i \in [\![1, n]\!]$. By the characterisation of NE outcomes in Theorem 6.8, all states in π are elements of $W_{-i}(\mathsf{coBüchi}(T_i))$.

Let $\pi' = s'_0 a'_0 s'_1 a_1 \dots$ be a play starting in s_0 that is consistent with the strategy profile σ_{-i}. We consider three cases. First, assume that the sequence of states of π' does not have the sequence of states of sg_0 as a prefix. Let $\ell < r$ be such that the sequence of states of $\pi'_{\leq \ell}$ has the longest common prefix with that of sg_0. We have that $\widehat{\mathsf{up}_{\mathfrak{M}}}(\pi'_{\leq \ell}) = s_{\ell+1}$. The definition of σ and the relation $s'_{\ell+1} \neq s_{\ell+1}$ imply that $s_\ell \in S_i$. It follows that $\pi'_{\geq \ell}$ is a play consistent with τ_{-i} starting in $s_\ell \in W_{-i}(\mathsf{coBüchi}(T_i))$, thus $\pi'_{\geq \ell} \in \mathsf{coBüchi}(T_i)$. We obtain that $\pi' \in \mathsf{coBüchi}(T_i)$, ending this first case.

For the two other cases, we assume that the sequence of states of sg_0 is a prefix of the sequence of states of π'. For the second case, we assume that for all states s occurring in $\pi'_{\geq r}$, there is some $j \geq 1$ such that s appears in some sg_j. Because there are no elements of T_i in these segments, it follows that $\pi' \in \mathsf{coBüchi}(T_i)$.

Finally, assume that some state appearing in $\pi'_{\geq r}$ does not occur in any of the segments sg_j with $j \geq 1$. It follows that the memory state of the players relying on $\mathfrak{M}_i$ eventually becomes of the form $\mathcal{P}_{i'}$. Let $\ell \in \mathbb{N}$ be the largest number such that $\widehat{\mathsf{up}_{\mathfrak{M}}}(\pi'_{\leq \ell})$ is of the form $(\mathcal{P}_{i'}, j)$. It holds that s'_ℓ occurs in

sg_j and $s'_{\ell+1}$ does not occur in sg_j by choice of ℓ. It follows that $s'_\ell \in S_i$ by definition of σ (otherwise, $s'_{\ell+1}$ would occur in sg_j). We obtain that $i' = i$ and that $\pi'_{\geq \ell}$ is a play consistent with τ_{-i} starting in s'_ℓ. Because s'_ℓ occurs in sg_j, we have $s'_\ell \in W_{-i}(\mathsf{coBüchi}(T_i))$. As in the first case, we obtain $\pi' \in \mathsf{coBüchi}(T_i)$, ending the proof. □

The classical approach to derive move-independent NEs from lasso NE outcomes (e.g., those provided by Lemma 7.11) is to encode the whole lasso in the memory. If $|S|$ is finite, the resulting strategies from this approach have a memory size of at most $(|S|+2)n$. When there are few players compared to states of the game, the bound given in Theorem 7.14 can be seen as preferable to the one obtained via this classical construction.

We now consider the case of NE such that no state occurs infinitely often in their outcome. Let σ' be an NE from s_0 in $\mathcal{G}$ such that no state occurs infinitely often in $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)$. By Lemma 7.12, we can derive an NE outcome π from s_0 such that π and $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)$ satisfy the same objectives of $\mathcal{G}$, and π admits a simple segment decomposition $\mathcal{S} = (\mathsf{sg}_0, \mathsf{sg}_1, \mathsf{sg}_2, \ldots)$ where no targets of losing players occur in segments other than sg_0, and any two segments with an index with the same parity traverse different sets of states.

Once more, we only describe the second phase of our two-phase approach. We adapt the definitions of Section 7.2.3 in another way. Intuitively, to construct a finite-memory strategy profile, we allocate infinitely many disjoint segments to a same group of memory state. Due to this, players may not react to someone exiting the current segment.

The update and next-move function in the original definitions are defined from on the segment sg_j for each memory state of the form $(\mathcal{P}_i, j)$. In this setting, we define the update and next-move functions in memory states of the form $(\mathcal{P}_i, 1)$ (resp. $(\mathcal{P}_i, 2)$) from all odd segments (resp. all even segments besides sg_0) of $\mathcal{S}$ simultaneously. When the end of an even segment is reached in a memory state of the form $(\mathcal{P}_i, 2)$, the memory is updated to a state of the form $(\mathcal{P}_{i'}, 1)$ (and similarly for odd segments). The choice that all odd (resp. even) segments traverse pairwise disjoint set of vertices ensures that the next-move function is well-defined.

If at some point in the second phase, a state that does not occur in an even segment is read in a memory state $(\mathcal{P}_i, 2)$, the memory is updated to a punishing state $\mathcal{P}_i$, such that players attempt to punish $\mathcal{P}_i$ with a memoryless strategy (Theorem 6.2). We proceed similarly for the odd case. The resulting finite-memory strategy profile is an NE from s_0. On the one hand, any deviation such that the memory never updates to a punishing state must only have vertices that occur in segment sg_j with $j \neq 0$ in the limit. By choice of $\mathcal{S}$, this deviation cannot be profitable. Otherwise, it can be argued that the punishing strategy does in fact sabotage the deviating player, so long as their objective is not satisfied in π, by Theorem 6.8. We formally describe the construction above and establish its correctness below.

Lemma 7.15. *Let σ' be a pure NE from a state s_0 such that no state occurs infinitely often in $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)$. There exists a finite-memory pure NE σ from s_0 such that all strategies of σ are induced by move-independent Mealy machines and for all $i \in [\![1, n]\!]$, $\pi \in \mathsf{Büchi}(T_i)$ if and only if $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0) \in \mathsf{Büchi}(T_i)$.*

Proof. Let π be an NE outcome obtained via Lemma 7.12 from $\mathsf{Out}_{\mathcal{A}}(\sigma', s_0)$. We let $(\mathsf{sg}_0, \mathsf{sg}_1, \ldots)$ be the decomposition provided by the lemma such that all sets of states traversed by segments with an even (resp. odd) index are pairwise disjoint. We prove that π is the outcome of a finite-memory move-independent NE. The construction below can be seen as an adaptation of the proof of Theorem 7.14.

We introduce some notation. Let $I \subseteq [\![1, n]\!]$ be $\{i \in [\![1, n]\!] \mid \pi \notin \mathsf{Büchi}(T_i)\}$ if this set is not empty, or $\{1\}$ otherwise. For all $i \in I$, we let τ_{-i} be a memoryless uniformly winning strategy for the second player of the coalition game $\mathcal{G}_i = (\mathcal{A}_i, \mathsf{Büchi}(T_i))$ (it exists by Theorem 6.2) and let $W_{-i}(\mathsf{coBüchi}(T_i))$ denote the winning region of this player in $\mathcal{G}_i$. We write $\mathsf{sg}_0 = s_0 a_0 \ldots s_r$. Finally, we define $S_I = \bigcup_{i \in I} S_i$, E_1 (resp. E_2) to be the states occurring in segments sg_j with odd j (resp. even $j \geq 2$), $L_1 = \{\mathsf{last}(\mathsf{sg}_j) \mid j \in 2\mathbb{N} + 1\}$ and $L_2 = \{\mathsf{last}(\mathsf{sg}_j) \mid j \in 2\mathbb{N} + 2\}$ be the set of last states of odd and positive even segments respectively.

For each $i \in [\![1, n]\!]$, we define a Mealy machine $\mathfrak{M}_i = (M, m_{\mathsf{init}}, \mathsf{up}_{\mathfrak{M}}, \mathsf{nxt}_{\mathfrak{M}_i})$

as follows. We define M as the union

$$\{s_\ell \in S \mid 0 \leq \ell \leq r\} \cup (\{\mathcal{P}_i \mid i \in I\} \times [\![1, 2]\!]) \cup \{\mathcal{P}_i \mid i \in I\}$$

and $m_{\mathsf{init}} = s_0$.

The update function $\mathsf{up}_{\mathfrak{M}}$ is defined as follows. For all $\ell < r$, we let $\mathsf{up}_{\mathfrak{M}}(s_\ell, s_\ell) = s_{\ell+1}$. We postpone the definition of $\mathsf{up}_{\mathfrak{M}}(s_r, s_r)$. For all $\ell \leq r$ and $s \neq s_\ell$, if $\ell \geq 1$ and there exists $i \in I$ such that $s_{\ell-1} \in S_i$ (i.e., $s \in S_I$), we let $\mathsf{up}_{\mathfrak{M}}(s_\ell, s) = \mathcal{P}_i$, and otherwise the update is arbitrary. Let $p \in [\![1, 2]\!]$. For all states of the form $(\mathcal{P}_i, p) \in M$ and all $s \in E_p$, we let $\mathsf{up}_{\mathfrak{M}}((\mathcal{P}_i, p), s) = (\mathcal{P}_{i'}, p')$ where (a) $\mathcal{P}_{i'}$ is the player controlling s if $s \in S_I$ and otherwise, $i' = i$, and (b) $p' = p$ if $s \notin L_p$, and otherwise we set $p' = 3 - p$ (i.e., if $p = 1$, it becomes 2 and vice-versa). For all states of the form $(\mathcal{P}_i, p) \in M$ and all $s \in S \setminus E_p$, we let $\mathsf{up}_{\mathfrak{M}}((\mathcal{P}_i, p), s) = \mathcal{P}_i$. We let $\mathsf{up}_{\mathfrak{M}}(s_r, s_r) = \mathsf{up}_{\mathfrak{M}}((\mathcal{P}_i, 1), s_r)$ where $i \in I$ is arbitrary (the definition does not depend on i because $s_r \in E_1$). Finally, for all $i \in I$ and $s \in S$, we let $\mathsf{up}_{\mathfrak{M}}(\mathcal{P}_i, s) = \mathcal{P}_i$.

Let $i \in I$. We now define $\mathsf{nxt}_{\mathfrak{M}_i}$. Let $s \in S_i$. We first consider memory states of the form s_ℓ. Fix $\ell \leq r$. If $s = s_\ell$, we let $\mathsf{nxt}_{\mathfrak{M}_i}(s_\ell, s) = a_\ell$ if $\ell \neq r$ and let $\mathsf{nxt}_{\mathfrak{M}_i}(s_r, s_r)$ be the first action of sg_1. Assume that $s \neq s_\ell$. If $\ell \geq 1$ and $s_{\ell-1} \in S_I$, we let $i' \in I$ such that $s \in S_{i'}$ and set $\mathsf{nxt}_{\mathfrak{M}_i}(s_\ell, s) = \tau_{-i'}(s)$ if $i' \neq i$. We let $\mathsf{nxt}_{\mathfrak{M}_i}(s_\ell, s)$ be arbitrary in all other cases.

We now deal with memory states of the form $(\mathcal{P}_{i'}, p)$. Fix $i' \in I$ and $p \in [\![1, 2]\!]$. Assume first that $s \in E_p \setminus L_p$. There is a unique $j \in 2\mathbb{N} + p$ such that s occurs in sg_j (it is unique because all segments with an index with the same parity traverse disjoint sets of vertices). We set $\mathsf{nxt}_{\mathfrak{M}_i}((\mathcal{P}_{i'}, p), s)$ to the action following s in sg_j. Next, assume that $s \in L_p$. There is a unique $j \in 2\mathbb{N} + (3 - p)$ such that $s = \mathsf{first}(\mathsf{sg}_j)$. We let $\mathsf{nxt}_{\mathfrak{M}_i}((\mathcal{P}_{i'}, p), s)$ be the first action of sg_j. Finally, if $s \notin E_p$ and $i' \neq i$, we let $\mathsf{nxt}_{\mathfrak{M}_i}((\mathcal{P}_{i'}, p), s) = \tau_{-i'}(s)$, and let it be arbitrary if $i' = i$. Finally, we let $\mathsf{nxt}_{\mathfrak{M}_i}(\mathcal{P}_{i'}, s) = \tau_{-i'}(s)$ if $i' \neq i$, and it is arbitrary otherwise.

We let σ_i be the strategy induced by $\mathfrak{M}_i$. It can be shown by induction that the outcome of $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$ is π. We omit the proof here; it is very close to the argument for coherence appearing in the proof of Lemma 7.6.

It remains to show that σ is an NE from s_0. It suffices to show that for all $i \in [\![1, n]\!]$, if $\pi \notin \mathsf{Büchi}(T_i)$, then $\mathcal{P}_i$ does not have a profitable deviation. We fix one such i. By the characterisation of NE outcomes in Theorem 6.8, all states in π are elements of $W_{-i}(\mathsf{coBüchi}(T_i))$.

Let $\pi' = s'_0a'_0s'_1\dots$ be a play starting in s_0 that is consistent with the strategy profile σ_{-i}. We consider three cases. First, assume that the sequence of states of π' does not have the sequence of states of sg_0 as a prefix. We can prove the absence of a profitable deviation of $\mathcal{P}_i$ as in the same way as in the proof of Theorem 7.14: π' has a suffix consistent with τ_{-i} starting in a state occurring in sg_0 that is therefore in $W_{-i}(\mathsf{coBüchi}(T_i))$.

For the other two cases, we assume that the sequence of states of $\pi'_{\leq r}$ matches the sequence of states of sg_0. Second, we assume that all states s occurring in $\pi'_{\geq r}$ are elements of $E_1 \cup E_2$. Because E_1 and E_2 do not intersect T_i by construction of π (see Lemma 7.12), it follows that $\pi' \in \mathsf{coBüchi}(T_i)$.

Finally, assume that some state appearing in $\pi'_{\geq r}$ is not an element of $E_1 \cup E_2$. It follows that the memory state of the players relying on $\mathfrak{M}_i$ eventually becomes of the form $\mathcal{P}_{i'}$. Let $\ell \in \mathbb{N}$ be the largest number such that $\widehat{\mathsf{up}_{\mathfrak{M}}}(\pi'_{\leq \ell})$ is of the form $(\mathcal{P}_{i'}, p)$. It holds that $s'_\ell \in E_p$ and $s'_{\ell+1} \notin E_p$ by choice of ℓ. It follows that $s'_\ell \in S_i$ by definition of σ (otherwise, $s'_{\ell+1}$ would have to be an element of E_p). We obtain that $i' = i$ and that $\pi'_{\geq \ell}$ is a play consistent with τ_{-i} starting in s'_ℓ. Because $s'_\ell \in E_p$, we have $s'_\ell \in W_{-i}(\mathsf{coBüchi}(T_i))$. We conclude that $\pi' \in \mathsf{coBüchi}(T_i)$. This ends the proof. □

Part III:

The expressiveness of different types of randomness

Chapter 8

Introduction

In this part, we present the results described in Chapter 3.2, originating from joint work with Mickaël Randour [MR22, MR24]. We study variations of stochastic Mealy machines and the classes of finite-memory strategies they induce. We provide a full description of the relationships between these classes of in terms of expressive power.

We refer the reader to Chapter 3.2 for an extended presentation of the context, and in particular for a description of our naming scheme for classes of finite-memory strategies via three-letter acronyms. This part contains three chapters. In the following, we summarise the contents of Chapter 9, in which we discuss the definition of outcome-equivalence and provide a proof of Kuhn's theorem. We then provide the lattices for the settings that were omitted from Chapter 3.2. These lattices are derived from the inclusion and separation results that are proven in Chapter 10 and Chapter 11 respectively. We close the chapter by discussing some related work.

Outcome-equivalence and Kuhn's theorem. We define two strategies of a player to be outcome-equivalent if they induce the same distributions over plays from all initial states and for all *pure strategy profiles* of the other players (Definition 2.46). In particular, a natural question is whether the quantification over pure strategy profiles is coherent with the intuitive notion of outcome-equivalence, which is to induce the same behaviour regardless of the behaviour of the other players.

To answer this question, we provide three outcome-equivalence criteria depending on the nature of the compared strategies. We show that two behavioural strategies are outcome-equivalent if and only if they agree over histories consistent with either strategy (Lemma 9.1). We provide similar criteria for the outcome-equivalence of two mixed strategies (Lemma 9.3) and for the outcome-equivalence of a mixed and a behavioural strategy (Lemma 9.5). All of these results essentially state that two strategy are outcome-equivalent if and only if they make the same decisions in consistent histories. Through these characterisations of outcome-equivalence, we can conclude that two outcome-equivalent strategies do indeed share the same behaviour (see the related Corollaries 9.2 and 9.4).

We then prove Kuhn's theorem in Chapter 9.2. We do not follow the proof of [Aum64], which considers a more general setting. We take inspiration from the proof of [OR94] in finite extensive form games and adapt their argument to the countably-branching infinite-horizon setting. Our outcome-equivalence criterion for mixed and behavioural strategies (Lemma 9.5) provides some intuition on how to construct outcome-equivalent strategies. We also provide a simple example illustrating that without perfect recall, mixed strategies need not admit outcome-equivalent behavioural strategies (Example 9.1).

We close this chapter by shifting our focus back to finite-memory strategies. We provide sufficient conditions under which an observation-based stochastic Mealy machine induces a behavioural strategy in games with imperfect information. We prove that, in a context with perfect recall, all observation-based Mealy machines induce behavioural observation-based strategies (Lemma 9.8), and that, regardless of perfect recall, DRD stochastic Mealy machines always induce behavioural strategies (Lemma 9.9).

Taxonomy of finite-memory strategies. Our results yield a full taxonomy of randomised finite-memory strategies in four settings:

- finite arenas with perfect recall;
- finite arenas with no assumption on recall;
- countable arenas with perfect recall;
- countable arenas with no assumption on recall.

In Chapter 3.2, we presented lattices that describe the relationships in terms of expressiveness between classes of randomised finite-memory strategies for the most general and least general setting among these four. In the following, we restate the lattice in the least general setting, and use it to derive the lattices for the two previously disregarded settings. While we do not repeat the lattice for the most general setting, we recall that, in that setting, all classes of strategies are pairwise distinct and the only inclusions that hold follow from one class of Mealy machines having more randomisation power than another (see Figure 3.3).

Figure 8.1 depicts the expressiveness relationships between classes of randomised finite-memory strategies in *finite arenas with perfect recall*. In this setting, three inclusions require a proof: RDD $\subseteq$ DRD, RRR $\subseteq$ DRR and RRR $\subseteq$ RDR. The proofs of these inclusions only use one of the two assumptions made in our least general setting: for RDD $\subseteq$ DRD and RRR $\subseteq$ DRR we only use *perfect recall* and for RRR $\subseteq$ RDR, we only require the *arena to be finite*. Without their required assumption, these inclusions fail, even if the other assumption holds.

In Figure 8.2, we depict the relevant lattices for countable arenas with perfect recall and for finite arenas with no assumption on perfect recall. First, we discuss the former, i.e., the lattice of Figure 8.2a. The differences with the lattice of Figure 8.1 are induced by the inclusion DRD $\subseteq$ RDR no longer holding in countable arenas: RDR strategies can only provide distributions over finite sets of actions, while DRD strategies do not have this restriction (see Lemma 11.10).

We now consider the latter setting, i.e., finite arenas with no assumption on recall, for which we obtain the lattice of Figure 8.2b. In this case, the differences with the lattice of Figure 8.1 follow from the failure of the inclusion RDD in DRR (Lemma 11.12): if actions are not observable, a DRR strategy cannot be used to emulate an RDD strategy that mixes two different constant strategies. In particular, the inclusion of RDD in DRD fails in this setting.

Chapters 10 and 11 contain all of the statements required to obtain all of our lattices.

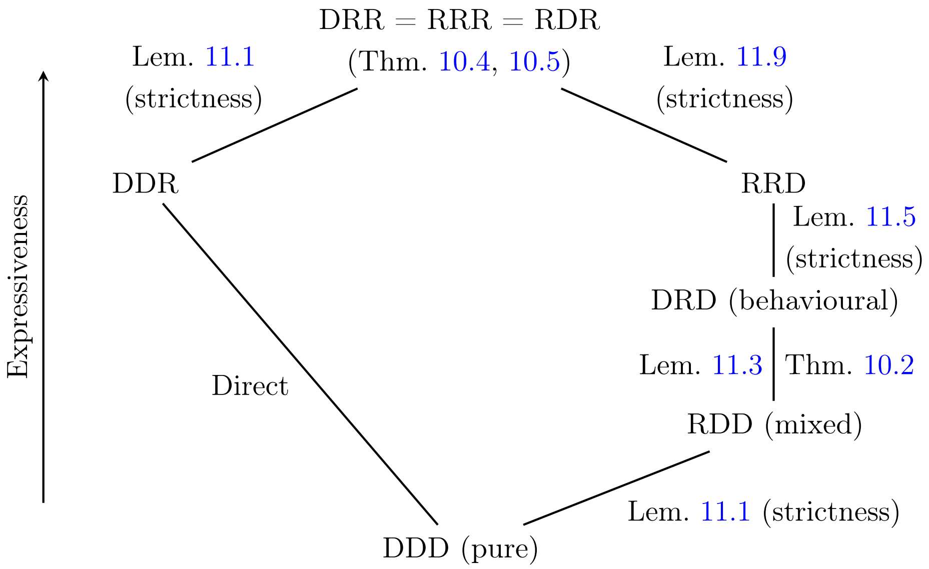
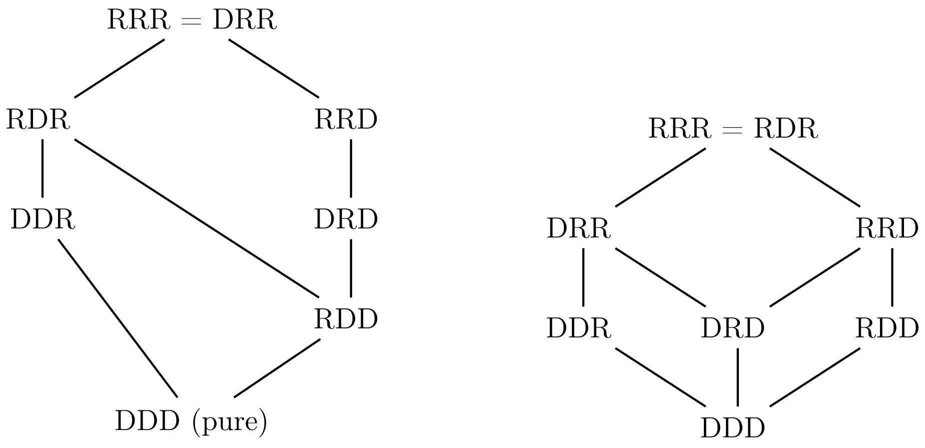


Figure 8.1: Lattice of finite-memory strategy classes in terms of expressive power in *finite multi-player arenas with perfect recall.*



(a) Lattice for *countable multi-player arenas with perfect recall.*

(b) Lattice for *finite multi-player arenas with no assumption on perfect recall.*

Figure 8.2: Lattices of finite-memory strategy classes in terms of expressive power.

Related work. We discuss two lines of work related to this work. The first one deals with the *various types of randomness* one can inject in strategies and their consequences. Obviously, Kuhn's theorem [Aum64] is a major inspiration, as well as the examples of differences between strategy models presented in [CDH10]. On a different but related note, [CDGH15] studies when randomness is not helpful in games nor strategies (as it can be simulated by other means or does not intervene).

The second axis concentrates on the use of *randomness as a means to simplify strategies* or to reduce their memory requirements. On the one hand, [CdH04, CHP08, CRR14, MPR20] study settings in which the performance of optimal strategies with memory can (almost) be matched by memoryless randomised strategies. On the other hand, [Cha07] and [Hor09] provide finer (upper and lower) memory bounds for almost-surely winning strategies in zero-sum turn-based stochastic Muller games by using DRD and RRR strategies respectively.

These are further motivations to understand randomised strategies even in contexts where randomness is not needed a priori to play optimally.

Chapter 9

Outcome-equivalence and Kuhn's theorem

We discuss the definition of outcome-equivalence: we provide equivalent definition that show that comparing two strategies only with respect to pure strategies of the other players implies that they induce the same behaviour even when the other players follow randomised strategies. We then provide a proof of Kuhn's theorem. The proof of one of its implications is inspired by our equivalent reformulation of outcome-equivalence of mixed and behavioural strategies. We close this chapter by providing sufficient conditions that ensure that the strategies induced by observation-based Mealy machines are observation-based behavioural strategies.

We fix an n-player arena $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$ and an arena with imperfect information $\mathfrak{P} = (\mathcal{A}, (\mathcal{Z}_i, \mathsf{Obs}_i)_{i\in[\![1,n]\!]})$ built on $\mathcal{A}$ for the whole chapter.

Contents

9.1 Outcome-equivalence

Two strategies are outcome-equivalent if and only if they induce the same distribution over plays from any state for all *pure strategies* of the other players. We restrict the strategies of the other players to pure strategies for a syntactic reason, as we have defined distributions over plays only for mixed strategy profiles and for behavioural strategy profiles.

We provide reformulations of outcome-equivalence for the comparison of two behavioural strategies (Section 9.1.1), of two mixed strategies (Section 9.1.2) and of a behavioural and a mixed strategy (Section 9.1.3). These reformulations provide conditions in which the strategies of the other players do not intervene. Intuitively, it is necessary and sufficient to check some condition over all histories consistent with one of the strategies to establish the outcome-equivalence of two randomised strategies.

9.1.1 Behavioural strategies

We first provide an equivalent formulation for the outcome-equivalence of behavioural strategies. We use this reformulation in Chapter 10 to establish inclusions between classes of finite-memory randomised strategies.

When comparing two behavioural strategies of a player, the distributions they induce depend only on the suggestions these strategies provide in histories that are consistent with them. Therefore, if these two strategies disagree only in inconsistent histories, then they yield the same distributions regardless of the strategy profile of the other players. Thus, the outcome-equivalence of two behavioural strategies can be restated as them having to agree over the histories that are consistent with (one of) the strategies.

Lemma 9.1. *Let $i \in [\![1, n]\!]$. Let σ_i and τ_i be behavioural strategies of $\mathcal{P}_i$ in $\mathcal{A}$. Then σ_i and τ_i are outcome-equivalent if and only if for all $h \in \mathsf{Hist}(\mathcal{A})$, if h consistent with σ_i, then $\sigma_i(h) = \tau_i(h)$.*

Proof. To simplify notation, we assume that $i = 1$; the general argument is recovered by renaming the players and adapting the notation in the proof below.

First, we assume that σ_1 and τ_1 are outcome-equivalent. Let $h \in \mathsf{Hist}(\mathcal{A})$ be a history that is consistent with σ_1. Let $s_{\mathsf{init}} = \mathsf{first}(h)$ and let σ_{-1} be a pure strategy profile of the players other than $\mathcal{P}_1$ with which h is consistent. Let $\bar{a} = (a^{(1)}, \sigma_{-1}(h)) \in \bar{A}(s)$. Let $s \in \mathsf{supp}(\delta(\mathsf{last}(h)), \bar{a}))$. By definition of the probability of a cylinder set and consistency of h with both σ_1 and σ_{-1}, we have

$$\sigma_1(h)(a^{(1)}) = \frac{\mathbb{P}^{\sigma_1,\sigma_{-1}}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h\bar{a}s))}{\mathbb{P}^{\sigma_1,\sigma_{-1}}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h)) \cdot \delta(\mathsf{last}(h), \bar{a})(s)}.$$

Furthermore, $\mathbb{P}^{\tau_1,\sigma_{-1}}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h)) = \mathbb{P}^{\sigma_1,\sigma_{-1}}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h)) > 0$ holds by outcome-equivalence of σ_1 and τ_1. Therefore, we have

$$\tau_1(h)(a^{(1)}) = \frac{\mathbb{P}^{\tau_1,\sigma_{-1}}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h\bar{a}s))}{\mathbb{P}^{\tau_1,\sigma_{-1}}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h)) \cdot \delta(\mathsf{last}(h), \bar{a})(s)}.$$

It follows from the equations above and the outcome-equivalence of σ_1 and τ_1 that $\sigma_1(h)(a^{(1)}) = \tau_1(h)(a^{(1)})$. We have shown that $\sigma_1(h) = \tau_1(h)$, which ends the proof of the first direction.

Let us now assume that σ_1 and τ_1 coincide over histories consistent with σ_1. Let σ_{-1} be a pure strategy profile of the players other than $\mathcal{P}_1$ and let $s_{\mathsf{init}} \in S$ be an initial state. It suffices to study the probability of cylinder sets. Let $h \in \mathsf{Hist}(\mathcal{A}, s_{\mathsf{init}})$ be a history starting in s_{init}. If h is consistent with σ_1, then all prefixes of h also are, therefore the definition of the probability of a cylinder ensures that $\mathbb{P}^{\sigma_1,\sigma_{-1}}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h)) = \mathbb{P}^{\tau_1,\sigma_{-1}}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h))$. Otherwise, if h is not consistent with σ_1, then h is necessarily of the form $h'\bar{a}h''$ with h' consistent with σ_1 and $\sigma_1(h')(a^{(1)}) = 0$. It follows that $\tau_1(h')(a^{(1)}) = 0$, thus $\mathbb{P}^{\sigma_1,\sigma_{-1}}_{s_{\mathsf{init}}}(h) = \mathbb{P}^{\tau_1,\sigma_{-1}}_{s_{\mathsf{init}}}(h) = 0$. This shows that σ_1 and τ_1 are outcome-equivalent, ending the proof. □

A corollary of Lemma 9.1 is that two behavioural strategies are outcome-equivalent if and only if they induce the same distribution over plays from any state for all *behavioural strategies* of the other players. This follows directly from the above and the definition of distributions over plays induced by behavioural strategies.

Corollary 9.2. *Let $i \in [\![1, n]\!]$, and let σ_i and τ_i be behavioural strategies of $\mathcal{P}_i$ in $\mathcal{A}$. Then σ_i and τ_i are outcome-equivalent if and only if for all behavioural strategy profiles σ_{-i} of players other than $\mathcal{P}_i$ and all $s \in S$, $\mathbb{P}_s^{\sigma_i,\sigma_{-i}} = \mathbb{P}_s^{\tau_i,\sigma_{-i}}$.*

9.1.2 Mixed strategies

We now state the counterpart of Lemma 9.1 for the outcome-equivalence of two mixed strategies. Similarly to the case of behavioural strategies, we need only check a property ranging over the set of histories consistent with the considered strategies. We must thus formalise the notion of consistency with a mixed strategy. Intuitively, a history is consistent with a mixed strategy if it may occur with positive probability under the strategy and some pure (or mixed) strategy profile of the other players. Formally, a history $h \in \mathsf{Hist}(\mathcal{A})$ is consistent with a mixed strategy μ_i of $\mathcal{P}_i$ if and only if the set of pure strategies of $\mathcal{P}_i$ with which h is consistent has a non-zero probability under μ_i.

Let $h \in \mathsf{Hist}(\mathcal{A})$. The contribution of a mixed strategy in the probability of the cylinder of h from its first state is the probability under the mixed strategy of the set of pure strategies with which h is consistent. It follows that two mixed strategies are outcome-equivalent if and only if these probabilities are the same for all histories that are consistent with the mixed strategies. This is formalised as follows.

Lemma 9.3. *Let $i \in [\![1, n]\!]$. For all $h \in \mathsf{Hist}(\mathcal{A})$, let $\Sigma^i_h \subseteq \Sigma^i_{\mathsf{pure}}(\mathcal{A})$ denote the set of pure strategies of $\mathcal{P}_i$ with which h is consistent. Let μ_i and ν_i be mixed strategies of $\mathcal{P}_i$ in $\mathcal{A}$. Then μ_i and ν_i are outcome-equivalent if and only if, for all histories $h \in \mathsf{Hist}(\mathcal{A})$ consistent with μ_i, we have $\mu_i(\Sigma^i_h) = \nu_i(\Sigma^i_h)$*

Proof. First, assume that μ_i and ν_i are outcome-equivalent. Let $h = s_0\bar{a}_0s_1\bar{a}_1 \dots \bar{a}_{r-1}s_r \in \mathsf{Hist}(\mathcal{A})$ be a history that is consistent with μ_i. Let σ_{-i} be a pure strategy profile of the players other than $\mathcal{P}_i$ with which h is consistent. By outcome-equivalence of μ and ν_i, we obtain that $\mathbb{P}_{s_0}^{\mu_i,\sigma_{-i}}(\mathsf{Cyl}(h)) = \mathbb{P}_{s_0}^{\nu_i,\sigma_{-i}}(\mathsf{Cyl}(h))$. By definition of probability distributions induced by mixed strategies this is equivalent to

$$\mu(\Sigma^i_h) \cdot \prod_{\ell=0}^{r-1} \delta(s_\ell, \bar{a}_\ell)(s_{\ell+1}) = \nu_i(\Sigma^i_h) \cdot \prod_{\ell=0}^{r-1} \delta(s_\ell, \bar{a}_\ell)(s_{\ell+1}).$$

In particular, since the product of transition probabilities appearing in the above expressions is non-zero, we obtain that $\mu_i(\Sigma^i_h) = \nu_i(\Sigma^i_h)$. This ends the proof of the first implication.

For the other implication, we prove its contrapositive. We now assume that μ_i and ν_i are not outcome-equivalent. We let $s \in S$ and σ_{-i} be a pure strategy profile of the players other than $\mathcal{P}_i$ such that $\mathbb{P}_s^{\mu_i,\sigma_{-i}} \neq \mathbb{P}_s^{\nu_i,\sigma_{-i}}$. There exists $h \in \mathsf{Hist}(\mathcal{A}, s)$ such that $\mathbb{P}_s^{\mu_i,\sigma_{-i}}(\mathsf{Cyl}(h)) \neq \mathbb{P}_s^{\nu_i,\sigma_{-i}}(\mathsf{Cyl}(h))$ (as otherwise the two distributions would coincide).

We may assume that $\mathbb{P}_s^{\mu_i,\sigma_{-i}}(\mathsf{Cyl}(h)) > \mathbb{P}_s^{\nu_i,\sigma_{-i}}(\mathsf{Cyl}(h))$: the set of cylinders of histories with the same length as h is a countable partition of the set of plays starting in s, and therefore we cannot have $\mathbb{P}_s^{\mu_i,\sigma_{-i}}(\mathsf{Cyl}(h')) \leq \mathbb{P}_s^{\nu_i,\sigma_{-i}}(\mathsf{Cyl}(h'))$ for all histories h' starting in s with the same length as h. It follows from this assumption that h is consistent with μ_i. Furthermore, $\mathbb{P}_s^{\mu_i,\sigma_{-i}}(\mathsf{Cyl}(h)) > \mathbb{P}_s^{\nu_i,\sigma_{-i}}(\mathsf{Cyl}(h))$ implies that $\mu_i(\Sigma^i_h) > \nu_i(\Sigma^i_h)$. This ends the proof of the second implication. □

Lemma 9.3 implies that two mixed strategies induce the same distributions over plays for all *mixed strategies* of the other players.

Corollary 9.4. *Let $i \in [\![1, n]\!]$, and let μ_i and ν_i be mixed strategies of $\mathcal{P}_i$ in $\mathcal{A}$. Then μ_i and ν_i are outcome-equivalent if and only if for all mixed strategy profiles μ_{-i} of players other than $\mathcal{P}_i$ and all $s \in S$, $\mathbb{P}_s^{\mu_i,\mu_{-i}} = \mathbb{P}_s^{\nu_i,\mu_{-i}}$.*

9.1.3 Mixed and behavioural strategies

We now consider the case in which we compare a mixed strategy to a behavioural strategy. On the one hand, for a mixed strategy μ, its contribution to the probability of a history cylinder is the probability under μ of the set of pure strategies with which the history is consistent. On the other hand, the contribution to the probability of a history cylinder of a behavioural strategy is the product of the probability of choosing the actions that appear along the history. Therefore, a mixed strategy and a behavioural strategy are outcome-equivalent if and only if these two contributions coincide over the set of histories that are consistent with one of the two strategies. Formally, we obtain the following result.

Lemma 9.5. *Let $i \in [\![1, n]\!]$. For all $h \in \mathsf{Hist}(\mathcal{A})$, let $\Sigma^i_h \subseteq \Sigma^i_{\mathsf{pure}}(\mathcal{A})$ denote the set of pure strategies of $\mathcal{P}_i$ with which h is consistent. Let μ_i be a mixed strategy of $\mathcal{P}_i$ in $\mathcal{A}$ and let σ_i be a behavioural strategy of $\mathcal{P}_i$ in $\mathcal{A}$. The three following assertions are equivalent:*

(i) μ_i and σ_i are outcome-equivalent;

(ii) for all histories $h = s_0\bar{a}_0s_1 \dots \bar{a}_{r-1}s_r \in \mathsf{Hist}(\mathcal{A})$ consistent with μ_i, we have $\mu_i(\Sigma^i_h) = \prod_{\ell=0}^{r-1} \sigma_i(h_{\leq \ell})(a^{(i)}_\ell)$;

(iii) for all histories $h = s_0\bar{a}_0s_1 \dots \bar{a}_{r-1}s_r \in \mathsf{Hist}(\mathcal{A})$ consistent with σ_i, we have $\mu_i(\Sigma^i_h) = \prod_{\ell=0}^{r-1} \sigma_i(h_{\leq \ell})(a^{(i)}_\ell)$.

Proof. We first prove that (i) implies (ii) and (iii), and then prove that the negation of (i) implies the negations of (ii) and (iii).

Assume that μ_i and σ_i are outcome-equivalent. It follows from the definition of consistency and of probability distributions induced by randomised strategies that a history is consistent with a mixed or behavioural strategy τ_i of $\mathcal{P}_i$ if and only it can occur with positive probability under τ_i with some pure strategy profile of the other players. Therefore, the set of histories consistent with μ_i and σ_i coincide due to their outcome-equivalence.

We let $h = s_0\bar{a}_0s_1 \dots \bar{a}_{r-1}s_r \in \mathsf{Hist}(\mathcal{A})$ be consistent with μ_i and σ_i. We let σ_{-i} be a pure strategy profile of the player other than $\mathcal{P}_i$ with which h is consistent. On the one hand, we have

$$\mathbb{P}^{\mu_i, \sigma_{-i}}_{s_0}(\mathsf{Cyl}\,(h)) = \mu_i(\Sigma^i_h) \cdot \prod_{\ell=0}^{r-1} \delta(s_\ell, \bar{a}_\ell)(s_{\ell+1}),$$

and on the other hand, we have

$$\mathbb{P}_{s_0}^{\mu_i,\sigma_{-i}}(\mathsf{Cyl}\,(h)) = \prod_{\ell=0}^{r-1} \sigma_i(h_{\leq \ell})(a_\ell^{(i)}) \cdot \delta(s_\ell, \bar{a}_\ell)(s_{\ell+1}).$$

It follows from $\prod_{\ell=0}^{r-1} \delta(s_\ell, \bar{a}_\ell)(s_{\ell+1}) > 0$ and the outcome-equivalence of μ_i and σ_i that $\mu_i(\Sigma_h^i) = \prod_{\ell=0}^{r-1} \sigma_i(h_{\leq \ell})(a_\ell^{(i)})$. This ends the proof that (i) implies (ii) and (iii).

We now assume that μ_i and σ_i are not outcome-equivalent and show that (ii) and (iii) do not hold. Let $s \in S$ and σ_{-i} be a pure strategy profile of the players other than $\mathcal{P}_i$ such that $\mathbb{P}_s^{\mu_i,\sigma_{-i}} \neq \mathbb{P}_s^{\sigma_i,\sigma_{-i}}$. There exists a history $h \in \mathsf{Hist}(\mathcal{A}, s)$ such that $\mathbb{P}_s^{\mu_i,\sigma_{-i}}(\mathsf{Cyl}\,(h)) \neq \mathbb{P}_s^{\sigma_i,\sigma_{-i}}(\mathsf{Cyl}\,(h))$. It follows that there exist histories $h^{(1)}$ and $h^{(2)}$ of the same length as h such that $\mathbb{P}_s^{\mu_i,\sigma_{-i}}(\mathsf{Cyl}\,(h^{(1)})) > \mathbb{P}_s^{\sigma_i,\sigma_{-i}}(\mathsf{Cyl}\,(h^{(1)}))$ and $\mathbb{P}_s^{\mu_i,\sigma_{-i}}(\mathsf{Cyl}\,(h^{(2)})) < \mathbb{P}_s^{\sigma_i,\sigma_{-i}}(\mathsf{Cyl}\,(h^{(2)}))$ (as the set of cylinders of histories with the same length as h is a countable partition of the set of plays). It follows that $h^{(1)}$ is consistent with μ_i and $h^{(2)}$ is consistent with σ_i. By following a reasoning similar to the first part of the proof, we can conclude $h^{(1)}$ and $h^{(2)}$ respectively witness that (ii) and (iii) do not hold. □

Through Lemma 9.5, we obtain that the outcome-equivalence of a mixed and behavioural strategy does not depend on the strategies of the other players. All that matters is that the two compared strategies contribute in the same way to the probability of cylinders of consistent histories. In particular, if we were to extend the definition of probabilities over plays to strategy profiles that contain both mixed and behavioural strategies, we would obtain that two strategies are outcome-equivalent if and only if the same distributions over plays are induced no matter the randomised strategies of the other players. This property can be seen as a generalisation of Corollary 9.2 and Corollary 9.4.

9.2 Kuhn's theorem

In this section, we prove Kuhn's theorem stating the equivalence of mixed and behavioural strategies. Our proof is based on that of [OR94] for finite extensive-form games (i.e., games played on finite trees); we adapt the argument they present to infinite-duration games on graphs. We compare the classes of randomised observation-based strategies in $\mathfrak{P}$ to state the theorem in its most general form. The result for the strategies of $\mathcal{A}$ follows by consider an arena with imperfect information built on $\mathcal{A}$ where the observation functions are the identify function.

Theorem 2.47 (Kuhn's theorem [Kuh53, Aum64])**.** *Let $i \in [\![1, n]\!]$. For every behavioural observation-based strategy σ_i of $\mathcal{P}_i$ in $\mathfrak{P}$, there exists an outcome-equivalent mixed strategy μ_i. If $\mathcal{P}_i$ has perfect recall, then for every mixed observation-based strategy μ_i of $\mathcal{P}_i$ in $\mathfrak{P}$, there exists an outcome-equivalent behavioural strategy σ_i.*

Lemma 9.5 provides some intuition on how to define mixed strategies from behavioural strategies and vice-versa. We show at the end of the section that without perfect recall, some mixed strategies need not admit an outcome-equivalent behavioural strategy.

We first prove that for all behavioural strategies, there is an equivalent mixed strategy.

Lemma 9.6. *Let $i \in [\![1, n]\!]$ and σ_i be a behavioural observation-based strategy of $\mathcal{P}_i$ in $\mathfrak{P}$. There exists a mixed observation-based strategy that is outcome-equivalent to σ_i.*

Proof. To simplify notation, we assume that $i = 1$. We define $P = \prod_{\bar{h} \in \mathsf{Obs}_1(\mathsf{Hist}(\mathcal{A}))} A^{(1)}(\mathsf{last}(\bar{h}))$, which can be seen as the set of pure observation-based strategies.

We derive a mixed observation-based strategy from a product measure over P. For all $h \in \mathsf{Hist}(\mathcal{A})$, we let $\mu_{\mathsf{Obs}_1(h)} \colon a^{(1)} \mapsto \sigma_i(h)(a^{(1)})$. This distribution is well-defined because σ_i is observation-based. We let μ'_1 be the (unique) product measure over P induced by the taking the distributions $\mu_{\mathsf{Obs}_1(h)}$ on each component.

We let $\mathcal{F} \colon P \to \Sigma^1_{\mathsf{pure}}(\mathcal{A})$ be the function that maps a pure observation-based strategy seen as a function over $\mathsf{Obs}_1(\mathsf{Hist}(\mathcal{A}))$ to the same strategy seen as a function over $\mathsf{Hist}(\mathcal{A})$. For all measurable $\Sigma \subseteq \Sigma^1_{\mathsf{pure}}(\mathcal{A})$, we define $\mu_1(\Sigma) = \mu'_1(\mathcal{F}^{-1}(\Sigma))$. The mixed strategy μ_1 is observation-based: the set of pure observation-based strategies has μ_1-probability 1 because its inverse image by $\mathcal{F}$ is P.

We now use the criterion of Lemma 9.5 to check the outcome-equivalence of σ_1 and μ_1. Let $h = s_0\bar{a}_0 s_1 \dots \bar{a}_{r-1}s_r \in \mathsf{Hist}(\mathcal{A})$ be consistent with σ_1. Let Σ^1_h denote the set of pure strategies of $\mathcal{P}_1$ in $\mathcal{A}$ with which h is consistent. By definition of μ_1, we obtain that

$$\mu_1(\Sigma^1_h) = \prod_{\ell=0}^{r-1} \sigma_1(h_{\leq \ell})(a^{(1)}_\ell).$$

Indeed, we have that $\mathcal{F}^{-1}(\Sigma^1_h)$ is the set of elements of P such its component for $h_{\leq \ell}$ is $a^{(1)}_\ell$ for all $\ell \in [\![r-1]\!]$. This ends the proof of the lemma. □

We now show that all mixed strategies of a player with perfect recall admit an outcome-equivalent behavioural strategy.

Lemma 9.7. *Let $i \in [\![1, n]\!]$ and σ_i be a behavioural observation-based strategy of $\mathcal{P}_i$ in $\mathfrak{P}$. Assume that $\mathcal{P}_i$ has perfect recall in $\mathfrak{P}$. Then there exists a mixed observation-based strategy that is outcome-equivalent to σ_i.*

Proof. We assume that $i = 1$ for convenience of notation.

Let μ_1 be a mixed observation-based strategy of $\mathcal{P}_1$. For all histories $h \in \mathsf{Hist}(\mathfrak{P})$, we let $\Sigma^1_h \subseteq \Sigma^1_{\mathsf{pure}}(\mathfrak{P})$ denote the set of pure strategies of $\mathcal{P}_1$ with which h is consistent. We consider the partially defined behavioural strategy σ_1 given by, for all $h \in \mathsf{Hist}(\mathfrak{P})$ consistent with μ_1, all actions $\bar{a} \in \bar{A}(\mathsf{last}(h))$ and all states $s \in \mathsf{supp}(\delta(\mathsf{last}(h), \bar{a}))$,

$$\sigma_1(h)(a^{(1)}) = \frac{\mu_1(\Sigma^1_{h\bar{a}s})}{\mu_1(\Sigma^1_h)}.$$

We claim that the above behavioural strategy is well-defined (over its domain) and is observation-based, and that any of its observation-based extensions are outcome-equivalent to μ_1. The claim regarding outcome-equivalence can be shown directly through the equivalent property (ii) of Lemma 9.5.

We observe that for any history $h \in \mathsf{Hist}(\mathfrak{P})$, $\bar{a} \in \bar{A}(\mathsf{last}(h))$ and $s \in \mathsf{supp}(\delta(\mathsf{last}(h), \bar{a}))$, the set $\Sigma^1_{h\bar{a}s}$ of pure strategies that are consistent $h\bar{a}s$ can be written as $\{\tau_1 \in \Sigma^1_h \mid \tau_1(h) = a^{(1)}\}$, i.e., this set does not depend on the actions of the players other than $\mathcal{P}_1$ in $\bar{a}$ and does not depend on s. Furthermore, $h \in \mathsf{Hist}(\mathfrak{P})$ is consistent with μ_1 if and only if $\mu_i(\Sigma^1_h) > 0$. It follows that σ_1 is well-defined over its domain.

It remains to show that σ_1 is observation-based. Let h and h' be histories that are indistinguishable for $\mathcal{P}_1$. Because μ_1 assigns a probability of zero to the set of pure strategies that are not observation-based, it is sufficient to show that for all pure observation-based strategies τ_1 of $\mathcal{P}_1$, h is consistent with τ_1 if and only if h' is to obtain that $\sigma_1(h) = \sigma_1(h')$. Let τ_1 be a pure observation-based strategy of $\mathcal{P}_1$. All prefixes of h and h' of the same length are indistinguishable, and thus share their image by τ_1. It follows that h is consistent with τ_1 if and only if h' is. □

We now provide a simple example illustrating that without perfect recall, mixed strategies need not admit outcome-equivalent behavioural strategies.

Example 9.1. We consider a one-state POMDP with two actions $\mathfrak{P}_{a,b}$ built from the MDP $\mathcal{M}_{a,b} = (\{s, \{a, b\}, \delta)$ – transitions in this MDP are self-loops. This MDP is used to witness the separations of classes of finite-memory strategies in Chapter 11; we depict it in said chapter, in Figure 11.1, Page 202. In $\mathfrak{P}$, we assign to s, a and b a shared observation o.

Let μ denote the finite-support mixed strategy that uniformly mixed the constant strategies a and b. No observation-based behavioural strategy is outcome-equivalent to μ. Let $\sigma\colon \{o\}(\{o\}^2)^* \to \mathcal{D}(\{a, b\})$ be an observation-based behavioural strategy of $\mathfrak{P}_{a,b}$. To obtain the outcome-equivalence of σ and μ, σ must be able to distinguish the histories sas and sbs and play action a and b respectively following these histories. However, because both histories are indistinguishable, this cannot be the case. ◁

The mixed strategy of Example 9.1 is a finite-memory strategy: it can be induced by a RDD Mealy machine whose randomised initialisation select a memory state that cannot be left that selects one of the actions. It follows that observation-based Mealy machines need not induced observation-based behavioural strategies. We provide sufficient conditions to ensure that Mealy machines induce observation-based strategies in the following section.

9.3 Observation-based Mealy machines

As illustrated in Example 9.1, the strategy induced by an observation-based Mealy machine need not be a behavioural strategy of $\mathfrak{P}$. We provide two sufficient conditions that ensure that observation-based Mealy machine induce an observation-based behavioural strategy. The first one we present introduces a restriction on the games. The second one involves no assumptions on games, but instead considers a restricted class of Mealy machines.

First, we show that all finite-memory strategies are behavioural in games with perfect recall. Intuitively, the distribution over memory states depends on the sequence of actions used by the considered player; the choice of actions conditions the distribution over memory states at each time it is updated. The visibility of actions makes it so the distribution over memory states depends only on the observations fed to the Mealy machine.

Lemma 9.8. *Let $i \in [\![1, n]\!]$. Let $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be an observation-based Mealy machine of $\mathcal{P}_i$. Assume that $\mathcal{P}_i$ has perfect recall in $\mathfrak{P}$. Then the strategy induced by $\mathfrak{M}$ is an observation-based behavioural strategy.*

Proof. Let μ_w denote the distribution over memory states of $\mathfrak{M}$ after w has taken place for $w \in (S\bar{A})^*$ consistent with $\mathfrak{M}$. By definition of the strategy induced by a Mealy machine, it suffices to show the following: for all $w, v \in (S\bar{A})^*$ that are indistinguishable and consistent with $\mathfrak{M}$, we have $\mu_w = \mu_v$. This can be shown by induction on the length of words in $(S\bar{A})^*$. On the one hand, for the base case (the empty word), there is nothing to show.

Let $w = w's\bar{a}$ and $v = v't\bar{b} \in (S\bar{A})^*$ be indistinguishable and consistent with $\mathfrak{M}$, and assume by induction that $\mu_{w'} = \mu_{v'}$. We show that $\mu_w = \mu_v$. Because $\mathcal{P}_i$ has perfect recall in $\mathfrak{P}$, $a^{(i)} = b^{(i)}$. For all $m \in M$, we have (from the inductive relationship of Equation (2.1)), that

$$\mu_w(m) = \frac{\sum_{m' \in M} \mu_{w'}(m') \cdot \mathsf{up}_{\mathfrak{M}}(m', s, \bar{a})(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s)(a^{(i)})}{\sum_{m' \in M} \mu_{w'}(m') \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s)(a^{(i)})},$$

and (since $a^{(i)} = b^{(i)}$)

$$\mu_v(m) = \frac{\sum_{m' \in M} \mu_{v'}(m') \cdot \mathsf{up}_{\mathfrak{M}}(m', t, \bar{b})(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m', t)(a^{(i)})}{\sum_{m' \in M} \mu_{v'}(m') \cdot \mathsf{nxt}_{\mathfrak{M}}(m', t)(a^{(i)})}.$$

Because $\mathfrak{M}$ is observation-based and w and v are indistinguishable, it follows from the previous equations that $\mu_w = \mu_v$. □

We now try to identify a subclass of Mealy machines that induce behavioural observation-based strategies regardless of whether the owner of the machine has perfect recall. Example 9.1 shows that this is not the case for Mealy machines with randomised initialisation. By adapting this example so the randomised initialisation is emulated by a stochastic memory update after the first round of the game, we can show that Mealy machines with randomised updates need not induce behavioural observation-based strategies either. In contrast to these types of Mealy machines, we can show that observation-based DRD Mealy machines always induce behavioural observation-based strategies.

Lemma 9.9. *Let $i \in [\![1, n]\!]$. Let $\mathfrak{M} = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be a DRD Mealy machine of $\mathcal{P}_i$ in $\mathfrak{P}$. Then the strategy induced by $\mathfrak{M}$ is a behavioural strategy.*

Proof. For a DRD strategy, the distribution over memory states at any point is a Dirac distribution. As explained by Definition 2.20, the memory state m_w reached after $w \in (S\bar{A})^*$ is defined by induction. We have $m_\varepsilon = m_{\mathsf{init}}$ and for $ws\bar{a} \in (S\bar{A})^+$, we have $m_{ws\bar{a}} = \mathsf{up}_{\mathfrak{M}}(m_w, s, \bar{a})$. Since $\mathfrak{M}$ is observation-based, it follows that m_w depends only on the observations assigned to $w \in (S\bar{A})^*$. This implies the claim of the lemma. □

Chapter 10

Inclusions between finite-memory strategy classes

Mealy machines, and therefore finite-memory strategies, can be differentiated depending on which components of the Mealy machine are randomised. This chapter presents the non-trivial inclusions between the different classes of finite-memory strategies arising from these variations.

We first establish, in Section 10.1, that if a class of finite-memory strategies is no more expressive than another in two-player arenas, then the same relation holds in multi-player arenas. Section 10.2 and Section 10.3 present the proofs that all RDD and RRR respectively admit outcome-equivalent DRD and DRR strategies whenever we have perfect recall. Finally, we show that RRR strategies admit outcome-equivalent RDR strategies in finite arenas in Section 10.4.

Contents

10.1 From two-player to multi-player arenas

To lighten notation, it is convenient to prove inclusions between different classes of finite-memory strategies in the two-player setting. In this section, we formally establish that if an inclusion holds between two classes of finite-memory strategies in two-player arenas of from a certain class, then this inclusion extends to all multi-player arenas of the same class.

Let $n \in \mathbb{N}_{>0}$ and let $i \in [\![1, n]\!]$. We consider four classes of n-player arenas with imperfect information with respect to $\mathcal{P}_i$:

(i) $\mathcal{C}^{FP}_{n,i}$, the class of finite arenas where $\mathcal{P}_i$ has perfect recall;

(ii) $\mathcal{C}^{IP}_{n,i}$, the class of countable arenas where $\mathcal{P}_i$ has perfect recall;

(iii) $\mathcal{C}^{FI}_{n,i}$, the class of finite arenas where $\mathcal{P}_i$ need not have perfect recall;

(iv) $\mathcal{C}^{II}_{n,i}$, the class of countable arenas where $\mathcal{P}_i$ need not have perfect recall.

We now fix a class $\mathcal{C}_{n,i}$ among the four above and the corresponding class $\mathcal{C}_{2,1}$ of two-player arenas.

Let $\mathrm{X}, \mathrm{Y}, \mathrm{Z}, \mathrm{A}, \mathrm{B}, \mathrm{C} \in \{\mathrm{D}, \mathrm{R}\}$. Assume that, in all arenas in $\mathcal{C}_{2,1}$, all observation-based XYZ Mealy machines of $\mathcal{P}_1$ admit an outcome-equivalent observation-based ABC Mealy machine. We claim that the same property holds for observation-based XYZ Mealy machines of $\mathcal{P}_i$ in arenas in $\mathcal{C}_{n,i}$. Given an arena in $\mathcal{C}_{n,i}$ and an XYZ Mealy machine of $\mathcal{P}_i$ in this arena, we group the players other than $\mathcal{P}_i$ into a coalition, and obtain a two-player arena in $\mathcal{C}_{2,1}$. In this arena, we can apply our assumption on $\mathcal{C}_{2,1}$ to obtain an outcome-equivalent ABC Mealy machine from our XYZ one. In the two-player arena, the coalition player has access to the same pure strategies as the coalition in the original arena. Therefore, it follows that the ABC Mealy machine obtained by the assumption on two-player arenas is outcome-equivalent to the original XYZ Mealy machine in the multi-player arena. We formalise this argument in the proof below.

Theorem 10.1. *Let $X, Y, Z, A, B, C \in \{D, R\}$. Let $n \in \mathbb{N}_{>0}$ and $i \in [\![1, n]\!]$. Let $(\mathcal{C}_{n,i}, \mathcal{C}_{2,1}) \in \{(\mathcal{C}^{FP}_{n,i}, \mathcal{C}^{FP}_{2,1}), (\mathcal{C}^{IP}_{n,i}, \mathcal{C}^{IP}_{2,1}), (\mathcal{C}^{FI}_{n,i}, \mathcal{C}^{FI}_{2,1}), (\mathcal{C}^{II}_{n,i}, \mathcal{C}^{II}_{2,1})\}$. Assume that, in all arenas of $\mathcal{C}_{2,1}$, all XYZ observation-based Mealy machines of $\mathcal{P}_1$ admit an* *outcome-equivalent ABC observation-based Mealy machine. Then, for all arenas in $\mathcal{C}_{n,i}$, all XYZ observation-based Mealy machines of $\mathcal{P}_i$ admit an outcome-equivalent ABC observation-based Mealy machine.*

Proof. We assume that $i = 1$ to simplify notation; the general case can be obtained by renaming the players in the following argument. Let $\mathfrak{P} \in \mathcal{C}_{n,1}$ where $\mathfrak{P} = (\mathcal{A}, (\mathcal{Z}_i, \mathsf{Obs}_i)_{i\in[\![1,n]\!]})$ and $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$.

We consider the two-player arena $\mathcal{A}_1 = (S, A^{(1)}, \prod_{2\leq i\leq n} A^{(i)}, \delta)$ obtained by grouping the players other than $\mathcal{P}_1$ into a coalition. We let $\mathfrak{P}_1$ be the arena of imperfect observation on $\mathcal{A}_1$ in which the observations of $\mathcal{P}_1$ match their observations in $\mathfrak{P}$ and the other player is fully informed. We have $\mathfrak{P}_1 \in \mathcal{C}_{2,1}$.

We identify histories and plays of $\mathcal{A}$ and $\mathcal{A}_1$. Therefore, strategies of $\mathcal{P}_1$ in $\mathcal{A}$ are strategies of $\mathcal{P}_1$ in $\mathcal{A}_1$ and vice-versa. Similarly, observation-based Mealy machines of $\mathcal{P}_1$ in $\mathfrak{P}$ are observation-based Mealy machines of $\mathcal{P}_1$ in $\mathfrak{P}_1$ and vice-versa.

Let $\mathfrak{M}$ be an XYZ observation-based Mealy machine of $\mathcal{P}_1$ in $\mathfrak{P}$. It is also an XYZ observation-based Mealy machine of $\mathcal{P}_1$ in $\mathfrak{P}_1$. By our assumption on arenas in $\mathcal{C}_{2,1}$, there exists an ABC observation-based Mealy machine $\mathfrak{N}$ of $\mathcal{P}_1$ such that $\mathfrak{M}$ and $\mathfrak{N}$ are outcome-equivalent in $\mathcal{A}_1$.

We claim that $\mathfrak{M}$ and $\mathfrak{N}$ are outcome-equivalent in $\mathcal{A}$. Let σ_1 and τ_1 respectively denote the strategies induced by $\mathfrak{M}$ and $\mathfrak{N}$. Let σ_2, …, σ_n be pure strategies of $\mathcal{P}_2$, …, $\mathcal{P}_n$ respectively in $\mathcal{A}$, and let $s \in S$ be an initial state. Let τ_2 be the pure strategy of the second player of $\mathcal{A}_1$ defined by $\tau_2(h) = (\sigma_2(h), \dots, \sigma_n(h))$ for all $h \in \mathsf{Hist}(\mathcal{A})$. By definition of distributions induced by plays and outcome-equivalence of σ_1 and τ_1 in $\mathcal{A}_1$, we obtain $\mathbb{P}^{\sigma_1,\sigma_2,\dots,\sigma_n}_{\mathcal{A},s} = \mathbb{P}^{\sigma_1,\tau_2}_{\mathcal{A}_1,s} = \mathbb{P}^{\tau_1,\tau_2}_{\mathcal{A}_1,s} = \mathbb{P}^{\tau_1,\sigma_2,\dots,\sigma_n}_{\mathcal{A},s}$. This shows the outcome-equivalence of σ_1 and τ_1 in $\mathfrak{P}$. □

Theorem 10.1 implies that we need only prove inclusions between classes of finite-memory strategies in two-player arenas to obtain a result for all multi-player arenas.

10.2 From mixed to behavioural strategies with finite memory

We prove that we can find an outcome-equivalent DRD strategy from any RDD strategy, i.e., a strategy that mixes finitely many pure finite-memory strategies can be emulated by using a Mealy machine with a deterministic initialisation, deterministic updates and randomised outputs. The converse inclusion is not true; we show this in Section 11.3. The construction use to establish our inclusion yields a DRD strategy that has a state space of size exponential in the size of the state space of the original RDD strategy. We complement our inclusion result with a family of examples illustrating that some RDD strategies for which this exponential blow-up in the number of states is necessary for any outcome-equivalent DRD strategy. We show that this blow-up is unavoidable in both deterministic turn-based two-player arenas and MDPs.

Let $\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$ be a two-player arena. Fix an RDD strategy $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ of $\mathcal{P}_i$. Let us sketch how to emulate $\mathfrak{M}$ with a DRD strategy $\mathfrak{N} = (N, n_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ built with a subset construction-like approach. The memory states of $\mathfrak{N}$ are functions $f\colon \mathsf{supp}(\mu_{\mathsf{init}}) \to M \cup \{\bot\}$. A memory state f is interpreted as follows. Let $m_0 \in \mathsf{supp}(\mu_{\mathsf{init}})$ be an initial memory state. We let $f(m_0) = \bot$ if the history seen up to now is not consistent with the pure finite-memory strategy $(M, m_0, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ obtained from $\mathfrak{M}$ by fixing its initial state to m_0. Otherwise $f(m_0)$ is the memory state reached in the same pure finite-memory strategy after processing the current history by iterating memory updates from m_0. Memory updates of $\mathfrak{N}$ are naturally derived from these semantics.

Using this state space and update scheme, we can compute the probability of each memory state of the mixed finite-memory strategy $\mathfrak{M}$ after some sequence $w \in (S\bar{A})^*$ has taken place. Indeed, we keep track of each initial memory state from which it was possible to be consistent with w, and, for each such initial memory state m_0, the memory state reached after w was processed starting in m_0. Therefore, this likelihood can be inferred from μ_{init}; the probability of $\mathfrak{M}$ being in $m \in M$ after w has been processed is given by the (normalised) sum of the probability of each initial memory state $m_0 \in \mathsf{supp}(\mu_{\mathsf{init}})$ such that $f(m_0) = m$.

The definition of the next-move function of $\mathfrak{N}$ is directly based on the distribution over states of $\mathfrak{M}$ described in the previous paragraph, and ensures that $\mathfrak{M}$ and $\mathfrak{N}$ select actions with the same probabilities after any history. For any action $a^{(i)} \in A^{(i)}(s)$, the probability of $a^{(i)}$ being chosen in arena state s and in memory state f is determined by the probability of $\mathfrak{M}$ being in some memory state m such that $\mathsf{nxt}_{\mathfrak{M}}(m, s) = a^{(i)}$, where this probability is inferred from f.

Intuitively, we postpone the initial randomisation and instead randomise at each step in an attempt of replicating the initial distribution in the long run. The following proof formalises the DRD strategy outlined above and establishes its outcome-equivalence with the RDD strategy it is based on. We also show that this construction extends to the imperfect information setting, as long as $\mathcal{P}_i$ has perfect recall.

Theorem 10.2. *Let $n \in \mathbb{N}_{>0}$, $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$ be an n-player arena, $\mathfrak{P} = (\mathcal{A}, (\mathcal{Z}_i, \mathsf{Obs}_i)_{i\in[\![1,n]\!]})$ be an arena with imperfect information and $i \in [\![1, n]\!]$. Let $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be an RDD strategy of $\mathcal{P}_i$ in $\mathcal{A}$. There exists a DRD strategy $\mathfrak{N} = (N, n_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ such that $\mathfrak{N}$ and $\mathfrak{M}$ are outcome-equivalent. Furthermore, if $\mathfrak{M}$ is observation-based in $\mathfrak{P}$ and $\mathcal{P}_i$ has perfect recall in $\mathfrak{P}$, then this outcome-equivalent DRD strategy $\mathfrak{N}$ is also observation-based.*

Proof. Theorem 10.1 implies that proving the theorem for two-player arenas implies it for arenas with any number of players. Thus, we assume that $n = 2$ and write $\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$.

We formalise the strategy described above. Let us write M_0 for the support of the initial distribution μ_{init} of $\mathfrak{M}$. We define the set of memory states N to be the set of functions $M_0 \to M \cup \{\bot\}$. The initial memory state of $\mathfrak{N}$ is given by the identity function $n_{\mathsf{init}} : m_0 \mapsto m_0$ over M_0. The update function $\mathsf{up}_{\mathfrak{N}}$ is as follows. For any $f \in N$, any $s \in S$ and $\bar{a} \in \bar{A}(s)$, we let $\mathsf{up}_{\mathfrak{N}}(f, s, \bar{a})$ be the function f' such that for all $m_0 \in M_0$, we have

$$f'(m_0) = \begin{cases} \mathsf{up}_{\mathfrak{M}}(f(m_0), s, \bar{a}) & \text{if } f(m_0) \in M \text{ and } \mathsf{nxt}_{\mathfrak{M}}(f(m_0), s) = a^{(i)} \\ \bot & \text{otherwise.} \end{cases}$$

Whenever we perform an update of the memory, we refine our knowledge on what the initial memory state could have been according to the actions selected by $\mathcal{P}_i$ prior to the update. We map to $\perp$ any initial memory states m_0 such that the played action would not have been selected in the memory state $f(m_0) \in M$, effectively removing m_0 from the set of initial memory states from which we could have started.

The next-move function $\mathsf{nxt}_{\mathfrak{N}}$ is defined as follows: for any memory state $f \in N$ and $s \in S$, we let $\mathsf{nxt}_{\mathfrak{N}}(f, s)$ be arbitrary if f maps $\perp$ to all memory states, and otherwise $\mathsf{nxt}_{\mathfrak{N}}(f, s)$ is the distribution over $A^{(i)}$ such that, for all $a^{(i)} \in A^{(i)}(s)$, we have

$$\mathsf{nxt}_{\mathfrak{N}}(f, s)(a^{(i)}) = \sum_{\substack{m_0 \in M_0 \\ \mathsf{nxt}_{\mathfrak{M}}(f(m_0), s) = a^{(i)}}} \frac{\mu_{\mathsf{init}}(m_0)}{\sum_{m_0' \in f^{-1}(M)} \mu_{\mathsf{init}}(m_0')}.$$

The memory state $f \in N$ mapping $\perp$ to all initial memory states is only reached whenever a history inconsistent with $\mathfrak{M}$ has taken place under $\mathfrak{M}$. Thanks to Lemma 9.1, we need not take in account histories inconsistent with $\mathfrak{M}$ to establish the outcome-equivalence of $\mathfrak{M}$ and $\mathfrak{N}$. This explains why the next-move function is left arbitrary in that case.

By construction, if $\mathfrak{M}$ is an observation-based Mealy machine of $\mathcal{P}_i$ in $\mathfrak{P}$, then $\mathfrak{N}$ can be made observation-based by adequately choosing actions in the memory state f mapping $\perp$ to all initial memory states.

We now show that $\mathfrak{M}$ and $\mathfrak{N}$ are outcome-equivalent via Lemma 9.1. To this end, we first show a relation, for each $w \in (S\bar{A})^*$ consistent with $\mathfrak{M}$, between the distribution $\mu_w \in \mathcal{D}(M)$ over the memory states of $\mathfrak{M}$ after processing w and the function $f_w = \widehat{\mathsf{up}_{\mathfrak{N}}}(w)$ (see Definition 2.20 for the definition of the iterated memory update function $\widehat{\mathsf{up}_{\mathfrak{N}}}(w)$) reached after $\mathfrak{N}$ reads w. Formally, this relation is as follows: for any $w \in (S\bar{A})^*$ consistent with $\mathfrak{M}$ and any memory state $m \in M$, we have

$$\mu_w(m) = \frac{\sum_{m_0 \in f_w^{-1}(m)} \mu_{\mathsf{init}}(m_0)}{\sum_{m_0 \in f_w^{-1}(M)} \mu_{\mathsf{init}}(m_0)}. \tag{10.1}$$

In the above, $f_w^{-1}(M)$ is the set of initial memory states $m_0 \in M_0$ of $\mathfrak{M}$ that are compatible with w taking place. This equation intuitively expresses that

$\mathfrak{N}$ accurately keeps track of the current distribution over memory states of $\mathfrak{M}$ along a play. A corollary of the above is that whenever we follow histories consistent with $\mathfrak{M}$, we are assured to never reach the memory state of $\mathfrak{N}$ that assigns $\perp$ to all states in M_0.

We prove Equation (10.1) with an inductive argument. The case of $w = \varepsilon$ is trivial: by definition $\mu_\varepsilon = \mu_{\mathsf{init}}$ and f_ε is the identity function over M_0. Now, let us assume that Equation (10.1) holds for $w' \in (S\bar{A})^*$ consistent with $\mathfrak{M}$, and let us prove it for $w = w's\bar{a}$ consistent with $\mathfrak{M}$.

To write the inductive relation between $\mu_{w'}$ and μ_w, we use an adapted (albeit equivalent in this context) form of Equation (2.1) of Section 2.4.4. In this case, the update function $\mathsf{up}_{\mathfrak{M}}$ and next-move $\mathsf{nxt}_{\mathfrak{M}}$ of $\mathfrak{M}$ are deterministic. Thus, instead considering sums weighted by Dirac distributions, we only sum over relevant states for clarity.

First, we remark that it may be the case that $f_w^{-1}(M) \neq f_{w'}^{-1}(M)$. In light of this, we must take care not to have $f_w^{-1}(M) = \emptyset$, in which case the denominator of the right-hand side of Equation (10.1) evaluates to zero. From the definition of $\mathsf{up}_{\mathfrak{N}}$, it follows that $f_w^{-1}(M)$ is formed of the memory elements $m_0 \in f_{w'}^{-1}(M)$ such that $\mathsf{nxt}_{\mathfrak{M}}(f_{w'}(m_0), s) = a^{(i)}$. We know that $w = w's\bar{a}$ is consistent with $\mathfrak{M}$. This implies there is some $m \in M$ such that $\mathsf{nxt}_{\mathfrak{M}}(m, s) = a^{(i)}$ and $\mu_{w'}(m) > 0$. From the inductive hypothesis (Equation (10.1) with w'), we obtain that there is some $m_0 \in f_{w'}^{-1}(M)$ such that $f_{w'}(m_0) = m$, otherwise the right-hand side of the equation would evaluate to zero. The equality $f_{w'}(m_0) = m$ implies $m_0 \in f_w^{-1}(M)$, thus we have shown that $f_w^{-1}(M)$ is non-empty.

Now that we have shown that both sides of Equation (10.1) are well-defined for w, we move on to its proof. Let us write $\mathsf{nxt}_{\mathfrak{M}}(\cdot, s)^{-1}(a^{(i)})$ for the set $\{m \in M \mid \mathsf{nxt}_{\mathfrak{M}}(m, s) = a^{(i)}\}$. From the inductive relation between μ_w and $\mu_{w'}$, we obtain that

$$\mu_w(m) = \frac{\sum_{\substack{m' \in \mathsf{nxt}_{\mathfrak{M}}(\cdot, s)^{-1}(a^{(i)}) \\ \mathsf{up}_{\mathfrak{M}}(m', s, \bar{a}) = m}} \mu_{w'}(m')}{\sum_{m' \in \mathsf{nxt}_{\mathfrak{M}}(\cdot, s)^{-1}(a^{(i)})} \mu_{w'}(m')}.$$

For the numerator, we obtain from the inductive hypothesis that

$$\begin{aligned}\sum_{\substack{m' \in \mathsf{nxt}_{\mathfrak{M}}(\cdot, s)^{-1}(a^{(i)}) \\ \mathsf{up}_{\mathfrak{M}}(m', s, \bar{a}) = m}} \mu_{w'}(m') &= \sum_{\substack{m' \in \mathsf{nxt}_{\mathfrak{M}}(\cdot, s)^{-1}(a^{(i)}) \\ \mathsf{up}_{\mathfrak{M}}(m', s, \bar{a}) = m}} \sum_{m_0 \in f_{w'}^{-1}(m')} \frac{\mu_{\mathsf{init}}(m_0)}{\sum_{m'_0 \in f_{w'}^{-1}(M)} \mu_{\mathsf{init}}(m'_0)} \\ &= \sum_{m_0 \in f_w^{-1}(m)} \frac{\mu_{\mathsf{init}}(m_0)}{\sum_{m'_0 \in f_{w'}^{-1}(M)} \mu_{\mathsf{init}}(m'_0)}.\end{aligned}$$

To derive the simple sum from the double sum, we rely on the fact that $f_w(m_0) = m$ holds if and only if $\mathsf{up}_{\mathfrak{M}}(f_{w'}(m_0), s, \bar{a}) = m$ and $\mathsf{nxt}_{\mathfrak{M}}(f_{w'}(m_0), s) = a^{(i)}$, by definition of $\mathsf{up}_{\mathfrak{N}}$.

For the denominator, from the inductive hypothesis, we obtain that

$$\begin{aligned}\sum_{m' \in \mathsf{nxt}_{\mathfrak{M}}(\cdot, s)^{-1}(a^{(i)})} \mu_{w'}(m') &= \sum_{m' \in \mathsf{nxt}_{\mathfrak{M}}(\cdot, s)^{-1}(a^{(i)})} \sum_{m_0 \in f_{w'}^{-1}(m')} \frac{\mu_{\mathsf{init}}(m_0)}{\sum_{m'_0 \in f_{w'}^{-1}(M)} \mu_{\mathsf{init}}(m'_0)} \\ &= \sum_{m_0 \in f_w^{-1}(M)} \frac{\mu_{\mathsf{init}}(m_0)}{\sum_{m'_0 \in f_{w'}^{-1}(M)} \mu_{\mathsf{init}}(m'_0)}.\end{aligned}$$

The last equality is a consequence of the definition of $\mathsf{up}_{\mathfrak{N}}$: recall that $f_w^{-1}(M)$ consists of the elements $m_0 \in f_{w'}^{-1}(M)$ such that $\mathsf{nxt}_{\mathfrak{M}}(f_{w'}(m_0), s) = a^{(i)}$. By combining the two equations above, we immediately obtain Equation (10.1), ending the inductive argument.

We now establish the outcome-equivalence of $\mathfrak{M}$ and $\mathfrak{N}$. Let $h = ws \in \mathsf{Hist}(\mathcal{A})$ be a history of $\mathcal{A}$ consistent with $\mathfrak{M}$. Let $a^{(i)} \in A^{(i)}(s)$ be an action enabled in s. The probability of $a^{(i)}$ being played after h under $\mathfrak{M}$ is given by the weighted sum

$$\sum_{m \in \mathsf{nxt}_{\mathfrak{M}}(\cdot, s)^{-1}(a^{(i)})} \mu_w(m).$$

Under $\mathfrak{N}$, the probability of $a^{(i)}$ being played is $\mathsf{nxt}_{\mathfrak{N}}(f_w, s)(a^{(i)})$. It follows from Equation (10.1) that these two probabilities coincide. We have shown the outcome-equivalence of strategies $\mathfrak{M}$ and $\mathfrak{N}$, ending the proof. □

The construction of a DRD strategy provided in the proof of Theorem 10.2 leads to an exponential blow-up of the memory state space. For an RDD strategy $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$, we have constructed an outcome-equivalent

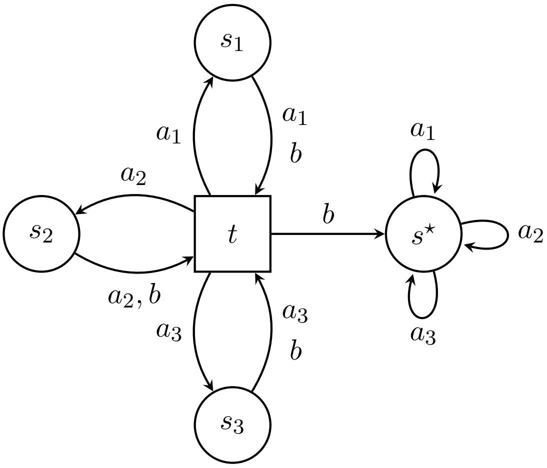


Figure 10.1: The arena $\mathcal{A}_3$ from the proof of Lemma 10.3. Circles and squares respectively represent states controlled by $\mathcal{P}_1$ and $\mathcal{P}_2$.

DRD strategy with a state space consisting of functions $\mathsf{supp}(\mu_{\mathsf{init}}) \to M \cup \{\bot\}$, therefore with a state space of size $(|M| + 1)^{|\mathsf{supp}(\mu_{\mathsf{init}})|}$. We show that an exponential blow-up in the number of initial memory states cannot be avoided in general, already in the turn-based setting.

Lemma 10.3. *Let $k \in \mathbb{N}_{>0}$. There exists a two-player turn-based deterministic arena (respectively an MDP) $\mathcal{A}_k$ with $k + 2$ states and $k + 1$ actions, and an RDD strategy $\mathfrak{M}_k$ of $\mathcal{P}_1$ with k states such that any outcome-equivalent DRD strategy must have at least $2^k - 1$ states.*

Proof. We construct a two-player turn-based deterministic arena $\mathcal{A}_k = (S_1^{(k)}, S_2^{(k)}, A^{(k)}, \delta^{(k)})$ as follows. We let $S_1^{(k)} = \{s_j \mid 1 \leq j \leq k\} \cup \{s^\star\}$, $S_2^{(k)} = \{t\}$ and $A^{(k)} = \{a_i \mid 1 \leq i \leq k\} \cup \{b\}$. As usual, we write $S^{(k)} = S_1^{(k)} \cup S_2^{(k)}$. We define the deterministic transition function $\delta^{(k)} \colon S^{(k)} \times A^{(k)} \to S^{(k)}$ as follows. For each $j \in [\![1, k]\!]$, only actions a_j and b are enabled in s_j, and all transitions from s_j move to t, i.e., $\delta^{(k)}(s_j, a_j) = \delta^{(k)}(s_j, b) = t$. All states besides t are reachable from t: we let, for all $j \in [\![1, k]\!]$, $\delta^{(k)}(t, a_j) = s_j$ and $\delta_k(t, b) = s^\star$.

In state $s^\star$, for all $j \in [\![1, k]\!]$, the action a_j labels a self-loop, i.e., we have $\delta^{(k)}(s^\star, a_j) = s^\star$. We illustrate the arena $\mathcal{A}_3$ in Figure 10.1.

We define an RDD strategy $\mathfrak{M}_k = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ of $\mathcal{P}_1$ as follows. We let $M = \{1, \dots, k\}$, and μ_{init} is the uniform distribution over M. The memory update function is trivial: we set $\mathsf{up}_{\mathfrak{M}}(m, s, a) = m$ for all $m \in M$, $s \in S^{(k)}$ and $a \in A^{(k)}$. For each memory state $m \in M$, we let $\mathsf{nxt}_{\mathfrak{M}}(m, s_m) = \mathsf{nxt}_{\mathfrak{M}}(m, s^\star) = a_m$ and, for all $j \neq m$, we let $\mathsf{nxt}_{\mathfrak{M}}(m, s_j) = b$. In $\mathfrak{M}$, once the initial state is decided, it no longer changes. In the memory state $m \in M$, the strategy prescribes action a_m in the states s_m and $s^\star$, and in states s_j with $j \neq m$, the strategy prescribes action b.

We now establish that all DRD strategies that are outcome-equivalent to $\mathfrak{M}$ must have at least $2^k - 1$ memory states. Let $\mathfrak{N} = (N, n_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ be one such DRD strategy. We give a lower bound on $|N|$ by showing that there must be at least $2^k - 1$ distinct distributions of the form $\mathsf{nxt}_{\mathfrak{N}}(\cdot, s^\star)$.

Let $E = \{j_1, \dots, j_\ell\} \subsetneq M$ be a proper subset of M. Consider the history (parentheses are provided to improve readability) $h_E = (t\, a_{j_1}\, s_{j_1}\, b)(t\, a_{j_2}\, s_{j_2}\, b) \dots (t\, a_{j_\ell}\, s_{j_\ell}\, b) t\, b\, s^\star$. Let $m \in E$. We see that along the history h_E, the action b is used in state s_m. Therefore, h_E is not consistent with the pure finite-memory strategy $(M, m, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ derived from $\mathfrak{M}$ by setting its initial state to m. Similarly, we see that for $m \notin E$, the history h_E is consistent with the pure finite-memory strategy $(M, m, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$. Thus, the set of actions that can be played after h_E when following $\mathfrak{M}_k$ is exactly the set $\{a_m \mid m \in M \setminus E\} \neq \emptyset$. Due to the deterministic initialisation and updates of DRD strategies, there must be some $n_E \in N$ such that $\mathsf{supp}(\mathsf{nxt}_{\mathfrak{N}}(n_E, s^\star)) = \{a_m \mid m \in M \setminus E\}$. Necessarily, we must have $\mathsf{supp}(\mathsf{nxt}_{\mathfrak{N}}(n_E, s^\star)) \neq \mathsf{supp}(\mathsf{nxt}_{\mathfrak{N}}(n_{E'}, s^\star))$ whenever $E \neq E'$, hence $n_E \neq n_{E'}$. Consequently, we must have at least one memory state in $\mathfrak{N}$ per proper subset of M, i.e., $|N| \geq 2^k - 1$.

It remains to show the existence of a suitable MDP and RDD strategy of this MDP. We explain how to adapt the deterministic arena $\mathcal{A}_k$ to a suitable MDP $\mathcal{M}_k$. Intuitively, we give replace the choices of $\mathcal{P}_2$ in $\mathcal{A}_k$ with random transitions. More precisely, we let $\mathcal{M}_k = (S^{(k)}, A^{(k)}, \delta^{(k)}_{\mathcal{M}})$ where $\delta^{(k)}_{\mathcal{M}}$ agrees with $\delta^{(k)}$ for any state-action pair $(s, a) \in S^{(k)} \times A^{(k)}$ such that $s \neq t$, and only action b is enabled in t in $\mathcal{M}_k$ and $\delta^{(k)}_{\mathcal{M}}(t, b)$ is a uniform distribution over $S^{(k)} \setminus \{t\}$.

We let $\mathfrak{M}'_k$ be the Mealy machine of $\mathcal{M}_k$ defined in the same way as $\mathfrak{M}_k$, with the additional output b in t. We can adapt the argument for $\mathcal{A}_k$ and $\mathfrak{M}_k$ to $\mathcal{M}_k$ and $\mathfrak{M}'_k$ to conclude that any DRD strategy that is outcome-equivalent to $\mathfrak{M}'_k$ in $\mathcal{M}_k$ requires at least $2^k - 1$ memory states. □

10.3 From randomised to deterministic initialisation

We now establish that DRR strategies are as expressive as RRR strategies, i.e., randomness in the initialisation can be removed. The general idea to remove randomisation in the initialisation is to simulate the behaviour of the RRR strategy at the start of the play using a new initial memory state and then move back into the RRR strategy we simulate.

We substitute the random selection of an initial memory element in two stages. To ensure the first action is selected in the same way under both the supplied strategy and the strategy we construct, we rely on randomised outputs. The probability of selecting an action $a^{(i)}$ in a given state s of the arena in our new initial memory state is given as the sum of selecting action $a^{(i)}$ in state s in each memory state m weighed by the initial probability of m.

We then leverage the stochastic updates to behave as though we had been using the original RRR strategy from the start. To achieve this, we base the update function of the constructed Mealy machine on the inductive relationship for the changes to the distribution over memory states after some sequence in $w \in (S\bar{A})^*$ takes place (Equation (2.1) of Chapter 2.4.4).

We now formalise this construction. It also applies in arenas with imperfect information where $\mathcal{P}_i$ has perfect recall. Perfect recall is useful to perform the first memory update described above. We thus obtain the following result.

Theorem 10.4. *Let $n \in \mathbb{N}_{>0}$, $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$ be an n-player arena, $\mathfrak{P} = (\mathcal{A}, (\mathcal{Z}_i, \mathsf{Obs}_i)_{i\in[\![1,n]\!]})$ be an arena with imperfect information and $i \in [\![1, n]\!]$. Let $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be an RRR strategy of $\mathcal{P}_i$ in $\mathcal{A}$. There exists a DRR strategy $\mathfrak{N} = (N, n_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ such that $\mathfrak{N}$ and $\mathfrak{M}$ are outcome-equivalent, and such that $|N| = |M| + 1$. Furthermore, if $\mathfrak{M}$ is observation-based in $\mathfrak{P}$ and $\mathcal{P}_i$ has perfect recall in $\mathfrak{P}$, then this outcome equivalent DRD strategy $\mathfrak{N}$ is also observation-based.*

Proof. By Theorem 10.1, it suffices to consider the case $n = 2$ to obtain the general case $n \in \mathbb{N}_{>0}$. Therefore, we assume that $n = 2$ and write $\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$.

We define a DRR strategy $\mathfrak{N} = (N, n_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ as follows. Let n_{init} be such that $n_{\mathsf{init}} \notin M$. We set $N = M \cup \{n_{\mathsf{init}}\}$. We let $\mathsf{up}_{\mathfrak{N}}$ and $\mathsf{nxt}_{\mathfrak{N}}$ coincide with $\mathsf{up}_{\mathfrak{M}}$ and $\mathsf{nxt}_{\mathfrak{M}}$ over $M \times S \times \bar{A}$ and $M \times S$ respectively (for the update function, we identify distributions over M to distributions over N that assign probability zero to n_{init}). It remains to define these two functions over $\{n_{\mathsf{init}}\} \times S \times \bar{A}$ and $\{n_{\mathsf{init}}\} \times S$ respectively.

First, we complete the definition of the memory update function $\mathsf{up}_{\mathfrak{N}}$. Let $s \in S$ and $\bar{a} \in \bar{A}$. We let $\mathsf{up}_{\mathfrak{N}}(n_{\mathsf{init}}, s, \bar{a})(n_{\mathsf{init}}) = 0$. We assume that there exists some $m_0 \in M$ such that $\mu_{\mathsf{init}}(m_0) > 0$ and $\mathsf{nxt}_{\mathfrak{M}}(m_0, s)(a^{(i)}) > 0$ (i.e., the action $a^{(i)}$ has a positive probability of being played in s at the start of a play under the strategy $\mathfrak{M}$). We set, for all $m \in M$,

$$\mathsf{up}_{\mathfrak{N}}(n_{\mathsf{init}}, s, \bar{a})(m) = \frac{\sum_{m' \in M} \mu_{\mathsf{init}}(m') \cdot \mathsf{up}_{\mathfrak{M}}(m', s, \bar{a})(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s)(a^{(i)})}{\sum_{m' \in M} \mu_{\mathsf{init}}(m') \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s)(a^{(i)})}.$$

Whenever we have $\mathsf{nxt}_{\mathfrak{M}}(m_0, s)(a^{(i)}) = 0$ for all $m_0 \in M_0$, we let $\mathsf{up}_{\mathfrak{N}}(n_{\mathsf{init}}, s, \bar{a})$ be a Dirac distribution over m for some arbitrary (but fixed independently of $a^{(i)}$) memory state $m \in M$.

For the next-move function $\mathsf{nxt}_{\mathfrak{N}}$, we define, for all states $s \in S$ and actions $a^{(i)} \in A^{(i)}(s)$,

$$\mathsf{nxt}_{\mathfrak{N}}(n_{\mathsf{init}}, s)(a^{(i)}) = \sum_{m \in M} \mu_{\mathsf{init}}(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)}).$$

By construction, $\mathfrak{N}$ is an observation-based Mealy machine in $\mathfrak{P}$ whenever $\mathfrak{M}$ is.

It remains to prove that $\mathfrak{M}$ and $\mathfrak{N}$ are outcome-equivalent. By Lemma 9.1, it suffices to show that both strategies suggest the same distributions over actions along histories consistent with $\mathfrak{M}$. We provide a proof in two steps. First, we consider histories with a single state. Second, we show that the distributions over memory states coincide in both Mealy machines after any $w \in S\bar{A}$ that is consistent with $\mathfrak{M}$ takes place. We conclude from this and the construction of $\mathfrak{N}$ that the strategies induced by $\mathfrak{M}$ and $\mathfrak{N}$ map all histories that are consistent with $\mathfrak{M}$ and have more than one state to the same distribution over actions of $\mathcal{P}_i$, ending the proof.

We show the first claim above. Let $s \in S$ and $a^{(i)} \in A^{(i)}(s)$. On the one hand, the probability of the action $a^{(i)}$ being played after the history s under $\mathfrak{M}$ is given by

$$\sum_{m \in M} \mu_{\mathsf{init}}(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)}).$$

On the other hand, the probability of this same action $a^{(i)}$ being played after the history s under $\mathfrak{N}$ is given by $\mathsf{nxt}_{\mathfrak{N}}(n_{\mathsf{init}}, s)(a^{(i)})$. These two probabilities coincide by construction.

Second, let $w = s\bar{a} \in S\bar{A}$ be consistent with $\mathfrak{M}$. Let μ_w and ν_w denote the distribution over memory states after w takes place under $\mathfrak{M}$ and $\mathfrak{N}$ respectively (ν_w is well-defined because the first claim implies that $s\bar{a}$ is consistent with $\mathfrak{N}$). Fix some $m \in M$, and let us prove that $\mu_w(m) = \nu_w(m)$. On the one hand, the relation between μ_{init} and μ_w given by Equation (2.1) (Section 2.4.4) states that

$$\begin{aligned}\mu_w(m) &= \frac{\sum_{m' \in M} \mu_{\mathsf{init}}(m') \cdot \mathsf{up}_{\mathfrak{M}}(m', s, \bar{a})(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s)(a^{(i)})}{\sum_{m' \in M} \mu_{\mathsf{init}}(m') \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s)(a^{(i)})} \\ &= \mathsf{up}_{\mathfrak{N}}(n_{\mathsf{init}}, s, a^{(i)})(m),\end{aligned}$$

and on the other hand, we have (because n_{init} is the sole initial state of $\mathfrak{N}$),

$$\nu_w(m) = \frac{\mathsf{up}_{\mathfrak{N}}(n_{\mathsf{init}}, s, \bar{a})(m) \cdot \mathsf{nxt}_{\mathfrak{N}}(n_{\mathsf{init}}, s)(a^{(i)})}{\mathsf{nxt}_{\mathfrak{N}}(n_{\mathsf{init}}, s)(a^{(i)})} = \mathsf{up}_{\mathfrak{N}}(n_{\mathsf{init}}, s, \bar{a})(m).$$

We have shown that $\mu_w = \nu_w$. Furthermore, because $\mathsf{nxt}_{\mathfrak{M}}$ and $\mathsf{nxt}_{\mathfrak{N}}$ agree over $M \times S$, and that $\mathsf{up}_{\mathfrak{M}}$ and $\mathsf{up}_{\mathfrak{N}}$ agree over $M \times S \times \bar{A}$, this equality generalises to all $w \in (S\bar{A})^+$ that are consistent with $\mathfrak{M}$. It follows that for any history $h \in (S\bar{A})^+S$ that is consistent with $\mathfrak{M}$, the images of h by the strategies induced by $\mathfrak{M}$ and $\mathfrak{N}$ match. We have shown that $\mathfrak{M}$ and $\mathfrak{N}$ are outcome-equivalent. □

10.4 From randomised to deterministic outputs in finite arenas

We are now concerned with the simulation of RRR strategies by RDR strategies, i.e., with substituting randomised outputs with deterministic outputs. The idea behind the removal of randomisation in outputs is to simulate said randomisation by means of both stochastic initialisation and updates. These are used to preemptively perform the random selection of an action, simultaneously with the selection of an initial or successor memory state. This construction assumes a *finite* arena. We can show that, in infinite arenas, some RRR strategies may not admit any outcome-equivalent RDR strategy. This discussion is postponed to Section 11.6.

Let $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$ be a finite n-player arena, $i \in [\![1, n]\!]$ and $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be an RRR strategy of $\mathcal{P}_i$. We construct an RDR strategy $\mathfrak{N} = (N, \nu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ that is outcome-equivalent to $\mathfrak{M}$ and such that $|N| \leq |M| \cdot |S| \cdot |A^{(i)}|$. The state space of $\mathfrak{N}$ consists of pairs (m, σ_i) where $m \in M$ and $\sigma_i \colon S \to A^{(i)}$ is a pure memoryless strategy of $\mathcal{P}_i$. To achieve our bound on the size of N, we cannot consider all pure memoryless strategies of $\mathcal{P}_i$, as there are exponentially many. We illustrate how we select pure memoryless strategies to achieve this bound through the following example. We apply the upcoming construction on a DRD strategy (which is a special case of RRR strategies) with a single memory state, i.e., a memoryless randomised strategy, in an MDP.

Example 10.1. We consider an MDP $\mathcal{M} = (S, A, \delta)$ where $S = \{s_1, s_2, s_3\}$, $A = \{a_1, a_2, a_3\}$ and all actions are enabled in all states. The transition function δ is irrelevant to this example, thus we leave unspecified. For our construction, we fix an order on the actions of $\mathcal{A}$: $a_1 < a_2 < a_3$.

Let $\mathfrak{M} = (\{m\}, m, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be the DRD strategy such that $\mathsf{nxt}_{\mathfrak{M}}(m, s_1)$ and $\mathsf{nxt}_{\mathfrak{M}}(m, s_2)$ are uniform distributions over $\{a_1, a_2\}$ and A respectively, and $\mathsf{nxt}_{\mathfrak{M}}(m, s_3)$ is defined by $\mathsf{nxt}_{\mathfrak{M}}(m, s_3)(a_1) = \frac{1}{3}$, $\mathsf{nxt}_{\mathfrak{M}}(m, s_3)(a_2) = \frac{1}{6}$ and $\mathsf{nxt}_{\mathfrak{M}}(m, s_3)(a_3) = \frac{1}{2}$.

Figure 10.2 illustrates the probability of each action being chosen in each state as the length of a segment. Let us write $0 = x_1 < x_2 < x_3 < x_4 < x_5 = 1$ for the endpoints of the segments appearing in the illustration. For each index

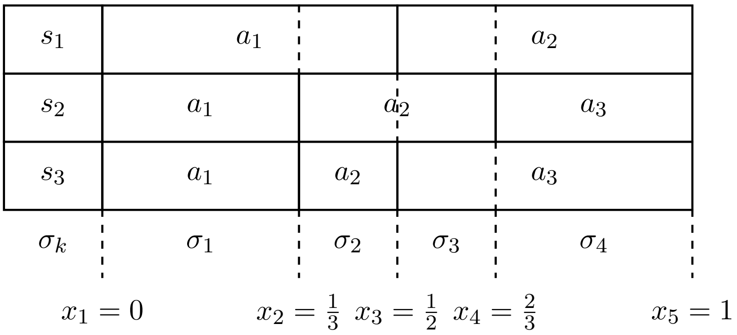


Figure 10.2: Representation of cumulative probability of actions under strategy $\mathfrak{M}$ and derived memoryless strategies.

$j \in [\![1, 4]\!]$, we define a pure memoryless strategy σ_j that assigns to each state the action lying in the segment above it in the figure. For instance, σ_2 is such that $\sigma_2(s_1) = a_1$ and $\sigma_2(s_2) = \sigma_2(s_3) = a_2$. Furthermore, for all $j \in [\![1, 4]\!]$, the length $x_{j+1} - x_j$ of its corresponding interval denotes the probability of the strategy being chosen during stochastic updates.

We construct an RDR strategy $\mathfrak{N} = (N, \nu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ that is outcome-equivalent to $\mathfrak{M}$ in the following way. We let $N = \{m\} \times \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$. The initial distribution is given by $\nu_{\mathsf{init}}(m, \sigma_j) = x_{j+1} - x_j$, i.e., the probability of σ_j in the illustration. We set, for any $j, j' \in [\![1, 4]\!]$, $s \in S$ and $a \in A$, $\mathsf{up}_{\mathfrak{N}}((m, \sigma_{j'}), s, a)((m, \sigma_j)) = x_{j+1} - x_j$. Finally, we let $\mathsf{nxt}_{\mathfrak{N}}((m, \sigma_j), s) = \sigma_j(s)$ for all $j \in [\![1, 4]\!]$ and $s \in S$.

The argument for the outcome-equivalence of $\mathfrak{N}$ and $\mathfrak{M}$ is the following: for any state $s \in S$, the probability of moving into a memory state (m, σ_j) such that $\sigma_j(s) = a$ is by construction the probability $\mathsf{nxt}_{\mathfrak{M}}(m, s)(a)$. ◁

In the previous example, we had a unique memory state m and we defined some memoryless strategies from the next-move function partially evaluated in this state (i.e., from $\mathsf{nxt}_{\mathfrak{M}}(m, \cdot)$). In general, each memory state may have a different partially evaluated next-move function. Therefore we define memoryless strategies for each individual memory state. For each memory state, we can bound the number of derived memoryless strategies by $|S| \cdot |A^{(i)}|$; we look at cumulative probabilities over actions (of which there are at most $|A^{(i)}|$) for each state. This explains our announced bound on $|N|$.

Furthermore, in general, the memory update function is not trivial. Generalising the construction above can be done in a straightforward manner to handle updates. Intuitively, the probability to move to some memory state of the form (m, σ_i) is given by the probability of moving into m in $\mathfrak{M}$ multiplied by the probability of σ (in the sense of Figure 10.2).

We now formally state our result in the general setting and provide its proof. The Mealy machine we construct has updates that do not depend on the actions of the player who owns it. We use this property to generalise the construction to finite arenas with imperfect information (perfect recall is not needed).

Theorem 10.5. *Let $n \in \mathbb{N}_{>0}$, $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$ be a finite n-player arena, $\mathfrak{P} = (\mathcal{A}, (\mathcal{Z}_i, \mathsf{Obs}_i)_{i\in[\![1,n]\!]})$ be an arena with imperfect information and $i \in [\![1, n]\!]$. Let $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be an RRR strategy of $\mathcal{P}_i$. There exists an RDR strategy $\mathfrak{N} = (N, \nu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ such that $\mathfrak{N}$ and $\mathfrak{M}$ are outcome-equivalent, and such that $|N| \leq |M| \cdot (|S| \cdot (|A^{(i)}| - 1) + 1)$. Furthermore, if $\mathfrak{M}$ is observation-based, then so is $\mathfrak{N}$.*

Proof. Theorem 10.1 implies that we need only consider the case $n = 2$ to obtain the general case $n \in \mathbb{N}_{>0}$. We assume that $n = 2$ and write $\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$ in the remainder of the proof.

Let us fix a linear order on the set of actions $A^{(i)}$, denoted by $<$. Fix some $m \in M$. We let $x_1^m < \ldots < x_{k(m)}^m$ denote the elements of the set

$$\left\{ \sum_{b^{(i)}<a^{(i)}} \mathsf{nxt}_{\mathfrak{M}}(m, s)(b^{(i)}) \mid s \in S,\, a^{(i)} \in A^{(i)} \right\}$$

that are strictly inferior to 1, and let $x_{k(m)+1}^m = 1$. These x_j^m represent the cumulative probability provided by $\mathsf{nxt}_{\mathfrak{M}}(m, \cdot)$ over actions of $\mathcal{P}_i$ taken in order, for each state of $\mathcal{A}$. For each $j \in [\![1, k(m)]\!]$, we define a memoryless strategy $\sigma_j^m \colon S \to A^{(i)}$ as follows: we have $\sigma_j^m(s) = a^{(i)}$ if

$$x_j^m \in \left[\sum_{b^{(i)} < a^{(i)}} \mathsf{nxt}_{\mathfrak{M}}(m, s)(b^{(i)}), \sum_{b^{(i)} \leq a^{(i)}} \mathsf{nxt}_{\mathfrak{M}}(m, s)(b^{(i)}) \right[.$$

In other words, for any state $s \in S$, we have $\sigma_j^m(s) = a^{(i)}$ whenever x_j^m is at least the cumulative probability of actions strictly inferior to $a^{(i)}$ in $\mathsf{nxt}_{\mathfrak{M}}(m, s)$ and at most the cumulative probability of actions up to action $a^{(i)}$ included. Refer to Figure 10.2 of Example 10.1 for an explicit illustration. We refer to $x_{j+1}^m - x_j^m$ as the probability of σ_j^m in the sequel.

Let $m \in M$, $s \in S$ and $a^{(i)} \in A^{(i)}(s)$. We show that we can relate $\mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)})$ and the sum of the probabilities of each σ_j^m such that $\sigma_j^m(s) = a^{(i)}$ as follows. First, we introduce some notation. Let $I(m, s, a^{(i)})$ denote the set of indices j such that $\sigma_j^m(s) = a^{(i)}$, i.e., the indices such that the jth strategy related to m prescribes action $a^{(i)}$ in s. It holds that

$$\sum_{j \in I(m,s,a^{(i)})} (x_{j+1}^m - x_j^m) = \mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)}). \tag{10.2}$$

Let $s \in S$ and $a^{(i)} \in A^{(i)}(s)$. Equation (10.2) can be proven as follows. First, note that all indices j appearing in the sum are consecutive by construction. Therefore, the sum above is telescoping and is equal to $x_{j^+ +1}^m - x_{j^-}^m$, where j^+ and j^- denote the largest and smallest indices in the sum respectively. By construction, we have $x_{j^-}^m = \sum_{b^{(i)} < a^{(i)}} \mathsf{nxt}_{\mathfrak{M}}(m, s)(b^{(i)})$ and $x_{j^+ +1}^m = \sum_{b^{(i)} \leq a^{(i)}} \mathsf{nxt}_{\mathfrak{M}}(m, s)(b^{(i)})$. We conclude that $x_{j^+ +1}^m - x_{j^-}^m = \mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)})$, proving Equation (10.2). This equation is used to establish the outcome-equivalence of $\mathfrak{M}$ with the strategy defined below.

We now define an RDR strategy $\mathfrak{N} = (N, \nu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$. We define

$$N = \{(m, \sigma_j^m) \mid m \in M,\, j \in [\![1, k(m)]\!]\}.$$

The initial distribution and update function of $\mathfrak{N}$ are derived from those of $\mathfrak{M}$ multiplied with the probability of the memoryless strategy that appears in the second component of the memory state of $\mathfrak{N}$ into which we move. The initial distribution ν_{init} is defined as

$$\nu_{\mathsf{init}}((m, \sigma_j^m)) = \mu_{\mathsf{init}}(m) \cdot (x_{j+1}^m - x_j^m)$$

for all $(m, \sigma_j^m) \in N$. The update function is defined as

$$\mathsf{up}_{\mathfrak{N}}((m, \sigma_j^m), s, \bar{a})((m', \sigma_k^{m'})) = \mathsf{up}_{\mathfrak{M}}(m, s, \bar{b})(m') \cdot (x_{k+1}^{m'} - x_k^{m'}),$$

where $\bar{b} = (\sigma_j^m(s), a^{(2)})$ if $i = 1$ (respectively $\bar{b} = (a^{(1)}, \sigma_j^m(s))$ if $i = 2$), for all $(m, \sigma_j^m), (m', \sigma_k^{m'}) \in N$, $s \in S$ and $\bar{a} \in \bar{A}$. We remark that this update function does not depend on the action of $\mathcal{P}_i$ given as input. Finally, the deterministic next-move function of $\mathfrak{N}$ is defined as $\mathsf{nxt}_{\mathfrak{N}}((m, \sigma_j^m), s) = \sigma_j^m(s)$ for all $(m, \sigma_j^m) \in N$ and all $s \in S$. By construction, $\mathfrak{N}$ is observation-based if $\mathfrak{M}$ is observation-based.

We now prove the outcome-equivalence of $\mathfrak{M}$ and $\mathfrak{N}$. For any $w \in (S\bar{A})^*$, let μ_w (resp. ν_w) denote the distribution over M (resp. N) after w has occurred under strategy $\mathfrak{M}$ (resp. $\mathfrak{N}$). The outcome-equivalence criterion of Lemma 9.1 and the definition of strategies derived from Mealy machines imply that, to end the proof, it suffices to establish, that the following holds for all histories $h = ws$ consistent with $\mathfrak{M}$:

$$\sum_{m \in M} \mu_w(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)}) = \sum_{m \in M} \sum_{j \in I(m, s, a^{(i)})} \nu_w((m, \sigma_j^m)). \tag{10.3}$$

To prove Equation (10.3), we first show that for any $w \in (S\bar{A})^*$ consistent with $\mathfrak{M}$, $\mu_w(m)$ is proportional to $\nu_w((m, \sigma_j^m))$. Specifically, for all $w \in (S\bar{A})^*$ consistent with $\mathfrak{M}$, we have

$$\nu_w((m, \sigma_j^m)) = (x_{j+1}^m - x_j^m) \cdot \mu_w(m). \tag{10.4}$$

To show Equation (10.4), we proceed by induction. Consider the empty word $w = \varepsilon$. Because $\mu_{\mathsf{init}} = \mu_\varepsilon$ and $\nu_{\mathsf{init}} = \nu_\varepsilon$, Equation (10.4) follows from the definition of ν_{init}. Let us now assume inductively that for $w' \in (S\bar{A})^*$ consistent with $\mathfrak{M}$, we have Equation (10.4) and let us prove it for $w = w's\bar{a}$ consistent with $\mathfrak{M}$. Fix $(m, \sigma_j^m) \in N$.

To invoke the inductive relation between ν_w and $\nu_{w'}$ (Equation (2.1), Section 2.4.4), w must be consistent with $\mathfrak{N}$. There exists $m' \in \mathsf{supp}(\mu_{w'})$ such that $\mathsf{nxt}_{\mathfrak{M}}(m', s)(a^{(i)}) > 0$ and $j \in I(m', s, a^{(i)})$ (this set is non-empty due to $\mathsf{nxt}_{\mathfrak{M}}(m', s)(a^{(i)}) > 0$). By the induction hypothesis, we obtain $\nu_{w'}((m', \sigma_j^{m'})) > 0$, which is sufficient to conclude that w is consistent with $\mathfrak{N}$.

We thus obtain, from the equation relating ν_w and $\nu_{w'}$, that $\nu_w((m, \sigma_j^m))$ is equal to

$$\frac{\sum_{m' \in M} \sum_{j' \in I(m', s, a^{(i)})} \nu_{w'}((m', \sigma_{j'}^{m'})) \cdot \mathsf{up}_{\mathfrak{N}}((m', \sigma_{j'}^{m'}), s, \bar{a})((m, \sigma_j^m))}{\sum_{m' \in M} \sum_{j' \in I(m', s, a^{(i)})} \nu_{w'}((m', \sigma_{j'}^{m'}))}.$$

The numerator of the above can be rewritten as follows, by successively using the definition of $\mathsf{up}_{\mathfrak{N}}$ followed by the inductive hypothesis and Equation (10.2):

$$\begin{aligned}
&\sum_{m' \in M} \sum_{j' \in I(m', s, a^{(i)})} \nu_{w'}((m', \sigma_{j'}^{m'})) \cdot \mathsf{up}_{\mathfrak{M}}(m', s, \bar{a})(m) \cdot (x_{j+1}^m - x_j^m) \\
=&(x_{j+1}^m - x_j^m) \cdot \sum_{m' \in M} \left(\mathsf{up}_{\mathfrak{M}}(m', s, \bar{a})(m) \cdot \mu_{w'}(m') \cdot \sum_{j' \in I(m', s, a^{(i)})} (x_{j'+1}^{m'} - x_{j'}^{m'}) \right) \\
=&(x_{j+1}^m - x_j^m) \cdot \sum_{m' \in M} \mathsf{up}_{\mathfrak{M}}(m', s, \bar{a})(m) \cdot \mu_{w'}(m') \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s)(a^{(i)}).
\end{aligned}$$

Following the same reasoning, the denominator can be rewritten as

$$\sum_{m' \in M} \mu_{w'}(m') \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s)(a^{(i)}).$$

By combining the equations above and the formula for the update of μ_w, we obtain that $\nu_w((m, \sigma_j^m)) = (x_{j+1}^m - x_j^m) \cdot \mu_w(m)$, ending the proof of Equation (10.4).

We now show that Equation (10.4) implies Equation (10.3), which will prove that $\mathfrak{M}$ and $\mathfrak{N}$ are outcome-equivalent. Let $h = ws \in \mathsf{Hist}(\mathcal{A})$ be a history consistent with $\mathfrak{M}$. Let $a^{(i)} \in A^{(i)}(s)$. The probability that the action $a^{(i)}$ is chosen after history h under $\mathfrak{M}$ is given by $\sum_{m \in M} \mu_w(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)})$. The probability that $a^{(i)}$ is selected after h under $\mathfrak{N}$, on the other hand, is given by

$$\begin{aligned}
\sum_{m \in M} \sum_{j \in I(m, s, a^{(i)})} \nu_w((m, \sigma_j^m)) &= \sum_{m \in M} \left(\mu_w(m) \cdot \sum_{j \in I(m, s, a^{(i)})} (x_{j+1}^m - x_j^m) \right) \\
&= \sum_{m \in M} \mu_w(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m, s)(a^{(i)}).
\end{aligned}$$

In the above, the first equation is obtained from Equation (10.4) and the second equation follows from Equation (10.2). This concludes the argument for the outcome-equivalence of our two finite-memory strategies.

To end the proof of this theorem, we prove the upper bound on $|N|$ given in the statement of the result. For any memory state $m \in M$, $k(m)$ is bounded by $|S| \cdot (|A^{(i)}| - 1) + 1$: by definition of the numbers x_j^m, we see that we must have $k(m) \leq |S| \cdot |A^{(i)}|$. To obtain the aforementioned bound, observe that for all $s \in S$, we have $\sum_{b^{(i)} < \min A^{(i)}} \mathsf{nxt}_{\mathfrak{M}}(m, s)(b^{(i)}) = 0$, i.e., 0 admits (at least) $|S|$ different writings in the set of the x_j^ms, hence $k(m) \leq |S| \cdot |A^{(i)}| - (|S| - 1) = |S| \cdot (|A^{(i)}| - 1) + 1$. Therefore, we have at most $|S| \cdot (|A^{(i)}| - 1) + 1$ pairs of the form (m, σ_j^m) per memory state $m \in M$. It follows that $|N| \leq |M| \cdot (|S| \cdot (|A^{(i)}| - 1) + 1)$. □

Remark 10.6. The choice of the order on the set of actions fixed at the start of the previous proof influences the size of the constructed strategy. It is not necessary to use the same ordering of actions for all memory states. The order is used to define all memoryless strategies of the form σ_j^m, which do not interact with strategies associated to other memory states. For this reason, it is possible to use different orderings on actions depending on the memory state m that is considered. ◁

Remark 10.7. The upper bound on the number of memory states given in the statement of Theorem 10.5 can be slightly improved in a turn-based setting. In general, we can replace the term $|S|$ in the bound by the number of states that $\mathcal{P}_i$ controls (more precisely, by the number of $\mathcal{P}_i$-controlled states with at least two enabled actions). ◁

Chapter 11

Separating classes of finite-memory randomised strategies

This chapter presents examples witnessing the non-inclusions of classes of randomised finite-memory strategies. We first present the separation results that hold in our most restricted setting: finite arenas with perfect information. These separations can be witnessed in an MDP with one state and two actions; we present this MDP in Section 11.1. We complement these examples with problem instances from the literature for which strategies from some class suffice whereas strategies from the compared class do not.

We first show that DDD is a strict subset of RDD in Section 11.2. Section 11.3 illustrates that DRD is not included in RDD. We then prove that DRD is strictly included in RRD in Section 11.4. Finally, we show that RRD and DDR strategies are incomparable in Section 11.5.

We then provide examples illustrating the non-inclusions of classes of strategies in more general settings. We first show that DRD is not included in RDR in infinite arenas 11.6. We then conclude by showing that RDD is not included in DRR in arenas without perfect recall in Section 11.7.

Contents

11.1 Separating classes of finite-memory strategies

We first formalise the MDP used throughout this chapter and explain the interpretation of the Mealy machine illustrations appearing in the chapter.

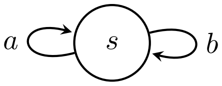


Figure 11.1: The MDP $\mathcal{M}_{a,b}$ with a single state and two actions.

We let $\mathcal{M}_{a,b}$ denote the MDP depicted in Figure 11.1. To prove the separation of two distinct classes of strategies, we provide witness strategies in $\mathcal{M}_{a,b}$ whenever possible. This is one of the simplest settings that allows us to distinguish classes of strategies. We accompany the separating examples of $\mathcal{M}_{a,b}$ with examples derived from problems from the literature.

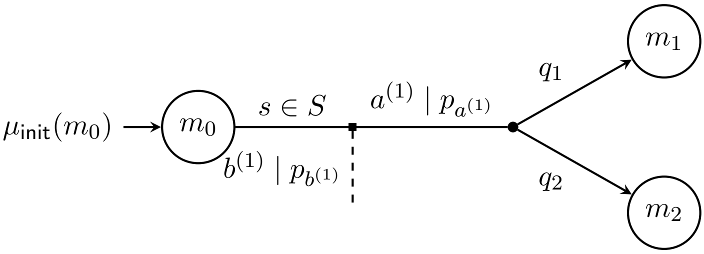

$\mu_{\mathsf{init}}(m_0)$
m_0
$s \in S$
$a^{(1)} \mid p_{a^{(1)}}$
$b^{(1)} \mid p_{b^{(1)}}$
q_1
q_2
m_1
m_2

Figure 11.2: A fragment of a stochastic Mealy machine of $\mathcal{P}_1$ the updates of which do not depend on the choices of other players. We have $p_{a^{(1)}} = \mathsf{nxt}_{\mathfrak{M}}(s)(a^{(1)})$, $p_{b^{(1)}} = \mathsf{nxt}_{\mathfrak{M}}(s)(b^{(1)})$, $q_1 = \mathsf{up}_{\mathfrak{M}}(m_0, s, \bar{a})(m_1)$ and $q_2 = \mathsf{up}_{\mathfrak{M}}(m_0, s, \bar{a})(m_2)$ (where $\bar{a}$ is an action profile where the action of $\mathcal{P}_1$ is $a^{(1)}$).

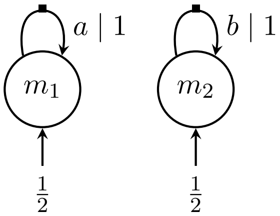


Figure 11.3: An RDD strategy of $\mathcal{M}_{a,b}$ that has no outcome-equivalent DDR counterpart.

We describe Mealy machines through illustrations. Figure 11.2 illustrates a fragment of a Mealy machine. Edges from a memory state are labelled by the current arena state, then split for randomised action choices, and finally, split again to represent stochastic memory updates. To lighten figures, we omit the first segment of an edge (labelled by arena states) when presenting examples for $\mathcal{M}_{a,b}$, as there is a single state in that MDP. When depicting a Mealy machine with deterministic updates, we omit the last edge subdivision.

11.2 DDD strategies are weaker than RDD ones

Pure finite-memory strategies are less powerful than RDD strategies. The latter class of strategies can induce non-Dirac distributions over the plays of $\mathcal{M}_{a,b}$ whereas the former cannot. We illustrate a strategy that has no outcome-equivalent DDD strategy in Figure 11.3. Furthermore, there is no DDR strategy that is outcome-equivalent to the strategy depicted in Figure 11.3: DDR strategies lack the ability to provide a randomised action at the first step of a game. We obtain the following result.

Lemma 11.1. *There exists an RDD strategy in $\mathcal{M}_{a,b}$ such that there is no outcome-equivalent DDR strategy. In particular, there is no outcome-equivalent DDD strategy.*

We now describe a setting in which RDD strategies suffice but DDD strategies do not. We consider multi-objective MDPs, i.e., MDPs with multiple objectives or payoff functions. Let $\mathcal{M} = (S, A, \delta)$ be an MDP and let

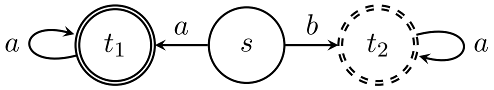


Figure 11.4: An MDP. The highlighted states are targets for the reachability objectives $\mathsf{Reach}(t_1)$ and $\mathsf{Reach}(t_2)$.

$\bar{f}\colon \mathsf{Plays}(\mathcal{M}) \to \mathbb{R}^d$ be a multi-dimensional payoff function We say that $\mathbf{q} \in \bar{\mathbb{R}}^d$ is *achievable* from $s \in S$ if there exists a strategy σ such that $\mathbb{E}^\sigma_s(\bar{f}) \geq \mathbf{q}$.

Randomisation may be necessary to achieve vectors in multi-objective MDPs (see, e.g., the example in Chapter 3.3). In Chapter 14, we prove that when considering universally integrable payoffs, any achievable vector can be achieved by using a mixed strategy with finite support. For certain classes of payoffs, this result can be strengthened to show that it suffices to *mix finitely many pure finite-memory strategies* to achieve any vector. We illustrate this property on MDPs with multiple reachability objectives (see, e.g., [EKVY08, RRS17]). We first provide an example where randomisation is necessary to achieve a vector.

Example 11.1. Consider the MDP depicted in Figure 11.4 and let s be the initial state. We consider the two targets $T_1 = \{t_1\}$ and $T_2 = \{t_2\}$ and the vector $\mathbf{q} = (\frac{1}{2}, \frac{1}{2})$. It is clear that no pure strategy witnesses the achievability of $\mathbf{q}$ from s; a pure strategy yields the vector $(1, 0)$ or $(0, 1)$ if it chooses action a or b in s respectively. However, there is an RDD strategy that witnesses the achievability of $\mathbf{q}$; any extension of the strategy depicted in Figure 11.3 that accounts for the new game states t_1 and t_2 achieves $\mathbf{q}$. $\lhd$

As claimed above, RDD strategies suffice to achieve any vector in an MDP with multiple reachability objectives. This follows from the results of [EKVY08]: they show that the set of achievable vectors in this setting is a polyhedral set. Their arguments imply that the vertices of the achievable set can be attained via pure finite-memory strategies. This implies that any vector can be achieved by a RDD strategy in this setting.

Lemma 11.2. *Let $\mathcal{M} = (S, A, \delta)$ be an MDP, $s \in S$, $T_1, \ldots, T_d \subseteq S$ be target sets. For all vectors $\mathbf{q}$ that are achievable from s, there exists an RDD strategy σ such that $\mathbf{q} \leq (\mathbb{P}^\sigma_s(\mathsf{Reach}(T_j))_{j \in [\![1,d]\!]}$.*

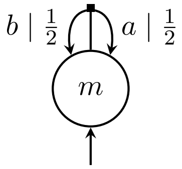


Figure 11.5: A DRD Mealy machine for the memoryless strategy playing uniformly at random at each step of a play. Witnesses that RDD $\subsetneq$ DRD.

11.3 DRD strategies are not included in RDD

We now show that there exists an DRD strategy that cannot be emulated by any RDD strategy in $\mathcal{M}_{a,b}$. Intuitively, an RDD strategy can only randomise once at the start between a finite number of pure finite-memory (DDD) strategies. After this initial randomisation, the sequence of actions prescribed by the RDD strategy is fixed relative to the play in progress. Any DRD strategy that chooses an action randomly at each step, such as the strategy depicted in Figure 11.5, i.e., the strategy playing actions a and b with uniform probability at each step in $\mathcal{M}_{a,b}$, cannot be reproduced by an RDD strategy. Indeed, this randomisation generates an infinite number of patterns of actions. These patterns cannot all be captured by an RDD strategy due to the fact that its initial randomisation is over a finite set.

Lemma 11.3. *There exists a DRD strategy in $\mathcal{M}_{a,b}$ such that there is no outcome-equivalent RDD strategy.*

Proof. Let $\sigma_1 \colon \{s\} \to \mathcal{D}(\{a, b\})$ be the memoryless strategy in $\mathcal{M}_{a,b}$ induced by the Mealy machine depicted in Figure 11.5. The distribution $\sigma_1(s)$ is the uniform distribution over $\{a, b\}$. The strategy σ_1 induces a probability distribution over plays of $\mathcal{M}_{a,b}$ such that all plays have a probability of zero. Indeed, let π be a play of $\mathcal{M}_{a,b}$. One can view the singleton $\{\pi\}$ as the decreasing intersection $\bigcap_{\ell\in\mathbb{N}} \mathsf{Cyl}\,(\pi_{\leq \ell})$. Hence, $\mathbb{P}_s^{\sigma_1}(\{\pi\}) = \lim_{\ell\to\infty} \mathbb{P}_s^{\sigma_1}(\mathsf{Cyl}\,(\pi_{\leq\ell}))$. For all $\ell \in \mathbb{N}$, we have $\mathbb{P}_s^{\sigma_1}(\mathsf{Cyl}\,(\pi_{\leq\ell})) = \frac{1}{2^\ell}$. It follows that the probability of $\{\pi\}$ is zero.

We now establish that there is no outcome-equivalent RDD strategy. First, let us recall that any RDD strategy can be presented as a distribution over

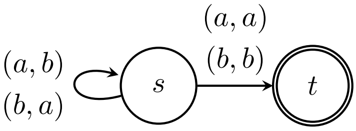


Figure 11.6: A concurrent arena. There exists a DRD almost-surely winning strategy of $\mathcal{P}_1$ from s in the reachability game with target $\{t\}$, but no almost-surely winning RDD strategy.

a finite number of pure finite-memory strategies. Given that there are no probabilities on the transitions of $\mathcal{M}_{a,b}$, for any pure strategy σ_1^{pure}, there is a single outcome under σ_1^{pure}. We can infer that, for any RDD strategy of $\mathcal{M}_{a,b}$, there must be at least one play that has a non-zero probability, and therefore this strategy cannot be outcome-equivalent to σ_1. □

We present a setting in which RDD strategies do not suffice, whereas DRD strategies suffice. We study two-player zero-sum concurrent reachability games. Let $\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$ be a two-player arena, $T \subseteq S$ and $\mathcal{G} = (\mathcal{A}, \mathsf{Reach}(T))$. In $\mathcal{G}$, the goal of $\mathcal{P}_1$ is to maximise the worst-case probability of $\mathsf{Reach}(T)$. The following example illustrates that RDD strategies may not suffice to win almost-surely in $\mathcal{G}$ from a given initial state, whereas DRD strategies do suffice.

Example 11.2. Consider the arena depicted in Figure 11.6 and the two-player zero-sum game $\mathcal{G} = (\mathcal{A}, \mathsf{Reach}(t))$. We first claim that there are no RDD strategies of $\mathcal{P}_1$ that win almost-surely from s. We fix an RDD Mealy machine $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ of $\mathcal{P}_1$ and let $\sigma_1^{\mathfrak{M}}$ denote the strategy it induces. For all $m_{\mathsf{init}} \in \mathsf{supp}(\mu_{\mathsf{init}})$, we consider the pure finite-memory strategy $\sigma_1^{m_{\mathsf{init}}}$ induced by $(M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$. We fix $m_{\mathsf{init}} \in \mathsf{supp}(\mu_{\mathsf{init}})$ and a pure strategy σ_2 of $\mathcal{P}_2$ such that for all histories h ending in s, $\sigma_2(h) \neq \sigma_1^{m_{\mathsf{init}}}(h)$. It follows that $\mathbb{P}_s^{\sigma_1^{m_{\mathsf{init}}},\sigma_2}(\mathsf{Reach}(T)) = 0$. This implies that $\mathfrak{M}$ is not almost-surely winning from s because, by the law of total probability, we have

$$\mathbb{P}_s^{\sigma_1^{\mathfrak{M}},\sigma_2}(\mathsf{Reach}(T)) = \sum_{m \in M} \mu_{\mathsf{init}}(m) \cdot \mathbb{P}_s^{\sigma_1^{m},\sigma_2}(\mathsf{Reach}(T)).$$

On the other hand, the memoryless randomised strategy depicted in Figure 11.5 is almost-surely winning: at each round prior to a visit of t, no matter

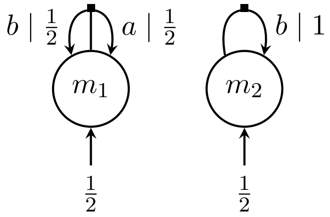


Figure 11.7: An RRD strategy of $\mathcal{M}_{a,b}$ with an infinite co-domain. It witnesses the strictness of the inclusion DRD $\subsetneq$ RRD.

the choices of $\mathcal{P}_2$, this strategy ensures a probability of $\frac{1}{2}$ of matching the action of $\mathcal{P}_2$. ◁

In full generality, there need not exist optimal strategies in concurrent reachability games (see Example 2.4, Page 48). Nonetheless, memoryless randomised strategies (which are a restricted class of DRD strategies) can be used to ensure any possible threshold in these games. In particular, if there exists an optimal strategy, there always exists one that is memoryless. We summarise these results in the following theorem.

Theorem 11.4 ([dAHK07, FBB⁺23])**.** *Let $\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$ be a two-player concurrent arena, $T \subseteq S$ and $\mathcal{G} = (\mathcal{A}, \mathsf{Reach}(T))$ be a zero-sum reachability game. If $\mathcal{P}_1$ can ensure $\theta \in [0, 1]$ in $\mathcal{G}$ from $s \in S$, then there exists a (randomised) memoryless strategy ensuring θ from s.*

11.4 DRD strategies are weaker than RRD ones

We highlight an RRD strategy in $\mathcal{M}_{a,b}$ that has no outcome-equivalent DRD strategy. On the one hand, a DRD strategy has a finite co-domain. Due to its deterministic updates and initialisation, a DRD strategy can only output distributions that are in the co-domain of its next-move function. This is not necessarily the case for RRD strategies.

We consider the Mealy machine of Figure 11.7. Intuitively, this Mealy machine attempts the action a at all steps with a positive probability due to memory state m_1. It also has a positive probability of never playing a due to memory state m_2. Therefore, a is played after a history $s(bs)^k$ with a probability that decreases to zero as k increases, as otherwise a would eventually occur almost-surely.

This behaviour cannot be achieved with a DRD strategy. The distribution over memory states of a DRD strategy following a history is a Dirac distribution due to the deterministic initialisation and deterministic updates. It follows that DRD strategies suggest actions with probabilities given directly by the next-move function, i.e., the image of a DRD strategy is finite. It follows that there is no DRD strategy that is outcome-equivalent to the strategy depicted in Figure 11.7. We formalise this argument in the proof of the following lemma.

Lemma 11.5. *There exists an RRD strategy in $\mathcal{M}_{a,b}$ such that there is no outcome-equivalent DRD strategy.*

Proof. We consider the RRD strategy σ_1 induced by the Mealy machine $\mathfrak{M} = (M, m_1, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ depicted in Figure 11.7. For any $w \in (\{s\}\{a,b\})^*$, let μ_w denote the distribution over M after w as taken place under $\mathfrak{M}$. It can be shown by induction that for any $k \in \mathbb{N}$, $\mu_{(sb)^k}(m_1) = 1 - \mu_{(sb)^k}(m_2) = \frac{1}{2^k+1}$ and for any $w \in (\{s\}\{a,b\})^*$ with at least one occurrence of a, $\mu_w(m_1) = 1$. It follows that for any $k \in \mathbb{N}$, $\sigma_1((sb)^k s)(a) = \frac{1}{2(2^k+1)}$ and $\sigma_1((sb)^k s)(b) = \frac{2^{k+1}+1}{2(2^k+1)}$, and for any history h containing an occurrence of a, $\sigma_1(h)(a) = \sigma_1(h)(b) = \frac{1}{2}$. We obtain that σ_1 plays the action a with positive probabilities that can be arbitrarily small and that all histories of $\mathcal{M}_{a,b}$ are consistent with σ_1.

We now show that no DRD strategy is outcome-equivalent to σ_1. Let $\mathfrak{N} = (N, n_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ denote a DRD strategy and let τ_1 denote its induced strategy. By Lemma 9.1, τ_1 is outcome-equivalent to σ_1 if and only if both strategies are equal, as all histories are consistent with σ_1. For all $h = ws \in \mathsf{Hist}(\mathcal{M}_{a,b})$, due to the deterministic initialisation and updates of $\mathfrak{N}$, we have $\tau_1(h) = \mathsf{nxt}_{\mathfrak{N}}(n, \mathsf{last}(h))$ for $n = \widehat{\mathsf{up}}_{\mathfrak{N}}(w)$. In particular, τ_1 cannot play the action a with arbitrarily small positive probabilities as it can only assign finitely many distributions to histories. We conclude that $\tau_1 \neq \sigma_1$, which ends the proof. □

The Mealy machine of Figure 11.7 is based on the finite-memory positively winning strategies of $\mathcal{P}_2$ of [CDH10] for the snowball game. We have presented

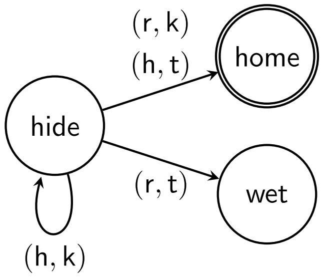


Figure 11.8: The arena of the snowball game of [dAHK07].

the snowball game in Example 2.4 (Page 48); we recall its arena $\mathcal{A}$ in Figure 11.8. Let $\mathcal{G} = (\mathcal{A}, \mathsf{Reach}(\mathsf{home}))$.

We have previously seen that $\mathcal{P}_1$ does not have an optimal strategy from hide in $\mathcal{G}$ by analysing the memoryless strategies of $\mathcal{P}_1$. Another approach to show this is to show the existence of a positively winning strategy of $\mathcal{P}_2$ from hide in $\mathcal{G}$, i.e., a strategy σ_2 such that, for all strategies σ_1 of $\mathcal{P}_1$, $\mathbb{P}^{\sigma_1,\sigma_2}_{\mathsf{hide}}(\mathsf{Reach}(\mathsf{home})) > 0$. A positively winning strategy of $\mathcal{P}_2$ is shown to exist in [dAHK07] in $\mathcal{G}$, although it is shown that DRD strategies do not suffice. An RRD positively winning strategy of $\mathcal{P}_2$ is provided in [CDH10]. This Mealy machine can be obtained by renaming the outputs a and b in the Mealy machine of Figure 11.7 by t and k respectively. This strategy is positively winning because it has a positive probability of never throwing the snowball while having a positive probability of throwing it at every round. Therefore, no matter when $\mathcal{P}_1$ chooses to run, there is a positive probability of them being hit by a snowball.

More generally, in a two-player zero-sum concurrent reachability game, $\mathcal{P}_2$ has a positively winning strategy from any state that is not almost-surely winning for $\mathcal{P}_1$. It is argued in [CDH10] that RRR strategies are sufficient for $\mathcal{P}_2$ in this setting. We build on their construction to show that RRD strategies suffice. We show the equivalent property that RRD strategies suffice to win positively in games with *safety objectives* for $\mathcal{P}_1$.

We let $\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$ be a two-player arena, $T \subseteq S$ and let $\mathcal{G} = (\mathcal{A}, \mathsf{Safe}(T))$. The following properties are a consequence of the (correctness proof of the) algorithm of [dAHK07] to compute states that are almost-surely winning in concurrent zero-sum reachability games. Each state in $\mathcal{G}$ can be assigned a rank. States of highest rank are those from which $\mathcal{P}_2$ wins almost-surely for their dual reachability objective $\mathsf{Reach}(T)$. States of minimal rank, if they are not simultaneously of maximal rank, are those from which $\mathcal{P}_1$ can surely enforce the safety objective no matter the strategy of $\mathcal{P}_2$, i.e., $\mathcal{P}_1$ has a (memoryless) strategy such that all plays consistent with this strategy that start from a state of minimal rank satisfy the safety objective.

Let $s \in S$ be a state that is positively winning. There exists an action of $\mathcal{P}_1$, which we will call a *sound action*, and a set $A^{(2)}_\star(s) \subseteq A^{(2)}(s)$ of actions of $\mathcal{P}_2$ such that the sound action surely prevents moving to states of higher rank against all actions in $A^{(2)}_\star(s)$. Furthermore, for actions of $\mathcal{P}_2$ outside of $A^{(2)}_\star(s)$, there is an action of $\mathcal{P}_1$ that moves to a state of strictly lower rank with positive probability. For instance, in the snowball game (Figure 11.8), seen as a safety game from the perspective of $\mathcal{P}_2$, the action k is a sound action for hide with respect to $A^{(2)}_\star(s) = \{\mathsf{h}\}$.

The property we require on our strategy to win positively is to use a strategy much like that of Figure 11.7. On the one hand, it must have a positive probability of only using sound actions from any point: this way, the safety objective is ensured whenever $\mathcal{P}_2$ only uses actions in the sets of the form $A^{(2)}_\star(s)$ in the remainder of the play. On the other hand, to account for the possibility of $\mathcal{P}_2$ taking an action outside of $A^{(2)}_\star(s)$ in state s, all actions should have a positive probability of occurring in all rounds, so a vertex of lower rank can be reached with positive probability in this case.

Because the state space is finite, one of two cases occurs. If $\mathcal{P}_2$ only resorts to actions compatible with sound actions from some point on, then the safety objective is satisfied with positive probability because sound actions are guaranteed to be always played from some point on with positive probability. Otherwise, states of minimal ranks are reached with positive probability, from which $\mathcal{P}_1$ can surely avoid T.

The idea of the RRR strategy proposed in [CDH10] to obtain the behaviour described above is to rely on pairs of memory states. In a pair, one memory state only proposes sound actions and the other memory state suggests all actions uniformly at random. When initialising the Mealy machine and each time there is a change in the rank of states, to ensure the resulting strategy has the property above, a stochastic memory update is used to give a uniform probability over such a pair of states.

We show that it suffices to randomise once at the start, for each rank (besides the maximum and minimum one), whether only sound actions should be suggested or whether we should play uniformly at random. This allows us to avoid stochastic updates and obtain an RRD strategy.

Theorem 11.6. *Let $\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$ be a two-player arena, $T \subseteq S$ and $\mathcal{G} = (\mathcal{A}, \mathsf{Safe}(T))$ be a zero-sum safety game. There exists an RRD strategy $\mathfrak{M}$ such that, for all $s_{\mathsf{init}} \in S$, if there exists a positively winning strategy from s_{init} for the objective $\mathsf{Safe}(T)$, then $\mathfrak{M}$ is positively winning from s_{init}.*

Proof. We assume that there exists at least some state from which $\mathcal{P}_1$ wins positively, otherwise the result is immediate. We use properties of [dAHK07, Algorithm 3], which computes the set of almost-surely winning states in a concurrent reachability game, i.e., the complement of the set of positively winning states for the player with a safety objective. Each iteration of this algorithm computes two sets of states that are positively winning for $\mathcal{P}_1$ and (essentially) removes them from the state space. Therefore, it yields a non-increasing sequence $S = U_0 \supseteq U_1 \ldots \supseteq U_k$ of sets of states ($k+2$ being double the number of iterations of the algorithm) such that $S \setminus U_k$ is the set of positively winning states for $\mathcal{P}_1$. In particular, note that $T \subseteq U_k$. Let, for all $s \in S$, $\mathsf{rk}(s)$ be the greatest j such that $s \in U_j$.

The sequence of sets $(U_j)_{1 \leq j \leq k}$ has the following property. For all states $s \in S$ such that $\mathsf{rk}(s) < k$, there exists a sound action $a^{(1)}_{\mathsf{sd}}(s) \in A^{(1)}(s)$ and a subset $A^{(2)}_{\star}(s) \subseteq A^{(2)}(s)$ such that

(i) for all $a^{(2)} \in A^{(2)}_{\star}(s)$ and all $s' \in \mathsf{supp}(\delta(s, a^{(1)}_{\mathsf{sd}}(s), a^{(2)}))$, $\mathsf{rk}(s') \leq \mathsf{rk}(s)$, and

(ii) for all $a^{(2)} \in A^{(2)}(s) \setminus A^{(2)}_{\star}(s)$, there exists an action $a^{(1)} \in A^{(1)}(s)$ and a state $s' \in \mathsf{supp}(\delta(s, a^{(1)}, a^{(2)}))$ such that $\mathsf{rk}(s') < \mathsf{rk}(s)$.

These conditions follow from the structure of the algorithm. In particular, the pure memoryless strategy of $\mathcal{P}_1$ that only plays sound actions, when played from states of rank 0, is such that all of its outcomes satisfy $\mathsf{Safe}(T)$ (i.e., states of rank 0 are surely winning for $\mathcal{P}_1$).

We now define an RRD strategy. Let $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ such that $M = \{\mathsf{sd}, \mathsf{un}\}^{k-1}$ (sd and un respectively stand for sound and uniform). We let μ_{init} be a uniform distribution over M. Let $m = (m_j)_{1 \leq j \leq k-1} \in M$ and $s \in S$. If $\mathsf{rk}(s) = k$, we let $\mathsf{nxt}_{\mathfrak{M}}(m, s)$ be arbitrary. Otherwise, if $\mathsf{rk}(s) = 0$ or $m_{\mathsf{rk}(s)} = \mathsf{sd}$, we let $\mathsf{nxt}_{\mathfrak{M}}(m, s)$ be a Dirac distribution on $a^{(1)}_{\mathsf{sd}}(s)$. Otherwise (if $0 < \mathsf{rk}(s) < k$ and $m_{\mathsf{rk}(s)} = \mathsf{un}$), we let $\mathsf{nxt}_{\mathfrak{M}}(m, s)$ be a uniform distribution over $A^{(1)}(s)$. The deterministic memory updates are trivial: for all $m \in M$, $s \in S$ and $\bar{a} \in \bar{A}(s)$, we let $\mathsf{up}_{\mathfrak{M}}(m, s, \bar{a}) = m$. Given $w \in (S\bar{A})^*$, we let μ_w denote the distribution over memory states of $\mathfrak{M}$ after w has taken place. For $m \in M$, we let σ_1^m be the strategy induced by the Mealy machine obtained by fixing the initial state of $\mathfrak{M}$ to m.

We now prove that $\mathfrak{M}$ induces a positively winning strategy from any state from which $\mathcal{P}_1$ has a positively winning strategy. Let s_0 be such a state and let σ_2 be an arbitrary strategy of $\mathcal{P}_2$. We use an inductive argument on histories, starting with the history $h_0 = s_0$. At step j of the induction, we assume that we have some history $h_j = w_j s_j$ consistent with σ_2 such that $\mathsf{rk}(s_j) < k - j$ and $\mathsf{supp}(\mu_{w_j}) = \{\mathsf{sd}, \mathsf{un}\}^{\mathsf{rk}(s_j)} \times M_j$ for some $M_j \subseteq \{\mathsf{sd}, \mathsf{un}\}^{k-\mathsf{rk}(s_j)}$ (this last hypothesis implies that h_j is consistent with $\mathfrak{M}$, otherwise μ_{w_j} would not be defined). This induction hypothesis is clearly satisfied at step 0 of the induction (positively winning states have rank at most $k-1$).

We consider two cases. First, we assume that, for all extensions $w_j h$ of h_j, if they are consistent with σ_2 and only sound actions are used by $\mathcal{P}_1$ in the suffix h, then $\mathsf{supp}(\sigma_2(w_j h)) \subseteq A^{(2)}_\star(\mathsf{last}(h))$. We remark that if $\mathsf{rk}(s_j) = 0$, we are necessarily in this case. We claim that for all extensions $w_j h$ of h_j consistent with σ_2 in which only sound $\mathcal{P}_1$ actions occur in h, it holds that all states in h have rank at most $\mathsf{rk}(s_j)$. This follows by a straightforward induction using the definition of sound actions and actions in sets $A^{(2)}_\star(s')$ (informally, the rank of states cannot increase at each step in this setting).

By the induction hypothesis, there exists some $m \in \mathsf{supp}(\mu_{w_j})$ such that $m_\ell = \mathsf{sd}$ for all $\ell \leq \mathsf{rk}(s_j)$. In particular, h_j is consistent with σ_1^m due to the definition of updates in $\mathfrak{M}$. It follows from the above that all extensions of h_j that are consistent with both σ_1^m and σ_2 satisfy $\mathsf{Safe}(T)$ (because all targets have rank k). Therefore, only a subset of $\mathsf{Cyl}(h_j)$ of $\mathbb{P}_s^{\sigma_1^m, \sigma_2}$-measure zero is not included in $\mathsf{Safe}(T)$. Therefore, $\mathbb{P}_s^{\sigma_1^m, \sigma_2}(\mathsf{Safe}(T)) \geq \mathbb{P}_s^{\sigma_1^m, \sigma_2}(\mathsf{Cyl}(h_j)) > 0$.

We conclude that $\mathbb{P}_s^{\sigma_1,\sigma_2}(\mathsf{Safe}(T)) > 0$ as $\mathbb{P}_s^{\sigma_1^m,\sigma_2}(\mathsf{Safe}(T))$ is the conditional probability of $\mathsf{Safe}(T)$ with respect to $\mathbb{P}_s^{\sigma_1,\sigma_2}$ assuming that the initial memory state is m.

Next, assume that there exists a history w_jh extending h_j that is consistent with σ_2, in which only sound actions are used by $\mathcal{P}_1$ in the suffix h and such that $\mathsf{supp}(\sigma_2(w_jh)) \nsubseteq A_\star^{(2)}(\mathsf{last}(h))$. We assume that w_jh is the shortest such extension of h_j. We fix $a^{(2)} \in \mathsf{supp}(\sigma_2)(w_jh) \setminus A_\star^{(2)}(\mathsf{last}(h))$, and $a^{(1)} \in A^{(1)}(\mathsf{last}(h))$ and $s_{j+1} \in \mathsf{supp}(\delta(\mathsf{last}(h), a^{(1)}, a^{(2)}))$ such that $\mathsf{rk}(s_{j+1}) < \mathsf{rk}(\mathsf{last}(h))$. We let $\bar{a} = (a^{(1)}, a^{(2)})$.

We define $h_{j+1} = w_jh\bar{a}s_{j+1}$ and show that it satisfies the induction hypothesis above. First, by construction, h_{j+1} is consistent with σ_2. Second, it holds that $\mathsf{rk}(\mathsf{last}(h)) \leq \mathsf{rk}(s_j)$. This can be shown by the same argument as in the first case, as only sound actions occur in h and all $\mathcal{P}_2$ actions taken in any state s in h are in $A_\star^{(2)}(s)$. It follows that $\mathsf{rk}(s_{j+1}) < \mathsf{rk}(s_j)$, implying that $\mathsf{rk}(s_{j+1}) < k - (j+1)$. Third, it can be shown by a straightforward induction that $\mathsf{supp}(\mu_w) = \mathsf{supp}(\mu_{w_j})$ for w such that $w_jh = w\mathsf{last}(h)$. The omitted inductive argument is based on the fact that all $\mathcal{P}_1$ actions are sound in h, are taken in states of rank at most $\mathsf{rk}(s_j)$ and $\mathsf{supp}(\mu_{w_j}) = \{\mathsf{sd}, \mathsf{un}\}^{\mathsf{rk}(s_j)} \times M_j$. Finally, it holds that $\mathsf{supp}(\mu_{w_jh\bar{a}}) = \{m \in \mathsf{supp}(\mu_{w_j}) \mid m_{\mathsf{rk}(\mathsf{last}(h))} = \mathsf{un}\}$ if $a^{(1)} \neq a_{\mathsf{sd}}^{(1)}(\mathsf{last}(h))$ and $\mathsf{supp}(\mu_{w_jh\bar{a}}) = \mathsf{supp}(\mu_{w_j})$ otherwise. By the inductive hypothesis, we obtain that

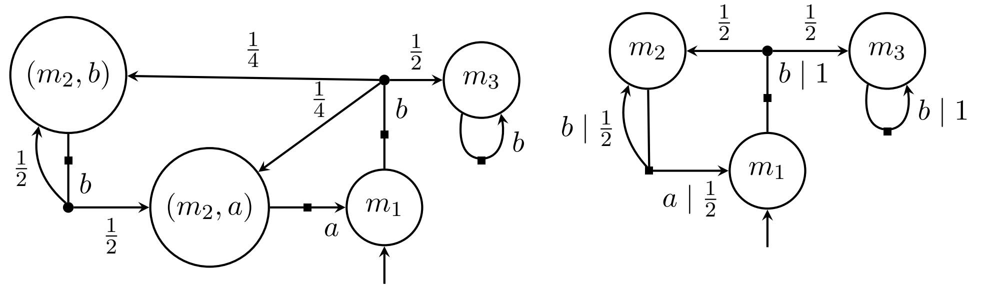
$$\mathsf{supp}(\mu_{w_jh\bar{a}}) = \{\mathsf{sd}, \mathsf{un}\}^{\mathsf{rk}(\mathsf{last}(h))-1} \times I \times \{\mathsf{sd}, \mathsf{un}\}^{\mathsf{rk}(s_j)-\mathsf{rk}(\mathsf{last}(h))} \times M_j,$$

where $I = \{\mathsf{un}\}$ in the first case, and $I = \{\mathsf{sd}, \mathsf{un}\}$ otherwise. This shows that we can continue the inductive argument with h_{j+1}.

The second case can occur in the worst case only in the $k-1$ first steps of the induction: at step k, s_k has rank 0, which guarantees we find ourselves in the first case. This concludes the proof that $\mathfrak{M}$ is positively winning from s_0. □

11.5 RRD and DDR strategies are incomparable

We prove in this section that the classes RRD and DDR of finite-memory strategies are incomparable. We have previously shown Lemma 11.1, which



(a) A DDR strategy witnessing DDR $\not\subseteq$ RRD.

(b) An outcome-equivalent RRR strategy with fewer states.

Figure 11.9: Outcome-equivalent strategies witnessing the non-inclusion DDR $\not\subseteq$ RRD. For the sake of readability, we do not label transitions by s. We omit the probability of actions in Figure 11.9a as outputs are deterministic.

states that RDD $\not\subseteq$ DDR and therefore implies that DRD $\not\subseteq$ DDR and RRD $\not\subseteq$ DDR. It remains to show that DDR $\not\subseteq$ RRD.

We illustrate a DDR strategy of $\mathcal{M}_{a,b}$ that has no outcome-equivalent RRD strategy in Figure 11.9a. For ease of analysis, we illustrate in Figure 11.9b a DRR strategy with fewer states that is outcome-equivalent to the Mealy machine depicted in Figure 11.9a. The DDR strategy of Figure 11.9a can be obtained by applying the construction of Theorem 10.5 to the Mealy machine of Figure 11.9b.

Intuitively, these strategies have a non-zero probability of never using action a after any history, while they have a positive probability of using action a at any time besides the first round and right after the action a occurs. We formally prove this property below.

Lemma 11.7. *Let σ_1 denote the strategy of $\mathcal{M}_{a,b}$ induced by the Mealy machines of Figure 11.9. For all histories $h \in \mathsf{Hist}(\mathcal{M}_{a,b})$ consistent with σ_1 in which no action appears or in which the last used action is a, $\mathbb{P}_s^{\sigma_1}(\{h(bs)^\omega\}) > 0$.*

Proof. First, we provide a partial definition of σ_1. For any $w \in (\{s\}\{a, b\})^*$, let μ_w denote the distribution over memory states of $\mathfrak{M}$ after w has taken place. It can be shown by induction that for any $w \in ((\{s\}\{b\})^+\{s\}\{a\})^*$ and $k \geq 1$, we have $\mu_w(m_1) = 1$ and $\mu_{w(sb)^k}(m_2) = 1 - \mu_{w(sb)^k}(m_3) = \frac{1}{2^{k-1}+1}$. It follows that for any history h consistent with σ_1 of the form s or $h'as$ and $k \geq 1$, we have $\sigma_1(h)(b) = 1$ and $\sigma_1(h(bs)^k)(a) = 1 - \sigma_1(h(bs)^k)(b) = \frac{1}{2^k+2}$.

Now, fix a history h consistent with σ_1 such that there are no actions or such that the last action is a. Next, we show that $\mathbb{P}_s^{\sigma_1}(\{h(bs)^\omega\}) = \mathbb{P}_s^{\sigma_1}(\mathsf{Cyl}\,(h)) \cdot \mathbb{P}_s^{\sigma_1}(\{(sb)^\omega\})$. We have, for any $k \in \mathbb{N}$, $\sigma_1(h(bs)^k)(b) = \sigma_1(s(bs)^k)(b)$ by definition of σ_1. Furthermore, the cylinder sequences $(\mathsf{Cyl}\left(s(bs)^k\right))_{k\in\mathbb{N}}$ and $(\mathsf{Cyl}\left(h(bs)^k\right))_{k\in\mathbb{N}}$ respectively decrease when taking their intersections to the singletons $\{(sb)^\omega\}$ and $\{h(bs)^\omega\}$. We obtain the following equations from the definition of $\mathbb{P}_s^{\sigma_1}$:

$$\begin{aligned}\mathbb{P}_s^{\sigma_1}(\{h(bs)^\omega\}) &= \lim_{k\to\infty} \mathbb{P}_s^{\sigma_1}\left(\mathsf{Cyl}\left(h(bs)^k\right)\right)\\ &= \lim_{k\to\infty} \mathbb{P}_s^{\sigma_1}(\mathsf{Cyl}\,(h)) \cdot \prod_{\ell=0}^{k-1} \sigma_1(h(bs)^\ell)(b)\\ &=\mathbb{P}_s^{\sigma_1}(\mathsf{Cyl}\,(h)) \cdot \lim_{k\to\infty} \cdot \prod_{\ell=0}^{k-1} \sigma_1(s(bs)^\ell)(b)\\ &=\mathbb{P}_s^{\sigma_1}(\mathsf{Cyl}\,(h)) \cdot \lim_{k\to\infty} \mathbb{P}_s^{\sigma_1}\left(\mathsf{Cyl}\left(s(bs)^k\right)\right)\\ &=\mathbb{P}_s^{\sigma_1}(\mathsf{Cyl}\,(h)) \cdot \mathbb{P}_s^{\sigma_1}(\{(sb)^\omega)\}).\end{aligned}$$

In light of the above, to show that $\mathbb{P}_s^{\sigma_1}(\{h(bs)^\omega\}) > 0$, it suffices to establish that $\mathbb{P}_s^{\sigma_1}(\{(sb)^\omega\}) > 0$ because h is assumed to be consistent with σ_1. It can be shown that $\mathbb{P}_s^{\sigma_1}(\{(sb)^\omega\}) = \frac{1}{2}$ as follows:

$$\begin{aligned}\mathbb{P}_s^{\sigma_1}(\{(sb)^\omega\}) &= \lim_{k\to\infty} \mathbb{P}_s^{\sigma_1}\left(\mathsf{Cyl}\left(s(bs)^k\right)\right)\\ &= \lim_{k\to\infty} 1 \cdot \prod_{j=1}^{k-1} \frac{2^j+1}{2^j+2}\\ &= \lim_{k\to\infty} \frac{1}{2^{k-1}} \cdot \prod_{j=1}^{k-1} \frac{2^j+1}{2^{j-1}+1}\\ &= \lim_{k\to\infty} \frac{1}{2^{k-1}} \cdot \frac{2^{k-1}+1}{2^{1-1}+1} = \frac{1}{2};\end{aligned}$$

the product of the probabilities of b being played in each round is simplified using the fact that the denominator of a term is double the numerator of the previous one. This closes the proof of our claimed inequality. □

The property stated in Lemma 11.7 cannot be reproduced by an RRD strategy. There are two reasons to this.

First, along any play consistent with an RRD strategy, the support of the distribution over memory states cannot increase in size. Because of deterministic updates, the probability carried by a memory state m can only be transferred to at most one state, and may be lost if the used action cannot be used while in m. We formally prove this observation below.

Lemma 11.8. *Let $n \in \mathbb{N}_{>0}$, $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$ be an n-player arena and $i \in [\![1, n]\!]$. Let $\mathfrak{N} = (N, \nu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ be an RRD strategy of $\mathcal{P}_i$ in $\mathcal{A}$. Let $w = w's\bar{a}$ be consistent with $\mathfrak{N}$. Let ν_w and $\nu_{w'}$ denote the distributions over N after w and w' have taken place under $\mathfrak{N}$. Then*

(i) $|\mathsf{supp}(\nu_w)| \leq |\mathsf{supp}(\nu_{w'})|$ *and*

(ii) if there exists $n \in \mathsf{supp}(\nu_{w'})$ such that $\mathsf{nxt}_{\mathfrak{N}}(n, s)(a^{(i)}) = 0$, then the previous inequality is strict.

Proof. For any memory state $n \in N$, recall that

$$\nu_w(n) = \frac{\sum_{n'\in N} \nu_{w'}(n') \cdot \mathsf{up}_{\mathfrak{N}}(n', s, \bar{a})(n) \cdot \mathsf{nxt}_{\mathfrak{N}}(n', s)(a^{(i)})}{\sum_{m'\in M} \nu_{w'}(n') \cdot \mathsf{nxt}_{\mathfrak{N}}(n', s)(a^{(i)})}.$$

In particular, $n \in \mathsf{supp}(\nu_w)$ if and only if there exists $n' \in \mathsf{supp}(\nu_w)$ such that $\mathsf{up}_{\mathfrak{N}}(n', s, \bar{a})(n) > 0$ and $\mathsf{nxt}_{\mathfrak{N}}(n', s)(a^{(i)}) > 0$, i.e., the probability of elements of $\mathsf{supp}(\nu_w)$ comes from elements of $\mathsf{supp}(\nu_{w'})$ in which $a^{(i)}$ is played with positive probability in s. Because updates are deterministic, for any given $n' \in N$, there is a unique $n \in N$ such that $\mathsf{up}_{\mathfrak{N}}(n', s, a^{(i)})(n) = 1$. Therefore, any element $n' \in \mathsf{supp}(\nu_{w'})$ transfers its probability to at most one memory state when deriving ν_w and this probability is transferred only if $\mathsf{nxt}_{\mathfrak{N}}(n', s)(a^{(i)}) > 0$. Both (i) and (i) follow. □

The property of RRD strategies presented in Lemma 11.8 does not hold for strategies that have stochastic updates, such as those of Figure 11.9.

Second, we can engineer situations in which the size of the support of the distribution over memory states of an RRD strategy must decrease. If after a given history h, the action a has a positive probability of never being used despite being assigned a positive probability at each round after h, then at some point there must be some memory state of the RRD strategy that has positive probability and that assigns (via the next-move function) probability zero to action a. For instance, this is the case from the start with the RRD strategy depicted in Figure 11.7. Intuitively, if at all times, all memory states in the support of the distribution over memory states after the current history assign a positive probability to action a, the probability of using a at each round after h would be bounded from below by the smallest positive probability assigned to a by the next-move function. Therefore a would eventually be played almost-surely assuming h has taken place, contradicting the fact that there was a positive probability of never using action a after h. By using action a at a point in which some memory state in the support of the distribution over memory states assigns probability zero to a, the size of the support of the memory state distribution decreases.

By design of our DDR strategy, if one assumes the existence of an outcome-equivalent RRD strategy, then it is possible to construct a play along which the size of the support of the distribution over memory states of the RRD strategy decreases infinitely often. Because this size cannot increase along a play, this is not possible, i.e., there is no such RRD strategy. We formalise the sketch above in the proof of the following lemma.

Lemma 11.9. *There exists a DDR strategy in $\mathcal{M}_{a,b}$ such that there is no outcome-equivalent RRD strategy.*

Proof. Consider the Mealy machine $\mathfrak{M} = (M, m_1, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ depicted in Figure 11.9b and let σ_1 denote the strategy induced by $\mathfrak{M}$. We recall that $\mathfrak{M}$ is a DRR Mealy machine that is outcome-equivalent to the DDR strategy illustrated in Figure 11.9a. It therefore suffices to show that there are no RRD strategies that are outcome-equivalent to σ_1 to end the proof. Let $\mathfrak{N} = (N, \nu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$ be an RRD Mealy machine and let τ_1 be the strategy it induces.

We assume that σ_1 and τ_1 are outcome-equivalent towards a contradiction. By Lemma 11.7, for all histories $h \in \mathsf{Hist}(\mathcal{M}_{a,b})$ that are consistent with σ_1 such that either there are no actions in h or such that the last action is a, we have $\mathbb{P}_s^{\sigma_1}(\{h(bs)^\omega\}) > 0$.

For any $w \in (\{s\}\{a,b\})^*$, let ν_w denote the distribution over memory states in N after w has taken place under $\mathfrak{N}$. We show that for any history h consistent with τ_1, if the probability of a never appearing again after h is non-zero, i.e., if $\mathbb{P}_s^{\tau_1}(\{h(bs)^\omega\}) > 0$, and for any $k \in \mathbb{N}$, we have $\tau_1(h(bs)^k)(a) > 0$, then there exists some $k_0 \in \mathbb{N}$ such that $|\mathsf{supp}(\nu_{h(bs)^{k_0}b})| > |\mathsf{supp}(\nu_{h(bs)^{k_0+1}a})|$.

Let h be consistent with τ_1. Assume that $\mathbb{P}_s^{\tau_1}(\{h(bs)^\omega\}) > 0$, and for any $k \in \mathbb{N}$, we have $\tau_1(h(bs)^k)(a) > 0$. By Lemma 11.8 (on the support of the distributions ν_w), we need only show that for some $k_0 \in \mathbb{N}$, there exists $n \in \mathsf{supp}(\nu_{h(bs)^{k_0}b})$ such that $\mathsf{nxt}_{\mathfrak{N}}(n,s)(a) = 0$. Assume towards a contradiction that this is not the case, i.e., for all $k \in \mathbb{N}$ and all $n \in \mathsf{supp}(\nu_{h(bs)^k b})$, we have $\mathsf{nxt}_{\mathfrak{N}}(n,s)(a) > 0$. Let $k \in \mathbb{N}$. The probability $\tau_1(h(bs)^{k+1})(a)$ is bounded below by the positive constant

$$\min\left\{\mathsf{nxt}_{\mathfrak{N}}(n,s)(a) \mid n \in N \text{ s. t. } \mathsf{nxt}_{\mathfrak{N}}(n,s)(a > 0\right\}.$$

It follows that the action a must be used almost-surely assuming h has taken place, contradicting the fact that $\mathbb{P}_s^{\tau_1}(\{h(bs)^\omega\}) > 0$. This ends the proof of the above claim.

We can repeatedly use the property shown above to construct a sequence of non-zero natural numbers $(k_\ell)_{\ell\in\mathbb{N}}$ such that $(|\mathsf{supp}(\nu_{w_\ell})|)_{\ell\in\mathbb{N}}$ is an infinite decreasing sequence, where $w_0 = \varepsilon$ and for all $\ell \in \mathbb{N}$, $w_{\ell+1} = w_\ell(sb)^{k_\ell}sa$. This contradicts the well-order of $\mathbb{N}$. This shows that there are no RRD strategies that are outcome-equivalent to σ_1. □

As in the previous sections, we provide a game and a specification that cannot be accomplished using an RRD strategy, but can be accomplished using a DDR strategy. In the following example, we consider a two-player turn-based game with several reachability objectives with absorbing targets. The goal is to construct, if it exists, a strategy that ensures given thresholds for several reachability objectives at once.

Example 11.3. Let $\mathcal{A} = (S_1, S_2, A, \delta)$ be the two-player turn-based arena

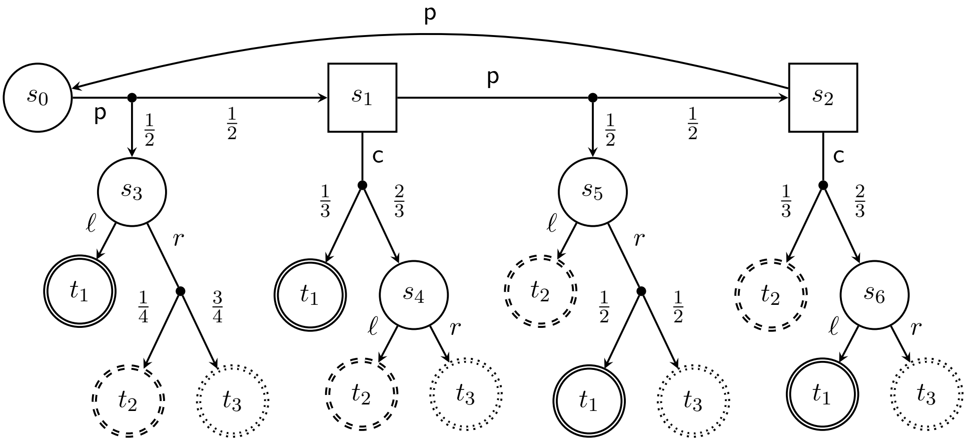


Figure 11.10: A turn-based stochastic game with multiple reachability objectives [CFK⁺13a]. Circles and squares respectively represent states controlled by $\mathcal{P}_1$ and $\mathcal{P}_2$. States t_1, t_2 and t_3 are drawn repeatedly for clarity (duplicates all represent the same state). Actions p and c stand for proceed and check respectively.

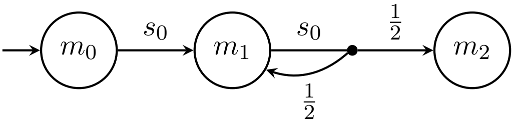


Figure 11.11: A Mealy machine update scheme for the arena of Figure 11.10. Updates depend only on states, not on actions. Updates that do not change the memory state are not depicted.

depicted in Figure 11.10, originating from [CFK+13a]. We consider three targets: $T_j = \{t_j\}$ for $j \in [\![1, 3]\!]$. In [CFK+13a], it is shown that there is no DRD strategy σ_1 of $\mathcal{P}_1$ such that for all strategies σ_2 of $\mathcal{P}_2$, $\mathbb{P}^{\sigma_1,\sigma_2}_{s_0}(\mathsf{Reach}(T_j)) \geq \frac{1}{3}$ for all $j \in [\![1, 3]\!]$, despite there existing an infinite-memory one. We prove that

(i) there is no RRD strategy that satisfies this specification and

(ii) there exists a DDR strategy that does.

We let, for $k \in \mathbb{N}$, $h_k = s_0(\mathsf{p}s_1\mathsf{p}s_2\mathsf{p}s_0)^k$. A description of satisfactory strategies is provided in the technical report [CFK+13b, App. B]. A strategy σ_1 of $\mathcal{P}_1$ ensures that all targets are visited with probability $\frac{1}{3}$ if for all $k \in \mathbb{N}$, $\sigma_1(h_k\mathsf{p}s_3)(\ell) = 1 - \frac{1}{3 \cdot 2^{k-1}}$, $\sigma_1(h_k\mathsf{p}s_1\mathsf{c}s_4)(\ell) = 1 - \frac{1}{2^{k+2}}$, $\sigma_1(h_k\mathsf{p}s_1\mathsf{p}s_5)(\ell) = 1 - \frac{1}{3 \cdot 2^k}$ and $\sigma_1(h_k\mathsf{p}s_1\mathsf{c}s_6)(\ell) = 1 - \frac{1}{2^{k+2}}$, and for all $k \in \mathbb{N}$, the first two equations are necessary to comply with the specification.

Let $\mathfrak{M}$ be an RRD strategy and let $\tau_1^{\mathfrak{M}}$ be its induced strategy. We show that $\tau_1^{\mathfrak{M}}$ cannot satisfy the multi-objective query by showing that the set of distributions $\{\tau_1^{\mathfrak{M}}(h_k\mathsf{p}s_3) \mid k \in \mathbb{N}\}$ must be a finite set, which is incompatible with the requirements given above.

Let μ_w denote the distribution over memory states after $w \in (SA)^*$ has taken place under $\mathfrak{M}$. For all $k \in \mathbb{N}$ and $m \in M$, it holds that $\mu_{h_k\mathsf{p}}(m) = \sum_{m' \in M'} \mu_{\mathsf{init}}(m')$ for some $M' \subseteq M$ (which depends on both k and m). This follows from the equations for the updates of the distributions μ_w. In all states along $h_k\mathsf{p}$, $\mathcal{P}_1$ only has a single action. Furthermore, $\mathfrak{M}$ has deterministic updates. Therefore, if w and wsa are prefixes of $h_k\mathsf{p}$, for all memory states $m \in M$, we obtain that $\mu_{wsa}(m)$ is the sum of $\mu_w(m')$ for all memory states m' such that $\mathsf{up}_{\mathfrak{M}}(m', s, a) = m$. In particular, this implies that the set of distributions $\{\mu_{h_k\mathsf{p}} \mid k \in \mathbb{N}\}$ is finite, which shows that $\{\tau_1^{\mathfrak{M}}(h_k\mathsf{p}s_3) \mid k \in \mathbb{N}\}$ is a finite set by definition of the strategy induced by a Mealy machine.

We now describe a Mealy machine $\mathfrak{N}$ that induces a strategy that coincides with σ_1 over $\mathsf{Cyl}(s_0)$, i.e., that ensures a probability of $\frac{1}{3}$ for all three reachability objectives. As in the proof of Lemma 11.9, we provide a DRR strategy that can be transformed into an outcome-equivalent DDR strategy via the construction underlying Theorem 10.5. We depict the relevant update scheme in Figure 11.11; updates that do not change the current memory state are omitted from the figure. Let ν_w denote the distribution over memory states of $\mathfrak{N}$ after $w \in (SA)^*$

has taken place under $\mathfrak{N}$. Let $k \in \mathbb{N}$. Below, we are interested in the distribution over memory states only for $w_k \in \{h_k\mathsf{p}, h_k\mathsf{p}s_1\mathsf{c}, h_k\mathsf{p}s_1\mathsf{p}, h_k\mathsf{p}s_1\mathsf{p}s_2\mathsf{c}\}$: it can be shown by a straightforward induction that we have $\nu_{w_k}(m_1) = 1 - \nu_{w_k}(m_2) = \frac{1}{2^k}$.

We now specify the next-move function of $\mathfrak{N}$ and describe the strategy $\sigma_1^{\mathfrak{N}}$ induced by $\mathfrak{N}$. We let $\mathsf{nxt}_{\mathfrak{M}}(m_0, s)$ be an arbitrary Dirac distribution for all states $s \in \{s_3, s_4, s_5, s_6\}$ (we require Dirac distributions so our Mealy machine has an outcome-equivalent DDR strategy). For s_3, we let $\mathsf{nxt}_{\mathfrak{M}}(m_1, s_3)(r) = \frac{2}{3}$ and $\mathsf{nxt}_{\mathfrak{M}}(m_2, s_3)(\ell) = 1$. It follows that for all $k \in \mathbb{N}$, we have $\sigma_1^{\mathfrak{N}}(h_k\mathsf{p}s_3)(r) = \frac{2}{3 \cdot 2^k} = \frac{1}{3 \cdot 2^{k-1}}$. For s_4, we let $\mathsf{nxt}_{\mathfrak{M}}(m_1, s_4)(r) = \frac{1}{4}$ and $\mathsf{nxt}_{\mathfrak{M}}(m_2, s_4)(\ell) = 1$. We obtain that for all $k \in \mathbb{N}$, we have $\sigma_1^{\mathfrak{N}}(h_k\mathsf{p}s_2\mathsf{c}s_4)(r) = \frac{1}{4 \cdot 2^k} = \frac{1}{2^{k+2}}$. For s_5, we let $\mathsf{nxt}_{\mathfrak{M}}(m_1, s_5)(r) = \frac{1}{3}$ and $\mathsf{nxt}_{\mathfrak{M}}(m_2, s_5)(\ell) = 1$. For all $k \in \mathbb{N}$, it holds that $\sigma_1^{\mathfrak{N}}(h_k\mathsf{p}s_2\mathsf{p}s_5)(r) = \frac{1}{3 \cdot 2^k}$. Finally, for s_6, we let $\mathsf{nxt}_{\mathfrak{M}}(m_1, s_6)(r) = \frac{1}{4}$ and $\mathsf{nxt}_{\mathfrak{M}}(m_2, s_6)(\ell) = 1$. We conclude that for all $k \in \mathbb{N}$, $\sigma_1^{\mathfrak{N}}(h_k\mathsf{p}s_2\mathsf{p}s_2\mathsf{c}s_6)(r) = \frac{1}{4 \cdot 2^k} = \frac{1}{2^{k+2}}$. This shows that $\sigma_1^{\mathfrak{N}}$ ensures all reachability objectives are satisfied with probability at least $\frac{1}{3}$. $\lhd$

Consider a turn-based stochastic arena $\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$ and targets $T_1, \ldots, T_d \subseteq S$. The general form of the problem from Example 11.3 is to decide, given an initial state $s_{\mathsf{init}} \in S$ and a threshold vector $\mathbf{q} = (q_j)_{1 \leq j \leq d} \in ([0,1] \cap \mathbb{Q})^d$ whether $\mathcal{P}_1$ can *ensure* $\mathbf{q}$ *from* s_{init}, i.e., whether there exists a strategy σ_1 of $\mathcal{P}_1$ such that for all strategies σ_2 of $\mathcal{P}_2$, we have $\mathbb{P}_{s_{\mathsf{init}}}^{\sigma_1,\sigma_2}(\mathsf{Reach}(T_j)) \geq q_j$ for all $j \in [\![1, d]\!]$.

It is not known whether RRR strategies of $\mathcal{P}_1$ suffice to provide a positive answer whenever possible in general. However, finite-memory strategies suffice to approximate any vector for which the problem has a positive answer. More precisely, if $\mathcal{P}_1$ can ensure $\mathbf{q} = (q_j)_{1 \leq j \leq d}$ from $s_{\mathsf{init}} \in S$, then for all $\varepsilon > 0$, $\mathcal{P}_1$ has an DRD strategy such that for all strategies σ_2 of $\mathcal{P}_2$ and all $j \in [\![1, d]\!]$, it holds that $\mathbb{P}_{s_{\mathsf{init}}}^{\sigma_1,\sigma_2}(\mathsf{Reach}(T_j)) \geq q_j - \varepsilon$ [CFK+13a, ACK+20].

11.6 RDR strategies are weaker than DRD ones in infinite arenas

By Theorem 10.5, any RRR strategy in a finite arena admits an outcome-equivalent RDR strategy, i.e., randomisation in outputs can be removed without reducing expressiveness. The construction presented in the proof of Theorem 10.5 yields a Mealy machine the size of which depends on both the size of the arena and the size of the action space, and therefore does not work in arenas with infinitely many states or infinitely many actions. We provide two examples on deterministic MDPs: one with finitely many states and infinitely many actions, and another with infinitely many states but only two actions.

For the first example, we consider a one-state MDP, which can be seen extension of $\mathcal{M}_{a,b}$ with additional actions. Any memoryless strategy that randomises over a given set of actions of the MDP require a RDR Mealy machine with as many states as there are actions, due to the deterministic outputs. It follows that a memoryless strategy that randomises over infinitely many actions does not admit an equivalent RDR strategy.

Lemma 11.10. *Let $\mathcal{M} = (\{s\}, \mathbb{N}, \delta)$ be a deterministic MDP in which all actions are enabled in all states. There exists a (memoryless) DRD strategy in $\mathcal{M}$ such that there is no outcome-equivalent RDR strategy.*

Proof. Let σ_1 be the memoryless strategy of $\mathcal{M}$ defined by $\sigma_1(s)(\ell) = \frac{1}{2^{\ell+1}}$. We claim that no RDR strategy in $\mathcal{M}$ is outcome-equivalent to σ_1. Let $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be an RDR Mealy machine and let τ_1 be the strategy induced by $\mathfrak{M}$. By definition, for all $\ell \in \mathbb{N}$, we have

$$\tau_1(s)(\ell) = \sum_{\substack{m \in M \\ \mathsf{nxt}_{\mathfrak{M}}(m,s)=\ell}} \mu_{\mathsf{init}}(m).$$

Because M is finite, we conclude that there are infinitely many $\ell \in \mathbb{N}$ such that $\tau_1(s)(\ell) = 0$. It follows that τ_1 cannot be outcome-equivalent to σ_1. □

Lemma 11.11. *Let $\mathcal{M} = (\mathbb{N}, \{a, b\}, \delta)$ be a deterministic MDP over $\mathbb{N}$ in which all actions are enabled in s. There exists a (memoryless) DRD strategy in $\mathcal{M}$ such that there is no outcome-equivalent RDR strategy.*

Proof. Let σ_1 be the memoryless strategy of $\mathcal{M}$ defined by $\sigma_1(\ell)(a) = \frac{1}{2^{\ell+1}}$ for all $\ell \in \mathbb{N}$. We claim that no RDR strategy in $\mathcal{M}$ is outcome-equivalent to σ_1. Let $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be an RDR Mealy machine and let τ_1 be the strategy induced by $\mathfrak{M}$. By definition, for all $\ell \in \mathbb{N}$, we have

$$\tau_1(\ell)(a) = \sum_{\substack{m \in M \\ \mathsf{nxt}_{\mathfrak{M}}(m,\ell)=a}} \mu_{\mathsf{init}}(m).$$

Because M is finite, it follows from the above equation that the set $\{\tau_1(\ell) \mid \ell \in \mathbb{N}\}$ is also finite. By definition of σ_1, the set $\{\sigma_1(\ell) \mid \ell \in \mathbb{N}\}$ is infinite. It follows that there exists some $\ell \in \mathbb{N}$ such that $\sigma_1(\ell) \neq \tau_1(\ell)$. This shows that σ_1 and τ_1 are not outcome-equivalent. □

11.7 RDD and DRR strategies are incomparable with imperfect recall

Theorem 10.2 and Theorem 10.4 respectively state that, if perfect recall holds, any RDD strategy has an outcome-equivalent DRD counterpart and any RRR strategy has an outcome-equivalent DRR counterpart. We illustrate that without perfect recall, neither of these results hold. Our example is the same as Example 9.1, which we have used to show that behavioural strategies are less expressive than mixed strategies without perfect recall. We consider the POMDP $\mathfrak{P}_{a,b}$ built on $\mathcal{M}_{a,b}$ such that s, a and b are assigned the same observation o. We show below that the RDD strategy that uniformly mixes the two pure constant strategies of $\mathcal{M}_{a,b}$ has no observation-based DRR outcome-equivalent counterpart in $\mathfrak{P}_{a,b}$.

Lemma 11.12. *There exists an RDD observation-based Mealy machine in $\mathfrak{P}_{a,b}$ such that there is no outcome-equivalent DRR observation-based strategy.*

Proof. Let σ_1 be a (behavioural) history-based strategy (i.e., a strategy in $\mathcal{M}_{a,b}$) that is outcome-equivalent to the RDD strategy of $\mathfrak{P}_{a,b}$ obtained by uniformly mixing the constant pure strategies a and b. Let $\mathfrak{M} = (M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ be an observation-based DRR strategy of $\mathfrak{P}_{a,b}$ and let τ_1 be the history-based strategy it induces. We assume towards a contradiction that τ_1 and σ_1 are outcome-equivalent.

We have $\mathsf{nxt}_{\mathfrak{M}}(m_{\mathsf{init}}, o)(a) = \tau_1(s)(a) = \sigma_1(s)(a) = \frac{1}{2}$. It follows that the distributions μ_{sa} and μ_{sb} over M after sa and sb have respectively occurred are, by definition, for all $m \in M$,

$$\mu_{sa}(m) = \frac{\mathsf{up}_{\mathfrak{M}}(m_{\mathsf{init}}, o, o)(m) \cdot \frac{1}{2}}{\frac{1}{2}} = \mu_{sb}(m).$$

We conclude that $\tau_1(sas) = \tau_1(sbs)$. However, the outcome-equivalence of σ_1 and τ_1 implies that $\tau_1(sas)(a) = \sigma_1(sas)(a) = 1$ and $\tau_1(sbs)(b) = \sigma_1(sbs)(b) = 1$, which constitutes a contradiction. □

Part IV:

The structure of payoff sets in multi-objective Markov decision processes

Chapter 12

Introduction

In this part, we present the results described in Chapter 3.3, based on joint work with Mickaël Randour [MR25]. We study Markov decision processes with multi-dimensional payoff functions, which we call *multi-objective Markov decision processes*. On the one hand, we study the structure of sets of expected payoff vectors in countable multi-objective MDPs and the impact of their structure on randomisation requirements in this framework. On the other hand, we study finite multi-objective MDPs with continuous payoff functions and identify a class of continuous payoffs for which sets of expected payoff vectors are closed.

We refer the reader to Chapter 3.3 for an extended presentation of the context. We divide this part into three chapters. We summarise their contents below, and comment on related work at the end of this chapter.

Expected payoffs in MDPs. Chapter 13 introduces our notation for multi-objective MDPs. We also present general results of universally unambiguously integrable payoffs, some of which are particularly useful in the following chapter, Chapter 14.

We show that, for a subclass of universally unambiguously integrable payoffs, their expectation from a state under a mixed strategy is the integral with respect to the mixed strategy of the expectation of the payoff under all pure strategies (Lemma 13.4). We exploit this result to prove several properties of universally unambiguously integrable payoffs.

First, we use the above result to obtain a simple proof of the convexity of sets of expected payoff vectors from a state (Theorem 13.7). The key to our argument is to observe that the expected payoff of a convex combination of pure strategies is the convex combination of the expectations of the individual strategies (Lemma 13.6).

Second, we prove a characterisation of universally integrable payoffs: a one-dimensional payoff is universally integrable if and only if, for all initial states, its set of expected payoffs from the state is bounded (Lemma 13.8). We use this result to prove a technical result on universally unambiguously integrable payoffs (Lemma 13.9) with which we can somewhat broaden the application range of Lemma 13.4.

Payoff sets in multi-objective MDPs. Chapter 14 presents our general results on the structure of expected payoff sets in multi-objective MDPs. We focus on the relationship between sets of expected payoffs of pure strategies and sets of expected payoffs of general strategies.

For universally integrable payoffs, we show that all expected payoff vectors are convex combinations of expected payoffs of pure strategies (Theorem 14.4). Our reasoning relies on lexicographic optimisation: a key observation is that, for universally integrable payoffs, randomisation does not provide any additional power for lexicographic optimisation (Theorem 14.1). It follows that finite-support mixed strategies suffice to obtain any expected payoff vector.

We show that neither of the above properties extend to universally unambiguously integrable payoffs (Examples 14.1 and 14.4). Instead, we show that, for such payoffs, convex combinations of pure payoffs can be used to approximate any expected payoff vector (Theorem 14.7) in the sense of the topology of $\bar{\mathbb{R}}^d$. In other words, finite-support mixed strategies suffice to approximate any expected payoff vector.

We close the chapter by providing bounds on the support size of finite-support mixed strategies. We build on Carathéodory's theorem for convex hulls (Theorem 2.1) to show that the expected payoff of any finite-support mixed strategy can be obtained exactly by mixing no more than $d + 1$ strategies in a d-dimensional setting (Theorem 14.8). Together with our previous results, this implies that it suffices to mix no more than $d + 1$ pure strategies to match or approximate the expectation of any strategy.

Continuous payoffs in finite multi-objective MDPs. Chapter 15 focuses on finite multi-objective MDPs with continuous payoffs. We focus on universally square integrable payoffs, i.e., payoffs whose square is universally integrable. We show that for multi-dimensional universally square integrable payoffs, the set of expected payoffs from each state is closed (Theorem 15.8).

We mainly follow a topological approach to establishing this result. First, we introduce a topology on the set of behavioural strategies (Chapter 15.1). We then show that, for one-dimensional square integrable payoffs, the function from the space of strategies to $\mathbb{R}$ that maps a strategy to its expectation is continuous: we first show this for real-valued payoffs (Theorem 15.6) then extend it to universally square integrable payoffs (Theorem 15.7).

We then show that for continuous payoffs that are not universally integrable, the function mapping a strategy to its expected payoff need not be continuous (Example 15.1) and expected payoff sets need not be closed (Example 15.2).

Finally, we show that universally integrable shortest-path costs are universally square integrable in finite MDPs (Lemma 15.11). This shows that our results for continuous universally square integrable payoffs applies to universally integrable shortest-path costs defined with a positive weight function.

Related work. We provide a few references, complementing those cited in Chapter 3.3. In our proof of Theorem 14.4, we invoke the separating hyperplane theorem (Theorem 2.3). This theorem also plays a role in approximation schemes of the set of achievable vectors: see, e.g., [FKP12, QK21]; a unifying approach is presented in [Qua23].

Specifications with multiple objectives have also been considered in the context of two-player games on finite turn-based deterministic arenas (e.g., [FH13, CRR14]) and two-player games on finite turn-based arenas (e.g., [CFK$^+$13a, ACK$^+$20]). Closely related to multi-objective specifications are approaches that provide guarantees simultaneously in the worst case and the expected case [BFRR17].

Regarding randomisation in strategies, [CDGH15] studies when randomisation is helpful in strategies or in transitions of games. In particular, the authors show that if there exists an optimal strategy to maximise the probability of an event in a finite MDP, then there exists a pure optimal strategy. This property is generalised by our result on lexicographic MDPs.

Chapter 13

Expected payoffs in Markov decision processes

In this chapter, we introduce our notation for multi-objective Markov decision processes and establish technical results regarding expected payoffs in Markov decision processes. Notation is introduced in Section 13.1. In Section 13.2, we show that expected payoffs with respect to mixed strategies can be written as the integral with respect to the mixed strategy of expected payoffs with respect to pure strategies. This result generalises Lemma 2.17, which states the same property in the special case of objective indicators. This result plays a major role in our proofs of some of the results of Chapter 14. We use the generalisation of Lemma 2.17 to show that convex combinations of expected payoffs are yet again expected payoffs in Section 13.3. Finally, in Section 13.4, we provide a characterisation of universally integrable payoffs and a technical property of universally unambiguously integrable payoffs.

We fix a countable MDP $\mathcal{M} = (S, A, \delta)$ for this entire chapter.

Contents

13.1 Terminology and notation

This section introduces terminology and notation for MDPs with multiple payoff functions. Let $d \in \mathbb{N}_{>0}$. We summarise d payoffs $f_1, \dots, f_d \colon \mathsf{Plays}(\mathcal{M}) \to \bar{\mathbb{R}}$ as a multi-dimensional payoff $\bar{f} \colon \mathsf{Plays}(\mathcal{M}) \to \bar{\mathbb{R}}^d$, and write $\bar{f} = (f_j)_{j \in [\![1,d]\!]}$.

Let $\bar{f} = (f_j)_{j \in [\![1,d]\!]}$ and let $s \in S$. We say that $\bar{f}$ is universally (resp. unambiguously) integrable whenever f_j is universally (resp. unambiguously) integrable for all $j \in [\![1, d]\!]$. We now assume that $\bar{f}$ is universally unambiguously integrable.

We use the following notation for the set of expected payoff vectors from an initial state.

Definition 13.1. Let $\Sigma \subseteq \Sigma(\mathcal{M})$ be a set of strategies of $\mathcal{M}$ and $s \in S$. We let $\mathsf{Pay}_s^{\Sigma}(\bar{f}) = \{\mathbb{E}_s^{\sigma}(\bar{f}) \mid \sigma \in \Sigma\}$ denote the *set of (expected) payoffs* of the strategies in Σ from s. We let $\mathsf{Pay}_s(\bar{f})$ and $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ be shorthand for $\mathsf{Pay}_s^{\Sigma(\mathcal{M})}(\bar{f})$ and $\mathsf{Pay}_s^{\Sigma_{\mathsf{pure}}(\mathcal{M})}(\bar{f})$ respectively.

We refer to elements of $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ as *pure expected payoffs*. A set of expected payoffs need not have a maximum for the component-wise order, e.g., there can be several Pareto-optimal payoffs.

In multi-objective optimisation, the goal is to ensure a given threshold on each dimension. This is formalised by the notion of achievable vectors.

Definition 13.2. A vector $\mathbf{q} \in \bar{\mathbb{R}}^d$ is *achievable* (from s) if there exists a strategy σ such that $\mathbf{q} \leq \mathbb{E}_s^{\sigma}(\bar{f})$. We say that σ *witnesses* that $\mathbf{q}$ is achievable.

For any class of strategies $\Sigma \subseteq \Sigma(\mathcal{M})$, we let $\mathsf{Ach}_s^{\Sigma}(\bar{f}) = \mathsf{down}(\mathsf{Pay}_s^{\Sigma}(\bar{f}))$ denote the set of vectors for which there exists a strategy of Σ witnesses that they are achievable. We define $\mathsf{Ach}_s(\bar{f})$ and $\mathsf{Ach}_s^{\mathsf{pure}}(\bar{f})$ as above.

As a technical tool in our analysis of expected payoff sets, we also consider the lexicographic optimisation of multiple objectives. We can define ensuring a threshold similarly than in the one-dimensional context (see Section 2.6.1), with the order over $\bar{\mathbb{R}}$ replaced with the lexicographic order. We also define an analogue of optimal strategies from the one-dimensional setting.

Definition 13.3. A strategy σ is *lexicographically optimal* if $\mathbb{E}_s^\sigma(\bar{f})$ is the lexicographic maximum of $\mathsf{Pay}_s(\bar{f})$.

13.2 Payoffs under mixed strategies

Let $f\colon \mathsf{Plays}(\mathcal{M}) \to \bar{\mathbb{R}}$ be a one-dimensional universally unambiguously integrable payoff. We generalise Lemma 2.17 from probabilities of objectives to expectations of payoffs. Lemma 2.17 states for all $s \in S$, the probability of an objective Ω under a mixed strategy μ from s is $\int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{P}_s^\tau(\Omega)\mathrm{d}\mu(\tau)$. We extend this result from indicators to general payoffs by considering payoffs of increasing complexity, analogously to the construction of the Lebesgue integral. In the following statement, we impose restrictions on f that ensure that we deal with a well-defined integral.

Lemma 13.4. *Let μ be a mixed strategy, $s \in S$ and f be a universally unambiguously integrable payoff. If $\inf_\tau \mathbb{E}_s^\tau(f) \geq 0$ or $\sup_\tau \mathbb{E}_s^\tau(f) \leq 0$ or f is $\mathbb{P}_s^\mu$-integrable, then the mapping $\Sigma_{\mathsf{pure}}(\mathcal{M}) \to \bar{\mathbb{R}}\colon \tau \mapsto \mathbb{E}_s^\tau(f)$ is measurable and*

$$\mathbb{E}_s^\mu(f) = \int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}_s^\tau(f)\mathrm{d}\mu(\tau).$$

Proof. Fix a state $s \in S$ and let f be a universally unambiguously integrable payoff. Throughout this proof, we use τ to (implicitly) denote pure strategies. We show the result for payoffs of increasing complexity. First, we prove it for indicators of objectives. Second, we show that it also holds for non-negative simple functions (i.e., linear combinations of indicators) by linearity of the integral. Third, we deal with non-negative payoffs with the monotone convergence theorem. Fourth, we consider $\mathbb{P}_s^\mu$-integrable payoffs. Finally, we close the proof by considering universally unambiguously integrable payoffs such that $\inf_\tau \mathbb{E}_s^\tau(f) \geq 0$ or $\sup_\tau \mathbb{E}_s^\tau(f) \leq 0$.

If f is the indicator of an objective, the result follows from Lemma 2.16, which guarantees that $\Sigma_{\mathsf{pure}}(\mathcal{M}) \to \mathbb{R}\colon \tau \mapsto \mathbb{P}_s^\tau(\Omega)$ is measurable, and Lemma 2.17, which yields the integral.

For the second step of the argument, we assume that f is a non-negative simple function. Let $\alpha_1, \dots, \alpha_n \geq 0$ and $\Omega_1, \dots, \Omega_n \subseteq \mathsf{Plays}(\mathcal{M})$ be objectives such that $f = \sum_{j=1}^{n} \alpha_j \mathbb{1}_{\Omega_j}$. The function $\Sigma_{\mathsf{pure}}(\mathcal{M}) \to \mathbb{R} \colon \tau \mapsto \mathbb{E}_s^{\tau}(f)$ is measurable: it is a non-negative linear combination of measurable functions by the above. It follows from the result for indicators applied for all $j \in [\![1, n]\!]$ and the linearity of the Lebesgue integral that

$$\begin{aligned}\mathbb{E}_s^{\mu}(f) &= \sum_{j=1}^{n} \alpha_j \mathbb{P}_s^{\mu}(\Omega_j) \\ &= \sum_{j=1}^{n} \alpha_j \int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{P}_s^{\tau}(\Omega_j) \mathrm{d}\mu(\tau) \\ &= \int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}_s^{\tau}(f) \mathrm{d}\mu(\tau).\end{aligned}$$

Third, we assume that f is a non-negative measurable function. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable simple functions increasing to f (i.e., for all plays $\pi \in \mathsf{Plays}(\mathcal{M})$, $f_n(\pi) \leq f_{n+1}(\pi)$ and $\lim_{n\to\infty} f_n(\pi) = f(\pi)$). By the monotone convergence theorem and the previous point on simple functions, we have

$$\mathbb{E}_s^{\mu}(f) = \lim_{n\to\infty} \mathbb{E}_s^{\mu}(f_n) = \lim_{n\to\infty} \int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}_s^{\tau}(f_n) \mathrm{d}\mu(\tau). \tag{13.1}$$

For all pure strategies τ, by the monotone convergence theorem, $\lim_{n\to\infty} \mathbb{E}_s^{\tau}(f_n) = \mathbb{E}_s^{\tau}(f)$. Therefore, the sequence of functions $(\tau \mapsto \mathbb{E}_s^{\tau}(f_n))_{n\in\mathbb{N}}$ over $\Sigma_{\mathsf{pure}}(\mathcal{M})$ increases (i.e., is non-decreasing and converges pointwise) to $\tau \mapsto \mathbb{E}_s^{\tau}(f)$, implying that this function is measurable. The monotone convergence theorem allows us to exchange the limit and integral in the rightmost term of Equation (13.1), and implies that:

$$\mathbb{E}_s^{\mu}(f) = \int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} \lim_{n\to\infty} \mathbb{E}_s^{\tau}(f_n) \mathrm{d}\mu(\tau) = \int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}_s^{\tau}(f) \mathrm{d}\mu(\tau).$$

We introduce some notation for the two last cases. We let $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$ denote the non-negative and non-positive parts of f respectively; we have $f = f^+ - f^-$. From the above, we obtain that for all universally unambiguously integrable payoffs, the function $\tau \mapsto \mathbb{E}_s^{\tau}(f)$ over $\Sigma_{\mathsf{pure}}(\mathcal{M})$ is measurable; it is the difference of the measurable non-negative functions $\tau \mapsto \mathbb{E}_s^{\tau}(f^+)$ and $\tau \mapsto \mathbb{E}_s^{\tau}(f^-)$.

For the second-to-last case, we assume that f is $\mathbb{P}^{\mu}_{s}$-integrable. We prove that the mappings $\tau \mapsto \mathbb{E}^{\tau}_{s}(f^{+})$ and $\tau \mapsto \mathbb{E}^{\tau}_{s}(f^{-})$ are μ-integrable. We proceed by bounding these functions by a μ-integrable function. For all $\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})$, we have $\mathbb{E}^{\tau}_{s}(f^{+}), \mathbb{E}^{\tau}_{s}(f^{-}) \leq \mathbb{E}^{\tau}_{s}(|f|)$. By the above, we have $\int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}^{\tau}_{s}(|f|)\mathrm{d}\mu(\tau) = \mathbb{E}^{\mu}_{s}(|f|)$, which is a real number since f is $\mathbb{P}^{\mu}_{s}$-integrable. We have shown that $\tau \mapsto \mathbb{E}^{\tau}_{s}(|f|)$ is μ-integrable, which implies that $\tau \mapsto \mathbb{E}^{\tau}_{s}(f^{+})$ and $\tau \mapsto \mathbb{E}^{\tau}_{s}(f^{-})$ also are. It follows that $\tau \mapsto \mathbb{E}^{\tau}_{s}(f)$ is μ-integrable.

By definition, we have $\mathbb{E}^{\mu}_{s}(f) = \mathbb{E}^{\mu}_{s}(f^{+}) - \mathbb{E}^{\mu}_{s}(f^{-})$ and $\mathbb{E}^{\tau}_{s}(f) = \mathbb{E}^{\tau}_{s}(f^{+}) - \mathbb{E}^{\tau}_{s}(f^{-})$ for all strategies $\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})$. Combining this with the linearity of the Lebesgue integral and the result for non-negative payoffs yields the following sequence of equalities:

$$\begin{aligned}
\mathbb{E}^{\mu}_{s}(f) &= \mathbb{E}^{\mu}_{s}(f^{+}) - \mathbb{E}^{\mu}_{s}(f^{-}) \\
&= \int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}^{\tau}_{s}(f^{+})\mathrm{d}\mu(\tau) - \int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}^{\tau}_{s}(f^{-})\mathrm{d}\mu(\tau) \\
&= \int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}^{\tau}_{s}(f^{+}) - \mathbb{E}^{\tau}_{s}(f^{-})\mathrm{d}\mu(\tau) \\
&= \int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}^{\tau}_{s}(f)\mathrm{d}\mu(\tau).
\end{aligned}$$

To deal with the last case, we assume that $\inf_{\tau} \mathbb{E}^{\tau}_{s}(f) \geq 0$. The analogous case $\sup_{\tau} \mathbb{E}^{\tau}_{s}(f) \leq 0$ can be recovered from the case $\inf_{\tau} \mathbb{E}^{\tau}_{s}(f) \geq 0$ by considering $-f$ as the payoff function. We assume that f is not $\mathbb{P}^{\mu}_{s}$-integrable, as this case has been examined above, i.e., that $\mathbb{E}^{\mu}_{s}(f) = +\infty$. The integral $\int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}^{\tau}_{s}(f)\mathrm{d}\mu(\tau)$ is formally well-defined by the assumption that $\inf_{\tau} \mathbb{E}^{\tau}_{s}(f) \geq 0$. To end the argument, we must show that this integral is $+\infty$. Assume towards a contradiction that this is not the case. This implies that $\tau \mapsto \mathbb{E}^{\tau}_{s}(f)$ is μ-integrable. From the result for non-negative payoffs, we obtain that $\int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}^{\tau}_{s}(f^{+})\mathrm{d}\mu(\tau) = \mathbb{E}^{\mu}_{s}(f^{+}) = +\infty$ and $\int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}^{\tau}_{s}(f^{-})\mathrm{d}\mu(\tau) = \mathbb{E}^{\mu}_{s}(f^{-}) \in \mathbb{R}$. By linearity of the Lebesgue integral (for μ-integrable payoffs), we obtain that

$$\begin{aligned}\mathbb{E}_s^{\mu}(f^+) &= \int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}_s^{\tau}(f^+)\mathrm{d}\mu(\tau)\\ &= \int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} (\mathbb{E}_s^{\tau}(f)+\mathbb{E}_s^{\tau}(f^-))\mathrm{d}\mu(\tau)\\ &= \int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}_s^{\tau}(f)\mathrm{d}\mu(\tau) + \int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}_s^{\tau}(f^-)\mathrm{d}\mu(\tau).\end{aligned}$$

This is a contradiction: on the one hand, we have $\mathbb{E}_s^{\mu}(f^+) = +\infty$ and, on the other hand, the sum in the last term is a real number. This ends the argument for the case $\inf_\tau \mathbb{E}_s^{\tau}(f) \geq 0$. □

We highlight two major consequences of Lemma 13.4 when combined with Kuhn's theorem. On the one hand, for all randomised (i.e., mixed or behavioural) strategies whose expected payoff is real, there exists a pure strategy with a greater expected payoff. On the other hand, if there exists a randomised strategy with an infinite expected payoff, then there are pure strategies with arbitrarily large expected payoffs in absolute value. We note that even if a randomised strategy has an infinite expectation, there need not exist a pure strategy with infinite expectation. This is analogous to the fact that real-valued random variables can have an infinite expectation.

Remark 13.5 (Partially observation MDPs). Lemma 13.4 holds for all mixed strategies. In particular, it applies to observation-based mixed strategies in POMDPs. In the sequel, we only apply Lemma 13.4 in the perfect-information setting. However, we remark that all of our arguments involving this property and Kuhn's theorem extend to *countable perfect-recall POMDPs*. In particular, the results of Chapter 14 extend to this more general setting. ◁

13.3 Convexity of expected payoff sets

Let $\bar{f} = (f_j)_{j\in[\![1,d]\!]}$ be a universally unambiguously integrable payoff of $\mathcal{M}$. The goal of this section is to show that, for all $s \in S$, convex combinations of elements of $\mathsf{Pay}_s(\bar{f})$ are in $\mathsf{Pay}_s(\bar{f})$ and that $\mathsf{Pay}_s(\bar{f}) \cap \mathbb{R}^d$ and $\mathsf{Ach}_s(\bar{f}) \cap \mathbb{R}^d$ are convex. We provide an argument that relies on mixed strategies; a proof of this result using behavioural strategies can be found in [Qua23].

The first step in our argument is to prove that the expected payoff of a convex combination of mixed strategies is the convex combination of the expected payoffs of these mixed strategies. We obtain this property as a consequence of Lemma 13.4.

Lemma 13.6. *Let $s \in S$, $\mu_1, \dots, \mu_n$ be mixed strategies and $\alpha_1, \dots, \alpha_n \in]0, 1[$ be convex combination coefficients. Let $\mu = \sum_{m=1}^{n} \alpha_m \mu_m$. For all universally unambiguously integrable payoffs f, we have $\mathbb{E}_s^{\mu}(f) = \sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\mu_m}(f)$.*

Proof. To obtain the result for non-negative payoffs, by Lemma 13.4, it suffices to establish that, for all measurable non-negative functions $\mathcal{F} \colon \Sigma_{\mathsf{pure}}(\mathcal{M}) \to \bar{\mathbb{R}}$,

$$\int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} \mathcal{F}(\tau) \mathrm{d}\mu(\tau) = \sum_{m=1}^{n} \alpha_m \int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} \mathcal{F}(\tau) \mathrm{d}\mu_m(\tau).$$

For indicators of measurable subsets of $\Sigma_{\mathsf{pure}}(\mathcal{M})$, this follows from the definition of μ and the linearity of the Lebesgue integral. It generalises to non-negative simple functions over $\Sigma_{\mathsf{pure}}(\mathcal{M})$ by linearity, and to all non-negative measurable functions by using the monotone convergence theorem with sequences of measurable non-negative simple functions.

The result for non-negative payoffs extends to universally unambiguously integrable payoffs by definition of unambiguous integrals with respect to the distribution induced by a strategy. □

Lemma 13.6 and Kuhn's theorem imply that, for all universally unambiguously integrable payoffs, convex combinations (with non-zero coefficients) of expected payoffs also are expected payoffs. We obtain that the set of vectors of reals in sets of expected payoffs and achievable vectors both are convex.

Theorem 13.7. *Assume that $\bar{f}$ is universally unambiguously integrable. Let $s \in S$. For all non-zero convex combination coefficients $\alpha_1, \dots, \alpha_n \in]0, 1]$ and expected payoff vectors $\mathbf{q}_1, \dots, \mathbf{q}_n \in \mathsf{Pay}_s(\bar{f})$, we have $\sum_{m=1}^{n} \alpha_m \mathbf{q}_m \in \mathsf{Pay}_s(\bar{f})$. In particular, $\mathsf{Pay}_s(\bar{f}) \cap \mathbb{R}^d$ and $\mathsf{Ach}_s(\bar{f}) \cap \mathbb{R}^d$ are convex sets.*

Proof. The claim regarding convex combinations of expected payoffs follow from Lemma 13.6 and Kuhn's theorem. It directly implies that $\mathsf{Pay}_s(\bar{f}) \cap \mathbb{R}^d$ is convex.

It remains to show that $\mathsf{Ach}_s(\bar{f}) \cap \mathbb{R}^d$ is convex. Let $\mathbf{q}$, $\mathbf{p} \in \mathsf{Ach}_s(\bar{f}) \cap \mathbb{R}^d$ and $\alpha \in]0,1[$. We must show that $\alpha\mathbf{q} + (1-\alpha)\mathbf{p} \in \mathsf{Ach}_s(\bar{f})$. There exist (by Kuhn's theorem) mixed strategies $\mu_{\mathbf{q}}$ and $\mu_{\mathbf{p}}$ such that $\mathbb{E}_s^{\mu_{\mathbf{q}}}(\bar{f}) \geq \mathbf{q}$ and $\mathbb{E}_s^{\mu_{\mathbf{p}}}(\bar{f}) \geq \mathbf{p}$. Let $\mu = \alpha\mu_{\mathbf{q}} + (1-\alpha)\mu_{\mathbf{p}}$; it is easy to see that

$$\alpha\mathbf{q} + (1-\alpha)\mathbf{p} \leq \mathbb{E}_s^{\mu}(\bar{f}) = \alpha\mathbb{E}_s^{\mu_{\mathbf{q}}} + (1-\alpha)\mathbb{E}_s^{\mu_{\mathbf{p}}}(\bar{f}),$$

where the latter equality follows from Lemma 13.6. We have shown that $\alpha\mathbf{q} + (1-\alpha)\mathbf{p} \in \mathsf{Ach}_{\bar{f}}(s)$. □

13.4 Unambiguously integrable payoffs

We focus on one-dimensional payoffs in this section. We apply Lemma 13.4 to obtain a characterisation of universally integrable payoffs. We use the property underlying this characterisation to show that, for all universally unambiguously integrable payoffs and all initial states, there exists a real lower or upper bound on the possible expectations of the payoff from the initial state.

We characterise universally integrable payoffs of $\mathcal{M}$ as follows. Let $\Sigma \in \{\Sigma(\mathcal{M}), \Sigma_{\mathsf{pure}}(\mathcal{M})\}$. A payoff f is universally integrable if and only if for all $s \in S$, $\sup_{\sigma\in\Sigma} \mathbb{E}_s^{\sigma}(|f|)$ is real. The non-trivial part of the proof is showing that the definition of universally integrable and the property when the supremum ranges over pure strategies both imply the property with the supremum ranging over all strategies. We show the contrapositive of both implications. We assume that for some $s \in S$, $\sup_{\sigma\in\Sigma} \mathbb{E}_s^{\sigma}(|f|) = +\infty$. Lemma 13.4 then implies that there are pure strategies τ with arbitrarily large $\mathbb{E}_s^{\tau}(|f|)$, which implies the validity of one of the implications. For the other, we mix countably many of these pure strategies to construct a mixed strategy μ such that $\mathbb{E}_s^{\mu}(|f|) = +\infty$.

Lemma 13.8. *Let f be a payoff. Let $s \in S$. The following assertions are equivalent.*

(i) f is $\mathbb{P}_s^{\sigma}$-integrable for all $\sigma \in \Sigma(\mathcal{M})$.

(ii) We have $\sup\{\mathbb{E}^\sigma_s(|f|) \mid \sigma \in \Sigma(\mathcal{M})\} \in \mathbb{R}$.

(iii) We have $\sup\{\mathbb{E}^\sigma_s(|f|) \mid \sigma \in \Sigma_{\mathsf{pure}}(\mathcal{M})\} \in \mathbb{R}$.

In particular, f is universally integrable if and only if (ii) (resp. (iii)) holds for all $s \in S$.

Proof. Item (ii) directly implies the other two items. We now show that the other two items imply (ii) via the contrapositive of these implications.

Assume that (ii) does not hold, i.e., that $\sup\{\mathbb{E}^\sigma_s(|f|) \mid \sigma \in \Sigma(\mathcal{M})\} = +\infty$. If there exists a pure strategy σ such that $\mathbb{E}^\sigma_s(|f|) = +\infty$, the negations of (i) and (iii) follow directly. In the remainder of the proof, we assume that this is not the case.

First, we show that (iii) does not hold. By Lemma 13.4 (and Kuhn's theorem), for all strategies $\sigma \in \Sigma(\mathcal{M})$, if $\mathbb{E}^\sigma_s(|f|) \in \mathbb{R}$, there exists a pure strategy τ such that $\mathbb{E}^\tau_s(|f|) \geq \mathbb{E}^\sigma_s(|f|)$ and, otherwise, if $\mathbb{E}^\sigma_s(|f|) = +\infty$, then for all $M \in \mathbb{R}$, there exists a pure strategy τ such that $\mathbb{E}^\tau_s(|f|) \geq M$. In particular, (iii) does not hold.

We now construct a strategy σ such that $\mathbb{E}^\sigma_s(|f|) = +\infty$ from the pure strategies above. For all $r \in \mathbb{N}$, let τ_r be a pure strategy such that $\mathbb{E}^{\tau_r}_s(|f|) \geq 2^r$. Let μ be the mixed strategy that randomises over the set $\{\tau_r \mid r \in \mathbb{N}\}$ and selects strategy τ_r with probability $\frac{1}{2^{r+1}}$. Kuhn's theorem implies that there exists a behavioural strategy σ that is outcome-equivalent to μ. We obtain that $\mathbb{E}^\sigma_s(|f|) = +\infty$. This ends the proof that (i) does not hold. □

Let f be a universally unambiguously integrable payoff. We now use Lemma 13.8 to prove a useful property of universally unambiguously integrable payoffs. We prove that, for all states $s \in S$, there either exists a real lower or upper bound on the expectation of strategies from s. We establish this result by showing that if no upper bound and no lower bound exist, then we can construct a mixed strategy whose expected payoff is ill-defined.

Lemma 13.9. *Let f be a universally unambiguously integrable payoff function. For all $s \in S$, we have* $\inf_{\sigma\in\Sigma(\mathcal{M})} \mathbb{E}^\sigma_s(f) \in \mathbb{R}$ *or* $\sup_{\sigma\in\Sigma(\mathcal{M})} \mathbb{E}^\sigma_s(f) \in \mathbb{R}$.

Proof. Let $s \in S$. Let $f^+ = \max(f, 0)$ and $f^- = \max(0, -f)$. Assume towards a contradiction that $\inf_{\sigma\in\Sigma(\mathcal{M})} \mathbb{E}^{\sigma}_{s}(f) \notin \mathbb{R}$ and $\sup_{\sigma\in\Sigma(\mathcal{M})} \mathbb{E}^{\sigma}_{s}(f) \notin \mathbb{R}$. Because $\Sigma(\mathcal{M})$ is non-empty, we have $\inf_{\sigma\in\Sigma(\mathcal{M})} \mathbb{E}^{\sigma}_{s}(f) = -\infty$ and $\sup_{\sigma\in\Sigma(\mathcal{M})} \mathbb{E}^{\sigma}_{s}(f) = +\infty$. We show that this implies that f is not universally unambiguously integrable, i.e., there exists a strategy $\sigma \in \Sigma(\mathcal{M})$ such that $\mathbb{E}^{\sigma}_{s}(f^+) = \mathbb{E}^{\sigma}_{s}(f^-) = +\infty$.

We observe that for all $\sigma \in \Sigma(\mathcal{M})$, we have $\mathbb{E}^{\sigma}_{s}(f) \leq \mathbb{E}^{\sigma}_{s}(f^+)$ and $\mathbb{E}^{\sigma}_{s}(-f) \leq \mathbb{E}^{\sigma}_{s}(f^-)$. It follows from Lemma 13.8 and Kuhn's theorem that there exists a mixed strategy μ_+ (resp. μ_-) such that f^+ (resp. f^-) is not $\mathbb{P}^{\mu_+}_{s}$-integrable (resp. $\mathbb{P}^{\mu_-}_{s}$-integrable). In particular, we obtain that $\mathbb{E}^{\mu_+}_{s}(f^+) = \mathbb{E}^{\mu_-}_{s}(f^-) = +\infty$. The mixed strategy $\mu = \frac{1}{2}\mu_+ + \frac{1}{2}\mu_-$ satisfies $\mathbb{E}^{\mu}_{s}(f^+) = \mathbb{E}^{\mu}_{s}(f^-) = +\infty$ by Lemma 13.6. This shows that f is not universally unambiguously integrable: f does not have an unambiguous $\mathbb{P}^{\mu}_{s}$-integral and Kuhn's theorem guarantees that μ is outcome-equivalent to some behavioural strategy. □

Lemma 13.4 cannot be applied to all universally unambiguously integrable payoffs: we impose some constraints on the expected payoffs to ensure that the integral in the statement is well-defined. We can use Lemma 13.9 to circumvent this restriction: it implies that by adding or subtracting a constant to an unambiguously universally integrable payoff, we can obtain a payoff satisfying the assumptions of Lemma 13.4.

Chapter 14

Payoff sets in multi-objective Markov decision processes

We present in detail the results presented in Chapter 3.3 related to the structure of payoff sets in multi-objective Markov decision processes and its impact on randomisation requirements in this setting.

We first present an example with two discounted payoff functions that highlights the potential complexity of payoff sets. This example illustrates that expected payoff sets need not be convex polytopes, and that even when finite memory suffices to obtain all expected payoffs, there need not exist a uniform bound on the necessary amount of memory.

We then discuss lexicographic optimisation in multi-objective MDPs. We prove that pure strategies suffice for universally integrable payoffs, but not in general. We use this result to show that *finite-support mixed strategies* suffice to (exactly) obtain any expected payoff vector when dealing with universally integrable payoffs. We show that this result does not generalise to universally unambiguously integrable payoffs, and prove that, for such payoffs, finite-support mixed strategies can be used to approximate any expected payoff vector. We close the section by giving upper bounds on the necessary size for the supports of the finite-support mixed-strategies of the previous results.

We recall that randomised strategies are necessary to achieve vectors in multi-objective Markov decision processes, e.g., see Chapter 3.3 or Example 11.1. For this section, we fix an MDP $\mathcal{M} = (S, A, \delta)$ and a d-dimensional payoff function $\bar{f} = (f_j)_{j\in[\![1,d]\!]}$ for the whole chapter.

Contents

14.1 Expected payoff sets need not be simple

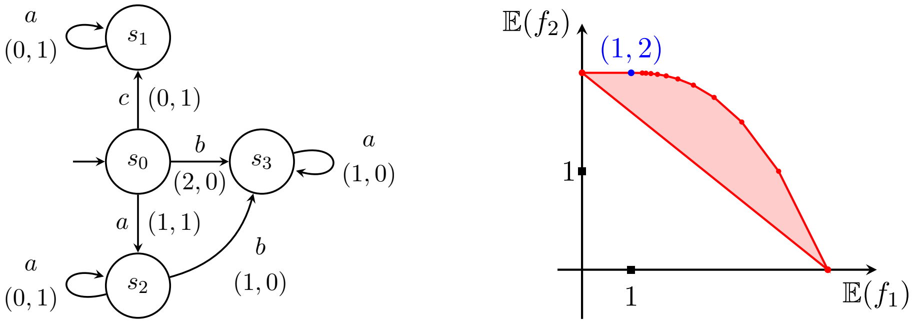
The goal of this section is to illustrate that expected payoff sets in multi-objective MDPs may be complex. We provide a two-dimensional example illustrating that sets of expected payoffs are not necessarily polytopes, even when the payoffs are universally integrable. In this section, we introduce our example and comment on several of its properties. To lighten the presentation, we defer the formal proofs of these properties to Appendix B.

We consider the MDP $\mathcal{M}$ depicted in Figure 14.1a and let w denote the two-dimensional weight function from the illustration. On this MDP, we consider the two-dimensional payoff $\bar{f} = (f_1, f_2)$ given by the discounted-sum payoffs $f_1 = \mathsf{DSum}^{3/4}_{w_1}$ and $f_2 = \mathsf{DSum}^{1/2}_{w_2}$. We note that MDPs with several discounted-sum payoffs with different discount factors have been studied in [CFW13].

Due to the absence of randomness in transitions, the expected payoff of any pure strategy from s_0 is the payoff of a play from s_0. Therefore, we obtain that

$$\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) = \{(0,2),(1,2)\} \cup \left\{\left(1+\frac{3^r}{4^{r-1}}, 2-\frac{1}{2^{r-1}}\right) \mid r \in \mathbb{N}\right\}.$$

On the one hand, the payoffs $(0,2)$ and $(1,2)$ are obtained by moving from s_0 to s_1 and s_2 respectively and looping in these states forever. On the other hand, for all $r \in \mathbb{N}$, the payoff $\left(1+\frac{3^r}{4^{r-1}}, 2-\frac{1}{2^{r-1}}\right)$ is obtained by spending r



(a) An MDP with deterministic transitions. Pairs next to actions represent two-dimensional weights.

(b) The set of expected payoffs for the MDP of Figure 14.1a for the payoff $f_1 = \mathsf{DSum}_{w_1}^{3/4}$ and $f_2 = \mathsf{DSum}_{w_2}^{1/2}$.

Figure 14.1: An MDP with a two-dimensional discounted-sum payoff $\bar{f}$ such that $\mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f}))$ is infinite.

rounds in s_2 then moving to s_3; for $r = 0$, we move from s_0 to s_3 directly. We provide detailed computations in the proof of Lemma B.1.

We approximately illustrate $\mathsf{Pay}_{s_0}(\bar{f})$ in Figure 14.1b. This illustration is based on the equality $\mathsf{Pay}_{s_0}(\bar{f}) = \mathsf{conv}(\mathsf{Pay}_{s_0}^{\mathsf{pure}}(\bar{f}))$. This equality follows from Theorem 14.4, which states that if $\bar{f}$ is universally integrable, then this equality holds. Furthermore, $\mathsf{Pay}_{s_0}(\bar{f})$ and $\mathsf{Pay}_{s_0}^{\mathsf{pure}}(\bar{f})$ are both closed; this can be shown directly or seen as a special case of Theorem 15.8, which implies that expected payoff sets for multi-dimensional real-valued continuous payoffs are closed.

Since $\mathsf{Pay}_{s_0}(\bar{f}) = \mathsf{conv}(\mathsf{Pay}_{s_0}^{\mathsf{pure}}(\bar{f}))$, we conclude that all extreme points of $\mathsf{Pay}_{s_0}(\bar{f})$ can be obtained by using pure strategies. Indeed, any vector of $\mathsf{Pay}_{s_0}(\bar{f})$ that cannot be obtained by a pure strategy is a convex combination of the expected payoffs of pure strategies, and thus is not extreme. In fact, we can show that the set of extreme points of $\mathsf{Pay}_{s_0}(\bar{f})$ is exactly $\mathsf{Pay}_{s_0}^{\mathsf{pure}}(\bar{f})$ (Lemma B.7). In particular, $\mathsf{Pay}_{s_0}(\bar{f})$ is not a convex polytope. Furthermore, it can be shown that all pure expected payoffs except $(0,2)$ are Pareto-optimal (Lemma B.6). It follows that even the set $\mathsf{Ach}_{s_0}(\bar{f})$ of achievable vectors has a complex structure.

Finally, we comment on the memory required to obtain certain expected payoffs. All but three extreme points of $\mathsf{Pay}_{s_0}(\bar{f})$ are obtained by moving from s_0 to s_2 and looping there finitely many times before moving to s_3. In other words, these extreme points require pure strategies that count up to some arbitrarily large number. We show that we can only obtain these payoffs by using these specific pure strategies (Lemma B.8). Intuitively, the expected payoff of a randomised strategy that induces more than one play is a non-trivial convex combination of payoffs of several plays, and therefore not in $\mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f}))$. This implies that some expected payoffs need strategies with arbitrarily large albeit finite memory to be obtained in this instance.

14.2 Lexicographic Markov decision processes

We consider the lexicographic optimisation of multiple payoff functions in MDPs. We first prove that, if $\bar{f}$ is universally integrable, then for all initial states $s \in S$ and all strategies $\sigma \in \Sigma(\mathcal{M})$, there exists a pure strategy τ such that $\mathbb{E}^{\sigma}_{s}(\bar{f}) \leq_{\mathsf{lex}} \mathbb{E}^{\tau}_{s}(\bar{f})$. We then show that this is not necessarily the case without the assumption that $\bar{f}$ is universally integrable.

Assume that $\bar{f}$ is universally integrable. The crux of our proof is to show that the Lebesgue integral is compatible with the lexicographic order over $\bar{\mathbb{R}}^d$. Once this is shown, we assume towards a contradiction that there is no suitable pure strategy τ. By Lemma 13.4 and Kuhn's theorem, we can write $\mathbb{E}^{\sigma}_{s}(\bar{f})$ as an integral over pure expected payoffs. We then reach the contradiction that $\mathbb{E}^{\sigma}_{s}(\bar{f}) <_{\mathsf{lex}} \mathbb{E}^{\sigma}_{s}(\bar{f})$. We formalise this argument below.

Theorem 14.1. *Assume that $\bar{f}$ is universally integrable. Let σ be a strategy and $s \in S$. There exists a pure strategy τ such that $\mathbb{E}^{\sigma}_{s}(\bar{f}) \leq_{\mathsf{lex}} \mathbb{E}^{\tau}_{s}(\bar{f})$.*

Proof. Let μ be a mixed strategy that is outcome-equivalent to σ (whose existence follows from Kuhn's theorem). To prove the theorem, we reason on the μ-integral of random variables of $(\Sigma_{\mathsf{pure}}(\mathcal{M}), \mathcal{F}_{\Sigma_{\mathsf{pure}}(\mathcal{M})})$; see Chapter 2.4.3, Page 36, for the definition of the σ-algebra $\mathcal{F}_{\Sigma_{\mathsf{pure}}(\mathcal{M})}$. For any real or multivariate random variable Y over $\Sigma_{\mathsf{pure}}(\mathcal{M})$, we write $\int Y \mathrm{d}\mu$ for $\int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} Y(\tau)\mathrm{d}\mu(\tau)$ to lighten notation.

For all $1 \leq j \leq d$, we consider the real random variable $X_j \colon x \mapsto \mathbb{E}^{\tau_x}_{s}(f_j)$ over $\Sigma_{\mathsf{pure}}(\mathcal{M})$. We let $\mathcal{X} = (X_1, \dots, X_d)$. Because $\bar{f}$ is universally integrable, $\mathcal{X}$ is μ-integrable and, by Lemma 13.4, we have $\mathbb{E}^{\mu}_{s}(\bar{f}) = \int \mathcal{X} \mathrm{d}\mu$.

Let $\mathcal{Y} = (Y_1, \dots, Y_d)$ be an integrable multi-variate real random variable over $\Sigma_{\mathsf{pure}}(\mathcal{M})$. We first show that if $\mathcal{X} <_{\mathsf{lex}} \mathcal{Y}$, then $\mathbb{E}^{\sigma}_{s}(\bar{f}) = \int \mathcal{X} \mathrm{d}\mu <_{\mathsf{lex}} \int \mathcal{Y} \mathrm{d}\mu$. We use this claim below to prove the theorem.

Assume that $\mathcal{X} <_{\mathsf{lex}} \mathcal{Y}$. We partition $\Sigma_{\mathsf{pure}}(\mathcal{M})$ as follows. For all $1 \leq j \leq d$, we let

$$E_j = \left\{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M}) \mid X_j(\tau) < Y_j(\tau) \text{ and } \forall j' < j,\, X_{j'}(\tau) = Y_{j'}(\tau)\right\}.$$

Intuitively, E_j is the set of elements such that the strict lexicographic ordering of their respective images by $\mathcal{X}$ and $\mathcal{Y}$ is witnessed in component j. The sets E_1, …, E_d partition $\Sigma_{\mathsf{pure}}(\mathcal{M})$ because $\mathcal{X} <_{\mathsf{lex}} \mathcal{Y}$. It follows that, for $\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\}$, we have $\int \mathcal{Z} \mathrm{d}\mu = \sum_{j=1}^{d} \int \mathcal{Z} \cdot \mathbb{1}_{E_j} \mathrm{d}\mu$.

Let $j^\star = \min\{1 \leq j \leq d \mid \mu(E_j) > 0\}$. To obtain that $\int \mathcal{X} \mathrm{d}\mu <_{\mathsf{lex}} \int \mathcal{Y} \mathrm{d}\mu$, we show that, for $j < j^\star$, we have $\int X_j \mathrm{d}\mu = \int Y_j \mathrm{d}\mu$ and $\int X_{j^\star} \mathrm{d}\mu < \int Y_{j^\star} \mathrm{d}\mu$. To prove these relations, we formulate two observations. First, we observe that for all $1 \leq j < j' \leq d$, X_j and Y_j agree over $E_{j'}$ (by definition), and thus we have

$$\int X_j \cdot \mathbb{1}_{E_{j'}} \mathrm{d}\mu = \int Y_j \cdot \mathbb{1}_{E_{j'}} \mathrm{d}\mu. \tag{14.1}$$

Second, since $\mu(E_{j'}) = 0$ for all $j' < j^\star$, it follows that for all $1 \leq j \leq d$ and all $Z_j \in \{X_j, Y_j\}$ that

$$\int Z_j \mathrm{d}\mu = \sum_{j'=j^\star}^{d} \int Z_j \cdot \mathbb{1}_{E_{j'}} \mathrm{d}\mu. \tag{14.2}$$

Let $1 \leq j < j^\star$. By combining Equations (14.1) and (14.2), we obtain that

$$\int X_j \mathrm{d}\mu = \sum_{j'=j^\star}^{d} \int X_j \cdot \mathbb{1}_{E_{j'}} \mathrm{d}\mu = \sum_{j'=j^\star}^{d} \int Y_j \cdot \mathbb{1}_{E_{j'}} \mathrm{d}\mu = \int Y_j \mathrm{d}\mu.$$

To end the proof that $\int \mathcal{X} \mathrm{d}\mu <_{\mathsf{lex}} \int \mathcal{Y} \mathrm{d}\mu$, it remains to show that $\int X_{j^\star} \mathrm{d}\mu < \int Y_{j^\star} \mathrm{d}\mu$. This inequality is equivalent to $\int X_{j^\star} \cdot \mathbb{1}_{E_{j^\star}} \mathrm{d}\mu < \int Y_{j^\star} \cdot \mathbb{1}_{E_{j^\star}} \mathrm{d}\mu$ by Equations (14.1) and (14.2). This inequality holds by compatibility of the Lebesgue integral with the order of $\mathbb{R}$ (recall that $X_{j^\star}(\tau) < Y_{j^\star}(\tau)$ for all $\tau \in E_{j^\star}$) and because $\mu(E_{j^\star}) > 0$. We have shown that $\int \mathcal{X} \mathrm{d}\mu <_{\mathsf{lex}} \int \mathcal{Y} \mathrm{d}\mu$ whenever $\mathcal{X} <_{\mathsf{lex}} \mathcal{Y}$ holds.

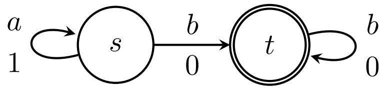


Figure 14.2: A deterministic MDP in which randomisation is need to accumulate an infinite total reward while reaching the target state t almost-surely.

Now assume that for all pure strategies τ, we have $\mathbb{E}^{\tau}_{s}(\bar{f}) <_{\mathsf{lex}} \mathbb{E}^{\sigma}_{s}(\bar{f})$ towards a contradiction. It follows that $\mathcal{X}(\tau) <_{\mathsf{lex}} \mathbb{E}^{\sigma}_{s}(\bar{f})$ for all $\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})$. From the above, we obtain that $\mathbb{E}^{\sigma}_{s}(\bar{f}) = \int \mathcal{X} \mathrm{d}\mu <_{\mathsf{lex}} \mathbb{E}^{\sigma}_{s}(\bar{f})$, which is a contradiction. □

A direct corollary of Theorem 14.1 is the following.

Corollary 14.2. *Assume that $\bar{f}$ is universally integrable. For all $s \in S$, if there exists a lexicographically optimal strategy from s, then there exists a pure lexicographically optimal strategy.*

We now provide an example illustrating that Theorem 14.1 does not hold without assuming that $\bar{f}$ is universally integrable.

Example 14.1. We consider the MDP $\mathcal{M}$ depicted in Figure 14.2. Let w denote the illustrated weight function. We consider the two-dimensional payoff function $\bar{f} = (f_1, f_2)$ such that $f_1 = \mathbb{1}_{\mathsf{Reach}(\{t\})}$ and $f_2 = \mathsf{TRew}_w$. We prove that randomisation is necessary to play lexicographically optimally from s.

Since transitions of $\mathcal{M}$ are deterministic, $\mathsf{Pay}^{\mathsf{pure}}_{s}(\bar{f})$ is the set of payoffs of plays from s. We introduce notation for these plays: for all $r \in \mathbb{N}$, let $\pi_r = (sa)^r s(bt)^\omega$ and let $\pi_\infty = (sa)^\omega$. It follows that $\mathsf{Pay}^{\mathsf{pure}}_{s}(\bar{f}) = \{\bar{f}(\pi_r) \mid r \in \mathbb{N} \cup \{\infty\}\} = \{(1, r) \mid r \in \mathbb{N}\} \cup \{(0, +\infty)\}$. In particular, no pure strategy has an expected payoff of $(1, +\infty)$, which is the greatest that could occur with $\bar{f}$.

However, there exists a randomised strategy whose expected payoff from s is $(1, +\infty)$. For each $r \in \mathbb{N}$, let τ_r be a pure strategy whose outcome from s is π_r. We consider the mixed strategy μ such that, for all $r \in \mathbb{N}$, μ assigns probability $2^{-(r+1)}$ to τ_{2^r}. It follows that $\mathbb{E}^{\mu}_{s}(\bar{f}) = \sum_{r=0}^{\infty} \mu(\tau_{2^r}) \cdot \bar{f}(\pi_{2^r}) = (1, \infty)$, i.e., μ is lexicographically optimal from s. ◁

14.3 Universally integrable payoffs

The goal of this section is to show that if $\bar{f}$ is universally integrable, then for all $s \in S$, $\mathsf{Pay}_s(\bar{f}) = \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$. We provide an overview of the proof of this property in Section 14.3.1. We formally prove the result in Section 14.3.2.

Throughout this section, we assume that $\bar{f}$ is universally integrable.

14.3.1 Proof overview

Let $s \in S$. By convexity of $\mathsf{Pay}_s(\bar{f})$, the inclusion $\mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})) \subseteq \mathsf{Pay}_s(\bar{f})$ holds. Therefore, the main difficulty of the proof is to prove the other inclusion.

Let $\mathbf{q} \in \mathsf{Pay}_s(\bar{f})$. The first step of the proof is to construct a linear map $L_{\mathbf{q}} \colon \mathbb{R}^d \to \mathbb{R}^{d'}$ with $d' \leq d$ such that

(i) the vector $L_{\mathbf{q}}(\mathbf{q})$ is the lexicographic maximum of $L_{\mathbf{q}}(\mathsf{Pay}_s(\bar{f}))$ and

(ii) we have $\mathbf{q} \in \mathsf{ri}(\mathsf{Pay}_s(\bar{f}) \cap V)$ where $V = L_{\mathbf{q}}^{-1}(L_{\mathbf{q}}(\mathbf{q}))$ denotes the set of vectors that share their image by $L_{\mathbf{q}}$ with $\mathbf{q}$.

The mapping $L_{\mathbf{q}}$ is constructed as follows (Examples 14.2 and 14.3 below illustrate this construction). If $\mathbf{q}$ is in the relative interior of $\mathsf{Pay}_s(\bar{f})$, we let $L_{\mathbf{q}}$ be the zero-valued linear form. Otherwise, by the supporting hyperplane theorem (Theorem 2.4), there exists a linear form x_1^* such that $x_1^*(\mathbf{q}) \geq x_1^*(\mathbf{p})$ for all $\mathbf{p} \in \mathsf{Pay}_s(\bar{f})$. Let $H_1 = (x_1^*)^{-1}(x_1^*(\mathbf{q}))$ denote the supporting hyperplane given by x_1^*. We check whether $\mathbf{q}$ is in the relative interior of $\mathsf{Pay}_s(\mathbf{q}) \cap H_1$. If it is the case, we obtain the desired linear mapping by letting $L_{\mathbf{q}} = x_1^*$. Otherwise, we continue: there is a linear form x_2^* describing a hyperplane H_2 that supports $\mathsf{Pay}_s(\mathbf{q}) \cap H_1$ at $\mathbf{q}$. We choose x_2^* such that x_1^* and x_2^* have distinct kernels, i.e., $H_1 \neq H_2$. To ensure that the kernels are distinct, we construct x_2^* as an extension to $\mathbb{R}^d$ of a non-zero linear form over $\mathsf{ker}(x_1^*)$. By induction, we continue constructing linear forms with pairwise distinct kernels (i.e., defining pairwise distinct supporting hyperplanes) until $\mathbf{q} \in \mathsf{ri}(\mathsf{Pay}_s(\bar{f}) \cap \bigcap_{1 \leq j \leq d'} H_j)$. When this condition is satisfied, we define, for all $\mathbf{v} \in \mathbb{R}^d$, $L_{\mathbf{q}}(\mathbf{v}) = (x_1^*(\mathbf{v}), \ldots, x_{d'}^*(\mathbf{v}))$. The invocation of the supporting hyperplane theorem at each iteration guarantees that $L_{\mathbf{q}}(\mathbf{q})$ is a lexicographic maximum of $L_{\mathbf{q}}(\mathsf{Pay}_s(\bar{f}))$. We remark that, when the stopping condition is fulfilled, the supporting hyperplane theorem is no longer applicable.

Let $V = L_{\mathbf{q}}^{-1}(L_{\mathbf{q}}(\mathbf{q}))$. By construction, $\mathbf{q} \in \mathsf{ri}(V \cap \mathsf{Pay}_s(\bar{f}))$. To conclude, we establish that $\mathsf{ri}(V \cap \mathsf{Pay}_s(\bar{f})) = \mathsf{ri}(V \cap \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$. This suffices because the latter set is a subset of $\mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$. This is equivalent to showing that $\mathsf{cl}(V \cap \mathsf{Pay}_s(\bar{f})) = \mathsf{cl}(V \cap \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$: convex subsets of $\mathbb{R}^d$ have the same relative interior as their closure [Roc70, Thm. 6.3].

On the one hand, the inclusion $\mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})) \subseteq \mathsf{Pay}_s(\bar{f})$ implies that $\mathsf{cl}(V \cap \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))) \subseteq \mathsf{cl}(V \cap \mathsf{Pay}_s(\bar{f}))$. For the other inclusion, we need only show that

$$V \cap \mathsf{Pay}_s(\bar{f}) \subseteq \mathsf{cl}(V \cap \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))).$$

We assume towards a contradiction that this is not the case. We fix a vector $\mathbf{p} \in V \cap \mathsf{Pay}_s(\bar{f}) \setminus \mathsf{cl}(V \cap \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$. Using the hyperplane separation theorem (Theorem 2.3), we obtain $x_\star^*$ such that $x_\star^*(\mathbf{p}) > x_\star^*(\mathbf{v})$ for all $\mathbf{v} \in \mathsf{cl}(V \cap \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$. Because $\mathbf{p} \in \mathsf{Pay}_s(\bar{f})$, Theorem 14.1 implies that there exists a pure strategy σ such that $(L_{\mathbf{q}}(\mathbf{p}), x_\star^*(\mathbf{p})) \leq_{\mathsf{lex}} (L_{\mathbf{q}}(\mathbb{E}_s^\sigma(\bar{f})), x_\star^*(\mathbb{E}_s^\sigma(\bar{f})))$. Furthermore, $\mathbf{p} \in V$ implies that $L_{\mathbf{q}}(\mathbf{p})$ is the lexicographic maximum of $L_{\mathbf{q}}(\mathsf{Pay}_s(\bar{f}))$, and thus so is $L_{\mathbf{q}}(\mathbb{E}_s^\sigma(\bar{f}))$, i.e., $\mathbb{E}_s^\sigma(\bar{f}) \in V \cap \mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$. It follows that $x_\star^*(\mathbf{p}) \leq x_\star^*(\mathbb{E}_s^\sigma(\bar{f}))$, which is contradictory. This closes the argument that $\mathsf{cl}(V \cap \mathsf{Pay}_s(\bar{f})) = \mathsf{cl}(V \cap \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$, which implies that $\mathbf{q} \in \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$, ending the sketch.

We complement the sketch above with two examples that illustrate the construction of the mapping $L_{\mathbf{q}}$. In the first example, we select $\mathbf{q}$ as an extreme point of the set of expected payoffs. In this case, the constructed linear mapping $L_{\mathbf{q}}$ is such that $\mathbf{q}$ is the unique vector $\mathbf{p} \in \mathbb{R}^d$ such that $L_{\mathbf{q}}(\mathbf{p})$ is the lexicographic maximum of $L_{\mathbf{q}}(\mathsf{Pay}_s(\bar{f}))$. In our second example, $\mathsf{Pay}_s(\bar{f})$ is not closed and we choose $\mathbf{q}$ as a non-extreme point such that no pure strategy has expected payoff $\mathbf{q}$. In this case, we observe that the uniqueness property of the first example cannot be obtained no matter which linear forms are used to construct $L_{\mathbf{q}}$.

Example 14.2 (Extreme point)**.** We consider the MDP depicted in Figure 14.3. Throughout this example, we (implicitly) consider s_0 as the initial state. We study indicators of reachability objectives: we let $\bar{f} = (\mathbb{1}_{\mathsf{Reach}(T_1)}, \mathbb{1}_{\mathsf{Reach}(T_2)})$ where $T_1 = \{s_1, s_4\}$ and $T_2 = \{s_2, s_4\}$. The set $\mathsf{Pay}_{s_0}(\bar{f})$ is depicted in Figure 14.4: it is the convex hull of the expected payoffs of the pure strategies

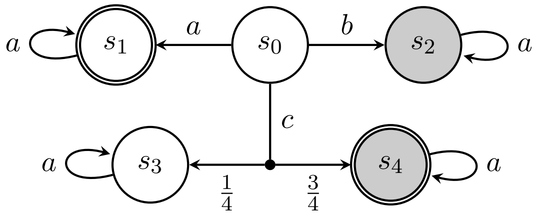
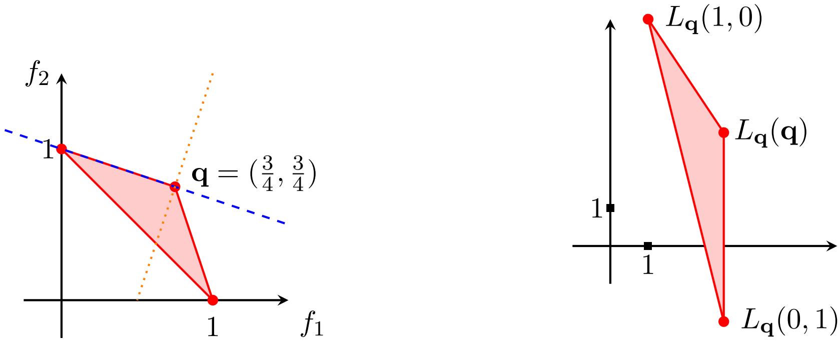


Figure 14.3: An MDP. The doubly circled and filled states respectively highlight the targets of the reachability objectives $\mathsf{Reach}(\{s_1, s_4\})$ and $\mathsf{Reach}(\{s_2, s_4\})$.



(a) The blue dashed line ($x + 3y = 3$) is a hyperplane supporting $\mathsf{Pay}_s(\bar{f})$ at $\mathbf{q}$. The orange dotted line ($6x - 2y = 3$) is a hyperplane obtained from an extension of the linear form defining the hyperplane $\{\mathbf{q}\}$ of the blue line.

(b) The image of the set on the left by the linear mapping $L_\mathbf{q} \colon (v_1, v_2) \mapsto (v_1 + 3v_2, 6v_1 - 2v_2)$ obtained through the equations of the hyperplanes on the left. Remark that the image of $\mathbf{q}$ is lexicographically optimal.

Figure 14.4: The set of expected payoffs for Example 14.2 and its image by a linear mapping.

that play one of the actions a, b or c in s_0. We focus on the extreme point $\mathbf{q} = (\frac{3}{4}, \frac{3}{4})$ in the remainder of this example.

The construction of $L_\mathbf{q}$ is in two steps. First, we consider the linear form x_1^* defined by, for all $\mathbf{v} \in \mathbb{R}^d$, $x_1^*(\mathbf{v}) = v_1 + 3v_2$. The hyperplane $H = (x_1^*)^{-1}(3)$ supports $\mathsf{Pay}_{s_0}(\bar{f})$ at $\mathbf{q}$ and is depicted by the blue dashed line in Figure 14.4a. It is not satisfactory to set $L_\mathbf{q} = x_1^*$: $\mathbf{q}$ is an endpoint of the segment $[(0,1), \mathbf{q}] = \mathsf{Pay}_{s_0}(\bar{f}) \cap H$, and is therefore not in its relative interior.

We recall that for any linear form y^* of $\mathsf{ker}(x_1^*)$ (i.e., the vector space corresponding to H), there exists $\mathbf{v} \in \mathsf{ker}(x_1^*)$ such that $y^*(\mathbf{w}) = \langle \mathbf{w}, \mathbf{v} \rangle$ for all $\mathbf{w} \in \mathsf{ker}(x_1^*)$. Since any non-zero linear form of $\mathsf{ker}(x_1^*)$ is bijective, all of them induce a hyperplane of H supporting $\mathsf{Pay}_s(\bar{f}) \cap H$ at $\mathbf{q}$. We proceed with the linear form $x_2^* \colon \mathbb{R}^2 \to \mathbb{R}$ defined by $x_2^*(\mathbf{v}) = 6v_1 - 2v_2$ for all $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$ (derived from the vector $\mathbf{w} = (6, -2) \in \mathsf{ker}(x_1^*)$). Observe that $\mathsf{ker}(x_1^*) \cap \mathsf{ker}(x_2^*) = \{\mathbf{0}\}$.

We define $L_\mathbf{q}(\mathbf{v}) = (x_1^*(\mathbf{v}), x_2^*(\mathbf{v}))$. Since $L_\mathbf{q}$ is bijective, $L_\mathbf{q}^{-1}(L_\mathbf{q}(\mathbf{q}))$ is a singleton set. Therefore, $\mathbf{q}$ is in the relative interior of $L_\mathbf{q}^{-1}(L_\mathbf{q}(\mathbf{q})) \cap \mathsf{Pay}_{s_0}(\bar{f})$. By linearity of the expectation, it holds that $\mathsf{Pay}_{s_0}(L_\mathbf{q} \circ \bar{f}) = L_\mathbf{q}(\mathsf{Pay}_{s_0}(\bar{f}))$. This set is illustrated in Figure 14.4b; it is easy to check that $L_\mathbf{q}(\mathbf{q})$ is the lexicographic maximum of this set.

In this case, $L_\mathbf{q}$ allows us to deduce that there exists a pure strategy σ such that $\mathbb{E}_{s_0}^{\sigma}(\bar{f}) = \mathbf{q}$. On the one hand, by Theorem 14.1, there exists a pure strategy σ that is lexicographically optimal from s_0 for $L_\mathbf{q} \circ \bar{f}$. On the other hand, the only payoff vector $\mathbf{p} \in \mathsf{Pay}_{s_0}(\bar{f})$ such that $L_\mathbf{q}(\mathbf{p})$ is the lexicographic maximum of $L_\mathbf{q}(\mathsf{Pay}_{s_0}(\bar{f}))$ is $\mathbf{q}$. This implies that $\mathbf{q}$ is the payoff of the pure strategy σ, and thus $\mathbf{q} \in \mathsf{Pay}_{s_0}^{\mathsf{pure}}(\bar{f})$. ◁

Remark 14.3 (The necessity of induction). In Example 14.2, it is possible to isolate $\mathbf{q}$ by a supporting hyperplane of $\mathsf{Pay}_{s_0}(\bar{f})$. We could thus choose $L_\mathbf{q}$ as a linear form and bypass the induction step of the construction of $L_\mathbf{q}$ here. We avoid using a linear form for the sake of illustration, as we cannot use linear forms to isolate extreme points in general. In fact, this can be shown via the example of Section 14.1.

We recall the set of expected payoffs of this example in Figure 14.5. Precisely,

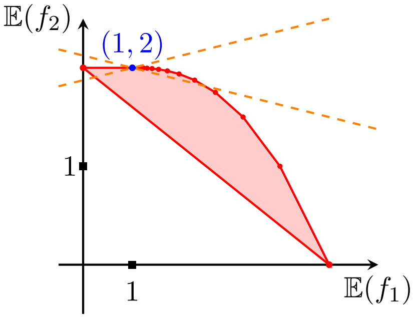


Figure 14.5: The set of expected payoffs of the example presented in Section 14.1. The line passing through $(0,2)$ and $(1,2)$ is the unique hyperplane supporting D at the point $(1,2)$.

it is the set

$$D = \mathsf{conv}\left(\{(0,2),(1,2)\} \cup \left\{\left(1+\frac{3^r}{4^{r-1}}, 2-\frac{1}{2^{r-1}}\right) \mid r \in \mathbb{N}\right\}\right).$$

We show that the only hyperplane supporting D at $\mathbf{q} = (1,2)$ is the line carrying the segment $[(0,2),(1,2)]$. The slanted lines passing through $\mathbf{q}$ in Figure 14.5 suggest that any other hyperplane is not a supporting hyperplane of D. We formalise this idea.

Assume towards a contradiction that there exists a linear form $x^*\colon \mathbb{R}^2 \to \mathbb{R}$ such that for all $\mathbf{p} \in D$, $x^*(\mathbf{q}) \geq x^*(\mathbf{p})$ and $x^*(\mathbf{q}) \neq x^*((0,2))$. Let $\alpha, \beta \in \mathbb{R}$ such that for all $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$, $x^*(\mathbf{v}) = \alpha v_1 + \beta v_2$. We observe that $x^*(\mathbf{q}) > x^*((0,2))$ is equivalent to $\alpha > 0$. We also have, for all $\ell \in \mathbb{N}$, $x^*(\mathbf{q}) \geq x^*((1+\frac{3^\ell}{4^{\ell-1}}, 2-\frac{1}{2^{\ell-1}}))$, i.e., $\beta \geq \frac{3^\ell}{2^{\ell-1}} \cdot \alpha$. The previous properties imply that β must be greater than all real numbers, which is a contradiction.

We have shown that $\mathbf{q}$ cannot be isolated from the other elements of D with a linear form, which implies that the induction step in the construction of $L_\mathbf{q}$ cannot be bypassed when dealing with extreme points in general. $\lhd$

The construction outlined in Example 14.2 can be generalised to show that all extreme points of the set of expected payoffs of $\bar{f}$ from a given state are the expected payoff of a pure strategy. However, this property is not sufficient to show that all expected payoffs are convex combinations of pure expected payoffs. In particular, the following example highlights that some expected

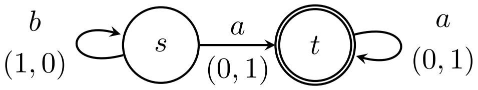
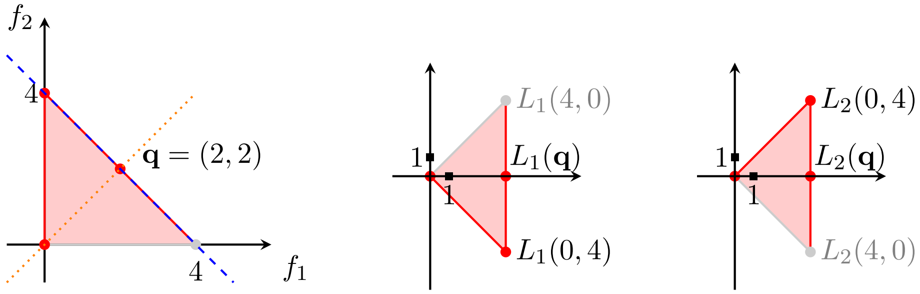


Figure 14.6: An MDP with deterministic transitions. The pairs beneath actions represent a two-dimensional weight function. State t is doubly circled because it is a target.



(a) The blue dashed line ($x+y=4$) is a hyperplane supporting $\mathsf{Pay}_s(\bar{f})$ at $\mathbf{q}$. The orange dotted line ($x-y=0$) is an orthogonal hyperplane included for reference for the adjacent figures.

(b) Image of the payoff set in Figure 14.7a by the linear mapping L_1 such that $(v_1, v_2) \mapsto (v_1 + v_2, v_1 - v_2)$.

(c) Image of the payoff set in Figure 14.7a by the linear mapping L_2 such that $(v_1, v_2) \mapsto (v_1 + v_2, v_2 - v_1)$.

Figure 14.7: The set of expected payoffs for Example 14.3 and its image by two (related) linear functions. The segment $](0,0),(4,0)]$ in grey does not intersect $\mathsf{Pay}_s(\bar{f})$. Its image is similarly coloured in the two other illustrations.

payoffs are not convex combinations of extreme points of the set of expected payoffs.

Example 14.3 (Non-extreme point)**.** We consider the MDP depicted in Figure 14.6. We assume that state s is the initial state throughout this example. Let $w = (w_1, w_2)$ denote the two-dimensional weight function given on the illustration. We consider a two-dimensional payoff function $\bar{f}$. The payoff of a play, for each dimension, is zero if t is not visited and, otherwise, its payoff is given by a discounted-sum payoff. We formalise this as the product of a discounted-sum payoff and an indicator. Therefore, formally, $\bar{f} = (f_1, f_2)$ is such that, for $j \in \{1,2\}$, $f_j = \mathbb{1}_{\mathsf{Reach}(\{t\})} \cdot \mathsf{DSum}^{\frac{3}{4}}_{w_j}$. We observe that, by definition of w, $f_2 = \mathsf{DSum}_{w_2}^{\frac{3}{4}}$.

The set $\mathsf{Pay}_s(\bar{f})$ is illustrated in Figure 14.7a. Any vector in $\mathsf{Pay}_s(\bar{f})$ is a convex combination of $\mathbf{0}$ and a vector in the segment $[(0,4),(4,0)[$. In particular, no strategy has an expected payoff of $(4,0)$ from s. We can derive $\mathsf{Pay}_s(\bar{f})$ from $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}) = \{\mathbf{0}\} \cup \{(4 - \frac{3^\ell}{4^{\ell-1}}, \frac{3^\ell}{4^{\ell-1}}) \mid \ell \in \mathbb{N}\}$. To obtain $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$, we note that any pure strategy in this MDP induces a single play from s, because all transitions are deterministic. On the one hand, we can obtain the payoff $\mathbf{0}$ with the play $(sb)^\omega$ (the payoff is zero on the first dimension because t is not visited). On the other hand, for all $\ell \in \mathbb{N}$, we have $\bar{f}((sb)^\ell s(at)^\omega) = \left(4 - \frac{3^\ell}{4^{\ell-1}}, \frac{3^\ell}{4^{\ell-1}}\right)$.

We consider the payoff vector $\mathbf{q} = (2,2)$ and construct $L_\mathbf{q}$. We remark that the vector $\mathbf{q}$ is not a convex combination of extreme points of $\mathsf{Pay}_s(\bar{f})$. Therefore, is not possible to conclude that $\mathbf{q} \in \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$ by adapting the argument of Example 14.2 to deal with all extreme points. The only hyperplane H that support $\mathsf{Pay}_s(\bar{f})$ at $\mathbf{q}$ is the line depicted in blue in Figure 14.7a. We let $x_1^* \colon \mathbb{R}^2 \to \mathbb{R}$ be the linear form defined by $x_2^*(\mathbf{v}) = v_1 + v_2$ for all $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$. We have $H = (x_1^*)^{-1}(4)$. We observe (via Figure 14.7a) that $\mathbf{q}$ is in the relative interior of $\mathsf{Pay}_s(\bar{f}) \cap H$. We define $L_\mathbf{q} = x_1^*$.

To close this example, we provide an argument based on $L_\mathbf{q}^{-1}(L_\mathbf{q}(\mathbf{q}))$ being a line to show that $\mathbf{q}$ is a convex combination of expected payoffs of pure strategies. While this argument differs from the general proof provided below, it can be generalised to show that $\mathbf{q} \in \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$ whenever $L_\mathbf{q}^{-1}(L_\mathbf{q}(\mathbf{q}))$ is a line. This argument consists in showing that there are payoffs of pure strategies on either side of $\mathbf{q}$ on the line segment $\mathsf{Pay}_s(\bar{f}) \cap H = [(0,4),(4,0)[$. This implies that $\mathbf{q} \in \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$.

We fix a direction vector $\mathbf{v}_H = (1,-1)$ of H. For all vectors $\mathbf{p}$ of $\mathsf{Pay}_s(\bar{f})$ in $[\mathbf{q},(4,0)[$ (resp. $[\mathbf{q},(0,4)]$), we have $\langle \mathbf{p}, \mathbf{v}_H\rangle \geq \langle \mathbf{q}, \mathbf{v}_H\rangle$ (resp. $\langle \mathbf{p}, -\mathbf{v}_H\rangle \geq \langle \mathbf{q}, -\mathbf{v}_H\rangle$). Consider the linear mappings $L_1 \colon \mathbf{w} \mapsto (x_1^*(\mathbf{w}), \langle \mathbf{w}, \mathbf{v}_H\rangle)$ and $L_2 \colon \mathbf{w} \mapsto (x_1^*(\mathbf{w}), \langle \mathbf{w}, -\mathbf{v}_H\rangle)$ over $\mathbb{R}^2$. We illustrate the image of $\mathsf{Pay}_s(\bar{f})$ by L_1 and L_2 respectively in Figure 14.7b and Figure 14.7c.

Theorem 14.1 implies that, for $i \in \{1,2\}$, there exists a pure strategy σ_i such that $L_i(\mathbf{p}_i) \geq_{\mathsf{lex}} L_i(\mathbf{q})$ where $\mathbf{p}_i = \mathbb{E}_s^{\sigma_i}(\bar{f})$. We obtain, by definition of L_i, that $x_1^*(\mathbf{p}_i) = x_1^*(\mathbf{q})$ for $i \in \{1,2\}$, as x_1^* supports $\mathsf{Pay}_s(\bar{f})$ at $\mathbf{q}$. Therefore, $\alpha_1 := \langle \mathbf{p}_1, \mathbf{v}_H\rangle \geq 0$ and $\alpha_2 := \langle \mathbf{p}_2, \mathbf{v}_H\rangle \leq 0$. Furthermore, for $i \in \{1,2\}$, it holds that $\mathbf{p}_i = \mathbf{q} + \frac{\alpha_i}{\|\mathbf{v}_H\|_2}\mathbf{v}_H$ because $(\frac{1}{\|\mathbf{q}\|_2}\mathbf{q}, \frac{1}{\|\mathbf{v}_H\|_2}\mathbf{v}_H)$

is an orthonormal basis of $\mathbb{R}^2$, $\langle \mathbf{p}_i, \frac{1}{\|\mathbf{q}\|_2}\mathbf{q}\rangle = \frac{2}{\|\mathbf{q}\|_2} \cdot x_1^*(\mathbf{p}_i) = 2\sqrt{2} = \|\mathbf{q}\|_2$ (as $x_1^*(\mathbf{p}_i) = x_1^*(\mathbf{q})$) and $\langle \mathbf{p}_i, \frac{1}{\|\mathbf{v}_H\|_2}\mathbf{v}_H\rangle = \frac{\alpha_i}{\|\mathbf{v}_H\|_2}$ (by definition of α_i). Thus, $\mathbf{q} \in [\mathbf{p}_1, \mathbf{p}_2] \subseteq \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$. $\lhd$

14.3.2 Theorem statement and proof

We now formally state the main theorem of this section and prove it.

Theorem 14.4. *Assume that $\bar{f}$ is universally integrable. For all $s \in S$, we have $\mathsf{Pay}_s(\bar{f}) = \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$. In other words, the expected payoff of any strategy is also the expected payoff of a finite-support mixed strategy.*

Proof. Throughout this proof, we assume that for all $1 \leq j \leq d$, f_j is a real-valued payoff. This is without loss of generality: these payoffs are universally integrable, and thus are $\mathbb{P}_s^\sigma$-almost-surely real-valued for all $\sigma \in \Sigma(\mathcal{M})$ and $s \in S$.

It is sufficient to show that $\mathsf{Pay}_s(\bar{f}) \subseteq \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$. Let $\mathbf{q} \in \mathsf{Pay}_s(\bar{f})$. We construct the linear mapping $L_\mathbf{q}$ as explained in the sketch, i.e., such that $L_\mathbf{q}(\mathbf{q})$ is the lexicographic maximum of $L_\mathbf{q}(\mathsf{Pay}_s(\bar{f}))$ and $\mathbf{q} \in \mathsf{ri}(L_\mathbf{q}^{-1}(L_\mathbf{q}(\mathbf{q})) \cap \mathsf{Pay}_s(\bar{f}))$.

Let $D = \mathsf{Pay}_s(\bar{f}) - \mathbf{q}$. We observe that $L_\mathbf{q}$ satisfies the conditions above if and only if $L_\mathbf{q}(\mathbf{0})$ is the lexicographic maximum of $L_\mathbf{q}(D)$ and $\mathbf{0}$ is in the relative interior of $D \cap \mathsf{ker}(L_\mathbf{q})$. We construct $L_\mathbf{q}$ by working with D and $\mathbf{0}$ instead of $\mathsf{Pay}_s(\bar{f})$ and $\mathbf{q}$. This allows us to work with vector sub-spaces instead of affine spaces, overall simplifying the presentation.

We let $y_0^* \colon \mathbb{R}^d \to \mathbb{R}$ be the constant zero function. If $\mathbf{0} \in \mathsf{ri}(D)$, we let $L_\mathbf{q} = y_0^*$. This function satisfies the desired properties. We now assume that $\mathbf{0} \notin \mathsf{ri}(D)$. We inductively define a sequence of non-zero linear forms $y_1^*, \ldots, y_{d'}^*$ such that $y_j^* \colon \mathsf{ker}(y_{j-1}^*) \to \mathbb{R}$ for all $j \in [\![1, d']\!]$. Next, for all $j \in [\![1, d']\!]$, we extend y_j^* to a form $x_j^* \colon \mathbb{R}^d \to \mathbb{R}$. Finally, we define the mapping $L_\mathbf{q}$ as $L_\mathbf{q}(\mathbf{v}) = (x_1^*(\mathbf{v}), \ldots, x_{d'}^*(\mathbf{v}))$ for all $\mathbf{v} \in \mathbb{R}^d$ and show that it satisfies the desired properties.

Let $j \geq 1$. By induction, assume that y_{j-1}^* is defined (this is the case even for $j = 1$). We distinguish two cases. If $\mathbf{0} \in \mathsf{ri}(\mathsf{ker}(y_{j-1}^*) \cap D)$, we stop the construction. We remark that if $j = d+1$, then $\mathsf{ker}(y_{j-1}^*)$ is a singleton set and we are necessarily in this case (i.e., $d' \leq d$). Now, assume that $\mathbf{0} \notin \mathsf{ri}(\mathsf{ker}(y^*_{j-1}) \cap D)$. The supporting hyperplane theorem (Theorem 2.4) implies that there exists a linear form $y^*_j \colon \mathsf{ker}(y^*_{j-1}) \to \mathbb{R}$ such that for all $\mathbf{p} \in D \cap \mathsf{ker}(y^*_{j-1})$, we have $y^*_j(\mathbf{p}) \leq 0 = y^*_j(\mathbf{0})$. This allows us to continue with the induction.

Assume that the procedure above has provided linear forms $y^*_1, \dots, y^*_{d'}$. We now extend them to $\mathbb{R}^d$. Let $j \in [\![1, d']\!]$. There exists $\mathbf{w}_j \in \mathsf{ker}(y^*_{j-1}) \subseteq \mathbb{R}^d$ such that for all $\mathbf{v} \in \mathsf{ker}(y^*_{j-1})$, we have $y^*_j(\mathbf{v}) = \langle \mathbf{v}, \mathbf{w}_j \rangle$. We define $x^*_j \colon \mathbb{R}^d \to \mathbb{R}$ by, for all $\mathbf{v} \in \mathbb{R}^d$, $x^*_j(\mathbf{v}) = \langle \mathbf{v}, \mathbf{w}_j \rangle$. We let $L_{\mathbf{q}} \colon \mathbb{R}^d \to \mathbb{R}^{d'}$ be such that $L_{\mathbf{q}}(\mathbf{v}) = (x^*_1(\mathbf{v}), \dots, x^*_{d'}(\mathbf{v}))$ for all $\mathbf{v} \in \mathbb{R}^d$.

We now show that $L_{\mathbf{q}}$ satisfies the required properties. By construction, for all $\mathbf{p} \in D$ and all $j \in [\![1, d']\!]$, if $x^*_{j'}(\mathbf{p}) = 0$ for all $j' \in [\![1, j-1]\!]$, then necessarily $\mathbf{p} \in \mathsf{ker}(y^*_{j-1})$, and thus $x^*_j(\mathbf{p}) \leq 0$. This implies that for all $\mathbf{p} \in D$, $L_{\mathbf{q}}(\mathbf{p}) \leq_{\mathsf{lex}} L_{\mathbf{q}}(\mathbf{0})$. This shows that $L_{\mathbf{q}}(\mathbf{0})$ is the lexicographic maximum of $L_{\mathbf{q}}(D)$. Next, we show that $\mathbf{0} \in \mathsf{ri}(D \cap \mathsf{ker}(L_{\mathbf{q}}))$. This follows from the stopping condition in the construction of $L_{\mathbf{q}}$ and the equality $\mathsf{ker}(L_{\mathbf{q}}) = \bigcap_{1 \leq j \leq d'} \mathsf{ker}(x^*_j) = \mathsf{ker}(y^*_{d'})$ (the second equality follows from $x^*_1 = y^*_1$ and $\mathsf{ker}(y^*_1) \supsetneq \dots \supsetneq \mathsf{ker}(y^*_{d'})$).

Let $V = L^{-1}_{\mathbf{q}}(L_{\mathbf{q}}(\mathbf{q}))$. We have shown that $\mathbf{q} \in \mathsf{ri}(V \cap \mathsf{Pay}_s(\bar{f}))$. It suffices to show that $\mathsf{ri}(V \cap \mathsf{Pay}_s(\bar{f})) = \mathsf{ri}(V \cap \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_s(\bar{f})))$ to conclude that $\mathbf{q} \in \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_s(\bar{f}))$. Since all convex subsets of $\mathbb{R}^d$ have the same relative interior as their closure [Roc70, Theorem 6.3], the equality of the relative interiors stated before is implied by the relation $\mathsf{cl}(V \cap \mathsf{Pay}_s(\bar{f})) = \mathsf{cl}(V \cap \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_s(\bar{f})))$. To end the proof, we show this equality of closures. The inclusion $\mathsf{cl}(V \cap \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_s(\bar{f}))) \subseteq \mathsf{cl}(V \cap \mathsf{Pay}_s(\bar{f}))$ is direct.

For the other inclusion, it suffices to show that $V \cap \mathsf{Pay}_s(\bar{f}) \subseteq \mathsf{cl}(V \cap \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_s(\bar{f})))$. Let $\mathbf{p} \in V \cap \mathsf{Pay}_s(\bar{f})$. Assume, by contradiction, that $\mathbf{p} \notin \mathsf{cl}(V \cap \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_s(\bar{f})))$. By the hyperplane separation theorem (Theorem 2.3), there exists a linear form x^* over $\mathbb{R}^d$ such that for all $\mathbf{p}' \in V \cap \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_s(\bar{f}))$, we have $x^*(\mathbf{p}) > x^*(\mathbf{p}')$. Let $\mathcal{L}_{\mathbf{q}} \colon \mathbb{R}^d \to \mathbb{R}^{d'+1}$ such that for all $\mathbf{v} \in \mathbb{R}^d$, we have $\mathcal{L}_{\mathbf{q}}(\mathbf{v}) = (L_{\mathbf{q}}(\mathbf{v}), x^*(\mathbf{v}))$.

Let σ such that $\mathbf{p} = \mathbb{E}^{\sigma}_s(\bar{f})$. By Theorem 14.1, there exists a pure strategy τ such that $\mathbb{E}^{\tau}_s(\mathcal{L}_{\mathbf{q}} \circ \bar{f}) \geq_{\mathsf{lex}} \mathbb{E}^{\sigma}_s(\mathcal{L}_{\mathbf{q}} \circ \bar{f}) = \mathcal{L}_{\mathbf{q}}(\mathbf{p})$. We have $\mathbb{E}^{\tau}_s(\bar{f}) \in V$ because $L_{\mathbf{q}}(\mathbb{E}^{\tau}_s(\bar{f})) = \mathbb{E}^{\tau}_s(L_{\mathbf{q}} \circ \bar{f}) \geq_{\mathsf{lex}} L_{\mathbf{q}}(\mathbf{p}) = L_{\mathbf{q}}(\mathbf{q})$ and $L_{\mathbf{q}}(\mathbf{q})$ is lexicographically optimal in $L_{\mathbf{q}}(\mathsf{Pay}_s(\bar{f}))$. It follows that $\mathbb{E}^{\tau}_s(x^* \circ \bar{f}) = x^*(\mathbb{E}^{\tau}_s(\bar{f})) \geq x^*(\mathbf{p})$. This is a contradiction with x^* defining a strongly separating hyperplane. □

We now formulate two corollaries of Theorem 14.4. The first one relates to extreme points of payoffs sets.

Corollary 14.5. *Assume that $\bar{f}$ is universally integrable. For all $s \in S$, $\mathsf{extr}(\mathsf{Pay}_s(\bar{f})) \subseteq \mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$, i.e., all extreme points of $\mathsf{Pay}_s(\bar{f})$ are payoffs of pure strategies.*

Proof. By Theorem 14.4, we have $\mathsf{Pay}_s(\bar{f}) = \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$. All extreme points of $\mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$ must be in $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ by definition of the convex hull. □

Second, we establish that, for all $s \in S$, $\mathsf{Pay}_s(\bar{f})$ is closed whenever $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ is closed. We note that, in general, the convex hull of a closed set need not be closed. However, the convex hull of a compact subset of $\mathbb{R}^d$ is closed (Lemma 2.2). The set of expected payoffs of a universally integrable payoff function is bounded by the characterisation of universally integrable payoffs in Lemma 13.8. Therefore, it is thus compact, implying the claimed property.

Corollary 14.6. *Assume that $\bar{f}$ is universally integrable. For all $s \in S$, if $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ is closed, then $\mathsf{Pay}_s(\bar{f})$ is compact.*

Proof. Let $s \in S$ such that $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ is closed. By Lemma 13.8, $\bar{f}$ is universally integrable if and only if $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ is bounded. It follows that $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ is compact. Theorem 14.4 ensures that $\mathsf{Pay}_s(\bar{f}) = \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$, and thus $\mathsf{Pay}_s(\bar{f})$ is compact by Lemma 2.2. □

We close this section by showing that Theorem 14.4 does not generalise to universally unambiguously integrable payoffs. We build on Example 14.1, which illustrates that randomisation may be necessary to play lexicographically optimally for universally unambiguously integrable payoffs.

Example 14.4 (Example 14.1 continued)**.** We consider the MDP $\mathcal{M}$ depicted in Figure 14.2 and the payoff function $\bar{f} = (\mathbb{1}_{\mathsf{Reach}(\{t\})}, \mathsf{TRew}_w)$ where w is the weight function of Figure 14.2. In Example 14.1, we have shown that there exists a randomised strategy σ such that $\mathbb{E}_s^{\sigma}(\bar{f}) = (1, +\infty)$ and that

$\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}) = \{(0, +\infty)\} \cup \{(1, \ell) \mid \ell \in \mathbb{N}\}$.

We show that $(1, +\infty) \notin \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$. On the one hand, convex combinations of these vectors that give a non-zero coefficient to the vector $(0, +\infty)$ have a first component is not equal to 1. On the other hand, convex combinations that assign a zero coefficient to $(0, +\infty)$ have a finite second component. We obtain that $(1, +\infty) \notin \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$.

Although we cannot have a payoff of $(1, +\infty)$ with finite-support mixed strategies, we can approximate it with such strategies. We generalise this observation in the next section. ◁

14.4 Universally unambiguously integrable payoffs

We now relax the assumption that $\bar{f}$ is universally integrable from the previous section, and assume that $\bar{f}$ is universally unambiguously integrable. We formulate an approximate variant of Theorem 14.4: from a given state, any expected payoff of a strategy can be approached by convex combinations of expected payoffs of pure strategies (in the sense of limits in $\bar{\mathbb{R}}^d$).

We provide a proof sketch in Section 14.4.1. The theorem statement and proof are formalised in Section 14.4.2.

14.4.1 Proof overview

Fix $s \in S$ and a strategy σ. The goal is to show that all neighbourhoods of $\mathbb{E}_s^{\sigma}(\bar{f})$ intersect $\mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$. In other words, we must prove that for all $\varepsilon > 0$ and all $M \in \mathbb{R}$, there exist finitely many pure strategies $\tau_1, \ldots, \tau_n$ and convex combination coefficients $\alpha_1, \ldots, \alpha_n \in [0, 1]$ and, for all $j \in [\![1, d]\!]$:

- if $\mathbb{E}_s^{\sigma}(f_j) = +\infty$, then $\sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \geq M$,
- if $\mathbb{E}_s^{\sigma}(f_j) = -\infty$, then $\sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \leq -M$ and,
- otherwise, $\sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \geq \mathbb{E}_s^{\sigma}(f_j) - \varepsilon$.

We fix $\varepsilon > 0$ and $M \in \mathbb{R}$. The proof is based on the reformulation of $\mathbb{E}_s^{\sigma}(\bar{f})$ as an integral of pure expected payoffs from Lemma 13.4 and the manipulation of random variables. Even though Lemma 13.4 is not applicable to all payoffs, Lemma 13.9 implies that there exists a vector $\mathbf{v}$ such that the payoff $\bar{f} + \mathbf{v}$

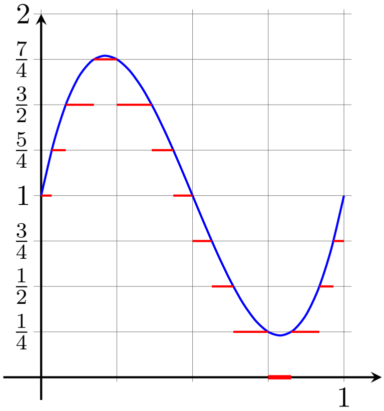


Figure 14.8: Illustration of the approach used to construct the approximation $\mathcal{Y}$ of $\mathcal{X}$ adapted to a function over $[0, 1]$. We round the blue function down to the closest multiple of $\frac{1}{4}$ to obtain the red function. This yields a linear combination of indicators that is $\frac{1}{4}$-close in all points to the function in blue.

satisfies the assumptions of Lemma 13.4. We can then recover the result for the original payoff using the linearity of the expectation.

We thus assume without loss of generality that Lemma 13.4 applies to all payoffs $f_1, \ldots, f_d$. We consider a mixed strategy μ that is outcome-equivalent to σ, whose existence is guaranteed by Kuhn's theorem. For all $j \in [\![1, d]\!]$, we let $X_j \colon \Sigma_{\mathsf{pure}}(\mathcal{M}) \to \bar{\mathbb{R}} \colon \tau \mapsto \mathbb{E}^{\tau}_{s}(f_j)$. We let $\mathcal{X} = (X_1, \ldots, X_d)$. By Lemma 13.4, we have $\mathbb{E}^{\sigma}_{s}(\bar{f}) = \int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} \mathcal{X}(\tau) \mathrm{d}\mu(\tau)$. We sketch the proof when the X_j are (μ-almost-surely) real-valued functions. We comment on the generalisation at the end of the sketch.

The broad idea is as follows. First, we approximate $\mathcal{X}$ with a multivariate random variable $\mathcal{Y} = (Y_1, \ldots, Y_d)$ over $\Sigma_{\mathsf{pure}}(\mathcal{M})$. We then approximate the integral of $\mathcal{Y}$ with a convex combination $\sum_{m=0}^{n} \alpha_m \mathcal{Y}(x_m) \in \mathsf{conv}(\mathsf{Im}(\mathcal{Y}))$. Finally, we derive the convex combination $\sum_{m=0}^{n} \alpha_m \mathcal{X}(x_m)$ from the previous one. The successive approximations above ensure that the last convex combination respects the claims of the theorem.

We now expand on the broad idea above. First, we construct $\mathcal{Y}$ such that $\mathcal{X} - \frac{\varepsilon}{3}\mathbf{1} \leq \mathcal{Y} \leq \mathcal{X} + \frac{\varepsilon}{3}\mathbf{1}$. It follows that, for all $j \in [\![1, d]\!]$, $\int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} Y_j \mathrm{d}\mu(\tau)$ is equal to $\mathbb{E}^{\sigma}_{s}(f_j)$ whenever $\mathbb{E}^{\sigma}_{s}(f_j) \in \{-\infty, +\infty\}$ and otherwise is $\frac{\varepsilon}{3}$-close. Intuitively, we construct $\mathcal{Y}$ as an infinite linear combination of indicators, following the rounding idea illustrated in Figure 14.8 (where the rounding precision depends on ε). Its integral is thus (informally) an infinite convex combination of images of $\mathcal{Y}$: there are sequences $(\beta_m)_{m\in\mathbb{N}}$ and $(x_m)_{m\in\mathbb{N}}$ respectively of coefficients and elements of $\Sigma_{\mathsf{pure}}(\mathcal{M})$ such that $\sum_{m=0}^{\infty}\beta_m = 1$ and $\int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})}\mathcal{Y}\mathrm{d}\mu(\tau) = \sum_{m\in\mathbb{N}}\beta_m\mathcal{Y}(x_m)$. We derive a sequence $(\mathbf{p}^{(n)})_{n\in\mathbb{N}}$ in $\mathsf{conv}(\mathsf{Im}(\mathcal{Y}(x_m)))$ that converges to $\int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})}\mathcal{Y}\mathrm{d}\mu(\tau)$ from this series: we let

$$\mathbf{p}^{(n)} = \sum_{m=0}^{n}\beta_m\mathcal{Y}(x_m) + \left(1-\sum_{m=0}^{n}\beta_m\right)\mathcal{Y}(x_0)$$

for all $n\in\mathbb{N}$.

Fix $n\in\mathbb{N}$ large enough such that, for all $j\in[\![1,d]\!]$, component j of $\mathbf{p}^{(n)}$ is $\frac{\varepsilon}{3}$-close to $\int_{\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})}Y_j\mathrm{d}\mu(\tau)$ if it is a real number or greater than $M+\varepsilon$ in absolute value otherwise. The convex combination $\mathbf{q} = \sum_{m=0}^{n}\beta_m\mathbb{E}_s^{\tau_{x_m}}(\bar{f}) + (1-\sum_{m=0}^{n}\beta_m)\mathbb{E}_s^{\tau_{x_0}}(\bar{f})$ is a satisfactory convex combination with respect to the claim of the theorem. Let $j\in[\![1,d]\!]$. If $\mathbb{E}_s^{\sigma}(f_j) = +\infty$, we obtain that $q_j \geq M$. Similarly, if $\mathbb{E}_s^{\sigma}(f_j) = -\infty$, we obtain that $q_j \leq -M$. Otherwise, we have that q_j is ε-close to $\mathbb{E}_s^{\sigma}(f_j)$.

We now briefly discuss the case where some X_j is not μ-almost-surely real-valued. For the sake of illustration, we assume that this only applies to $j=d$ and that $X_d \geq 0$. Therefore, we have $\mathbb{E}_s^{\sigma}(f_d) = +\infty$ and there exists some $\tau\in\Sigma_{\mathsf{pure}}(\mathcal{M})$ such that $\mathbb{E}_s^{\tau}(f_d) = +\infty$ and $\mathbb{E}_s^{\tau}(f_j)\in\mathbb{R}$ for all $j\in[\![1,d-1]\!]$. Let $\tau_1,\ldots,\tau_n$ and $\alpha_1,\ldots,\alpha_n$ given by the theorem for $(f_1,\ldots,f_{d-1})$, $\frac{\varepsilon}{2}$ and $M+\varepsilon$. Let $\mathbf{q} = \sum_{m=1}^{n}\alpha_m\mathbb{E}_s^{\tau_m}(\bar{f})$. By choosing $\eta\in\,]0,1[$ such that all components of $\mathbb{E}_s^{\tau}(\bar{f})$ other than the last have absolute value no more than $\frac{\varepsilon}{2}$ and the finite components of $(1-\eta)\mathbf{q}$ are $\frac{\varepsilon}{2}$-close to the corresponding components of $\mathbf{q}$, we obtain a suitable convex combination in the form of $\eta\mathbb{E}_s^{\tau}(\bar{f}) + (1-\eta)\mathbf{q}$.

14.4.2 Theorem statement and proof

We now formally state the main theorem of this section and prove it.

Theorem 14.7. *Assume that $\bar{f}$ is universally unambiguously integrable. Let $s\in S$. We have $\mathsf{cl}(\mathsf{Pay}_s(\bar{f})) = \mathsf{cl}(\mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$. In particular, for all strategies σ, all $\varepsilon > 0$ and all $M\in\mathbb{R}$, there exist finitely many pure strategies τ_1, …, τ_n and convex combination coefficients $\alpha_1,\ldots,\alpha_n\in[0,1]$ such that for* *all* $1 \leq j \leq d$*:*

- *if* $\mathbb{E}_s^\sigma(f_j) = +\infty$*, then* $\sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \geq M$*,*
- *if* $\mathbb{E}_s^\sigma(f_j) = -\infty$*, then* $\sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \leq -M$*, and,*
- *otherwise, if* $\mathbb{E}_s^\sigma(f_j) \in \mathbb{R}$*,* $\mathbb{E}_s^\sigma(f_j) - \varepsilon \leq \sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \leq \mathbb{E}_s^\sigma(f_j) + \varepsilon$*.*

Proof. The inclusion $\mathsf{cl}(\mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))) \subseteq \mathsf{cl}(\mathsf{Pay}_s(\bar{f}))$ follows from the convexity of $\mathsf{Pay}_s(\bar{f})$ (see Theorem 13.7). For other inclusion, it suffices to show that $\mathsf{Pay}_s(\bar{f}) \subseteq \mathsf{cl}(\mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$. This inclusion is equivalent to the last property of the theorem statement.

Let σ be a strategy, $\varepsilon > 0$ and $M \in \mathbb{R}$. Let μ be a mixed strategy that is outcome-equivalent to σ (whose existence follows from Kuhn's theorem). To prove the theorem, we reason on the μ-integral of random variables of $(\Sigma_{\mathsf{pure}}(\mathcal{M}), \mathcal{F}_{\Sigma_{\mathsf{pure}}(\mathcal{M})})$; see Chapter 2.4.3, Page 36, for the definition of the σ-algebra $\mathcal{F}_{\Sigma_{\mathsf{pure}}(\mathcal{M})}$. For any real or multivariate random variable Y over $\Sigma_{\mathsf{pure}}(\mathcal{M})$, we write $\int Y \mathrm{d}\mu$ for $\int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} Y(\tau)\mathrm{d}\mu(\tau)$ to lighten notation.

We make two assumptions without loss of generality. We defer the proof that these assumptions are without loss of generality to the end of the proof. First, we assume that for all $j \in [\![1, d]\!]$, either $\mathbb{E}_s^\tau(f_j) \geq 0$ for all strategies τ or $\mathbb{E}_s^\tau(f_j) \leq 0$ for all strategies τ. This assumption guarantees that $\mathbb{E}_s^\sigma(\bar{f}) = \int_{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})} \mathbb{E}_s^\tau(\bar{f})\mathrm{d}\mu(\tau)$ by Lemma 13.4. Second, we assume that for all $j \in [\![1, d]\!]$, the set $\{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M}) \mid \mathbb{E}_s^\tau(f_j) \notin \mathbb{R}\}$ has μ-measure zero.

For all $j \in [\![1, d]\!]$, we consider the random variable $X_j \colon \tau \mapsto \mathbb{E}_s^\tau(f_j)$ over $\Sigma_{\mathsf{pure}}(\mathcal{M})$. We let $\mathcal{X} = (X_1, \ldots, X_d)$. The first assumption above implies that for all $j \in [\![1, d]\!]$, X_j is a non-negative or non-positive random variable. The second assumption implies that $\mathcal{X}$ is almost-surely $\mathbb{R}^d$-valued. We thus interpret the random variable $\mathcal{X}$ as a function $\Sigma_{\mathsf{pure}}(\mathcal{M}) \to \mathbb{R}^d$.

We now prove the result under the assumptions above. The first step of the proof is to construct a random variable $\mathcal{Y} = (Y_1, \ldots, Y_d) \colon \Sigma_{\mathsf{pure}}(\mathcal{M}) \to \mathbb{R}^d$ which approximates $\mathcal{X}$. To ensure that the integral of $\mathcal{Y}$ is an infinite convex combination of images of $\mathcal{Y}$ (in the sense of the proof overview), we define $\mathcal{Y}$ as a function similar to a simple function. More precisely, we define $\mathcal{Y}$ as a series of indicators multiplied by coefficients.

We fix $k \in \mathbb{N}$ such that $\frac{1}{2^k} \leq \frac{\varepsilon}{3}$. We define $\mathcal{Y}$ component by component first, and introduce its series form later. Let $j \in [\![1, d]\!]$. We generalise the construction illustrated in Figure 14.8. For all $k \in \mathbb{N}$, we let Y_j be the random variable over $\Sigma_{\mathsf{pure}}(\mathcal{M})$ such that

$$Y_j = \sum_{\ell=0}^{\infty} \frac{\ell}{2^k} \cdot \mathbb{1}_{\left[\frac{\ell}{2^k}, \frac{\ell+1}{2^k}\right[}(X_j)$$

if X_j is non-negative, i.e., we round X_j down to the closest multiple of $\frac{1}{2^k}$, and, otherwise,

$$Y_j = \sum_{\ell=0}^{\infty} \frac{-\ell}{2^k} \cdot \mathbb{1}_{\left]\frac{-\ell-1}{2^k}, \frac{-\ell}{2^k}\right]}(X_j),$$

i.e., we round X_j up to the closest multiple of $\frac{1}{2^k}$. We have $X_j - \frac{1}{2^k} \leq Y_j \leq X_j + \frac{1}{2^k}$. This implies that $X_j - \frac{\varepsilon}{3} \leq Y_j^{(k)} \leq X_j + \frac{\varepsilon}{3}$. In particular, Y_j is integrable if and only if X_j is and, and, if both are integrable:

$$\mathbb{E}_s^{\sigma}(f_j) - \frac{\varepsilon}{3} \leq \int Y_j \mathrm{d}\mu \leq \mathbb{E}_s^{\sigma}(f_j) + \frac{\varepsilon}{3}. \tag{14.3}$$

We also have $\mathbb{E}_s^{\sigma}(f_j) = +\infty$ (resp. $-\infty$) if and only if $\int Y_j \mathrm{d}\mu = +\infty$ (resp. $-\infty$).

Now that we have shown that $\mathcal{Y}$ approximates $\mathcal{X}$, we prove that the integral of $\mathcal{Y}$ can be written as an infinite convex combination of elements of $\mathsf{Im}(\mathcal{Y})$. To this end, we rewrite $\mathcal{Y}$ in the series form mentioned above.

Let $j \in [\![1, d]\!]$. If $X_j \geq 0$, we define, for all $\ell \in \mathbb{N}$, $I_j(\ell) = \left[\frac{\ell}{2^k}, \frac{\ell+1}{2^k}\right[$ and $v_j(\ell) = \frac{\ell}{2^k}$. Otherwise, if $X_j \geq 0$ does not hold, we define, for all $\ell \in \mathbb{N}$, $I_j(\ell) = \left]\frac{-\ell-1}{2^k}, \frac{-\ell}{2^k}\right]$ and $v_j(\ell) = \frac{-\ell}{2^k}$. We fix an enumeration $(\bar{\ell}^{(m)})_{m \in \mathbb{N}}$ of $\mathbb{N}^d$. For all $m \in \mathbb{N}$, we let $\bar{\ell}^{(m)} = (\ell_1^{(m)}, \ldots, \ell_d^{(m)}) \in \mathbb{N}^d$. We define $B_m = \prod_{j=1}^{d} I_j(\ell_j^{(m)})$ and $\mathbf{v}_m = (v_j(\ell_j^{(m)}))_{j \in [\![1,d]\!]}$. We can rewrite $\mathcal{Y}$ as follows, using this notation:

$$\mathcal{Y} = \sum_{m \in \mathbb{N}} \mathbf{v}_m \mathbb{1}_{B_m}(\mathcal{X}).$$

Through this, we obtain that the integral of $\mathcal{Y}$ is an infinite convex combination: the monotone convergence theorem ensures that

$$\int \mathcal{Y} \mathrm{d}\mu = \sum_{m \in \mathbb{N}} \mathbf{v}_m \mu\left(\mathcal{X}^{-1}(B_m)\right).$$

Next, to determine the coefficients and pure strategies we seek, we define a sequence in $\mathsf{conv}(\mathsf{Im}(\mathcal{Y}))$ converging to $\int \mathcal{Y} \mathrm{d}\mu$. For all $m \in \mathbb{N}$, we let $\beta_m = \mu\left(\mathcal{X}^{-1}(B_m)\right)$ and let $\tau_m \in \Sigma_{\mathsf{pure}}(\mathcal{M})$ such that $\mathbf{v}_m = \mathcal{Y}(\tau_m)$ if $\beta_m \neq 0$, and τ_m is left arbitrary otherwise. We consider the sequence $(\mathbf{p}^{(n)})_{n\in\mathbb{N}}$ defined by, for all $n \in \mathbb{N}$,

$$\mathbf{p}^{(n)} = \sum_{m=1}^{n} \beta_m \mathcal{Y}(\tau_m) + \left(1 - \sum_{m=1}^{n} \beta_m\right) \mathcal{Y}(\tau_0).$$

We have $\lim_{n\to\infty} \mathbf{p}^{(n)} = \int \mathcal{Y} \mathrm{d}\mu$ and for all $n \in \mathbb{N}$, $\mathbf{p}^{(n)} \in \mathsf{conv}(\mathsf{Im}(\mathcal{Y}))$. For all $n \in \mathbb{N}$, we let $\mathbf{p}^{(n)} = (p_1^{(n)}, \dots, p_d^{(n)})$.

We now fix n such that, for all $j \in [\![1, d]\!]$, $\int Y_j \mathrm{d}\mu \in \mathbb{R}$ implies that

$$-\frac{\varepsilon}{3} \leq p_j^{(n)} - \int Y_j \mathrm{d}\mu \leq \frac{\varepsilon}{3}, \tag{14.4}$$

$\int Y_j \mathrm{d}\mu = +\infty$ implies that $p_j^{(n)} \geq M + \varepsilon$ and $\int Y_j \mathrm{d}\mu = -\infty$ implies that $p_j^{(n)} \leq -M - \varepsilon$. We set $\alpha_0 = 1 - \sum_{m=1}^{n} \beta_m$ and for $m \in [\![1, n]\!]$, $\alpha_m = \beta_m$. We remark that $\mathbf{p}^{(n)} = \sum_{m=0}^{n} \alpha_m \mathcal{Y}(\tau_m)$. We show that the convex combination $\mathbf{q} = \sum_{m=0}^{n} \alpha_m \mathcal{X}(\tau_m) = \sum_{m=0}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(\bar{f})$ satisfies the claim of the theorem.

We write $\mathbf{q} = (q_1, \dots, q_d)$. It follows from the inequalities $Y_j - \frac{\varepsilon}{3} \leq X_j \leq Y_j + \frac{\varepsilon}{3}$ for all $j \in [\![1, d]\!]$ and the definitions of $\mathbf{q}$ and $\mathbf{p}^{(n)}$ that

$$\mathbf{p}^{(n)} - \frac{\varepsilon}{3}\mathbf{1} \leq \mathbf{q} \leq \mathbf{p}^{(n)} + \frac{\varepsilon}{3}\mathbf{1}. \tag{14.5}$$

Let $j \in [\![1, d]\!]$. First, assume that $\mathbb{E}_s^{\sigma}(f_j) = +\infty$ (i.e., $\int Y_j \mathrm{d}\mu = +\infty$). In this case, it follows from Equation (14.5) and $p_j^{(n)} \geq M + \varepsilon$ that

$$q_j = \sum_{m=0}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \geq p_j^{(n)} - \frac{\varepsilon}{3} \geq M + \frac{2\varepsilon}{3} \geq M.$$

The argument for the case $\mathbb{E}_s^{\sigma}(f_j) = -\infty$ follows an analogous reasoning and is omitted.

We now assume that $\mathbb{E}_s^{\sigma}(f_j) \in \mathbb{R}$ (i.e., $\int Y_j \mathrm{d}\mu \in \mathbb{R}$). By applying Equations (14.5), (14.4) and (14.3) in succession, we obtain that

$$\begin{aligned} q_j &\leq p_j^{(n)} + \frac{\varepsilon}{3} \\ &\leq \int Y_j \mathrm{d}\mu + \frac{2\varepsilon}{3} \\ &\leq \mathbb{E}_s^{\sigma}(f_j) + \varepsilon. \end{aligned}$$

A similar succession of inequalities (referring to the same equations), yields $q_j \geq \mathbb{E}_s^\sigma(f_j) - \varepsilon$. This ends the argument that $\mathbf{q}$ satisfies the conditions outlined in the statement of the theorem.

To end the proof, it remains to show that the assumptions made above are without loss of generality. We recall them first:

1. for all $j \in [\![1, d]\!]$, $\mathbb{E}_s^\tau(f_j) \geq 0$ for all $\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})$ or $\mathbb{E}_s^\tau(f_j) \leq 0$ for all $\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M})$;
2. for all $j \in [\![1, d]\!]$, $\mu(\{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M}) \mid X_j(\tau) \notin \mathbb{R}\}) = 0$.

In the above, we have shown that the claim of the theorem holds with Assumptions 1 and 2. In the following, we first show that the theorem with both Assumptions 1 and 2 implies the theorem with only Assumption 2. After this, we show that the theorem with Assumption 2 implies the theorem with neither additional assumption.

Assume that Assumption 2 holds and let us show that the claim of the theorem holds. To obtain the result for $\bar{f}$, we derive a payoff $\bar{g} = (g_1, \ldots, g_d)$ from $\bar{f}$ such that $\bar{g}$ satisfies the conditions outlined in Assumptions 1 and 2, so that we can apply the variant of the theorem with Assumptions 1 and 2 to $\bar{g}$.

Let $j \in [\![1, d]\!]$. If $\inf_{\tau \in \Sigma(\mathcal{M})} \mathbb{E}_s^\tau(f_j) \in \mathbb{R}$, we let $g_j = f_j - \inf_{\tau \in \Sigma(\mathcal{M})} \mathbb{E}_s^\tau(f_j)$. We obtain that $\mathbb{E}_s^\tau(g_j) \geq 0$ for all strategies τ. Indeed, for all strategies τ, this follows by linearity of $\mathbb{E}$ if f_j is $\mathbb{P}_s^\tau$-integrable (which is equivalent to g_j being $\mathbb{P}_s^\tau$-integrable) and otherwise the non-negative parts of f_j and g_j are $|\inf_{\tau' \in \Sigma(\mathcal{M})} \mathbb{E}_s^{\tau'}(f_j)|$-close to one another, thus share their infinite integral and we obtain $\mathbb{E}_s^\tau(g_j) = \mathbb{E}_s^\tau(f_j) = +\infty \geq 0$. Otherwise, by Lemma 13.9, we have $\sup_{\tau \in \Sigma(\mathcal{M})} \mathbb{E}_s^\tau(f_j) \in \mathbb{R}$ and we let $g_j = f_j - \sup_{\tau \in \Sigma(\mathcal{M})} \mathbb{E}_s^\tau(f_j)$. By adapting the argument of the previous case, we obtain that for all strategies τ, we have $\mathbb{E}_s^\tau(g_j) \leq 0$ and the equivalence $\mathbb{E}_s^\tau(g_j) = -\infty$ if and only if $\mathbb{E}_s^\tau(f_j) = -\infty$.

Let $M' = M + \max_{j \in [\![1,d]\!]} |\gamma_j|$ where γ_j is the constant such that $g_j = f_j - \gamma_j$ for all $j \in [\![1, d]\!]$. We let $\tau_1, \ldots, \tau_n$ be the strategies and $\alpha_1, \ldots, \alpha_n \in [0, 1]$ be the coefficients given by the theorem with Assumptions 1 and 2 for $\bar{g}$, σ, ε and M'. Let $j \in [\![1, d]\!]$. For all $m \in [\![1, n]\!]$, f_j is $\mathbb{P}_s^{\tau_m}$-integrable (by the remaining additional assumption). We thus have $\sum_{m=1}^n \alpha_m \mathbb{E}^{\tau_m}(f_j) = \gamma_j + \sum_{m=1}^n \alpha_m \mathbb{E}^{\tau_m}(g_j)$. If $\mathbb{E}_s^\sigma(f_j) \in \mathbb{R}$, i.e., f_j is $\mathbb{P}_s^\sigma$-integrable, then so is g_j and we directly obtain $\mathbb{E}_s^\sigma(f_j) - \varepsilon \leq \sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \leq \mathbb{E}_s^\sigma(f_j) + \varepsilon$ from the similar inequality for g_j. Next, assume that $\mathbb{E}_s^\sigma(f_j) = +\infty$. By the above, this implies that $\mathbb{E}_s^\sigma(g_j) = +\infty$. We obtain (from the application of the theorem to g_j) that $\sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \geq \gamma_j + M' \geq M$. Finally, if $\mathbb{E}_s^\sigma(g_j) = -\infty$, we obtain that $\sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \leq \gamma_j - M' \leq -M$ in the same way as the previous case.

It remains to show that the theorem with Assumption 2 implies the theorem with no assumption. For convenience of notation, we assume that there exists $d' \in [\![1, d]\!]$ such that, for all $j \in [\![1, d']\!]$, $\mu(\{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M}) \mid \mathbb{E}_s^\tau(f_j) \notin \mathbb{R}\}) > 0$ and, for all $j \in [\![d'+1, d]\!]$ and $\mu(\{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M}) \mid \mathbb{E}_s^\tau(f_j) \in \mathbb{R}\}) = 1$.

For all $j \in [\![1, d']\!]$, $\mathbb{E}_s^\sigma(f_j)$ is infinite and there exists $\tau_j \in \Sigma_{\mathsf{pure}}(\mathcal{M})$ such that $\mathbb{E}_s^{\tau_j}(f_j) = \mathbb{E}_s^\sigma(f_j)$ (because $\mu(\{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M}) \mid \mathbb{E}_s^\tau(f_j) \notin \mathbb{R}\}) > 0$) and, for all $j \in [\![d'+1, d]\!]$, $\mathbb{E}_s^\tau(f_j) \in \mathbb{R}$ (because $\mu(\{\tau \in \Sigma_{\mathsf{pure}}(\mathcal{M}) \mid \mathbb{E}_s^\tau(f_{j'}) \in \mathbb{R}\}) = 1$). If $d' = d$, we conclude using the convex combination $\sum_{j=1}^d \frac{1}{d}\mathbb{E}_s^{\tau_j}(\bar{f}) = \mathbb{E}_s^\sigma(\bar{f})$. Therefore, we assume that $d' < d$. The payoff $(f_{d'+1}, \ldots, f_d)$ satisfies Assumption 2. We apply the theorem with Assumption 2 to $(f_{d'+1}, \ldots, f_d)$, strategy σ, $\frac{\varepsilon}{3}$ and $M + \varepsilon$ to obtain pure strategies $\tau_{d'+1}$, …, τ_n and convex combination coefficients $\beta_{d'+1}$, …, β_n that satisfy the implications given in the last property of the statement of the theorem.

We now define convex combination coefficients α_1, …, α_n to obtain the claim of the theorem for $\bar{f}$. Fix $\eta \in \,]0, 1[$ such that (i) for all $j \in [\![1, d']\!]$, the real components of $\eta\mathbb{E}_s^{\tau_j}(\bar{f})$ are no more than $\frac{\varepsilon}{3}$ in absolute value and (ii) $(1-\eta)\sum_{m=d'+1}^n \beta_m \mathbb{E}_s^{\tau_m}((f_{d'+1}, \ldots, f_d))$ is $\frac{\varepsilon}{3}$-close to $\sum_{m=d'+1}^n \beta_m \mathbb{E}_s^{\tau_m}((f_{d'+1}, \ldots, f_d))$. For all $1 \leq m \leq d'$, we set $\alpha_m = \frac{\eta}{d'}$ and for $m \in [\![d'+1, n]\!]$, we set $\alpha_m = (1-\eta)\beta_m$. It follows from $\eta \in \,]0, 1[$ that $1 - \eta > 0$ and $\sum_{m=1}^n \alpha_m = 1$.

We show that the pure strategies τ_1, …, τ_m and coefficients α_1, …, α_n are witnesses to the implications in the theorem statement. Let $j \in [\![1, d]\!]$. If $j \leq d'$, it follows from $\eta > 0$ that $\sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) = \eta\mathbb{E}_s^{\tau_j}(f_j) = \mathbb{E}_s^\sigma(f_j) \in \{-\infty, +\infty\}$, and the required inequality is trivially satisfied. We assume from here that $j \geq d' + 1$. It follows from properties (i) and (ii) above that $\sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j)$ is $\frac{2\cdot\varepsilon}{3}$-close to $\sum_{m=d'+1}^n \beta_m \mathbb{E}_s^{\tau_m}(f_j)$. Assume that $\mathbb{E}_s^\sigma(f_j) = +\infty$. Then, we have

$$\sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \geq \sum_{m=d'+1}^n \beta_m \mathbb{E}_s^{\tau_m}(f_j) - \frac{2\cdot\varepsilon}{3} \geq M + \varepsilon - \frac{2\cdot\varepsilon}{3} \geq M.$$

We obtain that $\mathbb{E}_s^\sigma(f_j) = -\infty$ implies $\sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \leq -M$ similarly.

Finally, assume that $\mathbb{E}_s^{\sigma}(f_j) \in \mathbb{R}$. We recall that $|\sum_{m=d'+1}^{n} \beta_m \mathbb{E}_s^{\tau_m}(f_j) - \mathbb{E}_s^{\sigma}(f_j)| \leq \frac{\varepsilon}{3}$ by definition of the strategies $\tau_{d'+1}$, …, τ_n and coefficients $\beta_{d'+1}$, …, β_n. We obtain, by the triangular inequality, that

$$\begin{aligned} \left|\sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(f_j) - \mathbb{E}_s^{\sigma}(f_j)\right| &\leq \left|\sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(f_j) - \sum_{m=d'+1}^{n} \beta_m \mathbb{E}_s^{\tau_m}(f_j)\right| \\ &+ \left|\sum_{m=d'+1}^{n} \beta_m \mathbb{E}_s^{\tau_m}(f_j) - \mathbb{E}_s^{\sigma}(f_j)\right| \\ &\leq \frac{2\cdot\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

This ends the proof that theorem with Assumption 2 implies the theorem without any additional assumption. □

14.5 Bounding the support of mixed strategies

Theorems 14.4 and 14.7 state that it suffices to mix finitely many pure strategies to respectively match or approximate the expected payoff of any strategy. We provide bounds on the number of pure strategies to mix in this section, in the same vein as Carathéodory's theorem for convex hulls (Theorem 2.1). We show that the expected payoff of a finite-support mixed strategy can be obtained exactly by mixing no more than $d+1$ strategies that are in its support and that a greater or equal payoff can be obtained by mixing no more than d of these strategies.

Let $s \in S$, $\sigma_1, \ldots, \sigma_n$ be pure strategies and α_1, …, $\alpha_n \in [0,1]$ be convex combination coefficients. Let $\mathbf{q} = (q_j)_{1\leq j\leq d} = \sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\sigma_m}(\bar{f})$. First, let us discuss the case when $\mathbf{q} \in \mathbb{R}^d$, i.e., when all considered pure strategies have a finite expected payoff on all dimensions. In this case, the first bound is direct by Carathéodory's theorem for convex hulls.

For the second bound, we observe that $\mathbf{q}$ is an element of the compact convex polytope $D = \mathsf{conv}(\{\mathbb{E}_s^{\sigma_m}(\bar{f}) \mid m \in [\![1,n]\!]\})$. In particular, $\mathbf{q}$ is dominated by an element $\mathbf{p}$ lying on a proper face of D (i.e., the intersection of D and a hyperplane). For instance, we can consider $\mathbf{q} + \beta \cdot \mathbf{1}$ for $\beta = \max\{\gamma \geq 0 \mid \mathbf{q} + \gamma\mathbf{1} \in D\}$. Because proper faces have dimension no more than $d-1$, Carathéodory's theorem implies that $\mathbf{p}$ is a convex combination of no more than d vectors of the form $\mathbb{E}_s^{\sigma_m}(\bar{f})$, taken among those lying on the considered proper face.

Assume now that $\mathbf{q} \notin \mathbb{R}^d$. In this case, Carathéodory's theorem does not apply directly. The idea is to reduce ourselves to the previous case. For each $j \in [\![1, d]\!]$, if q_j is infinite, there is $m \in [\![1, n]\!]$ such that $\mathbb{E}_s^{\sigma_m}(f_m) = q_j$. For each infinite component of $\mathbf{q}$, we fix $\alpha_m \mathbb{E}_s^{\sigma_m}(\bar{f})$ in the convex combination for one such m. We then obtain the sought bounds from the above; we reason on the real-valued components of $\mathbf{q}$ in the non-fixed part of the convex combination (after normalising its coefficients to sum to one).

We formalise our theorem and the previous proof sketch below.

Theorem 14.8. *Assume that $\bar{f}$ is universally unambiguously integrable. Let $s \in S$, $\sigma_1, \ldots, \sigma_n$ be pure strategies and $\alpha_1, \ldots, \alpha_n \in [0, 1]$ be convex combination coefficients. There exist convex combination coefficients $\beta_1, \ldots, \beta_n \in [0, 1]$ with $|\{1 \leq m \leq n \mid \beta_m \neq 0\}| \leq d + 1$ and convex combination coefficients $\gamma_1, \ldots, \gamma_n \in [0, 1]$ with $|\{1 \leq m \leq n \mid \gamma_m \neq 0\}| \leq d$ such that*

$$\sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\sigma_m}(\bar{f}) = \sum_{m=1}^{n} \beta_m \mathbb{E}_s^{\sigma_m}(\bar{f}) \leq \sum_{m=1}^{n} \gamma_m \mathbb{E}_s^{\sigma_m}(\bar{f}).$$

Proof. For all $m \in [\![1, n]\!]$, let $\mathbf{p}^{(m)} = (p_j^{(m)})_{j \in [\![1,d]\!]} = \mathbb{E}_s^{\sigma_m}(\bar{f})$ and let $\mathbf{q} = (q_j)_{1 \leq j \leq d} = \sum_{m=1}^{n} \alpha_m \mathbf{p}^{(m)}$. We assume that for all $m \in [\![1, n]\!]$, $\alpha_m \neq 0$.

First, we assume that $\mathbf{q} \in \mathbb{R}^d$. The existence of coefficients $\beta_1, \ldots, \beta_n$ obeying the required conditions is direct by Carathéodory's theorem for convex hulls (Theorem 2.1). We let $D = \mathsf{conv}(\{\mathbf{p}^{(m)} \mid m \in [\![1, n]\!]\})$, which is a compact set (see Lemma 2.2).

We define $\mathbf{p} = \mathbf{q} + \beta \cdot \mathbf{1}$ for $\beta = \sup\{\gamma \geq 0 \mid \mathbf{q} + \gamma \cdot \mathbf{1} \in D\}$. We remark that β is a real number. On the one hand, $0 \in \{\gamma \geq 0 \mid \mathbf{q} + \gamma \cdot \mathbf{1} \in D\}$, and thus $\beta \neq -\infty$. On the other hand, D is bounded, therefore $\beta \neq +\infty$. Furthermore, $\mathbf{p} \in D$. By convexity of D, $\mathbf{q} + \gamma \cdot \mathbf{1} \in D$ for all $0 \leq \gamma < \beta$. It follows that $\mathbf{p} \in \mathsf{cl}(D) = D$.

We have $\mathbf{q} \leq \mathbf{p}$. To end the case $\mathbf{q} \in \mathbb{R}^d$, we show that there exists a hyperplane H such that $\mathbf{p}$ is a convex combination of the vectors $\mathbf{p}^{(m)}$ that lie in H. This suffices, because Carathéodory's theorem then ensures that $\mathbf{p}$ is a convex combination of no more than d vectors among the $\mathbf{p}^{(m)}$.

If D is included in a hyperplane, there is nothing to show. We thus assume that D is not included in a hyperplane and obtain a hyperplane using the supporting hyperplane theorem (Theorem 2.4). By construction, $\mathbf{p} \notin \mathsf{int}(D) = \mathsf{ri}(D)$: for all $\gamma > 0$, $\mathbf{p} + \gamma \cdot \mathbf{1} \notin D$. Therefore, there exists a linear form $x^* \colon \mathbb{R}^d \to \mathbb{R}$ such that for all $\mathbf{v} \in D$, $x^*(\mathbf{v}) \leq x^*(\mathbf{p})$. We claim that $\mathbf{p}$ is a convex combination of the $\mathbf{p}^{(m)}$ that lie in the hyperplane $(x^*)^{-1}(x^*(\mathbf{p}))$. Write $\mathbf{p}$ as a convex combination $\sum_{m=1}^{n} \alpha'_m \mathbf{p}^{(m)}$. We observe that for all $m \in [\![1, n]\!]$, $\alpha'_m \neq 0$ implies $x^*(\mathbf{p}^{(m)}) = x^*(\mathbf{p})$, as otherwise we would obtain that $x^*(\mathbf{p}) < x^*(\mathbf{p})$. This ends the proof in the case $\mathbf{q} \in \mathbb{R}^d$.

We now assume that some components of $\mathbf{q}$ are infinite. For convenience of notation, we assume that there is $d' \in [\![1, d]\!]$ such that, for all $j \in [\![1, d']\!]$, $q_j \in \{-\infty, +\infty\}$ and, for all $j \in [\![d' + 1, d]\!]$, $q_j \in \mathbb{R}$. For all $j \in [\![1, d']\!]$, let $m_j \in [\![1, n]\!]$ such that $p_j^{(m_j)} = q_j$. If $d' = d$, we have $\mathbf{q} = \frac{1}{d} \sum_{j=1}^{d} \mathbf{p}^{(m_j)}$. We now assume that $d' < d$.

Let $I_\infty = \{m_j \mid j \in [\![1, d']\!]\}$, $I_\mathbb{R} = [\![1, n]\!] \setminus I_\infty$. We have $|I_\infty| \leq d' \leq d$. In particular, if $I_\mathbb{R}$ is empty, the sought result is direct. We assume that $I_\mathbb{R} \neq \emptyset$. Let $\alpha_\mathbb{R} = \sum_{m \in I_\mathbb{R}} \alpha_m > 0$, and let $\mathsf{proj}_{>d'} \colon (\bar{\mathbb{R}})^d \to (\bar{\mathbb{R}})^{d-d'}$ denote the projection of a vector onto its $d - d'$ last components. We apply the result for vectors in $\mathbb{R}^{d-d'}$ to $\mathsf{proj}_{>d'}(\sum_{m \in I_\mathbb{R}} \frac{\alpha_m}{\alpha_\mathbb{R}} \mathbf{p}^{(m)})$ with respect to the payoff function $(f_{d'+1}, \ldots, f_d)$. It follows that there exist convex combination coefficients $(\beta'_m)_{m \in I_\mathbb{R}}$ of which at most $d - d' + 1$ are positive and convex combination coefficients $(\gamma'_m)_{m \in I_\mathbb{R}}$ of which at most $d - d'$ are positive such that

$$\mathsf{proj}_{>d'}\left(\sum_{m \in I_\mathbb{R}} \frac{\alpha_m}{\alpha_\mathbb{R}} \mathbf{p}^{(m)}\right) = \sum_{m \in I_\mathbb{R}} \beta'_m \mathsf{proj}_{>d'}(\mathbf{p}^{(m)}) \leq \sum_{m \in I_\mathbb{R}} \gamma'_m \mathsf{proj}_{>d'}(\mathbf{p}^{(m)}).$$

We conclude by observing that

$$\begin{aligned}
\mathbf{q} &= \sum_{m \in I_\infty} \alpha_m \mathbf{p}^{(m)} + \alpha_\mathbb{R} \cdot \sum_{m \in I_\mathbb{R}} \frac{\alpha_m}{\alpha_\mathbb{R}} \mathbf{p}^{(m)} \\
&= \sum_{m \in I_\infty} \alpha_m \mathbf{p}^{(m)} + \sum_{m \in I_\mathbb{R}} \alpha_\mathbb{R} \beta'_m \mathbf{p}^{(m)} \\
&\leq \sum_{m \in I_\infty} \alpha_m \mathbf{p}^{(m)} + \sum_{m \in I_\mathbb{R}} \alpha_\mathbb{R} \gamma'_m \mathbf{p}^{(m)},
\end{aligned}$$

i.e., for all $m \in I_\infty$, we let $\beta_m = \gamma_m = \alpha_m$, and for all $m \in I_\mathbb{R}$, we let $\beta_m = \alpha_\mathbb{R} \cdot \beta'_m$ and $\gamma_m = \alpha_\mathbb{R} \cdot \gamma'_m$. □

We now highlight two corollaries of Theorem 14.8. First, we obtain bounds when dealing with universally integrable payoffs that follow from Theorem 14.4.

Corollary 14.9. *Assume that $\bar{f}$ is universally integrable. Let $s \in S$.*

- *For all $\mathbf{q} \in \mathsf{Pay}_s(\bar{f})$, $\mathbf{q}$ is a convex combination of at most $d+1$ elements of $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$.*
- *For all $\mathbf{q} \in \mathsf{Ach}_s(\bar{f})$, there exists a convex combination $\mathbf{p}$ of at most d elements of $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ such that $\mathbf{q} \leq \mathbf{p}$.*

Proof. By Theorem 14.4, $\mathsf{Pay}_s(\bar{f}) = \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$. Furthermore, we recall that $\mathbf{q} \in \mathsf{Ach}_s(\bar{f})$ if and only if there exists a strategy σ such that $\mathbf{q} \leq \mathbb{E}_s^\sigma(\bar{f})$. We obtain both claims of the corollary directly by Theorem 14.8. □

Example 14.4 implies that Corollary 14.9 does not directly extend to universally unambiguously integrable payoffs. Nonetheless, we can identify a class of vectors that can be achieved by mixing no more than d strategies: the vectors in $\mathsf{int}(\mathsf{Ach}_s(\bar{f}) \cap \mathbb{R}^d)$. These are vectors $\mathbf{q} \in \mathsf{Ach}_s(\bar{f}) \cap \mathbb{R}^d$ that can be improved in all dimensions simultaneously, i.e., such that $\mathbf{q} + \varepsilon\mathbf{1} \in \mathsf{Ach}_s(\bar{f})$ for some $\varepsilon > 0$. To prove our result, intuitively, we approximate, via Theorem 14.7, the payoff of a strategy achieving $\mathbf{q} + \varepsilon\mathbf{1}$ with a finite-support mixed strategy then invoke Theorem 14.8.

Corollary 14.10. *Assume that $\bar{f}$ is universally unambiguously integrable. Let $s \in S$. For $\mathbf{q} \in \mathsf{int}(\mathsf{Ach}_s(\bar{f}) \cap \mathbb{R}^d)$, there exists a convex combination $\mathbf{p}$ of at most d elements of $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ such that $\mathbf{q} \leq \mathbf{p}$.*

Proof. Let $\mathbf{q} = (q_j)_{1 \leq j \leq d} \in \mathsf{int}(\mathsf{Ach}_s(\bar{f}) \cap \mathbb{R}^d)$. By definition of the interior of a subset of $\mathbb{R}^d$, there exists $\varepsilon > 0$ such that $\mathbf{q} + \varepsilon\mathbf{1} \in \mathsf{Ach}_s(\bar{f})$. Therefore, there exists a strategy σ such that, for all $j \in [\![1, d]\!]$, $q_j < \mathbb{E}_s^\sigma(f_j)$. For all $j \in [\![1, d]\!]$, we have $\mathbb{E}_s^\sigma(f_j) \neq -\infty$ because $\mathbb{E}_s^\sigma(f_j) > q_j \in \mathbb{R}$. Let

$$\eta = \min\left(\{1\} \cup \{\mathbb{E}_s^\sigma(f_j) - q_j \mid \mathbb{E}_s^\sigma(f_j) \in \mathbb{R},\, j \in [\![1, d]\!]\}\right)$$

and

$$M = \max\left(\{1\} \cup \{q_j + 1 \mid \mathbb{E}_s^\sigma(f_j) = +\infty,\, j \in [\![1, d]\!]\}\right).$$

Theorem 14.7 implies that there exist pure strategies $\tau_1, \ldots, \tau_n$ and convex combination coefficients $\alpha_1, \ldots, \alpha_n \in [0, 1]$ such that if $\mathbb{E}_s^\sigma(f_j) = +\infty$, then $\sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \geq M$ and, otherwise, $\sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) - \mathbb{E}_s^\sigma(f_j) \geq -\eta$. By Theorem 14.8, we can assume that $n \leq d$. Let $\mathbf{p} = (p_j)_{1 \leq j \leq d} = \sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(\bar{f})$.

We show that $\mathbf{q} \leq \mathbf{p}$. Let $j \in [\![1, d]\!]$. First, assume that $\mathbb{E}_s^\sigma(f_j) = +\infty$. In this case, $M \geq q_j$, and we obtain $q_j \leq M \leq p_j$. Second, assume that $\mathbb{E}_s^\sigma(f_j) \in \mathbb{R}$. In that case, we have $\eta \leq \mathbb{E}_s^\sigma(f_j) - q_j$, and therefore $q_j \leq \mathbb{E}_s^\sigma(f_j) - \eta \leq p_j$. □

Chapter 15

Continuous payoffs in finite multi-objectives Markov decision processes

We study the topological properties of expected payoff sets when considering continuous payoff functions in finite multi-objective MDPs. The main result of this section applies to continuous payoffs whose square is universally integrable, or *universally square integrable* for short. We prove that the set of expected payoffs and the set of achievable vectors are closed for such payoff functions. This class of payoffs subsumes real-valued continuous payoffs (including discounted-sum payoffs) because such payoffs are bounded (since the set of plays is compact) and universally integrable continuous shortest-path costs (see Section 15.4).

We divide this chapter in four parts. In Section 15.1, we introduce a topology on the space of behavioural strategies such that the resulting topological space is metrisable and compact to formalise a notion of convergence of strategies. In Section 15.2, we formulate our result for continuous universally square integrable payoffs. Section 15.3 provides examples illustrating that the results of Section 15.2 do not hold for continuous payoffs that are not universally integrable. Finally, we show in Section 15.4 that the results of Section 15.2 apply to continuous universally integrable shortest-path costs.

For this whole chapter, we fix a *finite MDP* $\mathcal{M} = (S, A, \delta)$ and a d-

dimensional continuous payoff $\bar{f} = (f_j)_{j \in [\![1,d]\!]}$.

Contents

15.1 A topology on the space of strategies

In this section, we define a topology on the space of strategies to formalise the convergence of sequences of strategies.

First, we define a topology over $\mathcal{D}(A(s))$ for all $s \in S$. Let $s \in S$. The set $\mathcal{D}(A(s))$ is a compact subset of $\mathbb{R}^{A(s)}$. In the sequel, we consider the metric (induced by the Euclidean norm) $\mathsf{dist}_{\mathsf{proba}}(\mu, \mu') = \sqrt{\sum_{a \in A(s)} |\mu(a) - \mu'(a)|^2}$ for all μ, $\mu' \in \mathcal{D}(A(s))$ on distributions over actions.

We endow the set $\Sigma(\mathcal{M}) = \prod_{h \in \mathsf{Hist}(\mathcal{M})} \mathcal{D}(A(\mathsf{last}(h)))$ of all strategies with the product topology. We obtain that $\Sigma(\mathcal{M})$ is a compact metrisable topological space. We do not define a metric over strategies, as it is not necessary in the sequel. Instead, we recall that a sequence of strategies $(\sigma^{(n)})_{n \in \mathbb{N}}$ converges to a strategy σ if and only if, for all $h \in \mathsf{Hist}(\mathcal{M})$, $(\sigma^{(n)}(h))_{n \in \mathbb{N}}$ converges to $\sigma(h)$.

We now formulate a result intuitively stating that close strategies induce distributions that assign similar probabilities to cylinders of histories of bounded length. We first require the following technical lemma.

Lemma 15.1. *Let $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \in [0, 1]$. Then*

$$\left| \prod_{m=1}^{n} \alpha_m - \prod_{m=1}^{n} \beta_m \right| \leq \sum_{m=1}^{n} |\alpha_m - \beta_m| .$$

Proof. We proceed by induction. The claim trivially holds for $n = 1$. To lighten notation, we prove the case $n = 2$ separately first. We perform the general induction step below by building on this simpler case. We have:

$$\begin{aligned}|\alpha_1\alpha_2 - \beta_1\beta_2| &= |\frac{1}{2}(\alpha_1 - \beta_1)(\alpha_2 + \beta_2) + \frac{1}{2}(\alpha_1 + \beta_1)(\alpha_2 - \beta_2)| \\ &\leq |\alpha_1 - \beta_1|\frac{|\alpha_2 + \beta_2|}{2} + |\alpha_2 - \beta_2|\frac{|\alpha_1 + \beta_1|}{2} \\ &\leq |\alpha_1 - \beta_1| + |\alpha_2 - \beta_2|.\end{aligned}$$

The second line is obtained by the triangular inequality and the last line is obtained by using the fact that $\alpha_1, \alpha_2, \beta_1, \beta_2 \in [0, 1]$.

We now perform the general induction step. We assume that the result holds for some $n \in \mathbb{N}_{>0}$ and prove that it holds for $n + 1$. By applying the simpler case proven above and then the induction hypothesis, we obtain that

$$\begin{aligned}\left|\prod_{m=1}^{n+1} \alpha_m - \prod_{m=1}^{n+1} \beta_m\right| &\leq \left|\prod_{m=1}^{n} \alpha_m - \prod_{m=1}^{n} \beta_m\right| + |\alpha_{n+1} - \beta_{n+1}| \\ &\leq \sum_{m=1}^{n+1} |\alpha_m - \beta_m| .\end{aligned}$$

□

We now state the lemma regarding induced distributions.

Lemma 15.2. *Let σ, τ be two strategies, $k \in \mathbb{N}_{>0}$ and $\eta > 0$. Assume that, for all histories h that are at most k states long, $\mathsf{dist}_{\mathsf{proba}}(\sigma(h), \tau(h)) \leq \frac{\eta}{k}$. Then, for all histories h that are at most $k + 1$ states long and all $s \in S$, we have $|\mathbb{P}^{\sigma}_{s}(\mathsf{Cyl}\,(h)) - \mathbb{P}^{\tau}_{s}(\mathsf{Cyl}\,(h))| \leq \eta$.*

Proof. Let $h = s_0a_0s_1 \dots s_r$ with $r \leq k$. We only prove the claim for $s = s_0$. The other case is direct because, for $s \in S \setminus \{s_0\}$, we have $\mathbb{P}^{\sigma}_{s}(\mathsf{Cyl}\,(h)) = \mathbb{P}^{\tau}_{s}(\mathsf{Cyl}\,(h)) = 0$.

The proof follows from the following sequence of inequations.

$$\begin{aligned}
|\mathbb{P}^{\sigma}_{s_0}(\mathsf{Cyl}\,(h)) - \mathbb{P}^{\tau}_{s_0}(\mathsf{Cyl}\,(h))| &= \prod_{\ell=0}^{r-1} \delta(s_\ell, a_\ell)(s_{\ell+1}) \cdot \left| \prod_{\ell=0}^{r-1} \sigma(h_{\leq \ell}) - \prod_{\ell=0}^{r-1} \tau(h_{\leq \ell}) \right| \\
&\leq \left| \prod_{\ell=0}^{r-1} \sigma(h_{\leq \ell}) - \prod_{\ell=0}^{r-1} \tau(h_{\leq \ell}) \right| \\
&\leq \sum_{\ell=0}^{r-1} |\sigma(h_{\leq \ell}) - \tau(h_{\leq \ell})| \\
&\leq k \cdot \frac{\eta}{k} = \eta.
\end{aligned}$$

The first line is by definition of probability measures induced by the strategies from s_0. The second line uses the fact that transition probabilities are at most 1. The third line is obtained by Lemma 15.1. The last line follows from the assumption of the lemma and $r \leq k$. □

In the following section, we prove that if $\bar{f}$ is universally square integrable and $\Sigma \subseteq \Sigma(\mathcal{M})$ is closed, then $\mathsf{Pay}^{\Sigma}_{s}(\bar{f})$ is compact. We close this section by highlighting classes of strategies that are closed, i.e., classes of strategies for which we can apply the previous result.

First, we show that $\Sigma_{\mathsf{pure}}(\mathcal{M})$ is a closed subset of $\Sigma(\mathcal{M})$. The proof boils down to showing that the limit of a converging sequence of Dirac distributions is a Dirac distribution. This follows from such sequences being ultimately constant.

Lemma 15.3. *The set $\Sigma_{\mathsf{pure}}(\mathcal{M})$ is a closed subset of $\Sigma(\mathcal{M})$.*

Proof. Let $(\sigma^{(n)})_{n\in\mathbb{N}}$ be a sequence of pure strategies that converges to a strategy σ, i.e., for all $h \in \mathsf{Hist}(\mathcal{M})$, the sequence $(\sigma^{(n)}(h))_{n\in\mathbb{N}}$ converges to $\sigma(h)$. We must show that, for all $h \in \mathsf{Hist}(\mathcal{M})$, $\sigma(h)$ is a Dirac distribution.

Let $h \in \mathsf{Hist}(\mathcal{M})$. Because $(\sigma^{(n)}(h))_{n\in\mathbb{N}}$ is convergent, it is a Cauchy sequence. Thus, there exists some $n_0 \in \mathbb{N}$ such that for all $n, m \geq n_0$, $\mathsf{dist}_{\mathsf{proba}}(\sigma^{(n)}(h), \sigma^{(m)}(h)) < 1$. It follows from $(\sigma^{(n)}(h))_{n\in\mathbb{N}}$ being a sequence of a Dirac distributions and the definition of $\mathsf{dist}_{\mathsf{proba}}$ that the sequence $(\sigma^{(n)}(h))_{n\geq n_0}$

is constant. It follows that $\sigma(h) = \sigma^{(n_0)}(h)$, and thus $\sigma(h)$ is a Dirac distribution. □

We can also show that the set of finite-memory strategies with at most $B \in \mathbb{N}_{>0}$ memory states is closed. Furthermore, when taking the limit of a converging sequence of finite-memory strategies induced by Mealy machines with deterministic initialisation (resp. outputs, updates), the limit strategy is also induced by a Mealy machine with deterministic initialisation (resp. outputs, updates).

The proof reasons on the Mealy machines inducing the elements of the sequence. First, we extract a subsequence of Mealy machines that all have the same topology. Second, we extract a subsequence such that each component of the Mealy machine converges. To conclude, we prove that the limit Mealy machine induces the limit strategy of the original sequence.

The XYZ acronym in the following statement refers to the classification of finite-memory strategies studied in Part III (described in Chapter 3.2).

Lemma 15.4. *Let* $\mathrm{X}, \mathrm{Y}, \mathrm{Z} \in \{\mathrm{D}, \mathrm{R}\}$ *and* $B \in \mathbb{N}_{>0}$. *The set of finite-memory strategies induced by* XYZ *Mealy machines with at most* B *states is closed with respect to the metric* $\mathsf{dist}_{\mathsf{strat}}$.

Proof. Let σ be the limit of a sequence of finite-memory strategies $(\sigma_n)_{n\in\mathbb{N}}$ induced by XYZ Mealy machines with a most B states. For all $n \in \mathbb{N}$, we let $\mathfrak{M}_n = (M_n, \mu^n_{\mathsf{init}}, \mathsf{nxt}^n_{\mathfrak{M}}, \mathsf{up}^n_{\mathfrak{M}})$ be a Mealy machine with a most B states inducing σ_n. We assume that $M_n \subseteq \{1, \dots, B\}$.

We make several assumptions on the sequence of Mealy machines $(\mathfrak{M}_n)_{n\in\mathbb{N}}$ that can be enforced by working with a subsequence if necessary (using compactness of sets of distributions over finite sets). First, we assume that M_n is the same for all $n \in \mathbb{N}$ and let M denote this set. Second, we assume that $\mathsf{supp}(\mu^n_{\mathsf{init}})$ is constant for all $n \in \mathbb{N}$ and that $(\mu^n_{\mathsf{init}})_{n\in\mathbb{N}}$ is a convergent sequence. Third, we assume that, for all $m \in M$, $s \in S$ and $a \in A$, the supports $\mathsf{supp}(\mathsf{nxt}^n_{\mathfrak{M}}(m, s))$ and $\mathsf{supp}(\mathsf{up}_{\mathfrak{M}}(m, s, a))$ are the same for all $n \in \mathbb{N}$ and the sequences $(\mathsf{nxt}^n_{\mathfrak{M}}(m, s))_{n\in\mathbb{N}}$ and $(\mathsf{up}^n_{\mathfrak{M}}(m, s, a))_{n\in\mathbb{N}}$ are convergent.

We define a Mealy machine $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ as follows. We let

$\mu_{\mathsf{init}} = \lim_{n\to\infty} \mu_{\mathsf{init}}^n$ and define $\mathsf{nxt}_{\mathfrak{M}}$ and $\mathsf{up}_{\mathfrak{M}}$ such that, for all $m \in M$, $s \in S$ and $a \in A$, $\mathsf{nxt}_{\mathfrak{M}}(m, s) = \lim_{n\to\infty} \mathsf{nxt}_{\mathfrak{M}}^n(m, s)$ and $\mathsf{up}_{\mathfrak{M}}(m, s, a) = \lim_{n\to\infty} \mathsf{up}_{\mathfrak{M}}^n(m, s, a)$. We obtain that $\mathfrak{M}$ is an XYZ Mealy machine because sequences of Dirac distributions can only converge to Dirac distributions.

As a consequence of this definition of $\mathfrak{M}$ and the assumptions above, there are fewer initial states and transitions in $\mathfrak{M}$ than in the machines $\mathfrak{M}_n$. Formally, for all $m \in M$, $s \in S$, $a \in A$ and $n \in \mathbb{N}$, we have $\mathsf{supp}(\mu_{\mathsf{init}}) \subseteq \mathsf{supp}(\mu_{\mathsf{init}}^n)$, $\mathsf{supp}(\mathsf{nxt}_{\mathfrak{M}}(m, s)) \subseteq \mathsf{supp}(\mathsf{nxt}_{\mathfrak{M}}^n(m, s))$ and $\mathsf{supp}(\mathsf{up}_{\mathfrak{M}}(m, s, a)) \subseteq \mathsf{supp}(\mathsf{up}_{\mathfrak{M}}^n(m, s, a))$.

We let $\sigma^{\mathfrak{M}}$ be the partially-defined strategy induced by $\mathfrak{M}$, and prove that for all histories h in the domain of $\sigma^{\mathfrak{M}}$, $\sigma^{\mathfrak{M}}(h) = \sigma(h)$, i.e., that $\sigma^{\mathfrak{M}}(h) = \lim_{n\to\infty} \sigma_n(h)$.

For $w \in (SA)^*$, we reuse the notation μ_w introduced in Chapter 2.4.4 for the distribution over memory states when playing according to $\mathfrak{M}$ after w has taken place. For all $n \in \mathbb{N}$ and $w \in (SA)^*$, we denote by μ_w^n the similarly-defined distribution for $\mathfrak{M}_n$. We first show a technical claim before proceeding with the proof, to ensure that all objects manipulated below are well-defined. We show that for all $w \in (SA)^*$ and $n \in \mathbb{N}$, if w is consistent with $\mathfrak{M}$, then w is consistent with $\mathfrak{M}_n$.

Let $n \in \mathbb{N}$. We show this claim by induction on the number of MDP states in w (i.e., the length of w as a word over the alphabet SA). For the base case, i.e., the empty word ε, we note that ε is consistent with all Mealy machines. We now let $w = w'sa$ consistent with $\mathfrak{M}$ and assume by induction that the claim holds for w' (which is necessarily consistent with $\mathfrak{M}$, as otherwise, w could not be). The consistency of w with $\mathfrak{M}$ implies that there is some $m \in \mathsf{supp}(\mu_{w'})$ such that $a \in \mathsf{supp}(\mathsf{nxt}_{\mathfrak{M}}(m, s))$. By the induction hypothesis, we have $m \in \mathsf{supp}(\mu_{w'}^n(m, s))$ and w' consistent with $\mathfrak{M}_n$. It follows from these properties and the fact that $a \in \mathsf{supp}(\mathsf{nxt}_{\mathfrak{M}}(m, s)) \subseteq \mathsf{supp}(\mathsf{nxt}_{\mathfrak{M}}^n(m, s))$, that w is consistent with $\mathfrak{M}_n$. This ends the proof of the claim.

We now show that for all histories h consistent with $\mathfrak{M}$, we have $\sigma(h) = \sigma^{\mathfrak{M}}(h)$. For all histories $h = ws$ consistent with $\mathfrak{M}$, we recall that $\sigma^{\mathfrak{M}}(h)(a) = \sum_{m\in M} \mu_w(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m, s)(a)$ for all $a \in A$. Therefore, to conclude, we need only prove that for all $w \in (SA)^*$ consistent with $\mathfrak{M}$, $\lim_{n\to\infty} \mu_w^n = \mu_w$ because for all $m \in M$, $s \in S$, $\mathsf{nxt}_{\mathfrak{M}}(m, s) = \lim_{n\to\infty} \mathsf{nxt}_{\mathfrak{M}}^n(m, s)$.

Let w be consistent with $\mathfrak{M}$. For all $n \in \mathbb{N}$, w is consistent with $\mathfrak{M}_n$, i.e., μ^n_w is well-defined. The proof is by induction on the length of w. The base case $w = \varepsilon$ is direct via the assumption $\lim_{n\to\infty} \mu^n_{\mathsf{init}} = \mu_{\mathsf{init}}$. Next, assume that $w = w'sa$ and, by induction, that $\lim_{n\to\infty} \mu^n_{w'} = \mu_{w'}$. It suffices to show that for all $m \in \mathsf{supp}(\mu_w)$, $\lim_{n\to\infty} \mu^n_w(m) = \mu_w(m)$, i.e., that

$$\lim_{n\to\infty} \frac{\sum_{m'\in M} \mu^n_{w'}(m') \cdot \mathsf{nxt}^n_{\mathfrak{M}}(m', s)(a) \cdot \mathsf{up}^n_{\mathfrak{M}}(m', s, a)(m)}{\sum_{m'\in M} \mu^n_{w'}(m') \cdot \mathsf{nxt}^n_{\mathfrak{M}}(m', s)(a)} = \mu_w(m).$$

On the one hand, by the induction hypothesis, the numerator converges to $\sum_{m'\in M} \mu_{w'}(m') \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s)(a) \cdot \mathsf{up}_{\mathfrak{M}}(m', s, a)(m)$. Similarly, the denominator converges to $\sum_{m'\in M} \mu_{w'}(m') \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s)(a)$, which is non-zero by consistency of w with $\mathfrak{M}$. This implies that $\lim_{n\to\infty} \mu^n_w(m) = \mu_w(m)$. We have thus shown that $\mathfrak{M}$ induces σ. □

The set of all finite-memory strategies itself is not closed. In fact, it can be shown that the closure of the set of strategies induced by Mealy machines with randomised outputs (i.e., DRD strategies) is the set of all strategies. Intuitively, with such a finite-memory strategy, it is possible to imitate any strategy for a finite number of steps in the MDP. By considering increasingly greater number of steps, we can converge to any strategy in the limit.

Lemma 15.5. *The set of finite-memory strategies induced by Mealy machines with randomised outputs is dense in* $\Sigma(\mathcal{M})$.

Proof. Fix an arbitrary strategy σ. We use ε to denote the empty word. For each $n \in \mathbb{N}$, we define a Mealy machine $\mathfrak{M}_n = (M_n, \varepsilon, \mathsf{nxt}^n_{\mathfrak{M}}, \mathsf{up}^n_{\mathfrak{M}})$ where $M_n = \bigcup_{m=0}^{n} (SA)^m$, and for each $m \in M_n$, $s \in S$ and $a \in A$, we let $\mathsf{up}^n_{\mathfrak{M}}(m, s, a) = msa$ if $msa \in M_n$ and $\mathsf{up}^n_{\mathfrak{M}}(m, s, a) = m$ otherwise, and $\mathsf{nxt}^n_{\mathfrak{M}}(m, s) = \sigma(ms)$.

For all $n \in \mathbb{N}$, the strategy induced by $\mathfrak{M}_n$ agrees with σ for all histories that are at most $n + 1$ states long. It follows that the sequence of strategies induced by the Mealy machines $\mathfrak{M}_n$ converges to σ. □

15.2 Universally square integrable continuous payoffs

We prove that if $\bar{f}$ is universally square integrable, then for all $s \in S$, $\mathsf{Pay}_s(\bar{f})$ and $\mathsf{Ach}_s(\bar{f})$ are closed. Our proof relies on the following property: if $\bar{f}$ is universally square integrable, then for all sequences $(\sigma^{(n)})_{n\in\mathbb{N}}$ of strategies converging to a strategy σ, and for all $s \in S$, $(\mathbb{E}_s^{\sigma^{(n)}}(\bar{f}))_{n\in\mathbb{N}}$ converges to $\mathbb{E}_s^{\sigma}(\bar{f})$. It suffices to prove the convergence on each dimension to obtain this property. Therefore, we need only consider one-dimensional payoffs for now.

We split the proof into two parts: first, we consider the particular case of real-valued continuous payoffs and then generalise to universally square integrable payoffs. We remark that we may not assume that a universally integrable continuous payoff is real-valued without loss of generality: changing the payoff of plays that have an infinite payoff to a real number will violate the continuity property. Therefore, we do not generalise the property for real-valued continuous payoffs through such an assumption.

Let $f\colon \mathsf{Plays}(\mathcal{M}) \to \mathbb{R}$ be a continuous real-valued payoff. The proof in this case relies on the uniform continuity of f. Recall that f is uniformly continuous if for all $\varepsilon > 0$, there exists $\ell \in \mathbb{N}$ such that for all plays π, π', $\pi_{\leq\ell} = \pi'_{\leq\ell}$ implies that $|f(\pi) - f(\pi')| < \varepsilon$. In particular, for all $\varepsilon > 0$, f can be ε-approximated by a linear combination of indicator functions of cylinders of histories of a fixed length. This provides a means to ε-approximate $\mathbb{E}_s^{\tau}(f)$ for any strategy τ as a linear combination of probabilities of cylinders of histories of bounded length. Since Lemma 15.2 guarantees that the distributions over plays induced by strategies that are intuitively close assign similar probabilities to such cylinders, through the approximations above, we can obtain that $\mathbb{E}_s^{\sigma^{(n)}}(\bar{f})$ is 3ε-close $\mathbb{E}_s^{\sigma}(\bar{f})$ for all $s \in S$ for large values of n. We formalise this argument below.

We do not need to make assumptions regarding whether f is universally integrable. Continuous real-valued payoffs on finite MDPs are bounded (as the image of a compact set by a continuous function is compact), and thus the continuity of a real-valued payoff implies that it is universally integrable.

Theorem 15.6. *Let $s \in S$. Assume that $\mathcal{M}$ is finite, $\bar{f}\colon \mathsf{Plays}(\mathcal{M}) \to \mathbb{R}^d$ and that $\bar{f}$ is continuous. Then the function $\Sigma(\mathcal{M}) \to \mathbb{R}^d\colon \sigma \mapsto \mathbb{E}_s^{\sigma}(\bar{f})$ is continuous.*

In other words, for all sequences $(\sigma^{(n)})_{n\in\mathbb{N}}$ of strategies that converge to a strategy σ, $\lim_{n\to\infty} \mathbb{E}_s^{\sigma^{(n)}}(\bar{f}) = \mathbb{E}_s^{\sigma}(\bar{f})$.

Proof. It suffices to prove the theorem in the case $d = 1$ to obtain the general case. For this reason, we consider a one-dimensional real-valued continuous payoff f below. Let $(\sigma^{(n)})_{n\in\mathbb{N}}$ be a sequence of strategies converging to a strategy σ. We start with some notation. By continuity of f, f is bounded. We let $\|f\|_\infty = \sup_{\pi\in\mathsf{Plays}(\mathcal{M})} |f(\pi)|$. For any history $h = s_0a_0 \dots s_r$, we let $|h| = r$ denote the index of the last state of the history. We also fix, for all histories $h \in \mathsf{Hist}(\mathcal{M})$, a play $\pi(h) \in \mathsf{Cyl}\,(h)$ which is a continuation of h.

We must prove that $(\mathbb{E}_s^{\sigma^{(n)}}(f))_{n\in\mathbb{N}}$ converges to $\mathbb{E}_s^{\sigma}(f)$. Let $\varepsilon > 0$. We assume that $\|f\|_\infty > 0$, as otherwise the result is direct: if $\|f\|_\infty = 0$, then $\mathbb{E}_s^{\sigma^{(n)}}(f) = \mathbb{E}_s^{\sigma}(f) = 0$ for all $n \in \mathbb{N}$.

We start by constructing a simple function which $\frac{\varepsilon}{3}$-approximates f by exploiting the uniform continuity of f. By uniform continuity of f, there exists some $k \in \mathbb{N}_{>0}$ such that, for any two plays $\pi, \pi' \in \mathsf{Plays}(\mathcal{M})$, if $\pi_{\leq k} = \pi'_{\leq k}$, then $|f(\pi) - f(\pi')| \leq \frac{\varepsilon}{3}$. It follows that

$$\left| f - \sum_{|h|=k} f(\pi(h)) \cdot \mathbb{1}_{\mathsf{Cyl}(h)} \right| \leq \frac{\varepsilon}{3}.$$

Since f is universally integrable, it follows that, for all $\tau \in \Sigma(\mathcal{M})$,

$$\left| \mathbb{E}_s^{\tau}(f) - \sum_{|h|=k} f(\pi(h)) \cdot \mathbb{P}_s^{\tau}(\mathsf{Cyl}\,(h)) \right| \leq \frac{\varepsilon}{3}. \tag{15.1}$$

To end the argument, we now determine n_0 such that, for all $n \geq n_0$, we have

$$\left| \sum_{|h|=k} f(\pi(h)) \cdot \left(\mathbb{P}_s^{\sigma}(\mathsf{Cyl}\,(h)) - \mathbb{P}_s^{\sigma^{(n)}}(\mathsf{Cyl}\,(h)) \right) \right| \leq \frac{\varepsilon}{3}. \tag{15.2}$$

Let M denote the number of histories h such that $|h| = k$. Since $\lim_{n\to\infty} \sigma^{(n)} = \sigma$, there exists $n_0 \in \mathbb{N}$ such that for all histories h such that $|h| \leq k$ (i.e., with at most $k+1$ states), we have $\mathsf{dist}_{\mathsf{proba}}(\sigma^{(n)}(h), \sigma(h)) \leq \frac{\varepsilon}{3\cdot M\cdot\|f\|_\infty\cdot k}$ (here, we use the fact that there are finitely many such histories). Equation (15.2) follows from the triangular inequality and Lemma 15.2: we have, for all $n \geq n_0$,

$$\begin{aligned}
&\left|\sum_{|h|=k} f(\pi(h)) \cdot \left(\mathbb{P}_s^{\sigma}(\mathsf{Cyl}\,(h)) - \mathbb{P}_s^{\sigma^{(n)}}(\mathsf{Cyl}\,(h))\right)\right| \\
&\leq \|f\|_\infty \cdot \sum_{|h|=k} \left|\mathbb{P}_s^{\sigma}(\mathsf{Cyl}\,(h)) - \mathbb{P}_s^{\sigma^{(n)}}(\mathsf{Cyl}\,(h))\right| \\
&\leq \|f\|_\infty \cdot \sum_{|h|=k} \frac{\varepsilon}{3 \cdot M \cdot \|f\|_\infty} \\
&\leq \frac{\varepsilon}{3}.
\end{aligned}$$

Let $n \geq n_0$. We now show that $|\mathbb{E}_s^{\sigma}(f) - \mathbb{E}_s^{\sigma_n}(f)| \leq \varepsilon$. From the triangular inequality, Equation (15.1) and Equation (15.2), we obtain

$$\begin{aligned}
\left|\mathbb{E}_s^{\sigma}(f) - \mathbb{E}_s^{\sigma^{(n)}}(f)\right| \leq{}& \left|\mathbb{E}_s^{\sigma}(f) - \sum_{|h|=k} f(\pi(h)) \cdot \mathbb{P}_s^{\sigma}(\mathsf{Cyl}\,(h))\right| \\
&+ \left|\sum_{|h|=k} f(\pi(h)) \cdot \left(\mathbb{P}_s^{\sigma}(\mathsf{Cyl}\,(h)) - \mathbb{P}_s^{\sigma^{(n)}}(\mathsf{Cyl}\,(h))\right)\right| \\
&+ \left|\sum_{|h|=k} f(\pi(h)) \cdot \mathbb{P}_s^{\sigma^{(n)}}(\mathsf{Cyl}\,(h)) - \mathbb{E}_s^{\sigma^{(n)}}(f)\right| \\
\leq{}& \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.
\end{aligned}$$

We have thus shown that $(\mathbb{E}_s^{\sigma^{(n)}}(f))_{n\in\mathbb{N}}$ converges to $\mathbb{E}_s^{\sigma}(f)$. □

We now generalise Theorem 15.6 to universally square integrable payoffs. We note that the previous proof relies on the boundedness of continuous payoffs and their uniform continuity, and thus is not valid in this more general case.

Similarly to above, it suffices to consider one-dimensional non-negative payoffs; the general case can be recovered by writing payoffs on each dimension as the difference of their non-negative and non-positive parts. The non-negative and non-positive parts of a continuous payoff are also continuous (see Lemma A.1, Page 403). We let $f\colon \mathsf{Plays}(\mathcal{M}) \to \bar{\mathbb{R}}$ be a non-negative continuous square integrable payoff, $s \in S$, let $(\sigma^{(n)})_{n\in\mathbb{N}}$ be a sequence of strategies that converges to a strategy σ and $\varepsilon > 0$. Given $M \in \mathbb{R}$, we abbreviate $\{f(\pi) \geq M \mid \pi \in \mathsf{Plays}(\mathcal{M})\}$ by $\{f \geq M\}$, for all strategies τ, we write $\mathbb{P}^\tau_s(f \geq M)$ instead of $\mathbb{P}^\tau_s(\{f \geq M\})$ and we let $\min(f, M)$ denote the payoff $\pi \mapsto \min\{f(\pi), M\}$.

The goal is to show that $|\mathbb{E}^{\sigma^{(n)}}_s(f) - \mathbb{E}^\sigma_s(f))| \leq \varepsilon$ for all large enough values of n. We show that there exists a constant $M \geq 0$ (dependent on ε) such that the real-valued continuous payoff $\min(f, M)$ (whose continuity follows from Lemma A.1) satisfies $0 \leq \mathbb{E}^\tau_s(f) - \mathbb{E}^\tau_s(\min(f, M)) \leq \frac{\varepsilon}{3}$ for all strategies $\tau \in \Sigma(\mathcal{M})$. By the triangular inequality, $|\mathbb{E}^{\sigma^{(n)}}_s(f) - \mathbb{E}^\sigma_s(f))|$ is no more than

$$|\mathbb{E}^{\sigma^{(n)}}_s(f) - \mathbb{E}^{\sigma^{(n)}}_s(\min(f, M))| + |\mathbb{E}^{\sigma^{(n)}}_s(\min(f, M)) - \mathbb{E}^\sigma_s(\min(f, M))| \\ + |\mathbb{E}^\sigma_s(\min(f, M)) - \mathbb{E}^\sigma_s(f)|.$$

The first and last terms are no more than $\frac{\varepsilon}{3}$ by choice of M, and the second term is smaller than $\frac{\varepsilon}{3}$ for large values of n by Theorem 15.6.

Thus, the main hurdle of the proof is establishing the existence of a suitable M. The first step consists in showing that the probability of f being large can be made arbitrarily small, in the sense that for all $\eta > 0$, there exists $M(\eta)$ such that $\mathbb{P}^\tau_s(f \geq M(\eta)) \leq \eta$ for all strategies τ. The negation of this property would allow the existence of strategies with an arbitrarily large expected payoff from s, contradicting Lemma 13.8. It then follows from the Cauchy-Schwarz inequality (e.g., [Dur19, Thm. 1.5.2.]) that, for all strategies τ, $\mathbb{E}^\tau_s(f - \min(f, M)) \leq \sqrt{\mathbb{E}^\tau_s(f^2) \cdot \eta}$ because $f - \min(f, M) \leq f \cdot \mathbb{1}_{f \geq M}$. Because f^2 is universally integrable, Lemma 13.8 guarantees that $\sup_\tau \mathbb{E}^\tau_s(f^2)$ is finite, i.e., we obtain the desired inequality by choosing η small enough.

We provide the details of the above sketch in the following proof.

Theorem 15.7. *Let $s \in S$. Assume that $\mathcal{M}$ is finite, and that $\bar{f}$ is continuous and universally square integrable. Then the function $\Sigma(\mathcal{M}) \to \mathbb{R}^d \colon \sigma \mapsto \mathbb{E}^\sigma_s(\bar{f})$ is continuous. In other words, for all sequences $(\sigma^{(n)})_{n\in\mathbb{N}}$ of strategies that converge to a strategy σ, $\lim_{n\to\infty} \mathbb{E}^{\sigma^{(n)}}_s(\bar{f}) = \mathbb{E}^\sigma_s(\bar{f})$.*

Proof. It suffices to prove the theorem in the case $d = 1$ to obtain the general case. For this reason, we consider a one-dimensional universally square integrable continuous payoff f. Let $(\sigma^{(n)})_{n\in\mathbb{N}}$ be a sequence of strategies that converges to a strategy σ. We assume that f is non-negative. We show that this implies the general case at the end of the proof.

We show that $\lim_{n\to\infty}(\mathbb{E}_s^{\sigma^{(n)}}(\bar{f}))_{n\in\mathbb{N}} = \mathbb{E}_s^{\sigma}(\bar{f})$ by the standard definition of convergence. Let $\varepsilon > 0$. Our goal is to determine some $M \geq 0$ and some n_0 such that for all $n \geq n_0$, we have

$$|\mathbb{E}_s^{\sigma^{(n)}}(f) - \mathbb{E}_s^{\sigma^{(n)}}(\min(f, M))| + |\mathbb{E}_s^{\sigma^{(n)}}(\min(f, M)) - \mathbb{E}_s^{\sigma}(\min(f, M))| \\ + |\mathbb{E}_s^{\sigma}(\min(f, M)) - \mathbb{E}_s^{\sigma}(f)| \leq \varepsilon$$

This is sufficient: the sum highlighted above is greater or equal to $|\mathbb{E}_s^{\sigma^{(n)}}(f) - \mathbb{E}_s^{\sigma}(f)|$ by the triangular inequality. We bound each term of the above sum by $\frac{\varepsilon}{3}$ in the following.

The crux of the proof is determining a bound $M \geq 0$ such that for all strategies τ, $\mathbb{E}_s^{\tau}(f) - \mathbb{E}_s^{\tau}(\min(f, M)) < \frac{\varepsilon}{3}$. To establish this, we show the following property: for all $\eta > 0$, there exists a bound $M(\eta) \geq 0$ such that, for all strategies τ, $\mathbb{P}_s^{\tau}(f \geq M(\eta)) \leq \eta$. This last property is shown by contradiction. Assume that there exists some $\eta > 0$ such that for all $M \geq 0$, there exists a strategy τ_M such that $\mathbb{P}_s^{\tau_M}(f \geq M) > \eta$. Then, we obtain that for all $M \geq 0$, $\mathbb{E}_s^{\tau_M}(f) \geq \mathbb{E}_s^{\tau_M}(M \cdot \mathbb{1}_{\{f\geq M\}}) = M \cdot \mathbb{P}_s^{\tau_M}(f \geq M) \geq \eta \cdot M$. This contradicts the fact that f is universally integrable (Lemma 13.8).

Let $\alpha = 1 + \sup_{\tau\in\Sigma(\mathcal{M})} \mathbb{E}_s^{\tau}(f^2) > 0$. The value α is real by universal square integrability of f (Lemma 13.8). For the remainder of the proof, we fix $M \geq 0$ such that for all strategies τ, we have $\mathbb{P}_s^{\tau}(f \geq M) \leq \frac{\varepsilon^2}{9\cdot\alpha}$. We prove that M is the bound sought above. First, we observe that $f - \min(f, M) = (f - M)\cdot\mathbb{1}_{\{f\geq M\}} \leq f \cdot \mathbb{1}_{\{f\geq M\}}$. Because indicators are equal to their square, they are universally square integrable. By applying the Cauchy-Schwarz inequality, we obtain that, for all strategies τ,

$$\mathbb{E}_s^{\tau}(f - \min(f, M)) \leq \mathbb{E}_s^{\tau}(f \cdot \mathbb{1}_{\{f\geq M\}}) \leq \sqrt{\mathbb{E}_s^{\tau}(f^2) \cdot \mathbb{E}_s^{\tau}(\mathbb{1}_{\{f\geq M\}})} \leq \frac{\varepsilon}{3}.$$

To conclude (for non-negative payoffs), it remains to show that there exists n_0 such that for all $n \geq n_0$, we have $|\mathbb{E}_s^{\sigma^{(n)}}(\min(f, M)) - \mathbb{E}_s^{\sigma}(\min(f, M))| \leq \frac{\varepsilon}{3}$. To this end, we observe that the payoff $\min(f, M)$ is continuous (Lemma A.1). Theorem 15.6 then implies that a suitable n_0 exists ($\min(f, M)$ is real-valued).

We have shown that the theorem holds for non-negative continuous universally square integrable payoffs. We now generalise to arbitrary continuous universally square integrable payoffs. Let $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$ denote the non-negative and non-positive parts of f. These payoffs are continuous by Lemma A.1. We observe that $f^2 = (f^+)^2 + (f^-)^2$, and, in particular, $f^2 \geq (f^+)^2 + (f^-)^2$. It follows that $(f^+)^2$ and $(f^-)^2$ are universally integrable. We obtain the claim of the theorem from the above and the following sequence of equations:

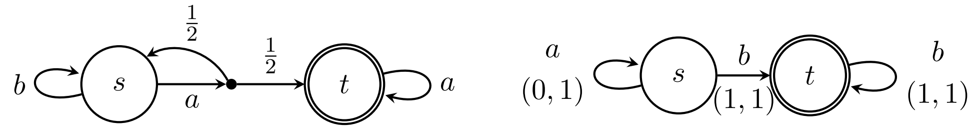
$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}_s^{\sigma^{(n)}}(f) &= \lim_{n\to\infty} \mathbb{E}_s^{\sigma^{(n)}}(f^+) - \lim_{n\to\infty} \mathbb{E}_s^{\sigma^{(n)}}(f^-)\\ &= \mathbb{E}_s^{\sigma}(f^+) - \mathbb{E}_s^{\sigma}(f^-)\\ &= \mathbb{E}_s^{\sigma}(f).\end{aligned}$$

□

We now prove that for multi-dimensional universally square integrable payoffs that are continuous, given a closed set of strategies, its set of expected payoffs and achievable set from a state are compact. For sets of expected payoffs, this follows from Theorem 15.7: since the image of compact set by a continuous function is compact, the result is direct. For achievable sets, it follows from the property that the downward closure of any compact set is compact (see Lemma A.10, Appendix A.8).

Theorem 15.8. *Let $s \in S$ and $\Sigma \subseteq \Sigma(\mathcal{M})$ be a closed set of strategies. Assume that $\mathcal{M}$ is finite and that $\bar{f}$ is a continuous universally square integrable payoff. Then $\mathsf{Pay}_s^{\Sigma}(\bar{f})$ and $\mathsf{Ach}_s^{\Sigma}(\bar{f})$ are compact subsets of $\mathbb{R}^d$ and $\bar{\mathbb{R}}^d$ respectively. In particular, $\mathsf{Pay}_s(\bar{f})$, $\mathsf{Ach}_s(\bar{f})$, $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ and $\mathsf{Ach}_s^{\mathsf{pure}}(\bar{f})$ are compact.*

Proof. Since Σ is closed and $\Sigma(\mathcal{M})$ is compact, we obtain that Σ is compact. It follows from Theorem 15.6 that $\mathsf{Pay}_s^{\Sigma}(\bar{f})$ is the image of Σ by a continuous function, and thus is compact. The claim regarding $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ follows from Lemma 15.3. The claims related to achievable sets follow from the property that for all compact $D \subseteq \bar{\mathbb{R}}^d$, $\mathsf{down}(D)$ is compact (Lemma A.10) and the fact that $\mathsf{Ach}_s^{\Sigma}(\bar{f}) = \mathsf{down}(\mathsf{Pay}_s^{\Sigma}(\bar{f}))$ by definition. □



(a) A weighted MDP with one randomised transition. Weights are omitted and are all 1.

(b) An MDP with two-dimensional weights.

Figure 15.1: MDPs for counter-examples to Theorems 15.7 (left) and 15.8 (right) without the universally square integrable assumption.

Theorem 15.8 provides a sufficient condition on payoffs which guarantees that we can dominate any expected payoff by a Pareto-optimal expected payoff. This property does not hold for all universally integrable payoffs in full generality, e.g., in the one-dimensional setting, optimal strategies need not exist.

We now turn our attention to finite-memory strategies. Theorem 15.8 and Lemma 15.4 imply that the set of expected payoffs of strategies induced by Mealy machines of a bounded size is closed when considering universally square integrable continuous payoffs. Furthermore, by density of the set of finite-memory strategies in the set of strategies of $\mathcal{M}$ (Lemma 15.5), it follows from Theorem 15.7that any expectation of a universally square integrable continuous payoff can be approximated with finite-memory strategies.

Corollary 15.9. *Let $s \in S$ and $\sigma \in \Sigma(\mathcal{M})$. Assume that $\mathcal{M}$ is finite and that $\bar{f}$ is a continuous universally square integrable payoff. Then for all $\varepsilon > 0$, there exists a finite-memory strategy τ such that $|\mathbb{E}_s^\sigma - \mathbb{E}_s^\tau| \leq \varepsilon$.*

15.3 Non-universally integrable continuous payoffs

In this section, we present counterexamples to Theorems 15.7 and 15.8 when the assumption of universal square integrability is relaxed. The MDPs used in these counterexamples are depicted in Figure 15.1.

For the first example, we provide an MDP with a shortest-path cost witnessing that Theorem 15.7 does not generalise to shortest-path costs that are not universally (square) integrable, even when only considering pure strategies and when $\mathsf{Pay}_s(\bar{f})$ is closed.

Example 15.1. We consider the MDP $\mathcal{M}$ depicted in Figure 15.1a, the constant weight function $w = 1$ and the shortest-path cost function $\mathsf{SPath}_w^{\{t\}}$. We provide a sequence of pure strategies $(\sigma^{(n)})_{n\in\mathbb{N}}$ converging to a pure strategy σ such that $\lim_{n\to\infty}\mathbb{E}_s^{\sigma^{(n)}}(\mathsf{SPath}_w^{\{t\}}) \neq \mathbb{E}_s^{\sigma}(\mathsf{SPath}_w^{\{t\}})$. For all $n \in \mathbb{N}$, we define $\sigma^{(n)}$ as the pure strategy such that, for all $h \in \mathsf{Hist}(\mathcal{M})$, $\sigma^{(n)}(h) = b$ if $\mathsf{last}(h) = s$ and there are at least n actions in h, and otherwise, $\sigma^{(n)}(h) = a$. Intuitively, when starting in s, $\sigma^{(n)}$ uses action a for the first n rounds of the play and then uses b for all subsequent rounds. The pure memoryless strategy σ assigning action a to all states is easily checked to be $\lim_{n\to\infty}\sigma^{(n)}$.

Let $n \in \mathbb{N}$. Let π_n denote the play $(sa)^n(sb)^\omega$. We have $\mathbb{P}_s^{\sigma^{(n)}}(\{\pi_n\}) = \frac{1}{2^n}$. Indeed, for all $r \in \mathbb{N}$, the definition of the probability distribution over plays under a strategy implies that

$$\mathbb{P}_s^{\sigma^{(n)}}(\mathsf{Cyl}\left((sa)^n(sb)^r s\right)) = \mathbb{P}_s^{\sigma^{(n)}}(\mathsf{Cyl}\left((sa)^n s\right)) = \frac{1}{2^n}.$$

It follows from $\mathsf{SPath}_w^{\{t\}}(\pi_n) = +\infty$ that $\mathbb{E}_s^{\sigma^{(n)}}(\mathsf{SPath}_w^{\{t\}}) = +\infty$. We conclude that $\lim_{n\to\infty}\mathbb{E}_s^{\sigma^{(n)}}(\mathsf{SPath}_w^{\{t\}}) = +\infty$.

We now show that $\mathbb{E}_s^{\sigma}(\mathsf{SPath}_w^{\{t\}}) \in \mathbb{R}$. First, we note that $\mathbb{P}_s^{\sigma}(\mathsf{Reach}(\{t\})) = 1$. Therefore, $\mathbb{E}_s^{\sigma}(\mathsf{SPath}_w^{\{t\}})$ is the unique solution of the equation $x = 1 + \frac{1}{2}x$ (see, e.g., [BK08]), i.e., $\mathbb{E}_s^{\sigma}(\mathsf{SPath}_w^{\{t\}}) = 2 \in \mathbb{R}$. We have shown that $\lim_{n\to\infty}\mathbb{E}_s^{\sigma^{(n)}}(\mathsf{SPath}_w^{\{t\}}) \neq \mathbb{E}_s^{\sigma}(\mathsf{SPath}_w^{\{t\}})$.

We now show that $\mathsf{Pay}_s(\mathsf{SPath}_w^{\{t\}})$ is closed. The memoryless strategy playing a is optimal when adopting a minimisation point of view. Furthermore, there exist strategies with arbitrarily large but finite expected payoffs from s. For instance, for all $n \in \mathbb{N}$, the strategy that plays b for the first n rounds in s and then only uses a after ensures a finite payoff greater than n. By convexity of $\mathsf{Pay}_s(\mathsf{SPath}_w^{\{t\}}) \cap \mathbb{R}$, it follows that $\mathsf{Pay}_s(\mathsf{SPath}_w^{\{t\}}) = [2, +\infty]$ and thus is closed. ◁

We now present a counter-example to Theorem 15.8 when the considered payoffs are not universally integrable: we provide a continuous payoff such that its set of expected payoffs and achievable sets from a given state are not closed.

Example 15.2. We consider the MDP $\mathcal{M}$ and the weight function $w = (w_1, w_2)$ depicted in Figure 15.1b. Let $\bar{f} = (f_1, f_2)$ be the two-dimensional payoff such that $f_1 = \mathsf{DSum}_{w_1}^{1/2}$ and $f_2 = \mathsf{SPath}_{w_2}^{\{t\}}$. To show that $\mathsf{Pay}_s(\bar{f})$ is not closed, we show that $(2, +\infty) \in \mathsf{cl}(\mathsf{Pay}_s(\bar{f})) \setminus \mathsf{Pay}_s(\bar{f})$.

First, we show that $(2, +\infty) \in \mathsf{cl}(\mathsf{Pay}_s(\bar{f}))$. We have $\bar{f}((sa)^\omega) = (0, \infty)$ and $\bar{f}(s(bt)^\omega) = (2, 1)$. By convexity of $\mathsf{Pay}_s(\bar{f})$, we obtain that for all $\eta \in]0, 1[$, $(2\eta, +\infty) \in \mathsf{Pay}_s(\bar{f})$. It follows that $(2, +\infty) \in \mathsf{cl}(\mathsf{Pay}_s(\bar{f}))$.

Next, we argue that $(2, +\infty) \notin \mathsf{Pay}_s(\bar{f})$. We observe that the only play π from s such that $f_1(\pi) = 2$ is the play $s(bt)^\omega$. All other plays π' starting in s are such that $f_1(\pi') < 2$. Indeed, for all $\ell \in \mathbb{N}$, we have $f_1((sa)^\ell s(bt)^\omega) = \sum_{\ell' \geq \ell} \frac{1}{2^{\ell'}} = \frac{1}{2^{\ell-1}} < 2$. It follows that, to obtain a payoff of 2 on the first dimension from s, we require a strategy whose only outcome is $s(bt)^\omega$. This implies that $(2, +\infty) \notin \mathsf{Pay}_s(\bar{f})$. We have shown that $\mathsf{Pay}_s(\bar{f})$ is not closed. $\lhd$

15.4 Applicability to universally integrable shortest-path costs

The goal of this section is to prove that all universally integrable shortest-path costs are universally square integrable in finite MDPs. In particular, this implies that Theorem 15.7 is applicable to universally integrable continuous shortest-path costs. We assume that $\mathcal{M}$ is finite throughout this entire section.

Let $T \subseteq S$ be a target set of states. We show that the payoff SPath_w^T is universally square integrable for all weight functions $w \colon S \times A \to \mathbb{R}$ if and only if, for all strategies τ and all initial states $s \in S$, we have $\mathbb{P}_s^\tau(\mathsf{Reach}(T)) = 1$. For any weight function w, if SPath_w^T is universally (square) integrable, then the set of plays with payoff $+\infty$ has zero $\mathbb{P}_s^\sigma$-probability for all strategies σ and $s \in S$, i.e., a target state is reached almost surely no matter the strategy and initial state.

To show the converse, it suffices to show that SPath_1^T, i.e., the shortest-path cost where all weights are 1, is universally square integrable, because all shortest-path costs in finite arenas are bounded in absolute value by a multiple of SPath_1^T (as there are finitely many weights in a finite MDP). Because all weights are 1, the expectation of $(\mathsf{SPath}_1^T)^2$ from $s \in S$ under a strategy σ can be written as a series over $r \in \mathbb{N}$ of r^2 multiplied by the $\mathbb{P}_r^\sigma$-probability of reaching T after exactly r actions are used. Convergence of this series is guaranteed by the fact that the sequence of probabilities of reaching the target after exactly r actions is in $\mathcal{O}(\alpha^r)$ for some $\alpha \in [0, 1[$. This asymptotic bound on these probabilities follows from the linear convergence of value iteration for reachability in MDPs of a specific form [HM18].

The formal proof below requires the notion of end-components of MDPs.

Definition 15.10. An *end-component* of $\mathcal{M}$ is a pair $\mathcal{E} = (E, A_{\mathcal{E}})$ such that

(i) $E \subseteq S$ is a non-empty set of states,

(ii) $A_{\mathcal{E}} \colon E \to 2^A \setminus \{\emptyset\}$ is a mapping assigning, to all states $s \in E$, a non-empty set of actions such that for all $a \in A_{\mathcal{E}}(s)$, $\mathsf{supp}(\delta(s, a)) \subseteq E$ and

(iii) from any s, $s' \in E$, there exists a history $s_0 a_0 s_1 \dots a_{r-1} s_r$ with $r \geq 1$ such that $s_0 = s$, $s_r = s'$ and $a_\ell \in A_{\mathcal{E}}(s_\ell)$ for all $0 \leq \ell \leq r - 1$.

Condition (iii) in the previous deviation requires that an end-component be strongly connected.

Lemma 15.11. *Let $T \subseteq S$ be a set of targets and $s_{\mathsf{init}} \in S$ be an initial state. Assume that $\mathcal{M}$ is finite. We have $\mathbb{P}^\sigma_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) = 1$ for all strategies σ if and only if, for all strategies σ and all weight functions $w \colon S \times A \to \mathbb{R}$, SPath^T_w is $\mathbb{P}^\sigma_{s_{\mathsf{init}}}$-square integrable. In particular, for any weight function w, if SPath^T_w is universally integrable, then SPath^T_w is universally square integrable.*

Proof. If all shortest-path costs with target T are $\mathbb{P}^\sigma_{s_{\mathsf{init}}}$-integrable for all strategies σ, then under all strategies and from all initial states, the set of plays with payoff $+\infty$ (which is independent of the weight function) must have probability zero, i.e., the set of targets is reached almost-surely. This implies that the equivalence stated in the lemma entails the claim made in particular. In the remainder of the proof, we assume that for all strategies σ, we have $\mathbb{P}^\sigma_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) = 1$ and show that for all weight functions w and strategies σ, SPath^T_w is $\mathbb{P}^\sigma_{s_{\mathsf{init}}}$-square integrable.

If $s_{\mathsf{init}} \in T$, then for all plays $\pi \in \mathsf{Cyl}(s_{\mathsf{init}})$, $\mathsf{SPath}^T_w(\pi) = 0$. Therefore, we directly obtain that SPath^T_w is $\mathbb{P}^\sigma_{s_{\mathsf{init}}}$-square integrable for all strategies σ. We thus assume that $s_{\mathsf{init}} \notin T$.

We assume without loss of generality that all states of $\mathcal{M}$ are reachable from s_{init} (i.e., for all states $s \in S$, there exists a history starting in s_{init} and ending in s). Furthermore, we also assume that there is a unique absorbing target state, i.e., that $T = \{t\}$ such that for all $a \in A(t)$, $\delta(t, a)(t) = 1$. We can always reduce to this case by considering the MDP $\mathcal{M}'$ obtained from $\mathcal{M}$ by merging all target states into a single absorbing state in which all actions are enabled. For any strategy σ and all weight functions w, SPath_w^T is $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}$-square integrable if and only if its counterpart in $\mathcal{M}'$ is $\mathbb{P}^{\sigma'}_{s_{\mathsf{init}}}$-square integrable, where $\sigma' \in \Sigma(\mathcal{M}')$ agrees with σ over $\mathsf{Hist}(\mathcal{M}) \cap \mathsf{Hist}(\mathcal{M}')$.

We now fix a strategy σ. First, we consider the constant weight function $w = 1$ and show that SPath_1^T is $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}$-square integrable. The general case follows from this one.

We preface the argument with some notation. For all $r \in \mathbb{N}$, we let $\mathsf{Reach}^{\leq r}(T) = \{s_0a_0s_1 \ldots \in \mathsf{Plays}(\mathcal{M}) \mid \exists \ell \leq r,\, s_\ell \in T\}$ denote the set of plays in which a target is reached in at most r transitions. For all $r \in \mathbb{N}_{>0}$, we let $\mathsf{Reach}^{=r}(T) = \mathsf{Reach}^{\leq r}(T) \setminus \mathsf{Reach}^{\leq r-1}(T)$ and $\mathsf{Reach}^{=0}(T) = \mathsf{Reach}^{\leq 0}(T)$. Given $\Omega \in \{\mathsf{Reach}(T)\} \cup \{\mathsf{Reach}^{\leq r}(T) \mid r \in \mathbb{N}\}$, we let, $\mathbb{P}^{\min}_{s_{\mathsf{init}}}(\Omega) = \min_{\tau \in \Sigma(\mathcal{M})} \mathbb{P}^{\tau}_{s_{\mathsf{init}}}(\Omega)$ denote the least probabilities that strategies can obtain for Ω from s. These minima are well-defined and $(\mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}^{\leq r}(T)))_{r\in\mathbb{N}}$ is an increasing sequence that converges to $\mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) = 1$ (see, e.g., [BK08]).

The equality $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) = 1$ implies that

$$\mathbb{E}^{\sigma}_{s_{\mathsf{init}}}((\mathsf{SPath}_1^T)^2) = \sum_{r\in\mathbb{N}} r^2 \cdot \mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\mathsf{Reach}^{=r}(T)).$$

It suffices to show that there exists constants $\alpha \in]0, 1[$ and $\beta \in [0, +\infty[$ such that for all $r \in \mathbb{N}$, $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\mathsf{Reach}^{=r}(T)) \leq \beta \cdot \alpha^r$ to obtain convergence. Indeed, the convergence of the above series is implied by that of the series $\sum_{r\in\mathbb{N}} \beta \cdot r^2 \cdot \alpha^r$, whose convergence is guaranteed by Cauchy's convergence test for non-negative series (because $\lim_{r\to\infty} \sqrt[r]{\beta \cdot r^2 \cdot \alpha^r} = \alpha < 1$).

We now seek the constants α and β satisfying the above condition. We obtain from $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) = \mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) = 1$ that, for all $r \in \mathbb{N}$, we have

$$\begin{aligned}\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\mathsf{Reach}^{=r}(T)) &\leq \mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\mathsf{Reach}(T) \setminus \mathsf{Reach}^{\leq r-1}(T)) \\ &= \mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) - \mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\mathsf{Reach}^{\leq r-1}(T)) \\ &\leq \mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) - \mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}^{\leq r-1}(T)).\end{aligned}$$

To bound the last term, we prove that $\mathcal{M}$ is min-reduced in the sense of [HM18]. An MDP is min-reduced if there are two states s^+ and s^- such that $T = \{s^+\}$, $s^- \neq s^+$ and all end-components have a singleton $\{s^+\}$ or $\{s^-\}$ as their set of states. In our case, the only end-components of $\mathcal{M}$ are of the form $(\{t\}, A')$ where $A' \subseteq A$ is non-empty. By contradiction, assume that there is an end-component $\mathcal{E} = (E, A_{\mathcal{E}})$ such that $E \neq \{t\}$. Then $t \notin E$ because t is absorbing and an end-component is strongly connected. Because E is be reachable from s_{init}, there exists a strategy reaching E with positive probability and then surely remaining in E from there (by only using actions authorised by $A_{\mathcal{E}}$). This contradicts the fact that for all strategies τ, $\mathbb{P}^{\tau}_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) = 1$.

It follows from [HM18, Proof of Thm. 2] that there exists $I \in \mathbb{N}_{>0}$ such that, for all $\ell \in \mathbb{N}$, $\mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) - \mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}^{\leq \ell I}(T)) \leq (1-\eta^I)^{\ell}$ where η is the smallest positive transition probability in $\mathcal{M}$. Let $r \in \mathbb{N}$, $\ell \in \mathbb{N}$ and $0 \leq p < I$ such that $r = \ell I + p$. We obtain, using the fact that the sequence $(\mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}^{\leq r'}(T)))_{r' \in \mathbb{N}}$ is non-decreasing, that

$$\begin{aligned}\mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) - \mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}^{\leq r}(T)) &\leq \mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}(T)) - \mathbb{P}^{\min}_{s_{\mathsf{init}}}(\mathsf{Reach}^{\leq \ell I}(T)) \\ &\leq (1-\eta^I)^{\ell} \\ &= \left(\sqrt[I]{1-\eta^I}\right)^{r-p} \\ &\leq \left(\sqrt[I]{1-\eta^I}\right)^{r-(I-1)}.\end{aligned}$$

We conclude that it suffices to set $\alpha = \sqrt[I]{1-\eta^I}$ and $\beta = \left(\sqrt[I]{1-\eta^I}\right)^{-I}$ to conclude the proof that SPath^T_1 is $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}$-square integrable.

We end the proof by establishing that for all weight functions w, the shortest-path cost SPath^T_w is $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}$-square integrable. Let $W = \max_{(s,a) \in S \times A} |w(s,a)|$. By the triangular inequality, we obtain that $|\mathsf{SPath}^T_w| \leq W \cdot \mathsf{SPath}^T_1$. We have thus bounded SPath^T_w in absolute value by a multiple of a $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}$-square integrable function, thus implying that SPath^T_w is also $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}$-square integrable. □

Part V:

Beyond Mealy machines: counter-based strategies in one-counter Markov decision processes

Chapter 16

Introduction

In this part, we detail the results presented in Chapter 3.4, originating from joint work with Michal Ajdarów, Petr Novotný and Mickaël Randour [AMNR25]. We study decision problems in one-counter Markov decision processes (Definition 2.49) for interval strategies, a class of memoryless strategies described by interval partitions of the set of counter values. We focus on the state-reachability (Definition 2.53) and selective termination (Definition 2.54) objectives, which respectively ask to visit a target regardless of counter value and to hit counter value zero in a target state.

We refer the reader to Chapter 3.4 for an extended presentation of the context. We divide this part into five chapters. We summarise their contents below, and comment on related work at the end of this chapter.

Interval strategies. Chapter 17 introduces interval strategies, presents some of their properties and formalises our three interval strategy decision problems. Recall that we consider two semantics for MDPs induced by OC-MDPs (see Definition 2.51): *bounded OC-MDPs*, in which we impose a counter upper bound such plays that reach it are interrupted, and *unbounded OC-MDPs*, in which no counter upper bound is imposed.

An *interval strategy* is a memoryless strategy of the MDP over configurations induced by an OC-MDP that admits a finite description based on interval partitions of the set of counter values. We consider two variants of interval strategies (Definition 17.2): *open-ended interval strategies* (OEIS), defined both in bounded and unbounded OC-MDPs and *cyclic interval strategies* (CIS), defined in unbounded OC-MDPs. On the one hand, an OEIS is a strategy such that, for all states, the decisions made in this state are the same for all sufficiently large counter values. On the other hand, a CIS is a strategy for which there exists a positive integer period such that, for all states, the decisions made in this state are identical given two counter values that differ by the period.

Strategies of either type can be described by some finite interval partition of all counter values for OEISs and of counter values up to the period for CISs, and by one memoryless strategy of the finite MDP underlying the OC-MDP per interval. We investigate the relationship between interval strategies and the corresponding strategies of the underlying MDP in Chapter 17.2. We obtain that OEISs in bounded OC-MDPs and CISs correspond to finite-memory strategies, but that the relevant Mealy machines may require as many memory states as the counter upper bound or the period respectively (Example 17.1). We also obtain that OEISs in unbounded OC-MDPs need not correspond to finite-memory strategies.

We then study how powerful interval strategies are with respect to our two objectives. In bounded OC-MDPs, the existence of optimal OEIS follows from the existence of pure memoryless optimal strategies for reachability in finite MDPs (e.g., [BK08]). For unbounded OC-MDPs, we show that interval strategies can be used to approximate the value (Lemma 17.5). We then illustrate that an optimal OEIS (resp. CIS) need not exist, even if an optimal CIS (resp. OEIS) does (Examples 17.2 and 17.3). This shows that interval strategies do not suffice to play optimally in general, although they can be used to play almost-optimally.

Finally, we formalise the three decision problems we are interested in. We first define the *interval strategy decision problem*, which asks whether a given interval strategy ensures an objective from an initial configuration with a probability greater than a given threshold (Definition 17.6). We then define two realisability problems for structurally-constrained interval strategies. On the one hand, the *fixed-interval OEIS (resp. CIS) realisability problem* asks whether there exists an OEIS (resp. a CIS) based on a given interval partition that ensures an objective from an initial configuration with a probability greater than a given threshold (Definitions 17.7 and 17.8). On the other hand, the *parameterised*

OEIS (resp. CIS) realisability problem asks whether there exists an interval partition whose size and number of intervals is bounded from above by given parameters and an OEIS (resp. a CIS) based on this partition that ensures an objective from an initial configuration with a probability greater than a given threshold (Definitions 17.9 and 17.10). For these two realisability problems, we show that randomised strategies may yield better maximum probabilities than pure strategies (Example 17.4), and thus consider two variants: one where we ask whether a suitable *pure strategy* exists and another for *randomised strategies*.

Compressed Markov chains. Markov chains induced (Definition 2.14) by interval strategies are large in bounded OC-MDPs and are infinite in unbounded OC-MDPs. We analyse such induced Markov chains by means of a *compression approach*: we build a *compressed Markov chain* with fewer configurations and adjusted transitions. We define compressed Markov chains in Chapter 18, then prove some key properties to solve our interval strategy decision problems with them.

The compression of an induced Markov chain preserves selective termination probabilities (Theorem 18.4). While this is not the case for state-reachability probabilities, we can apply a modification to the OC-MDP such that state-reachability probabilities are preserved for a given target (Theorem 18.5).

We obtain that verification can be reduced to the analysis of compressed Markov chains. However, we cannot directly analyse compressed Markov chains for two reasons. First, transition probabilities in compressed Markov chains can be either irrational or can require large representations (Example 18.1). Second, compressed Markov chains are obtained by removing configurations locally for each interval over which the strategy behaves uniformly, and thus the compression of the Markov chain induced by a CIS is infinite.

For the first issue, we show that we can represent transition probabilities as the least solutions of *polynomial size quadratic equation systems* (Theorems 18.6 and 18.9). Furthermore, we show that our systems for transitions over bounded intervals, we can refine the system to have a unique solution in polynomial time, and the refined system can be solved in polynomial time in the Blum-Shub-Smale (BSS) model (see Chapter 2.9) of computation (Theorem 18.11).

For the second issue, we exploit the periodic nature of CISs to show that compressed Markov chains for CISs are induced by one-counter Markov chains (Chapter 18.5).

Verification algorithms. We present our verification algorithms in Chapter 19. We provide a specific approach for OEIS in bounded OC-MDPs, and two similar approaches to deal with OEISs (that works in both the unbounded and bounded settings) and CISs.

The key to solving the verification problem for OEISs in bounded OC-MDPs is the ability to compute the transition probabilities of the compressed Markov chain in polynomial time in the BSS model. Therefore, in the BSS model, our problem boils down to computing reachability probabilities in finite Markov chains, which can be done in polynomial time by solving a linear system of equations (see Appendix A.2.1). We obtain a $\mathsf{P}^{\mathsf{PosSLP}}$ complexity upper bound (by the results of [ABKM09]) through this approach (Theorem 19.1).

For OEISs and CISs, we reduce to checking whether a universal formula holds in the (first-order) theory of the reals. We briefly summarise the main idea for OEISs.

We build formulae from our characterisation of transition probabilities in compressed Markov chains and the characterisation of reachability probabilities in Markov chains as the least solution of a linear system. The resulting formulae are such that any vector satisfying their conjunction is an over-estimation of the intended values (Lemma 19.2). To answer the verification problem, it suffices to check that all over-estimations of the probability of interest are greater or equal to the threshold (Theorem 19.3). We thus obtain a $\mathsf{co\text{-}ETR}$ upper complexity bound (Theorem 19.5).

We adapt this approach to CISs as follows. The main difference is that we apply the compression approach twice. First, we compress the Markov chain induced by the CIS and derive its one-counter Markov chain description. We then compress the Markov chain induced by this one-counter Markov chain with respect to a finite interval partition (i.e., similarly to an OEIS). We then define an adaptation of the OEIS verification formula that uses the characterisation of transition probabilities in a compression twice (Theorem 19.7). This approach also yields a $\mathsf{co\text{-}ETR}$ upper bound for the CIS verification problem

(Theorem 19.9).

Realisability algorithms. Chapter 20 presents our interval strategy realisability algorithms. We use different approaches depending on whether the goal is to check the existence of a well-performing pure or randomised strategy.

For *pure interval strategy realisability*, we propose algorithms based on a guess-and-check approach. For the fixed-interval problem, we guess actions for each state-interval pair, and, for the parameterised problem, we guess both an appropriate interval partitions and actions for each state-interval pair. We then check the resulting strategy by verifying it. Through this approach, we obtain a $\mathsf{NP}^{\mathsf{PosSLP}}$ upper bound for OEISs in bounded OC-MDPs (Theorem 20.2) and a $\mathsf{NP}^{\mathsf{ETR}} = \mathsf{NP}^{\mathsf{co\text{-}ETR}}$ upper bound for OEISs in unbounded OC-MDPs (Theorem 20.5) and for CISs (Theorem 20.8).

For *randomised interval strategy realisability*, we build on the formulae for verification by treating strategy probabilities in these formulae as variables instead of parameters. By prefacing the verification formulae with existential quantifiers for strategy probabilities (see the formulae of Theorems 20.6 and 20.9), we obtain a PSPACE upper bound for fixed-interval OEIS and CIS realisability (Theorems 20.7 and 20.10). We obtain the same upper bounds for the parameterised problem, as it suffices to non-deterministically guess a partition compatible with parameters and then run a fixed-interval algorithm.

For our randomised OEIS realisability problems in bounded OC-MDPs, we show that we can use non-determinism to reduce to checking the validity of an existential formula (Theorem 20.3). This avoids the quantifier alternation of the above approach, and yields an $\mathsf{NP}^{\mathsf{ETR}}$ upper bound (Theorem 20.4).

Lower bounds. We close this part with lower complexity bounds in Chapter 21. We first present *square-root-sum* hardness results for all of our problems. The square-root-sum problem asks, given some natural numbers and a natural threshold, whether the sum of the square roots of the first numbers is greater or equal to the threshold. In unbounded OC-MDPs, we obtain the square-root-sum hardness of our three interval strategy problems from a square-root-sum hardness result for one-counter Markov chains [EWY10]. In bounded OC-MDPs, we adapt the reduction of [EWY10] so that, intuitively, the reduction remains valid when imposing a large enough counter upper bound (Lemma 21.6). This yields the square-root-sum hardness of our three problems in bounded OC-MDPs (Theorem 21.7).

Finally, we show the NP-hardness of the realisability problem for (pure and randomised) *counter-oblivious strategies* for selective termination, a special case of our two interval strategy realisability problems. A counter-oblivious strategy is a memoryless strategy that disregards counter values, i.e., a one-interval OEIS and a CIS of period one. We provide a reduction from the NP-complete Hamiltonian cycle problem, which asks whether there exists a simple cycle visiting all vertices in a finite directed graph. Intuitively, we build an OC-MDP from the graph by duplicating an initial vertex in such a way that there exists a Hamiltonian cycle if and only if it is possible, in the OC-MDP, to reach the duplicate initial vertex from the original one in as many steps as there are vertices. Counter-oblivious strategies cannot almost-surely close the cycle in the required amount of steps unless there is a Hamiltonian cycle in the graph.

Related work. In addition to the main references cited previously, we mention some (non-exhaustive) related work. The closest is [BBN$^+$20]: interval strategies are similar to counter selector strategies, studied in *consumption* MDPs. These differ from OC-MDPs in key aspects: all transitions consume resources (i.e., bear negative weights), and recharges can only be done in special reload states, where it is considered as an atomic action up to a given capacity. Consumption and counter-based (or energy) models have different behaviors (e.g., [BCKN12]). The authors of [BBN$^+$20] also study incomparable objectives: almost-sure Büchi objectives.

Another related model is *solvency games* [BKSV08], which are stateless variants of OC-MDPs with binary counter updates. The goal in a solvency game is to never go bankrupt (i.e., the complement objective of termination). Berger et al. identify a natural class of *rich man's strategies,* which correspond to our OEISs. While [BKSV08] shows that optimal rich man's strategies do not always exist, if they do, their existence substantially simplifies the model analysis.

Finally, we remark that restricting oneself to subclasses of strategies that prove to be of practical interest is a common endeavor in synthesis: e.g., strategies represented by decision trees [BCC+15, BCKT18, AJK+21, JKW23], pure strategies [Gim07, DKQR20, BORV23] and finite-memory strategies [CRR14, BRV22, BORV23].

Chapter 17

Interval strategies

This chapter introduces *interval strategies*, highlights some of their properties and formalises our *interval strategy decision problems*. Definitions are given in Section 17.1. Section 17.2 discusses the conciseness of interval strategy representations with respect to Mealy machine equivalents in the finite MDP underlying the considered OC-MDP. We then study the power of interval strategies in Section 17.3. On the one hand, we show that the values (i.e., the supremum probabilities) for state-reachability and selective termination for pure interval strategies coincides with the value for arbitrary strategies. On the other hand, we show that our two classes of interval strategies are not sufficient to play optimally in general. Finally, we formalise our interval strategy verification and realisability problems in Section 17.4.

For this whole chapter, we fix an OC-MDP $\mathcal{Q} = (Q, A, \delta, w)$ and a bound $B \in \bar{\mathbb{N}}$ on counter values.

Contents

17.1 Definition

Interval strategies are a subclass of memoryless strategies of $\mathcal{M}^{\leq B}(\mathcal{Q})$ that admit finite compact representations based on families of intervals. Intuitively, a strategy is an interval strategy if there exists an *interval partition* (i.e., a partition containing only intervals) of the set of counter values such that decisions taken in a state for two counter values in the same interval coincide. We require that this partition has a finite representation to formulate verification and synthesis problems for interval strategies.

The set $[\![1, B-1]\!]$ contains all counter values for which decisions are relevant. Let $\mathcal{I}$ be an interval partition of $[\![1, B-1]\!]$. We use the following terminology to relate interval partitions and memoryless strategies.

Definition 17.1. A memoryless strategy σ of $\mathcal{M}^{\leq B}(\mathcal{Q})$ is *based on the partition* $\mathcal{I}$ if for all $q \in Q$, all $I \in \mathcal{I}$ and all $k, k' \in I$, we have $\sigma(q, k) = \sigma(q, k')$.

All memoryless strategies are based on the trivial partition of $[\![1, B-1]\!]$ into singleton sets. In practice, we are interested in strategies based on partitions with a small number of large intervals.

We define two classes of interval strategies: strategies that are based on finite partitions and, in unbounded OC-MDPs, strategies that are based on periodic partitions. An interval partition $\mathcal{I}$ of $\mathbb{N}_{>0}$ is *periodic* if there exists a period $\rho \in \mathbb{N}_{>0}$ such that for all $I \in \mathcal{I}$, $I + \rho = \{k + \rho \mid k \in I\} \in \mathcal{I}$. A periodic interval partition $\mathcal{I}$ with period ρ *induces* the interval partition $\mathcal{J} = \{I \in \mathcal{I} \mid I \subseteq [\![1, \rho]\!]\}$ of $[\![1, \rho]\!]$. Conversely, for any $\rho \in \mathbb{N}_{>0}$, an interval partition $\mathcal{J}$ of $[\![1, \rho]\!]$ *generates* the periodic partition $\mathcal{I} = \{I + k \cdot \rho \mid I \in \mathcal{J}, k \in \mathbb{N}\}$.

We define two types of interval strategies as follows.

Definition 17.2 (Interval strategies)**.** Let σ be a memoryless strategy of $\mathcal{M}^{\leq B}(\mathcal{Q})$. The strategy σ is an *open-ended interval strategy* (OEIS) if it is based on a finite interval partition of $[\![1, B-1]\!]$. When $B = \infty$, σ is a *cyclic interval strategy* (CIS) if there exists a period $\rho \in \mathbb{N}_{>0}$ such that for all $q \in Q$ and all $k \in \mathbb{N}_{>0}$, we have $\sigma(q, k) = \sigma(q, k + \rho)$.

The qualifier open-ended for OEISs follows from there being an unbounded interval in any finite interval partition of $\mathbb{N}_{>0}$ if the counter is unbounded. A strategy is a CIS with period ρ if and only if it is based on a periodic interval partition of $\mathbb{N}_{>0}$ with period ρ.

Remark 17.3. We do not consider strategies based on ultimately periodic interval partitions. However, our techniques can be adapted to analyse such strategies. We analyse OEISs and CISs through so-called compressed Markov chains (see Chapter 18). A compressed Markov chain can be defined for any memoryless strategy of $\mathcal{M}^{\leq B}(\mathcal{Q})$, and by combining our approaches for the analysis of OEISs and CISs via compressed Markov chains, we can analyse strategies based on ultimately periodic partitions. Furthermore, it can shown that our complexity bounds for the decision problems we study extend to these strategies. $\lhd$

We represent interval strategies as follows. First, assume that σ is an OEIS and let $\mathcal{I}$ be the coarsest finite interval partition of $[\![1, B-1]\!]$ on which σ is based. We can represent σ by a table that lists, for each $I \in \mathcal{I}$, the bounds of I and a memoryless strategy of $\mathcal{Q}$ dictating the choices to be made when the current counter value lies in I.

Next, assume that $B = \infty$ and that σ is a CIS with period ρ. Let $\mathcal{J}$ be an interval partition of $[\![1, \rho]\!]$ such that σ is based on the partition $\mathcal{I}$ generated by $\mathcal{J}$. We represent σ by ρ and an OEIS of $\mathcal{M}^{\leq \rho+1}(\mathcal{Q})$ based on $\mathcal{J}$ that specifies the behaviour of σ for counter values up to ρ.

Remark 17.4. In practice, in the bounded setting, it is not necessary to encode the counter upper bound B in the representation of an OEIS; it is implicit from $\mathcal{M}^{\leq B}(\mathcal{Q})$. Nonetheless, we assume that B is part of the strategy representation for the sake of convenience: this allows us to treat all bounded intervals uniformly in complexity analyses, as though all interval bounds are in the encoding of considered OEIS. This has no impact on our complexity results, as B is part of the description of $\mathcal{M}^{\leq B}(\mathcal{Q})$. $\lhd$

Interval strategies subsume *counter-oblivious strategies*, i.e., memoryless strategies that make choices based only on the state and disregard the current counter value. Counter-oblivious strategies can be viewed as memoryless strategies $\sigma \colon Q \to \mathcal{D}(A)$ of $\mathcal{Q}$.

17.2 Interval strategies and Mealy machines

We now discuss the relationship between strategies of $\mathcal{Q}$ with memory and memoryless strategies of $\mathcal{M}^{\leq B}(\mathcal{Q})$. Intuitively, memoryless strategies of $\mathcal{M}^{\leq B}(\mathcal{Q})$ can be seen as strategies of $\mathcal{Q}$ when an initial counter value is fixed: we can use memory to keep track of the counter instead of observing it. We can thus compare our representation of interval strategies to the classical Mealy machine representation of the corresponding strategies of $\mathcal{Q}$.

In this section, we formalise a translation from memoryless strategies of $\mathcal{M}^{\leq B}(\mathcal{Q})$ to strategies of $\mathcal{Q}$ with memory. We then show that the strategies derived from OEISs in the bounded setting and the strategies derived from CISs are finite-memory strategies of $\mathcal{Q}$, and that their Mealy machine representation may require a size exponential in the binary encoding size of the counter upper bound for a CIS and of the smallest period of the CIS otherwise. For OEISs in unbounded OC-MDPs, we obtain that their counterpart in $\mathcal{Q}$ need not be a finite-memory strategy.

We first formalise how to derive a strategy of $\mathcal{Q}$ from a memoryless strategy of $\mathcal{M}^{\leq B}(\mathcal{Q})$ and an initial counter value. Let $k_{\mathsf{init}} \in [\![1, B-1]\!]$ be an initial counter value and let σ be a memoryless strategy of $\mathcal{M}^{\leq B}(\mathcal{Q})$. We build a (partially-defined) strategy $\tau^{\sigma}_{k_{\mathsf{init}}}$ of $\mathcal{Q}$ from σ. Intuitively, instead of having the counter value as an input of the strategy, we store the current counter value in memory. For any history $h_{\mathcal{Q}} = q_1a_1 \dots a_{r-1}q_r \in \mathsf{Hist}(\mathcal{Q})$, we define $w(h_{\mathcal{Q}}) = \sum_{\ell=0}^{r-1} w(q_\ell, a_\ell)$. The strategy $\tau^{\sigma}_{k_{\mathsf{init}}}$ is defined, for any history $h_{\mathcal{Q}} \in \mathsf{Hist}(\mathcal{Q})$, by $\tau^{\sigma}_{k_{\mathsf{init}}}(h_{\mathcal{Q}}) = \sigma((\mathsf{last}(h_{\mathcal{Q}}), k_{\mathsf{init}} + w(h_{\mathcal{Q}})))$ when $k_{\mathsf{init}} + w(h_{\mathcal{Q}}) \in [\![1, B-1]\!]$ and is left undefined otherwise.

Similarly, the state-reachability and selective termination objectives of $\mathcal{M}^{\leq B}(\mathcal{Q})$ can be translated into objectives of $\mathcal{Q}$ when an initial counter value k_{init} is specified. We can show that the probability in $\mathcal{M}^{\leq B}(\mathcal{Q})$ of such an objective under σ from a configuration (q, k_{init}) matches the probability of the counterpart objective in $\mathcal{Q}$ under the strategy $\tau^{\sigma}_{k_{\mathsf{init}}}$ from q.

If $B \in \mathbb{N}$, the counterpart in $\mathcal{Q}$ of any OEIS has finite memory: there are only finitely many counter values. Assume that $B \in \mathbb{N}$ and let σ be an OEIS. Formally, the strategy $\tau^{\sigma}_{k_{\mathsf{init}}}$ is induced by the following Mealy machine: we let $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ where $M = [\![1, B-1]\!]$, $m_{\mathsf{init}} = k_{\mathsf{init}}$, and for all $q \in Q$,

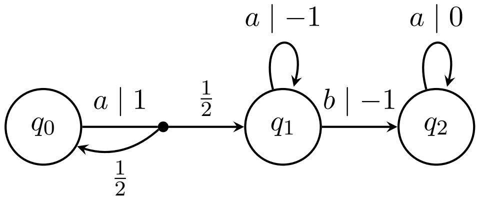


Figure 17.1: An OC-MDP. Edge splits following actions indicate probabilistic transitions. We indicate the weight of a state-action pair next to each action.

$k \in M$ and $a \in A$, we let $\mathsf{up}_{\mathfrak{M}}(k, q, a) = k + w(q, a)$ and $\mathsf{nxt}_{\mathfrak{M}}(k, q) = \sigma((q, k))$. Intuitively, $\mathfrak{M}$ induces $\tau^{\sigma}_{k_{\mathsf{init}}}$ because it keeps track of the weight of the current history and this weight determines the choice prescribed by $\tau^{\sigma}_{k_{\mathsf{init}}}$.

This construction yields a Mealy machine whose size is exponential in the binary encoding size of B. The following example illustrates that such exponential-size Mealy machines may be required, even for OEISs based on the partition $\{\{1\}, [\![2, B-1]\!]\}$, i.e., OEISs that only have to distinguish 1 from other counter values. Additionally, this same example shows that the counterpart in $\mathcal{Q}$ of an OEIS may require infinite memory in the unbounded setting.

Example 17.1. We consider the OC-MDP $\mathcal{Q}$ illustrated in Figure 17.1. Let $B \in \bar{\mathbb{N}}_{>0}$, $B \geq 3$. We consider the OEIS σ defined by $\sigma(q_1, k) = a$ for all $k \in [\![2, B-1]\!]$ and $\sigma(q_1, 1) = b$. The strategy σ maximises the probability of terminating in q_2 from the configuration $(q_0, 1)$.

We let τ^{σ}_1 denote the counterpart of σ in $\mathcal{Q}$ for the initial counter value 1. We show that for all $k \in [\![2, B-2]\!]$, a deterministic Mealy machine with at most k states cannot induce τ^{σ}_1. In particular, if B is finite, it means that any Mealy machine inducing τ^{σ}_1 must have at least $B-1$ states, and if B is infinite, it means that there is no (finite) Mealy machine inducing τ^{σ}_1.

Let $k \in [\![2, B-2]\!]$. We proceed by contradiction. Assume that there exists a Mealy machine $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ inducing τ^{σ}_1 such that $|M| \leq k$. For all $\ell \in [\![k]\!]$, we let $m_\ell = \widehat{\mathsf{up}_{\mathfrak{M}}}((q_0a)^k(q_1a)^\ell)$ and let $h_\ell = (q_0a)^k(q_1a)^\ell q_1$. For all $\ell \in [\![k]\!]$, h_ℓ is a history of $\mathcal{Q}$ in the domain of τ^{σ}_1 (because $k < B-1$ and the initial counter value is 1) that is consistent with τ^{σ}_1. Since $\mathfrak{M}$ induces τ^{σ}_1, it follows that for all $\ell \in [\![k-1]\!]$, we have $\tau^{\sigma}_1(h_\ell) = \mathsf{nxt}_{\mathfrak{M}}(m_\ell, q_1) = a$ and that $\tau^{\sigma}_1(h_k) = \mathsf{nxt}_{\mathfrak{M}}(m_k, q_1) = b$. However, m_k must occur more than once in the sequence $(m_\ell)_{\ell\in[\![k]\!]}$ because this sequence is obtained by repeating the update rule $\mathsf{up}_{\mathfrak{M}}(\cdot, q_1, a)$ from m_0 and there are no more than k memory states in $\mathfrak{M}$. This is a contradiction. $\lhd$

We now assume that $B = \infty$. The counterpart in $\mathcal{Q}$ of any CIS is a finite-memory strategy: it suffices to keep track of the remainder of division of the counter value by a period. Let σ be a CIS, ρ be a period of σ and k_{init} be an initial counter value. Formally, the strategy $\tau^{\sigma}_{k_{\mathsf{init}}}$ is induced by the Mealy machine $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ defined as follows. We let $M = [\![\rho - 1]\!]$ and $m_{\mathsf{init}} = k_{\mathsf{init}} \bmod \rho$. Updates are defined, for all $q \in Q$, $k \in M$ and $a \in A$, by $\mathsf{up}_{\mathfrak{M}}(k, q, a) = (k + w(q, a)) \bmod \rho$. The next-move function is defined differently following whether the memory state is zero or not. We let, for all $q \in Q$, $k \in M$ and $a \in A$, $\mathsf{nxt}_{\mathfrak{M}}(k, q)(a) = \sigma((q, k))$ if $k \neq 0$ and $\mathsf{nxt}_{\mathfrak{M}}(0, q)(a) = \sigma((q, \rho))$. Intuitively, $\mathfrak{M}$ induces $\tau^{\sigma}_{k_{\mathsf{init}}}$ because the remainder of the current counter value for its division by the period is sufficient to mimic σ.

By adapting Example 17.1, we can show that a CIS representation may be exponentially more succinct than any Mealy machine for its counterpart in $\mathcal{Q}$.

17.3 The power of interval strategies

Interval strategies are a proper subclass of memoryless strategies of $\mathcal{M}^{\leq B}(\mathcal{Q})$ whenever $B = \infty$. Therefore, a natural question is to ask whether interval strategies can be used to approximate the value (in the sense of Definition 2.39) from a given initial state for selective termination and state-reachability objectives. The question is not relevant in the bounded setting: all memoryless strategies are OEISs and uniformly optimal pure memoryless strategies always exist in finite MDPs for reachability [BK08].

In Section 17.3.1, we show that pure OEISs and CISs in unbounded OC-MDPs are sufficient to play almost-optimally in $\mathcal{M}^{\leq B}(\mathcal{Q})$ from all states for selective termination and state-reachability. We then provide examples in Section 17.3.2 to illustrate that our two classes of interval strategies are not sufficient to play optimally when optimal strategies exist. We provide an example where there is an optimal OEIS but no optimal CIS and another where the situation is reversed.

17.3.1 Approximating values with interval strategies

We study the question of whether OEISs and CISs in unbounded OC-MDPs attain the same value from a given initial state as general strategies for selective termination and state-reachability objectives. We assume that $B = \infty$ for the remainder of the section. In countable MDPs, for reachability objectives, pure memoryless strategies are as powerful as general ones [Orn69, KMS+20]. To show that the value for interval strategies coincides with the classical value, we derive interval strategies from pure memoryless strategies such that the interval strategies perform almost as well as the initial strategy.

Let σ be a pure memoryless strategy of $\mathcal{M}^{\leq\infty}(\mathcal{Q})$ and let $s \in Q \times \mathbb{N}$ be a configuration. Let $T \subseteq Q$ be a target and let $\Omega \in \{\mathsf{Term}(T), \mathsf{Reach}(T)\}$. Let $\varepsilon > 0$ denote our approximation precision. We can write Ω as the union, for $n \in \mathbb{N}$, of the plays that reach a target configuration of Ω without the counter exceeding n. The continuity of probability implies that we can ε-approximate $\mathbb{P}^{\sigma}_{\mathcal{M}^{\leq\infty}(\mathcal{Q}),s}(\Omega)$ by only considering the subset of Ω in which counter values are bounded by some large enough n prior to reaching a target configuration. It follows that any memoryless strategy τ that agrees with σ on all configurations with a counter value of at most n will be such that

$$\mathbb{P}^{\tau}_{\mathcal{M}^{\leq\infty}(\mathcal{Q}),s}(\Omega) \geq \mathbb{P}^{\sigma}_{\mathcal{M}^{\leq\infty}(\mathcal{Q}),s}(\Omega) - \varepsilon.$$

We use this observation to construct our sought OEIS and CIS.

Lemma 17.5. *Assume that $B = \infty$. Let $T \subseteq Q$ and $\Omega \in \{\mathsf{Term}(T), \mathsf{Reach}(T)\}$. Let Σ denote the set of pure OEISs (resp. the set of pure CISs). Then, for all $s \in Q \times \mathbb{N}$, the supremum*

$$\sup_{\sigma\in\Sigma} \mathbb{P}^{\sigma}_{\mathcal{M}^{\leq\infty}(\mathcal{Q}),s}(\Omega)$$

coincides with the value from s for Ω.

Proof. Let $s = (q, k) \in Q \times \mathbb{N}$. In the following, by a target configuration, we mean a configuration in the target set of configurations of Ω viewed as a reachability objective. Throughout this proof, all probabilities over play are with respect to $\mathcal{M}^{\leq\infty}(\mathcal{Q})$; we omit it from the notation henceforth.

In this context, the value can be approached with pure memoryless strategies [Orn69, KMS+20]. Therefore, it suffices to show that, for all pure memoryless strategies σ of $\mathcal{M}^{\leq\infty}(\mathcal{Q})$ and all $\varepsilon > 0$, there exists an OEIS (resp. a CIS) τ such that

$$\mathbb{P}_s^\tau(\Omega) \geq \mathbb{P}_s^\sigma(\Omega) - \varepsilon.$$

Let σ be a memoryless strategy of $\mathcal{M}^{\leq B}(\mathcal{Q})$ and let $\varepsilon > 0$. Let, for all $n \in \mathbb{N}$, $\Omega_{<n} \subseteq \Omega$ denote the set of plays in Ω such that a target configuration is reached without ever visiting a configuration with counter value greater or equal to n. Because Ω is a reachability objective, we have $\Omega = \bigcup_{n\in\mathbb{N}} \Omega_n$. By the continuity of probability, we obtain that

$$\lim_{n\to\infty} \mathbb{P}_s^\sigma(\Omega_{<n}) = \mathbb{P}_s^\sigma(\Omega).$$

We fix $n \geq k$ large enough such that

$$\mathbb{P}_s^\sigma(\Omega_{<n}) \geq \mathbb{P}_s^\sigma(\Omega) - \varepsilon.$$

We now let τ denote an OEIS or a CIS agreeing with σ over $Q \times [\![n]\!]$. We claim that

$$\mathbb{P}_s^\tau(\Omega) \geq \mathbb{P}_s^\sigma(\Omega) - \varepsilon.$$

Let $\mathcal{H} \subseteq \mathsf{Hist}(\mathcal{M}^{\leq\infty}(\mathcal{Q}))$ be the set of histories in which counter values are strictly less than n that end in a target configuration and such that no target configuration appears prior to the last configuration. We obtain that $\mathcal{H}$ is prefix-free and that $\mathsf{Cyl}(\mathcal{H}) = \Omega_{<n}$. Because τ and σ agree over $Q \times [\![n]\!]$, we obtain that:

$$\mathbb{P}_s^\tau(\Omega) \geq \mathbb{P}_s^\tau(\Omega_{<n}) = \mathbb{P}_s^\tau(\mathsf{Cyl}(\mathcal{H})) = \mathbb{P}_s^\sigma(\mathsf{Cyl}(\mathcal{H})) = \mathbb{P}_s^\sigma(\Omega_{<n}) \geq \mathbb{P}_s^\tau(\Omega) - \varepsilon.$$

This ends the proof. □

17.3.2 Limitations of interval strategies

Lemma 17.5 shows that interval strategies can be used to approximate the optimal probability of the objectives we study from any state. In this section, we provide an example where an OEIS suffices to play optimally but no CIS does, and another where a CIS suffices but no OEIS does. These two examples can

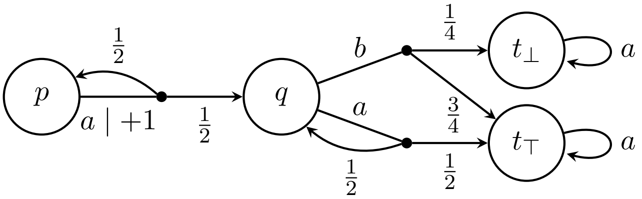


Figure 17.2: An OC-MDP where all unspecified weights are -1. An OEIS is sufficient to maximise the probability of reaching t from $(p, 1)$, but no CIS is.

be combined to obtain an example in which a strategy based on an ultimately periodic partition of $\mathbb{N}$ is required; we present them separately to simplify their presentation.

Our two examples share a similar structure: we have an initial state with a single enabled action with weight 1 from which we either loop back or move into a finite-horizon MDP, i.e., an OC-MDP where all weights are -1, in which termination is guaranteed. We obtain a uniformly optimal strategy in the finite-horizon MDPs by using value iteration to determine optimal actions for each remaining number of steps. We provide a brief description of value iteration in Appendix A.2.2.

We first provide an example where an optimal OEIS exists, but no optimal CIS does. In finite-horizon MDPs, when the remaining number of steps is large, the optimal actions are chosen to be safe actions. More precisely, one only uses actions that could be prescribed by an optimal memoryless strategy for (infinite-horizon) reachability. However, when the number of steps gets low, it can be worthwhile to take an actions that would be deemed too risky otherwise. In the following example, we require a specific action when the number of remaining steps gets low, and otherwise we use another action.

Example 17.2. We consider the OC-MDP $\mathcal{Q}$ depicted in Figure 17.2 in which all weights are -1 besides the one on the transition leading out of p. We consider the objective $\Omega = \mathsf{Reach}(t_\top) = \mathsf{Term}(t_\top)$, the counter bound $B = \infty$. We claim that there exists an OEIS that is optimal from $(p, 1)$, but no CIS.

We first analyse the optimal decisions to be made in q depending on the counter value. From $(q, 1)$, action a leads to Ω being satisfied with probability $\frac{1}{2}$ whereas choosing b yields a probability of $\frac{3}{4}$. Therefore, we must choose b in

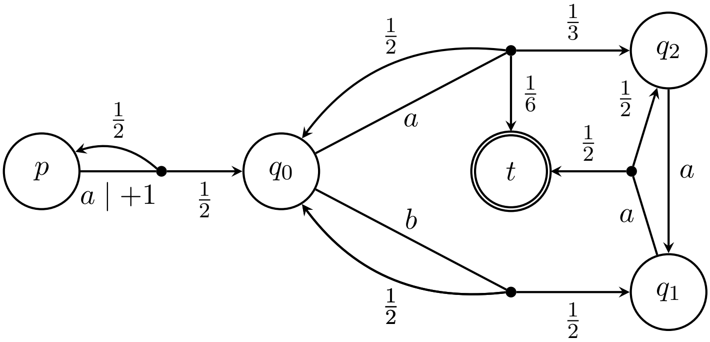


Figure 17.3: An OC-MDP where all unspecified weights are -1. A CIS is sufficient to maximise the probability of reaching t from $(p, 1)$, but no OEIS is: one must alternate the actions a and b in q_0 to be optimal.

$(q, 1)$. On the other hand, for any $k \geq 2$, a is preferable to b in (q, k): action a leads to a satisfaction probability of $\frac{1}{2} + \frac{1}{2}\theta_{k-1}$, where $\theta_{k-1} \geq \frac{3}{4}$ is the optimal probability from $(q, k-1)$ (the inequality $\theta_{k-1} \geq \frac{3}{4}$ follows from $k - 1 \geq 1$), whereas b yields the smaller probability $\frac{3}{4}$.

We claim that the only optimal memoryless strategy from $(p, 1)$ is the pure OEIS σ defined by $\sigma(q, 1) = b$ and $\sigma(q, k) = a$ for all $k \geq 2$. On the one hand, regardless of the used strategy, all configurations (q, k) with $k \geq 2$ will be reached from $(p, 1)$. It follows that any memoryless strategy that is optimal from $(p, 1)$ must agree with σ over all configurations with counter value greater or equal to 2. Any strategy that agrees with σ over these configurations visits $(q, 1)$ with positive probability and thus must also play optimally from there. It follows that the only optimal memoryless strategy from $(p, 1)$ is σ. In particular, no CIS is optimal from $(p, 1)$. ◁

We now consider an example that show that OEISs do not suffice in general to play optimally. In our example, we describe a finite-horizon MDP in which it is necessary to alternate between two enabled actions to play optimally. It follows that there is an optimal CIS but no optimal OEIS.

Example 17.3. We consider the OC-MDP $\mathcal{Q} = (Q, A, \delta, w)$ depicted in Figure 17.3 in which all weights are -1 besides the one on the transition leading out of p. We consider objective $\Omega = \mathsf{Reach}(t) = \mathsf{Term}(t)$, the counter bound

$B = \infty$. We claim that there exists a CIS that is optimal from $(p, 1)$, but no OEIS. Similarly to Example 17.2, we first analyse the optimal choices in (q_0, k) for all $k \in \mathbb{N}_{>0}$, and derive an optimal CIS from them. We let, for all $k \in \mathbb{N}$, $\mathbf{v}^{(k)} = (v_q^{(k)})_{q \in q}$ where, for all $q \in Q$, $v_q^{(k)}$ denotes the maximum probability of Ω from (q, k) for all strategies.

First, we note that the value in all configurations in $\{q_1, q_2\} \times \mathbb{N}$ is independent of the chosen strategy: if one removes p and q_0 from the OC-MDP, we obtain a one-counter Markov chain. For all $k \in \mathbb{N}$, we can show by induction that $v_{q_1}^{(2k)} = 1 - \frac{1}{2^k}$, $v_{q_1}^{(2k+1)} = 1 - \frac{1}{2^{k+1}}$ and $v_{q_2}^{(2k)} = v_{q_2}^{(2k+1)} = 1 - \frac{1}{2^k}$.

Let $\sigma\colon Q \times \mathbb{N} \to A$ be an optimal strategy; such a strategy exists because the only decisions are made in q_0 and we can define pure uniformly optimal strategies in finite-horizon MDPs through value iteration. We establish that for any $k \in \mathbb{N}_{>0}$, it must be the case that $\sigma(q_0, k) = a$ if k is odd and $\sigma(q_0, k) = b$ otherwise. It follows from σ being optimal that

$$\sigma(q_0, k) \in \operatorname*{argmax}_{c \in \{a,b\}} \left(\sum_{q \in Q} \delta(q_0, c)(q) \cdot v_q^{(k-1)} \right). \tag{17.1}$$

Let $k \in \mathbb{N}_0$. First, let us assume that $k = 2 \cdot \ell + 2$ is even (note that $k \geq 1$, i.e., we can write k in this way while handling all cases). It follows from Equation (17.1) that it suffices that the inequality

$$\frac{1}{6} + \frac{1}{3} \cdot \left(1 - \frac{1}{2^\ell}\right) < \frac{1}{2} \cdot \left(1 - \frac{1}{2^{\ell+1}}\right),$$

holds to imply $\sigma(q_0, k) = b$. Given that this inequality is equivalent to $4 > 3$, we obtain $\sigma(q_0, k) = b$. Now, let us assume that $k = 2 \cdot \ell + 1$ is odd. It follows from Equation (17.1) that if suffices to have

$$\frac{1}{6} + \frac{1}{3} \cdot \left(1 - \frac{1}{2^\ell}\right) > \frac{1}{2} \cdot \left(1 - \frac{1}{2^\ell}\right),$$

hold to have $\sigma(q_0, k) = a$. Given that this inequality is equivalent to $2 < 3$, we obtain $\sigma(q_0, k) = a$.

It follows from the above that the CIS that selects action a in q_0 for odd counter values and b for even counter values is optimal from $(p, 1)$ for Ω. Furthermore, the above shows that no OEIS can be optimal from $(p, 1)$, as for all $k \geq 2$, we reach (q_0, k) from $(p, 1)$ with positive probability regardless of the used strategy (and we must play optimally from these configurations). ◁

17.4 Interval strategy decision problems

We formally define the decision problems we are interested in regarding interval strategies. The common inputs of these problems are an OC-MDP $\mathcal{Q}$ with rational transition probabilities, a counter bound $B \in \bar{\mathbb{N}}_{>0}$ (encoded in binary if it is finite), a target $T \subseteq Q$, an objective $\Omega \in \{\mathsf{Reach}(T), \mathsf{Term}(T)\}$, an initial configuration $s_{\mathsf{init}} = (q_{\mathsf{init}}, k_{\mathsf{init}})$ and a threshold $\theta \in [0, 1] \cap \mathbb{Q}$ against which we compare the probability of Ω. Problems that are related to CISs assume that $B = \infty$ (as all strategies in the bounded case are OEISs). We lighten the probability notation below by omitting $\mathcal{M}^{\leq B}(\mathcal{Q})$ from it.

First, we are concerned with the verification of interval strategies, i.e., whether Ω is satisfied with probability at least θ from s_{init} for a given strategy.

Definition 17.6. The *interval strategy verification problem* asks to decide, given an interval strategy σ, whether $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\Omega) \geq \theta$.

When studying the verification problem, we assume that the encoding of the input interval strategy matches the description of Section 17.1.

The other two problems relate to interval strategy synthesis. The corresponding decision problem is called *realisability*. We provide algorithms checking the existence of well-performing structurally-constrained interval strategies. We formulate two variants of this problem.

For the first variant, we fix an interval partition $\mathcal{I}$ of $[\![1, B-1]\!]$ beforehand and ask to check if there is a good strategy based on $\mathcal{I}$.

Definition 17.7. The *fixed-interval OEIS realisability problem* asks, given a finite interval partition $\mathcal{I}$ of $[\![1, B-1]\!]$, whether there exists an OEIS σ based on $\mathcal{I}$ such that $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\Omega) \geq \theta$.

The variant for CISs is defined similarly.

Definition 17.8. The *fixed-interval CIS realisability problem* asks, given a period $\rho \in \mathbb{N}_{>0}$ and an interval partition $\mathcal{J}$ of $[\![1, \rho]\!]$, whether there exists a CIS σ based on the partition generated by $\mathcal{J}$ such that $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\Omega) \geq \theta$.

For the second variant, we parameterise the number of intervals in the partition and the size of bounded intervals.

Definition 17.9. The *parameterised OEISs realisability problem* asks, given a bound $d \in \mathbb{N}_{>0}$ on the number of intervals and a bound $n \in \mathbb{N}_{>0}$ on the size of bounded intervals, whether there exists an OEIS σ such that $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\Omega) \geq \theta$ and σ is based on an interval partition $\mathcal{I}$ of $[\![1, B-1]\!]$ with $|\mathcal{I}| \leq d$ and, for all bounded $I \in \mathcal{I}$, $|I| \in [\![1, n]\!]$.

We note that, in the above definition, the parameter n does not constrain the required infinite interval in the unbounded setting $B = \infty$. For CISs, the parameterised realisability problem is defined as follows.

Definition 17.10. The *parameterised CIS realisability problem* asks, given a bound $d \in \mathbb{N}_{>0}$ on the number of intervals and a bound $n \in \mathbb{N}_{>0}$ on the size of intervals, whether there exists a CIS σ such that $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\Omega) \geq \theta$ and σ is based on an interval partition $\mathcal{I}$ of $\mathbb{N}_{>0}$ with period ρ such that $|I| \leq n$ for all $I \in \mathcal{I}$ and $\mathcal{I}$ induces a partition of $[\![1, \rho]\!]$ with at most d intervals.

For both parameterised realisability problems, we assume that the number d is given in unary. This ensures that witness strategies, when they exist, are based on interval partitions that have a representation of size polynomial in the size of the inputs.

Remark 17.11. In bounded OC-MDPs, instances of the parameterised OEIS realisability problem such that no partitions are compatible with the input parameters d and n are trivially negative. If $B \in \mathbb{N}$, then there are no interval partitions of $[\![1, B-1]\!]$ compatible with the parameters d and n whenever $B - 1 > d \cdot n$. This issue does not arise for OEISs in unbounded OC-MDPs or for CISs, as counter-oblivious strategies are always possible witnesses no matter the parameters. $\lhd$

For both realisability problems, we consider two variants, depending on whether we want the answer with respect to the set of *pure* or *randomised* interval strategies. For many objectives in MDPs (e.g., reachability, parity objectives [BK08, BORV23]), the maximum probability of satisfying the objective

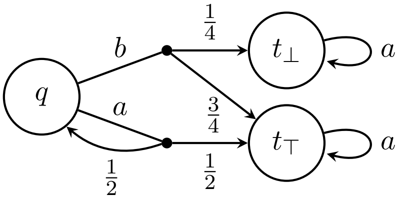


Figure 17.4: A variant of the OC-MDP of Figure 17.2. All weights are -1 and are omitted from the figure.

is the same for pure and randomised strategies. As highlighted by the following example (a variant of Example 17.2), when we restrict the structure of the sought interval strategy, there may exist randomised strategies that perform better than all pure ones. Intuitively, randomisation can somewhat alleviate the loss in flexibility caused by structural restrictions [CRR14].

Example 17.4. The fixed-interval and parameterised realisability problems subsume the realisability problem for counter-oblivious strategies. We provide an OC-MDP in which there exists a randomised counter-oblivious strategy that performs better than any pure counter-oblivious strategy from a given configuration.

We consider the OC-MDP $\mathcal{Q}$ depicted in Figure 17.4 in which all weights are -1, the objective $\mathsf{Reach}(t_\top) = \mathsf{Term}(t_\top)$, a counter bound $B \geq 3$ and the initial configuration $(q, 2)$. In Example 17.2, we have considered a variant of this OC-MDP and have shown that to maximise the probability of reaching $t_\top$ from $(q, 2)$ with an unrestricted strategy, we must select action a in $(q, 2)$ and then b in $(q, 1)$.

We now limit our attention to counter-oblivious strategies. For pure strategies, no matter whether action a or b is chosen in q, $t_\top$ is reached with probability $\frac{3}{4}$ from $(q, 2)$. However, when playing both actions uniformly at random in q, the resulting reachability probability from $(q, 2)$ is $\frac{25}{32} > \frac{3}{4}$. This shows that randomised counter-oblivious strategies can achieve better reachability (resp. selective termination) probabilities than pure strategies. ◁

Chapter 18

Compressing induced Markov chains in one-counter Markov decision processes

This chapter introduces compressed Markov chains. Compressed Markov chains are the main tool underlying our algorithms for the interval-strategy-related decision problems formalised in the previous chapter. We use compressed Markov chains to analyse the (possibly infinite) Markov chains induced by memoryless strategies over the space of configurations of an OC-MDP. A compressed Markov chain is defined with respect to a memoryless strategy and an interval partition on which the strategy is based. This construction is generic, in the sense that it can formally be defined not only for interval strategies, but for all memoryless strategies.

Section 18.1 illustrates and formalises compressed Markov chains. A compressed Markov chain can only be constructed with respect to interval partitions with intervals respecting some size constraint; Section 18.2 explains how to efficiently refine interval partitions to enforce these constraints. In Section 18.3, we prove that termination probabilities and the probability of hitting a counter upper bound are preserved following compression. We show that the transition probabilities of the compressed Markov chain can be represented as solutions of systems of quadratic equations in Section 18.4. Finally, we close the chapter by proving that compressed Markov chains for CISs are induced by one-counter Markov chains in Section 18.5.

For this whole chapter, we fix an OC-MDP $\mathcal{Q} = (Q, A, \delta, w)$, a bound $B \in \bar{\mathbb{N}}_{>0}$ on counter values and a memoryless strategy σ of $\mathcal{M}^{\leq B}(\mathcal{Q})$ based on an interval partition $\mathcal{I}$ of $[\![1, B-1]\!]$.

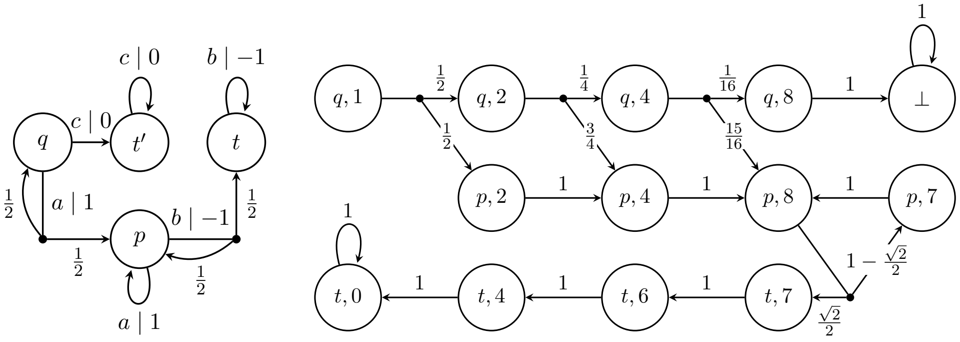
Contents

18.1 Definition

We define the *compressed Markov chain* $\mathcal{C}^\sigma_\mathcal{I}(\mathcal{Q})$ derived from the Markov chain induced on $\mathcal{M}^{\leq B}(\mathcal{Q})$ by σ and the partition $\mathcal{I}$. We write $\mathcal{C}^\sigma_\mathcal{I}$ instead of $\mathcal{C}^\sigma_\mathcal{I}(\mathcal{Q})$ whenever $\mathcal{Q}$ is clear from the context. Intuitively, we keep some configurations in the state space of $\mathcal{C}^\sigma_\mathcal{I}$ and, to balance this, transitions of $\mathcal{C}^\sigma_\mathcal{I}$ aggregate several histories of the induced Markov chain. We also apply compression for bounded intervals: interval bounds are represented in binary and thus the size of an interval is exponential in its encoding size. We open with an example.

Example 18.1. We consider the OC-MDP depicted in Figure 18.1a and counter upper bound $B = +\infty$. Let σ denote the randomised OEIS based on $\mathcal{I} = \{[\![1, 7]\!], [\![8, \infty]\!]\}$ such that $\sigma(q, 1)(a) = \sigma(p, 1)(a) = \sigma(q, 8)(c) = 1$ and $\sigma(p, 8)(a) = \sigma(p, 8)(b) = \frac{1}{2}$.

We define the compressed Markov chain $\mathcal{C}^\sigma_\mathcal{I}$ depicted in Figure 18.1b by processing each interval individually. First, we consider the bounded interval $I = [\![1, 7]\!]$. When we enter I from its minimum or maximum, we only consider counter jumps by powers of 2, starting with $1 = 2^0$. If a counter value in I is reached by jumping by 2^β, we consider counter updates of $2^{\beta+1}$ from it; Figure 18.2 illustrates this counter update rule. This explains the counter progressions from $(q, 1)$ to $(q, 8)$ and from $(t, 7)$ to $(t, 0)$. The state space of $\mathcal{C}^\sigma_\mathcal{I}$ with respect to I contains the configurations whose counter values can be reached via this scheme. Transitions aggregate several histories of $\mathcal{M}^{\leq\infty}(\mathcal{Q})$, e.g., the probability from $s = (q, 2)$ to $s' = (p, 4)$ is the probability under σ



(a) An OC-MDP. Weights are written next to actions.

(b) Fragment of the compressed Markov chain of Example 18.1 reachable from $(q, 1)$. Configuration parentheses are omitted to lighten the figure.

Figure 18.1: An illustration of an OC-MDP and its compression for a specific strategy.

of all histories of $\mathcal{M}^{\leq \infty}(\mathcal{Q})$ from s to s' along which counter values elsewhere than in s' remain between $\min I = 1$ and 3 (i.e., the counter value before the next step). The encoding of transition probabilities may be exponential in the number of retained configurations; this is highlighted by the progression of probability denominators between $(q, 1)$ and $(q, 8)$.

For the unbounded interval $I = [\![8, \infty]\!]$, we only consider configurations with counter value $\min I = 8$ and consider transitions to configurations with counter value $\min I - 1 = 7$. In this case, for instance, the transition probability from $(p, 8)$ to $(p, 7)$, corresponds the probability under σ in $\mathcal{M}^{\leq \infty}(\mathcal{Q})$ of hitting counter value 7 for the first time in p from $(p, 8)$. This example illustrates that this probability can be irrational. Here, the probability of moving from $(p, 8)$ to $(p, 7)$ is a solution of the quadratic equation $x = \frac{1}{4} + \frac{1}{2}x^2$ (see [KEM06]): $\frac{1}{4}$ is the probability of directly moving from $(p, 8)$ to $(p, 7)$ and $\frac{1}{2}x^2$ is the probability of moving from $(p, 8)$ to $(p, 7)$ by first going through the intermediate configuration $(p, 9)$.

Finally, we comment on the absorbing state $\perp$. The rules making up transitions of $\mathcal{C}^{\sigma}_{\mathcal{I}}$ outlined above require a change in counter value. We redirect the probability of never seeing such a change to $\perp$. In this example, σ does not allow a counter decrease from $(q, 8)$. $\triangleleft$

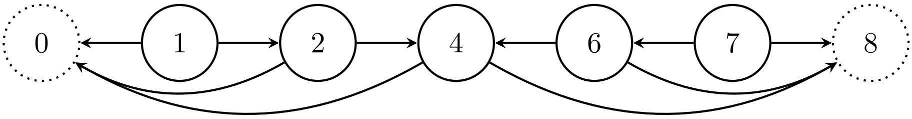


Figure 18.2: An illustration of the counter update scheme in a compressed Markov chain for the interval $[\![1,7]\!]$.

Example 18.1 outlines the main ideas to construct $\mathcal{C}^{\sigma}_{\mathcal{I}}$. To ensure that compressed Markov chains are well-defined, we impose the following assumption on $\mathcal{I}$ which guarantees that, in general, bounded intervals of $\mathcal{I}$ can only be entered by one of their bounds.

Assumption 18.1. For all bounded $I \in \mathcal{I}$, $\log_2(|I|+1) \in \mathbb{N}$, i.e., $|I| = 2^{\beta_I} - 1$ for some $\beta_I \in \mathbb{N}$.

Assumption 18.1 is not prohibitive: we prove in Section 18.2 that, given a bounded interval, we can partition it into sub-intervals satisfying the required size constraint in polynomial time. We assume that Assumption 18.1 is satisfied for $\mathcal{I}$ for the remainder of the chapter.

We now formalise $\mathcal{C}^{\sigma}_{\mathcal{I}} = (S_{\mathcal{I}}, \delta^{\sigma}_{\mathcal{I}})$. We start by defining its state space $S_{\mathcal{I}}$ which does not depend on σ. We first formalise the configurations that are retained for each interval.

Let $I \in \mathcal{I}$. First, let us assume that I is unbounded and let b_I^- denote its minimum. We set $S_I = Q \times \{b_I^-\}$, i.e., we only retain the configurations with minimal counter value in I.

Next, let us assume that I is bounded and of the form $[\![b_I^-, b_I^+]\!]$. Let $\beta_I = \log_2(|I|+1)$ (this is an integer by Assumption 18.1). We retain the counter values that can be reached via a generalisation of the scheme depicted in Figure 18.2. The set of retained configurations for I is given by

$$S_I = Q \times \left(\{b_I^- + 2^\alpha - 1 \mid \alpha \in [\![\beta_I - 1]\!]\} \cup \{b_I^+ - (2^\alpha - 1) \mid \alpha \in [\![\beta_I - 1]\!]\}\right).$$

Finally, we consider absorbing configurations and the new state $\perp$. We let $S^{\perp}_{\mathcal{I}} = \{\perp\} \cup (Q \times \{0, B\})$ if $B \in \mathbb{N}$ and $S^{\perp}_{\mathcal{I}} = \{\perp\} \cup (Q \times \{0\})$ otherwise. We define $S_{\mathcal{I}} = S^{\perp}_{\mathcal{I}} \cup \bigcup_{I \in \mathcal{I}} S_I$.

We now define the transition function $\delta^\sigma_{\mathcal{I}}$. For all $s \in S^\perp_{\mathcal{I}}$, we let $\delta^\sigma_{\mathcal{I}}(s)(s) = 1$. For configurations whose counter value lies in one of the intervals $I \in \mathcal{I}$, we provide a unified definition based on a notion of *successor counter values*, generalising the ideas of Example 18.1 and Figure 18.2.

Let $I \in \mathcal{I}$. If I is unbounded, we define the successor of $b^-_I = \min I$ to be $b^-_I - 1$. We now assume that $I = [\![b^-_I, b^+_I]\!]$ is bounded and let $\beta_I = \log_2(|I| + 1)$ and $\alpha \in [\![\beta_I - 1]\!]$. The successors of $b^-_I + 2^\alpha - 1$ are $b^-_I - 1$ and $b^-_I + 2^{\alpha+1} - 1$. Symmetrically, for $b^+_I - (2^\alpha - 1)$, its successors are $b^+_I + 1$ and $b^+_I - (2^{\alpha+1} - 1)$. Both cases entail a counter change by 2^α. Assumption 18.1 ensures that all successor counter values appear in $S_{\mathcal{I}}$.

Let $s = (q, k) \in S_{\mathcal{I}} \setminus S^\perp_{\mathcal{I}}$ and $s' = (q', k') \in S_{\mathcal{I}} \setminus \{\perp\}$. If k' is not a successor of k, we set $\delta^\sigma_{\mathcal{I}}(s)(s') = 0$. Assume now that k' is a successor of k. We let $\mathcal{H}_{\mathsf{succ}}(s, s') \subseteq \mathsf{Hist}(\mathcal{M}^{\leq B}(\mathcal{Q}))$ be the set of histories h such that $\mathsf{first}(h) = s$, $\mathsf{last}(h) = s'$ and for all configurations s'' along h besides s', the counter value of s'' is not a successor of k; outside of s' along h, the counter remains, in the bounded case, strictly between the two successors of k, and, in the unbounded case, strictly above the successor $k - 1$ of k. We set $\delta^\sigma_{\mathcal{I}}(s)(s') = \mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\mathsf{Cyl}\,(\mathcal{H}_{\mathsf{succ}}(s, s')))$. To ensure that $\delta^\sigma_{\mathcal{I}}(s)$ is a distribution we let $\delta^\sigma_{\mathcal{I}}(s)(\perp) = 1 - \sum_{s'' \neq \perp} \delta^\sigma_{\mathcal{I}}(s)(s'')$; this transition captures the probability of the counter never hitting a successor of k.

Remark 18.2. Although we have formalised compressed Markov chains for OC-MDPs, the construction can be applied to one-counter Markov chains. In particular, the properties outlined below transfer to the compression of a one-counter Markov chain. ◁

In the following, we differentiate histories of $\mathcal{C}^\sigma_{\mathcal{I}}$ from histories of $\mathcal{M}^{\leq B}(\mathcal{Q})$ by denoting them with a bar, e.g., $\bar{h}$ indicates a history of $\mathcal{C}^\sigma_{\mathcal{I}}$.

18.2 Efficiently refining interval partitions

To define a compressed Markov chain with respect to an interval partition $\mathcal{J}$ of $[\![1, B - 1]\!]$, we require that the size constraints of Assumption 18.1 hold, i.e., that for all bounded $I \in \mathcal{J}$, $\log_2(|I| + 1) \in \mathbb{N}$. We present a refinement procedure for interval partitions that enforces this property while generating few intervals. At the end of this section, we provide an additional procedure that

Algorithm 18.1: Procedure Refine to split an interval into intervals of size in $\{2^\beta - 1 \mid \beta \in \mathbb{N}\}$.

```
Data: A bounded interval $I = [\![b^-, b^+]\!]$.
$\ell \leftarrow \lfloor \log_2(b^+ - b^- + 2) \rfloor (= \lfloor \log_2(|I| + 1) \rfloor)$;
if $|I| = 2^\ell - 1$ then
    return $\{I\}$;
else
    $I' \leftarrow [\![b^-, b^- + 2^\ell - 2]\!]$; $I'' \leftarrow [\![b^- + 2^\ell - 1, b^+]\!]$;
    return $\{I'\} \cup \mathsf{Refine}(I'')$;
```

can be used to retain specific configurations in the state space of compressed Markov chains.

To refine an interval partition, we subdivide its bounded intervals one by one. Breaking up these intervals into singleton sets is not a valid approach for complexity reasons; any input interval partition is such that its interval bounds are represented in binary, i.e., the size of intervals is exponential in the size of their representation. We provide a polynomial-time refinement procedure that divides an interval into sub-intervals of the required size in Algorithm 18.1. To refine an interval, we determine a largest sub-interval of a suitable size and then continue by recursively partitioning its complement. This algorithm enables us, in the context of verification, to modify the interval partition from the representation of an interval strategy into one suitable for compressed Markov chains.

We show that, for all bounded intervals I of $\mathbb{N}_{>0}$, the partition $\mathsf{Refine}(I)$ (from Algorithm 18.1) has a polynomial size (with respect to the binary encoding of the bounds of I) and all of its elements J satisfy $\log_2(|J| + 1) \in \mathbb{N}$.

Lemma 18.3. *Let $I = [\![b^-, b^+]\!]$ be a bounded interval of $\mathbb{N}_{>0}$. The interval partition $\mathsf{Refine}(I)$ of I obtained via Algorithm 18.1 satisfies $|\mathsf{Refine}(I)| \leq \log_2(|I|+1)+1 \leq \log_2(b^+)+1$ and, for all $J \in \mathsf{Refine}(I)$, we have $\log_2(|J|+1) \in \mathbb{N}$.*

Proof. Let $\ell = \lfloor \log_2(|I|+1) \rfloor$. We show both statements by induction.

For the size of the elements in $\mathsf{Refine}(I)$, we proceed by induction on $|I|$. If $|I| = 1$, then $\mathsf{Refine}(I) = \{I\}$ (since $1 = 2^1 - 1$) and the result follows. Now, assume that for all intervals smaller than I, the statement holds. If $I = 2^\ell - 1$, we have $\mathsf{Refine}(I) = \{I\}$ which satisfies the condition. Otherwise, we let $I' = \llbracket b^-, b^- + 2^\ell - 2 \rrbracket$ and $I'' = \llbracket b^- + 2^\ell - 1, b^+ \rrbracket$. In particular, we have $|I'| = b^- + 2^\ell - 2 - b^- + 1 = 2^\ell - 1$ and thus $\log_2(|I'|+1) \in \mathbb{N}$. We conclude that all elements of $\mathsf{Refine}(I) = \{I'\} \cup \mathsf{Refine}(I'')$ satisfy the required constraint on their size.

We now show that $|\mathsf{Refine}(I)| \leq \ell + 1$. We proceed by induction on ℓ. If $\ell = 1$, then $|I| = 1$ and we have $|\mathsf{Refine}(I)| = 1$. This closes the base case.

We assume by induction that for all I' such that $\ell' = \lfloor \log_2(|I'|+1) \rfloor < \ell$, we have $|\mathsf{Refine}(I')| \leq \ell' + 1$. If $|I| = 2^\ell - 1$, we have $|\mathsf{Refine}(I)| = 1 \leq \ell + 1$. We thus assume that $2^\ell - 1 < |I| < 2^{\ell+1} - 1$ (the upper bound follows from the definition of ℓ). We let $I' = \llbracket b^-, b^- + 2^\ell - 2 \rrbracket$ and $I'' = \llbracket b^- + 2^\ell - 1, b^+ \rrbracket$. It remains to show that $|\mathsf{Refine}(I'')| \leq \ell$ to conclude. It holds that $|I''| = |I| - (2^\ell - 1) < 2^{\ell+1} - 1 - (2^\ell - 1) = 2^\ell$. We distinguish two cases in light of this. First, we assume that $|I''| = 2^\ell - 1$. In this case, we have $|\mathsf{Refine}(I'')| = 1$, which implies that $|\mathsf{Refine}(I)| = 2 \leq \ell + 1$. Second, we assume that $|I''| < 2^\ell - 1$. By the induction hypothesis, we obtain that $|\mathsf{Refine}(I'')| \leq \ell$, ensuring that $|\mathsf{Refine}(I)| \leq \ell + 1$ and ending the argument. □

For the sake of conciseness, we extend the Refine operator to infinite intervals and interval partitions. For any infinite interval I of $\mathbb{N}_{>0}$, we let $\mathsf{Refine}(I) = \{I\}$. Let J be an interval of $\mathbb{N}_{>0}$ and let $\mathcal{J}$ be a partition of J. We let $\mathsf{Refine}(\mathcal{J}) = \bigcup_{I \in \mathcal{J}} \mathsf{Refine}(I)$. Lemma 18.3 implies that the constraints of Assumption 18.1 are satisfied by $\mathsf{Refine}(\mathcal{J})$. This result also yields bounds on the size of $\mathsf{Refine}(\mathcal{J})$ when $\mathcal{J}$ is finite.

We remark that if an interval partition $\mathcal{J}$ of $\mathbb{N}_{>0}$ has period ρ and is generated by an interval partition $\mathcal{J}'$ of $\llbracket 1, \rho \rrbracket$, then $\mathsf{Refine}(\mathcal{J})$ is generated by $\mathsf{Refine}(\mathcal{J}')$.

We now introduce an operator ensuring that a specific counter value is retained in a compression by making it an interval bound. For any interval $I = \llbracket b^-, b^+ \rrbracket$ of $\mathbb{N}_{>0}$ (not necessarily bounded) and $k \in \mathbb{N}$, we let $\mathsf{Isolate}(I, k)$

denote $\{I\}$ if $k \notin I$ and $\{[\![b^-, k]\!], [\![k+1, b^+]\!]\} \setminus \{\emptyset\}$ if $k \in I$. We extend the $\mathsf{Isolate}$ operator to interval partitions as follows. For all intervals J of $\mathbb{N}_{>0}$, interval partitions $\mathcal{J}$ of J and $k \in \mathbb{N}$, we let $\mathsf{Isolate}(\mathcal{J}, k) = \bigcup_{I \in \mathcal{J}} \mathsf{Isolate}(I, k)$.

18.3 Validity of the compression approach

We establish that for all configurations $s \in S_{\mathcal{I}}$ and states $q \in Q$, the probability of terminating or reaching the counter upper bound B in q coincides in $\mathcal{C}^{\sigma}_{\mathcal{I}}$ and in the Markov chain induced on $\mathcal{M}^{\leq B}(\mathcal{Q})$ by σ. There is not such a direct correspondence for state-reachability probabilities. We prove that for all targets $T \subseteq Q$, there exists an OC-MDP $\mathcal{Q}'$ with state space Q derived by changing transitions of $\mathcal{Q}$ such that, for all $q \in T$, the probability of visiting T for the first time via a configuration with state q in $\mathcal{M}^{\leq B}(\mathcal{Q})$ under σ coincides with the probability of terminating or hitting the counter upper bound in q in $\mathcal{M}^{\leq B}(\mathcal{Q}')$ under σ.

For the first property, we rely on a relation between histories of $\mathcal{C}^{\sigma}_{\mathcal{I}}$ and of $\mathcal{M}^{\leq B}(\mathcal{Q})$. Let $h = s_0 a_0 \dots a_{r-1} s_r \in \mathsf{Hist}(\mathcal{M}^{\leq B}(\mathcal{Q}))$ such that $\mathsf{last}(h) \in Q \times \{0, B\}$ and $\mathsf{last}(h)$ occurs only once in h. By induction, we identify a sequence of configurations in $S_{\mathcal{I}}$ along h that is a well-formed history of $\mathcal{C}^{\sigma}_{\mathcal{I}}$. Let $\ell_0 = 0$. Assume that we have constructed an increasing sequence $\ell_0 < \dots < \ell_\iota$ such that $s_{\ell_0} \dots s_{\ell_\iota} \in \mathsf{Hist}(\mathcal{C}^{\sigma}_{\mathcal{I}})$. If $\ell_\iota \neq r$, we let $\ell_{\iota+1}$ be the least index $\ell > \ell_\iota$ such that $s_\ell \in S_{\mathcal{I}}$ and $\delta^{\sigma}_{\mathcal{I}}(s_{\ell_\iota})(s_\ell) > 0$ and continue the induction. Such an index is guaranteed to exist. Since weights are in $\{-1, 0, 1\}$, we witness all counter values between that of s_{ℓ_ι} and s_r in the suffix $s_{\ell_\iota} \dots s_r$. Furthermore, all counter values have a smaller successor, and those from a bounded interval have a greater successor. If $\ell_\iota = r$, the induction ends and we let $\bar{h} = s_0 s_{\ell_1} \dots s_{\ell_{r'-1}} s_{\ell_{r'}}$ be the resulting history. We say that $\bar{h}$ *abstracts* h, and it is the unique history of $\mathcal{C}^{\sigma}_{\mathcal{I}}$ that does so.

We now state the first theorem of this section. The crux of its proof is to establish that, for all histories $\bar{h}$ of $\mathcal{C}^{\sigma}_{\mathcal{I}}$ ending in $Q \times \{0, B\}$, the probability of its cylinder in $\mathcal{C}^{\sigma}_{\mathcal{I}}$ matches the probability that a history abstracted by $\bar{h}$ occurs in the Markov chain induced by σ on $\mathcal{M}^{\leq B}(\mathcal{Q})$. We conclude by writing reachability objectives as countable unions of history cylinders. For the sake of clarity, in the following statement, we indicate the relevant MDP or Markov chain for each objective.

Theorem 18.4. *Let $s \in S_{\mathcal{I}} \setminus \{\bot\}$. For all $q \in Q$, $\mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\mathsf{Term}(q)) = \mathbb{P}_{\mathcal{C}^{\sigma}_{\mathcal{I}},s}(\mathsf{Reach}_{\mathcal{C}^{\sigma}_{\mathcal{I}}}(q,0))$ and, if $B \in \mathbb{N}$, $\mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\mathsf{Reach}_{\mathcal{M}^{\leq B}(\mathcal{Q})}(q,B)) = \mathbb{P}_{\mathcal{C}^{\sigma}_{\mathcal{I}},s}(\mathsf{Reach}_{\mathcal{C}^{\sigma}_{\mathcal{I}}}(q,B))$.*

Proof. Let $q \in Q$. We only prove the result for the target configuration $(q,0)$. The argument is the same when the target is (q,B). To lighten notation, we let $\mathcal{M} = \mathcal{M}^{\leq B}(\mathcal{Q})$ for the remainder of the proof.

Let $\bar{h} = s_0 s_1 \dots s_r \in \mathsf{Hist}(\mathcal{C}^{\sigma}_{\mathcal{I}})$ be such that $\mathsf{first}(\bar{h}) = s$ and $\mathsf{last}(\bar{h}) = (q,0)$. We let $\mathcal{H}_{\mathsf{abs}}(\bar{h}) \subseteq \mathsf{Hist}(\mathcal{M})$ be the set of histories abstracted by $\bar{h}$. We show that $\mathbb{P}_{\mathcal{C}^{\sigma}_{\mathcal{I}},s}(\mathsf{Cyl}_{\mathcal{C}^{\sigma}_{\mathcal{I}}}(\bar{h})) = \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Cyl}_{\mathcal{M}}(\mathcal{H}_{\mathsf{abs}}(\bar{h})))$. By construction of the abstraction relation, all elements of $\mathcal{H}_{\mathsf{abs}}(\bar{h})$ are a uniquely defined concatenation of an element of $\mathcal{H}_{\mathsf{succ}}(s_0, s_1)$ with an element of $\mathcal{H}_{\mathsf{succ}}(s_1, s_2)$, ..., with an element of $\mathcal{H}_{\mathsf{succ}}(s_{r-1}, s_r)$. Conversely, any such concatenation is an element of $\mathcal{H}_{\mathsf{abs}}(\bar{h})$. We obtain:

$$\begin{aligned}
&\mathbb{P}_{\mathcal{C}^{\sigma}_{\mathcal{I}},s}(\mathsf{Cyl}_{\mathcal{C}^{\sigma}_{\mathcal{I}}}(\bar{h})) \\
&\quad= \prod_{\ell=0}^{r-1} \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Cyl}_{\mathcal{M}}(\mathcal{H}_{\mathsf{succ}}(s_\ell, s_{\ell+1}))) \\
&\quad= \prod_{\ell=0}^{r-1} \left(\sum_{h_\ell \in \mathcal{H}_{\mathsf{succ}}(s_\ell, s_{\ell+1})} \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Cyl}_{\mathcal{M}}(h_\ell)) \right) \\
&\quad= \sum_{h_0 \in \mathcal{H}_{\mathsf{succ}}(s_0,s_1)} \cdots \sum_{h_{r-1} \in \mathcal{H}_{\mathsf{succ}}(s_{r-1},s_r)} \left(\prod_{\ell=0}^{r-1} \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Cyl}_{\mathcal{M}}(h_\ell)) \right) \\
&\quad= \sum_{h_0 \in \mathcal{H}_{\mathsf{succ}}(s_0,s_1)} \cdots \sum_{h_{r-1} \in \mathcal{H}_{\mathsf{succ}}(s_{r-1},s_r)} \left(\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Cyl}_{\mathcal{M}}(h_0 \cdot \ldots \cdot h_{r-1})) \right) \\
&\quad= \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Cyl}_{\mathcal{M}}(\mathcal{H}_{\mathsf{abs}}(\bar{h}))).
\end{aligned}$$

The first line follows by definition of $\delta^{\sigma}_{\mathcal{I}}$ and the definition of the probability distribution over plays of Markov chains. For the second line, we first observe that for all $\ell \in [\![r-1]\!]$, the set $\mathcal{H}_{\mathsf{succ}}(s_\ell, s_{\ell+1})$ is prefix-free and thus the cylinders of elements of $\mathcal{H}_{\mathsf{succ}}(s_\ell, s_{\ell+1})$ are pairwise disjoint. The third line is a rewriting of the second. The fourth line is obtained by definition of the probability distribution induced by a strategy in an MDP, using the fact that σ is a memoryless strategy. The last line is obtained because $\mathcal{H}_{\mathsf{abs}}(\bar{h})$ is the set of all concatenations occurring in the previous line.

We can now end the argument. Let $\mathcal{H}$ and $\bar{\mathcal{H}}$ denote the set of histories of $\mathcal{M}$ and $\mathcal{C}^{\sigma}_{\mathcal{I}}$ respectively that start in s and end in $(q, 0)$ with only one occurrence of $(q, 0)$. These two sets are prefix-free and we have $\mathcal{H} = \bigcup_{\bar{h}\in\bar{\mathcal{H}}} \mathcal{H}_{\mathsf{abs}}(\bar{h})$. Using the above, we conclude that:

$$\begin{aligned}\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Term}(q)) &= \sum_{h\in\mathcal{H}} \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Cyl}_{\mathcal{M}}(h)) \\ &= \sum_{\bar{h}\in\bar{\mathcal{H}}} \mathbb{P}_{\mathcal{C}^{\sigma}_{\mathcal{I}},s}(\mathsf{Cyl}_{\mathcal{C}^{\sigma}_{\mathcal{I}}}(\bar{h})) \\ &= \mathbb{P}_{\mathcal{C}^{\sigma}_{\mathcal{I}},s}(\mathsf{Reach}((q, 0))).\end{aligned}$$

This is the claim of the theorem. □

We now discuss state-reachability probabilities. Let $T \subseteq Q$ be a target. Transitions of $\mathcal{C}^{\sigma}_{\mathcal{I}}$ group together (possibly infinitely many) transitions of $\mathcal{M}^{\leq B}(\mathcal{Q})$. In particular, this compression may result in some visits to T not being observed in $\mathcal{C}^{\sigma}_{\mathcal{I}}$ despite occurring in $\mathcal{M}^{\leq B}(\mathcal{Q})$. By slightly modifying $\mathcal{Q}$, we can guarantee that all of these visits are witnessed in the new compressed Markov chain.

The idea is to make all target states absorbing with self-loops of weight -1. Formally, we let $\mathcal{Q}' = (Q, A, \delta', w')$ be the OC-MDP defined by letting, for all $q \in Q$ and all $a \in A(q)$, $\delta'(q, a) = \delta(q, a)$ and $w'(q, a) = w(q, a)$ if $q \notin T$ and, otherwise, $\delta'(q, a)(q) = 1$ and $w'(q, a) = -1$. We remark that σ is a well-defined memoryless strategy of $\mathcal{M}^{\leq B}(\mathcal{Q}')$.

By design, any history of $\mathcal{M}^{\leq B}(\mathcal{Q})$ that ends in a configuration in $T \times [\![B]\!]$ and that does not visit this set before is also a history of $\mathcal{M}^{\leq B}(\mathcal{Q}')$. The cylinders of these histories in both MDPs have the same probability under σ, as transitions are the same in states outside of T. This implies that, under σ, the probability of terminating or hitting the counter upper bound in T in $\mathcal{M}^{\leq B}(\mathcal{Q}')$ is equal to the probability of reaching T in $\mathcal{M}^{\leq B}(\mathcal{Q})$. We conclude by Theorem 18.4 that the compressed Markov chain $\mathcal{C}^{\sigma}_{\mathcal{I}}(\mathcal{Q}')$ captures the state-reachability probabilities for the target T in $\mathcal{M}^{\leq B}(\mathcal{Q})$ under σ. We formalise this by the following theorem, in which, for the sake of clarity, we indicate the relevant MDP or Markov chain for each objective.

Theorem 18.5. *Let $T \subseteq Q$. Let $\mathcal{Q}' = (Q, A, \delta', w')$ be defined as above. For all $s \in S_{\mathcal{I}}$, $\mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\mathsf{Reach}_{\mathcal{M}^{\leq B}(\mathcal{Q})}(T)) = \mathbb{P}_{\mathcal{C}^{\sigma}_{\mathcal{I}}(\mathcal{Q}'),s}(\mathsf{Reach}_{\mathcal{C}^{\sigma}_{\mathcal{I}}(\mathcal{Q}')}(T \times \{0, B\}))$.*

Proof. To lighten notation, we let $\mathcal{M} = \mathcal{M}^{\leq B}(\mathcal{Q})$ and $\mathcal{M}' = \mathcal{M}^{\leq B}(\mathcal{Q}')$ for the remainder of the proof. By Theorem 18.4, it suffices to show that $\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Reach}_{\mathcal{M}}(T)) = \mathbb{P}^{\sigma}_{\mathcal{M}',s}(\mathsf{Reach}_{\mathcal{M}'}(T \times \{0, B\}))$.

Let $\mathcal{H} \subseteq \mathsf{Hist}(\mathcal{M})$ be the set of histories $h \in \mathsf{Hist}(\mathcal{M})$ such that $\mathsf{last}(h) \in T \times [\![B]\!]$ and no prior configuration of h is in $T \times [\![B]\!]$. The state-reachability objective $\mathsf{Reach}_{\mathcal{M}}(T)$ can be written as $\mathsf{Cyl}_{\mathcal{M}}(\mathcal{H})$. Since $\mathcal{H}$ is prefix-free, we have $\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Reach}_{\mathcal{M}}(T)) = \sum_{h \in \mathcal{H}} \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Cyl}_{\mathcal{M}}(h))$. Furthermore, for all $h \in \mathcal{H}$, by definition of δ', since no configuration with a state in T occurs along h besides the last one, we have $h \in \mathsf{Hist}(\mathcal{M}')$ and $\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathsf{Cyl}_{\mathcal{M}}(h)) = \mathbb{P}^{\sigma}_{\mathcal{M}',s}(\mathsf{Cyl}_{\mathcal{M}'}(h))$.

To end the proof, it suffices to show that $\mathsf{Cyl}_{\mathcal{M}'}(\mathcal{H}) = \mathsf{Reach}_{\mathcal{M}'}(T \times \{0, B\})$. We show both inclusions. Let $\pi \in \mathsf{Cyl}_{\mathcal{M}'}(\mathcal{H})$. By definition of $\mathcal{H}$, there exists a configuration of π with a state in T. If the counter value of this configuration is B, then we have $\pi \in \mathsf{Reach}_{\mathcal{M}'}(T \times \{0, B\})$. If not, we are guaranteed to have a configuration in $T \times \{0\}$ along π because states of T are absorbing in $\mathcal{Q}'$ and their self-loops have weight -1. Conversely, let $\pi \in \mathsf{Reach}_{\mathcal{M}'}(T \times \{0, B\})$. By definition of the reachability objective, there must be a configuration with a state in T along π. The earliest occurrence of a state of T (regardless of the counter value) witnesses that $\pi \in \mathsf{Cyl}_{\mathcal{M}'}(\mathcal{H})$. This ends the proof. □

18.4 Characterising transition probabilities

Example 18.1 illustrates that the transition probabilities of a compressed Markov chain may require large representations or be irrational. This section presents characterisations of these transition probabilities via equation systems.

For the remainder of this section, we fix an interval $I \in \mathcal{I}$. We present a system characterising the outgoing transition probabilities from configurations of $\mathcal{C}^{\sigma}_{\mathcal{I}}$ with counter value in I. In Section 18.4.1, we assume that I is unbounded, and we handle the bounded case in Section 18.4.2. We also provide bounds on the size of the systems.

18.4.1 Unbounded intervals

We assume that I is an infinite interval and let $b^- = \min I$. This implies that $B = \infty$. We characterise the transition probabilities from the configurations of $S_{\mathcal{I}}$ with counter value b^- via existing results on termination probabilities in one-counter Markov chains.

For any q, $p \in Q$, the transition probability $\delta_{\mathcal{I}}^{\sigma}((q, b^-))((p, b^- - 1))$ can be seen as a termination probability in a one-counter Markov chain. Let τ denote the counter-oblivious strategy $\sigma(\cdot, b^-)$. We consider the one-counter Markov chain $\mathcal{R} = (Q, \delta^{\tau})$, where, for all q, $p \in Q$ and all $u \in \{-1, 0, 1\}$, we let $\delta^{\tau}(q)(p, u) = \sum_{a \in A(q), w(q,a)=u} \tau(q)(a) \cdot \delta(q, a)(p)$.

Let q, $p \in Q$ and let $s = (q, b^-)$. There is a bijection between $h \in \mathcal{H}_{\mathsf{succ}}(s, (p, b^- - 1))$ and the set of histories of $\mathcal{C}^{\leq\infty}(\mathcal{R})$ that start in $(q, 1)$ and end in $(p, 0)$: one omits all actions and subtracts $b^- - 1$ to all counter values in the history. By definition of σ and δ^{τ}, this bijection preserves the probability of cylinders. This implies that $\delta_{\mathcal{I}}^{\sigma}((q, b^-))((p, b^- - 1))$ is exactly the probability, in $\mathcal{C}^{\leq\infty}(\mathcal{R})$, of terminating in p from $(q, 1)$.

It follows that, in our case, we can characterise our transitions probabilities as termination probabilities in one-counter Markov chains. We use the characterisation of termination probabilities in *probabilistic pushdown automata*, a generalisation of one-counter Markov chains in which termination equates to reaching an empty stack, of [KEM06]. We specialise this characterisation to the setting of one-counter Markov chains in the following theorem.

Theorem 18.6 ([KEM06]). *For each $q, p \in Q$, we consider a variable $\langle q \searrow p\rangle$ and the system of equations formed by the equations, for all $q, p \in Q$,*

$$\langle q \searrow p\rangle = \delta^I(q)(p, -1) + \sum_{t \in Q} \delta^I(q)(t, 0) \cdot \langle t \searrow p\rangle + \sum_{t \in Q} \delta^I(q)(t, 1) \cdot \left(\sum_{t' \in Q} \langle t \searrow t'\rangle \cdot \langle t' \searrow p\rangle\right),$$

where $\delta^I(t)(t', u) = \sum_{a \in A(t), w(t,a)=u} \sigma(t, b^-)(a) \cdot \delta(t, a)(t')$ for all t, $t' \in Q$ and *all* $u \in \{-1, 0, 1\}$*. The least non-negative solution of this system is obtained by substituting each variable* $\langle q \searrow p\rangle$ *by* $\delta^{\sigma}_{\mathcal{I}}((q, b^-))((p, b^- - 1))$*.*

The equation system of Theorem 18.6 has one variable of the form $\langle q \searrow p\rangle$ for every two states $q, p \in Q$ and there is one equation per variable. Furthermore, the equations have length polynomial in the sizes of $|A|$ and $|Q|$. Indeed, if we distribute all products in the right-hand sides of the equations to rewrite them as a sum of products, there are at most $|A| \cdot |Q|^2$ products of at most four variables or constants. We obtain the following result.

Lemma 18.7. *The equation system of Theorem 18.6 has* $|Q|^2$ *variables and equations. Its equations have length polynomial in* $|Q|$ *and* $|A|$*.*

18.4.2 Bounded intervals

We now assume that I is bounded. We write $I = [\![b^-, b^+]\!]$ and let $\beta = \log_2(|I| + 1) \in \mathbb{N}_{>0}$. To improve readability, we assume that $b^- = 1$ and $b^+ = 2^\beta - 1$. All results below can be recovered for the general case by adding $b^- - 1$ to the counter values in configurations.

Counter values of I that are kept in $S_{\mathcal{I}}$ can be partitioned in two sets: the set $\{2^\alpha \mid \alpha \in [\![\beta - 1]\!]\}$ of values reachable from b^- and the set $Q \times \{2^\beta - 2^\alpha \mid \alpha \in [\![\beta - 1]\!]\}$ of values reachable from b^+ (in the sense of Figure 18.2). By symmetry of the transition structure of the compressed Markov chain, the outgoing transitions from a configuration $(q, 2^\alpha)$ correspond to outgoing transitions from the configuration $(q, 2^\beta - 2^\alpha)$.

Lemma 18.8. *Let* $q, p \in Q$ *and* $\alpha \in [\![\beta - 1]\!]$*. It holds that* $\delta^{\sigma}_{\mathcal{I}}(q, 2^\alpha)(p, 2^{\alpha+1}) = \delta^{\sigma}_{\mathcal{I}}(q, 2^\beta - 2^\alpha)(p, 2^\beta)$ *and* $\delta^{\sigma}_{\mathcal{I}}(q, 2^\alpha)(p, 0) = \delta^{\sigma}_{\mathcal{I}}(q, 2^\beta - 2^\alpha)(p, 2^\beta - 2^{\alpha+1})$

Proof. We only prove that $\delta^{\sigma}_{\mathcal{I}}(q, 2^\alpha)(p, 2^{\alpha+1}) = \delta^{\sigma}_{\mathcal{I}}(q, 2^\beta - 2^\alpha)(p, 2^\beta)$ as the other case is similar. We define a bijection

$$\mathcal{F} \colon \mathcal{H}_{\mathsf{succ}}((q, 2^\alpha), (p, 2^{\alpha+1})) \to \mathcal{H}_{\mathsf{succ}}((q, 2^\beta - 2^\alpha), (p, 2^\beta))$$

and prove that $\mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}),(q,2^\alpha)}(\mathsf{Cyl}\,(h)) = \mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}),(q,2^\beta - 2^\alpha)}(\mathsf{Cyl}\,(\mathcal{F}(h)))$ for all

$h \in \mathcal{H}_{\mathsf{succ}}((q, 2^\alpha), (p, 2^{\alpha+1}))$. This is sufficient to obtain our result.

Let $h \in \mathcal{H}_{\mathsf{succ}}((q, 2^\alpha), (p, 2^{\alpha+1}))$. We let $\mathcal{F}(h)$ be the history obtained by adding $2^\beta - 2^{\alpha+1}$ to all counter values along h. We must show that $\mathcal{F}(h) \in \mathcal{H}_{\mathsf{succ}}((q, 2^\beta - 2^\alpha), (p, 2^\beta))$. For the first and last configurations, we observe that $2^\alpha + 2^\beta - 2^{\alpha+1} = 2^\beta - 2^\alpha$ and $2^{\alpha+1} + 2^\beta - 2^{\alpha+1} = 2^\beta$. For the other configurations, their counter values are in the interval $[\![1, 2^{\alpha+1} - 1]\!]$, thus their counterparts in $\mathcal{F}(h)$ have a counter value in $[\![2^\beta - 2^{\alpha+1} + 1, 2^\beta - 1]\!]$.

We now establish that

$$\mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}),(q,2^\alpha)}(\mathsf{Cyl}\,(h)) = \mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}),(q,2^\beta-2^\alpha)}(\mathsf{Cyl}\,(\mathcal{F}(h))).$$

Let $h = (q_0, k_0)a_0(q_1, k_1)\dots a_r(q_r, k_r)$. Because σ is memoryless, based on $\mathcal{I}$ and $I \in \mathcal{I}$, we obtain

$$\begin{aligned}\mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}),(q,2^\alpha)}(\mathsf{Cyl}\,(h)) &= \prod_{\ell=0}^{r-1} \delta(q_\ell, a_\ell)(q_{\ell+1}) \cdot \sigma(s_\ell, k_\ell)(a_\ell) \\ &= \prod_{\ell=0}^{r-1} \delta(q_\ell, a_\ell)(q_{\ell+1}) \cdot \sigma(s_\ell, k_\ell + 2^\beta - 2^{\alpha+1})(a_\ell) \\ &= \mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}),(q,2^\beta-2^\alpha)}(\mathsf{Cyl}\,(\mathcal{F}(h))).\end{aligned}$$

To prove that $\mathcal{F}$ is bijective, we define its inverse. We let $\mathcal{F}^{-1}$ be the function over $\mathcal{H}_{\mathsf{succ}}((q, 2^\beta - 2^\alpha), (p, 2^\beta))$ that subtracts $2^\beta - 2^{\alpha+1}$ to all counter values along histories. It is easy to verify that $\mathcal{F}^{-1}$ is well-defined and that it is the inverse of $\mathcal{F}$. □

Due to Lemma 18.8, it is sufficient for us to characterise the outgoing transition probabilities for the configurations in $Q \times \{2^\alpha \mid \alpha \in [\![\beta - 1]\!]\}$. We do so via a quadratic system of equations. We provide intuition on how to derive this system for our interval $I = [\![1, 2^\beta - 1]\!]$ by using Markov chains: our systems can be derived from linear systems for reachability probabilities in the Markov chains illustrated below. We recall the general form of these linear systems in Appendix A.2.1.

Let us first consider transitions from $Q \times \{1\}$. We illustrate the situation in Figure 18.3: we consider a Markov chain over $Q \times \{0, 1, 2\}$ where states in $Q \times \{0, 2\}$ are absorbing and transitions from other states are inherited

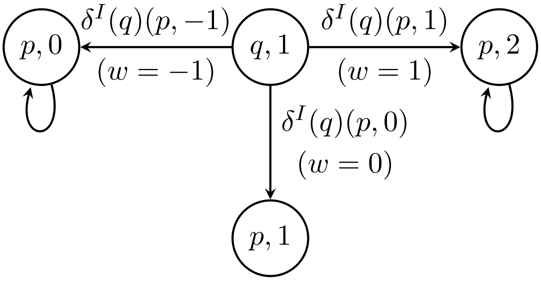


Figure 18.3: Markov chain transition scheme used to derive a characterisation of transitions from $Q \times \{1\}$ in $\mathcal{C}^{\sigma}_{\mathcal{I}}$ for a bounded interval of the form $[\![1, 2^{\beta} - 1]\!]$. In this figure, $\delta^I(q)(p,u) = \sum_{\substack{a \in A(q) \\ w(q,a)=u}} \sigma((q,1))(a) \cdot \delta(q,a)(p)$.

from the Markov chain induced by σ on $\mathcal{M}^{\leq B}(\mathcal{Q})$. We represent transitions in this Markov chain from a configuration $(q,1) \in Q \times \{1\}$ to configurations with a state $p \in Q$. For any $q \in Q$, the probability of reaching a configuration $s' \in Q \times \{0,2\}$ from $(q,1)$ in this Markov chain is $\delta^{\sigma}_{\mathcal{I}}((q,1))(s')$ by definition.

Next, we let $\alpha \in [\![1, \beta - 1]\!]$ and consider configurations in $Q \times \{2^{\alpha}\}$. The situation is depicted in Fig. 18.4. We divide a counter change by 2^{α} into counter changes by $2^{\alpha-1}$ and, thus, rely on the transition probabilities from $Q \times \{2^{\alpha-1}\}$ in $\mathcal{C}^{\sigma}_{\mathcal{I}}$. In this case, we can see transition probabilities from $Q \times \{2^{\alpha}\}$ in $\mathcal{C}^{\sigma}_{\mathcal{I}}$ as reachability probabilities in a Markov chain over $Q \times \{0, 2^{\alpha-1}, 2^{\alpha}, 3 \cdot 2^{\alpha-1}, 2^{\alpha+1}\}$.

By putting together the reachability systems for $Q \times \{2^{\alpha}\}$ for all $\alpha \in [\![\beta - 1]\!]$, we obtain a quadratic system of equations. To formalise our system and prove its validity, we introduce some notation.

Let $\alpha \in [\![\beta - 1]\!]$, $q, p \in Q$ and $k \in [\![1, 2^{\alpha+1} - 1]\!]$. We let $\mathcal{H}_{\alpha}((q,k) \nearrow p)$ (resp. $\mathcal{H}_{\alpha}((q,k) \searrow p)$) denote the set of histories h of $\mathcal{M}^{\leq B}(\mathcal{Q})$ such that $\mathsf{first}(h) = (q,k)$, $\mathsf{last}(h) = (p, 2^{\alpha+1})$ (resp. $(p,0)$) and no configuration along h besides its last one has a counter value in $\{0, 2^{\alpha+1}\}$. These sets are prefix-free. We let $[(q,k) \nearrow p]_{\alpha} = \mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}),(q,k)}(\mathsf{Cyl}\,(\mathcal{H}_{\alpha}((q,k) \nearrow p)))$ and $[(q,k) \searrow p]_{\alpha} = \mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}),(q,k)}(\mathsf{Cyl}\,(\mathcal{H}_{\alpha}((q,k) \searrow p)))$. In all of this notation, if α is the subscript, then an upwards (resp. downwards) arrow indicates that the counter of the target configuration is $2^{\alpha+1}$ (resp. 0).

The transition probabilities of $\mathcal{C}^{\sigma}_{\mathcal{I}}$ can be written with the above notation. For all $q, p \in Q$ and $\alpha \in [\![\beta - 1]\!]$, we have $\delta^{\sigma}_{\mathcal{I}}((q, 2^{\alpha}))((p,0)) = [(q, 2^{\alpha}) \searrow p]_{\alpha}$

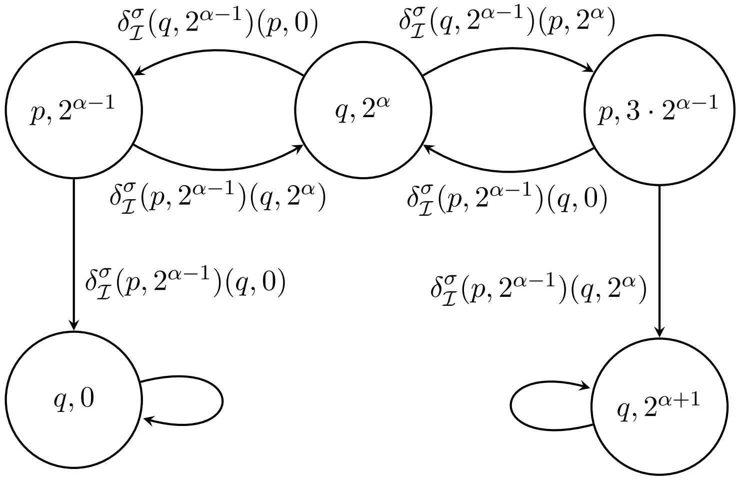

$\delta^\sigma_\mathcal{I}(q, 2^{\alpha-1})(p, 0)$
$\delta^\sigma_\mathcal{I}(q, 2^{\alpha-1})(p, 2^\alpha)$
$p, 2^{\alpha-1}$
$q, 2^\alpha$
$p, 3 \cdot 2^{\alpha-1}$
$\delta^\sigma_\mathcal{I}(p, 2^{\alpha-1})(q, 2^\alpha)$
$\delta^\sigma_\mathcal{I}(p, 2^{\alpha-1})(q, 0)$
$\delta^\sigma_\mathcal{I}(p, 2^{\alpha-1})(q, 0)$
$\delta^\sigma_\mathcal{I}(p, 2^{\alpha-1})(q, 2^\alpha)$
$q, 0$
$q, 2^{\alpha+1}$

Figure 18.4: Markov chain transition scheme used to derive a characterisation of transitions from $Q \times \{2^\alpha\}$ for $0 < \alpha < \beta$ in $\mathcal{C}^\sigma_\mathcal{I}$ for a bounded interval of the form $[\![1, 2^\beta - 1]\!]$.

and $\delta^\sigma_\mathcal{I}((q, 2^\alpha))((p, 2^{\alpha+1})) = [(q, 2^\alpha) \nearrow p]_\alpha$.

The following theorem formalises our characterisation of the transition probabilities of $\mathcal{C}^\sigma_\mathcal{I}$ for configurations in $S_\mathcal{I} \cap (Q \times I)$. The size of this system is polynomial in $|Q|$ and β. We provide a self-contained proof that does not refer to the Markov chains described in Figures 18.3 and 18.4. This proof is inspired from the reasoning used to establish Theorem 18.6 in [KEM06]. A corollary of this proof is that the Markov chains above yield an accurate characterisation of the transition probabilities.

Theorem 18.9. *For each $q, p \in Q$, we consider variables $\langle (q, 1) \nearrow p\rangle_0$ and $\langle (q, 1) \searrow p\rangle_0$, and for all $\alpha \in [\![1, \beta - 1]\!]$ and $k \in \{2^{\alpha-1}, 2^\alpha, 3 \cdot 2^{\alpha-1}\}$, we consider variables $\langle (q, k) \nearrow p\rangle_\alpha$ and $\langle (q, k) \searrow p\rangle_\alpha$. For all q, $p \in Q$ and all $u \in \{-1, 0, 1\}$, let $\delta^I(q)(p, u) = \sum_{a \in A(q), w(q,a) = u} \sigma(q, 1)(a) \cdot \delta(q, a)(p)$.*

Consider the system of equations given by, for all $q, p \in Q$:

$$\langle (q,1) \nearrow p\rangle_0 = \delta^I(q)(p,1) + \sum_{t\in Q} \delta^I(q)(t,0) \cdot \langle (t,1) \nearrow p\rangle_0 \tag{18.1}$$

$$\langle (q,1) \searrow p\rangle_0 = \delta^I(q)(p,-1) + \sum_{t\in Q} \delta^I(q)(t,0) \cdot \langle (t,1) \searrow p\rangle_0$$

and for all $\alpha \in [\![1, \beta-1]\!]$,

$$\langle (q,2^{\alpha-1}) \nearrow p\rangle_\alpha = \sum_{t\in Q} \langle (q,2^{\alpha-1}) \nearrow t\rangle_{\alpha-1} \cdot \langle (t,2^{\alpha}) \nearrow p\rangle_\alpha, \tag{18.2}$$

$$\begin{aligned}\langle (q,2^{\alpha}) \nearrow p\rangle_\alpha = \sum_{t\in Q} \Big(&\langle (q,2^{\alpha-1}) \nearrow t\rangle_{\alpha-1} \cdot \langle (t,3\cdot 2^{\alpha-1}) \nearrow p\rangle_\alpha \\ &+ \langle (q,2^{\alpha-1}) \searrow t\rangle_{\alpha-1} \cdot \langle (t,2^{\alpha-1}) \nearrow p\rangle_\alpha \Big),\end{aligned} \tag{18.3}$$

$$\begin{aligned}\langle (q,3\cdot 2^{\alpha-1}) \nearrow p\rangle_\alpha = &\sum_{t\in Q} \Big(\langle (q,2^{\alpha-1}) \searrow t\rangle_{\alpha-1} \cdot \langle (t,2^{\alpha}) \nearrow p\rangle_\alpha \Big) \\ &+ \langle (q,2^{\alpha-1}) \nearrow p\rangle_{\alpha-1},\end{aligned} \tag{18.4}$$

$$\begin{aligned}\langle (q,3\cdot 2^{\alpha-1}) \searrow p\rangle_\alpha =& \sum_{t\in Q} \langle (q,2^{\alpha-1}) \searrow t\rangle_{\alpha-1} \cdot \langle (t,2^{\alpha}) \searrow p\rangle_\alpha, \\ \langle (q,2^{\alpha}) \searrow p\rangle_\alpha =& \sum_{t\in Q} \Big(\langle (q,2^{\alpha-1}) \searrow t\rangle_{\alpha-1} \cdot \langle (t,2^{\alpha-1}) \searrow p\rangle_\alpha \\ &+ \langle (q,2^{\alpha-1}) \nearrow t\rangle_{\alpha-1} \cdot \langle (t,3\cdot 2^{\alpha-1}) \searrow p\rangle_\alpha \Big), \\ \langle (q,2^{\alpha-1}) \searrow p\rangle_\alpha =& \sum_{t\in Q} \Big(\langle (q,2^{\alpha-1}) \nearrow t\rangle_{\alpha-1} \cdot \langle (t,2^{\alpha}) \searrow p\rangle_\alpha \Big) \\ &+ \langle (q,2^{\alpha-1}) \searrow p\rangle_{\alpha-1}.\end{aligned}$$

The least non-negative solution of this system is obtained by substituting each variable $\langle (q,k) \nearrow p\rangle_\alpha$ by $[(q,k) \nearrow p]_\alpha$ and $\langle (q,k) \searrow p\rangle_\alpha$ by $[(q,k) \searrow p]_\alpha$.

Proof. Occurrences of $\mathbb{P}$ in this proof refer to $\mathcal{M}^{\leq B}(\mathcal{Q})$, thus we omit it from the notation. We show a claim to shorten our arguments. Let s, $s' \in Q \times [\![B]\!]$. Let $\mathcal{H} \subseteq \mathsf{Hist}(\mathcal{M}^{\leq B}(\mathcal{Q}))$ be a prefix-free set of histories starting in s. Assume that there exist two prefix-free sets of histories $\mathcal{H}^{(1)}$ and $\mathcal{H}^{(2)}$ such that the last (resp. first) configuration of all elements of $\mathcal{H}^{(1)}$ (resp. $\mathcal{H}^{(2)}$) is s' and we have $\mathcal{H} = \{h_1 \cdot h_2 \mid h_1 \in \mathcal{H}^{(1)}, h_2 \in \mathcal{H}^{(2)}\}$. Then it holds that

$$\mathbb{P}^\sigma_s(\mathsf{Cyl}\,(\mathcal{H})) = \mathbb{P}^\sigma_s\left(\mathsf{Cyl}\left(\mathcal{H}^{(1)}\right)\right) \cdot \mathbb{P}^\sigma_{s'}\left(\mathsf{Cyl}\left(\mathcal{H}^{(2)}\right)\right). \tag{18.5}$$

Equation (18.5) can be proven as follows:

$$\begin{aligned}
\mathbb{P}^\sigma_s(\mathsf{Cyl}\,(\mathcal{H})) &= \sum_{h_1\in\mathcal{H}^{(1)}} \sum_{h_2\in\mathcal{H}^{(2)}} \mathbb{P}^\sigma_s(\mathsf{Cyl}\,(h_1 \cdot h_2)) \\
&= \sum_{h_1\in\mathcal{H}^{(1)}} \sum_{h_2\in\mathcal{H}^{(2)}} \mathbb{P}^\sigma_s(h_1) \cdot \mathbb{P}^\sigma_{s'}(h_2) \\
&= \left(\sum_{h_1\in\mathcal{H}^{(1)}} \mathbb{P}^\sigma_s(\mathsf{Cyl}\,(h_1))\right) \cdot \left(\sum_{h_2\in\mathcal{H}^{(2)}} \mathbb{P}^\sigma_s(\mathsf{Cyl}\,(h_2))\right) \\
&= \mathbb{P}^\sigma_s(\mathsf{Cyl}\left(\mathcal{H}^{(1)}\right)) \cdot \mathbb{P}^\sigma_{s'}(\mathsf{Cyl}\left(\mathcal{H}^{(2)}\right))
\end{aligned}$$

The first line follows from $\mathcal{H}$ being prefix-free. The second line is obtained from the definition of $\mathbb{P}^\sigma_s$, using the fact that σ is memoryless. We obtain the third line by algebraic manipulations and the last one using the fact that $\mathcal{H}^{(1)}$ and $\mathcal{H}^{(2)}$ are prefix-free.

We now prove the theorem. We start by proving that the asserted solution is a solution of the system. We only verify Equations (18.1)–(18.4), i.e., the equations in which the left-hand side of the equation has a variable with an upwards arrow $\nearrow$. Arguments for the others are analogous.

Let $q, p \in Q$. First, we consider the case $\alpha = 0$. We recall that $\mathcal{H}_0((q,1) \nearrow p)$ is prefix-free. We partition $\mathcal{H}_0((q,1) \nearrow p)$ into two sets $\mathcal{H}$ and $\mathcal{H}'$ such that $\mathcal{H}$ is the set of histories starting in $(q,1)$ whose second configuration is $(p,0)$ and $\mathcal{H}' = \mathcal{H}_0((q,1) \nearrow p) \setminus \mathcal{H}$. For all histories of $\mathcal{H}'$, their second configuration has counter value 1. We rewrite $\mathbb{P}^\sigma_{(q,1)}(\mathsf{Cyl}\,(\mathcal{H}))$ and $\mathbb{P}^\sigma_{(q,1)}(\mathsf{Cyl}\,(\mathcal{H}'))$ to prove the desired equality. On the one hand, we have

$$\mathbb{P}^\sigma_{(q,1)}(\mathsf{Cyl}\,(\mathcal{H})) = \sum_{\substack{a\in A(q)\\ w(q,a)=1}} \mathbb{P}^\sigma_{(q,1)}(\mathsf{Cyl}\,((q,1)a(p,2))) = \delta^I(q,1)(p).$$

For the other set, we partition $\mathcal{H}'$ according to the second configuration of the histories. We further partition the resulting sets following the first action and apply Equation (18.5) to obtain

$$\begin{aligned}
&\mathbb{P}^{\sigma}_{(q,1)}(\mathsf{Cyl}\,(\mathcal{H}')) \\
&\quad= \sum_{t\in Q}\sum_{\substack{a\in A(q)\\ w(q,a)=0}}\sum_{h\in\mathcal{H}_0((t,1)\nearrow p)} \mathbb{P}^{\sigma}_{(q,1)}(\mathsf{Cyl}\,((q,1)a(t,1)\cdot h)) \\
&\quad= \sum_{t\in Q}\Bigg(\sum_{\substack{a\in A(q)\\ w(q,a)=0}} \sigma(q,1)(a)\cdot\delta(q,a)(t)\Bigg)\cdot \mathbb{P}^{\sigma}_{(t,1)}(\mathsf{Cyl}\,(\mathcal{H}_0((t,1)\nearrow p))) \\
&\quad= \sum_{t\in Q}\delta^I(q,0)(t)\cdot[(t,1)\nearrow p]_0.
\end{aligned}$$

Equation (18.1) thus follows from the above and

$$[(q,1)\nearrow p]_0 = \mathbb{P}^{\sigma}_{(q,1)}(\mathsf{Cyl}\,(\mathcal{H})) + \mathbb{P}^{\sigma}_{(q,1)}(\mathsf{Cyl}\,(\mathcal{H}')).$$

This ends the case where $\alpha = 0$.

Let $\alpha \geq 1$. We start by considering Equation (18.2). All histories in $\mathcal{H}_\alpha((q,2^{\alpha-1})\nearrow p)$ have a configuration with counter value 2^α. We let $(U_t)_{t\in Q}$ be a partition of $\mathcal{H}_\alpha((q,2^{\alpha-1})\nearrow p)$ based on the state of the first configuration with counter value 2^α that is reached. For all $t\in Q$ and all $h\in U_t$, we let h_1 and h_2 such that $h = h_1\cdot h_2$ where h_1 is the prefix of h up to the first occurrence of $(t,2^\alpha)$, and let, for $i\in\{1,2\}$, $U_t^{(i)} = \{h_i \mid h_1\cdot h_2\in U_t\}$. We have $U_t^{(1)} = \mathcal{H}_{\alpha-1}((q,2^{\alpha-1})\nearrow t)$ and $U_t^{(2)} = \mathcal{H}_\alpha((q,2^\alpha)\nearrow t)$ by construction. We conclude that Equation (18.2) is satisfied by the candidate solution via the following equations (the second line uses Equation (18.5)):

$$\begin{aligned}
[(q,2^{\alpha-1})\nearrow p]_\alpha &= \sum_{t\in Q}\mathbb{P}^{\sigma}_{(q,2^{\alpha-1})}\left(\mathsf{Cyl}\,(U_t)\right) \\
&= \sum_{t\in Q}\mathbb{P}^{\sigma}_{(q,2^{\alpha-1})}\left(\mathsf{Cyl}\left(U_t^{(1)}\right)\right)\cdot\mathbb{P}^{\sigma}_{(t,2^\alpha)}\left(\mathsf{Cyl}\left(U_t^{(2)}\right)\right) \\
&= \sum_{t\in Q}[(q,2^{\alpha-1})\nearrow t]_{\alpha-1}\cdot[(t,2^\alpha)\nearrow p]_\alpha.
\end{aligned}$$

We now move on to Equation (18.3). We partition $\mathcal{H}_\alpha((q,2^\alpha)\nearrow p)$ as follows. Let $t\in Q$. We let U_t (resp. D_t) denote the subset of $\mathcal{H}_\alpha((q,2^\alpha)\nearrow p)$

containing the histories such that the first configuration with counter value $3 \cdot 2^{\alpha-1}$ (resp. $2^{\alpha-1}$) that is visited has state t and no prior configuration has a counter value in $\{3 \cdot 2^{\alpha-1}, 2^{\alpha-1}\}$. The sets D_t and U_t, $t \in Q$ partition, $\mathcal{H}_\alpha((q, 2^\alpha) \nearrow p)$. Indeed, all of these sets are disjoint by definition and any history from $(q, 2^\alpha)$ to $(p, 2^{\alpha+1})$ must traverse a configuration with counter value $3 \cdot 2^{\alpha-1}$. Similarly to above (for Equation (18.2)), for all $t \in Q$ and all $h \in U_t$ (resp. D_t), we let $h = h_1 \cdot h_2$ such that h_1 ends in the configuration witnessing that $h \in U_t$ (resp. D_t). For $i \in \{1, 2\}$, we let $U_t^{(i)} = \{h_i \mid h_1 \cdot h_2 \in U_t\}$ and $D_t^{(i)} = \{h_i \mid h_1 \cdot h_2 \in D_t\}$. By applying Equation (18.5), we obtain:

$$\begin{aligned}[(q, 2^\alpha) \nearrow p]_\alpha = &\sum_{t \in Q} \mathbb{P}^\sigma_{(q,2^\alpha)}\left(\mathsf{Cyl}\left(U_t^{(1)}\right)\right) \cdot \mathbb{P}^\sigma_{(t,3\cdot 2^{\alpha-1})}\left(\mathsf{Cyl}\left(U_t^{(2)}\right)\right) \\ &+ \sum_{t \in Q} \mathbb{P}^\sigma_{(q,2^\alpha)}\left(\mathsf{Cyl}\left(D_t^{(1)}\right)\right) \cdot \mathbb{P}^\sigma_{(t,2^{\alpha-1})}\left(\mathsf{Cyl}\left(D_t^{(2)}\right)\right)\end{aligned}$$

We now prove that the cylinder probabilities match the terms in Equation (18.3). Let $t \in Q$. We have $\mathbb{P}^\sigma_{(t,3\cdot 2^{\alpha-1})}\left(\mathsf{Cyl}\left(U_t^{(2)}\right)\right) = [(t, 3 \cdot 2^{\alpha-1}) \nearrow p]_\alpha$ and $\mathbb{P}^\sigma_{(t,2^{\alpha-1})}\left(\mathsf{Cyl}\left(D_t^{(2)}\right)\right) = [(t, 2^{\alpha-1}) \nearrow p]_\alpha$ because $U_t^{(2)}$ and $D_t^{(2)}$ are respectively the sets $\mathcal{H}_\alpha((t, 3 \cdot 2^{\alpha-1}) \nearrow p)$ and $\mathcal{H}_\alpha((t, 2^{\alpha-1}) \nearrow p)$.

The sets $U_t^{(1)}$ and $D_t^{(1)}$ do not directly match relevant sets of histories as above. However, there are bijections from $U_t^{(1)}$ to $\mathcal{H}_{\alpha-1}((q, 2^{\alpha-1}) \nearrow t)$ and from $D_t^{(1)}$ to $\mathcal{H}_{\alpha-1}((q, 2^{\alpha-1}) \searrow t)$. Both bijections map a history to the history obtained by subtracting $2^{\alpha-1}$ to the counter values in all configurations along the history. All counter values in a history in $U_t^{(1)}$ or $D_t^{(1)}$ and its image lies in the interval I. Therefore, for all $h_1 \in U_t^{(1)} \cup D_t^{(1)}$ with $\mathsf{first}(h_1) = s$, if its image by the relevant bijection is h_1' such that $\mathsf{first}(h_1') = s'$, then $\mathbb{P}^\sigma_s(\mathsf{Cyl}(h_1)) = \mathbb{P}^\sigma_{s'}(\mathsf{Cyl}(h_1'))$. We conclude that $\mathbb{P}^\sigma_{(q,2^\alpha)}\left(\mathsf{Cyl}\left(U_t^{(1)}\right)\right) = [(q, 2^{\alpha-1}) \nearrow t]_{\alpha-1}$ and $\mathbb{P}^\sigma_{(q,2^\alpha)}\left(\mathsf{Cyl}\left(D_t^{(1)}\right)\right) = [(q, 2^{\alpha-1}) \searrow t]_{\alpha-1}$ (a similar argument is more detailed in the proof of Lemma 18.8). We have shown that the asserted solution satisfies Equation (18.3).

We now move on to Equation (18.4). We follow the same scheme as above, i.e., we partition $\mathcal{H}_\alpha((q, 3 \cdot 2^{\alpha-1}) \nearrow p)$. First, we let U_p be the subset with all histories that never hit counter value 2^α. For any $t \in Q$, we let D_t be the subset of $\mathcal{H}_\alpha((q, 3 \cdot 2^{\alpha-1}) \nearrow p) \setminus U_p$ consisting of histories that reach counter value 2^α for the first time in a configuration with state t. The sets D_t, $t \in Q$, and U_p

partition $\mathcal{H}_\alpha((q, 3 \cdot 2^{\alpha-1}) \nearrow p)$. As above, for any $t \in Q$ and $h \in D_t$, we write $h = h_1 \cdot h_2$ such that h_1 is the prefix of h up to the first occurrence of $(t, 2^\alpha)$. For $i \in \{1, 2\}$, we let $D_t^{(i)} = \{h_i \mid h_1 \cdot h_2 \in D_t\}$. Like before, we obtain, from Equation (18.5),

$$[(q, 3 \cdot 2^{\alpha-1}) \nearrow p]_\alpha = \sum_{t \in Q} \mathbb{P}^\sigma_{(q,3\cdot 2^{\alpha-1})}\left(\mathsf{Cyl}\left(D_t^{(1)}\right)\right) \cdot \mathbb{P}^\sigma_{(t,2^\alpha)}\left(\mathsf{Cyl}\left(D_t^{(2)}\right)\right) + \mathbb{P}^\sigma_{(q,3\cdot 2^\alpha)}\left(\mathsf{Cyl}\left(U_p\right)\right).$$

Let $t \in Q$. It follows from $D_t^{(2)} = \mathcal{H}_\alpha((t, 2^\alpha) \nearrow p)$ that $\mathbb{P}^\sigma_{(t,2^\alpha)}\left(\mathsf{Cyl}\left(D_t^{(2)}\right)\right) = [(t, 2^\alpha) \nearrow p]_\alpha$. We can adapt the bijection-based argument used for Equation (18.3) to conclude that $\mathbb{P}^\sigma_{(q,3\cdot 2^{\alpha-1})}\left(\mathsf{Cyl}\left(D_t^{(1)}\right)\right) = [(q, 2^{\alpha-1}) \searrow t]_{\alpha-1}$ and $\mathbb{P}^\sigma_{(q,3\cdot 2^{\alpha-1})}\left(\mathsf{Cyl}\left(U_p\right)\right) = [(q, 2^{\alpha-1}) \nearrow p]_{\alpha-1}$. This shows that Equation (18.4) is verified by the asserted solution, and ends the argument that all equations hold.

It remains to show that the asserted solution is the least non-negative solution of the system. Once again, we only consider the case of variables with ascending arrows as the other case can be handled similarly. We fix an arbitrary non-negative solution of the system. We denote its component corresponding to a variable x by $x^\star$.

All probabilities in the asserted solution can be written as the probability of a cylinder of a set of histories. In particular, these probabilities can be approximated by only considering the histories with at most r actions (for $r \in \mathbb{N}$). It suffices therefore to show that each approximation is lesser or equal to the fixed arbitrary solution to end the proof.

For all $r \in \mathbb{N}$, $q, p \in Q$, $\alpha \in [\![\beta - 1]\!]$ and $k \in \{2^{\alpha-1}, 2^\alpha, 3 \cdot 2^{\alpha-1}\}$ if $\alpha \neq 0$ and $k = 1$ otherwise, we let $[(q, k) \nearrow p]_\alpha^{\leq r} = \mathbb{P}^\sigma_{(q,k)}(\mathsf{Cyl}\left(\mathcal{H}^{\leq r}\right))$ where $\mathcal{H}^{\leq r}$ is the subset of $\mathcal{H}_\alpha((q, k) \nearrow p)$ containing all histories with at most r actions. We define $[(q, k) \searrow p]_\alpha^{\leq r}$ similarly.

Let q and $p \in Q$. We use nested induction arguments in the remainder of the proof: an outer induction on α and an inner induction on r.

First, we deal with the case $\alpha = 0$. Let $r \in \mathbb{N}$. For the base case $r = 0$, we have $[(q, 1) \nearrow p]_0^{\leq 0} = 0 \leq \langle (q, 1) \nearrow p \rangle_0^\star$ because we consider a non-negative solution. We now assume that $[(q, 1) \nearrow p]_0^{\leq r-1} \leq \langle (q, 1) \nearrow p \rangle_0^\star$ by induction. We can apply the reasoning used when considering Equation (18.1) in the first part of the proof (taking in account the length of histories) and then apply the induction hypothesis to obtain:

$$\begin{aligned}[(q,1)\nearrow p]_0^{\leq r} &= \delta^I(q)(p,1) + \sum_{t\in Q}\delta^I(q)(t,0)\cdot[(t,1)\nearrow p]_0^{\leq r-1}\\ &\leq \delta^I(q)(p,1) + \sum_{t\in Q}\delta^I(q)(t,0)\cdot\langle(t,1)\nearrow p\rangle_0^\star\\ &= \langle(q,1)\nearrow p\rangle_0^\star.\end{aligned}$$

This closes the proof for the case $\alpha = 0$.

Next, let $\alpha \geq 1$. We assume, by induction on α, that we have shown that for all $t, t' \in Q$, we have $[(t,2^{\alpha-1})\nearrow t']_{\alpha-1} \leq \langle(t,2^{\alpha-1})\nearrow t'\rangle_{\alpha-1}^\star$ and $[(t,2^{\alpha-1})\searrow t']_{\alpha-1} \leq \langle(t,2^{\alpha-1})\searrow t'\rangle_{\alpha-1}^\star$. The base case $r = 0$ of the inner induction is direct because for all $k \in \{2^{\alpha-1}, 2^\alpha, 3\cdot 2^{\alpha-1}\}$, we have $[(q,k)\nearrow p]_\alpha^{\leq 0} = 0$.

We assume by induction that $[(t,k)\nearrow t']_\alpha^{\leq r} \leq \langle(t,k)\nearrow t'\rangle_\alpha^\star$ for all $t, t' \in Q$ and all $k \in \{2^{\alpha-1}, 2^\alpha, 3\cdot 2^{\alpha-1}\}$. All required inequalities are obtained by an adaptation of the argument used in the first part of the proof for Equations (18.2), (18.3) and (18.4), i.e., partitioning the set of histories while taking in account the length of histories and invoking Equation (18.5). For this reason, we omit some details. From configuration $(q, 2^{\alpha-1})$, we obtain that

$$[(q,2^{\alpha-1})\nearrow p]_\alpha^{\leq r} \leq \sum_{t\in Q}[(q,2^{\alpha-1})\nearrow t]_{\alpha-1}\cdot[(t,2^\alpha)\nearrow p]_\alpha^{\leq r-1}.$$

By the induction hypotheses and the fact we are dealing with a solution of the system, we obtain $[(q,2^{\alpha-1})\nearrow p]_\alpha^{\leq r} \leq \langle(q,2^{\alpha-1})\nearrow p\rangle_\alpha^\star$. Next, for configuration $(q, 2^\alpha)$, we obtain that

$$\begin{aligned}[(q,2^\alpha)\nearrow p]_\alpha^{\leq r} \leq \sum_{t\in Q}\Big(&[(q,2^{\alpha-1})\nearrow t]_{\alpha-1}\cdot[(t,3\cdot 2^{\alpha-1})\nearrow p]_\alpha^{\leq r-1}\\ &+[(q,2^{\alpha-1})\searrow t]_{\alpha-1}\cdot[(t,2^{\alpha-1})\nearrow p]_\alpha^{\leq r-1}\Big).\end{aligned}$$

It follows from the induction hypotheses and the fact we deal with a solution that $[(q,2^\alpha)\nearrow p]_\alpha^{\leq r} \leq \langle(q,2^\alpha)\nearrow p\rangle_\alpha^\star$. Finally, for configuration $(q, 3\cdot 2^{\alpha-1})$, we have

$$\begin{aligned}[(q,3\cdot 2^{\alpha-1})\nearrow p]_\alpha^{\leq r} \leq &\sum_{t\in Q}\Big([(q,2^{\alpha-1})\searrow t]_{\alpha-1}\cdot[(t,2^\alpha)\nearrow p]_\alpha^{\leq r-1}\Big)\\ &+[(q,2^{\alpha-1})\nearrow p]_{\alpha-1}.\end{aligned}$$

The induction hypotheses imply that $[(q, 3 \cdot 2^{\alpha-1}) \nearrow p]_{\alpha}^{\leq r} \leq \langle(q, 3 \cdot 2^{\alpha-1}) \nearrow p\rangle_{\alpha}^{\star}$.

We have shown that the asserted solution is the least non-negative solution of the system. □

We now analyse the size of the equation system of Theorem 18.9. There are as many equations as there are variables. There are $2 \cdot |Q|^2$ equations in the system for which the variable of the left-hand side is indexed by 0, and, for all $\alpha \in [\![1, \beta - 1]\!]$, there are $6 \cdot |Q|^2$ equations in the system for which the variable of the left-hand side is indexed by α. We can also show that these equations have length polynomial in $|Q|$ and $|A|$. We obtain the following result.

Lemma 18.10. *The equation system of Theorem 18.9 has $2 \cdot |Q|^2 \cdot (3\beta - 2)$ variables and equations. Its equations have length polynomial in $|Q|$ and $|A|$.*

Proof. The argument regarding the number of variables and equations is given above. We thus provide an analysis of the length of the equations. We analyse Equations (18.1)–(18.4) from Theorem 18.9. A similar analysis applies to the other equations. We need only comment on the right-hand side of each equation, as the left-hand side contains a single variable.

We start with Equation (18.1). If we rewrite its right-hand side as a sum of products (we substitute references to δ^I by the corresponding sum), we obtain a sum of no more than $|Q| \cdot |A|$ products of at most three variables or constants. For Equations (18.2)–(18.4), we observe that their right-hand side are respectively sums of no more than $2|Q|$ products of two variables. □

Theorem 18.9 provides a system of equations that may not have a unique solution. We describe how to alter this system to have a unique solution based on the supports of the distributions assigned by σ.

We rely on the Markov chains described in Figures 18.3 and 18.4. By Theorem 18.9, the transition probabilities of $\mathcal{C}_{\mathcal{I}}^{\sigma}$ are reachability probabilities in these Markov chains. More precisely, the system of Theorem 18.9 is a collection of systems for reachability probabilities in these Markov chains. It follows that modifying the equation system of Theorem 18.9 by setting all relevant probabilities to zero will ensure uniqueness of the solution.

It remains to determine how to identify the probabilities that are zero in the least solution of the system. The probabilities that are zero only depend on the transitions (with non-zero probability) between configurations (see Appendix A.2.1). Therefore, we need not compute the transition probabilities of the Markov chains (which would have an important computational cost, see Example 18.1) and need only determine the transitions qualitatively.

The idea of the procedure is to proceed gradually increasing the counter step size. First, we can study the Markov chain for counter values $\{0, 1, 2\}$, as illustrated in Figure 18.3, and perform a graph-based analysis to determine which probabilities to set to zero for outgoing transitions from $Q \times \{1\}$ in $\mathcal{C}^{\sigma}_{\mathcal{I}}$. Then, for all $\alpha \in [\![\beta - 1]\!]$, assuming that the non-zero transition probabilities in $\mathcal{C}^{\sigma}_{\mathcal{I}}$ have been determined for configurations in $Q \times \{2^{\alpha-1}\}$, we can perform another graph-based analysis on the Markov chain described in Figure 18.4 to determine the non-zero transition probabilities from $Q \times \{2^{\alpha}\}$ in $\mathcal{C}^{\sigma}_{\mathcal{I}}$.

In this way, we obtain a procedure that runs in time polynomial in $|Q|$ and β: we perform a reachability analysis on one graph of size $3 \cdot |Q|$ for the base case and on $\beta - 1$ graphs of size $5 \cdot |Q|$ for the other cases. This analysis does not require the precise probabilities given by σ, and it is sufficient to only know which actions are chosen with positive probabilities in $Q \times I$. When given the precise probabilities, the system resulting from this procedure can be solved in polynomial time in the BSS model; by construction, its unique solution can be computed by solving β linear systems. We summarise this result in the following theorem.

Theorem 18.11. *There exists an algorithm modifying the system of Theorem 18.9 such that*

(i) the least solution of the original system is the unique solution of the modified one and

(ii) the algorithm runs in time polynomial in β and the representation size of $\mathcal{Q}$.

This algorithms only relies on the support of the distributions in the image of σ and not the precise probabilities. The resulting system can be solved in polynomial time in the BSS model.

Proof. We first formalise the Markov chains of Figures 18.3 and 18.4. The remainder of our argument is based on these β Markov chains.

We let $\mathcal{C}_0 = (\{\bot\} \cup (Q \times \{0, 1, 2\}), \delta^I_0)$ be the Markov chain such that all states in $\{\bot\} \cup (Q \times \{0, 2\})$ are absorbing and for all $q, p \in Q$ and $u \in \{-1, 0, 1\}$, $\delta^I_0((q, 1))((p, 1+u)) = \sum_{a \in A(q), w(q,a)=u} \sigma(q, 1)(a) \cdot \delta(q, a)(p)$ (unattributed probability goes to $\bot$).

For all $\alpha \in [\![1, \beta-1]\!]$, we let $\mathcal{C}_\alpha = (\{\bot\} \cup (Q \times \{0, 2^{\alpha-1}, 2^\alpha, 3 \cdot 2^{\alpha-1}, 2^{\alpha+1}\}), \delta^I_\alpha)$ where the states in $\{\bot\} \cup (Q \times \{0, 2^{\alpha+1}\})$ are absorbing and, for all $q, p \in Q$ and $k \in \{2^{\alpha-1}, 2^\alpha, 3 \cdot 2^{\alpha-1}\}$, we let $\delta_\alpha((q, k))((p, k - 2^{\alpha-1})) = \delta^\sigma_{\mathcal{I}}((q, 2^{\alpha-1}))((p, 0))$ and $\delta_\alpha((q, k))((p, k + 2^{\alpha-1})) = \delta^\sigma_{\mathcal{I}}((q, 2^{\alpha-1}))((p, 2^\alpha))$.

We observe that for all $p \in Q$ and all $\alpha \in [\![\beta-1]\!]$, the subset of equations from Theorem 18.9 with the variables of the form $\langle (q, k) \nearrow p \rangle_\alpha$ (resp. $\langle (q, k) \searrow p \rangle_\alpha$) in the left-hand side coincides with a system for reachability probabilities in $\mathcal{C}_\alpha$ for target $\{(p, 2^{\alpha+1})\}$ (resp. $\{(p, 0)\}$) when substituting variables indexed by $\alpha - 1$ by their assignment in the least solution of the system. We devise an algorithm that individually modifies every such system so it has a unique solution. Because we are dealing with systems for reachability probabilities, we obtain a system with a unique solution by setting the variables whose assignment in the least solution of the system is zero to zero. This set of variables can be determined using a qualitative reachability analysis of the Markov chains $\mathcal{C}_\alpha$.

We analyse the Markov chains in order, i.e., we start with $\mathcal{C}_0$, then continue with $\mathcal{C}_1$ and so on. This is necessary: the transitions with non-zero probabilities in a Markov chain $\mathcal{C}_\alpha$ with $\alpha \geq 1$ are not known beforehand, but can be inferred from the analysis of $\mathcal{C}_{\alpha-1}$. We prove the following invariant of our procedure: after processing $\mathcal{C}_\alpha$, the least non-negative solution of the modified system is the least non-negative solution of the original system and all variables indexed by $\alpha' \leq \alpha$ have a unique valid assignment in any solution of the new system. We modify the system by adding constraints that are satisfied by the least non-negative solution of the original system. Thus, the first part of the invariant follows directly and we do not comment on it.

The transition structure of the Markov chain $\mathcal{C}_0$ can be constructed directly as follows: there exists a transition from a state $(q, 1)$ to a state $(p, 1 + u)$ if and only if there exists an action $a \in \mathsf{supp}(\sigma(q, 1))$ such that $w(q, a) = u$ and $p \in \mathsf{supp}(\delta(q, a))$ (in particular, the numerical probabilities do not matter).

This yields a directed graph G_0 over $Q \times \{0,1,2\}$. For all $q, p \in Q$, we have $[(q,1) \nearrow p]_0 = 0$ (resp. $[(q,1) \searrow p]_0 = 0$) if and only if $(p,2)$ (resp. $(p,0)$) cannot be reached from $(q,1)$ in G_0, and in this case, we add $\langle (q,1) \nearrow p \rangle_0 = 0$ (resp. $\langle (q,1) \searrow p \rangle_0 = 0$) to the system. After analysing $\mathcal{C}_0$, the invariant is satisfied. Indeed, following the addition of the new equations, there is only one possible assignment of the variables indexed by 0 in any solution: all of these variables are involved in a Markov chain reachability probability system with a unique solution.

We now let $\alpha \in [\![1, \beta - 1]\!]$ and assume that $\mathcal{C}_{\alpha-1}$ has been processed. We assume that the invariant holds by induction. Via the analysis of $\mathcal{C}_{\alpha-1}$, we know which transitions of $\mathcal{C}_\alpha$ have non-zero probability because these probabilities are reachability probabilities in $\mathcal{C}_{\alpha-1}$ by Theorem 18.9. We construct a directed graph G_α over the state space of $\mathcal{C}_\alpha$ similarly to above. Let $s = (q,k) \in Q \times \{2^{\alpha-1}, 2^\alpha, 3 \cdot 2^{\alpha-1}\}$ and $p \in Q$. In G_α, there is an edge from s to $(p, k + 2^{\alpha-1})$ if $[(q, 2^{\alpha-1}) \nearrow p]_{\alpha-1} > 0$ and there is an edge from s to $(p, k - 2^{\alpha-1})$ if $[(q, 2^{\alpha-1}) \searrow p]_{\alpha-1} > 0$. Whether these probabilities are positive is known from the analysis of $\mathcal{C}_{\alpha-1}$. As above, we have $[(q,k) \nearrow p]_\alpha = 0$ (resp. $[(q,k) \searrow p]_\alpha = 0$) if and only if $(p, 2^{\alpha+1})$ (resp. $(p,0)$) cannot be reached from (q,k) in G_α, and in this case, we add $\langle (q,k) \nearrow p \rangle_\alpha = 0$ (resp. $\langle (q,k) \searrow p \rangle_\alpha = 0$) to the system.

We prove that the invariant is preserved after this iteration. By induction, all variables indexed by $\alpha - 1$ have only one possible valid assignment. Given a solution of the system obtained after processing $\mathcal{C}_\alpha$, the variables indexed by α must satisfy an equation system with a unique solution, the coefficients of which are given by the unique valid assignment of the variables indexed by $\alpha - 1$. It follows that there can only be one valuation for the variables indexed by α in any solution of this system. This, in addition to the inductive hypothesis, guarantees that the invariant holds after the analysis of $\mathcal{C}_\alpha$. In the end, after analysing $\mathcal{C}_{\beta-1}$, the invariant guarantees that the resulting system has a unique solution.

To end the proof, it remains to show that the above algorithm respects the asserted complexity bounds. For all $\alpha \in [\![\beta - 1]\!]$, constructing the graph G_α takes time polynomial in the representation of $\mathcal{Q}$. Indeed, for G_0, to find all successors of a configuration $(q,1)$, it suffices to iterate over all actions $a \in \mathsf{supp}(\sigma(q,1))$ and then build on the set of successors $\mathsf{supp}(\delta(q,a))$. For the other graphs G_α, their structure is inferred from the analysis of $G_{\alpha-1}$. Each graph G_α can be analysed in polynomial time by performing a backward reachability analysis from each configuration on the right (i.e., after the arrow) of a variable indexed by α, and there are $2|Q|$ such configurations per graph. As we analyse β graphs, the overall time required to implement the procedure above respects the announced complexity bounds.

It remains to prove that the unique solution of the system provided by the procedure above can be computed in polynomial time in the BSS model. It suffices to solve linear systems for reachability probabilities in each of the Markov chains $\mathcal{C}_\alpha$ for $\alpha \in [\![\beta - 1]\!]$. This can be done in polynomial time with unit-cost arithmetic: these Markov chains have no more than $5 \cdot |Q|$ states each. □

18.5 Representing compressed Markov chains

The definition of the compressed Markov chain does not impose any conditions on the memoryless strategy σ: $\mathcal{C}^\sigma_\mathcal{I}$ can be defined without assuming that σ is an OEIS or a CIS. However, for algorithmic purposes, we require that $\mathcal{C}^\sigma_\mathcal{I}$ has a finite representation that is amenable to verification algorithms. In this section, we focus on the representation of the state space of $\mathcal{C}^\sigma_\mathcal{I}$, as the results of Section 18.4 provide a finite representation of transition probabilities for each interval.

By construction, $\mathcal{C}^\sigma_\mathcal{I}$ is finite if and only if $\mathcal{I}$ is finite. Thus $\mathcal{C}^\sigma_\mathcal{I}$ can only be finite when σ is an OEIS. In the remainder of this section, our goal is to show that $\mathcal{C}^\sigma_\mathcal{I}$ has a finite representation when σ is a CIS and $\mathcal{I}$ is periodic. We assume that $B = \infty$ and that σ is a CIS that for the remainder of the section. We let ρ denote a common period of σ and $\mathcal{I}$. We let $\mathcal{J}$ be the partition of $[\![1, \rho]\!]$ induced by $\mathcal{I}$.

We claim that $\mathcal{C}^\sigma_\mathcal{I}$ is induced by a one-counter Markov chain $\mathcal{R}^\sigma_\mathcal{J} = (R_\mathcal{J}, \delta^\sigma_\mathcal{J})$ where $R_\mathcal{J} = S_\mathcal{I} \cap (\{\bot\} \cup (Q \times [\![1, \rho]\!]))$ and $\delta^\sigma_\mathcal{J}$ is described below. We first explain the interpretation of configurations before giving intuition on $\delta^\sigma_\mathcal{J}$. Let $((q, k), k')$ be a configuration of $\mathcal{C}^{\leq\infty}(\mathcal{R}^\sigma_\mathcal{J})$ such that $k' \geq 1$ or $k = \rho$ (configurations that do not satisfy these conditions will be unreachable and thus ignored). This configuration corresponds to the configuration $(q, \rho \cdot (k' - 1) + k) \in S_\mathcal{I}$. The counter value k keeps track of where in the period we are and the counter value k' indicates how many multiples of ρ the counter has (strictly) exceeded. This correspondence guarantees that the configuration $((q,\rho),0)$ of $\mathcal{C}^{\leq\infty}(\mathcal{R}^\sigma_\mathcal{J})$ represents the configuration $(q,0) \in S_\mathcal{I}$.

Transitions are defined so that successors in $\mathcal{C}^{\leq\infty}(\mathcal{R}^\sigma_\mathcal{J})$ correspond to successors in $\mathcal{C}^\sigma_\mathcal{I}$. We formalise $\delta^\sigma_\mathcal{J}\colon R_\mathcal{J} \to \mathcal{D}(R_\mathcal{J} \times \{-1,0,1\})$ as follows. Like before, $\perp$ is absorbing and we give a weight of zero to its self-loop to ensure that we cannot terminate in $\perp$. In other words, we set $\delta^\sigma_\mathcal{J}(\perp)(\perp,0) = 1$. Let $s = (q,k) \in R_\mathcal{J}$. Each transition from s in $\mathcal{C}^\sigma_\mathcal{I}$ to a state in $R_\mathcal{J}$ yields a transition with weight zero in $\mathcal{R}^\sigma_\mathcal{J}$, i.e., for all $s' \in R_\mathcal{J}$, we let $\delta^\sigma_\mathcal{J}(s)(s',0) = \delta^\sigma_\mathcal{I}(s)(s')$. In particular, all incoming transitions of $\perp$ have weight zero. Any transition from s to a configuration $(p,0)$ in $\mathcal{C}^\sigma_\mathcal{I}$ induces a transition from s to (p,ρ) in $\mathcal{R}^\sigma_\mathcal{J}$ with a weight of -1, i.e., we let $\delta^\sigma_\mathcal{J}(s)((p,\rho),-1) = \delta^\sigma_\mathcal{I}(s)((p,0))$. Intuitively, in this case, we go back to the previous period. Finally, any transition from s to the configuration $(p,\rho+1)$ in $\mathcal{C}^\sigma_\mathcal{I}$ yields a transition with a weight of 1 in $\mathcal{R}^\sigma_\mathcal{J}$ from s to $(p,1) \in R_\mathcal{J}$ (this configuration is guaranteed to be in $S_\mathcal{I}$ because 1 is the minimum of the first interval and $\mathcal{I}$ has period ρ), i.e., we let $\delta^\sigma_\mathcal{J}(s)((p,1),1) = \delta^\sigma_\mathcal{I}(s)((p,\rho+1))$. Intuitively, in this case, we have passed a multiple of ρ. We obtain a well-defined transition function with the above: for all counter values k of configurations in $R_\mathcal{J}$, the successor counter values of k are a counter value of a configuration in $R_\mathcal{J}$, 0 or $\rho+1$, i.e., the upper and lower bound respectively of the intervals adjacent to $[\![1,\rho]\!]$ in $\mathcal{I} \cup \{[\![0]\!]\}$.

We now show that the termination probabilities in $\mathcal{C}^\sigma_\mathcal{I}$ and in $\mathcal{C}^{\leq\infty}(\mathcal{R}^\sigma_\mathcal{J})$ match from all initial configurations with the correspondence outlined previously.

Theorem 18.12. *For all $(q,k) \in R_\mathcal{J} \setminus \{\perp\}$ and $k' \in \mathbb{N}$ such that $k' \geq 1$ or $k = \rho$ and all $p \in Q$, we have*

$$\mathbb{P}_{\mathcal{C}^\sigma_\mathcal{I},(q,\rho\cdot(k'-1)+k)}(\mathsf{Reach}((p,0))) = \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{R}^\sigma_\mathcal{J}),((q,k),k')}(\mathsf{Term}((p,\rho))).$$

Proof. We define an injective mapping $\mathcal{F}\colon S_\mathcal{I} \setminus \{\perp\} \to R_\mathcal{J} \times \mathbb{N}$ such that, for any $s = (q,k) \in S_\mathcal{I}$, if k is divisible by ρ, we let $\mathcal{F}(s) = ((q,\rho),\frac{k}{\rho})$, and otherwise, we let $\mathcal{F}(s) = ((q,k \bmod \rho),\lfloor\frac{k}{\rho}\rfloor + 1)$. We observe that the image of

$\mathcal{F}$ is the set of configurations $((q,k),k')$ of $\mathcal{C}^{\leq\infty}(\mathcal{R}^{\sigma}_{\mathcal{J}})$ such that $k' \geq 1$ or $k = \rho$. The configurations of $\mathcal{C}^{\leq\infty}(\mathcal{R}^{\sigma}_{\mathcal{J}})$ with a state other than $\perp$ that are not in the image of $\mathcal{F}$ have no incoming transitions in $\mathcal{C}^{\leq\infty}(\mathcal{R}^{\sigma}_{\mathcal{J}})$ and are absorbing by construction. The statement of the theorem is equivalent to showing that for all $s \in S_{\mathcal{I}}$ and all $p \in Q$, we have $\mathbb{P}_{\mathcal{C}^{\sigma}_{\mathcal{I}},s}(\mathsf{Reach}(p,0)) = \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{R}^{\sigma}_{\mathcal{J}}),\mathcal{F}(s)}(\mathsf{Term}(p,\rho))$.

Let $[\delta^{\sigma}_{\mathcal{J}}]^{\leq\infty}$ denote the transition function of $\mathcal{C}^{\leq\infty}(\mathcal{R}^{\sigma}_{\mathcal{J}})$. The crux of the proof is to establish that for all s, $s' \in S_{\mathcal{I}}$, we have $\delta^{\sigma}_{\mathcal{I}}(s)(s') = [\delta^{\sigma}_{\mathcal{J}}]^{\leq\infty}(\mathcal{F}(s))(\mathcal{F}(s'))$. To refer to this property, we say that $\mathcal{F}$ preserves transitions.

Let $s = (q,k) \in S_{\mathcal{I}}$. If $k = 0$, we have $\delta^{\sigma}_{\mathcal{I}}(s)(s) = 1 = [\delta^{\sigma}_{\mathcal{J}}]^{\leq\infty}(\mathcal{F}(s))(\mathcal{F}(s)) = [\delta^{\sigma}_{\mathcal{J}}]^{\leq\infty}(((q,\rho),0))((q,\rho),0))$ since configurations with counter value zero are absorbing. We thus assume that $k > 0$.

We distinguish two cases below. First, we assume that $\rho = 1$, i.e., the strategy σ is counter-oblivious. It follows that for all $(p,k') \in Q \times \mathbb{N}$, we have $\mathcal{F}((p,k')) = ((p,1),k')$. The successor counter values of k are $k-1$ and $k+1$ because $\rho = 1$. Let $p \in Q$, $u \in \{-1,1\}$ and $s' = (p,k+u)$. By definition of $\mathcal{R}^{\sigma}_{\mathcal{J}}$ and of $\mathcal{C}^{\sigma}_{\mathcal{I}}$ (in particular, its periodic structure), we have

$$\begin{aligned}
[\delta^{\sigma}_{\mathcal{J}}]^{\leq\infty}(\mathcal{F}(s))(\mathcal{F}(s')) &= \delta^{\sigma}_{\mathcal{J}}((q,1))((p,1),u) \\
&= \delta^{\sigma}_{\mathcal{I}}((q,1))((p,1+u)) \\
&= \delta^{\sigma}_{\mathcal{I}}(s)(s').
\end{aligned}$$

This ends the proof that $\mathcal{F}$ preserves transitions when $\rho = 1$.

Now, we assume that $\rho > 1$. First, we assume that k is divisible by ρ, i.e., that $\mathcal{F}(s) = ((q,\rho),\frac{k}{\rho})$. Successor counter values of k are $k+1$ and $k-1$, since multiples of ρ are the maximum of their interval in $\mathcal{I}$. Let $p \in Q$. First, we consider $s' = (p,k+1)$. In this case, we have $\mathcal{F}(s') = ((p,1),\frac{k}{\rho}+1)$. By definition of $\mathcal{R}^{\sigma}_{\mathcal{J}}$ and of $\mathcal{C}^{\sigma}_{\mathcal{I}}$, we have

$$\begin{aligned}
[\delta^{\sigma}_{\mathcal{J}}]^{\leq\infty}(\mathcal{F}(s))(\mathcal{F}(s')) &= \delta^{\sigma}_{\mathcal{J}}((q,\rho))((p,1),1) \\
&= \delta^{\sigma}_{\mathcal{I}}((q,\rho))((p,\rho+1)) \\
&= \delta^{\sigma}_{\mathcal{I}}(s)(s').
\end{aligned}$$

Now, we consider $s' = (p, k-1)$. We obtain $\mathcal{F}(s') = ((p, \rho - 1), \frac{k}{\rho})$ and

$$\begin{aligned}[\delta^{\sigma}_{\mathcal{J}}]^{\leq\infty}(\mathcal{F}(s))(\mathcal{F}(s')) &= \delta^{\sigma}_{\mathcal{J}}((q,\rho))((p,\rho-1),0)\\ &= \delta^{\sigma}_{\mathcal{I}}((q,\rho))((p,\rho-1))\\ &= \delta^{\sigma}_{\mathcal{I}}(s)(s').\end{aligned}$$

We have shown that $\mathcal{F}$ preserves the transitions from s whenever k is a multiple of ρ.

We now assume that k is not a multiple of ρ, and thus that $\mathcal{F}(s) = ((q, k \bmod \rho), \lfloor\frac{k}{\rho}\rfloor + 1)$. Let $p \in Q$, k' be a successor counter value of k and $s' = (p, k')$. Let $I = [\![b^-, b^+]\!] \in \mathcal{I}$ such that $k \in I$. It follows that $k' \in [\![b^- - 1, b^+ + 1]\!]$. Since multiples of ρ are upper bounds of intervals, this implies that $k' \in [\![\lfloor\frac{k}{\rho}\rfloor \cdot \rho, (\lfloor\frac{k}{\rho}\rfloor + 1) \cdot \rho + 1]\!]$. In light of this, we distinguish four cases:

(i) $k' = \lfloor\frac{k}{\rho}\rfloor \cdot \rho$,

(ii) $k' = (\lfloor\frac{k}{\rho}\rfloor + 1) \cdot \rho$

(iii) $k' = (\lfloor\frac{k}{\rho}\rfloor + 1) \cdot \rho + 1$, and

(iv) k' is in none of the previous cases.

First, we assume that $k' = \lfloor\frac{k}{\rho}\rfloor \cdot \rho$, which implies that $\mathcal{F}(s') = ((p, \rho), \lfloor\frac{k}{\rho}\rfloor)$. We have

$$\begin{aligned}[\delta^{\sigma}_{\mathcal{J}}]^{\leq\infty}(\mathcal{F}(s))(\mathcal{F}(s')) &= \delta^{\sigma}_{\mathcal{J}}((q,k \bmod \rho))((p,\rho),-1)\\ &= \delta^{\sigma}_{\mathcal{I}}((q,k \bmod \rho))((p,0))\\ &= \delta^{\sigma}_{\mathcal{I}}(s)(s').\end{aligned}$$

Second, we assume that $k' = (\lfloor\frac{k}{\rho}\rfloor + 1) \cdot \rho$. This implies $\mathcal{F}(s') = ((p, \rho), \lfloor\frac{k}{\rho}\rfloor + 1)$. It holds that

$$\begin{aligned}[\delta^{\sigma}_{\mathcal{J}}]^{\leq\infty}(\mathcal{F}(s))(\mathcal{F}(s')) &= \delta^{\sigma}_{\mathcal{J}}((q,k \bmod \rho))((p,\rho),0)\\ &= \delta^{\sigma}_{\mathcal{I}}((q,k \bmod \rho))((p,\rho))\\ &= \delta^{\sigma}_{\mathcal{I}}(s)(s').\end{aligned}$$

Third, we assume that $k' = (\lfloor \frac{k}{\rho} \rfloor + 1) \cdot \rho + 1$. This implies that $\mathcal{F}(s') = ((p, 1), \lfloor \frac{k}{\rho} \rfloor + 2)$. It follows that

$$\begin{aligned}[\delta^\sigma_\mathcal{J}]^{\leq\infty}(\mathcal{F}(s))(\mathcal{F}(s')) &= \delta^\sigma_\mathcal{J}((q, k \bmod \rho))((p, 1), 1) \\ &= \delta^\sigma_\mathcal{I}((q, k \bmod \rho))((p, \rho + 1)) \\ &= \delta^\sigma_\mathcal{I}(s)(s').\end{aligned}$$

Finally, we assume that none of the previous cases holds. We conclude that $\mathcal{F}(s') = ((p, k' \bmod \rho), \lfloor \frac{k}{\rho} \rfloor + 1)$. We obtain

$$\begin{aligned}[\delta^\sigma_\mathcal{J}]^{\leq\infty}(\mathcal{F}(s))(\mathcal{F}(s')) &= \delta^\sigma_\mathcal{J}((q, k \bmod \rho))((p, k' \bmod \rho), 0) \\ &= \delta^\sigma_\mathcal{I}((q, k \bmod \rho))((p, k' \bmod \rho)) \\ &= \delta^\sigma_\mathcal{I}(s)(s').\end{aligned}$$

This ends the proof that $\mathcal{F}$ preserves transitions.

We lift $\mathcal{F}$ to histories by letting, for all $\bar{h} = s_1 \dots s_r \in \mathsf{Hist}(\mathcal{C}^\sigma_\mathcal{I})$ in which $\bot$ does not occur, $\mathcal{F}(\bar{h}) = \mathcal{F}(s_1) \dots \mathcal{F}(s_r)$ and we obtain, since $\mathcal{F}$ preserves transitions, $\mathbb{P}_{\mathcal{C}^\sigma_\mathcal{I}, \mathsf{first}(\bar{h})}(\mathsf{Cyl}\left(\bar{h}\right)) = \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{R}^\sigma_\mathcal{J}), \mathcal{F}(\mathsf{first}(\bar{h}))}(\mathsf{Cyl}\left(\mathcal{F}(\bar{h})\right))$. The claim of the theorem follows by writing the objectives as disjoint unions of history cylinders and using the fact that $\mathcal{F}$ is injective. □

Chapter 19

Interval strategy verification algorithms

We present algorithms for the interval strategy verification problem (Definition 17.6) based on the compressed Markov chains of Chapter 18. In Section 19.1, we present a polynomial time algorithm in the BSS model of computation for the verification of OEISs in bounded OC-MDPs. Sections 19.2 and 19.3 present a reduction from the verification problem for OEISs and CISs respectively to checking the validity of a universal formula in the theory of the reals.

Throughout this chapter, we use the Refine and $\mathsf{Isolate}$ operators over interval partitions defined in Chapter 18.2. In several places, we reference linear systems for reachability probabilities in Markov chains; we refer the reader to Appendix A.2.1 for a description of these systems.

We fix the following inputs for the whole chapter: an OC-MDP $\mathcal{Q} = (Q, A, \delta, w)$, a counter upper bound $B \in \bar{\mathbb{N}}_{>0}$, an OEIS or CIS σ of $\mathcal{M}^{\leq B}(\mathcal{Q})$, an initial configuration $s_{\mathsf{init}} = (q_{\mathsf{init}}, k_{\mathsf{init}}) \in Q \times [\![B]\!]$, a set of targets $T \subseteq Q$ and a threshold $\theta \in [0, 1] \cap \mathbb{Q}$.

To avoid redundancy, we describe the algorithms in a unified fashion for both the selective termination objective $\mathsf{Term}(T)$ and the state-reachability objective $\mathsf{Reach}(T)$. We let $\Omega \in \{\mathsf{Term}(T), \mathsf{Reach}(T)\}$ denote the objective. The major difference between the algorithms for selective termination and state-reachability is with respect to the studied OC-MDP: analysing the state-reachability probabilities requires a (polynomial-time) modification of $\mathcal{Q}$ beforehand (see Theorem 18.5). We assume that this modification has been applied if $\Omega = \mathsf{Reach}(T)$.

To further unify notation, we let $T_\Omega = T \times \{0\}$ if $\Omega = \mathsf{Term}(T)$ or $B = \infty$ and $T_\Omega = T \times \{0, B\}$ otherwise. This choice is motivated by the fact that, for all partitions $\mathcal{I}$ of $[\![1, B-1]\!]$ for which $\mathcal{C}^\sigma_\mathcal{I} = (S_\mathcal{I}, \delta^\sigma_\mathcal{I})$ is well-defined and $s_{\mathsf{init}} \in S_\mathcal{I}$, Theorems 18.4 and 18.5 ensure that $\mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}), s_{\mathsf{init}}}(\Omega) = \mathbb{P}_{\mathcal{C}^\sigma_\mathcal{I}, s_{\mathsf{init}}}(\mathsf{Reach}(T_\Omega))$.

Contents

19.1 Verification in bounded one-counter Markov decision processes

We provide a $\mathsf{P}^{\mathsf{PosSLP}}$ upper bound on the complexity of the OEIS verification problem in bounded OC-MDPs. We assume that $B \in \mathbb{N}_{>0}$. Let $\mathcal{I}'$ be the partition of $[\![1, B-1]\!]$ given by the description of σ. We let $\mathcal{I} = \mathsf{Refine}(\mathsf{Isolate}(\mathcal{I}', k_{\mathsf{init}}))$. It follows that σ is based on $\mathcal{I}$ and that $s_{\mathsf{init}} \in S_\mathcal{I}$ (because k_{init} is a bound of an interval in $\mathcal{I}$).

To obtain a $\mathsf{P}^{\mathsf{PosSLP}}$ complexity upper bound, we need only show that we can decide whether $\mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}), s_{\mathsf{init}}}(\Omega) \geq \theta$ in polynomial time in the BSS model [ABKM09]. In this model of computation, we can explicitly compute the transition probabilities of $\mathcal{C}^\sigma_\mathcal{I}$ in polynomial time (by Theorem 18.11) and use them to compute the probability of reaching T_Ω from s_{init} in $\mathcal{C}^\sigma_\mathcal{I}$. This reachability probability is exactly $\mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}), s_{\mathsf{init}}}(\Omega)$ by Theorems 18.4 and 18.5. We conclude by comparing it to θ. We obtain the following result.

Theorem 19.1. *The OEIS verification problem for state-reachability and selective termination in bounded OC-MDPs is in $\mathsf{P}^{\mathsf{PosSLP}}$.*

Proof. In this proof, we reason in the BSS model of computation. Our goal is to clarify the algorithm outlined above and prove that it runs in polynomial time. We let $\mathcal{I} = \mathsf{Refine}(\mathsf{Isolate}(\mathcal{I}', k_{\mathsf{init}}))$ where $\mathcal{I}'$ is the interval partition of $[\![1, B-1]\!]$ in the representation of σ. Lemma 18.3 guarantees that $\mathcal{I}$ can be computed in polynomial time and has a polynomial-size representation with respect to $\mathcal{I}'$, and that $\mathcal{C}^{\sigma}_{\mathcal{I}}$ is well-defined (cf. Assumption 18.1). It follows from Theorem 18.11 that the transition probabilities of $\mathcal{C}^{\sigma}_{\mathcal{I}}$ can be computed in polynomial time.

We have $\mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s_{\mathsf{init}}}(\Omega) = \mathbb{P}_{\mathcal{C}^{\sigma}_{\mathcal{I}},s_{\mathsf{init}}}(\mathsf{Reach}(T_{\Omega}))$ by Theorems 18.4 and 18.5. It follows that $\mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s_{\mathsf{init}}}(\Omega)$ can be computed in polynomial time by solving a linear system for Markov chain reachability probabilities (with $|S_{\mathcal{I}}| \leq 2 \cdot |\mathcal{I}| \cdot |Q| \cdot \log_2(B)$ variables) and then can be compared to θ in constant time. We conclude that the OEIS verification problem for Ω can be solved in polynomial time in the BSS model and thus lies in $\mathsf{P}^{\mathsf{PosSLP}}$ [ABKM09]. □

19.2 Open-ended interval strategies

We describe a co-ETR algorithm for the OEIS verification problem. This algorithm applies both in the bounded and unbounded settings. Recall that co-ETR is the class of decision problems that can be reduced (in polynomial time) to checking whether a universal sentence holds in the theory of the reals and that co-ETR is included in PSPACE [Can88]. The algorithm of Section 19.1 provides a finer bound when dealing with bounded OC-MDPs.

We construct logic formulae in the signature of ordered fields to decide the verification problem. We also use these formulae in the interval strategy realisability algorithms presented in Chapter 20. Therefore, we provide formulae that depend only on $\mathcal{Q}$ and the structure of σ, i.e., a finite interval partition $\mathcal{I}$ of $[\![1, B-1]\!]$ with respect to which compressed Markov chains are well-defined. This allows to build on these formulae to check the existence of well-performing strategies based on the considered interval partition.

We fix a finite interval partition $\mathcal{I}$ of $[\![1, B-1]\!]$ satisfying Assumption 18.1, i.e., such that, for all $I \in \mathcal{I}$, $\log_2(|I|+1) \in \mathbb{N}$. We build a formula with respect to $\mathcal{Q}$ and $\mathcal{I}$ and show that we can answer the verification problem via this formula for all OEISs based on $\mathcal{I}$ from any initial configuration in $S_{\mathcal{I}}$. We postpone the definition of a relevant partition for σ and s_{init} to the end of the section.

Our formula uses three sets of variables. First, for all $q \in Q$, $a \in A(q)$ and $I \in \mathcal{I}$, we introduce a variable $z^{I}_{q,a}$ to represent $\tau(q, \min I)(a)$ for any OEIS

τ based on $\mathcal{I}$. For all $I \in \mathcal{I}$, we let $\mathbf{z}^I = (z^I_{q,a})_{q \in Q, a \in A(q)}$ and let $\mathbf{z} = (\mathbf{z}^I)_{I \in \mathcal{I}}$. We let $\tau_{\mathbf{z}}$ be the parametric OEIS based on $\mathcal{I}$ defined by $\tau_{\mathbf{z}}(q, \min I)(a) = z^I_{q,a}$ for all $q \in Q$, $a \in A(q)$ and $I \in \mathcal{I}$. The notation $\tau_{\mathbf{z}}$ allows us to refer to the compressed Markov chain $\mathcal{C}^{\tau_{\mathbf{z}}}_{\mathcal{I}}$ parameterised by $\mathbf{z}$ in the following. To lighten notation, we write $\mathcal{C}^{\mathbf{z}}_{\mathcal{I}}$ instead of $\mathcal{C}^{\tau_{\mathbf{z}}}_{\mathcal{I}}$ and $\delta^{\mathbf{z}}_{\mathcal{I}}$ instead of $\delta^{\tau_{\mathbf{z}}}_{\mathcal{I}}$.

The second set of variables comes from Theorems 18.6 and 18.9 for each interval of $\mathcal{I}$ and are used to represent (and to characterise) the transition probabilities of $\mathcal{C}^{\mathbf{z}}_{\mathcal{I}}$ from configurations in $S_{\mathcal{I}} \setminus S^{\perp}_{\mathcal{I}}$ (recall that $S^{\perp}_{\mathcal{I}}$ is the set of absorbing states of $\mathcal{C}^{\mathbf{z}}_{\mathcal{I}}$). We let $\mathbf{x}$ denote the vector of all of these variables. For all configurations $s = (q, k) \in S_{\mathcal{I}} \setminus S^{\perp}_{\mathcal{I}}$ and $s' = (p, k') \in S_{\mathcal{I}} \setminus \{\perp\}$ such that k' is a successor counter value of k, we let $x_{s,s'}$ denote the variable corresponding to $\delta^{\sigma}_{\mathcal{I}}(s)(s')$. Some variables represent the outgoing probabilities from two configurations of the compressed Markov chain (see Lemma 18.8).

The last set of variables represents the probability of the counterpart of Ω in $\mathcal{C}^{\mathbf{z}}_{\mathcal{I}}$ from each configuration. For all $s \in S_{\mathcal{I}} \setminus S^{\perp}_{\mathcal{I}}$, we introduce a variable y_s where y_s represents $\mathbb{P}_{\mathcal{C}^{\mathbf{z}}_{\mathcal{I}},s}(\mathsf{Reach}(T_\Omega))$. We let $\mathbf{y}$ denote the vector of these variables.

We now construct, for all $s \in S_{\mathcal{I}} \setminus S^{\perp}_{\mathcal{I}}$, a quantifier-free formula such that when substituting $\mathbf{z}$ by a vector $\mathbf{z}^\star$ that yields a well-defined strategy $\tau_{\mathbf{z}^\star}$ and quantifying the other variables universally, the resulting sentence holds if and only if $\mathbb{P}^{\tau_{\mathbf{z}^\star}}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\Omega) \geq \theta$. We rely on universal quantification because we do not have a unique characterisation of the transition probabilities of $\mathcal{C}^{\mathbf{z}}_{\mathcal{I}}$ when $B = \infty$. We construct a quantifier-free conjunction (parameterised by the choices of the strategy) that only holds for (some) over-estimations of $\mathbb{P}^{\tau_{z^\star}}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\Omega)$. This allows us to check that $\mathbb{P}^{\tau_{z^\star}}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\Omega)$ exceeds θ by checking that all of its over-estimations do.

Our formula has two major sub-formulae. First, we define a formula depending on $\mathbf{z}$ such that the least vector satisfying it includes the transition probabilities of $\mathcal{C}^{\mathbf{z}}_{\mathcal{I}}$ from configurations to other configurations (i.e., not to $\perp$). For each $I \in \mathcal{I}$, we define $\Phi^I_\delta(\mathbf{x}, \mathbf{z}^I)$ as the conjunction of all the equations in the system characterising the transition probabilities from $S_{\mathcal{I}} \cap (Q \times I)$ in $\mathcal{C}^{\mathbf{z}}_{\mathcal{I}}$ given by Theorems 18.6 and 18.9 (the invoked theorem depends on whether I

is finite or not). We define

$$\Phi_{\delta}^{\mathcal{I}}(\mathbf{x},\mathbf{z}) = \bigwedge_{x\in\mathbf{x}} x \geq 0 \wedge \bigwedge_{I\in\mathcal{I}} \Phi_{\delta}^{I}(\mathbf{x},\mathbf{z}^{I}) \wedge \bigwedge_{s\in S_{\mathcal{I}}\setminus S_{\mathcal{I}}^{\perp}} \sum_{x_{s,s'}\in x_{s,\cdot}} x_{s,s'} \leq 1, \qquad (19.1)$$

where for all $s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}$, $x_{s,\cdot}$ denotes the set of well-defined variables of the form $x_{s,s'}$ ($s' \in S_{\mathcal{I}}$). The first conjunction ensures that any vector satisfying $\Phi_{\delta}^{\mathcal{I}}$ is non-negative while the rightmost conjunction ensures that for all $s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}$, $s' \mapsto x_{s,s'}$ is a sub-probability distribution. It follows that, for all configurations $s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}$ and all vectors $\mathbf{x}^{\star}$ and $\mathbf{z}^{\star}$ such that $\Phi_{\delta}^{\mathcal{I}}(\mathbf{x}^{\star},\mathbf{z}^{\star})$ holds, we can define a distribution $\delta_{\mathbf{x}^{\star}}(s) \in \mathcal{D}(S_{\mathcal{I}})$ such that for $s' \in S_{\mathcal{I}} \setminus \{\perp\}$, if $x_{s,s'}$ is a well-defined variable then $\delta_{\mathbf{x}^{\star}}(s)(s') = x^{\star}_{s,s'}$ and, otherwise, $\delta_{\mathbf{x}^{\star}}(s)(s') = 0$. We use these distributions in our correctness proof: they allow us to reason on Markov chains over $S_{\mathcal{I}}$.

The second block of the formula describes the probability of reaching T_{Ω} in $\mathcal{C}_{\mathcal{I}}^{\mathbf{z}}$. We consider the following formula, derived from a linear system for reachability probabilities in the finite Markov chain $\mathcal{C}_{\mathcal{I}}^{\mathbf{z}}$,

$$\Phi_{\Omega}^{\mathcal{I}}(\mathbf{x},\mathbf{y}) = \bigwedge_{s\in S_{\mathcal{I}}\setminus S_{\mathcal{I}}^{\perp}} \left(y_s \geq 0 \wedge y_s = \sum_{\substack{x_{s,s'}\in x_{s,\cdot}\\ s'\in S_{\mathcal{I}}\setminus S_{\mathcal{I}}^{\perp}}} x_{s,s'}y_{s'} + \sum_{\substack{x_{s,s'}\in x_{s,\cdot}\\ s'\in T_{\Omega}}} x_{s,s'} \right). \qquad (19.2)$$

We now state that for all well-defined instances $\tau_{\mathbf{z}^{\star}}$ of $\tau_{\mathbf{z}}$, the conjunction $\Phi_{\delta}^{\mathcal{I}}(\mathbf{x},\mathbf{z}^{\star}) \wedge \Phi_{\Omega}^{\mathcal{I}}(\mathbf{x},\mathbf{y})$ only holds for over-estimations of the values represented by the variables. This mainly follows from the construction of the formulae (in particular, by Theorems 18.6 and 18.9).

Lemma 19.2. *Let $\mathbf{z}^{\star}$ be a vector such that $\tau_{\mathbf{z}^{\star}}$ is a well-defined OEIS of $\mathcal{M}^{\leq B}(\mathcal{Q})$ based on $\mathcal{I}$. Let $\mathbf{x}^{\star},\mathbf{y}^{\star}$ be vectors such that $\mathbb{R} \models \Phi_{\delta}^{\mathcal{I}}(\mathbf{x}^{\star},\mathbf{z}^{\star}) \wedge \Phi_{\Omega}^{\mathcal{I}}(\mathbf{x}^{\star},\mathbf{y}^{\star})$. Then, for all $s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}$, we have $y^{\star}_s \geq \mathbb{P}_{\mathcal{C}_{\mathcal{I}}^{\mathbf{z}^{\star}},s}(\mathsf{Reach}(T_{\Omega}))$, and, for all $s' \in S_{\mathcal{I}} \setminus \{\perp\}$ such that $x_{s,s'}$ is a well-defined variable, $x^{\star}_{s,s'} \geq \delta_{\mathcal{I}}^{\mathbf{z}^{\star}}(s)(s')$.*

Proof. For all $s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}$ and all $s' \in S_{\mathcal{I}} \setminus \{\perp\}$ such that $x_{s,s'}$ is defined, we have $x^{\star}_{s,s'} \geq \delta_{\mathcal{I}}^{\mathbf{z}^{\star}}(s)(s')$ by construction of $\Phi_{\delta}^{\mathcal{I}}$ through Theorems 18.6 and 18.9. It remains to show that $y^{\star}_s \geq \mathbb{P}_{\mathcal{C}_{\mathcal{I}}^{\mathbf{z}^{\star}},s}(\mathsf{Reach}(T_{\Omega}))$ for all $s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}$.

Our argument relies on the finite Markov chain $\mathcal{C}_{\mathbf{x}^\star} = (S_{\mathcal{I}}, \delta_{\mathbf{x}^\star})$ where for all $s \in S_{\mathcal{I}}^{\perp}$, we have $\delta_{\mathbf{x}^\star}(s)(s) = 1$ and for all $s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}$ and all $s' \in S_{\mathcal{I}} \setminus \{\perp\}$, we have $\delta_{\mathbf{x}^\star}(s)(s') = x^\star_{s,s'}$ whenever $x_{s,s'}$ is defined and direct the probability mass that is not assigned to a successor of s in the previous way to $\perp$.

The least vector satisfying $\Phi^{\mathcal{I}}_{\Omega}(\mathbf{x}^\star, \mathbf{y})$ is $(\mathbb{P}_{\mathcal{C}_{\mathbf{x}^\star},s}(\mathsf{Reach}(T_\Omega)))_{s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}}$. Therefore, it suffices to show that for all $s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}$, we have $\mathbb{P}_{\mathcal{C}_{\mathbf{x}^\star},s}(\mathsf{Reach}(T_\Omega)) \geq \mathbb{P}_{\mathcal{C}^{\mathbf{z}^\star}_{\mathcal{I}},s}(\mathsf{Reach}(T_\Omega))$ to end the proof. It suffices to establish that for all histories $\bar{h} \in \mathsf{Hist}(\mathcal{C}^{\mathbf{z}^\star}_{\mathcal{I}})$ with $\mathsf{last}(\bar{h}) \in T_\Omega$ and no prior configuration in T_Ω, we have $\mathbb{P}_{\mathcal{C}^{\mathbf{z}^\star}_{\mathcal{I}},\mathsf{first}(\bar{h})}(\mathsf{Cyl}\,(h)) \leq \mathbb{P}_{\mathcal{C}_{\mathbf{x}^\star},\mathsf{first}(\bar{h})}(\mathsf{Cyl}\,(h))$. Let $\bar{h} \in \mathsf{Hist}(\mathcal{C}^{\mathbf{z}^\star}_{\mathcal{I}})$ be such a history. We assume that $\mathsf{first}(\bar{h}) \notin T_\Omega$, as otherwise the result is trivial. Since $\perp$ is absorbing (in both Markov chains), it follows that all states along $\bar{h}$ are configurations in $S_{\mathcal{I}}$. The desired inequality follows from $\delta_{\mathbf{x}^\star}(s)(s') = x^\star_{s,s'} \geq \delta^{\mathbf{z}^\star}_{\mathcal{I}}(s)(s')$ holding for all $s, s' \in S_{\mathcal{I}} \setminus \{\perp\}$. □

The following theorem provides the formula we use to solve the OEIS verification problem based on the intuition given above. Its correctness follows from Lemma 19.2.

Theorem 19.3. *Let $\mathbf{z}^\star$ be a vector such that $\tau_{\mathbf{z}^\star}$ is a well-defined OEIS. For all $s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}$, we have $\mathbb{P}^{\tau_{\mathbf{z}^\star}}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\Omega) \geq \theta$ if and only if $\mathbb{R} \models \forall \mathbf{x}\, \forall \mathbf{y}((\Phi^{\mathcal{I}}_{\delta}(\mathbf{x}, \mathbf{z}^\star) \wedge \Phi^{\mathcal{I}}_{\Omega}(\mathbf{x}, \mathbf{y})) \implies y_s \geq \theta)$.*

Proof. Let $s \in S_{\mathcal{I}} \setminus S_{\mathcal{I}}^{\perp}$. By Theorems 18.4 and 18.5, we have $\mathbb{P}^{\tau_{\mathbf{z}^\star}}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\Omega) = \mathbb{P}_{\mathcal{C}^{\mathbf{z}^\star}_{\mathcal{I}},s}(\mathsf{Reach}(T_\Omega))$. First, we assume that $\mathbb{P}_{\mathcal{C}^{\mathbf{z}^\star}_{\mathcal{I}},s}(\mathsf{Reach}(T_\Omega)) \geq \theta$. Let $\mathbf{x}^\star$ and $\mathbf{y}^\star$ such that $\mathbb{R} \models \Phi^{\mathcal{I}}_{\delta}(\mathbf{x}^\star, \mathbf{z}^\star) \wedge \Phi^{\mathcal{I}}_{\Omega}(\mathbf{x}^\star, \mathbf{y}^\star)$. By Lemma 19.2, we obtain $y^\star_s \geq \mathbb{P}_{\mathcal{C}^{\mathbf{z}^\star}_{\mathcal{I}},s}(\mathsf{Reach}(T_\Omega)) \geq \theta$. This shows the first implication.

Conversely, assume that $\mathbb{R} \models \forall \mathbf{x}\, \forall \mathbf{y}((\Phi^{\mathcal{I}}_{\delta}(\mathbf{x}, \mathbf{z}^\star) \wedge \Phi^{\mathcal{I}}_{\Omega}(\mathbf{x}, \mathbf{y})) \implies y_s \geq \theta)$. Let $\mathbf{x}^\star$ be the least non-negative satisfying assignment of $\mathbf{x}$ in $\Phi^{\mathcal{I}}_{\delta}(\mathbf{x}, \mathbf{z}^\star)$. The existence of $\mathbf{x}^\star$ is guaranteed by Theorems 18.6 and 18.9, which also imply that for all variables $x_{s',s''}$, we have $x^\star_{s',s''} = \delta^{\mathbf{z}^\star}_{\mathcal{I}}(s')(s'')$. We then let $\mathbf{y}^\star$ be the least satisfying assignment of $\mathbf{y}$ in the formula with parameters $\Phi^{\mathcal{I}}_{\Omega}(\mathbf{x}^\star, \mathbf{y})$. By construction of $\Phi^{\mathcal{I}}_{\Omega}$, we conclude that $\mathbf{y}^\star$ exists and that $\mathbb{P}_{\mathcal{C}^{\mathbf{z}^\star}_{\mathcal{I}},s}(\mathsf{Reach}(T_\Omega)) = y^\star_s \geq \theta$. □

We now analyse the size of the formula of Theorem 19.3. We show that this formula is of size polynomial in the encoding of $\mathcal{Q}$ and the natural representation of $\mathcal{I}$, i.e., as a finite set of intervals whose bounds are described in binary. We use this to show that we can build a formula to solve the verification problem in polynomial time. This analysis is also relevant to obtain complexity bounds for realisability.

Lemma 19.4. *The formula* $(\Phi^{\mathcal{I}}_{\delta}(\mathbf{x},\mathbf{z}) \wedge \Phi^{\mathcal{I}}_{\Omega}(\mathbf{x},\mathbf{y})) \implies y_s \geq \theta$ *has a number of variables and atomic sub-formulae polynomial in* $|Q|$, $|A|$, $|\mathcal{I}|$ *and the binary encoding of the largest integer bound in* $\mathcal{I}$.

Proof. Let $\beta = \max_{I\in\mathcal{I},|I|<\infty} \log_2(|I|+1)$ if there is a bounded interval in $\mathcal{I}$ and, otherwise, let $\beta = 1$. We note that $\beta \leq \log_2(b^+ + 1)$ where b^+ is the largest integer interval bound of $\mathcal{I}$.

First, we have, by definition of $\mathbf{z}$ and $\mathbf{y}$, $|\mathbf{z}| \leq |\mathcal{I}|\cdot|Q|\cdot|A|$ and $|\mathbf{y}| \leq |S_{\mathcal{I}} \setminus S^{\perp}_{\mathcal{I}}|$. By definition of $S_{\mathcal{I}}$, we have $|S_{\mathcal{I}} \setminus S^{\perp}_{\mathcal{I}}| \leq 2\cdot\beta\cdot|\mathcal{I}|\cdot|Q|$. Second, Lemmas 18.7 and 18.10 imply that $|\mathbf{x}|$ and the number and length of the atomic sub-formulae of $\Phi^{\mathcal{I}}_{\delta}(\mathbf{x},\mathbf{z})$ derived from Theorems 18.6 and 18.9 are polynomial in $|Q|$, $|A|$, $|\mathcal{I}|$ and β. It follows from the above that the number and length of the atomic sub-formulae of $(\Phi^{\mathcal{I}}_{\delta}(\mathbf{x},\mathbf{z}) \wedge \Phi^{\mathcal{I}}_{\Omega}(\mathbf{x},\mathbf{y})) \implies y_s \geq \theta$ is polynomial in $|Q|$, $|A|$, $|\mathcal{I}|$ and β. □

We now assume that σ is an OEIS and define the interval partition used to construct a verification formula. Let $\mathcal{I}'$ be the interval partition of $[\![1, B-1]\!]$ given in the representation of σ. We let $\mathcal{I} = \mathsf{Refine}(\mathsf{Isolate}(\mathcal{I}', k_{\mathsf{init}}))$. The partition $\mathcal{I}$ satisfies Assumption 18.1 and we have $s_{\mathsf{init}} \in S_{\mathcal{I}}$. Let $\mathbf{z}^{\sigma}$ denote the valuation of $\mathbf{z}$ defined by $z^{I}_{q,a} = \sigma(q, \min I)(a)$ for all $q \in Q$, $a \in A(q)$ and $I \in \mathcal{I}$. To decide the verification problem, we check whether the formula of Theorem 19.3 for s_{init} holds for this valuation $\mathbf{z}^{\sigma}$ of $\mathbf{z}$. We obtain the following complexity result.

Theorem 19.5. *The OEIS verification problem for selective termination and state-reachability objectives is in* co-ETR.

Proof. We prove that the formula used to answer the verification problem can be constructed in polynomial time. The structure of the formula is fixed. Therefore, it can be constructed in time polynomial in its size. It follows from Lemma 19.4 that we can construct our formula in polynomial time if $\mathcal{I}$ admits a representation of size polynomial in the number of inputs to the verification problem.

By definition of the Refine and Isolate operators, all interval bounds of $\mathcal{I}$ are either dominated by a bound in the representation of σ or by $k_{\mathsf{init}} + 1$. Therefore, all bounds admit a polynomial-size representation. Furthermore, Lemma 18.3 guarantees that when applying the refinement procedure to obtain $\mathcal{I}$, we obtain a partition of size polynomial in the size of the inputs to the verification problem. □

19.3 Cyclic interval strategies

We provide a co-ETR algorithm for the CIS verification problem that follows the same ideas as in Section 19.2. We assume throughout this section that $B = \infty$. To analyse CISs, we compress their induced Markov chain twice. We first apply the compression technique to the Markov chain induced by the strategy to be verified (for a well-chosen periodic partition of $\mathbb{N}_{>0}$). We represent this infinite compression as a one-counter Markov chain, as described in Chapter 18.5. We then use the compression approach to analyse this one-counter Markov chain.

As in the previous section, we provide formulae that are used in the fixed-interval and parameterised CIS realisability algorithms of Chapter 20: we design formulae that apply to all strategies based on a given periodic partition of $\mathbb{N}_{>0}$. We let $\rho \in \mathbb{N}_{>0}$ be a period, $\mathcal{J}$ be an interval partition of $[\![1, \rho]\!]$ into intervals and let $\mathcal{I}$ be the periodic partition generated by $\mathcal{J}$. We fix a finite interval partition $\mathcal{K}$ of $\mathbb{N}_{>0}$ for the second compression. For all intervals $I \in \mathcal{J} \cup \mathcal{K}$, we assume that $\log_2(|I| + 1) \in \mathbb{N}$ to guarantee that compressed Markov chains are well-defined with respect to these partitions (see Assumption 18.1). We design formulae for all CISs based on $\mathcal{I}$ whose structure depends only on $\mathcal{Q}$, $\mathcal{J}$ and $\mathcal{K}$. We let $\bar{T} = (T \times \{\rho\}) \times \{0\}$ denote the target of interest in the compression of the one-counter Markov chain (see Theorems 18.4 and 18.5).

Our formula for the verification problem uses four sets of variables; we require a new set of variables comparatively to Section 19.2 for the additional compression. First, we introduce variables for the choices of strategies. For all $q \in Q$, $a \in A(q)$ and $I \in \mathcal{J}$, we introduce a variable $z^I_{q,a}$ to represent $\sigma(q, \min I)(a)$. For all $I \in \mathcal{J}$, we let $\mathbf{z}^I = (z^I_{q,a})_{q \in Q, a \in A(q)}$ and let $\mathbf{z} = (\mathbf{z}^I)_{I \in \mathcal{J}}$. We let $\tau_{\mathbf{z}}$ be the parametric CIS of period ρ based on $\mathcal{I}$ defined by $\tau_{\mathbf{z}}(q, \min I)(a) = z^I_{q,a}$ for all $q \in Q$, $a \in A(q)$ and $I \in \mathcal{J}$. To lighten notation, we write $\mathcal{C}^{\mathbf{z}}_{\mathcal{I}}$ for the compressed Markov chain $\mathcal{C}^{\tau_{\mathbf{z}}}_{\mathcal{I}}$ associated to $\tau_{\mathbf{z}}$ (parameterised by $\mathbf{z}$) and $\mathcal{R}^{\mathbf{z}}_{\mathcal{J}} = (R_{\mathcal{J}}, \delta^{\tau_{\mathbf{z}}}_{\mathcal{J}})$ for the one-counter Markov chain $\mathcal{R}^{\tau_{\mathbf{z}}}_{\mathcal{J}}$ inducing $\mathcal{C}^{\mathbf{z}}_{\mathcal{I}}$ in the sense of Theorem 18.12. We let $R^{\top}_{\mathcal{J}} = R_{\mathcal{J}} \setminus \{\bot\}$. We let $\mathcal{C}_{\mathcal{K}}(\mathcal{R}^{\mathbf{z}}_{\mathcal{J}}) = (S_{\mathcal{K}}(R_{\mathcal{J}}), \delta_{\mathcal{K}}[\mathcal{R}^{\mathbf{z}}_{\mathcal{J}}])$ denote the compression of $\mathcal{C}^{\leq\infty}(\mathcal{R}^{\mathbf{z}}_{\mathcal{J}})$ with respect to $\mathcal{K}$.

We then introduce a new set of variables $\mathbf{v}$ for the transitions probabilities of $\mathcal{R}^{\mathbf{z}}_{\mathcal{J}}$ between configurations in $R^{\top}_{\mathcal{J}}$; these variables come from the system of Theorem 18.9. For any two $s, s' \in R^{\top}_{\mathcal{J}}$ and weight $u \in \{-1, 0, 1\}$, we let $v_{s,s',u}$ denote the variable corresponding to $\delta^{\tau_{\mathbf{z}}}_{\mathcal{J}}(s)(s', u)$ whenever this variable is well-defined. Third, we consider a set of variables $\mathbf{x}$ for the transitions probabilities of $\mathcal{C}_{\mathcal{K}}(\mathcal{R}^{\mathbf{z}}_{\mathcal{J}})$ taken from the systems of Theorems 18.6 and 18.9. For all $\bar{s}, \bar{s}' \in S_{\mathcal{K}}(R_{\mathcal{J}})$ such that $\bar{s} \in R^{\top}_{\mathcal{J}} \times \mathbb{N}_{>0}$, we write $x_{\bar{s},\bar{s}'}$ for the variable corresponding to $\delta_{\mathcal{K}}[\mathcal{R}^{\mathbf{z}}_{\mathcal{J}}](\bar{s})(\bar{s}')$ whenever this variable is defined. Finally, we introduce a variable $y_{\bar{s}}$ for all configurations $\bar{s} \in S_{\mathcal{K}}(R_{\mathcal{J}}) \cap (R^{\top}_{\mathcal{J}} \times \mathbb{N}_{>0})$ to represent the probability $\mathbb{P}_{\mathcal{C}_{\mathcal{K}}(\mathcal{R}^{\mathbf{z}}_{\mathcal{J}}),\bar{s}}(\mathsf{Reach}(\bar{T}))$. We let $\mathbf{y} = (y_{\bar{s}})_{\bar{s} \in S_{\mathcal{K}}(R_{\mathcal{J}}) \cap (R^{\top}_{\mathcal{J}} \times \mathbb{N}_{>0})}$.

We now formulate three sub-formulae of the formula used in our decision procedure. For all $I \in \mathcal{J}$, we let $\Psi^I_{\delta}(\mathbf{v}, \mathbf{z}^I)$ be the conjunction of the equations obtained by Theorem 18.9 for the outgoing transitions of $R_{\mathcal{J}} \cap (Q \times I)$ in $\mathcal{R}^{\tau_{\mathbf{z}}}_{\mathcal{J}}$. Similarly to Equation (19.1), we define a formula for the transitions of $\mathcal{R}^{\mathbf{z}}_{\mathcal{J}}$ by

$$\Psi^{\mathcal{J}}_{\delta}(\mathbf{v}, \mathbf{z}) = \bigwedge_{v \in \mathbf{v}} v \geq 0 \wedge \bigwedge_{I \in \mathcal{J}} \Psi^I_{\delta}(\mathbf{v}, \mathbf{z}^I) \wedge \bigwedge_{s \in R_{\mathcal{J}} \setminus \{\bot\}} \sum_{v_{s,s',u} \in v_{s,\cdot,\cdot}} v_{s,s',u} \leq 1. \quad (19.3)$$

We then construct the counterpart $\Phi^{\mathcal{K}}_{\delta}(\mathbf{x}, \mathbf{v})$ of the formula of Equation (19.1) for the compressed Markov chain $\mathcal{C}_{\mathcal{K}}(\mathcal{R}^{\sigma}_{\mathcal{J}})$. In this case, the sub-formulae derived from the systems of Theorem 18.6 and Theorem 18.9 for each interval of $\mathcal{K}$ depend on $\mathbf{v}$ instead of $\mathbf{z}$. We also build a counterpart $\Phi^{\mathcal{K}}_{\Omega}(\mathbf{x}, \mathbf{y})$ of the formula given in Equation (19.2) for $\mathcal{C}_{\mathcal{K}}(\mathcal{R}^{\sigma}_{\mathcal{J}})$ with respect to the target $\bar{T}$.

To decide the verification problem, we rely on a formula similar to that of Theorem 19.3: we check that over-estimations of the probability of interest exceed the threshold θ. To validate this approach, we establish a counterpart of Lemma 19.2 for the conjunction $\Psi^{\mathcal{J}}_{\delta}(\mathbf{v}, \mathbf{z}^\star) \wedge \Phi^{\mathcal{K}}_{\delta}(\mathbf{x}, \mathbf{v}) \wedge \Phi^{\mathcal{K}}_{\Omega}(\mathbf{x}, \mathbf{y})$ given a vector $\mathbf{z}^\star$ such that $\tau_{\mathbf{z}^\star}$ is a well-defined CIS based on $\mathcal{I}$.

Lemma 19.6. *Let $\mathbf{z}^\star$ be a vector such that $\tau_{\mathbf{z}^\star}$ is a well-defined CIS of $\mathcal{M}^{\leq\infty}(\mathcal{Q})$ based on $\mathcal{I}$. Let $\mathbf{v}^\star$, $\mathbf{x}^\star$ and $\mathbf{y}^\star$ be vectors such that $\mathbb{R} \models \Psi^{\mathcal{J}}_{\delta}(\mathbf{v}^\star, \mathbf{z}^\star) \wedge \Phi^{\mathcal{K}}_{\delta}(\mathbf{x}^\star, \mathbf{v}^\star) \wedge \Phi^{\mathcal{K}}_{\Omega}(\mathbf{x}^\star, \mathbf{y}^\star)$. Then, it holds that*

(i) for all $\bar{s} \in S_{\mathcal{K}}(R_{\mathcal{J}}) \cap (R^{\top}_{\mathcal{J}} \times \mathbb{N}_{>0})$, we have $y^\star_{\bar{s}} \geq \mathbb{P}_{\mathcal{C}_{\mathcal{K}}(\mathcal{R}^{\mathbf{z}^\star}_{\mathcal{J}}),\bar{s}}(\mathsf{Reach}(\bar{T}))$;

(ii) for all $\bar{s} \in S_{\mathcal{K}}(R_{\mathcal{J}}) \cap (R^{\top}_{\mathcal{J}} \times \mathbb{N}_{>0})$ and all $\bar{s}' \in S_{\mathcal{K}}(R_{\mathcal{J}})$ such that $x_{\bar{s},\bar{s}'}$ is defined, we have $x^\star_{\bar{s},\bar{s}'} \geq \delta_{\mathcal{K}}[\mathcal{R}^{\mathbf{z}^\star}_{\mathcal{J}}](\bar{s})(\bar{s}')$;

(iii) for all s, $s' \in R_{\mathcal{J}} \setminus \{\bot\}$ and $u \in \{-1, 0, 1\}$ such that $v_{s,s',u}$ is defined, we have $v^\star_{s,s',u} \geq \delta^{\mathbf{z}^\star}_{\mathcal{J}}(s)(s', u)$.

Proof. Item (iii) follows from the construction of $\Psi^{\mathcal{J}}_{\delta}$ based on Theorem 18.9. To prove (i) and (ii), we consider the one-counter Markov chain $\mathcal{R}_{\mathbf{v}^\star} = (R_{\mathcal{J}}, \delta_{\mathbf{v}^\star})$ where for all s, $s' \in R_{\mathcal{J}}$ and all $u \in \{-1, 0, 1\}$, $\delta_{\mathbf{v}^\star}(s)(s', u) = v^\star_{s,s',u}$ whenever the variable $v_{s,s',u}$ is defined and any probability that is not assigned in this way is attributed to $\delta_{\mathbf{v}^\star}(s)(\bot, 0)$. We show that in the compression $\mathcal{C}_{\mathcal{K}}(\mathcal{R}_{\mathbf{v}^\star}) = (S_{\mathcal{K}}(R_{\mathcal{J}}), \delta_{\mathcal{K}}[\mathcal{R}_{\mathbf{v}^\star}])$ of $\mathcal{R}_{\mathbf{v}^\star}$ with respect to $\mathcal{K}$, we have the two following properties:

(a) for all $\bar{s}, \bar{s}' \in S_{\mathcal{K}}(R_{\mathcal{J}}) \cap (R^{\top}_{\mathcal{J}} \times \mathbb{N}_{>0})$, $\delta_{\mathcal{K}}[\mathcal{R}_{\mathbf{v}^\star}](\bar{s})(\bar{s}') \geq \delta^{\mathbf{z}^\star}_{\mathcal{J}}(\bar{s})(\bar{s}')$;

(b) for all $\bar{s} \in S_{\mathcal{K}}(R_{\mathcal{J}}) \cap (R^{\top}_{\mathcal{J}} \times \mathbb{N}_{>0})$, we have $\mathbb{P}_{\mathcal{C}_{\mathcal{K}}(\mathcal{R}_{\mathbf{v}^\star}),\bar{s}}(\mathsf{Reach}(\bar{T})) \geq \mathbb{P}_{\mathcal{C}_{\mathcal{K}}(\mathcal{R}^{\mathbf{z}^\star}_{\mathcal{J}}),\bar{s}}(\mathsf{Reach}(\bar{T}))$.

These two properties along with Lemma 19.2 (with respect to the parameterised formula $\Phi^{\mathcal{K}}_{\delta}(\mathbf{x}, \mathbf{v}^\star) \wedge \Phi^{\mathcal{K}}_{\Omega}(\mathbf{x}, \mathbf{y})$) yield Items (i) and (ii).

We first prove (a). Let $\bar{s}, \bar{s}' \in S_{\mathcal{K}}(R_{\mathcal{J}}) \setminus \{\bot\}$. Let s, $s \in R_{\mathcal{J}}$ and $k, k' \in \mathbb{N}$ such that $\bar{s} = (s, k)$ and $\bar{s}' = (s', k')$. If k' is not a successor counter value of k with respect to $\mathcal{K}$, then we have $\delta_{\mathcal{K}}[\mathcal{R}_{\mathbf{v}^\star}](\bar{s})(\bar{s}') = \delta_{\mathcal{K}}[\mathcal{R}^{\mathbf{z}^\star}_{\mathcal{J}}](\bar{s})(\bar{s}') = 0$. Otherwise, $\delta_{\mathcal{K}}[\mathcal{R}^{\mathbf{z}^\star}_{\mathcal{J}}](\bar{s})(\bar{s}')$ is the probability of reaching $\bar{s}'$ from $\bar{s}$ in $\mathcal{C}^{\leq\infty}(\mathcal{R}^{\mathbf{z}^\star}_{\mathcal{J}})$ without visiting another configuration with a successor value of k beforehand. Along the relevant plays for this probability, there are no $\bot$ configurations. Since $\delta_{\mathcal{K}}[\mathcal{R}_{\mathbf{v}^\star}](\bar{s})(\bar{s}')$ is similarly defined, the desired inequality follows from (iii).

Item (a) implies (b), as there are no $\bot$ configuration that can occur on plays ending in $\bar{T}$ (of which all states are absorbing). $\square$

We obtain an adaptation of Theorem 19.3 for CISs via Lemma 19.6. We use the correspondence between configurations of $\mathcal{R}^{\mathbf{z}}_{\mathcal{J}}$ and $\mathcal{C}_{\mathcal{K}}(\mathcal{R}^{\mathbf{z}}_{\mathcal{J}})$ established in Theorem 18.12 in this result.

Theorem 19.7. *Let $\mathbf{z}^\star$ be a vector such that $\tau_{\mathbf{z}^\star}$ is a well-defined CIS of $\mathcal{M}^{\leq\infty}(\mathcal{Q})$ based on $\mathcal{I}$. For all $\bar{s} = ((q,k),k') \in S_{\mathcal{K}}(R_{\mathcal{J}}) \cap (R^{\top}_{\mathcal{J}} \times \mathbb{N}_{>0})$, we have $\mathbb{P}^{\tau_{\mathbf{z}^\star}}_{\mathcal{M}^{\leq\infty}(\mathcal{Q}),(q,(k'-1)\cdot\rho+k)}(\mathsf{Reach}(\bar{T})) \geq \theta$ if and only if $\mathbb{R} \models \forall\mathbf{x}\,\forall\mathbf{y}\forall\mathbf{v}((\Psi^{\mathcal{J}}_{\delta}(\mathbf{v},\mathbf{z}^\star) \wedge \Phi^{\mathcal{K}}_{\delta}(\mathbf{x},\mathbf{v}) \wedge \Phi^{\mathcal{K}}_{\Omega}(\mathbf{x},\mathbf{y})) \implies y_{\bar{s}} \geq \theta)$.*

Proof. Let $\bar{s} = ((q,k),k') \in S_{\mathcal{K}}(R_{\mathcal{J}}) \cap (R^{\top}_{\mathcal{J}} \times \mathbb{N}_{>0})$. By Theorems 18.12 and 18.4, we have $\mathbb{P}^{\tau_{\mathbf{z}^\star}}_{\mathcal{M}^{\leq\infty}(\mathcal{Q}),(q,(k'-1)\rho+k)}(\mathsf{Reach}(\bar{T})) = \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{R}^{\mathbf{z}^\star}_{\mathcal{J}}),\bar{s}}(\mathsf{Reach}(\bar{T})) = \mathbb{P}_{\mathcal{C}_{\mathcal{K}}(\mathcal{R}^{\mathbf{z}^\star}_{\mathcal{J}}),\bar{s}}(\mathsf{Reach}(\bar{T}))$

If $\mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{R}^{\sigma}_{\mathcal{J}}),\bar{s}}(\mathsf{Reach}(\bar{T})) \geq \theta$, then we have $\mathbb{R} \models \forall\mathbf{x}\,\forall\mathbf{y}\forall\mathbf{v}((\Psi^{\mathcal{J}}_{\delta}(\mathbf{v},\mathbf{z}^\star) \wedge \Phi^{\mathcal{K}}_{\delta}(\mathbf{x},\mathbf{v}) \wedge \Phi^{\mathcal{K}}_{\Omega}(\mathbf{x},\mathbf{y})) \implies y_{\bar{s}} \geq \theta)$ by Lemma 19.6.

Conversely, assume that $\mathbb{R} \models \forall\mathbf{x}\,\forall\mathbf{y}\forall\mathbf{v}((\Psi^{\mathcal{J}}_{\delta}(\mathbf{v},\mathbf{z}^\star) \wedge \Phi^{\mathcal{K}}_{\delta}(\mathbf{x},\mathbf{v}) \wedge \Phi^{\mathcal{K}}_{\Omega}(\mathbf{x},\mathbf{y})) \implies y_{\bar{s}} \geq \theta)$. Let $\mathbf{v}^\star$ be the least vector satisfying $\Psi^{\mathcal{J}}_{\delta}(\mathbf{v},\mathbf{z}^\star)$, $\mathbf{x}^\star$ be the least vector satisfying $\Phi^{\mathcal{K}}_{\delta}(\mathbf{x},\mathbf{v}^\star)$ and $\mathbf{y}^\star$ be the least vector satisfying $\Phi^{\mathcal{K}}_{\Omega}(\mathbf{x}^\star,\mathbf{y})$. By construction of these three formulae (and Theorems 18.6 and 18.9), these vectors are well-defined and we obtain $y^\star_{\bar{s}} = \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{R}^{\sigma}_{\mathcal{J}}),\bar{s}}(\mathsf{Reach}(\bar{T})) \geq \theta$. $\square$

We now study the size of the formula of Theorem 19.7 for our complexity analysis. We obtain a conclusion similar to that provided by Lemma 19.4.

Lemma 19.8. *The formula $(\Psi^{\mathcal{J}}_{\delta}(\mathbf{v},\mathbf{z}) \wedge \Phi^{\mathcal{K}}_{\delta}(\mathbf{x},\mathbf{v}) \wedge \Phi^{\mathcal{K}}_{\Omega}(\mathbf{x},\mathbf{y})) \implies y_{\bar{s}} \geq \theta$ has a number of variables and atomic sub-formulae polynomial in $|Q|$, $|A|$, $|\mathcal{J}|$, $|\mathcal{K}|$, the binary encoding of ρ and the binary encoding of the largest integer bound in $\mathcal{K}$. The length of its atomic sub-formulae is polynomial in the same parameters.*

Proof. Lemma 19.4 implies that the number of variables and atomic formulae of $\Phi^{\mathcal{K}}_{\delta}(\mathbf{x}, \mathbf{v}) \wedge \Phi^{\mathcal{K}}_{\Omega}(\mathbf{x}, \mathbf{y})$, well as the length of these atomic formulae, is polynomial in $|R^{\top}_{\mathcal{J}}|$, $|\mathcal{K}|$ and the binary encoding of the largest bound in $\mathcal{K}$ (actions are not relevant; remark that we deal with a Markov chain). By construction of $R^{\top}_{\mathcal{J}}$, we have $|R^{\top}_{\mathcal{J}}| \leq 2 \cdot \log_2(\rho + 1) \cdot |\mathcal{J}| \cdot |Q|$ (interval bounds of $\mathcal{J}$ are at most ρ). Regarding $\Psi^{\mathcal{J}}_{\delta}(\mathbf{v}, \mathbf{z})$, it suffices to adapt the analysis performed in the proof of Lemma 19.4 for the formula $\Phi^{\mathcal{I}}_{\delta}$ to obtain the desired bounds. □

We now assume that the input strategy σ is a CIS. We close the section by explaining how to construct $\mathcal{J}$ and $\mathcal{K}$ in polynomial time from the representation of σ and s_{init} to prove that the verification problem is in $\mathsf{co\text{-}ETR}$ via the previous results of this section. Let $\mathcal{J}'$ denote the partition of $[\![1, \rho]\!]$ given in the representation of σ. We let $\mathcal{J} = \mathsf{Refine}(\mathsf{Isolate}(\mathcal{J}, k_{\mathsf{init}} \bmod \rho))$. For the partition for the second compression, we let $\mathcal{K} = \mathsf{Refine}([\![1, \lfloor \frac{k_{\mathsf{init}}}{\rho} \rfloor]\!]) \cup \{[\![\lfloor \frac{k_{\mathsf{init}}}{\rho} \rfloor + 1, +\infty]\!]\}$ of $\mathbb{N}_{>0}$. We observe that the counterpart of s_{init} in $S_{\mathcal{K}}(R_{\mathcal{J}})$ (in the sense of Theorem 18.12) is guaranteed to exist: if $k_{\mathsf{init}} \bmod \rho = 0$, then we have $((q_{\mathsf{init}}, \rho), \frac{k_{\mathsf{init}}}{\rho}) \in S_{\mathcal{K}}(R_{\mathcal{J}})$, and, otherwise, we have $((q_{\mathsf{init}}, k_{\mathsf{init}} \bmod \rho), \lfloor \frac{k_{\mathsf{init}}}{\rho} \rfloor + 1) \in S_{\mathcal{K}}(R_{\mathcal{J}})$. We let $\mathbf{z}^{\sigma}$ denote the substitution of $\mathbf{z}$ such that $z^{I}_{q,a}$ is set to $\sigma(q, \min I)(a)$ for all $q \in Q$, $a \in A(q)$ and $I \in \mathcal{J}$. With the partitions $\mathcal{J}$ and $\mathcal{K}$ given above and the formula of Theorem 19.7 with respect to the counterpart of s_{init} in $S_{\mathcal{K}}(R_{\mathcal{J}})$ and the parameter $\mathbf{z}^{\sigma}$, we can decide our instance of the verification problem. We thus obtain the following complexity result.

Theorem 19.9. *The CIS verification problem for selective termination and reachability objectives is in* $\mathsf{co\text{-}ETR}$.

Proof. We observe, in the same way as in the proof of Theorem 19.5, that Lemmas 18.3 and 19.8 imply that the formula of Theorem 19.7 can be constructed in polynomial time for the partitions described above. □

Chapter 20

Structurally-constrained interval strategy realisability algorithms

We provide complexity upper bounds for the fixed-interval and parameterised realisability problems for interval strategies. Our algorithms are built on the verification techniques presented in Chapter 19. We first provide a technical result for analysing the complexity of the parameterised realisability problem in Section 20.1. In Section 20.2, we focus on the case of bounded OC-MDPs. We consider OEISs in general, i.e., we provide an approach applicable for both the bounded and unbounded setting, in Section 20.3. We close the chapter with CISs in Section 20.4.

We use similar approaches for all settings. For fixed-interval realisability for pure strategies, we non-deterministically construct strategies and verify them. This approach is not viable for fixed-interval realisability for randomised strategies; instead, we quantify existentially over strategy variables in the logical formulae used for verification. For the parameterised realisability problem, we build on our algorithms for the fixed-interval case. The main idea is use non-determinism to find an interval partition compatible with the input parameters and then use fixed-interval algorithms with this partition to answer our problem. All complexity bounds provided in this chapter are in PSPACE.

We consider the following inputs for the whole section: an OC-MDP $\mathcal{Q} = (Q, A, \delta, w)$, a counter upper bound $B \in \bar{\mathbb{N}}_{>0}$, an initial configuration $s_{\mathsf{init}} = (q_{\mathsf{init}}, k_{\mathsf{init}}) \in Q \times [\![B]\!]$, a set of targets $T \subseteq Q$, an objective $\Omega \in \{\mathsf{Reach}(T), \mathsf{Term}(T)\}$ and a threshold $\theta \in [0, 1] \cap \mathbb{Q}$. We specify the other inputs below. As in Chapter 19, we assume that we work with the modified OC-MDP of Theorem 18.5 if $\Omega = \mathsf{Reach}(T)$ to allow for a uniform presentation.

Contents

20.1 Parameters and compatible interval partitions

We study the representation size of interval partitions that arise in the study of the parameterised interval strategy realisability problems. We let $d \in \mathbb{N}_{>0}$ denote the parameter bounding the number of intervals in the partition and $n \in \mathbb{N}_{>0}$ be the parameter bounding the size of bounded intervals of the partition. We recall that d is assumed to be encoded in unary. Formally, we say that an interval partition $\mathcal{I}$ of $[\![1, k]\!]$ (where $k \in \mathbb{N}_{>0}$) is *compatible* with d and n if such that $|\mathcal{I}| \leq d$ and, for all bounded $I \in \mathcal{I}$, $|I| \leq n$.

Our algorithms for the parameterised interval strategy realisability problems rely on non-determinism to find an interval partition that is compatible with the input parameters for which there exists a strategy based on it that ensures Ω with probability at least θ. To guarantee a PSPACE complexity upper bound, our approach requires that the interval partitions that are compatible with d and n admit a representation that is polynomial in the unary encoding of d and the binary encoding size of d. We show this below.

Lemma 20.1. *Let $d \in \mathbb{N}_{>0}$, $n \in \mathbb{N}_{>0}$ and $k \in \bar{\mathbb{N}}$. Let $\mathcal{I}$ be an interval partition of $[\![1, k]\!]$ such that $|\mathcal{I}| \leq d$ and for all bounded $I \in \mathcal{I}$, $|I| \leq n$. Then $\mathcal{I}$ can be explicitly represented in space $\mathcal{O}(d \cdot (\log_2(d) + \log_2(n)))$.*

Proof. We can represent each interval $[\![b^-, b^+]\!] \in \mathcal{I}$ by the pair (b^-, b^+) (where b^+ can be $+\infty$). We prove that each finite interval bound in these pairs is at most $n \cdot d$.

First, assume that $k \in \mathbb{N}$. In this case, the interval $[\![1, k]\!]$ is the union of at most d sets of at most n elements (by the assumption on $\mathcal{I}$). It follows that $k \leq d \cdot n$, and thus, that all finite interval bounds of $\mathcal{I}$ are no more than $d \cdot n$.

Second, assume that $k = \infty$. Let I_∞ denote the unbounded interval in $\mathcal{I}$. It holds (by the same reasoning as above) that the bounds of all intervals in $\mathcal{I} \setminus \{I_\infty\}$ are no more than $(d-1) \cdot n$. This implies that $\min I_\infty - 1 \leq (d-1) \cdot n$. We obtain that in this second case, all finite interval bounds in $\mathcal{I}$ are no more than $(d-1) \cdot n + 1 \leq d \cdot n$.

We conclude that, in both cases, we can represent $\mathcal{I}$ using no more than d pairs of numbers whose binary encoding is in $\mathcal{O}(\log_2(d) + \log_2(n))$. □

Lemma 20.1 implies that, under the assumption that the parameter d for the number of intervals is given in unary, the interval partitions that are compatible with the parameters admit a polynomial-size representation with respect to the inputs to the parameterised interval strategy realisability problems.

20.2 Realisability in bounded one-counter Markov decision processes

Assume that $B \in \mathbb{N}_{>0}$. We provide algorithms for the fixed-interval and parameterised OEIS realisability problems in bounded OC-MDPs. We first discuss the variants of these problems for pure strategies in Section 20.2.1. We then discuss the variants for randomised strategies in Section 20.2.2.

20.2.1 Pure strategies

First, we consider the fixed-interval pure OEIS realisability problem. Fix an input interval partition $\mathcal{I}'$ of $[\![1, B-1]\!]$. We obtain a straightforward non-deterministic algorithm: we guess a pure interval strategy σ based on the partition $\mathcal{I}'$ and then use our verification algorithm for OEISs in bounded OC-MDPs as a $\mathsf{P}^{\mathsf{PosSLP}}$ sub-procedure (Theorem 19.1) to check whether $\mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}), s_{\mathsf{init}}}(\Omega) \geq \theta$. This realisability algorithm runs in non-deterministic polynomial time with a PosSLP oracle: we non-deterministically choose $d \cdot |Q|$ actions, i.e., one per state-interval pair, and then run a deterministic polynomial-time algorithm with a PosSLP oracle. This yields an $\mathsf{NP}^{\mathsf{PosSLP}}$ upper bound for this problem.

For the parameterised pure OEIS realisability problem in bounded OC-MDPs, we proceed similarly. Let $d \in \mathbb{N}_{>0}$ and $n \in \mathbb{N}_{>0}$ respectively denote the input parameters bounding the number and size of intervals. We non-deterministically guess an interval partition $\mathcal{I}'$ of $[\![1, B-1]\!]$ that is compatible with d and n (these partitions can be represented in polynomial space by Lemma 20.1), guess a pure strategy based on $\mathcal{I}'$ and verify it. In this case, by adapting the analysis made above, we also obtain an $\mathsf{NP}^{\mathsf{PosSLP}}$ upper complexity bound. We summarise the above upper bounds in the following theorem.

Theorem 20.2. *The fixed-interval and parameterised pure OEIS realisability problems for selective termination and state-reachability objectives in bounded OC-MDPs are in* $\mathsf{NP}^{\mathsf{PosSLP}}$.

20.2.2 Randomised strategies

We now consider the fixed-interval and parameterised randomised OEIS realisability problem and describe an $\mathsf{NP}^{\mathsf{ETR}}$ algorithm. We start with the fixed-interval realisability problem. Let $\mathcal{I}' = (I_j)_{j \in [\![1,d]\!]}$ denote an input interval partition of $[\![1, B-1]\!]$. Prefacing the formula we used in our verification algorithm (cf. Theorem 19.3) with existential quantifiers for the strategy probabilities yields a polynomial-space procedure (cf. Section 20.3). We provide an alternative approach for bounded OC-MDPs which yields a more precise bound.

The key is to rely on a unique characterisation of the transition and reachability probabilities in the compressed Markov chain that we consider. Theorem 18.11 (Page 338) provides the means to do this: it provides systems whose only solution contains the transition probabilities of a compressed Markov chain. These systems also indicate which transitions have positive probability in the compressed Markov chain. We can thus refine the reachability probability system (described in the formula $\Phi_\Omega^{\mathcal{I}}$ in Equation (19.2), Page 351) to have a unique solution.

To adequately refine systems using Theorem 18.11, we must know the supports of the distributions assigned by the considered strategy. For this, we use non-determinism; we guess the action supports for each state-interval pair. We then construct an existential formula that is dependent on these supports. This formula holds if and only if there exists a strategy witnessing a positive answer to the fixed-interval randomised OEIS realisability problem that uses these supports.

We let $T_\Omega = T \times \{0\}$ if $\Omega = \mathsf{Term}(T)$ and $T_\Omega = T \times \{0, B\}$ if $\Omega = \mathsf{Reach}(T)$. Let $\mathcal{I} = \mathsf{Refine}(\mathsf{Isolate}(\mathcal{I}', k_{\mathsf{init}}))$. We note that $s_{\mathsf{init}} \in S_{\mathcal{I}}$. For all $j \in [\![1, d]\!]$, we let $\mathcal{I}_j = \{I \in \mathcal{I} \mid I \subseteq I_j\}$. We use the same variables as the verification formula of Theorem 19.3, Chapter 19.2. We briefly recall these variables. We have a variable vector $\mathbf{z}$ for the probabilities assigned by strategies and a vector $\mathbf{z}^I$ for all $I \in \mathcal{I}$ for the strategy probabilities specific to I. We let $\tau_{\mathbf{z}}$ denote a parametric strategy given by $\tau_{\mathbf{z}}(q, \min I)(a) = z^I_{q,a}$ for all $q \in Q$, $a \in A(q)$ and $I \in \mathcal{I}$. We also have variable vectors $\mathbf{x}$ for the transition probabilities of the compressed Markov chain $\mathcal{C}^{\mathbf{z}}_{\mathcal{I}}$ and $\mathbf{y}$ for the probability of reaching T_Ω from each configuration in $S_{\mathcal{I}} \setminus S^{\perp}_{\mathcal{I}}$.

We call functions $\mathcal{B}\colon Q \times [\![1, d]\!] \to 2^A \setminus \{\emptyset\}$ such that for all $q \in Q$ and $j \in [\![1, d]\!]$, the inclusion $\mathcal{B}(q, j) \subseteq A(q)$ holds *support-assigning functions*. An OEIS σ based on $(I_j)_{j \in [\![1,d]\!]}$ is $\mathcal{B}$-*supported* if for all $q \in Q$ and $j \in [\![1, d]\!]$, we have $\mathsf{supp}(\sigma(q, \min I_j)) = \mathcal{B}(q, j)$.

We now define the required sub-formulae used in our algorithm. The first formula checks that the substitution of $\mathbf{z}$ results in an interpretation of the symbolic strategy $\tau_{\mathbf{z}}$ that is based on $\mathcal{I}'$ and is $\mathcal{B}$-supported. For each $j \in [\![1, d]\!]$, we fix $I_j^\star \in \mathcal{I}_j$. We define the formula $\Phi_\sigma^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{z})$ as the conjunction of

$$\bigwedge_{j\in[\![1,d]\!]} \left(\bigwedge_{q\in Q} \sum_{a\in A(q)} z_{q,a}^{I_j^\star} = 1 \wedge \bigwedge_{I\in\mathcal{I}_j} \mathbf{z}^I = \mathbf{z}^{I_j^\star} \right). \tag{20.1}$$

which requires that the interpretation of $\tau_\mathbf{z}$ is a well-defined OEIS based on $\mathcal{I}'$, and

$$\bigwedge_{j\in[\![1,d]\!]} \bigwedge_{q\in Q} \left(\bigwedge_{a\in\mathcal{B}(q,j)} z_{q,a}^{I_j^\star} > 0 \wedge \bigwedge_{a\notin\mathcal{B}(q,j)} z_{q,a}^{I_j^\star} = 0 \right),$$

which requires that the interpretation of $\tau_\mathbf{z}$ be $\mathcal{B}$-supported.

The other sub-formulae are built under the assumption that we consider an interpretation of $\tau_\mathbf{z}$ with the supports described by $\mathcal{B}$. The second sub-formula is a parallel of the formula of Equation (19.1), which describes the transition probabilities of a compressed Markov chain. For each $I \in \mathcal{I}$, we let $\Phi_\delta^{I,\mathcal{I}',\mathcal{B}}(\mathbf{x}, \mathbf{z}^I)$ be the conjunction of the equations in the system with a unique solution obtained from Theorem 18.11 for the transitions from $S_\mathcal{I} \cap (Q \times I)$ in $\mathcal{C}_\mathcal{I}^\mathbf{z}$. We let $\Phi_\delta^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x}, \mathbf{z}) = \bigwedge_{I\in\mathcal{I}} \Phi_\delta^{I,\mathcal{I}',\mathcal{B}}(\mathbf{x}, \mathbf{z}^I)$. From these equations, we can deduce the transition structure of $\mathcal{C}_\mathcal{I}^\mathbf{z}$ and construct a linear system with a unique solution describing the probability of reaching T_Ω in $\mathcal{C}_\mathcal{I}^\mathbf{z}$; we let $\Phi_\Omega^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x}, \mathbf{y})$ denote the conjunction of the equations of this system. Using the fact that these last two formulae have unique satisfying assignments for a valuation of $\mathbf{z}$ satisfying $\Phi_\sigma^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{z})$, we obtain the following theorem.

Theorem 20.3. *Let* $\mathcal{B}\colon Q \times [\![1, d]\!] \to 2^A \setminus \{\emptyset\}$ *be a support-assigning function and* $s \in S_\mathcal{I} \setminus S_\mathcal{I}^\perp$. *There exists a* $\mathcal{B}$*-supported strategy* σ *based on* $\mathcal{I}'$ *such that* $\mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\Omega) \geq \theta$ *if and only if* $\mathbb{R} \models \exists\mathbf{z}\exists\mathbf{x}\,\exists\mathbf{y}(\Phi_\sigma^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{z}) \wedge \Phi_\delta^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x}, \mathbf{z}) \wedge \Phi_\Omega^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x}, \mathbf{y}) \wedge y_s \geq \theta)$.

Proof. Assume that there exists a $\mathcal{B}$-supported strategy σ based on $\mathcal{I}'$ such that $\mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\Omega) \geq \theta$. Let $\mathbf{z}^\sigma$ denote the valuation of $\mathbf{z}$ given by $z_{q,a}^I = \sigma(q, \min I)(a)$ for all $q \in Q$, $a \in A(q)$ and $I \in \mathcal{I}$. It is easy to see that $\Phi_\sigma^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{z}^\sigma)$ because σ is $\mathcal{B}$-supported. By construction of $\Phi_\delta^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x}, \mathbf{z})$ (via Theorem 18.11), there is a unique vector $\mathbf{x}^\star$ such that $\Phi_\delta^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x}^\star, \mathbf{z}^\sigma)$ holds which contains the transition probabilities of $\mathcal{C}_\mathcal{I}^\sigma$. In turn, this implies that there is a unique vector $\mathbf{y}^\star$ such that $\Phi_\Omega^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x}^\star, \mathbf{y}^\star)$ holds, and this vector is $(\mathbb{P}_{\mathcal{C}^\sigma_{\mathcal{I}},s'}(\mathsf{Reach}(T_\Omega)))_{s'\in S_{\mathcal{I}}\setminus S_{\mathcal{I}}^\perp}$. By Theorems 18.4 and 18.5, we obtain that $y_s^\star = \mathbb{P}_{\mathcal{C}^\sigma_{\mathcal{I}},s}(\mathsf{Reach}(T_\Omega)) = \mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\Omega) \geq \theta$.

Conversely, let $\mathbf{z}^\star$, $\mathbf{x}^\star$ and $\mathbf{y}^\star$ witnessing that the existential formula above holds and define $\sigma = \tau_{\mathbf{z}^\star}$. The strategy σ is well-defined and $\mathcal{B}$-supported because $\Phi_\sigma^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{z}^\star)$ holds. Furthermore, by construction of the formulae $\Phi_\delta^{\mathcal{I},\mathcal{I}',\mathcal{B}}$ and $\Phi_\Omega^{\mathcal{I},\mathcal{I}',\mathcal{B}}$, we deduce that $y_s^\star = \mathbb{P}_{\mathcal{C}^\sigma_{\mathcal{I}},s}(\mathsf{Reach}(T_\Omega)) \geq \theta$. We conclude that $\mathbb{P}^\sigma_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\Omega) = \mathbb{P}_{\mathcal{C}^\sigma_{\mathcal{I}},s}(\mathsf{Reach}(T_\Omega)) \geq \theta$ by Theorems 18.4 and 18.5. □

To decide the fixed-interval randomised OEIS realisability problem, it suffices to check that there exists a support-assigning function $\mathcal{B}$ (using non-determinism) such that the formula of Theorem 20.3 for holds for s_{init} (by construction of $\mathcal{I}$, $s_{\mathsf{init}} \in S_{\mathcal{I}}$). We thus obtain an $\mathsf{NP}^{\mathsf{ETR}}$ upper bound for this variant of the fixed-interval realisability problem.

For the parameterised realisability problem, we obtain an $\mathsf{NP}^{\mathsf{ETR}}$ upper bound by altering the fixed-interval algorithm slightly. In this case, we use non-determinism to guess an interval partition $\mathcal{I}'$ that is compatible with the input parameters and a support-assigning function. We then check the validity of the formula of Theorem 20.3 for the interval partition $\mathcal{I} = \mathsf{Refine}(\mathsf{Isolate}(\mathcal{I}', k_{\mathsf{init}}))$ and the initial configuration s_{init}. We obtain the following result.

Theorem 20.4. *The fixed-interval and parameterised randomised OEIS realisability problems for selective termination and state-reachability objectives in bounded OC-MDPs are in $\mathsf{NP}^{\mathsf{ETR}}$.*

Proof. We present a unified argument for both the fixed-interval and parameterised realisability problems. The only difference between our complexity analysis for these two problems is how the interval partition $\mathcal{I}'$ is obtained. In the fixed-interval case, the interval partition $\mathcal{I}'$ is a part of the input. In the parameterised case, $\mathcal{I}'$ is of size polynomial in the input parameters by Lemma 20.1 (recall that the parameter bounding the number of intervals is assumed to be given in unary).

We let $\mathcal{I} = \mathsf{Refine}(\mathsf{Isolate}(\mathcal{I}', k_{\mathsf{init}}))$ and $\mathcal{B}\colon Q \times [\![1, |\mathcal{I}'|]\!] \to 2^A \setminus \{\emptyset\}$ be a support-assigning function. Lemma 18.3, which bounds the number of intervals generated by Refine, guarantees that $\mathcal{I}$ has a representation of size polynomial in that of $\mathcal{I}'$ and k_{init}. The support-assigning function $\mathcal{B}$ can be explicitly represented in space polynomial in $|Q|$, $|A|$ and $|\mathcal{I}'|$. Therefore, to prove that the $\mathsf{NP}^{\mathsf{ETR}}$ upper bound holds, it remains to prove that the formula $\Phi_{\sigma}^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{z}) \wedge \Phi_{\delta}^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x},\mathbf{z}) \wedge \Phi_{\Omega}^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x},\mathbf{y}) \wedge y_{s_{\mathsf{init}}} \geq \theta$ can be constructed in deterministic polynomial time from $\mathcal{I}'$, $\mathcal{I}$, $\mathcal{B}$ and the other inputs to the considered realisability problem.

Lemma 19.4 implies that the number of variables in this formula is polynomial in $|Q|$, $|A|$ and $|\mathcal{I}|$, and that the formulae $\Phi_{\delta}^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x},\mathbf{z})$ and $\Phi_{\Omega}^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x},\mathbf{y})$ have size polynomial in $|Q|$, $|A|$ and $|\mathcal{I}|$. Indeed, for these two formulae, we observe that they can be derived from the original formulae $\Phi_{\delta}^{\mathcal{I}}$ and $\Phi_{\Omega}^{\mathcal{I}}$ for verification by taking their conjunction with atomic propositions requiring that some variables in $\mathbf{x}$ and $\mathbf{y}$ are equal to zero (these variables can be identified in polynomial time through the algorithm of Theorem 18.11 and with a reachability analysis of the compressed Markov chain). Finally, we observe that, in the formula $\Phi_{\sigma}^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{z})$, there are no more than $|\mathcal{I}'| \cdot (2 \cdot |Q| + |\mathcal{I}| \cdot |A| \cdot |Q|)$ atomic formulae of length in $\mathcal{O}(|A|)$. We have thus shown that the formula $\Phi_{\sigma}^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{z}) \wedge \Phi_{\delta}^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x},\mathbf{z}) \wedge \Phi_{\Omega}^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{x},\mathbf{y}) \wedge y_{s_{\mathsf{init}}} \geq \theta$ can be constructed in polynomial time. □

20.3 Open-ended interval strategies

We now consider the fixed-interval and parameterised realisability problems for OEISs in general OC-MDPs. We first consider the variant for pure strategies, then the variant for randomised strategies.

20.3.1 Pure strategies

For pure strategies, we adapt the approach of Section 20.2. In the fixed-interval case, we guess a pure OEIS based on the input interval partition of the set of counter values and verify it. In the parameterised case, we guess an interval partition compatible with the input parameters (its representation is of size polynomial in the representation of the parameters by Lemma 20.1) and a pure OEIS based on it, then verify it. Theorem 19.5 thus implies that the fixed-interval and parameterised realisability problems for pure OEISs can be solved by a non-deterministic polynomial-time algorithm that uses a $\mathsf{co\text{-}ETR}$

oracle (which is the same as using an ETR oracle). We obtain the following upper bound.

Theorem 20.5. *The fixed-interval and parameterised pure OEIS realisability problems for selective termination and state-reachability objectives are in* $\mathsf{NP}^{\mathsf{ETR}}$.

20.3.2 Randomised strategies

We now consider the variant of our realisability problems for randomised OEISs. First, we discuss the fixed-interval problem and let $\mathcal{I}' = (I_j)_{j\in[\![1,d]\!]}$ be an input partition. Let $\mathcal{I} = \mathsf{Refine}(\mathsf{Isolate}(\mathcal{I}', k_{\mathsf{init}}))$. We let, for all $j \in [\![1,d]\!]$, $\mathcal{I}_j = \{I \in \mathcal{I} \mid I \subseteq I_j\}$.

As we have done in Section 20.2.2 for realisability in the bounded setting, we consider the sets of variables $\mathbf{z}$, $\mathbf{x}$ and $\mathbf{y}$ and the formulae $\Phi^{\mathcal{I}}_{\delta}(\mathbf{x},\mathbf{z})$ and $\Phi^{\mathcal{I}}_{\Omega}(\mathbf{x},\mathbf{y})$ from Chapter 19.2. We introduce a new formula to constrain the variables $\mathbf{z}$ similarly to the formula of Equation (20.1): we let $\Phi^{\mathcal{I},\mathcal{I}'}_{\sigma}(\mathbf{z})$ be the formula

$$\bigwedge_{j\in[\![1,d]\!]} \bigwedge_{I\in\mathcal{I}_j} \left(\bigwedge_{q\in Q} \left(\bigwedge_{a\in A(q)} z^I_{q,a} \geq 0 \wedge \sum_{a\in A(q)} z^I_{q,a} = 1 \right) \wedge \bigwedge_{I'\in\mathcal{I}_j} \mathbf{z}^I = \mathbf{z}^{I'} \right). \quad (20.2)$$

Any vector $\mathbf{z}^\star$ satisfying $\Phi^{\mathcal{I},\mathcal{I}'}_{\sigma}(\mathbf{z}^\star)$ defines a strategy $\tau_{\mathbf{z}^\star}$ based on $\mathcal{I}'$ and any such strategy induces such a vector. From the formula and equivalence presented in Theorem 19.3, we obtain the following result.

Theorem 20.6. *Let $s \in S_{\mathcal{I}} \setminus S^{\perp}_{\mathcal{I}}$. There exists an OEIS σ based on the partition $\mathcal{I}'$ such that $\mathbb{P}^{\sigma}_{\mathcal{M}^{\leq B}(\mathcal{Q}),s}(\Omega) \geq \theta$ if and only if $\mathbb{R} \models \exists\mathbf{z}\forall\mathbf{x}\,\forall\mathbf{y}(\Phi^{\mathcal{I},\mathcal{I}'}_{\sigma}(\mathbf{z}) \wedge ((\Phi^{\mathcal{I}}_{\delta}(\mathbf{x},\mathbf{z}) \wedge \Phi^{\mathcal{I}}_{\Omega}(\mathbf{x},\mathbf{y})) \implies y_s \geq \theta))$.*

We obtain that the fixed-interval realisability problem for randomised OEISs can be reduced to deciding a sentence in $\mathbb{R}$ with two blocks of quantifiers. This shows that the problem is decidable in polynomial space [BPR06, Rmk. 13.10]. For the parameterised randomised OEIS realisability problem, we obtain an $\mathsf{NPSPACE} = \mathsf{PSPACE}$ [Sav70] upper bound through the following algorithm: we use non-determinism to obtain a partition $\mathcal{I}'$ of $[\![1, B-1]\!]$ compatible with the input parameters and then check the validity of the formula of Theorem 20.6 for this partition. We obtain the following.

Theorem 20.7. *The fixed-interval and parameterised randomised OEIS realisability problems for selective termination and state-reachability objectives are in* PSPACE*.*

Proof. Let $\mathcal{I}'$ denote the input interval partition of $[\![1, B-1]\!]$ if we consider the fixed-interval realisability problem or the partition obtained using non-determinism if we consider the parameterised realisability problem. In the latter, the representation of $\mathcal{I}'$ is of size polynomial in that of the input parameters by Lemma 20.1.

Whether a formula with two blocks of quantifiers holds in the theory of the reals can be decided in polynomial space [BPR06, Rmk. 13.10]. Therefore, to obtain the claim of the theorem, it suffices to show that the formula $\Phi_\sigma^{\mathcal{I},\mathcal{I}'}(\mathbf{z}) \wedge ((\Phi_\delta^{\mathcal{I}}(\mathbf{x},\mathbf{z}) \wedge \Phi_\Omega^{\mathcal{I}}(\mathbf{x},\mathbf{y})) \implies y_s \geq \theta)$ can be constructed in polynomial time with respect to the representation of $\mathcal{Q}$, k_{init} and $\mathcal{I}'$.

Lemma 18.3 implies that $\mathcal{I} = \mathsf{Refine}(\mathsf{Isolate}(\mathcal{I}', k_{\mathsf{init}}))$ admits a representation of size polynomial in the representation of $\mathcal{I}'$ and k_{init}. It follows from Lemma 19.4 that the sub-formula $(\Phi_\delta^{\mathcal{I}}(\mathbf{x},\mathbf{z}) \wedge \Phi_\Omega^{\mathcal{I}}(\mathbf{x},\mathbf{y})) \implies y_s \geq \theta$ can be constructed in polynomial time. For $\Phi_\sigma^{\mathcal{I},\mathcal{I}'}(\mathbf{z})$, we observe that it is a conjunction of no more than $|\mathcal{I}| \cdot (|Q| \cdot (|A|+1) + |\mathcal{I}| \cdot |Q| \cdot |A|)$ atomic formulae of length in $\mathcal{O}(|A|)$. □

20.4 Cyclic interval strategies

We now consider the fixed-interval and parameterised realisability problems for CISs. We assume that $B = \infty$ for the remainder of the section. We adapt the techniques of Section 20.3. We first discuss the problem variants for pure strategies, then for randomised strategies.

20.4.1 Pure strategies

First, we provide non-deterministic algorithms for the variants of these problems for pure CISs. In the fixed-interval case, it suffices to non-deterministically select an action for each state of the OC-MDP and interval from the input interval partition and then verify the resulting CIS. We now consider the parameterised case. Let $d \in \mathbb{N}_{>0}$ and $n \in \mathbb{N}_{>0}$ respectively denote the parameter bounding the number of intervals and the size of intervals for the desirable interval partitions. To solve the parameterised realisability problem, we guess a period $\rho \leq d \cdot n$, an interval partition $\mathcal{J}'$ of $[\![1, \rho]\!]$ that is compatible with d and n and actions for all pairs in $Q \times \mathcal{J}'$, then verify the obtained CIS. Our $\mathsf{co\text{-}ETR}$ upper bound for the CIS verification problem of Theorem 19.9, along with Lemma 20.1 in the parameterised case, imply that both of these problems can be solved in non-deterministic polynomial time with an ETR oracle.

Theorem 20.8. *The fixed-interval and parameterised pure CIS realisability problems for selective termination and state-reachability objectives are in* $\mathsf{NP}^{\mathsf{ETR}}$.

20.4.2 Randomised strategies

We now consider the problem variants for randomised strategies. As before, we first focus on the fixed-interval case. Let $\rho \in \mathbb{N}_{>0}$ denote the input period and $\mathcal{J}' = (I_j)_{j \in [\![1,d]\!]}$ denote the input interval partition of $[\![1, \rho]\!]$. We define $\mathcal{J} = \mathsf{Refine}(\mathsf{Isolate}(\mathcal{J}', k_{\mathsf{init}} \bmod \rho))$ of $[\![1, \rho]\!]$. We let, for all $j \in [\![1, d]\!]$, $\mathcal{J}_j = \{I \in \mathcal{J} \mid I \subseteq I_j\}$. We also let $\mathcal{K} = \mathsf{Refine}([\![1, \lfloor \frac{k_{\mathsf{init}}}{\rho} \rfloor]\!]) \cup \{[\![\lfloor \frac{k_{\mathsf{init}}}{\rho} \rfloor + 1, \infty]\!]\}$. These choices guarantee that the counterpart in the sense of Theorem 18.12 of s_{init} in $S_{\mathcal{K}}(R_{\mathcal{J}})$ exists, where $S_{\mathcal{K}}(R_{\mathcal{J}})$ is the state space of the compression of the compression, in the sense of Chapter 19.3.

Next, we reintroduce the sets of variables $\mathbf{z}$, $\mathbf{v}$, $\mathbf{x}$ and $\mathbf{y}$ and the formulae $\Psi_{\delta}^{\mathcal{J}}(\mathbf{v}, \mathbf{z})$, $\Phi_{\delta}^{\mathcal{K}}(\mathbf{x}, \mathbf{v})$ and $\Phi_{\Omega}^{\mathcal{K}}(\mathbf{x}, \mathbf{y})$ from Chapter 19.3. We formulate an adaptation of the formulae of Equations (20.1) and (20.2) with respect to $\mathcal{J}$: we let $\Phi_{\sigma}^{\mathcal{J},\mathcal{J}'}(\mathbf{z})$ be the formula

$$\bigwedge_{j \in [\![1,d]\!]} \bigwedge_{I \in \mathcal{J}_j} \left(\bigwedge_{q \in Q} \left(\bigwedge_{a \in A(q)} z_{q,a}^{I} \geq 0 \wedge \sum_{a \in A(q)} z_{q,a}^{I} = 1 \right) \wedge \bigwedge_{I' \in \mathcal{J}_j} \mathbf{z}^{I} = \mathbf{z}^{I'} \right). \quad (20.3)$$

Once again, we obtain a natural correspondence between vectors $\mathbf{z}^{\star}$ satisfying $\Phi_{\sigma}^{\mathcal{J},\mathcal{J}'}(\mathbf{z}^{\star})$ and the CISs considered for our realisability problem. The following theorem is therefore implied by Theorem 19.7, which provides the formula used in our CIS verification algorithm.

Theorem 20.9. *Let $\bar{s} \in S_{\mathcal{K}}(R_{\mathcal{J}}) \cap (R_{\mathcal{J}} \times \mathbb{N}_{>0})$. There exists a CIS σ based on the periodic partition generated by $\mathcal{J}'$ such that $\mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{R}^{\sigma}_{\mathcal{J}}),\bar{s}}(\mathsf{Reach}(\bar{T})) \geq \theta$ if and only if $\mathbb{R} \models \exists\mathbf{z}\forall\mathbf{x}\,\forall\mathbf{y}\forall\mathbf{v}(\Phi^{\mathcal{J},\mathcal{J}'}_{\sigma}(\mathbf{z}) \wedge ((\Psi^{\mathcal{J}}_{\delta}(\mathbf{v},\mathbf{z}) \wedge \Phi^{\mathcal{K}}_{\delta}(\mathbf{x},\mathbf{v}) \wedge \Phi^{\mathcal{K}}_{\Omega}(\mathbf{x},\mathbf{y})) \implies y_{\bar{s}} \geq \theta))$.*

We thus obtain that the fixed-interval realisability problem for randomised CISs is reducible in polynomial time to deciding whether a sentence with two quantifier blocks holds in the theory of the reals. We obtain a PSPACE upper bound [BPR06, Rmk. 13.10]. For the parameterised case, we obtain an NPSPACE = PSPACE [Sav70] upper bound; we non-deterministically guess an interval partition $\mathcal{J}'$ as we have done for the parameterised pure CIS realisability problem, then reduce to checking the validity of the formula of Theorem 20.9 as in the fixed-interval case. The following theorem summarises our complexity bounds.

Theorem 20.10. *The fixed-interval and parameterised randomised CIS realisability problems for selective termination and state-reachability objectives are in PSPACE.*

Proof. This theorem follows from an adaptation of the proof of Theorem 20.7. The major difference, in this case, is that we refer to Lemma 19.8 instead of Lemma 19.4 for bounds on the size of the sentence used in to solve the realisability problem. □

Chapter 21

Hardness of interval strategy problems

We present lower complexity bounds for the interval strategy verification problem, the fixed-interval realisability problem and the parameterised realisability problem. In Section 21.1, we prove the square-root-sum hardness of all variants of these problems. In Section 21.2, we show that our interval strategy realisability problems are NP-hard when considering selective termination.

Contents

21.1 Square-root-sum hardness

We establish the square-root-sum hardness of our interval strategy problems via an existing reduction from the square-root-sum problem to (a variant of) the verification for one-counter Markov chains from [EWY10].

We formalise the definition of the square-root-sum problem and discuss our definition in Section 21.1.1. We adapt the reduction of [EWY10] to our formalism) in Section 21.1.2, and derive square-root-sum hardness for our problems in unbounded OC-MDPs. In Section 21.1.3, we present an adaptation of this reduction to the bounded setting: intuitively, we show that we can, in polynomial time, compute a large enough counter bound so the bounded one-counter Markov chain approximates the one used in the unbounded reduction well enough for the reduction to still be valid. We present some additional technical details for this second reduction separately in Section 21.1.4.

21.1.1 The square-root-sum problem

The square-root-sum problem consists in comparing a sum of square roots of natural numbers to some integer bound. It is formalised as follows.

Definition 21.1. The *square-root-sum problem* asks, given integers $x_1, \dots, x_n \in \mathbb{N}$ and $y \in \mathbb{N}$, whether $\sum_{i=1}^{n} \sqrt{x_i} \geq y$.

The square-root-sum problem is not known to be solvable in polynomial time in the Turing model of computation. It is known that the square-root-sum problem can be solved in polynomial time in the BSS model [Tiw92]. In particular, the square-root-sum problem is in $\mathsf{P}^{\mathsf{PosSLP}}$ [ABKM09] and thus in the counting hierarchy.

We discuss our definition of the square-root-sum problem in the following.

Remark 21.2. The square-root-sum problem is typically formulated as having to decide whether $\sum_{i=1}^{n} \sqrt{x_i} \leq y$, i.e., with the opposite inequality. For the sake of illustrating the hardness of a problem, both problems can be seen as equally suitable.

We argue this by briefly showing that an efficient solution to either variant of the problem yields an efficient solution to the other one. We observe that these two variants are almost the complement of one another. The only case in which the two problems have the same solution for the same inputs is when $\sum_{i=1}^{n} \sqrt{x}_i = y$. Deciding whether $\sum_{i=1}^{n} \sqrt{x}_i = y$ can be done in polynomial time [BFHT85]. Therefore, an efficient decision procedure for one variant of the square-root-sum problem would entail an efficient one for the other. ◁

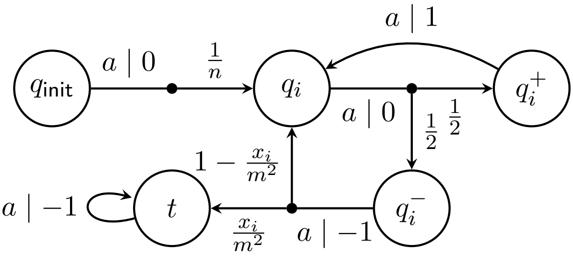


Figure 21.1: A fragment of $\mathcal{Q}_{\mathbf{x}}$. Transition probabilities are well-defined: q_{init} has n successors and its outgoing transition share the same probabilities and, for the states q_i^-, we have $x_i \leq m$.

21.1.2 Unbounded one-counter Markov decision processes

In this section, we adapt the reduction of [EWY10] from the square-root-sum to a verification problem in one-counter Markov chains to our OC-MDP formalism. We use this reduction to obtain lower-bounds for all of our interval strategy problems.

We fix inputs $x_1, \ldots, x_n$ and y to the square-root-sum problem, and let $m = \max_{1 \leq i \leq n} x_i$ and $\mathbf{x} = (x_1, \ldots, x_n)$. We define an OC-MDP $\mathcal{Q}_{\mathbf{x}}$ with only one action (i.e., a one-counter Markov chain) based on $\mathbf{x}$ such that the probability of terminating in a given state t is $\frac{1}{nm} \sum_{i=1}^{n} \sqrt{x_i}$ from a fixed initial state q_{init}.

We depict the fragment of $\mathcal{Q}_{\mathbf{x}}$ that is associated with x_i in Figure 21.1 for $i \in [\![1, n]\!]$. Formally, we define $\mathcal{Q}_{\mathbf{x}} = (Q_{\mathbf{x}}, \{a\}, \delta_{\mathbf{x}}, w_{\mathbf{x}})$ where $Q_{\mathbf{x}} = \{q_{\mathsf{init}}, t\} \cup \bigcup_{i=1}^{n} \{q_i, q_i^+, q_i^-\}$ and, for all $i \in [\![1, n]\!]$, transitions and weights to and from the state q_i, q_i^+, q_i^- and t match those in the illustration. For any $B \in \bar{\mathbb{N}}_{>0}$, we let $\mathcal{C}^{\leq B}(\mathcal{Q}_{\mathbf{x}})$ denote the Markov chain induced by the sole strategy of $\mathcal{M}^{\leq B}(\mathcal{Q}_{\mathbf{x}})$. We have the following theorem, which can also be seen as a corollary of Theorem 18.6.

Theorem 21.3 ([EWY10])**.** *We have* $\mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(q_{\mathsf{init}},1)}(\mathsf{Term}(t)) = \frac{1}{nm} \sum_{i=1}^{n} \sqrt{x_i}$ *and, for all* $1 \leq i \leq n$, $\mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(q_i,1)}(\mathsf{Term}(t)) = \frac{1}{m}\sqrt{x_i}$.

Due to the structure of $\mathcal{Q}_{\mathbf{x}}$, reaching t and terminating in t are equivalent.

Thus, Theorem 21.3 implies that we can reduce the square-root sum instance fixed above to the verification problem for selective termination and state-reachability on $\mathcal{Q}_{\mathbf{x}}$ for the unique (counter-oblivious) strategy of $\mathcal{M}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}})$, which is both an OEIS and a CIS, and the threshold $\theta = \frac{y}{nm}$. Furthermore, as there is only a single strategy, the answer to the verification problem is the same as the answer to the realisability problem for (pure or randomised) counter-oblivious strategies, which is a special case of the fixed-interval and parameterised interval strategy realisability problems for OEISs and CISs. We obtain the following hardness result.

Theorem 21.4. *The interval strategy verification, fixed-interval realisability and parameterised realisability problems for state-reachability and selective termination are square-root-sum-hard in unbounded OC-MDPs.*

Proof. The OC-MDP $\mathcal{Q}_{\mathbf{x}}$ and the threshold $\theta = \frac{y}{nm}$ can be computed in polynomial time. Therefore, we need only comment on the correctness of the reduction.

By Theorem 21.3 and due to the structure of $\mathcal{Q}_{\mathbf{x}}$, we have $\mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(s_{\mathsf{init}},1)}(\mathsf{Reach}(t)) = \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(s_{\mathsf{init}},1)}(\mathsf{Term}(t)) = \frac{1}{nm}\sum_{i=1}^{n}\sqrt{x_i}$. Let $\Omega \in \{\mathsf{Term}(t), \mathsf{Reach}(t)\}$. We clearly have $\mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(s_{\mathsf{init}},1)}(\Omega) \geq \theta$ if and only if $\sum_{i=1}^{n}\sqrt{x_i} \geq y$ in light of the above, and thus the reduction is correct. □

21.1.3 Bounded one-counter Markov decision processes

We now establish the square-root sum hardness for the interval strategy verification, fixed-interval realisability and parameterised realisability problems in bounded OC-MDPs. Intuitively, in most cases, the reduction consists in adding a counter upper bound to the reduction of Section 21.1.2. However, we will see that this does not always suffice.

We fix inputs $x_1, \ldots, x_n$ and y to the square-root-sum problem for the remainder of the section and let $m = \max_{1\leq i\leq n} x_i$ and $\mathbf{x} = (x_1, \ldots, x_n)$. Let $\theta = \frac{y}{nm}$ denote the threshold used for the reduction. We assume that for all $i \in [\![1, n]\!]$, $x_i \neq 0$, and thus that $m \geq 1$.

We aim to determine a bound $B \in \mathbb{N}_{>0}$ such that

$$\mathbb{P}_{\mathcal{C}^{\leq B}(\mathcal{Q}_{\mathbf{x}}),(q_{\mathsf{init}},1)}(\mathsf{Term}(t)) \geq \theta \text{ if and only if } \sum_{i=1}^{n} \sqrt{x_i} \geq y.$$

For all $B \in \mathbb{N}_{>0}$, let $\varepsilon_B = \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(q_{\mathsf{init}},1)}(\mathsf{Reach}(Q \times \{B\}))$. Using Theorem 21.3, we obtain that for all $B \in \mathbb{N}_{>0}$, we have

$$\begin{aligned}\mathbb{P}_{\mathcal{C}^{\leq B}(\mathcal{Q}_{\mathbf{x}}),(q_{\mathsf{init}},1)}(\mathsf{Term}(t)) &= \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(q_{\mathsf{init}},1)}(\mathsf{Term}(t)) - \varepsilon_B \\ &= \frac{1}{nm}\sum_{i=1}^{n}\sqrt{x_i} - \varepsilon_B.\end{aligned}$$

Therefore, we require a bound $B \in \mathbb{N}_{>0}$ with a polynomial-size representation such that the positive error term ε_B above is small enough to ensure the correctness of the reduction.

There is a particular case for which it is clear that no suitable B exists: whenever $\sum_{i=1}^{n} \sqrt{x_i} = y$ holds. This is not an issue however: this equality can be decided in polynomial time [BFHT85], and thus we consider a reduction that is conditioned on it. First, we check if the equality holds in polynomial time. If it does, we reduce to a fixed positive instance of the considered interval strategy problem (in bounded OC-MDPs). Otherwise, we mirror the reduction of the unbounded case with a well-chosen counter upper bound B.

We now state the two main results that we prove to establish the viability of the approach described above. First, the following result provides a lower bound on the distance between $\sum_{i=1}^{n} \sqrt{x_i}$ and y whenever these two values differ.

Lemma 21.5. *Let λ be the sum of the bit-sizes of $x_1, \dots, x_n$ and y. If $\sum_{i=1}^{n} \sqrt{x_i} \neq y$, then $|\sum_{i=1}^{n} \sqrt{x_i} - y| > 2^{-2^n(\lambda+1)}$.*

This bound is adapted to our formulation of the square-root-sum problem from [Tiw92, Lem. 3]. We prove this result in the first half of Section 21.1.4 through field-theoretic reasoning.

Lemma 21.5 implies that our reduction is correct whenever we ensure that $n \cdot m \cdot \varepsilon_B \leq 2^{-2^n(\lambda+1)}$ where λ denotes the sum of bit-sizes of the inputs to the square-root-sum problem. Choosing $B = 2^n m \cdot (\lambda + 1) + nm^2 + 1$ is sufficient to obtain the required bound on the approximation error due to the counter upper bound. This value of B can be computed in polynomial time as n is the number of inputs. We show that choosing this value of B is sufficient in the second half of Section 21.1.4 by bounding the error ε_B from above. Formally, we establish the following result.

Lemma 21.6. *Assume that $\sum_{i=1}^{n} \sqrt{x_i} \neq y$. Let λ denote the sum of bit-sizes of x_1, ..., x_n and y. For all $B \geq 2^n m \cdot (\lambda + 1) + nm^2 + 1$, it holds that*

$$n \cdot m \cdot \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(q_{\mathsf{init}},1)}(\mathsf{Reach}(Q \times \{B\})) \leq 2^{-2^n(\lambda+1)}.$$

By using Lemma 21.6, we can prove the square-root-sum hardness of our interval strategy problems in OC-MDPs via the reduction sketched above.

Theorem 21.7. *The interval strategy verification, fixed-interval realisability and parameterised realisability problems for state-reachability and selective termination are square-root-sum-hard in bounded OC-MDPs.*

Proof. The reduction only differs slightly between the three considered problem; we discuss the verification problem and comment on the additional steps for the other two problems below. The reduction is the same for state-reachability and selective termination, and thus we only mention the target in the following without specifying the objective. We consider inputs x_1, ..., x_n and y to the square-root-sum problem. We assume that for all $i \in [\![1, n]\!]$, $x_i \neq 0$. Let $m = \max_{1 \leq i \leq n} x_i$ and $\mathbf{x} = (x_1, \ldots, x_n)$. We describe the reduction and prove its correctness.

First, we check in polynomial time whether $\sum_{i=1}^{n} \sqrt{x_i} = y$. If this equality holds, we construct an OC-MDP $\mathcal{Q}$ with a single state q and a single action a where the self-loop of q labelled by a has weight -1. We reduce our instance of the square-root-sum problem to the verification problem on $\mathcal{Q}$ with counter upper bound $B = 2$, initial configuration $(q, 1)$, target $\{q\}$ and threshold $\theta = 1$. The reduction is trivially correct in this case and is in polynomial time.

If $\sum_{i=1}^{n} \sqrt{x_i} \neq y$, we construct the OC-MDP $\mathcal{Q}_{\mathbf{x}}$. Let $B = 2^n m \cdot (\lambda + 1) + nm^2 + 1$ where λ is the sum of bit-sizes of the inputs of the square-root-sum instance. We reduce our instance of the square-root-sum problem to the verification problem on $\mathcal{Q}_{\mathbf{x}}$ with counter upper bound B, initial configuration $(q_{\mathsf{init}}, 1)$, target $\{t\}$ and threshold $\theta = \frac{y}{nm}$. This reduction is in polynomial time, and thus it remains to prove its correctness.

Let $\varepsilon_B = \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(q_{\mathsf{init}},1)}(\mathsf{Reach}(Q \times \{B\}))$. It follows from Theorem 21.3 that we must show that $\frac{1}{nm}\sum_{i=1}^{n}\sqrt{x_i} - \varepsilon_B \geq \theta$ if and only if $\sum_{i=1}^{n}\sqrt{x_i} \geq y$. It is direct that $\frac{1}{nm}\sum_{i=1}^{n}\sqrt{x_i} - \varepsilon_B \geq \theta$ implies $\sum_{i=1}^{n}\sqrt{x_i} \geq y$. We prove the converse implication. Assume that $\sum_{i=1}^{n}\sqrt{x_i} \geq y$. By Lemmas 21.5 and 21.6, it holds that $\varepsilon_B \leq \frac{1}{nm}|\sum_{i=1}^{n}\sqrt{x_i} - y| = \frac{1}{nm}(\sum_{i=1}^{n}\sqrt{x_i} - y)$. We obtain that $\frac{1}{nm}\sum_{i=1}^{n}\sqrt{x_i} - \varepsilon_B \geq \frac{1}{nm}\cdot y = \theta$. This shows that the reduction is correct.

It remains to comment on how to adapt the above reduction to the fixed-interval and parameterised realisability problems. Instead of specifying the strategy as an input, we specify the interval partition $\mathcal{I} = [\![1, B-1]\!]$ for the fixed-interval case, and the parameters $d = 1$ for the number of intervals and $n = B - 1$ for the size of intervals in the parameterised case. These inputs are such that we check the existence of a well-performing counter-oblivious strategy (with respect to the threshold θ specified above). □

21.1.4 Details for bounded one-counter Markov decision processes

The goal of this section is to prove Lemma 21.5 and Lemma 21.6. The proof of the former uses field-theoretic tools whereas the proof of the latter consists mainly of computations used to derive an upper bound. We split this section into two parts, one for each result.

We let $x_1, \dots, x_n \in \mathbb{N}$ and $y \in \mathbb{N}$ be fixed for the remainder of this section.

Proof of Lemma 21.5

The goal of this section is to prove Lemma 21.5. This section has three parts. First, we recall some field-theoretic notions that are required for the proof. We refer the reader to [Lan02] for a reference on field theory. Second, we show that all roots of the minimal polynomial of $\sum_{i=1}^{n}\sqrt{x_i} - y$ are of a certain form. We end with a proof of Lemma 21.5.

Field-theoretic background. A complex number is *algebraic* (over $\mathbb{Q}$) if it is the root of a polynomial with rational coefficients. An *algebraic extension* of $\mathbb{Q}$ is a field K such that $\mathbb{Q} \subseteq K \subseteq \mathbb{C}$ (where $\mathbb{C}$ denotes the set of complex numbers) such that all elements of K are algebraic. Let $K \subseteq L \subseteq \mathbb{C}$ be algebraic extensions of $\mathbb{Q}$. We write L/K as shorthand to mean that L is an extension of K. The *minimum polynomial* of $\alpha \in L$ over K is the unique monic polynomial with coefficients in K of minimum degree that has α as a root. An algebraic number is an *algebraic integer* if its minimal polynomial over $\mathbb{Q}$ has integer coefficients. Algebraic integers form a sub-ring of the algebraic closure of $\mathbb{Q}$.

Given algebraic numbers $\alpha_1, \dots, \alpha_\ell$, we let $K(\alpha_1, \dots, \alpha_\ell)$ be the smallest algebraic extension of K containing α_1, ..., α_ℓ. The *degree* of the extension L/K, denoted by $[L : K]$, is the dimension of L as a vector space over K, and if $L = K(\alpha)$, $[L : K]$ is the degree of the minimal polynomial of α over K. Degrees of successive extensions multiply, in the sense that, given F/L, it holds that $[F : K] = [F : L] \cdot [L : K]$.

An extension L/K is *Galois* if and only if any embedding of K in the algebraic closure of $\mathbb{Q}$ induces an automorphism of K. Given a polynomial $P \in K[X]$, the *splitting field* of P is the smallest algebraic extension of K that contains all of the complex roots of P. If L/K is of finite degree, then L/K is Galois if and only if it is the splitting field of a polynomial in $K[X]$. If L/K is Galois, the Galois group $\mathsf{Gal}(L/K)$ of L/K is the group formed by the (field) automorphisms of L whose restriction to K is the identity function over K. The order of the Galois group of a Galois extension is the degree of the extension.

Minimal polynomials. The following lemma provides a set that is guaranteed to contain all roots of the minimal polynomial of $\sum_{i=1}^{n} \sqrt{x_i} - y$. Through this lemma, we can bound the coefficients of the minimal polynomial of $\sum_{i=1}^{n} \sqrt{x_i} - y$, which is the crux of the proof of Lemma 21.5.

Lemma 21.8. *Let $\beta = \sum_{i=1}^{n} \sqrt{x_i} - y$. The minimal polynomial of β over $\mathbb{Q}$ has at most 2^n roots and all are included in the set $\{\sum_{i=1}^{n} (-1)^{b_i} \sqrt{x_i} - y \mid (b_1, \dots, b_n) \in \{0, 1\}^n\}$.*

Proof. Let P_β denote the minimum polynomial of β and let $K = \mathbb{Q}(\sqrt{x_1}, \dots, \sqrt{x_n})$. To bound the number of roots of P_β, it suffices to bound its degree, i.e., $[\mathbb{Q}(\beta) : \mathbb{Q}]$. We have $\mathbb{Q}(\beta) \subseteq K$ because $\beta \in K$ by definition of K. It follows that $[\mathbb{Q}(\beta) : \mathbb{Q}]$ is a divisor of $[K : \mathbb{Q}]$. We have that $[K : \mathbb{Q}]$ is at most 2^n because

$$[K : \mathbb{Q}] = \prod_{0 \le i \le n-1} [\mathbb{Q}(\sqrt{x_1}, \dots, \sqrt{x_{i+1}}) : \mathbb{Q}(\sqrt{x_1}, \dots, \sqrt{x_i})]$$

and the degrees in the product are one or two; for all $1 \le i \le n$, $\sqrt{x_i}$ is a root of $X^2 - x_i$.

We now show that all roots of P_β are in $\{\sum_{i=1}^n (-1)^{b_i} \sqrt{x_i} - y \mid (b_1, \dots, b_n) \in \{0,1\}^n\}$. First, we note that K is the splitting field of the polynomial $\prod_{i=1}^n (X^2 - x_i)$. Therefore, $K/\mathbb{Q}$ is Galois (as its degree is finite).

We determine the Galois group of $K/\mathbb{Q}$. Let $R \subseteq \{\sqrt{x_1}, \dots, \sqrt{x_n}\}$ be a minimal set such that $K = \mathbb{Q}(R)$. Assume that $R = \{\sqrt{x_1}, \dots, \sqrt{x_{n'}}\}$. For all $1 \le i \le n'$, $[K : \mathbb{Q}(R \setminus \{\sqrt{x_i}\})] = 2$ and $K/\mathbb{Q}(R \setminus \{\sqrt{x_i}\})$ is Galois. It follows that the group $\mathsf{Gal}(K/\mathbb{Q})$ contains, for all $1 \le i \le n'$, the automorphism of K in $\mathsf{Gal}(K/\mathbb{Q}(R \setminus \{\sqrt{x_i}\}))$ that is such that $\sqrt{x_i} \mapsto -\sqrt{x_i}$ that leaves other elements of R unchanged. These different automorphisms commute. It follows that these automorphisms generate the Galois group $\mathsf{Gal}(K/\mathbb{Q})$ whose order is $2^{n'}$.

Let L denote the splitting field of P_β (thus $L/\mathbb{Q}$ is Galois). It holds that $L \subseteq K$, because $K/\mathbb{Q}$ is Galois and P_β has a root in K. On the one hand, elements of $\mathsf{Gal}(L/\mathbb{Q})$ are the restrictions of elements of $\mathsf{Gal}(K/\mathbb{Q})$. On the other hand, the action of $\mathsf{Gal}(L/\mathbb{Q})$ on the set of roots of P_β is transitive (because P_β is irreducible in $\mathbb{Q}[X]$). It follows that the roots are all of the claimed form. □

Proof of Lemma 21.5. We now provide a proof of Lemma 21.5.

Lemma 21.5. *Let λ be the sum of the bit-sizes of $x_1, \dots, x_n$ and y. If $\sum_{i=1}^n \sqrt{x_i} \neq y$, then $|\sum_{i=1}^n \sqrt{x_i} - y| > 2^{-2^n(\lambda+1)}$.*

Proof. Assume that $\sum_{i=1}^n \sqrt{x_i} \neq y$. Let P_β denote the minimal polynomial of $\beta = \sum_{i=1}^n \sqrt{x_i} - y$. It has integer coefficients, because β is an algebraic integer. Indeed, square roots of integers are algebraic integers and algebraic integers are a sub-ring of the algebraic closure of $\mathbb{Q}$. We bound the coefficients of P_β to conclude via the following result: the non-zero roots of a non-zero polynomial with k-bit integer coefficients are greater than 2^{-k} in absolute value [Hou70, Tiw92].

Let d denote the degree of P_β. We have $d \leq 2^n$ by Lemma 21.8. By the same result, it follows that the roots of P_β are of absolute value at most $\sum_{i=1}^{n} x_i + y < 2^\lambda$. From the decomposition of P_β in linear factors, we obtain that its coefficients are sums of at most 2^d products of roots of P_β, and thus are strictly less than $2^d \cdot (2^\lambda)^d = 2^{d(\lambda+1)}$ in absolute value, i.e., their bit-size is at most $d(\lambda+1)$. We obtain that $|\beta| > 2^{-d(\lambda+1)} \geq 2^{-2^n(\lambda+1)}$. □

Proof of Lemma 21.6

The goal of this subsection is to prove Lemma 21.6. For all $B \in \mathbb{N}_{>0}$, we let $\varepsilon_B = \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(q_{\mathsf{init}},1)}(\mathsf{Reach}(Q \times \{B\}))$.

First, we provide an explicit upper bound on the sequence $(\varepsilon_B)_{B\in\mathbb{N}_{>0}}$ that depends on m and B.

Lemma 21.9. *For all $B \in \mathbb{N}_{>0}$, $\varepsilon_B \leq (\frac{m}{m+1})^{B-1}$.*

Proof. All probability notation $\mathbb{P}$ in this proof refers to the Markov chain $\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}})$, and thus we omit the Markov chain from the notation to lighten it.

We have $\varepsilon_1 = 1 = (\frac{m}{m+1})^0$, and thus the inequality holds trivially for $B = 1$. To obtain the general result, we prove properties that hold for all $B \geq 2$.

For all $B \geq 2$ and $i \in [\![1, n]\!]$, let $\varepsilon_B^{(i)} = \mathbb{P}_{(q_i,1)}(\mathsf{Reach}((q_i, B)))$ denote the probability of hitting counter value B from $(q_i, 1)$. For all $B \geq 2$, we have $\varepsilon_B = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_B^{(i)}$ due to the structure of $\mathcal{Q}_{\mathbf{x}}$. To obtain the lemma, it suffices to show that for all $i \in [\![1, n]\!]$, we have $\varepsilon_B^{(i)} \leq (\frac{m}{m+1})^{B-1}$. We fix $i \in [\![1, n]\!]$.

For all $B \geq 2$, we let $\eta_B^{(i)} = \mathbb{P}_{(q_i,B-1)}(\mathsf{Reach}((q_i, B)))$ denote the probability of reaching counter value B from $(q_i, B-1)$. To conclude this proof, we establish three statements. First, we show that for all $B \geq 2$, we have $\varepsilon_{B+1}^{(i)} = \varepsilon_B^{(i)} \cdot \eta_{B+1}^{(i)}$. Second, we show that for all $B \geq 2$, we have $\eta_B^{(i)} \leq \frac{m}{m+1}$. Finally, we combine these two properties to conclude using an inductive argument.

For all $B \in \mathbb{N}_{>0}$ and $k \in [\![1, B]\!]$, we let $\mathcal{H}_{k\to B}$ denote the set of histories of

$\mathcal{M}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}})$ starting in (q_i, k) and ending in (q_i, B) with only one occurrence of this last configuration. The sets $\mathcal{H}_{k\to B}$ are prefix-free; the cylinders of their elements are pairwise disjoint.

We now show the first claim. Let $B \geq 2$. All histories of $\mathcal{H}_{1\to B}$ can be written as the concatenation of a history of $\mathcal{H}_{1\to B-1}$ and a history of $\mathcal{H}_{B-1\to B}$. We obtain

$$\begin{aligned}
\varepsilon_B^{(i)} &= \mathbb{P}_{(q_i,1)}(\mathsf{Cyl}\,(\mathcal{H}_{1\to B})) \\
&= \sum_{h_1\in\mathcal{H}_{1\to B-1}} \sum_{h_2\in\mathcal{H}_{B-1\to B}} \mathbb{P}_{(q_i,1)}(\mathsf{Cyl}\,(h_1\cdot h_2)) \\
&= \left(\sum_{h_1\in\mathcal{H}_{1\to B-1}} \mathbb{P}_{(q_i,1)}(\mathsf{Cyl}\,(h_1))\right)\cdot\left(\sum_{h_2\in\mathcal{H}_{B-1\to B}} \mathbb{P}_{(q_i,B-1)}(\mathsf{Cyl}\,(h_2))\right) \\
&= \varepsilon_{B-1}^{(i)}\cdot\eta_B^{(i)}.
\end{aligned}$$

This proves the first claim.

For the second claim, we show that the sequence $(\eta_B^{(i)})_{B\geq 2}$ is increasing and convergent, and that $\lim_{B\to\infty}\eta_B^{(i)} \leq \frac{m}{m+1}$. Let us prove that $(\eta_B^{(i)})_{B\geq 2}$ is increasing. Let $B \geq 2$. We consider the mapping $f_{+1}\colon \mathcal{H}_{B-1\to B} \to \mathcal{H}_{B\to B+1}$ that increases all counter values along a history by 1. This mapping is injective. Furthermore, for all $h \in \mathcal{H}_{B-1\to B}$, we have $\mathbb{P}_{(q_i,B-1)}(\mathsf{Cyl}\,(h)) = \mathbb{P}_{(q_i,B)}(\mathsf{Cyl}\,(f_{+1}(h)))$. It follows that

$$\eta_B^{(i)} = \mathbb{P}_{(q_i,B-1)}(\mathsf{Cyl}\,(\mathcal{H}_{B-1\to B})) \leq \mathbb{P}_{(q_i,B)}(\mathsf{Cyl}\,(\mathcal{H}_{B\to B+1})) = \eta_{B+1}^{(i)}.$$

This shows that $(\eta_B^{(i)})_{B\geq 2}$ is increasing.

The sequence $(\eta_B^{(i)})_{B\geq 2}$ is bounded and increasing, thus it converges. We prove that $\lim_{B\to\infty}\eta_B^{(i)} = \frac{m}{m+\sqrt{x_i}}$. To this end, we establish an inductive relation on the elements of this sequence: we prove that for all $B \geq 2$, we have $\eta_{B+1}^{(i)} = \frac{1}{2} + (1-\frac{x_i}{m^2})\eta_B^{(i)}\cdot\eta_{B+1}^{(i)}$. Let $B \geq 2$. By separating histories for which a counter increment occurs first from those for which a counter decrement occurs first, we obtain that

$$\eta_{B+1}^{(i)} = \frac{1}{2} + \frac{1}{2}\cdot\left(1-\frac{x_i}{m^2}\right)\cdot\mathbb{P}_{(q_i,B-1)}(\mathsf{Cyl}\,(\mathcal{H}_{B-1\to B+1})).$$

We obtain $\eta^{(i)}_{B+1} = \frac{1}{2}+(1-\frac{x_i}{m^2})\eta^{(i)}_{B-1}\cdot\eta^{(i)}_B$ by observing that elements of $\mathcal{H}_{B-1\to B+1}$ can be written as concatenations of elements of $\mathcal{H}_{B-1\to B}$ and $\mathcal{H}_{B\to B+1}$ and following the same reasoning as for $\varepsilon^{(i)}_B$ above.

By taking the limits on both sides of the above inductive relation, we obtain that $\lim_{B\to\infty}\eta^{(i)}_B = \frac{1}{2}+(1-\frac{x_i}{m^2})\cdot(\lim_{B\to\infty}\eta^{(i)}_B)^2$. If $m = 1$, then $x_i = 1$ (inputs are positive) and we directly obtain $\lim_{B\to\infty}\eta^{(i)}_B = \frac{1}{2} = \frac{m}{m+\sqrt{x_i}}$. Otherwise, if $m \geq 2$, we have $\frac{x_i}{m^2} < 1$ and we can solve a quadratic equation to deduce that $\lim_{B\to\infty}\eta^{(i)}_B \in \{\frac{m}{m+\sqrt{x_i}}, \frac{m}{m-\sqrt{x_i}}\}$. It follows from $\frac{m}{m-\sqrt{x_i}} > 1$ and $(\eta^{(i)}_B)_{B\geq 2}$ being a sequence of probabilities that $\lim_{B\to\infty}\eta^{(i)}_B = \frac{m}{m+\sqrt{x_i}}$. To end the proof of the second claim, we observe that since $(\eta^{(i)}_B)_{B\geq 2}$ is increasing and $x_i \geq 1$, we have, for all $B \geq 2$, $\eta^{(i)}_B \leq \frac{m}{m+\sqrt{x_i}} \leq \frac{m}{m+1}$.

We now combine the two claims to provide an inductive proof that $\varepsilon^{(i)}_B \leq (\frac{m}{m+1})^{B-1}$ for all $B \geq 2$. For $B = 2$, we have $\varepsilon^{(i)}_2 = \frac{1}{2} \leq \frac{m}{m+1}$ (because $m \geq 1$). We now assume that $\varepsilon^{(i)}_B \leq (\frac{m}{m+1})^{B-1}$ holds. Via the two claims, we conclude that $\varepsilon^{(i)}_{B+1} = \varepsilon^{(i)}_B \cdot \eta^{(i)}_{B+1} \leq (\frac{m}{m+1})^B$. □

The following result is a technical inequality required to show that the candidate for B for the reduction is well-chosen. We separate it from the main proof for the sake of clarity.

Lemma 21.10. *It holds that* $\frac{1}{\log_2(\frac{m+1}{m})} \leq m$.

Proof. The inequality above is equivalent to $1 + \frac{1}{m} \geq 2^{\frac{1}{m}}$. To prove this equivalent formulation, we show that the function $f\colon [1,+\infty[\to \mathbb{R}\colon z \mapsto 1+\frac{1}{z} - 2^{\frac{1}{z}}$ is non-negative. Let $z_0 = -\log_2(\ln(2))^{-1} > 1$. We show that f is increasing on $[1, z_0]$ and decreasing on $[z_0, +\infty[$. This property implies that f is non-negative. Indeed, on the one hand, we have $f(1) = 0$, implying that f is non-negative on $[1, z_0]$. On the other hand, because $\lim_{z\to+\infty} f(z) = 0$, f is necessarily non-negative on the interval $[z_0, +\infty[$.

We study the sign of the derivative of f to determine its intervals of monotonicity. We define $g\colon [1,+\infty[\to \mathbb{R}$ such that, for all $z \in [1,+\infty[$, $g(z) = \ln(2)\cdot 2^{\frac{1}{z}} - 1$. For all $z \in [1,+\infty[$, we have $f'(z) = \frac{1}{z^2}g(z)$. We obtain that for all $z \in [1, +\infty[$, the sign of $f'(z)$ depends only on the sign of $g(z)$. The function g is a decreasing function, because $]1, +\infty[\to \mathbb{R} \colon z \mapsto 2^{\frac{1}{z}}$ is a restriction of the composition of the decreasing function $]0, +\infty[\to \mathbb{R} \colon z \mapsto \frac{1}{z}$ and the increasing function $]0, +\infty[\to \mathbb{R} \colon z \mapsto 2^z$. Because $g(1) = 2\ln(2) - 1 > 0$ and $g(z_0) = 0$, it follows that f' is positive on the interval $]1, z_0]$ and negative on the interval $]z_0, +\infty[$. This implies the desired property for f, ending the proof. □

We can now show that choosing $B = 2^n m \cdot (\lambda + 1) + nm^2 + 1$ (where λ is the sum of bit-sizes of the inputs to our square-root sum instance) is sufficient to achieve the precision given by Lemma 21.5 that ensures the validity of the reduction.

Lemma 21.6. *Assume that $\sum_{i=1}^{n} \sqrt{x_i} \neq y$. Let λ denote the sum of bit-sizes of x_1, …, x_n and y. For all $B \geq 2^n m \cdot (\lambda + 1) + nm^2 + 1$, it holds that*

$$n \cdot m \cdot \mathbb{P}_{\mathcal{C}^{\leq\infty}(\mathcal{Q}_{\mathbf{x}}),(q_{\mathsf{init}},1)}(\mathsf{Reach}(Q \times \{B\})) \leq 2^{-2^n(\lambda+1)}.$$

Proof. We claim that it is sufficient to show that, for all $B \geq 2^n m \cdot (\lambda + 1) + nm^2 + 1$,

$$\left(\frac{m}{m+1}\right)^{B-1} \leq 2^{-2^n(\lambda+1)-nm}. \tag{21.1}$$

We observe that m, $n \in \mathbb{N}_{>0}$ implies that $2^{-nm} \leq \frac{1}{nm}$. Combining this with Lemma 21.9 implies that, for all $B \in \mathbb{N}_{>0}$ such that Equation (21.1) holds,

$$\varepsilon_B \leq \left(\frac{m}{m+1}\right)^{B-1} \leq 2^{-2^n(\lambda+1)-nm} \leq \frac{1}{nm} 2^{-2^n(\lambda+1)}.$$

This guarantees that establishing Equation (21.1) for the relevant upper bounds B is sufficient.

For all $B \in \mathbb{N}_{>0}$, Equation (21.1) is equivalent (by applying $\log_2$ on both sides then using algebraic manipulations) to

$$B - 1 \geq \frac{2^n(\lambda+1) + nm}{\log_2\left(\frac{m+1}{m}\right)}. \tag{21.2}$$

By Lemma 21.10, Equation (21.2) is guaranteed to hold whenever $B - 1 \geq$

$m(2^n(\lambda+1)+nm)$, and thus the same applies to Equation (21.1). This ends the proof of this lemma. □

21.2 NP-hardness for interval strategy realisability

We now prove that the realisability problem for counter-oblivious strategies is NP-hard for the selective termination objective. This implies the NP-hardness of the fixed-interval and parameterised realisability problems: counter-oblivious strategies are single-interval OEISs and are also CISs with a period of one. We prove this hardness result by a reduction from the problem of deciding if a directed graph has a Hamiltonian cycle. Formally, a Hamiltonian cycle is defined as follows.

Definition 21.11. Let $G=(V,E)$ denote a directed graph where V is a finite set of vertices and $E \subseteq V^2$ is a set of edges. A *Hamiltonian cycle* of G is a simple cycle $v_0v_1\dots v_r$ such that $r=|V|$, i.e., a cycle that passes through all vertices exactly once, except the first vertex which is visited twice.

Deciding whether a finite graph has a Hamiltonian cycle is NP-complete (e.g., [GJ79]).

We sketch a reduction from the problem of deciding if a graph has a Hamiltonian cycle to the counter-oblivious strategy realisability problem for selective termination. Let $G=(V,E)$ be a finite directed graph. We fix an initial vertex v_{init}. We derive an OC-MDP $\mathcal{Q}$ with deterministic transitions from G by adding vertices and redirecting transitions. We add a copy $v'_{\mathsf{init}} \notin V$ of the initial vertex and a fresh absorbing state $q \notin V$. All incoming transitions of v_{init} are redirected to v'_{init} and the only successor of v'_{init} is set to be q. All transitions are given a weight of -1. We assume a counter upper bound $B \in \{|V|+1,\infty\}$ that exceeds the initial counter value chosen below.

We claim that there is a Hamiltonian cycle in G if and only if there is a strategy guaranteeing (almost-)sure termination in v'_{init} from the initial configuration $(v_{\mathsf{init}}, |V|)$. Intuitively, all cycles of G from v_{init} with k edges (i.e., $k+1$ vertices) are equivalent to a history from $(v_{\mathsf{init}}, |V|)$ to $(v'_{\mathsf{init}}, |V|-k)$ in $\mathcal{M}^{\leq B}(\mathcal{Q})$. If there is a Hamiltonian cycle in G, because it is simple, we obtain a history of length $|V|$ in $\mathcal{M}^{\leq B}(\mathcal{Q})$ that can be obtained via a pure counter-oblivious strategy, and this strategy provides a positive answer to the realisability problem. For the converse, we observe that any other simple cycle from v_{init} of G yields a history terminating in the additional state q. Therefore, if there is no Hamiltonian cycle in G, all (randomised) counter-oblivious strategies have a history consistent with them that either terminates in q if v'_{init} is reached in under $|V|$ steps or terminates in V.

Theorem 21.12. *The problem of deciding whether there exists a counter-oblivious (pure or randomised) strategy ensuring almost-sure selective termination is* NP*-hard. In particular, the fixed-interval and parameterised realisability problems for selective termination are* NP*-hard.*

Proof. We provide a reduction from the NP-complete problem of deciding whether a finite directed graph contains a Hamiltonian cycle. We fix a finite directed graph $G = (V, E)$ and an initial vertex $v_{\mathsf{init}} \in V$ for the remainder of the proof.

We consider $\mathcal{Q} = (Q, A, \delta, w)$ such that $Q = V \cup \{v'_{\mathsf{init}}, q\}$ (where $v'_{\mathsf{init}}, q \notin V$) and $A = V$. The transition function is deterministic: we view it as a function $\delta \colon Q \times A \to Q$. We formalise δ as follows. First, for all $(v, v') \in E$ such that $v' \neq v_{\mathsf{init}}$, we let $\delta(v, v') = v'$. Second, for all $v \in V$ such that $(v, v_{\mathsf{init}}) \in E$, we let $\delta(v, v_{\mathsf{init}}) = v'_{\mathsf{init}}$. Finally, for all $v \in V$, we let $\delta(v'_{\mathsf{init}}, v) = \delta(q, v) = q$. All weights are -1. Recall that counter-oblivious strategies can be seen as functions $\sigma \colon Q \to A$.

We show that the three following assertions are equivalent:

(i) there exists a Hamiltonian cycle of G;

(ii) there exists a pure counter-oblivious strategy σ of $\mathcal{Q}$ such that $\mathbb{P}^{\sigma}_{(v_{\mathsf{init}}, |V|)}(\mathsf{Term}(v'_{\mathsf{init}})) = 1$;

(iii) there exists a counter-oblivious strategy σ of $\mathcal{Q}$ such that $\mathbb{P}^{\sigma}_{(v_{\mathsf{init}}, |V|)}(\mathsf{Term}(v'_{\mathsf{init}})) = 1$;

We prove that (i) implies (ii) and that (iii) implies (i). The implication from (ii) to (iii) is direct.

We assume that there exists a Hamiltonian cycle $v_0v_1\dots v_{|V|}$ of G. Assume without loss of generality that $v_0 = v_{\mathsf{init}}$. It is easy to see that the pure counter-oblivious strategy σ such that $\sigma(v_\ell) = v_{\ell+1}$ for all $\ell \in [\![|V|-1]\!]$ ensures that $\mathbb{P}^{\sigma}_{(v_{\mathsf{init}},|V|)}(\mathsf{Term}(v'_{\mathsf{init}})) = 1$. The strategy σ is well-defined because $v_0v_1\dots v_{|V|-1}$ is a simple path. This shows that (i) implies (ii).

We now prove the contrapositive of the implication from (iii) to (i). Assume that there is no Hamiltonian cycle in G. Let σ be a counter-oblivious strategy. We show that termination occurs in a state other than v'_{init} with positive probability. If $\mathbb{P}^{\sigma}_{(v_{\mathsf{init}},|V|)}(\mathsf{Term}(v'_{\mathsf{init}})) = 0$, then the claim is direct. We assume that $\mathbb{P}^{\sigma}_{(v_{\mathsf{init}},|V|)}(\mathsf{Term}(v'_{\mathsf{init}})) > 0$. Thus, there exists a history $h = (v_0,|V|)v_1(v_1,|V|-1)\dots(v_{|V|-1},1)v_{\mathsf{init}}(v'_{\mathsf{init}},0)$ consistent with σ such that $v_0 = v_{\mathsf{init}}$. Since there are no Hamiltonian cycles in G and h induces a cycle of G, there must be a state other than v_{init} (due to the structure of $\mathcal{Q}$) that is repeated in this induced cycle. Let $0 < \ell < \ell' < |V|$ such that $v_\ell = v_{\ell'}$. It is easy to see that the history starting in $(v_0,|V|)$ that follows h up to index ℓ' and then loops in the cycle between v_ℓ and $v_{\ell'}$ until termination is consistent with σ and therefore $\mathbb{P}^{\sigma}_{(v_{\mathsf{init}},|V|)}(\mathsf{Term}(v'_{\mathsf{init}})) < 1$. This ends the proof that (iii) implies (i).

To conclude, we note that $\mathcal{Q}$ can be constructed in polynomial time. This ends our NP-hardness proof. □

Part VI:

Concluding remarks

Chapter 22

Conclusion

We close this manuscript with a conclusion and a discussion of future works. We do not provide a summary in this chapter; we refer the reader to Chapter 3 for an extended overview of our contributions. Summaries of each parts can also be found in the first chapter of each part: see Chapter 4 for our results regarding Nash equilibria, Chapter 8 for our classification of finite-memory strategies, Chapter 12 for our results on the structure of payoff sets in multi-objective MDPs and Chapter 16 for interval strategies in one-counter MDPs.

Contents

22.1 Conclusion

We circle back to one of the key questions highlighted in Chapter 1: *what makes a strategy complex?* Our results lead us to believe that there are *several dimensions* to strategy complexity. In this manuscript, we have explored three faces of strategy complexity. We first focused on the classical notion of *memory* measured by the size of Mealy machines implementing strategies. We have then moved on to *randomisation*, exploring both the *differences in expressiveness* of randomised strategies and *randomisation requirements* in multi-objective MDPs. Finally, we have considered *concise representations* of strategies in one-counter MDPs and provided verification and realisability algorithms for them. From a strategy complexity standpoint, this suggests that *alternative representations* of strategies can provide insight into the structure of their memory and how they make decisions. We briefly comment on each of these measures.

Memory. We first focus on *memory* via *Mealy machines*, the complexity measure considered in Part II. Before we discuss strategy complexity, let us take a brief step back and highlight why we believe that Mealy machines constitute a *natural model for memory*, without referring to their well-established relevance. Each state of a Mealy machine represents a piece of information summarising the past of the ongoing play. The update function of the Mealy machine models how this information progresses throughout the play: given the current knowledge and the latest observation, it combines them into a summary of the new history. Finally, the next-move function models the idea that decisions should depend only on the current knowledge and the latest observation. In fact, all strategies follow this scheme: for an infinite-memory strategy, it suffices to keep track of the *entire history* of the ongoing play.

The above suggests that Mealy machines can be used to model all strategies that conform to the *intuitive idea* of a *finite-memory* strategy. In a Mealy machine, each memory state corresponds to a set of histories after which the strategy behaves in the same way. Therefore, the memory of a strategy quantifies the number of different (long-term) behaviours that the strategy exhibits from some point on. The amount of memory of a strategy is a natural way of quantifying strategy complexity. It is particularly useful from a theoretical standpoint, as it enables the formulation of general results that hold for all arenas (in a given class), regardless of their specific traits.

Nonetheless, memory by itself does not fully describe the complexity of a strategy. This is due to this measure attributing *the same complexity* to all strategies that can be implemented with the same amount of memory. This can lead to an underestimation or an overestimation of strategy complexity for practical applications. On the one hand, when reasoning with memory, memoryless strategies are the *base unit*, and all are considered to be *equally simple*. However, this is not necessarily the case: constant strategies are simpler than injective strategies in general. On the other hand, even if a large memory state space is required, it may admit a concise well-structured representation. Let us consider winning strategies in zero-sum energy-parity games on deterministic arenas for the sake of illustration [CD12a]. Winning strategies in these games require an exponential memory. However, it suffices to use strategies that alternate between increasing a counter value up to a threshold and reaching a good state. In other words, with our interval strategy terminology, one can win in these games with two-interval OEISs. Therefore, there is a gap between the complexity given by memory and a practical implementation of the strategy. Such observations motivate our multi-dimensional vision of strategy complexity, to complement the information given by memory measured through Mealy machines.

Randomised strategies. We now move on to randomisation. We have seen that there exist several ways of integrating randomisation in strategic decision making. We can distinguish *various classes of randomised strategies*, some of which appear more simple than others. The most general classes are the classical *mixed and behavioural strategies*. Beyond these strategies, several classes of randomised strategies can be defined through *variants of stochastic Mealy machines*, the expressiveness of which we studied in Part III. Finally, we can define subclasses of randomised strategies independently of Mealy machines. In particular, in Part IV, we showed that *finite-support mixed strategies* often suffice in multi-objective MDPs.

These different classes highlight that the *randomisation requirements* to win, reach an equilibrium or achieve a vector in a game do not boil down to simply evaluating whether randomisation is necessary or not, and can be studied in a finer way. A finer understanding of randomisation requirements can help in understanding *why it is required*, i.e., its purpose in the game, and to what extent. It can also be useful to minimise randomness in decision making if randomness is best avoided; think, e.g., of medical applications in which randomness is undesirable.

Randomisation can fulfil several roles depending on the application at hand. For multi-objective MDPs, we saw that the only interest of randomisation is to *balance different objectives*. In the rock paper scissors game in Example 2.3 (Page 46), randomisation is necessary to make oneself *unpredictable*. In games with imperfect information, randomisation can prove useful to *compensate a* *lack of information* (see, e.g., [RCDH07]). In the latter two cases, randomisation is used to optimise the performance of the strategy, and thus a natural question is to understand how little randomisation suffices to this end.

Alternative strategy representations. As explained above, Mealy machine representations may obscure the structure of the memory of a strategy; this is undesirable when it is well-structured, e.g., based on counters. Furthermore, finite memory does not guarantee the existence of a finite representation of the strategy if the state space is infinite. This latter observation motivates and justifies our restriction to *interval strategies* in one-counter MDPs in Part V.

We believe that it is a worthwhile endeavour to understand when Mealy machines are *well-structured* and to take advantage of this structure to obtain more *compact representations of strategies*. Having a small representation of the elements of the memory state space, e.g., as vectors of numbers that can be represented concisely in binary (e.g., as in multi-dimensional energy games [CRR14, JLS15]), is not sufficient in general to be able to concisely represent strategies: it is also imperative that the *next-move function* can also be concisely represented. This additional constraint increases the challenge of identifying relevant models.

Nonetheless, representing strategies through approaches other than Mealy machines can provide more concise strategy representations. Another advantage that can be obtained through such representations is *explainability*, i.e., the ability to explain the decisions made by a strategy. Understanding the structure of the memory of a strategy also benefits us from an explainability standpoint. Explainability constitutes another motivation of works on decision tree representations of memoryless strategies (e.g., [BCC$^+$15, BCKT18, JKW23]).

We remark that *conciseness* and *explainability* are two different distinct aspects of strategy complexity. For instance, *neural networks* are used in reinforcement learning to learn and represent strategies in (discounted-sum) MDPs (e.g., [SB18]). While neural networks yield relatively small strategies, their behaviour can be quite opaque due to their numerous parameters.

The many faces of strategy complexity. Strategy complexity has many different components and finding simple strategies can be seen as a *multi-objective optimisation problem.* This is highlighted by the *trade-offs* that can arise between different aspects of strategy complexity. For instance, in some games, memory can be traded for randomisation, i.e., we can reduce memory requirements by increasing the randomisation power of the strategy (e.g., [CdH04, Hor09, CRR14, MPR20]), or there can be a trade-off between explainability and conciseness. Measures of strategy complexity can be *quantitative*, e.g., the amount of memory required or the size of a representation, or *qualitative*, e.g., the randomisation model or the level of explainability.

More generally, we believe that we would benefit from a *comprehensive theory* of strategy complexity based on this *multi-dimensional vision.* A refined understanding of strategy complexity is key to design simpler controllers for practical applications. Through the work presented in this thesis, we provide a first step in the direction of a general theory of strategy complexity.

22.2 Future works

We briefly comment on some future works that arise from each part of this thesis.

Memory requirements for constrained equilibria. A natural variation of the main question tackled in Part II is to transpose it to *other classes of equilibria* than Nash equilibria. We briefly comment on the challenges that arise for a classical alternative to NEs: *subgame perfect equilibria* (SPE) [Sel65] in non-zero-sum reachability games on deterministic arenas. Intuitively, an SPE from an initial state is a strategy profile such that, for all histories starting in the initial state, there is no profitable deviation from the profile when we assume that the history has taken place (i.e., in the subgame starting from the history). SPEs are a refinement of NEs that avoid the issue of *non-credible threats*, i.e., it forbids players from threatening something that would negatively impact their payoff.

In Part II, we constructed finite-memory NEs from NE outcome. For SPEs, instead of a single outcome, we have to deal with a tree-like structure [BBGR21] that accounts for all histories. This makes it more challenging to obtain finite-memory SPEs in infinite arenas, and to obtain arena-independent memory upper bounds for reachability, even when considering move-dependent Mealy machines. An avenue to establishing such results would be to prove that the tree-like description of SPEs can be simplified as we have done for NE outcomes (e.g., by removing redundancies between branches of the tree), in such a way that we can derive small strategies from it.

The power of randomised strategies. In Part III, we have investigated the expressiveness of different models of stochastic Mealy machines. Outcome-equivalence is specification-agnostic: two outcome-equivalent strategies induce the same behaviour. Therefore, a variant of the problem would be to identify classes of *specifications* and *arenas* for which there are additional inclusions or equalities in our lattice. We are currently investigating one of these variants: we study how our lattice is affected when considering the *value* of a one-dimensional payoff in zero-sum *turn-based* stochastic games.

It is crucial that some restrictions are made on the setting and range of considered specifications. Indeed, for multi-objective specifications on subclasses of finite concurrent stochastic two-player arenas, the examples presented in Chapter 11 imply that our lattice in finite perfect information arenas would remain unchanged with this comparison criterion.

Memory in multi-objective Markov decision processes. The results of Part IV provide insight into *randomisation* requirements in multi-objective MDPs: finite-support mixed strategies can match the expectation of any strategy when dealing with universally integrable payoffs, and these strategies can be used to approximate any expected payoff for universally unambiguously integrable payoffs. These results assume strategies with possibly *infinite memory*. This leads to the question of understanding when *finite memory* suffices in multi-objective MDPs to achieve vectors. In general, such questions are addressed by devising *sufficient conditions* on payoffs or through *characterisations* of such payoffs (see, e.g., [GZ05, Gim07, BLO$^+$22, BORV23] for such results for similar questions). In addition to this question, one can also ask when *RDD strategies* are as powerful as general strategies; this would yield a natural *finite-memory* analogue of our result for universally integrable payoffs.

Structure of strategies in finite-horizon MDPs. In Part V, we have defined and studied interval strategies in OC-MDPs. In general, optimal strategies need not exist in OC-MDPs. However, there is a special subclass of OC-MDPs in which there are uniformly optimal strategies for state-reachability: *finite-horizon MDPs*, i.e., OC-MDPs in which all weights are negative. Optimal actions for all states with a given counter value can be determined via *value iteration* (see Appendix A.2.2). Since uniformly optimal strategies are guaranteed to exist in this setting, one can ask whether they have any *regular structure*. Examples 17.2 and 17.3 can be adapted to illustrate that OEISs and CISs are not sufficient to play optimally in finite-horizon MDPs. These examples do not exclude the possibility that strategies built on an *ultimately periodic* partition suffice to play optimally from all configurations in a finite-horizon MDP; it is open whether this is the case or not. This question can be generalised to the study of the *structure of optimal strategies* in unbounded OC-MDPs whenever optimal strategies exist.

Going further. As mentioned in the previous section, there is value in a general *theoretical framework* for strategy complexity, e.g., to enable the design of simple controllers. Developing such a framework can be done through different lines of work. For instance, identifying *relevant measures* of strategy complexity, both for theory and practice, appears necessary to thoroughly grasp complexity. This also requires understanding the *relationships* between these different measures, e.g., how they affect one another. For individual specifications, determining *ad hoc models* that yield *small controllers* also falls within the scope of this line of work: they can yield insight into the structure of the memory (in the general sense) of strategies. We believe in the interest of these questions, and intend to continue studying them.

Part VII:

Appendices

Chapter A

Additional preliminaries

This chapter complements Chapter 2. In Section A.1, we recall some topological notions, which are of particular usefulness for Chapter 15, in which we study continuous payoffs in multi-objective MDPs. We recall classical results regarding maximal probabilities of reachability objectives in Markov chains and MDPs in Section A.2. Sections A.3–A.7 contain proofs that were omitted from Chapter 2. In Section A.8, we show that the downward closure of a compact subset of $\bar{\mathbb{R}}^d$ is compact. Finally, we establish the continuity of some payoff functions in Section A.9.

Contents

A.1 Topology

This section recalls topological definitions as well as some classical results, mainly used above and in Chapter 15. We refer the reader to [Mun97] for a reference on topology.

A.1.1 Topology

Let X be a non-empty set. A *topology* over X is a set $\mathcal{T} \subseteq 2^X$ of subsets of X such that (i) $\emptyset$, $X \in \mathcal{T}$, (ii) for any family $(U_i)_{i\in I}$ such that $U_i \in \mathcal{T}$ for all $i \in I$, $\bigcup_{i\in I} U_i \in \mathcal{T}$ and (iii) if U, $U' \in \mathcal{T}$, then $U \cap U' \in \mathcal{T}$. The pair $(X, \mathcal{T})$ is called a *topological space*. Elements of $\mathcal{T}$ are *open sets*. A set $F \subseteq X$ is *closed* if it is the complement of an open set, i.e., if there exists $U \in \mathcal{T}$ such that $F = X \setminus U$.

We say that $(X, \mathcal{T})$ is a *Hausdorff space* when for any two distinct elements x and $y \in X$, there exists disjoint open sets U_x and U_y such that $x \in U_x$ and $y \in U_y$. We assume that all topological spaces below are Hausdorff.

Simple examples of topologies include the discrete topology $\mathcal{T}_{\mathsf{dis}} = 2^X$ and the trivial topology $\{\emptyset, X\}$. The discrete topology is Hausdorff. The trivial topology is not Hausdorff whenever X has at least two elements.

A set $N \subseteq X$ is a *neighbourhood* of $x \in X$ if there exists an open set $U_x \in \mathcal{T}$ such that $x \in U_x \subseteq N$. A set is open if and only if it is a neighbourhood of all of its elements. A point $x \in X$ is an *isolated point* if $\{x\}$ is a neighbourhood of x.

Let $Y \subseteq X$. The *closure* $\mathsf{cl}(Y)$ of Y is the smallest closed set in which Y is included. An element $x \in X$ is in $\mathsf{cl}(Y)$ if and only if all (open) neighbourhoods of x intersect Y. The *interior* $\mathsf{int}(Y)$ of Y is the greatest open set that is included in Y. An element $x \in X$ is in $\mathsf{int}(Y)$ if and only if there exists an open neighbourhood N_x of x such that $N_x \subseteq Y$. A set is closed (resp. open) if and only if it is equal to its closure (resp. interior).

A *base* of $\mathcal{T}$ is a set $\mathcal{B} \subseteq \mathcal{T}$ such that all elements of $\mathcal{T}$ are (arbitrary) unions of elements of $\mathcal{B}$. For instance, a topology is a base of itself. A base of the discrete topology is the set of all singleton sets. Another example is the usual topology of the extended real line $\bar{\mathbb{R}}$; a base of this topology is given by the set of intervals

$$\{]\alpha, \beta[\,, [-\infty, \alpha[\,,]\alpha, +\infty] \mid \alpha, \beta \in \mathbb{R},\ \alpha < \beta\}.$$

This topological space is Hausdorff.

Given a non-empty set $Y \subseteq X$, we define the *induced* (or subspace) topology $\mathcal{T}_{\mathsf{ind}}$ on Y as the topology defined by $\mathcal{T}_{\mathsf{ind}} = \{U \cap Y \mid U \in \mathcal{T}\}$. An element $x \in Y$ is an *isolated point of* Y if it is an isolated point in $(Y, \mathcal{T}_{\mathsf{ind}})$.

A.1.2 Metric and normed spaces

Let X be a non-empty set. A *metric* over X is a function $\mathsf{dist}\colon X \times X \to [0, +\infty[$ such that, for all x, y, $z \in X$, (i) $\mathsf{dist}(x, y) = 0$ if and only if $x = y$, (ii) $\mathsf{dist}(x, y) = \mathsf{dist}(y, x)$ and (iii) $\mathsf{dist}(x, z) \leq \mathsf{dist}(x, y) + \mathsf{dist}(y, z)$. The last condition is called the *triangle inequality*. An *open ball* centred in $x \in X$ of radius $\varepsilon > 0$ is the set $B(x, \varepsilon) = \{y \in X \mid \mathsf{dist}(x, y) < \varepsilon\}$. A base of the topology induced by a metric is the set of open balls. A topological space $(X, \mathcal{T})$ is *metrisable* if there exists a metric that induces $\mathcal{T}$. We remark that all metrisable spaces are necessarily Hausdorff.

For instance, the usual topology of $\mathbb{R}$ is metrisable and is induced by the metric dist defined by $\mathsf{dist}(x, y) = |x - y|$ for all x, $y \in \mathbb{R}$. The extended real line $\bar{\mathbb{R}}$ with its usual topology is also metrisable; it is homeomorphic (i.e., topologically isomorphic) to $[0, 1]$ with the induced topology inherited from $\mathbb{R}$. Topological spaces X with the discrete topology are also metrisable; the *discrete metric* $\mathsf{dist}_{\mathsf{disc}}$ defined by $\mathsf{dist}_{\mathsf{disc}}(x, y) = 1$ whenever $x \neq y$ induces the discrete topology (observe that all singleton sets are open balls).

Let $d \in \mathbb{N}_{>0}$. Any norm $\|\cdot\|$ on $\mathbb{R}^d$ induces a topology via the metric $\mathsf{dist}(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$ for all $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^d$. All norms of $\mathbb{R}^d$ are *equivalent*, i.e., induce the same topology, which is the usual topology of $\mathbb{R}^d$. For infinite-dimensional spaces, which we do not consider here, some norms may not be equivalent, and can induce different topologies.

A.1.3 Convergence

Let $(X, \mathcal{T})$ be a Hausdorff topological space. A sequence $(x_n)_{n\in\mathbb{N}}$ of elements of X is said to converge to $x \in X$ if for all open neighbourhoods U_x of x, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $x_n \in U_x$. It is equivalent to universally quantify only over open neighbourhoods of x in a fixed base of $\mathcal{T}$ instead of all open neighbourhoods. The uniqueness of the limit is guaranteed by the Hausdorff assumption. We note that in general spaces, sequences can have several limits: e.g., for the trivial topology $\{\emptyset, X\}$, all sequences converge to all elements of the set.

In a metric space (X, dist), this definition of convergence is equivalent to the usual definition recalled hereafter: for all $\varepsilon > 0$, there exists some $n_0 \in \mathbb{N}$ such that for all $n > n_0$, $\mathsf{dist}(x_n, x) < \varepsilon$ (i.e., $x_n \in B(x, \varepsilon)$).

Convergence of (real) sequences in $\bar{\mathbb{R}}$ in the above sense is equivalent to the classical definitions for convergence to a real limit, $+\infty$ or $-\infty$.

A.1.4 Continuity

A function $f\colon (X, \mathcal{T}) \to (Y, \mathcal{T}')$ is *continuous* at $x \in X$ if for all neighbourhoods $N_{f(x)} \subseteq Y$ of $f(x)$, $f^{-1}(N_{f(x)})$ is a neighbourhood of x. If $\mathcal{B}_Y$ is a basis of $(Y, \mathcal{T}')$, continuity can be checked by looking only at elements of the basis in the following sense: f is continuous at x if and only if for all $U_{f(x)} \in \mathcal{B}_Y$ such that $f(x) \in U_{f(x)}$, $f^{-1}(U_{f(x)})$ is a neighbourhood of x. The function f is said to be continuous if it is continuous at x for all $x \in X$.

If $f\colon (X, \mathsf{dist}_X) \to (Y, \mathsf{dist}_Y)$ is a function between metric spaces, the definition above is directly equivalent to the usual ε-δ definition of continuity: f is continuous at $x \in X$ if for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all

$x' \in X$, $\mathsf{dist}_X(x, x') < \delta$ implies that $\mathsf{dist}_Y(f(x), f(x')) < \varepsilon$.

For functions between metric spaces, there exists a stronger variant of continuity, called uniform continuity. A function $f\colon (X, \mathsf{dist}_X) \to (Y, \mathsf{dist}_Y)$ is *uniformly continuous* if for all $\varepsilon > 0$, there exists some $\delta > 0$ such that for all $x, x' \in X$, $\mathsf{dist}_X(x, x') < \delta$ implies that $\mathsf{dist}_Y(f(x), f(x')) < \varepsilon$. The difference with continuity is the quantification order. For continuity, δ may depend on both ε and the point at which we check continuity, whereas for uniform continuity, δ may only depend on ε and must work for all points.

For instance, the function $[0, +\infty[\to [0, +\infty[: x \mapsto \sqrt{x}$ is uniformly continuous. This can be shown via the observation that errors at a neighbourhood of some $x \geq 0$ can be bounded independently of x, i.e., we have that for all $x \geq 0$ and $|h| \leq x$ (this ensures that $\sqrt{x+h}$ is well-defined), we have $|\sqrt{x+h} - \sqrt{x}| \leq \sqrt{|h|}$. On the other hand, the function $\mathbb{R} \to \mathbb{R}\colon x \mapsto x^2$ is not uniformly continuous. For any $x \in \mathbb{R}$ and $h \in \mathbb{R}$, we have $|(x+h)^2 - x^2| = |h^2 - 2xh|$, and thus, intuitively, we cannot choose δ independently of x in the definition of continuity because errors depend on x (which cannot be bounded).

For functions from a metric space to another, continuity at x is equivalent to *sequential continuity* at x. A function $f\colon (X, \mathcal{T}) \to (Y, \mathcal{T}')$ is sequentially continuous at $x \in X$ if for all sequences $(x_n)_{n\in\mathbb{N}}$ that converge to x, the sequence $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(x)$.

We now prove a result implying that the non-negative and non-positive parts of a continuous function are continuous (for later use).

Lemma A.1. *Let $(X, \mathcal{T})$ be a topological space, $x \in X$, $f\colon X \to \bar{\mathbb{R}}$ be a function that is continuous at x and $M \in \mathbb{R}$. Then the functions $\min(f, M)\colon y \mapsto \min\{f(y), M\}$ and $\max(f, M)\colon y \mapsto \max\{f(y), M\}$ are continuous at x.*

Proof. We provide a proof only for $\min(f, M)$ as the argument is analogous for $\max(f, M)$. We distinguish three cases.

First, assume that $f(x) < M$. Let $N_{f(x)}$ be a neighbourhood of $f(x) = \min(f, M)(x)$. We must show that $\min(f, M)^{-1}(N_{f(x)})$ is a neighbourhood of x. By continuity of f at x, since $N_{f(x)} \cap [-\infty, M[$ is a neighbourhood of $f(x)$, $f^{-1}(N_{f(x)} \cap [-\infty, M[)$ is a neighbourhood of x. To close the first case, it suffices to establish that $f^{-1}(N_{f(x)} \cap [-\infty, M[) \subseteq \min(f, M)^{-1}(N_{f(x)})$. This inclusion follows from the fact that for all $y \in X$, if $f(y) < M$, then $f(y) = \min(f, M)(y)$. This ends the proof of the first case.

Second, assume that $f(x) > M$. Let N_M be a neighbourhood of $M = \min(f, M)(x)$. By definition of $\min(f, M)$, we obtain that for all $y \in f^{-1}(]M, +\infty])$, $\min(f, M)(y) = M$. It follows that $f^{-1}(]M, +\infty]) \subseteq \min(f, M)^{-1}(N_M)$. By continuity of f at x, $f^{-1}(]M, +\infty])$ is a neighbourhood of x. We conclude from the above that $\min(f, M)^{-1}(N_M)$ is a neighbourhood of x.

Finally, assume that $f(x) = M$ and let N_M be a neighbourhood of $M = \min(f, M)(x)$. It suffices to show that $f^{-1}(N_M) \subseteq \min(f, M)^{-1}(N_M)$ by continuity of f at x. Let $y \in X$ such that $f(y) \in N_M$. If $f(y) \leq M$, then $\min(f, M)(y) = f(y) \in N_M$. Otherwise, $\min(f, M)(y) = M \in N_M$. This shows the required inclusion and ends the proof. □

A.1.5 Compactness

A topological space $(X, \mathcal{T})$ is *compact* if for any open cover $(U_i)_{i \in I}$ of X (i.e., for all $i \in I$, U_i is open and $\bigcup_{i \in I} U_i = X$), one can extract a finite open cover of X, i.e., there exists $I' \subseteq I$ finite such that $\bigcup_{i \in I'} U_i = X$. A subset Y of a topological space $(X, \mathcal{T})$ is compact if $(Y, \mathcal{T}_{\mathsf{ind}})$ is compact, where $\mathcal{T}_{\mathsf{ind}}$ is the induced topology. A compact subset of a Hausdorff topological space is closed. For instance, any finite set with the discrete topology is compact. Any closed bounded interval of $\mathbb{R}$ is also compact. This can be used to show that the extended real line $\bar{\mathbb{R}}$ is compact (without relying on the fact that $\bar{\mathbb{R}}$ and $[0, 1]$ are homeomorphic).

In metrisable spaces, there is an equivalent characterisation of compactness based on sequences. A topological space $(X, \mathcal{T})$ is *sequentially compact* if for all sequences $(x_n)_{n \in \mathbb{N}}$ of X, there exists a convergent subsequence. A metrisable space is compact if and only if it is sequentially compact.

For subsets of $\mathbb{R}^d$ (and more generally, of finite-dimensional normed vector spaces), there is yet another equivalent formulation of compactness. A set $D \subseteq \mathbb{R}^d$ is compact if and only if it is closed and bounded.

Let $(X, \mathcal{T})$ and $(Y, \mathcal{T}')$ be topological spaces and let $f \colon X \to Y$. If $(X, \mathcal{T})$ is compact and f is continuous, then $f(X)$ is compact. Furthermore, if (X, dist_X) is a compact metric space and (Y, dist_Y) is a metric space, f is continuous if and only if f is uniformly continuous.

A.1.6 Product topology

The *product topology* is a topology defined over Cartesian products of topological spaces. Let I be an arbitrary non-empty set. For all $i \in I$, let $(X_i, \mathcal{T}_i)$ be a topological space. A base of open sets for the product topology over $\prod_{i\in I} X_i$ consists of the open sets $\prod_{i\in I} U_i$ where for all $i \in I$, $U_i \in \mathcal{T}_i$ and $U_i = X_i$ for all but finitely many $i \in I$. Such sets are called *cylinder sets*. The product topology is the coarsest topology for which the projections $\prod_{i'\in I} X_{i'} \to X_i \colon (x_{i'})_{i'\in I} \mapsto x_i$ are continuous for all $i \in I$.

A simple example of the product topology is $\mathbb{R}^d$; the usual topology of $\mathbb{R}^d$ corresponds to the product topology of d copies of $\mathbb{R}$ with its usual topology.

A sequence in a product of topological spaces converges with respect to the product topology if and only if it converges component-wise. This is formalised in the following result.

Lemma A.2. *Let I be a non-empty set. For all $i \in I$, let $(X_i, \mathcal{T}_i)$ be a topological space. Let $\mathcal{T}_\Pi$ denote the product topology on $\prod_{i\in I} X_i$. Let $(\mathbf{x}^{(n)})_{n\in\mathbb{N}}$ be a sequence of elements of $\prod_{i\in I} X_i$ and $\mathbf{x} = (x_i)_{i\in\mathbb{N}} \in \prod_{i\in I} X_i$. Then $(\mathbf{x}^{(n)})_{n\in\mathbb{N}}$ converges to $\mathbf{x}$ if and only if for all $i \in I$, $(x_i^{(n)})_{n\in\mathbb{N}}$ converges to x_i.*

A product of metrisable spaces need not be metrisable in general. However, countable products of metrisable spaces are metrisable. This can be shown in the same way that [Mun97, Chap 2, Thm. 9.5] proves that $\mathbb{R}^\omega$ with the product topology is metrisable (where the usual topology is assumed on $\mathbb{R}$).

Finally, we show that countable products of compact metrisable spaces are compact. In full generality, arbitrary products of compact topological spaces are compact; this result is known as Tychonoff's theorem. Tychonoff's theorem is equivalent to the axiom of choice. Because of this, we provide an alternative proof below for the case of countable products.

Theorem A.3. *For all $i \in \mathbb{N}$, let $(X_i, \mathcal{T}_i)$ be a sequentially compact topological space. Then $(\prod_{i\in\mathbb{N}} X_i, \mathcal{T}_\Pi)$ is sequentially compact, where $\mathcal{T}_\Pi$ denotes the product topology. In particular, countable products of compact metrisable spaces are* *compact.*

Proof. The second claim of the theorem follows from the first and the fact that, in metrisable spaces, compactness and sequential compactness are equivalent. We thus focus on the first claim of the theorem. Let $(\mathbf{x}^{(n)})_{n\in\mathbb{N}}$ be a sequence of elements of $\prod_{i\in\mathbb{N}} X_i$. Our goal is to show that $(\mathbf{x}^{(n)})_{n\in\mathbb{N}}$ has a convergent subsequence.

First, for all $i \in \mathbb{N}$, we construct a subsequence $(\mathbf{x}^{(n)})_{n\in I_i}$ of $(\mathbf{x}^{(n)})_{n\in\mathbb{N}}$ such that for all $j < i$, $(x_j^{(n)})_{n\in I_i}$ converges in X_j. We proceed by induction. We let $I_0 = \mathbb{N}$ for the base case. For the induction step, we assume that I_i is defined. By compactness of X_{i+1}, there exists $I_{i+1} \subseteq I_i$ such that $(x_i^{(n)})_{n\in I_{i+1}}$ converges. The induction hypothesis holds by construction because $(\mathbf{x}^{(n)})_{n\in I_{i+1}}$ is a subsequence of $(\mathbf{x}^{(n)})_{n\in I_i}$.

We now construct a convergent subsequence of $(\mathbf{x}^{(n)})_{n\in\mathbb{N}}$. Let $n_0 = \min I_1$ and, for all $i > 0$, let n_i be the least element of I_{i+1} strictly greater than n_{i-1}. It is easy to check that $(\mathbf{x}^{(n_i)})_{i\in\mathbb{N}}$ converges via Lemma A.2. □

A.2 Reachability in Markov chains and Markov decision processes

In this section, we recall properties of reachability objectives in Markov chains and Markov decision processes. We first describe a linear system characterising reachability probabilities in Markov chains. We use such linear systems in Chapters 18 and 19 to analyse memoryless strategies of MDPs induced by one-counter MDPs. We then describe a value iteration scheme to approximate the maximum reachability probabilities in MDPs. We use this scheme in Chapter 17.3.2 to analyse an example. We refer a reader to [BK08, Chap. 10] for details regarding the contents of this section.

A.2.1 Probabilities in Markov chains

Let $\mathcal{C} = (S, \delta)$ be a finite Markov chain and let $T \subseteq S$ be a set of targets. The probabilities $(\mathbb{P}_s(\mathsf{Reach}(T)))_{s\in S}$ are the least solution of a system of linear equations. We recall this system.

Theorem A.4. *Assume that $\mathcal{C}$ is finite. Let $\{S_{=0}, S_{=1}, S_?\}$ be a partition of S such that $S_{=0} \subseteq \{s \in S \mid \mathbb{P}_s(\mathsf{Reach}(T)) = 0\}$ and $T \subseteq S_{=1} \subseteq \{s \in S \mid \mathbb{P}_s(\mathsf{Reach}(T)) = 1\}$. We consider the system defined by $x_s = 0$ for all $s \in S_{=0}$, $x_s = 1$ for all $s \in S_{=1}$ and $x_s = \sum_{s' \in S} \delta(s)(s') \cdot x_{s'}$ for all $s \in S_?$. The least non-negative solution of this system is obtained by letting $x_s = \mathbb{P}_s(\mathsf{Reach}(T))$ for all $s \in S$. Furthermore, this system has a unique solution when $S_{=0} = \{s \in S \mid \mathbb{P}_s(\mathsf{Reach}(T)) = 0\}$.*

In the previous statement, the set $\{s \in S \mid \mathbb{P}_s(\mathsf{Reach}(T)) = 0\}$ only depends on the topology of the Markov chain, i.e., which states are connected by a transition. The transition probabilities do not matter: a state is in this set if and only if there are no histories starting in this state and ending in T.

A.2.2 Optimal probabilities in Markov decision processes

We now let $\mathcal{M} = (S, A, \delta)$ be a finite MDP and let $T \subseteq S$. The maximum reachability probability vector $(\max_{\sigma \in \Sigma(\mathcal{M})} \mathbb{P}^\sigma_s(\mathsf{Reach}(T)))_{s \in S}$ can be approximated by computing, for each state, the maximum probability of reaching T in no more than $k \in \mathbb{N}$ steps for increasing values of k. This maximum step-bounded reachability probability can be computed through an inductive technique called *value iteration*.

We introduce some notation for step-bounded reachability objectives. For all $k \in \mathbb{N}$, we let $\mathsf{Reach}^{\leq k}(T) = \{s_0 a_0 s_1 \ldots \in \mathsf{Plays}(\mathcal{M}) \mid \exists \ell \leq k,\, s_\ell \in T\}$ denote the set of plays in $\mathsf{Reach}(T)$ such that T is reached in no more than k transitions.

Theorem A.5. *Assume that $\mathcal{M}$ is finite and let $T \subseteq S$. For all $k \in \mathbb{N}$, let $\mathbf{v}^{(k)} = (v^{(k)}_s)_{s \in S} = (\max_{\sigma \in \Sigma(\mathcal{M})} \mathbb{P}^\sigma_s(\mathsf{Reach}^{\leq k}(T)))_{s \in S}$. For all $s \in T$ and $k \in \mathbb{N}$, $v^{(k)}_s = 1$ and for all $s \in S \setminus T$, $v^{(0)}_s = 0$ and, for all $k \in \mathbb{N}$,*

$$v^{(k+1)}_s = \max_{a \in A(s)} \sum_{t \in S} \delta(s, a)(t) \cdot v^{(k)}_t. \tag{A.1}$$

To play optimally with respect to a step-bounded reachability objective, it is necessary and sufficient to choose an action in the argument of the maximum in Equation (A.1) in s when $k+1$ steps remain.

A.3 Convex hulls of compact sets

We provide a proof of Lemma 2.2, which states that the convex hull of a compact subset of $\mathbb{R}^d$ is itself compact.

Lemma 2.2. *Let $d \in \mathbb{N}_{>0}$. Let $D \subseteq \mathbb{R}^d$. If D is compact, then $\mathsf{conv}(D)$ is also compact.*

Proof. Assume that D is compact. We assume that D is non-empty, as otherwise the result is direct.

We first show that $\mathsf{conv}(D)$ is bounded. Let $\mathbf{q} \in \mathsf{conv}(D)$. Let $\alpha_1, \dots, \alpha_n$ be convex combination coefficients and let $\mathbf{p}_1, \dots, \mathbf{p}_n \in D$ such that $\mathbf{q} = \sum_{m=1}^{n} \alpha_m \mathbf{p}_m$. By triangulation, we obtain that $\|\mathbf{q}\|_2 \leq \sum_{m=1}^{n} \alpha_m \|\mathbf{p}_m\|_2 \leq \sup\{\|\mathbf{p}\|_2 \mid \mathbf{p} \in D\} \in \mathbb{R}$, where the second inequality is a consequence of $\sum_{m=1}^{n} \alpha_m = 1$. It follows that $\mathsf{conv}(D)$ is bounded.

We now show that $\mathsf{conv}(D)$ is closed. Let $\mathbf{q} \in \mathbb{R}^d$ such that there exists a sequence $(\mathbf{q}^{(n)})_{n\in\mathbb{N}} \subseteq \mathsf{conv}(D)$ such that $\lim_{n\to\infty} \mathbf{q}^{(n)} = \mathbf{q}$. By Carathéodory's theorem for convex hulls (Theorem 2.1), all elements of the sequence $(\mathbf{q}^{(n)})_{n\in\mathbb{N}}$ are a convex combination of no more than $d+1$ elements of D. For all $n \in \mathbb{N}$, let $\alpha_1^{(n)}, \dots, \alpha_{d+1}^{(n)}$ be convex combination coefficients and let $\mathbf{p}_1^{(n)}, \dots, \mathbf{p}_{d+1}^{(n)}$ such that $\mathbf{q}^{(n)} = \sum_{j=1}^{d+1} \alpha_j^{(n)} \mathbf{p}_j^{(n)}$. By compactness of $[0,1]$ and D, we obtain an increasing sequence of natural numbers $(n_m)_{m\in\mathbb{N}}$, convex combination coefficients $\alpha_1, \dots, \alpha_{d+1}$ and $\mathbf{p}_1, \dots, \mathbf{p}_{d+1} \in D$ such that for all $1 \leq j \leq d+1$, $\lim_{m\to\infty} \alpha_j^{(n_m)} = \alpha_j$ and $\lim_{m\to\infty} \mathbf{p}_j^{(n_m)} = \mathbf{p}_j$. It follows (from the uniqueness of the limit) that $\mathbf{q} = \sum_{j=1}^{d+1} \alpha_j \mathbf{p}_j$. This shows that $\mathbf{q} \in \mathsf{conv}(D)$ and ends the proof that $\mathsf{conv}(D)$ is closed. □

A.4 Details regarding the topology over plays

Let $n \in \mathbb{N}_{>0}$ and $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$ be an $\mathbb{N}$-player arena. In this section, we present proofs of Lemma 2.9, which states that the set of history cylinders is a base of the topology of $\mathsf{Plays}(\mathcal{A})$ and Lemma 2.10, which implies that $\mathsf{Plays}(\mathcal{A})$ is compact whenever $\mathcal{A}$ is finite. We start with Lemma 2.9.

Lemma 2.9. *The set* $\{\mathsf{Cyl}\,(h) \mid h \in \mathsf{Hist}(\mathcal{A})\}$ *of history cylinders is a base of the topology of* $\mathsf{Plays}(\mathcal{A})$*.*

Proof. It suffices to show that any intersection of a (general) cylinder of $(S\bar{A})^\omega$ and $\mathsf{Plays}(\mathcal{M})$ can be written as a union of cylinders of histories.

Let $U = \left(\prod_{\ell=1}^{r} U_S^{(\ell)} \times U_{\bar{A}}^{(\ell)}\right) \times U_S^{(r+1)} \times (\bar{A}S)^\omega$ be a cylinder of $(S\bar{A})^\omega$, where $U_S^{(1)}, \ldots, U_S^{(r+1)}$ are (open) subsets of S and $U_{\bar{A}}^{(1)}, \ldots, U_{\bar{A}}^{(r)}$ are (open) subsets of $\bar{A}$ (all cylinders of $(S\bar{A})^\omega$ can be written this way). We obtain that $U \cap \mathsf{Plays}(\mathcal{A})$ is given by

$$\mathsf{Cyl}\left(\mathsf{Hist}(\mathcal{A}) \cap \left(\prod_{\ell=1}^{r} U_S^{(\ell)} \times U_{\bar{A}}^{(\ell)}\right) \times U_S^{(r+1)}\right),$$

which shows that $U \cap \mathsf{Plays}(\mathcal{A})$ is a union of history cylinders. □

We now prove Lemma 2.10.

Lemma 2.10. *The set* $\mathsf{Plays}(\mathcal{A})$ *is a closed subset of* $(S\bar{A})^\omega$*. In particular,* $\mathsf{Plays}(\mathcal{A})$ *is a compact space whenever* $\mathcal{A}$ *is finite.*

Proof. The first part of the statement implies the second, because all closed subsets of compact spaces are themselves compact [Mun97, Chap. 3, Thm. 5.2].

For all $w \in (S\bar{A})^*$, the set of continuations of w in $(S\bar{A})^\omega$ is an open set (it is a cylinder). Furthermore, any $u \in (S\bar{A})^\omega \setminus \mathsf{Plays}(\mathcal{A})$ has a prefix $w_u \in (S\bar{A})^*S$ that is not a history. We obtain that, for all $u \in (S\bar{A})^\omega \setminus \mathsf{Plays}(\mathcal{A})$, the set of continuations of w_u does not intersect $\mathsf{Plays}(\mathcal{A})$. Therefore, $(S\bar{A})^\omega \setminus \mathsf{Plays}(\mathcal{A})$ can be written as the union of the sets of continuations of each w_u, thus is an open set. We have shown that $\mathsf{Plays}(\mathcal{A})$ is closed. □

A.5 Distributions and mixed strategies

In this section, we prove two results. First, we prove Lemma 2.16, which states that the function mapping pure strategy profiles to the probability of a measurable set of plays is measurable. Second, we prove Lemma 2.17, which states that the probability of a measurable set of plays under a mixed strategy profile can be written as an integral over the probability of this set under pure strategies profiles.

Lemma 2.16. *Let $\Omega \subseteq \mathsf{Plays}(\mathcal{A})$ be measurable and let $s \in S$. The function $P_\Omega \colon \prod_{i=1}^{n} \Sigma^i_{\mathsf{pure}}(\mathcal{A}) \to [0,1] : \sigma \to \mathbb{P}^\sigma_s(\Omega)$ is measurable.*

Proof. We prove the above property in three steps. We first establish it directly for history cylinders. We extend the result to open subsets Ω of $\mathsf{Plays}(\mathcal{A})$ showing that P_Ω can be written as a sum or series of the form $\sum_{h\in\mathcal{H}} P_{\mathsf{Cyl}(h)}$ for a countable set of histories $\mathcal{H}$. Finally, we generalise to all measurable sets by induction on the Borel hierarchy (described below), by writing the function as a pointwise limit of measurable functions.

Let $h \in \mathsf{Hist}(\mathcal{A})$ and let $\Omega = \mathsf{Cyl}\,(h)$. We assume that $s = \mathsf{first}(h)$, as otherwise P_Ω is the constant zero function and the result is direct. By definition of probability measures over plays for pure strategies, there exists a constant θ such that, for all strategies profiles $\sigma \in \prod_{i=1}^{n} \Sigma^i_{\mathsf{pure}}(\mathcal{A})$ such that h is consistent (resp. inconsistent) with σ, we have $P_\Omega(\sigma) = \theta$ (resp. $P_\Omega(\sigma) = 0$). For each $i \in [\![1, n]\!]$, the set of pure strategies of $\mathcal{P}_i$ with which h is consistent is a generator of $\mathcal{F}_{\Sigma^i_{\mathsf{pure}}(\mathcal{A})}$. It follows that P_Ω is a linear combination of two indicators of measurable sets, and is therefore measurable.

We now assume that Ω is an open subset of $\mathsf{Plays}(\mathcal{A})$. By Lemma 2.9, we can write Ω as a countable union of histories, i.e., $\Omega = \mathsf{Cyl}\,(\mathcal{H})$ for a countable set of histories $\mathcal{H}$. We assume that the cylinders of histories in $\mathcal{H}$ are pairwise disjoint, as two cylinder sets have a non-empty intersection if and only if one is included in the other. It follows that $P_\Omega = \sum_{h\in\mathcal{H}} P_{\mathsf{Cyl}(h)}$ (by sigma-additivity of $\mathbb{P}^\sigma_s$ for all pure strategy profiles σ. This shows that P_Ω is the pointwise limit of a sequence of measurable functions and is thus measurable.

We now introduce the Borel hierarchy to handle general Borel sets. We refer the reader to [Kec95] for an extended exposition. Borel subsets of a metrisable topological space can be arranged in a hierarchy. Let ω_1 be the first uncountable ordinal. For $\mathsf{Plays}(\mathcal{A})$, this hierarchy is as follows. We let Σ^0_1 be the open subsets of $\mathsf{Plays}(\mathcal{A})$. For each ordinal $1 \leq \xi < \omega_1$, we let $\Pi^0_\xi = \{\mathsf{Plays}(\mathcal{A}) \setminus U \mid U \in \Sigma^0_\xi\}$

be the complements of the sets in Σ^0_ξ, and if $\xi > 1$, we let

$$\Sigma^0_\xi = \left\{ \bigcup_{\ell \in \mathbb{N}} U_\ell \mid U_\ell \in \Pi_{\xi_\ell}, \xi_\ell < \xi, \ell \in \mathbb{N} \right\}$$

be the set of countable unions of sets in $\bigcup_{\xi' < \xi} \Pi^0_{\xi'}$. Every Borel set is in one of the sets Σ^0_ξ for some $\xi < \omega_1$ and for all $1 < \xi' \leq \xi < \omega_1$, we have $\Sigma^0_{\xi'} \subseteq \Sigma^0_\xi$ and $\Pi^0_{\xi'} \subseteq \Pi^0_\xi$.

In the previous point, we have shown that P_Ω is measurable if $\Omega \in \Sigma^0_1$. This is our base case. For all ordinals $1 \leq \xi < \omega_1$, by showing that P_Ω is measurable for all $\Omega \in \Sigma^0_\xi$, we obtain that $P_\Omega = 1 - P_{\mathsf{Plays}(\mathcal{A}) \setminus \Omega}$ is measurable for all $\Omega \in \Pi^0_\xi$.

Let $1 < \xi < \omega_1$. We assume by induction that for all $1 \leq \xi' < \xi$ and for all $\Omega \in \Pi^0_{\xi'}$, P_Ω is measurable. Let $\Omega \in \Pi^0_\xi$. We show that P_Ω is measurable. Let $(\Omega_\ell)_{\ell \in \mathbb{N}}$ be a sequence of elements of $\bigcup_{\xi' < \xi} \Pi^0_{\xi'}$ such that $\Omega = \bigcup_{\ell \in \mathbb{N}} \Omega_\ell$. We let $\Omega_{\leq \ell} = \bigcup_{\ell' \leq \ell} \Omega_{\ell'}$. Since the sequence of sets $(\Omega_{\leq \ell})_{\ell \in \mathbb{N}}$ increases to Ω, it follows from the continuity of probability measures that P_Ω is the pointwise limit of $(P_{\Omega_{\leq \ell}})_{\ell \in \mathbb{N}}$. Thus, to conclude, it remains to show that for all $\ell \in \mathbb{N}$, $\Omega_{\leq \ell} \in \bigcup_{\xi' < \xi} \Pi^0_{\xi'}$ to conclude with the induction hypothesis. This property follows from the fact that for each $\xi' < \xi$, $\Pi^0_{\xi'}$ is stable by finite unions [Kec95, Prop. 22.1]. We have shown that P_Ω is measurable, which ends the inductive argument and the overall proof. □

Lemma 2.17. *Let $\mu = (\mu_i)_{i \in [\![1,n]\!]}$ be a mixed strategy profile and $s_{\mathsf{init}} \in S$ be an initial state. Let $\mu_1 \times \cdots \times \mu_n$ denote the (unique) product measure over $\prod_{i=1}^n \Sigma^i_{\mathsf{pure}}(\mathcal{A})$ obtained from $\mu_1, \cdots, \mu_n$. For all measurable $\Omega \subseteq \mathsf{Plays}(\mathcal{A})$, we have*

$$\mathbb{P}^\mu_{\mathcal{A}, s_{\mathsf{init}}}(\Omega) = \int_{\sigma \in \prod_{i=1}^n \Sigma^i_{\mathsf{pure}}(\mathcal{A})} \mathbb{P}^\sigma_{\mathcal{A}, s_{\mathsf{init}}}(\Omega) \mathrm{d}(\mu_1 \times \cdots \times \mu_n)(\sigma).$$

Proof. Let $\nu \in \mathcal{D}(\mathsf{Plays}(\mathcal{A}), \mathcal{F}_\mathcal{A})$ be such that for all measurable $\Omega \subseteq \mathsf{Plays}(\mathcal{A})$,

$$\nu(\Omega) = \int_{\sigma \in \prod_{i=1}^n \Sigma^i_{\mathsf{pure}}(\mathcal{A})} \mathbb{P}^\sigma_{\mathcal{A}, s_{\mathsf{init}}}(\Omega) \mathrm{d}(\mu_1 \times \cdots \times \mu_n)(\sigma).$$

The above integral is well-defined: we integrate a non-negative measurable function (see Lemma 2.16). It is easily checked that ν is a well-defined element of $\mathcal{D}(\mathsf{Plays}(\mathcal{A}), \mathcal{F}_{\mathcal{A}})$ (σ-additivity follows from linearity of the Lebesgue integral together with the monotone convergence theorem).

It suffices to show that $\mathbb{P}^{\mu}_{s_{\mathsf{init}}}(\Omega)$ agree over cylinders of histories starting in s_{init} to end the proof. Let $h \in \mathsf{Hist}(\mathcal{A}, s_{\mathsf{init}})$. For any pure strategy profile $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$, it follows from the definition of $\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h))$ that

$$\mathbb{P}^{\sigma}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h)) = \prod_{i=1}^{n} \mathbb{1}_{\Sigma^i_h}(\sigma_i) \cdot \prod_{\ell=0}^{r-1} \delta(s_\ell, \bar{a}_\ell)(s_{\ell+1}),$$

where, for all $i \in [\![1, n]\!]$, Σ^i_h is the set of pure strategies of $\mathcal{P}_i$ that is consistent with h. We obtain that $\nu(\mathsf{Cyl}\,(h)) = \mathbb{P}^{\mu}_{s_{\mathsf{init}}}(\mathsf{Cyl}\,(h))$ by injecting the above in the definition of ν and applying Fubini's theorem. □

A.6 Distributions over memory states of Mealy machines

The goal of this section is to prove Equation (2.1), which describes how the distribution over memory states of a Mealy machine changes from one step of a play to the next. To lighten notation, we only consider the two-player case. The argument is analogous if there are more than two players. We thus assume that $\mathcal{A}$ is a two-player arena, i.e., $\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$, for the remainder of the section.

We establish Equation (2.1) for a Mealy machine of $\mathcal{P}_1$. We fix a Mealy machine $\mathfrak{M} = (M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$ of $\mathcal{P}_1$ for the remainder of the section. We must first formalise what we mean by the distribution over M after $w \in (S\bar{A})^*$ occurs. We do this via the Markov chain over $\mathsf{Hist}(\mathcal{A}) \times M$ obtained when $\mathcal{P}_1$ plays according to $\mathfrak{M}$ and $\mathcal{P}_2$ plays according to a strategy σ_2 from an initial state s_{init}. We then prove Equation (2.1) by analysing this same Markov chain.

A.6.1 Induced Markov chain

We fix a strategy σ_2 of $\mathcal{P}_2$ and $s_{\mathsf{init}} \in S$ an initial state. We describe the Markov chain induced by playing $\mathfrak{M}$ and σ_2 from s_{init} in $\mathcal{A}$. The state space of this Markov chain is $\mathsf{Hist}(\mathcal{A}) \times M$. Transitions connect pairs of the form

$(h, m) \in \mathsf{Hist}(\mathcal{A}) \times M$ to pairs of the form $(h\bar{a}s, m')$; such a transition occurs with probability

$$\delta(\mathsf{last}(h), \bar{a})(s) \cdot \mathsf{up}_{\mathfrak{M}}(m, \mathsf{last}(h), \bar{a})(m') \cdot \\ \mathsf{nxt}_{\mathfrak{M}}(m, \mathsf{last}(h))(a^{(1)}) \cdot \sigma_2(h)(a^{(2)}).$$

A play of this Markov chain from (s_{init}, m) is of the form

$$(s_0, m_0)(s_0\bar{a}_0s_1, m_1)(s_0\bar{a}_0s_1\bar{a}_1s_2, m_2)\dots$$

and, therefore, we view it as a pair $(\pi, \mathbf{m}) \in \mathsf{Plays}(\mathcal{A}) \times M^\omega$ where $\pi = s_0\bar{a}_0s_1\dots$ and $\mathbf{m} = m_0m_1\dots$. We write $\mathbb{P}$ for the probability over $\mathsf{Plays}(\mathcal{A}) \times M^\omega$ induced by the above Markov chain with respect to the initial distribution $\nu_{\mathsf{init}} \in \mathcal{D}(\mathsf{Hist}(\mathcal{A}) \times M)$ defined by $\nu_{\mathsf{init}}(s_{\mathsf{init}}, m) = \mu_{\mathsf{init}}(m)$ for all $m \in M$ (and 0 elsewhere).[1]

We now introduce some random variables over $\mathsf{Plays}(\mathcal{A}) \times M^\omega$ to use in our derivation of Equation (2.1). Let $(\pi, \mathbf{m}) = (s_0\bar{a}_0s_1\dots, m_0m_1\dots)$. We use the following random variables. For all $\ell \in \mathbb{N}$, we let $S_\ell((\pi, \mathbf{m})) = s_\ell$, $\bar{A}_\ell((\pi, \mathbf{m})) = \bar{a}_\ell$, $M_\ell((\pi, \mathbf{m})) = m_\ell$. We let $A^{(1)}_\ell$ and $A^{(2)}_\ell$ be the random variables such that $\bar{A}_\ell = (A^{(1)}_\ell, A^{(2)}_\ell)$. We write W_ℓ for the random variable describing the sequence $W_\ell = S_0\bar{A}_0S_1\bar{A}_1\dots S_{\ell-1}\bar{A}_{\ell-1}$ which is the sequence read by $\mathfrak{M}$ prior to step ℓ. Similarly, we write H_ℓ for the random variable $H_\ell = W_\ell S_\ell$ that describes the history at step ℓ.

Next, we introduce some convenient notation. Let B denote a set. For any random variable $X\colon \mathsf{Plays}(\mathcal{A}) \times M^\omega \to B$ and $b \in B$, we write $\{X = b\}$ for $X^{-1}(\{b\})$ and omit the braces when evaluating $\mathbb{P}$ over such sets, e.g., we write $\mathbb{P}(X = b)$ for $\mathbb{P}(\{X = b\})$.

We now list three properties of the above random variables that are useful to formally derive Equation (2.1). First, we note that memory updates and state updates are independent. In particular, we have the following.

Claim A.6. *Let $h = s_0\bar{a}_0\dots\bar{a}_{\ell-1}s_\ell \in \mathsf{Hist}(\mathcal{A})$ such that $\mathbb{P}(H_\ell = h) > 0$. For*

[1] Given a Markov chain $\mathcal{C}' = (S', \delta')$, the distribution over plays of $\mathcal{C}'$ following an initial distribution $\nu_{\mathsf{init}} \in \mathcal{D}(S')$ is defined by $\mathbb{P}_{\mathcal{C}',\nu_{\mathsf{init}}}(\Omega) = \sum_{s\in S'} \nu_{\mathsf{init}}(s) \cdot \mathbb{P}_{\mathcal{C}',s}(\Omega)$ for all measurable $\Omega \subseteq \mathsf{Plays}(\mathcal{C}')$.

all $m \in M$, we have

$$\mathbb{P}(M_\ell = m \mid H_\ell = h) = \mathbb{P}(M_\ell = m \mid W_\ell = w),$$

where w denotes $s_0\bar{a}_0 \dots s_{\ell-1}\bar{a}_{\ell-1}$.

Proof. By definition of a conditional probability, it suffices to show that for all $m \in M$,

$$\mathbb{P}(M_\ell = m \wedge H_\ell = h) = \mathbb{P}(M_\ell = m \wedge W_\ell = w) \cdot \delta(s_{\ell-1}, \bar{a}_{\ell-1})(s_\ell) \tag{A.2}$$

and that

$$\mathbb{P}(H_\ell = h) = \mathbb{P}(W_\ell = w) \cdot \delta(s_{\ell-1}, \bar{a}_{\ell-1})(s_\ell).$$

We remark that the second equation can be obtained from Equation (A.2) by summing over all $m \in M$. Therefore, we focus on Equation (A.2) for the remainder of the proof. The result follows directly from the definition of probabilities in Markov chains.

We let $m_\ell \in M$. Let $s_\star \in S$ and $h_{s_\star} = ws$. We note that, $h_{s_\ell} = h$. We can write $\mathbb{P}(M_\ell = m_\ell \wedge H_\ell = h_{s_\star})$ as the sum

$$\sum_{(m_0,\dots,m_{\ell-1}) \in M^\ell} \mathbb{P}\left(\bigwedge_{j \in [\![\ell]\!]} M_j = m_j \wedge H_\ell = h_{s_\star}\right).$$

The sets described in the summed probabilities are cylinder sets of our Markov chain over $\mathsf{Hist}(\mathcal{A}) \times M$. Therefore, we can rewrite $\mathbb{P}(M_\ell = m_\ell \wedge H_\ell = h_{s_\star})$ as the following sum of products, by definition of distributions over plays of Markov chains:

$$\sum_{(m_0,\dots,m_{\ell-1}) \in M^\ell} \Bigg(\mathbb{P}\Bigg(\bigwedge_{j \in [\![\ell-1]\!]} M_j = m_j \wedge H_{\ell-1} = h_{\leq \ell-1}\Bigg) \cdot \\ \mathsf{nxt}_{\mathfrak{M}}(m, s_{\ell-1})(a^{(1)}_{\ell-1}) \cdot \sigma_2(h_{\leq \ell-1})(a^{(2)}_{\ell-1}) \cdot \\ \delta(s_{\ell-1}, \bar{a}_{\ell-1})(s_\star) \cdot \mathsf{up}_{\mathfrak{M}}(m_{\ell-1}, s_{\ell-1}, \bar{a}_{\ell-1})(m')\Bigg).$$

The term $\delta(s_{\ell-1}, \bar{a}_{\ell-1})(s_\star)$ can be factorised in front of the above sum. Equation (A.2) follows from the above equation and from the equality $\mathbb{P}(M_\ell = m_\ell \wedge W_\ell = w) = \sum_{s_\star \in S} \mathbb{P}(M_\ell = m_\ell \wedge H_\ell = h_{s_\star})$. □

Second, we have the following fact regarding memory updates.

Claim A.7. *Let $w = s_0\bar{a}_0 \dots s_\ell\bar{a}_\ell$ be a history prefix and $m' \in M$ such that $\mathbb{P}(W_\ell = w \wedge M_\ell = m') > 0$. Then, for all $m \in M$,*

$$\mathbb{P}(M_{\ell+1} = m \mid W_{\ell+1} = w \wedge M_\ell = m') = \mathsf{up}_{\mathfrak{M}}(m', s_\ell, \bar{a}_\ell)(m).$$

Proof. Let $m \in M$. We directly compute the above conditional probability to obtain the result. We first study $\mathbb{P}(M_{\ell+1} = m \wedge W_{\ell+1} = w \wedge M_\ell = m')$. We can write this probability as follows:

$$\sum_{(m_0,\dots,m_{\ell-1})\in M^\ell} \sum_{s\in S} \mathbb{P}\left(\bigwedge_{j\in[\![\ell+1]\!]} M_j = m_j \wedge H_{\ell+1} = ws\right),$$

where $m_\ell = m'$ and $m_{\ell+1} = m$. Fix $(m_0, \dots, m_{\ell-1}) \in M^\ell$ and $s \in S$. Let $h' = s_0\bar{a}_0 \dots s_\ell$ be the prefix of w obtained by removing its last action profile. By definition of $\mathbb{P}$, we obtain that

$$\begin{aligned}\mathbb{P}&\left(\bigwedge_{j\in[\![\ell+1]\!]} M_j = m_j \wedge H_{\ell+1} = ws\right)\\ &= \mathbb{P}\left(\bigwedge_{j\in[\![\ell]\!]} M_j = m_j \wedge H_\ell = h'\right) \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s_\ell)(a_\ell^{(1)}) \cdot \sigma_2(h')(a_\ell^{(2)})\cdot\\ &\qquad \delta(s_\ell, \bar{a}_\ell)(s) \cdot \mathsf{up}_{\mathfrak{M}}(m', s_\ell, \bar{a})(m).\end{aligned}$$

The factor $\mathsf{up}_{\mathfrak{M}}(m', s_\ell, \bar{a})(m)$ in this last equation does not depend on $(m_0, \dots, m_{\ell-1})$ nor on s. It thus follows (from all of the above) that $\mathbb{P}(M_{\ell+1} = m \wedge W_{\ell+1} = w \wedge M_\ell = m')$ can be factorised as an expression of the form $\mathsf{up}_{\mathfrak{M}}(m', s_\ell, \bar{a})(m) \cdot P(m', w)$ for some function P that only depends on m' and w, not on m.

Because the argument above can be adapted by replacing m by any $m'' \in M$, we obtain that $\mathbb{P}(W_{\ell+1} = w \wedge M_\ell = m')$ can be written as the sum

$$\sum_{m''\in M} \mathsf{up}_{\mathfrak{M}}(m', s_\ell, \bar{a})(m'') \cdot P(m', w) = P(m', w).$$

This yields the desired result. □

Finally, we observe the following equality regarding action choices of the two players.

Claim A.8. *Let* $h = s_0\bar{a}_0 \dots s_\ell \in \mathsf{Hist}(\mathcal{A})$ *and* $m \in M$ *such that* $\mathbb{P}(H_\ell = h \wedge M_\ell = m) > 0$. *For all action profiles* $\bar{a} = (a^{(1)}, a^{(2)}) \in \bar{A}(\mathsf{last}(h))$,

$$\mathbb{P}(\bar{A}_\ell = \bar{a} \mid H_\ell = h \wedge M_\ell = m) = \mathsf{nxt}_{\mathfrak{M}}(m, \mathsf{last}(h))(a^{(1)}) \cdot \sigma_2(h)(a^{(2)}).$$

Proof. We can rewrite $\mathbb{P}(\bar{A}_\ell = \bar{a} \mid H_\ell = h \wedge M_\ell = m)$ as the sum

$$\sum_{s\in S}\sum_{m'\in M} \mathbb{P}(\bar{A}_\ell = \bar{a} \wedge S_{\ell+1} = s \wedge M_{\ell+1} = m' \mid H_\ell = h \wedge M_\ell = m).$$

By definition of the transitions in our Markov chain, we can rewrite this sum as

$$\sum_{s\in S}\sum_{m'\in M} \Bigg(\mathsf{nxt}_{\mathfrak{M}}(m, s_\ell)(a^{(1)}) \cdot \sigma_2(h)(a^{(2)}) \cdot \\ \delta(s_\ell, \bar{a})(s) \cdot \mathsf{up}_{\mathfrak{M}}(m, s_\ell, \bar{a})(m')\Bigg).$$

By rearranging this sum and recalling that $\delta(s_\ell, \bar{a})$ and $\mathsf{up}_{\mathfrak{M}}(m, s_\ell, \bar{a})$ are distributions, we obtain the claim. □

A.6.2 Establishing Equation (2.1)

For any $m \in M$ and $w = s_0\bar{a}_1 s_1 \dots s_{\ell-1}\bar{a}_{\ell-1} \in (S\bar{A})^*$, the probability $\mu_w(m)$ over memory states after w occurs when $\mathcal{P}_1$ follows $\mathfrak{M}$ is formalised by the conditional probability $\mathbb{P}(M_\ell = m \mid W_\ell = w)$. This formulation suggests that this probability depends on σ_2. However, a by-product of the inductive relationship expressed by Equation (2.1) (recalled in Equation (A.3) below), is that it does not depend on σ_2. We now prove Equation (2.1).

Lemma A.9. *Let* $w' = s_0\bar{a}_0 s_1\bar{a}_1 \dots s_{\ell-1}\bar{a}_{\ell-1}$ *and* $w = w's_\ell\bar{a}_\ell \in (S\bar{A})^*$ *such that* $\mathbb{P}(W_{\ell+1} = w) > 0$. *For any* $m \in M$, *let* $\mu_{w'}(m) = \mathbb{P}(M_\ell = m \mid W_\ell = w')$

and $\mu_w(m) = \mathbb{P}(M_{\ell+1} = m \mid W_{\ell+1} = w)$*. For all* $m \in M$*, we have:*

$$\mu_w(m) = \frac{\sum_{m' \in M} \mu_{w'}(m') \cdot \mathsf{up}_{\mathfrak{M}}(m', s_\ell, \bar{a}_\ell)(m) \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s_\ell)(a_\ell^{(i)})}{\sum_{m' \in M} \mu_{w'}(m') \cdot \mathsf{nxt}_{\mathfrak{M}}(m', s_\ell)(a_\ell^{(i)})}. \tag{A.3}$$

In particular, μ_w *is independent of* σ_2*.*

Proof. We fix $m \in M$ for the whole proof.

It follows from the law of total probability (formulated for conditional probabilities) that

$$\begin{aligned}\mu_w(m) &= \mathbb{P}(M_{\ell+1} = m \mid W_{\ell+1} = w)\\ &= \sum_{m' \in M} \mathbb{P}(M_{\ell+1} = m \mid W_{\ell+1} = w \wedge M_\ell = m') \cdot \mathbb{P}(M_\ell = m' \mid W_{\ell+1} = w)\end{aligned}$$

Therefore, Claim A.7 implies that

$$\mu_w(m) = \sum_{m' \in M} \mathsf{up}_{\mathfrak{M}}(m', s_\ell, \bar{a}_\ell)(m) \cdot \mathbb{P}(M_\ell = m' \mid W_{\ell+1} = w). \tag{A.4}$$

We note that, for any $m' \in M$, the probability $\mathbb{P}(M_\ell = m' \mid W_{\ell+1} = w)$ is not $\mu_{w'}(m') = \mathbb{P}(M_\ell = m' \mid W_\ell = w')$. Using Bayes' theorem, we obtain a relation between $\mathbb{P}(M_k = m' \mid W_{k+1} = w)$ and $\mu_{w'}(m')$. We let $h' = w's_\ell$, i.e., h' is the history obtained by removing the last action pair of w. We note that $\{W_{\ell+1} = w\}$ and $\{H_\ell = h'\} \cap \{\bar{A}_\ell = \bar{a}_\ell\}$ both denote the same set. We obtain the following chain of equations:

$$\begin{aligned}&\mathbb{P}(M_\ell = m' \,|W_{\ell+1} = w)\\ &\qquad= \mathbb{P}(M_\ell = m' \wedge H_\ell = h' \mid W_{\ell+1} = w)\\ &\qquad= \frac{\mathbb{P}(W_{\ell+1} = w \mid M_\ell = m' \wedge H_\ell = h') \cdot \mathbb{P}(M_\ell = m' \wedge H_\ell = h')}{\mathbb{P}(W_{\ell+1} = w)}\\ &\qquad= \frac{\mathbb{P}(\bar{A}_\ell = \bar{a}_\ell \mid M_\ell = m' \wedge H_\ell = h') \cdot \mathbb{P}(M_\ell = m' \mid H_\ell = h')}{\mathbb{P}(\bar{A}_\ell = \bar{a}_\ell \mid H_\ell = h')}.\end{aligned}$$

The first equality is a consequence of $W_{\ell+1} = w$ implying $H_\ell = h'$. Bayes' theorem is used between lines two and three. To go from the third to the fourth line, both the numerator and denominator of the fraction have been multiplied by $\mathbb{P}(H_\ell = h')$ and the definition of conditional probabilities has been used to rewrite the denominator and the rightmost factor of the numerator.

We now analyse the three terms of the fraction above. First, by Claim A.6, we have $\mathbb{P}(M_\ell = m' \mid H_\ell = h') = \mathbb{P}(M_\ell = m' \mid W_\ell = w') = \mu_{w'}(m')$. Second, it follows from Claim A.8 that

$$\mathbb{P}(\bar{A}_\ell = \bar{a}_\ell \mid M_\ell = m' \wedge H_\ell = h') = \mathsf{nxt}_{\mathfrak{M}}(m', s_\ell)(a^{(1)}) \cdot \sigma_2(h')(a^{(2)}).$$

We now rewrite $\mathbb{P}(\bar{A}_\ell = \bar{a}_\ell \mid H_\ell = h')$ as

$$\sum_{\substack{m'' \in M \\ \mathbb{P}(M_\ell = m'' \mid H_\ell = h') > 0}} \mathbb{P}(\bar{A}_\ell = \bar{a}_\ell \mid M_\ell = m'' \wedge H_\ell = h') \cdot \mathbb{P}(M_\ell = m'' \mid H_\ell = h')$$

By Claim A.8, this is equal to

$$\sigma_2(h')(a_\ell^{(2)}) \cdot \sum_{m'' \in M} \mathsf{nxt}_{\mathfrak{M}}(m'', s_\ell)(a_\ell^{(1)}) \cdot \mu_{w'}(m'').$$

By injecting all of the above in Equation (A.4), we directly obtain Equation (A.3) (note that any term appearing in a denominator is non-zero by the assumption $\mathbb{P}(W_{\ell+1} = w) > 0$). □

A.7 Pure Nash equilibria in deterministic arenas

Let $n \in \mathbb{N}_{>0}$ and $\mathcal{A} = (S, (A^{(i)})_{i \in [\![1,n]\!]}, \delta)$ be a deterministic n-player arena. We provide a proof of Lemma 2.41, which states that in a multi-player game on $\mathcal{A}$, a player has a profitable deviation from an initial state with respect to a pure strategy profile if and only if they have a pure profitable deviation.

Lemma 2.41. *Assume that $\mathcal{A}$ is deterministic and that, for all $i \in [\![1,n]\!]$, f_i is a cost function. Let $s_{\mathsf{init}} \in S$ and $\sigma = (\sigma_i)_{i \in [\![1,n]\!]}$ be a pure strategy profile. Let $i \in [\![1,n]\!]$ and write $\sigma = (\sigma_i, \sigma_{-i})$. The following statements are equivalent:*

(i) $\mathcal{P}_i$ has a profitable deviation with respect to σ from s_{init};

(ii) there exists a play π from s_{init} consistent with σ_{-i} such that $f_i(\pi) < f_i(\mathsf{Out}_{\mathcal{A}}(\sigma, s_{\mathsf{init}}))$.

(iii) $\mathcal{P}_i$ has a pure profitable deviation with respect to σ from s_{init};

In particular, σ is an NE from s_{init} if and only if no player has a pure profitable deviation.

Proof. We observe that (iii) implies (i). It suffices to show that (i) implies (ii) and that (ii) implies (iii) to prove the lemma.

We first assume that there exists a profitable deviation τ_i of $\mathcal{P}_i$. By definition, we have that $\mathbb{E}^{\tau_i,\sigma_{-i}}_{s_{\mathsf{init}}}(f_i) < f_i(\mathsf{Out}_{\mathcal{A}}(\sigma, s_{\mathsf{init}}))$. It follows (from the compatibility of the Lebesgue integral with the order) that the set of plays $\{\pi \in \mathsf{Plays}(\mathcal{A}) \mid f_i(\pi) < f_i(\mathsf{Out}_{\mathcal{A}}(\sigma, s_{\mathsf{init}}))\}$ has positive $\mathbb{P}^{\tau_i,\sigma_{-i}}_{s_{\mathsf{init}}}$ probability. Since the set of plays that do not start in s_{init} or that are inconsistent with σ_{-i} have zero $\mathbb{P}^{\tau_i,\sigma_{-i}}_{s_{\mathsf{init}}}$-probability (see Remark 2.13), it follows that there exists a play π starting in s_{init} that is consistent with σ_{-i} such that $f_i(\pi) < f_i(\mathsf{Out}_{\mathcal{A}}(\sigma, s_{\mathsf{init}}))$. This proves that (i) implies (ii).

We now prove that (ii) implies (iii). Let $\pi = s_0\bar{a}_0s_1\bar{a}_1\ldots$ be a play from s_{init} consistent with σ_{-i} such that $f_i(\pi) < f_i(\mathsf{Out}_{\mathcal{A}}(\sigma, s_{\mathsf{init}}))$. We let $\tau_i \colon \mathsf{Hist}(\mathcal{A}) \to \bar{A}$ be a pure strategy such that $\tau_i(\pi_{\leq \ell}) = a^{(i)}_{\ell}$ for all $\ell \in \mathbb{N}$. We obtain that (τ_i, σ_{-i}) is a pure strategy profile and it is each to check that $\mathsf{Out}_{\mathcal{A}}((\tau_i, \sigma_{-i}), s_{\mathsf{init}}) = \pi$. By our assumption on π, we conclude that τ_i is a profitable deviation of $\mathcal{P}_i$ with respect to σ from s_{init}. This ends the proof. □

A.8 Downward closures of compact sets

We prove that the downward closure of a closed set of $\bar{\mathbb{R}}^d$ is a compact subset of $\bar{\mathbb{R}}^d$. The main tool used in the following proof is sequential compactness.

Lemma A.10. *Let $D \subseteq \bar{\mathbb{R}}^d$ be closed (i.e., compact) with respect to the topology of $\bar{\mathbb{R}}^d$. Then $\mathsf{down}(D)$ is compact with respect to the topology of $\bar{\mathbb{R}}^d$ and $\mathsf{down}(D) \cap \mathbb{R}^d$ is closed with respect to the topology of $\mathbb{R}^d$.*

Proof. Recall that $\bar{\mathbb{R}}^d$ is a compact metrisable space. Thus, it suffices to show that $\mathsf{down}(D)$ is closed with respect to the topology of $\bar{\mathbb{R}}^d$ to end the proof.

Let $\mathbf{q} \in \bar{\mathbb{R}}^d$ and $(\mathbf{q}_n)_{n\in\mathbb{N}}$ a sequence of elements of $\mathsf{down}(D)$ such that

$\mathbf{q}_n \to \mathbf{q}$ when $n \to \infty$. We must show that $\mathbf{q} \in \mathsf{down}(D)$. The idea of the proof is to bound $\mathbf{q}$ from above by a vector that is the limit of some sequence of elements of D.

For all $n \in \mathbb{N}$, we let $\mathbf{p}_n \in D$ such that $\mathbf{p}_n \geq \mathbf{q}_n$. By (sequential) compactness of D, $(\mathbf{p}_n)_{n\in\mathbb{N}}$ has a convergent subsequence. Let $\mathbf{p} \in D$ denote the limit of one such subsequence. It follows that $\mathbf{q} \leq \mathbf{p}$. This shows that $\mathbf{q} \in \mathsf{down}(D)$, and thus $\mathsf{down}(D)$ is a closed subset of $\bar{\mathbb{R}}^d$. Since the topology of $\mathbb{R}^d$ can be seen as the topology induced on $\mathbb{R}^d$ by that of $\bar{\mathbb{R}}^d$, it follows that $\mathsf{down}(D) \cap \mathbb{R}^d$ is a closed subset of $\mathbb{R}^d$. □

Remark A.11 (Assumption of Lemma A.10). The subsets of $\mathbb{R}^d$ that are closed with respect to the topology of $\bar{\mathbb{R}}^d$ are the compact subsets of $\mathbb{R}^d$. Therefore, the assumption of Lemma A.10 does not apply to all closed subsets of $\mathbb{R}^d$.

We present a closed subset of $\mathbb{R}^2$ the downward-closure of which is not a closed subset of $\mathbb{R}^2$. We consider $D = \{(-\frac{1}{n}, n) \mid n \in \mathbb{N}_0\}$. To see that D is $\mathbb{R}^2$-closed, consider a convergent sequence of elements of D. As it is a Cauchy sequence, from some point on, all subsequent elements of the sequence are at a distance of at most $\frac{1}{2}$ from one another. Because the distance between two different elements of D is at least 1, it follows the sequence that is considered is ultimately constant, thus its limit lies in D. This shows that D is $\mathbb{R}^2$-closed.

We now argue that $\mathsf{down}(D)$ and $\mathsf{down}(D) \cap \mathbb{R}^2$ are not closed: $(0, 0)$ is in the closure of these sets, but not in them. To show that $(0, 0) \in \mathsf{cl}(\mathsf{down}(D))$, we observe that the sequence $((-\frac{1}{n}, 0))_{n\in\mathbb{N}_0}$ is a sequence of elements of $\mathsf{down}(D)$ that converges to $(0, 0)$. On the other hand, we see that for all $n \in \mathbb{N}_0$, $(0, 0) \leq (-\frac{1}{n}, n)$ does not hold, i.e., $(0, 0) \notin \mathsf{down}(D)$. ◁

A.9 Examples of continuous payoffs

We present examples of continuous payoffs in this section. First, we show that (a generalisation of) the discounted-sum payoff is continuous. Second, we show that the shortest-path payoff is continuous whenever the considered weight function bounded from below by a positive constant. Finally, we provide characterisations in finite arenas of objectives whose indicator is continuous, and of prefix-independent payoffs that are continuous. We fix an arena $\mathcal{A} = (S, (A^{(i)})_{i\in[\![1,n]\!]}, \delta)$

for the remainder of the section.

A.9.1 Discounted-sum payoff

In the main text, we defined the discounted-sum payoff with a fixed discount factor. Certain authors study a variant of this payoff where the discount factor changes at each step of the play depending on the current state. We define a generalisation of such discounted-sum payoffs in which both the discount factor and the weights depend not only on the current state-action pair, but on the history and current action at each step. We provide sufficient conditions ensuring that this generalisation is well-defined and continuous.

We consider a history-dependent weight function $w\colon (S\bar{A})^+ \to \mathbb{R}$ and a history-dependent discount factor function $\lambda\colon (S\bar{A})^+ \to [0,1[$. We assume that w is no more than $W \in \mathbb{R}$ in absolute value and that λ is bounded away from 1, i.e., there exists $\lambda_\star \in [0,1[$ such that $\lambda(u) \leq \lambda_\star$ for all $u \in (S\bar{A})^+$. We define the *generalised discounted-sum payoff* as the function GDSum_w^λ defined, for all $\pi = s_0\bar{a}_0 s_1 \ldots \in \mathsf{Plays}(\mathcal{A})$, by

$$\mathsf{GDSum}_w^\lambda(\pi) = \sum_{r=0}^{\infty} \left(\prod_{\ell=0}^{r-1} \lambda(s_0\bar{a}_0 \ldots s_\ell \bar{a}_\ell) \right) w(s_0\bar{a}_0 \ldots s_r\bar{a}_r).$$

This function is well-defined for all plays: the defining series is absolutely convergent by the assumptions on w and λ. We now prove that GDSum_w^λ is continuous.

Lemma A.12. *Let $w\colon (S\bar{A})^+ \to \mathbb{R}$ be a history-dependent weight function such that w is no more than $W \in \mathbb{R}$ in absolute value and $\lambda\colon (S\bar{A})^+ \to [0,1[$ be a history-dependent discount factor function such that there exists $\lambda_\star \in [0,1[$ such that $\lambda(u) \leq \lambda_\star$ for all $u \in (S\bar{A})^+$. Then GDSum_w^λ is a continuous payoff.*

Proof. Let $\pi = s_0\bar{a}_1 s_1 \ldots \in \mathsf{Plays}(\mathcal{A})$ and $\varepsilon > 0$. Let $\ell \in \mathbb{N}$ such that $\frac{2\cdot W \cdot \lambda_\star^\ell}{1-\lambda_\star} < \varepsilon$ (whose existence is guaranteed by $\lambda_\star \in [0,1[$). Let $\pi' = t_0\bar{b}_0 t_0 \ldots \in \mathsf{Cyl}\,(\pi_{\leq \ell})$. For all $r \in \mathbb{N}$, let $u_r = s_0\bar{a}_0 \ldots s_r\bar{a}_r$ and $u'_r = t_0\bar{b}_0 \ldots t_r\bar{b}_r$. By definition of

GDSum_w^λ, we obtain that

$$\begin{aligned}
&\left|\mathsf{GDSum}_w^\lambda(\pi) - \mathsf{GDSum}_w^\lambda(\pi')\right| \\
&\qquad = \left|\sum_{r=\ell}^{\infty}\left(\left(\prod_{n=0}^{r-1}\lambda(u_n)\right)w(u_r) - \left(\prod_{n=0}^{r-1}\lambda(u'_n)\right)w(u'_r)\right)\right| \\
&\qquad \leq 2\cdot\sum_{r=\ell}^{\infty}\lambda_\star^r\cdot W \\
&\qquad = \frac{2\cdot W\cdot\lambda_\star^\ell}{1-\lambda_\star} \\
&\qquad < \varepsilon.
\end{aligned}$$

We have shown that GDSum_w^λ is continuous in π. □

A.9.2 Shortest-path payoff

We consider a weight function $w\colon S\times\bar{A}\to\mathbb{R}$ and a target $T\subseteq S$. The shortest-path payoff SPath_w^T is continuous over $\mathsf{Reach}(T)$, since the payoff of a play depends only on its prefix prior to the first visit to T. However, without imposing any conditions on w, the shortest-path payoff SPath_w^T is not necessarily continuous everywhere.

Example A.1. Consider a two-state MDP $\mathcal{M} = (\{s,t\},\{a\},\delta)$ (in fact, $\mathcal{M}$ is a Markov chain) such that $\delta(s,a)$ is the uniform distribution on $\{s,t\}$ and $\delta(t,a)(t) = 1$. Let w be a the constant zero weight function and consider the target $T = \{t\}$. The payoff SPath_w^T is not continuous at $\pi = (sa)^\omega$. Indeed, for all $\ell\in\mathbb{N}$, the play $\pi' = (sa)^{\ell+1}(ta)^\omega$ is such that $\pi_{\leq\ell} = \pi'_{\leq\ell}$ and $\mathsf{SPath}_w^T(\pi') = 0$. Therefore, for SPath_w^T to be continuous, 0 would have to be in all neighbourhoods of $\mathsf{SPath}_w^T(\pi) = +\infty$, which is not the case.

The same example can be used to show that $\mathbb{1}_{\mathsf{Reach}(T)}$ is not continuous in general. ◁

The problem occurring in the previous example is that there is a play with an infinite payoff that can be approached (in the sense of convergence) by plays that have a bounded payoff. A sufficient condition to avoid this phenomenon is to require that all weights are bounded from below by some positive constant; in finite arenas, this is equivalent to assuming that all weights are positive.

Under this condition, the payoff of a play is no less than the smallest weight multiplied by the number of actions occurring before the first visit to T. We show that this condition implies the continuity of SPath_w^T.

Lemma A.13. *Let $w\colon S \times \bar{A} \to \mathbb{R}$ be a weight function and $T \subseteq S$ be a target. Assume that there exists $\eta > 0$ such that $w(s, \bar{a}) \geq \eta$ for all $s \in S$ and all $\bar{a} \in \bar{A}$. Then SPath_w^T is continuous.*

Proof. Let $\pi = s_0\bar{a}_0s_1 \ldots \in \mathsf{Plays}(\mathcal{A})$. First, assume that $\mathsf{SPath}_w^T(\pi) \in \mathbb{R}$, i.e., $\pi \in \mathsf{Reach}(T)$. Let $r \in \mathbb{N}$ such that $s_r \in T$. By definition of SPath_w^T, for all $\pi' \in \mathsf{Cyl}\,(\pi_{\leq r})$, we have $\mathsf{SPath}_w^T(\pi) = \mathsf{SPath}_w^T(\pi')$. This implies that SPath_w^T is continuous in π.

Now, assume that $\mathsf{SPath}_w^T(\pi) = +\infty$. Let $M \in \mathbb{R}$. Let $r \in \mathbb{N}$ such that $M \leq r \cdot \eta$. We claim that for all $\pi' \in \mathsf{Cyl}\,(\pi_{\leq r})$, $\mathsf{SPath}_w^T(\pi') \geq M$. Let $\pi' \in \mathsf{Cyl}\,(\pi_{\leq r})$. If $\pi' \notin \mathsf{Reach}(T)$, the sought inequality is direct. We now assume that $\pi' \in \mathsf{Reach}(T)$. Because no state of T occurs in $\pi_{\leq r}$ and all weights are non-negative, we obtain that

$$\mathsf{SPath}_w^T(\pi') \geq \sum_{\ell=0}^{r} w(s_\ell, \bar{a}_\ell) \geq r \cdot \eta \geq M.$$

We have shown that SPath_w^T is continuous. □

A.9.3 Objectives and indicators

We now characterise objectives that have a continuous indicator function in a finite arena. Let Ω be an objective. The co-domain of an indicator function is $\{0, 1\}$. This implies that $\mathbb{1}_\Omega$ is continuous if and only if, for all plays π, there exist $\ell \in \mathbb{N}$ such that, for all plays $\pi' \in \mathsf{Cyl}\,(\pi_{\leq \ell})$, we have $\pi' \in \Omega$ if and only if $\pi \in \Omega$. Furthermore, if $\mathcal{A}$ is finite, it follows from uniform continuity that ℓ can be chosen independently of the play in the previous statement. Therefore, in finite arenas, $\mathbb{1}_\Omega$ is continuous if and only if membership in Ω depends only on a bounded prefix of plays. We show below that this is also equivalent to Ω being both open and closed.

Lemma A.14. *Assume that $\mathcal{A}$ is finite. Let Ω be an objective. The three following statements are equivalent:*

(i) *$\mathbb{1}_\Omega$ is continuous;*

(ii) *there exists $\ell \in \mathbb{N}$ such that for all $\pi \in \mathsf{Plays}(\mathcal{A})$, $\mathsf{Cyl}\,(\pi_{\leq \ell})$ is either included in or disjoint from Ω;*

(iii) *Ω is open and closed, i.e., there exist finitely many histories $h_1, \dots, h_n$ such that $\Omega = \bigcup_{m=1}^{n} \mathsf{Cyl}\,(h_m)$.*

Proof. We show that (i) and (ii) are equivalent, then that (ii) and (iii) are equivalent. In the interest of this proof being self-contained, we also show that the two properties in (iii) are equivalent at the end of the proof.

We first prove that (i) and (ii) are equivalent. Since $\mathbb{1}_\Omega$ is real-valued and $\mathcal{A}$ is finite, the payoff $\mathbb{1}_\Omega$ is uniformly continuous, i.e., for all $\varepsilon > 0$ there exists $\ell \in \mathbb{N}$ such that for all plays $\pi, \pi' \in \mathsf{Plays}(\mathcal{A})$, if $\pi_{\leq \ell} = \pi'_{\leq \ell}$, then $|\mathbb{1}_\Omega(\pi) - \mathbb{1}_\Omega(\pi')| < \varepsilon$. It follows from the co-domain of $\mathbb{1}_\Omega$ being $\{0, 1\}$ that $\mathbb{1}_\Omega$ is continuous if and only if there exists $\ell \in \mathbb{N}$ such that for all plays $\pi, \pi' \in \mathsf{Plays}(\mathcal{A})$, if $\pi' \in \mathsf{Cyl}\,(\pi_{\leq \ell})$, then $\mathbb{1}_\Omega(\pi) = \mathbb{1}_\Omega(\pi')$ (the non-trivial direction follows by choosing $\varepsilon = \frac{1}{2}$), i.e., $\mathsf{Cyl}\,(\pi_{\leq \ell})$ is included in or disjoint from Ω. This establishes the equivalence of (i) and (ii).

Next, we establish that (ii) and (iii) are equivalent. First, assume that (ii) holds and let $\ell \in \mathbb{N}$ be given by this property. We obtain that $\Omega = \bigcup_{\pi \in \Omega} \mathsf{Cyl}\,(\pi_{\leq \ell})$. There are finitely many histories of the form $\pi_{\leq \ell}$ ($\pi \in \mathsf{Plays}(\mathcal{A})$) because $\mathcal{A}$ is finite. This shows that (iii) holds. Conversely, assume that (iii) holds and let $h_1, \dots, h_n \in \mathsf{Hist}(\mathcal{A})$ such that $\Omega = \bigcup_{m=1}^{n} \mathsf{Cyl}\,(h_m)$. Property (ii) follows by letting ℓ be the greatest number of states in the histories $h_1, \dots, h_n$ or zero if there are no histories.

For the sake of completeness, we close the proof by showing that the two properties given in (iii) are equivalent. First, assume that Ω is open and closed. A base of the topology of $\mathsf{Plays}(\mathcal{A})$ is the set of history cylinders (Lemma 2.9), therefore Ω is a union of history cylinders. Since $\mathsf{Plays}(\mathcal{A})$ is compact and Ω is closed, it follows that Ω is compact. We conclude that Ω can be written as a finite union of cylinder sets by compactness.

Conversely, assume that $\Omega = \bigcup_{m=1}^{n} \mathsf{Cyl}\,(h_m)$ for some histories $h_1, \ldots, h_n$. It suffices to show that $\mathsf{Cyl}\,(h)$ is open and closed for all $h \in \mathsf{Hist}(\mathcal{A})$. Let $h = s_0\bar{a}_0 \ldots \bar{a}_{r-1}s_r \in \mathsf{Hist}(\mathcal{A})$. By definition of the product topology, $\mathsf{Cyl}\,(h)$ is open. To show that $\mathsf{Cyl}\,(h)$ is closed, we consider $\pi \in \mathsf{cl}(\mathsf{Cyl}\,(h))$ and show that $\pi \in \mathsf{Cyl}\,(h)$. By definition of closure, the cylinder $\mathsf{Cyl}\,(\pi_{\leq r})$ intersects $\mathsf{Cyl}\,(h)$, i.e., there exists a play with the prefixes $\pi_{\leq r}$ and h. It follows that $\pi_{\leq r} = h$, and thus $\pi \in \mathsf{Cyl}\,(h)$. We have shown that $\mathsf{Cyl}\,(h)$ is both open and closed, ending the proof. □

A.9.4 Prefix-independent payoffs

Prefix-independent functions assign the same payoff to any two plays that share a common suffix. Formally, a payoff $f\colon \mathsf{Plays}(\mathcal{A}) \to \bar{\mathbb{R}}$ is *prefix-independent* if for any play $\pi \in \mathsf{Plays}(\mathcal{A})$ and any $r \in \mathbb{N}$, $f(\pi) = f(\pi_{\geq r})$. The goal of this section is to characterise prefix-independent payoffs in *finite arenas*. We thus assume that $\mathcal{A}$ is finite for the remainder of the section.

We first introduce some notation. For any play $\pi = s_0\bar{a}_0s_1 \ldots \in \mathsf{Plays}(\mathcal{A})$, we let $\mathsf{inf}(\pi) = \{s \in S \mid \forall r \in \mathbb{N},\, \exists\, \ell \geq r,\, s_\ell = s\}$ denote the set of states that occur infinitely often in π.

Let $f\colon \mathsf{Plays}(\mathcal{A}) \to \bar{\mathbb{R}}$ be a prefix-independent payoff. Assume that f is continuous. We claim that the payoff of a play π is uniquely determined by $\mathsf{inf}(\pi)$. Let π and π' be two plays such that $\mathsf{inf}(\pi) = \mathsf{inf}(\pi')$. To prove that $f(\pi) = f(\pi')$, it suffices to show that in all neighbourhoods of π, there is a play whose payoff is $f(\pi')$. Indeed, this property together with the continuity of f implies that $f(\pi')$ is in all neighbourhoods of $f(\pi)$. By prefix-independence of f, we need only establish that in all neighbourhoods of π, there is a play that shares a common suffix with π'. Let $\ell \in \mathbb{N}$ such that $\mathsf{last}(\pi_{\leq \ell}) \in \mathsf{inf}(\pi)$. We construct a play π'' in $\mathsf{Cyl}\,(\pi_{\leq \ell})$ such that $f(\pi'') = f(\pi)$ as follows. Since $\mathsf{inf}(\pi) = \mathsf{inf}(\pi')$, there exists $r \in \mathbb{N}$ such that the first state of $\pi'_{\geq r}$ is $\mathsf{last}(\pi_{\leq \ell})$. We thus define $\pi'' = \pi_{\leq \ell} \cdot \pi'_{\geq r}$. By prefix-independence of f, we obtain that $f(\pi'') = f(\pi'_{\geq r}) = f(\pi')$. This shows that all neighbourhoods of π contain a play whose payoff is $f(\pi')$ and this closes our argument.

This argument above can be adapted to establish a more general property of continuous prefix-independent payoffs: for any two plays π and π', if $\mathsf{inf}(\pi)$ is reachable from $\mathsf{inf}(\pi')$, then both plays have the same payoff. We show that this property characterises continuous prefix-independent payoffs.

To formalise our characterisation, we define the *strongly connected components* (SCCs) of (the graph induced by) $\mathcal{A}$. An SCC is a maximal set of states $C \subseteq S$ such that, for all s, $t \in C$ there exists a history of $\mathcal{A}$ with at least one action starting in s and ending in t. An SCC C is reachable from an SCC C' if there exists a history starting in C' and ending in C.

We obtain the following characterisation. For the sake of conciseness, we apply the convention $0 \cdot (+\infty) = 0 \cdot (-\infty) = 0$ below.

Lemma A.15. *Assume that $\mathcal{A}$ is finite. Let f be a prefix-independent payoff function. Let C_1, …, C_k be the SCCs of $\mathcal{A}$. Then f is continuous if and only if there exist constants $\alpha_1, \ldots, \alpha_k \in \bar{\mathbb{R}}$ such that $f = \sum_{i=1}^{k} \alpha_i \cdot \mathbb{1}_{\mathsf{Büchi}(C_i)}$ and, for all $i, i' \in [\![1, k]\!]$, $\alpha_i = \alpha_{i'}$ whenever $C_{i'}$ is reachable from C_i.*

Proof. We first observe that the set of states visited infinitely often along a play is a subset of an SCC. In other words, for all plays $\pi \in \mathsf{Plays}(\mathcal{A})$, there exists a unique SCC C such that $\pi \in \mathsf{Büchi}(C)$; we say that π *stabilises* in C for short.

We first assume that f is continuous. Let π, $\pi' \in \mathsf{Plays}(\mathcal{A})$. We show that if π stabilises in an SCC C and π' stabilises in an SCC C' such that C is reachable from C', then $f(\pi) = f(\pi')$. Let $r \in \mathbb{N}$ such that $\pi_{\geq r}$ starts in a state of C. For all $\ell \in \mathbb{N}$, since C is reachable from C', there exists a play $\pi^{(\ell)} \in \mathsf{Cyl}\left(\pi'_{\leq \ell}\right)$ such that $\pi_{\geq r}$ is a suffix of $\pi^{(\ell)}$. By prefix-independence of f, for all $\ell \in \mathbb{N}$, we have $f(\pi^{(\ell)}) = f(\pi)$. Furthermore, all neighbourhoods of π' contain at least one play $\pi^{(\ell)}$ by construction. It follows from the continuity of f that $f(\pi)$ is in all neighbourhoods of $f(\pi')$, i.e., $f(\pi) = f(\pi')$.

The previous argument implies that the payoff of a play depends only on the SCC in which the play stabilises. For all $1 \leq i \leq k$, let α_i be the payoff of any play that stabilises in C_i. From the above, we obtain that $f = \sum_{i=1}^{k} \alpha_i \cdot \mathbb{1}_{\mathsf{Büchi}(C_i)}$ and that the coefficients α_i satisfy the required conditions. This ends the proof of the first implication.

We now let $\alpha_1, \ldots, \alpha_k \in \bar{\mathbb{R}}$ such that $f = \sum_{i=1}^{k} \alpha_i \cdot \mathbb{1}_{\mathsf{Büchi}(C_i)}$ and, for all $i, i' \in [\![1, k]\!]$, $\alpha_i = \alpha_{i'}$ whenever $C_{i'}$ is reachable from C_i. We show that f is continuous. Let $\pi \in \mathsf{Plays}(\mathcal{A})$. It suffices to show that f is constant over a neighbourhood of π. Let $1 \leq i \leq k$ such that π stabilises in C_i and let $r \in \mathbb{N}$

such that all states of $\pi_{\geq r}$ are in C_i. Let $\pi' \in \mathsf{Cyl}\left(\pi_{\geq r}\right)$. Then π' stabilises in an SCC that is reachable from C_i, thus $f(\pi') = \alpha_i = f(\pi)$. We have shown that f is constant over $\mathsf{Cyl}\left(\pi_{\geq r}\right)$, ending the proof of the second implication. □

Chapter B

Details of Section 14.1

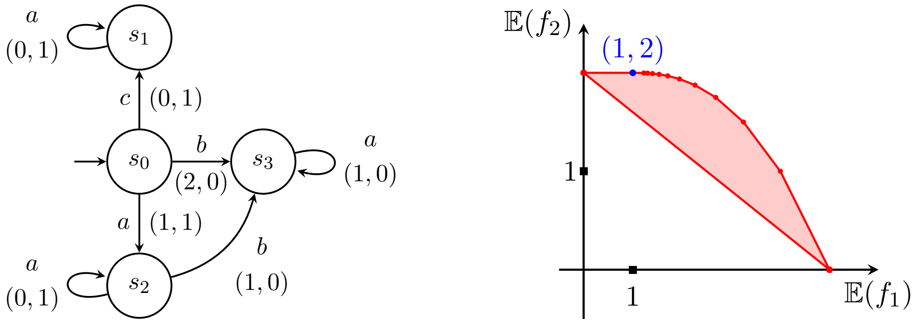
In this section, we formally prove the statements made with respect to the example presented in Chapter 14.1. We recall the MDP and its payoff set in Figures B.1a and B.1b. Throughout this section $\mathcal{M}$ refers to the MDP depicted in Figure B.1a and w denotes the two-dimensional weight function from the illustration. Recall that we consider the two-dimensional payoff $\bar{f} = (f_1, f_2)$ given by the discounted-sum payoffs $f_1 = \mathsf{DSum}_{w_1}^{3/4}$ and $f_2 = \mathsf{DSum}_{w_2}^{1/2}$.

First, we prove that the description of $\mathsf{Pay}_{s_0}^{\mathsf{pure}}(\bar{f})$ given in Chapter 14.1 is accurate. Second, we establish that all pure payoffs are extreme points and that all of these points except $(0, 2)$ are Pareto-optimal. Finally, we close the section by proving that all payoffs of pure strategies can *only* be obtained by playing without randomisation, and comment on the consequences in terms of strategy complexity.

Throughout this section, $\mathcal{M}$ refers to the MDP of Figure 14.1a.

Determining the set of payoffs of pure strategies

We provide the computations necessary to obtain the description of $\mathsf{Pay}_{s_0}^{\mathsf{pure}}(\bar{f})$ of Chapter 14.1. The argument is based on the fact that there are no randomised transitions in the MDP we consider. Therefore, the payoff of a pure strategy from s_0 is the payoff of a single play. In the following proof, we directly compute the payoff of each play from s_0.



(a) An MDP with deterministic transitions. Pairs next to actions represent two-dimensional weights.

(b) The set of expected payoffs for the MDP of Figure B.1a for the payoff $f_1 = \mathsf{DSum}^{3/4}_{w_1}$ and $f_2 = \mathsf{DSum}^{1/2}_{w_2}$.

Figure B.1: An MDP with a two-dimensional discounted-sum payoff $\bar{f}$ such that $\mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f}))$ is infinite.

Lemma B.1. *We have*

$$\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) = \{(0,2),(1,2)\} \cup \left\{ \left(1 + \frac{3^r}{4^{r-1}}, 2 - \frac{1}{2^{r-1}}\right) \mid r \in \mathbb{N} \right\}.$$

Proof. There are no randomised transitions in $\mathcal{M}$. For this reason, any pure strategy induces a single play in $\mathcal{M}$ from any starting state. We compute the payoff of all plays of $\mathcal{M}$ from s_0 to obtain the desired result.

We first consider the three plays $s_0c(s_1a)^\omega$, $s_0a(s_2a)^\omega$ and $s_0b(s_3a)^\omega$ that never leave their second state once it is reached. By definition of discounted-sum payoff functions, we have $\bar{f}(s_0c(s_1a)^\omega) = (0, \sum_{\ell=0}^{\infty} \frac{1}{2^\ell}) = (0,2)$, $\bar{f}(s_0a(s_2a)^\omega) = (1, \sum_{\ell=0}^{\infty} \frac{1}{2^\ell}) = (1,2)$ and $\bar{f}(s_0b(s_3a)^\omega) = (1 + \sum_{\ell=0}^{\infty} (\frac{3}{4})^\ell, 0) = (5,0) = (1 + \frac{3^0}{4^{0-1}}, 2 - \frac{1}{2^{0-1}})$.

It remains to deal with the plays that move from s_0 to s_2 and then eventually move to s_3. It suffices to show that for all $r \geq 1$, we have $\bar{f}(s_0(as_2)^r b(s_3a)^\omega) = (1 + \frac{3^r}{4^{r-1}}, 2 - \frac{1}{2^{r-1}})$. Let $r \geq 1$. We obtain, by definition of discounted-sum payoff functions, that

$$f_1(s_0(as_2)^r b(s_3a)^\omega) = 1 + \sum_{\ell=r}^{\infty} \frac{3^\ell}{4^\ell} = 1 + \frac{3^r}{4^r} \cdot \sum_{\ell=0}^{\infty} \frac{3^\ell}{4^\ell} = 1 + \frac{3^r}{4^{r-1}},$$

and

$$f_2(s_0(as_2)^r b(s_3a)^\omega) = \sum_{\ell=0}^{r-1} \frac{1}{2^\ell} = \frac{1-\frac{1}{2^r}}{1-\frac{1}{2}} = 2 - \frac{1}{2^{r-1}}.$$

This proves the required equality to end the proof. □

Extreme points and Pareto-optimality

We now show that $\mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f})) = \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$ in the context of this example (this property does not hold in full generality). It follows from $\mathsf{Pay}_{s_0}(\bar{f}) = \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}))$ that all extreme points of $\mathsf{Pay}_{s_0}(\bar{f})$ are the payoff of a pure strategy. It remains to show that $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \subseteq \mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f}))$. We first observe that $(0,2) \in \mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f}))$ because it is with the least first component among all elements of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$. The main difficulty lies with the elements of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$.

We handle the remaining points with a two-part argument. First, we show that any non-extreme element of the boundary of $\mathsf{Pay}_{s_0}(\bar{f})$ is a convex combination of two extreme points. This implies that all Pareto-optimal elements of $\mathsf{Pay}_{s_0}(\bar{f})$ are convex combinations of no more than two vectors. Second, we prove that for all vectors $\mathbf{q} \in \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$, $\mathbf{q}$ is either incomparable or strictly greater (with respect to the component-wise ordering) to convex combinations of any two vectors of $\mathsf{Pay}_{s_0}(\bar{f}) \setminus \{\mathbf{q}\}$. We prove this by reasoning on a strictly concave real function whose graph includes $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$. The graphical intuition is as follows: any segment joining two points of the graph of a strictly concave function is beneath the curve, so any points on the segment that are comparable to $\mathbf{q}$ must be smaller. This approach also yields that all vectors of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$ are Pareto-optimal elements of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$.

We now prove a generalisation of the first property formulated above: given a compact set $D \subseteq \mathbb{R}^2$, any vector in the boundary of $\mathsf{conv}(D)$ is a convex combination of no more than two elements of D. This can be seen as a refinement of Carathéodory's theorem for convex hulls (Theorem 2.1) when considering the boundary of the convex hull of a compact set in a two-dimensional setting.

Lemma B.2. *Let $D \subseteq \mathbb{R}^2$ be compact. For all $\mathbf{q} \in \mathsf{bd}(\mathsf{conv}(D))$, either $\mathbf{q} \in \mathsf{extr}(\mathsf{conv}(D))$ or $\mathbf{q}$ is a convex combination of two vectors of $D \setminus \{\mathbf{q}\}$.*

Proof. If D is empty or a singleton set, the result is direct. We thus assume that D has at least two elements. Let $\mathbf{q} \in \mathsf{bd}(\mathsf{conv}(D))$. Because D is compact, $\mathsf{conv}(D)$ is closed (Lemma 2.2) and thus $\mathsf{bd}(\mathsf{conv}(D)) \subseteq \mathsf{conv}(D)$. We let $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(n)} \in D$ and $\alpha_1, \ldots, \alpha_n \in]0,1]$ be non-zero convex combination coefficients such that $\mathbf{q} = \sum_{m=1}^{n} \alpha_n \mathbf{p}^{(m)}$.

We first show that the vectors $\mathbf{q}$, $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(n)}$ lie on a single line. If $\mathsf{aff}(D)$ is a line, then this is direct. We thus assume that $\mathsf{aff}(D)$ is not a line. We obtain that $\mathsf{aff}(D) = \mathbb{R}^2$: it contains a line because D has at least two elements, and therefore its dimension must be two. This implies that $\mathsf{ri}(D) = \mathsf{int}(D)$, and thus that $\mathbf{q} \notin \mathsf{ri}(D)$. By the supporting hyperplane theorem (Theorem 2.4), there exists a non-zero linear form x^* such that for all $\mathbf{p} \in \mathsf{conv}(D)$, $x^*(\mathbf{q}) \geq x^*(\mathbf{p})$. It follows that for all $1 \leq m \leq n$, we have $x^*(\mathbf{p}^{(m)}) = x^*(\mathbf{q})$ (by linearity, because $\alpha^{(m)} \neq 0$). In other words, $\mathbf{q}$ and the $\mathbf{p}^{(m)}$ lie on the line $(x^*)^{-1}(x^*(\mathbf{q}))$.

We have shown that $\mathbf{q}$ and the $\mathbf{p}^{(m)}$ lie on a single line. There exists a non-zero vector $\mathbf{v} \in \mathbb{R}^2$ such that for all $1 \leq m \leq n$, there exists $\beta_m \in \mathbb{R}$ such that $\mathbf{p}^{(m)} = \mathbf{q} + \beta_m \cdot \mathbf{v}$. There must exist $1 \leq m, m' \leq n$ such that $\beta_m \geq 0$ and $\beta_{m'} \leq 0$. We conclude that $\mathbf{q} \in \left[\mathbf{p}^{(m)}, \mathbf{p}^{(m')}\right]$. We have shown that $\mathbf{q}$ is a convex combination of at most two elements of D. □

We now introduce the strictly concave function $\mathcal{F}\colon [1,5] \to \mathbb{R}$ used in the remainder of our argument. For all $x \in [1,5]$, we let

$$\mathcal{F}(x) = 2 - \left(\frac{x-1}{3}\right)^{\log_{4/3}(2)}$$

We observe that $\mathcal{F}$ is well-defined in 0 because $\log_{4/3}(2) > 0$. We recall that $\mathcal{F}$ is strictly concave if and only if for all $x, x' \in [1,5]$ such that $x < x'$ and all $\beta \in]0,1[$, we have $\beta\mathcal{F}(x) + (1-\beta)\mathcal{F}(x') < \mathcal{F}(\beta x + (1-\beta)x')$. We now show that $\mathcal{F}$ is decreasing and strictly concave.

Lemma B.3. *The function $\mathcal{F}$ is decreasing and strictly concave.*

Proof. To prove that $\mathcal{F}$ is decreasing (resp. strictly concave), it suffices to show that it is differentiable over $]1,5[$ and its derivative $\mathcal{F}'$ is strictly negative (resp. decreasing). For the strict concavity of $\mathcal{F}$, in practice, we show that the second derivative $\mathcal{F}''$ of $\mathcal{F}$ (defined over $]1,5[$) is negative. For all $x \in]1,5[$, we have

$$\mathcal{F}'(x) = -\log_{4/3}(2) \cdot \left(\frac{x-1}{3}\right)^{\log_{4/3}(2)-1}$$

and

$$\mathcal{F}''(x) = -\log_{4/3}(2) \cdot \left(\log_{4/3}(2) - 1\right) \cdot \left(\frac{x-1}{3}\right)^{\log_{4/3}(2)-2}.$$

To prove that $\mathcal{F}'$ and $\mathcal{F}''$ are negative over $]1,5[$, it suffices to show that $\log_{4/3}(2) > 0$ and $\log_{4/3}(2) - 1 > 0$. The second inequality implies the first and can be shown to be equivalent to $2 > \frac{4}{3}$. This ends the proof that $\mathcal{F}$ is decreasing and strictly concave. □

Next, we prove that $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$ is included in the graph of $\mathcal{F}$.

Lemma B.4. *The graph of $\mathcal{F}$ includes $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$, i.e., for all $(x,y) \in \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$, $\mathcal{F}(x) = y$.*

Proof. First, we observe that $\mathcal{F}(1) = 2$, i.e., $(1,2)$ is in the graph of $\mathcal{F}$. It remains to show that for all $\ell \in \mathbb{N}$, we have

$$\mathcal{F}\left(1 + \frac{3^\ell}{4^{\ell-1}}\right) = 2 - \frac{1}{2^{\ell-1}}.$$

We let $\ell \in \mathbb{N}$. We obtain the following equalities:

$$\begin{aligned}\mathcal{F}\left(1 + \frac{3^\ell}{4^{\ell-1}}\right) &= 2 - \left(\frac{3^{\ell-1}}{4^{\ell-1}}\right)^{\log_{4/3}(2)} \\ &= 2 - \left(\frac{3}{4}\right)^{(\ell-1)\cdot\log_{4/3}(2)} \\ &= 2 - \left(\left(\frac{4}{3}\right)^{\log_{4/3}(2)}\right)^{-(\ell-1)} \\ &= 2 - \frac{1}{2^{\ell-1}}.\end{aligned}$$

We have shown that all elements of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$ are in the graph of $\mathcal{F}$. □

We now prove that all vectors $\mathbf{q} \in \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$ are not smaller than convex combinations of two elements of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\mathbf{q}\}$.

Lemma B.5. *Let* $\mathbf{q} \in \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$*. For all* $\mathbf{p}_1, \mathbf{p}_2 \in \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\mathbf{q}\}$ *and all* $\alpha \in [0,1]$*,* $\mathbf{q}$ *is incomparable to or strictly greater than* $\alpha \cdot \mathbf{p}_1 + (1-\alpha) \cdot \mathbf{p}_2$*.*

Proof. Let $\mathbf{q} = (q_1, q_2)$. Let $\mathbf{p}_1, \mathbf{p}_2 \in \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\mathbf{q}\}$. We fix $\alpha \in]0,1[$; the above statement for $\alpha \in \{0,1\}$ is covered by the cases $\mathbf{p}_1 = \mathbf{p}_2$. We let $\mathbf{p} = \alpha \cdot \mathbf{p}_1 + (1-\alpha) \cdot \mathbf{p}_2$. In the first part of this proof, we assume that $\mathbf{p}_1 \neq (0,2)$ and $\mathbf{p}_2 \neq (0,2)$ and discuss the case when this assumption is lifted at the end of the proof.

By Lemma B.4, we have $q_2 = \mathcal{F}(q_1)$ and we can write $\mathbf{p}_1 = (x_1, \mathcal{F}(x_1))$ and $\mathbf{p}_2 = (x_2, \mathcal{F}(x_2))$ where x_1 and x_2 are the first components of $\mathbf{p}_1$ and $\mathbf{p}_2$ respectively. We assume without loss of generality that $x_1 \leq x_2$. All elements of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$ have different x-components, hence $x_1 \neq q_1$ and $x_2 \neq q_1$.

We first assume that $x_1 \leq x_2 < q_1$. In this case, by Lemma B.3 ($\mathcal{F}$ is strictly decreasing), we have $\mathcal{F}(x_1) \geq \mathcal{F}(x_2) > q_1$. It follows that $\mathbf{q}$ and $\mathbf{p}$ are incomparable: the first component of $\mathbf{q}$ is greater than that of $\mathbf{p}$ but the second component of $\mathbf{q}$ is smaller than that of $\mathbf{p}$. In the case that $q_1 < x_1 \leq x_2$, we obtain in a similar fashion that $\mathbf{q}$ and $\mathbf{p}$ are incomparable.

We now assume that $x_1 < q_1 < x_2$. We let $\beta \in]0,1[$ such that $q_1 = \beta \cdot x_1 + (1-\beta)x_2$, which exists because $q_1 \in]x_1, x_2[$. If $\mathbf{p}$ is not comparable to $\mathbf{q}$, there is nothing to show. We assume that these two vectors are comparable and show that $\mathbf{p} < \mathbf{q}$. We proceed by contradiction and assume that $\mathbf{p} \geq \mathbf{q}$.

We consider the linear form $x^* \colon \mathbb{R}^2 \to \mathbb{R}$ defined by $x^*(\mathbf{v}) = -\frac{\mathcal{F}(x_2)-\mathcal{F}(x_1)}{x_2-x_1} \cdot v_1 + v_2$ for all $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$. We have $x^*(\mathbf{p}_1) = x^*(\mathbf{p}_2)$ and thus $(x^*)^{-1}(x^*(\mathbf{p}_1))$ is the line carrying the segment $[\mathbf{p}_1, \mathbf{p}_2]$. Because $\mathcal{F}$ is decreasing (Lemma B.3), we have $\frac{\mathcal{F}(x_2)-\mathcal{F}(x_1)}{x_2-x_1} < 0$. This implies that x^* is increasing in the sense that for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$, $\mathbf{v}_1 < \mathbf{v}_2$ implies that $x^*(\mathbf{v}_1) < x^*(\mathbf{v}_2)$.

Let $\mathbf{v} = (q_1, \beta \cdot \mathcal{F}(x_1) + (1-\beta)\mathcal{F}(x_2))$. We have $\mathbf{v} < \mathbf{q}$ by strict concavity of $\mathcal{F}$ (Lemma B.3). Furthermore, we have $x^*(\mathbf{v}) = x^*(\mathbf{p})$ because both of these vectors are in the segment $[\mathbf{p}_1, \mathbf{p}_2]$. We obtain, because x^* is increasing and $\mathbf{v} < \mathbf{q} < \mathbf{p}$, that $x^*(\mathbf{v}) < x^*(\mathbf{q}) < x^*(\mathbf{p}) = x^*(\mathbf{v})$. This is a contradiction. This ends the proof in the case that $\mathbf{p}_1 \neq (0,2)$ and $\mathbf{p}_2 \neq (0,2)$.

Next, we assume that $\mathbf{p}_1 = \mathbf{p}_2 = (0,2)$. We obtain that $\mathbf{p} = (0,2)$, which is smaller than $(1,2)$ and incomparable to all elements of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(1,2),(0,2)\}$. Finally, we assume that only one of $\mathbf{p}_1$ or $\mathbf{p}_2$ are $(0,2)$. We assume without loss of generality that $\mathbf{p}_1 = (0,2)$. If $\mathbf{p} > \mathbf{q}$, then we would have $\alpha \cdot (1,2) + (1-\alpha) \cdot \mathbf{p}_2 > \mathbf{p} > \mathbf{q}$, which would contradict the first part of the proof. □

With Lemma B.2 and Lemma B.5, we directly obtain that all elements of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$ are Pareto-optimal elements of $\mathsf{Pay}_{s_0}(\bar{f})$: if they were not Pareto-optimal, then they would be dominated by an element of the boundary of $\mathsf{Pay}_{s_0}(\bar{f})$ which would be the convex combination of two vectors in $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$.

Lemma B.6. *Let* $\mathbf{q} \in \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{(0,2)\}$. *Then* $\mathbf{q}$ *is a Pareto-optimal element of* $\mathsf{Pay}_{s_0}(\bar{f})$.

Proof. Assume towards a contradiction that there exists $\mathbf{p} \in \mathsf{Pay}_{s_0}(\bar{f})$ such that $\mathbf{q} < \mathbf{p}$. We show that there exists $\mathbf{p}' \in \mathsf{bd}(\mathsf{Pay}_{s_0}(\bar{f}))$ such that $\mathbf{p}' > \mathbf{q}$. This yields a contradiction with Lemma B.2 and Lemma B.5.

If $\mathbf{p} \in \mathsf{bd}(\mathsf{Pay}_{s_0}(\bar{f}))$, then we let $\mathbf{p}' = \mathbf{p}$. We thus assume that $\mathbf{p} \in \mathsf{int}(\mathsf{Pay}_{s_0}(\bar{f}))$. Let $\alpha = \sup\{\beta \geq 0 \mid \mathbf{p} + \beta\mathbf{1} \in \mathsf{Pay}_{s_0}(\bar{f})\}$. We have $\alpha \in \mathbb{R}$ because $\mathbf{p} \in \mathsf{Pay}_{s_0}(\bar{f})$ (i.e., $0 \in \{\beta \geq 0 \mid \mathbf{p} + \beta\mathbf{1} \in \mathsf{Pay}_{s_0}(\bar{f})\}$ thus $\alpha > -\infty$) and $\mathsf{Pay}_{s_0}(\bar{f})$ is bounded (thus $\alpha < +\infty$). Furthermore, because $\mathsf{Pay}_{s_0}(\bar{f})$ is closed, we have $\mathbf{p} + \alpha\mathbf{1} \in \mathsf{bd}(\mathsf{Pay}_{s_0}(\bar{f}))$. We obtain the announced contradiction by letting $\mathbf{p}' = \mathbf{p} + \alpha\mathbf{1} > \mathbf{q}$. □

We conclude this section by showing that the set of extreme points of $\mathsf{Pay}_{s_0}(\bar{f})$ is the set of pure payoffs.

Lemma B.7. *We have* $\mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f})) = \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$.

Proof. First, we show that $\mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f})) \subseteq \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$. It suffices to show that all vectors $\mathbf{q} \in \mathsf{Pay}_{s_0}(\bar{f}) \setminus \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$ are not extreme points of $\mathsf{Pay}_{s_0}(\bar{f})$. Let $\mathbf{q} \in \mathsf{Pay}_{s_0}(\bar{f}) \setminus \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$. By Theorem 14.4, we have $\mathbf{q} \in \mathsf{Pay}_{s_0}(\bar{f}) = \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}))$. This implies that $\mathbf{q} \in \mathsf{conv}(\mathsf{Pay}_{s_0}(\bar{f}) \setminus \{\mathbf{q}\})$ because $\mathbf{q} \notin \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$, and thus $\mathbf{q} \notin \mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f}))$.

Conversely, let $\mathbf{q} \in \mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$. First, we assume that $\mathbf{q} = (0, 2)$. The first coordinate of all other vectors of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$ is greater than or equal to 1. This therefore also applies to any convex combination of vectors in $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})\setminus\{(0, 2)\}$. It follows that $\mathbf{q} \in \mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f}))$. Second, we assume that $\mathbf{q} \neq (0, 2)$. Assume towards a contradiction that $\mathbf{q} \notin \mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f}))$. By Lemma B.6, $\mathbf{q}$ is a Pareto-optimal element of $\mathsf{Pay}_{s_0}(\bar{f})$, and thus lies on the boundary of this set. We obtain that $\mathbf{q}$ is the convex combination of two elements of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\mathbf{q}\}$ by Lemma B.2 (which is applicable because $\mathsf{Pay}_{s}(\bar{f})$ is closed) and the assumption that $\mathbf{q} \notin \mathsf{extr}(\mathsf{Pay}_{s_0}(\bar{f}))$. This yields a contradiction with Lemma B.5, which states that there is no convex combination of two vectors of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\mathbf{q}\}$ that is greater than or equal to $\mathbf{q}$. □

Memory cannot be traded for randomness

We prove that the only way to obtain an expected payoff in $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$ is through a strategy that induces a single play from s_0 (i.e., intuitively, a strategy that is pure in practice).

Lemma B.8. *Let $\sigma \in \Sigma(\mathcal{M})$ be a strategy such that at least two plays starting in s_0 are consistent with σ. Then $\mathbb{E}^{\sigma}_{s_0}(\bar{f}) \notin \mathsf{Pay}^{\mathsf{pure}}_{s}(\bar{f})$.*

Proof. We consider the following enumeration of the set of plays of $\mathcal{M}$ that start in s_0. We let $\pi_{-2} = s_0c(s_1a)^{\omega}$, $\pi_{-1} = s_0a(s_2a)^{\omega}$, and, for all $r \in \mathbb{N}$, we let $\pi_r = s_0(as_2)^r b(s_3a)^{\omega}$. We have $\bar{f}(\pi_{-2}) = (0, 2)$, $\bar{f}(\pi_{-1}) = (1, 2)$ and, for all $r \in \mathbb{N}$, $\bar{f}(\pi_r) = \left(1 + \frac{3^r}{4^{r-1}}, 2 - \frac{1}{2^{r-1}}\right)$ (refer to the proof of Lemma B.1 for the relevant computations). Due to the absence of randomised transitions in $\mathcal{M}$, $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$ is the set of payoffs of the plays starting in s_0. Therefore, to end the proof, we must show that for all $r \geq -2$, $\mathbb{E}^{\sigma}_{s_0}(\bar{f}) \neq \bar{f}(\pi_r)$.

For all $r \in \{-1, -2\} \cup \mathbb{N}$, let $\alpha_r = \mathbb{P}^{\sigma}_{s_0}(\{\pi_r\})$. Through this notation, we obtain that

$$\mathbb{E}^{\sigma}_{s_0}(\bar{f}) = \sum_{r \geq -2} \alpha_r \bar{f}(\pi_r). \tag{B.1}$$

First, we show that $\mathbb{E}^{\sigma}_{s_0}(\bar{f}) \notin \{(0,2),(1,2)\}$. If there exists $r \in \mathbb{N}$ such that $\alpha_r > 0$, then we have $\mathbb{E}^{\sigma}_{s_0}(f_2) < 2$, which implies that $\mathbb{E}^{\sigma}_{s_0}(\bar{f}) \notin \{(0,2),(1,2)\}$. Next, assume that $\alpha_r = 0$ for all $r \in \mathbb{N}$. Then α_{-2} and α_{-1} must sum to one. Furthermore, since σ has at least two outcomes, both α_{-2} and α_{-1} must be non-zero. It follows that $\mathbb{E}^{\sigma}_{s_0}(f_1) \in]0,1[$. We conclude that $\mathbb{E}^{\sigma}_{s_0}(\bar{f}) \notin \{(0,2),(1,2)\}$ in this case as well.

We now fix $r \in \mathbb{N}$ and show that $\mathbb{E}^{\sigma}_{s_0}(\bar{f}) \neq \bar{f}(\pi_r)$. We proceed by contradiction. Assume towards a contradiction that $\mathbb{E}^{\sigma}_{s_0}(\bar{f}) = \bar{f}(\pi_r)$. Our goal is to contradict Lemma B.7, i.e., to show that $\bar{f}(\pi_r)$ is not an extreme point of $\mathsf{Pay}_{s_0}(\bar{f})$. We do so in two steps. First, we show that $\mathbb{E}^{\sigma}_{s_0} \in \mathsf{cl}(\mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\bar{f}(\pi_r)\}))$. Next, we prove that $\mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\bar{f}(\pi_r)\})$ is closed. Together, these statements imply that $\bar{f}(\pi_r) \in \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\bar{f}(\pi_r)\})$, which is the sought contradiction.

We now establish that $\mathbb{E}^{\sigma}_{s_0} \in \mathsf{cl}(\mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\bar{f}(\pi_r)\}))$. It follows from $\mathbb{E}^{\sigma}_{s_0}(\bar{f}) = \bar{f}(\pi_r)$ and Equation (B.1) that $\mathbb{E}^{\sigma}_{s_0}(\bar{f}) = \sum_{\ell \neq r} \frac{\alpha_\ell}{1-\alpha_r} \bar{f}(\pi_\ell)$. We obtain from this last equality that $\mathbb{E}^{\sigma}_{s_0}(\bar{f}) \in \mathsf{cl}(\mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\bar{f}(\pi_r)\}))$.

It remains to show that $\mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\bar{f}(\pi_r)\})$ is closed. The vector $\bar{f}(\pi_r)$ is an isolated point of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$: all other elements of $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f})$ are at distance at least 2^{-r} of $\bar{f}(\pi_r)$. Therefore, $\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\bar{f}(\pi_r)\}$ is closed (when removing an isolated point from a closed set, the resulting set is still closed), and thus $\mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_{s_0}(\bar{f}) \setminus \{\bar{f}(\pi_r)\})$ is closed by Lemma 2.2. □

We close this section by commenting on the significance of Lemma B.8 in terms of memory requirements. When we only consider pure strategies, some Pareto-optimal payoffs may require strategies with an arbitrarily large memory, in the sense that we may need to count to some high (but finite) counter value to enact a given number of loops in s_2 to obtain certain expected payoffs. Lemma B.8 implies that we cannot substitute this counting by randomisation, even when only considering Pareto-optimal payoffs. In particular, for this example, we obtain that although all expected payoffs can be obtained with strategies that only count up to a finite number of steps (i.e., to count loops in

s_2), we cannot bound this number of steps uniformly for all expected payoff vectors.

Bibliography

[ABKM09] Eric Allender, Peter Bürgisser, Johan Kjeldgaard-Pedersen, and Peter Bro Miltersen. On the complexity of numerical analysis. *SIAM Journal on Computing*, 38(5):1987–2006, 2009. *6 citations in pages 59, 85, 296, 348, 349 and 372*

[ACK$^+$20] Pranav Ashok, Krishnendu Chatterjee, Jan Kretínský, Maximilian Weininger, and Tobias Winkler. Approximating values of generalized-reachability stochastic games. In Holger Hermanns, Lijun Zhang, Naoki Kobayashi, and Dale Miller, editors, *Proceedings of the 35th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2020, Saarbrücken, Germany, July 8–11, 2020*, pages 102–115. ACM, 2020. *2 citations in pages 221 and 229*

[ACKK24] Muqsit Azeem, Debraj Chakraborty, Sudeep Kanav, and Jan Kretínský. Explainable finite-memory policies for partially observable markov decision processes. *CoRR*, abs/2411.13365, 2024. *Cited in page 16*

[AD94] Rajeev Alur and David L. Dill. A theory of timed automata. *Theoretical Computer Science*, 126(2):183–235, 1994. *Cited in page 14*

[AJK$^+$21] Pranav Ashok, Mathias Jackermeier, Jan Kretínský, Christoph Weinhuber, Maximilian Weininger, and Mayank Yadav. dtControl 2.0: Explainable strategy representation via decision tree learning steered by experts. In Jan Friso Groote and Kim Guldstrand Larsen,

editors, *Proceedings (Part II) of the 27th International Conference on Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2021, Held as Part of ETAPS 2021, Luxemburg City, Luxemburg, March 27–April 1, 2021*, volume 12652 of *Lecture Notes in Computer Science*, pages 326–345. Springer, 2021.
Cited in page 299

[AMNR25] Michal Ajdarów, James C. A. Main, Petr Novotný, and Mickael Randour. Taming infinity one chunk at a time: Concisely represented strategies in one-counter MDPs. In Keren Censor-Hillel, Fabrizio Grandoni, Joël Ouaknine, and Gabriele Puppis, editors, *to appear in Proceedings of the 52th International Colloquium on Automata, Languages, and Programming, ICALP 2025, July 8–11, 2025, Aarhus, Denmark*, volume 334 of *LIPIcs*. Schloss Dagstuhl –Leibniz-Zentrum für Informatik, 2025.
4 citations in pages 12, 14, 82 and 293

[Aum64] Robert J . Aumann. Mixed and behavior strategies in infinite extensive games. In Melvin Dresher, Lloyd S. Shapley, and Albert William Tucker, editors, *Advances in Game Theory. (AM-52), Volume 52*, pages 627–650. Princeton University Press, 1964.
7 citations in pages 7, 37, 55, 68, 164, 167 and 176

[BBC$^+$14a] Tomás Brázdil, Václav Brozek, Krishnendu Chatterjee, Vojtech Forejt, and Antonín Kucera. Markov decision processes with multiple long-run average objectives. *Logical Methods in Computer Science*, 10(1), 2014. *Cited in page 70*

[BBC$^+$14b] Tomáš Brázdil, Václav Brožek, Krishnendu Chatterjee, Vojtěch Forejt, and Antonín Kučera. Markov decision processes with multiple long-run average objectives. *Logical Methods in Computer Science*, Volume 10, Issue 1, Feb 2014.
2 citations in pages 80 and 81

[BBE$^+$10] Tomás Brázdil, Václav Brozek, Kousha Etessami, Antonín Kucera, and Dominik Wojtczak. One-counter Markov decision processes.

In Moses Charikar, editor, *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 863–874. SIAM, 2010. *4 citations in pages 5, 11, 82 and 83*

[BBEK13] Tomás Brázdil, Václav Brozek, Kousha Etessami, and Antonín Kucera. Approximating the termination value of one-counter MDPs and stochastic games. *Information and Computation*, 222:121–138, 2013. *Cited in page 83*

[BBG$^+$20] Thomas Brihaye, Véronique Bruyère, Aline Goeminne, Jean-François Raskin, and Marie van den Bogaard. The complexity of subgame perfect equilibria in quantitative reachability games. *Logical Methods in Computer Science*, 16(4), 2020. *Cited in page 92*

[BBGR21] Thomas Brihaye, Véronique Bruyère, Aline Goeminne, and Jean-François Raskin. Constrained existence problem for weak subgame perfect equilibria with ω-regular Boolean objectives. *Information and Computation*, 278:104594, 2021. *Cited in page 393*

[BBGT21] Thomas Brihaye, Véronique Bruyère, Aline Goeminne, and Nathan Thomasset. On relevant equilibria in reachability games. *Journal of Computer and System Sciences*, 119:211–230, 2021. *7 citations in pages 10, 63, 64, 92, 99, 118 and 120*

[BBMU15] Patricia Bouyer, Romain Brenguier, Nicolas Markey, and Michael Ummels. Pure Nash equilibria in concurrent deterministic games. *Logical Methods in Computer Science*, 11(2), 2015. *Cited in page 92*

[BBN$^+$20] Frantisek Blahoudek, Tomás Brázdil, Petr Novotný, Melkior Ornik, Pranay Thangeda, and Ufuk Topcu. Qualitative controller synthesis for consumption Markov decision processes. In Shuvendu K. Lahiri and Chao Wang, editors, *Proceedings (Part II) of the 32nd International Conference on Computer Aided Verification, CAV 2020, Los Angeles, CA, USA, July 21–24, 2020*, volume 12225 of *Lecture Notes in Computer Science*, pages 421–447. Springer, 2020. *Cited in page 298*

[BCC$^+$15] Tomás Brázdil, Krishnendu Chatterjee, Martin Chmelik, Andreas Fellner, and Jan Kretínský. Counterexample explanation by learning small strategies in Markov decision processes. In Daniel Kroening and Corina S. Pasareanu, editors, *Proceedings (Part I) of the 27th International Conference on Computer Aided Verification, CAV 2015, San Francisco, CA, USA, July 18–24, 2015*, volume 9206 of *Lecture Notes in Computer Science*, pages 158–177. Springer, 2015. *4 citations in pages 8, 16, 299 and 392*

[BCJ18] Roderick Bloem, Krishnendu Chatterjee, and Barbara Jobstmann. Graph games and reactive synthesis. In Edmund M. Clarke, Thomas A. Henzinger, Helmut Veith, and Roderick Bloem, editors, *Handbook of Model Checking*, pages 921–962. Springer, 2018. *3 citations in pages 3, 15 and 90*

[BCKN12] Tomás Brázdil, Krishnendu Chatterjee, Antonín Kucera, and Petr Novotný. Efficient controller synthesis for consumption games with multiple resource types. In P. Madhusudan and Sanjit A. Seshia, editors, *Computer Aided Verification - 24th International Conference, CAV 2012, Berkeley, CA, USA, July 7-13, 2012 Proceedings*, volume 7358 of *Lecture Notes in Computer Science*, pages 23–38. Springer, 2012. *Cited in page 298*

[BCKT18] Tomás Brázdil, Krishnendu Chatterjee, Jan Kretínský, and Viktor Toman. Strategy representation by decision trees in reactive synthesis. In Dirk Beyer and Marieke Huisman, editors, *Proceedings (Part I) of the 24th International Conference on Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2018, Held as Part of ETAPS 2018, Thessaloniki, Greece, April 14–20, 2018*, volume 10805 of *Lecture Notes in Computer Science*, pages 385–407. Springer, 2018. *4 citations in pages 8, 16, 299 and 392*

[BCM$^+$23] Damien Busatto-Gaston, Debraj Chakraborty, Anirban Majumdar, Sayan Mukherjee, Guillermo A. Pérez, and Jean-François Raskin. Bi-objective lexicographic optimization in Markov decision processes with related objectives. In Étienne André and

Jun Sun, editors, *Proceedings (Part I) of the 21st Interval Symposium on Automated Technology for Verification and Analysis, ATVA 2023, Singapore, October 24–27, 2023*, volume 14215 of *Lecture Notes in Computer Science*, pages 203–223. Springer, 2023. *Cited in page 76*

[BDOR20] Thomas Brihaye, Florent Delgrange, Youssouf Oualhadj, and Mickael Randour. Life is random, time is not: Markov decision processes with window objectives. *Logical Methods in Computer Science*, 16(4), 2020. *2 citations in pages 14 and 80*

[BDS13] Thomas Brihaye, Julie De Pril, and Sven Schewe. Multiplayer cost games with simple Nash equilibria. In Sergei N. Artëmov and Anil Nerode, editors, *Proceedings of the International Symposium on Logical Foundations of Computer Science, LFCS 2013, San Diego, CA, USA, January 6–8, 2013*, volume 7734 of *Lecture Notes in Computer Science*, pages 59–73. Springer, 2013. *3 citations in pages 63, 64 and 121*

[BFHT85] Allan Borodin, Ronald Fagin, John E. Hopcroft, and Martin Tompa. Decreasing the nesting depth of expressions involving square roots. *Journal of Symbolic Computation*, 1(2):169–188, 1985. *2 citations in pages 372 and 375*

[BFRR17] Véronique Bruyère, Emmanuel Filiot, Mickael Randour, and Jean-François Raskin. Meet your expectations with guarantees: Beyond worst-case synthesis in quantitative games. *Information and Computation*, 254:259–295, 2017. *2 citations in pages 70 and 229*

[BGG17] Nathalie Bertrand, Blaise Genest, and Hugo Gimbert. Qualitative determinacy and decidability of stochastic games with signals. *Journal of the ACM*, 64(5):33:1–33:48, 2017. *3 citations in pages 4, 6 and 68*

[BGHM17] Thomas Brihaye, Gilles Geeraerts, Axel Haddad, and Benjamin Monmege. Pseudopolynomial iterative algorithm to solve total-payoff games and min-cost reachability games. *Acta Informatica*, 54(1):85–125, 2017. *3 citations in pages 15, 62 and 107*

[BGMR23] Thomas Brihaye, Aline Goeminne, James C. A. Main, and Mickael Randour. Reachability games and friends: A journey through the lens of memory and complexity (invited talk). In Patricia Bouyer and Srikanth Srinivasan, editors, *Proceedings of the 43rd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2023, IIIT Hyderabad, Telangana, India, December 18–20, 2023*, volume 284 of *LIPIcs*, pages 1:1–1:26. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023. *6 citations in pages 5, 15, 58, 63, 76 and 95*

[BHR16] Véronique Bruyère, Quentin Hautem, and Mickael Randour. Window parity games: an alternative approach toward parity games with time bounds. In Domenico Cantone and Giorgio Delzanno, editors, *Proceedings of the 7th International Symposium on Games, Automata, Logics, and Formal Verification, GandALF 2016, Catania, Italy, September 14–16, 2016*, volume 226 of *EPTCS*, pages 135–148, 2016. *Cited in page 14*

[BHRR19] Véronique Bruyère, Quentin Hautem, Mickael Randour, and Jean-François Raskin. Energy mean-payoff games. In Wan J. Fokkink and Rob van Glabbeek, editors, *Proceedings of the 30th International Conference on Concurrency Theory, CONCUR 2019, Amsterdam, the Netherlands, August 27–30, 2019*, volume 140 of *LIPIcs*, pages 21:1–21:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. *Cited in page 80*

[BK08] Christel Baier and Joost-Pieter Katoen. *Principles of model checking*. MIT Press, 2008. *10 citations in pages 2, 4, 15, 49, 285, 288, 294, 306, 313 and 406*

[BKK11] Tomás Brázdil, Stefan Kiefer, and Antonín Kucera. Efficient analysis of probabilistic programs with an unbounded counter. In Ganesh Gopalakrishnan and Shaz Qadeer, editors, *Computer Aided Verification - 23rd International Conference, CAV 2011, Snowbird, UT, USA, July 14-20, 2011. Proceedings*, volume 6806 of *Lecture Notes in Computer Science*, pages 208–224. Springer, 2011. *Cited in page 11*

[BKN$^+$19] Nikhil Balaji, Stefan Kiefer, Petr Novotný, Guillermo A. Pérez, and Mahsa Shirmohammadi. On the complexity of value iteration. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, Patras, Greece, July 9–12, 2019*, volume 132 of *LIPIcs*, pages 102:1–102:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. *2 citations in pages 12 and 83*

[BKSV08] Noam Berger, Nevin Kapur, Leonard J. Schulman, and Vijay V. Vazirani. Solvency games. In Ramesh Hariharan, Madhavan Mukund, and V. Vinay, editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2008, December 9-11, 2008, Bangalore, India*, volume 2 of *LIPIcs*, pages 61–72. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2008. *2 citations in pages 84 and 298*

[BL69] J Richard Büchi and Lawrence H Landweber. Solving sequential conditions by finite-state strategies. *Transactions of the American Mathematical Society*, 138:295–311, 1969. *Cited in page 2*

[BLO$^+$22] Patricia Bouyer, Stéphane Le Roux, Youssouf Oualhadj, Mickael Randour, and Pierre Vandenhove. Games where you can play optimally with arena-independent finite memory. *Logical Methods in Computer Science*, 18(1), 2022. *3 citations in pages 15, 16 and 394*

[BMR14] Véronique Bruyère, Noémie Meunier, and Jean-François Raskin. Secure equilibria in weighted games. In Thomas A. Henzinger and Dale Miller, editors, *Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), CSL-LICS'14, Vienna, Austria, July 14–18, 2014*, pages 26:1–26:26. ACM, 2014. *Cited in page 92*

[Bor21] Emile Borel. La théorie du jeu et les équations intégralesa noyau symétrique. *Comptes rendus de l'Académie des Sciences*, 173(1304-1308):58, 1921. *Cited in page 4*

[BORV23] Patricia Bouyer, Youssouf Oualhadj, Mickael Randour, and Pierre Vandenhove. Arena-independent finite-memory determinacy in stochastic games. *Log. Methods Comput. Sci.*, 19(4), 2023. *4 citations in pages 16, 299, 313 and 394*

[BPR06] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in Real Algebraic Geometry.* 1431-1550. Springer, 2nd edition, 2006. *4 citations in pages 59, 367, 368 and 370*

[BRV22] Patricia Bouyer, Mickael Randour, and Pierre Vandenhove. The true colors of memory: A tour of chromatic-memory strategies in zero-sum games on graphs (invited talk). In Anuj Dawar and Venkatesan Guruswami, editors, *Proceedings of the 42nd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2022, IIT Madras, Chennai, India, December 18–20, 2022*, volume 250 of *LIPIcs*, pages 3:1–3:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. *Cited in page 299*

[BRV23] Patricia Bouyer, Mickael Randour, and Pierre Vandenhove. Characterizing omega-regularity through finite-memory determinacy of games on infinite graphs. *TheoretiCS*, 2, 2023. *2 citations in pages 15 and 16*

[BRvdB22] Léonard Brice, Jean-François Raskin, and Marie van den Bogaard. The complexity of spes in mean-payoff games. In Mikolaj Bojanczyk, Emanuela Merelli, and David P. Woodruff, editors, *Proceedings of the 49th International Colloquium on Automata, Languages, and Programming, ICALP 2022, Paris, France, July 4–8, 2022*, volume 229 of *LIPIcs*, pages 116:1–116:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. *Cited in page 92*

[BSS89] Lenore Blum, Mike Shub, and Steve Smale. Over the real numbers: NP-completeness, recursive functions and universal machines. *Bulletin of the American Mathematical Society*, 21(1), 1989. *2 citations in pages 59 and 85*

[Büc62] J. Richard Büchi. On a decision method in restricted second order arithmetic. *Proceedings of the International Congress on Logic, Methodology and Philosophy of Science*, pages 1–11, 1962. *Cited in page 2*

[BvdBR23] Léonard Brice, Marie van den Bogaard, and Jean-François Raskin. Subgame-perfect equilibria in mean-payoff games (journal version). *Logical Methods in Computer Science*, 19(4), 2023. *Cited in page 92*

[Can88] John F. Canny. Some algebraic and geometric computations in PSPACE. In Janos Simon, editor, *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, STOC 1988, Chicago, Illinois, USA, May 2–4, 1988*, pages 460–467. ACM, 1988. *Cited in page 349*

[CD12a] Krishnendu Chatterjee and Laurent Doyen. Energy parity games. *Theoretical Computer Science*, 458:49–60, 2012. *3 citations in pages 15, 62 and 391*

[CD12b] Krishnendu Chatterjee and Laurent Doyen. Partial-observation stochastic games: How to win when belief fails. In *Proceedings of the 27th Annual IEEE Symposium on Logic in Computer Science, LICS 2012, Dubrovnik, Croatia, June 25–28, 2012*, pages 175–184. IEEE Computer Society, 2012. *3 citations in pages 4, 6 and 68*

[CDGH15] Krishnendu Chatterjee, Laurent Doyen, Hugo Gimbert, and Thomas A. Henzinger. Randomness for free. *Information and Computation*, 245:3–16, 2015. *2 citations in pages 167 and 229*

[CdH04] Krishnendu Chatterjee, Luca de Alfaro, and Thomas A. Henzinger. Trading memory for randomness. In *Proceedings of the 1st International Conference on Quantitative Evaluation of Systems, QEST 2004, Enschede, The Netherlands, 27–30 September 2004*, pages

206–217. IEEE Computer Society, 2004.
5 citations in pages 62, 75, 97, 167 and 393

[CDH10] Julien Cristau, Claire David, and Florian Horn. How do we remember the past in randomised strategies? In Angelo Montanari, Margherita Napoli, and Mimmo Parente, editors, *Proceedings of the First Symposium on Games, Automata, Logic, and Formal Verification, GANDALF 2010, Minori (Amalfi Coast), Italy, June 17–18, 2010*, volume 25 of *EPTCS*, pages 30–39, 2010.
8 citations in pages 7, 68, 69, 70, 167, 208, 209 and 210

[CE81] Edmund M. Clarke and E. Allen Emerson. Design and synthesis of synchronization skeletons using branching-time temporal logic. In Dexter Kozen, editor, *Logics of Programs, Workshop, Yorktown Heights, New York, USA, May 1981*, volume 131 of *Lecture Notes in Computer Science*, pages 52–71. Springer, 1981. *Cited in page 2*

[CFGR16] Rodica Condurache, Emmanuel Filiot, Raffaella Gentilini, and Jean-François Raskin. The complexity of rational synthesis. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, *Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, Rome, Italy, July 11–15, 2016*, volume 55 of *LIPIcs*, pages 121:1–121:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016. *Cited in page 117*

[CFK$^+$13a] Taolue Chen, Vojtech Forejt, Marta Z. Kwiatkowska, Aistis Simaitis, and Clemens Wiltsche. On stochastic games with multiple objectives. In Krishnendu Chatterjee and Jirí Sgall, editors, *Proceedings of the 38th International Symposium on Mathematical Foundations of Computer Science, MFCS 2013, Klosterneuburg, Austria, August 26–30, 2013*, volume 8087 of *Lecture Notes in Computer Science*, pages 266–277. Springer, 2013.
4 citations in pages 219, 220, 221 and 229

[CFK$^+$13b] Taolue Chen, Vojtěch Forejt, Marta Kwiatkowska, Aistis Simaitis, and Clemens Wiltsche. On stochastic games with multiple objectives. Technical Report RR-13-06, University of Oxford, Department of Computer Science, 2013. *Cited in page 220*

[CFW13] Krishnendu Chatterjee, Vojtech Forejt, and Dominik Wojtczak. Multi-objective discounted reward verification in graphs and MDPs. In Kenneth L. McMillan, Aart Middeldorp, and Andrei Voronkov, editors, *Proceedings of the 19th International Conference on Logic for Programming, Artificial Intelligence, and Reasoning, LPAR-19, Stellenbosch, South Africa, December 14–19, 2013*, volume 8312 of *Lecture Notes in Computer Science*, pages 228–242. Springer, 2013. *3 citations in pages 76, 80 and 242*

[Cha07] Krishnendu Chatterjee. Optimal strategy synthesis in stochastic müller games. In Helmut Seidl, editor, *Proceedings of the 10th International Conference on Foundations of Software Science and Computational Structures, FoSSaCS 2007, Held as Part of ETAPS 2007, Braga, Portugal, March 24 – April 1, 2007*, volume 4423 of *Lecture Notes in Computer Science*, pages 138–152. Springer, 2007. *4 citations in pages 7, 62, 70 and 167*

[CHP08] Krishnendu Chatterjee, Thomas A. Henzinger, and Vinayak S. Prabhu. Trading infinite memory for uniform randomness in timed games. In Magnus Egerstedt and Bud Mishra, editors, *Proceedings of the 11th International Workshop on Hybrid Systems: Computation and Control, HSCC 2008, St. Louis, MO, USA, April 22–24, 2008*, volume 4981 of *Lecture Notes in Computer Science*, pages 87–100. Springer, 2008. *2 citations in pages 97 and 167*

[Chu57] Alonzo Church. Application of recursive arithmetic to the problem of circuit synthesis. *Summaries of the Summer Institute of Symbolic Logic*, I:3–50, 1957. *Cited in page 2*

[CHVB18] Edmund M. Clarke, Thomas A. Henzinger, Helmut Veith, and Roderick Bloem, editors. *Handbook of Model Checking*. Springer, 2018. *4 citations in pages 2, 10, 63 and 99*

[CJT20] Steven Carr, Nils Jansen, and Ufuk Topcu. Verifiable rnn-based policies for POMDPs under temporal logic constraints. In Christian Bessiere, editor, *Proceedings of the 29th International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 4121–4127. ijcai.org, 2020. *Cited in page 16*

[CJW+19] Steven Carr, Nils Jansen, Ralf Wimmer, Alexandru Constantin Serban, Bernd Becker, and Ufuk Topcu. Counterexample-guided strategy improvement for POMDPs using recurrent neural networks. In Sarit Kraus, editor, *Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10–16, 2019*, pages 5532–5539. ijcai.org, 2019. *Cited in page 16*

[CKK17] Krishnendu Chatterjee, Zuzana Kretínská, and Jan Kretínský. Unifying two views on multiple mean-payoff objectives in Markov decision processes. *Logical Methods in Computer Science*, 13(2), 2017. *3 citations in pages 70, 76 and 80*

[CKM+23] Krishnendu Chatterjee, Joost-Pieter Katoen, Stefanie Mohr, Maximilian Weininger, and Tobias Winkler. Stochastic games with lexicographic objectives. *Formal methods in system design*, pages 1–41, 2023. *Cited in page 76*

[CN06] Thomas Colcombet and Damian Niwiński. On the positional determinacy of edge-labeled games. *Theoretical Computer Science*, 352(1-3):190–196, 2006. *Cited in page 16*

[Con92] Anne Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203–224, 1992. *2 citations in pages 4 and 6*

[CRR14] Krishnendu Chatterjee, Mickael Randour, and Jean-François Raskin. Strategy synthesis for multi-dimensional quantitative objectives. *Acta Informatica*, 51(3-4):129–163, 2014. *13 citations in pages 10, 15, 62, 63, 75, 97, 99, 167, 229, 299, 314, 392 and 393*

[dAFH+03] Luca de Alfaro, Marco Faella, Thomas A. Henzinger, Rupak Majumdar, and Mariëlle Stoelinga. The element of surprise in timed games. In Roberto M. Amadio and Denis Lugiez, editors, *Proceedings of the 14th International Conference on Concurrency Theory, CONCUR 2003, Marseille, France, September 3–5, 2003*, volume 2761 of *Lecture Notes in Computer Science*, pages 142–156. Springer, 2003. *Cited in page 14*

[dAH00] Luca de Alfaro and Thomas A. Henzinger. Concurrent omega-regular games. In *Proceedings of the 15th Annual IEEE Symposium on Logic in Computer Science, LICS 2000, Santa Barbara, California, USA, June 26–29, 2000*, pages 141–154. IEEE Computer Society, 2000. *Cited in page 4*

[dAHK07] Luca de Alfaro, Thomas A. Henzinger, and Orna Kupferman. Concurrent reachability games. *Theoretical Computer Science*, 386(3):188–217, 2007. *10 citations in pages 4, 6, 7, 48, 49, 68, 74, 207, 209 and 211*

[DJ23] Catalin Dima and Wojciech Jamroga. Computationally feasible strategies. In Noa Agmon, Bo An, Alessandro Ricci, and William Yeoh, editors, *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2023, London, United Kingdom, May 29 – June 2, 2023*, pages 784–792. ACM, 2023. *Cited in page 16*

[DKQR20] Florent Delgrange, Joost-Pieter Katoen, Tim Quatmann, and Mickael Randour. Simple strategies in multi-objective MDPs. In Armin Biere and David Parker, editors, *Proceedings (Part I) of the 26th International Conference on Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2020, Held as Part of ETAPS 2020, Dublin, Ireland, April 25–30, 2020*, volume 12078 of *Lecture Notes in Computer Science*, pages 346–364. Springer, 2020. *5 citations in pages 6, 68, 76, 84 and 299*

[Dur19] Rick Durrett. *Probability: Theory and Examples.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 5th edition, 2019. *2 citations in pages 81 and 281*

[EJ88] E. Allen Emerson and Charanjit S. Jutla. The complexity of tree automata and logics of programs (extended abstract). In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science, FOCS 1988, White Plains, New York, USA, October 24–26, 1988*, pages 328–337. IEEE Computer Society, 1988. *3 citations in pages 15, 66 and 107*

[EKVY08] Kousha Etessami, Marta Z. Kwiatkowska, Moshe Y. Vardi, and Mihalis Yannakakis. Multi-objective model checking of Markov decision processes. *Logical Methods in Computer Science*, 4(4), 2008. *8 citations in pages 6, 68, 74, 76, 79, 80, 81 and 204*

[EM79] Andrzej Ehrenfeucht and Jan Mycielski. Positional strategies for mean payoff games. *International Journal of Game Theory*, 8(2):109–113, 1979. *Cited in page 15*

[EWY10] Kousha Etessami, Dominik Wojtczak, and Mihalis Yannakakis. Quasi-birth-death processes, tree-like QBDs, probabilistic 1-counter automata, and pushdown systems. *Performance Evaluation*, 67(9):837–857, 2010. *5 citations in pages 12, 86, 297, 371 and 373*

[FBB$^+$23] Nathanaël Fijalkow, Nathalie Bertrand, Patricia Bouyer-Decitre, Romain Brenguier, Arnaud Carayol, John Fearnley, Hugo Gimbert, Florian Horn, Rasmus Ibsen-Jensen, Nicolas Markey, Benjamin Monmege, Petr Novotný, Mickael Randour, Ocan Sankur, Sylvain Schmitz, Olivier Serre, and Mateusz Skomra. Games on graphs. *CoRR*, abs/2305.10546, 2023. *4 citations in pages 3, 4, 15 and 207*

[FH13] Nathanaël Fijalkow and Florian Horn. Les jeux d'accessibilité généralisée. *Technique et Science Informatiques*, 32(9-10):931–949, 2013. *4 citations in pages 6, 15, 62 and 229*

[FKP12] Vojtech Forejt, Marta Z. Kwiatkowska, and David Parker. Pareto curves for probabilistic model checking. In Supratik Chakraborty and Madhavan Mukund, editors, *Proceedings of the 10th International Symposium on Automated Technology for Verification and Analysis, ATVA 2014, Thiruvananthapuram, India, October 3–6,* *2012*, volume 7561 of *Lecture Notes in Computer Science*, pages 317–332. Springer, 2012. *4 citations in pages 76, 80, 81 and 229*

[FL83] Drew Fudenberg and David Levine. Subgame-perfect equilibria of finite-and infinite-horizon games. *Journal of Economic Theory*, 31(2):251–268, 1983. *Cited in page 63*

[Fri71] James Friedman. A non-cooperative equilibrium for supergames. *Review of Economic Studies*, 38(1):1–12, 1971. *Cited in page 64*

[Gel14] Marcus Gelderie. *Strategy machines: representation and complexity of strategies in infinite games.* PhD thesis, RWTH Aachen University, 2014. *2 citations in pages 8 and 16*

[Gim07] Hugo Gimbert. Pure stationary optimal strategies in Markov decision processes. In Wolfgang Thomas and Pascal Weil, editors, *Proceedings of the 24th Annual Symposium on Theoretical Aspects of Computer Science, STACS 2007, Aachen, Germany, February 22–24, 2007*, volume 4393, pages 200–211. Springer, 2007. *3 citations in pages 16, 299 and 394*

[GJ79] M. R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman, 1979. *Cited in page 384*

[GK23] Hugo Gimbert and Edon Kelmendi. Submixing and shift-invariant stochastic games. *International Journal of Game Theory*, 52(4):1179–1214, 2023. *Cited in page 16*

[GO10] Hugo Gimbert and Youssouf Oualhadj. Probabilistic automata on finite words: Decidable and undecidable problems. In Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide, and Paul G. Spirakis, editors, *Proceedings (Part II) of the 37th International Colloquium on Automata, Languages and Programming, ICALP 2010, Bordeaux, France, July 6–10, 2010*, volume 6199 of *Lecture Notes in Computer Science*, pages 527–538. Springer, 2010. *Cited in page 7*

[GS53] David Gale and Frank M Stewart. Infinite games with perfect information. *Contributions to the Theory of Games*, 2(245-266):2–16, 1953. *2 citations in pages 107 and 108*

[GTW02] Erich Grädel, Wolfgang Thomas, and Thomas Wilke, editors. *Automata, Logics, and Infinite Games: A Guide to Current Research [outcome of a Dagstuhl seminar, February 2001]*, volume 2500 of *Lecture Notes in Computer Science*. Springer, 2002. *3 citations in pages 3, 6 and 90*

[GZ05] Hugo Gimbert and Wiesław Zielonka. Games where you can play optimally without any memory. In Martín Abadi and Luca de Alfaro, editors, *Proceedings of the 16th International Conference on Concurrency Theory, CONCUR 2005, San Francisco, CA, USA, August 23–26, 2005*, volume 3653 of *Lecture Notes in Computer Science*, pages 428–442. Springer, 2005. *3 citations in pages 15, 16 and 394*

[HM18] Serge Haddad and Benjamin Monmege. Interval iteration algorithm for MDPs and IMDPs. *Theoretical Computer Science*, 735:111–131, 2018. *2 citations in pages 287 and 289*

[Hor09] Florian Horn. Random fruits on the Zielonka tree. In Susanne Albers and Jean-Yves Marion, editors, *Proceedings of the 26th International Symposium on Theoretical Aspects of Computer Science, STACS 2009, Freiburg, Germany, February 26–28, 2009*, volume 3 of *LIPIcs*, pages 541–552. Schloss Dagstuhl –Leibniz-Zentrum für Informatik, Germany, 2009. *6 citations in pages 7, 62, 75, 97, 167 and 393*

[Hou70] Alston Scott Householder. *The Numerical Treatment of a Single Nonlinear Equation.* McGraw-Hill, 1970. *Cited in page 380*

[HP85] David Harel and Amir Pnueli. On the development of reactive systems. In Krzysztof R. Apt, editor, *Logics and Models of Concurrent Systems - Conference proceedings, Colle-sur-Loup (near Nice), France, 8-19 October 1984*, volume 13 of *NATO ASI Series*, pages 477–498. Springer, 1985. *Cited in page 1*

[HPS$^+$21] Ernst Moritz Hahn, Mateo Perez, Sven Schewe, Fabio Somenzi, Ashutosh Trivedi, and Dominik Wojtczak. Model-free reinforcement learning for lexicographic omega-regular objectives. In Marieke Huisman, Corina S. Pasareanu, and Naijun Zhan, editors, *Proceedings of the 24th International Symposium on Formal Methods, FM 2021, Virtual Event, November 20–26, 2021*, volume 13047 of *Lecture Notes in Computer Science*, pages 142–159. Springer, 2021. *Cited in page 76*

[JKW23] Florian Jüngermann, Jan Kretínský, and Maximilian Weininger. Algebraically explainable controllers: decision trees and support vector machines join forces. *International Journal on Software Tools for Technology Transfer*, 25(3):249–266, 2023. *4 citations in pages 8, 16, 299 and 392*

[JLS15] Marcin Jurdzinski, Ranko Lazic, and Sylvain Schmitz. Fixed-dimensional energy games are in pseudo-polynomial time. In Magnús M. Halldórsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, editors, *Proceedings (Part II) of the 42nd International Colloquium on Automata, Languages, and Programming, ICALP 2015, Kyoto, Japan, July 6–10, 2015*, volume 9135 of *Lecture Notes in Computer Science*, pages 260–272. Springer, 2015. *Cited in page 392*

[Kal21] Olav Kallenberg. *Foundations of Modern Probability.* Springer International Publishing, Cham, 2021. *Cited in page 34*

[Kec95] A. Kechris. *Classical Descriptive Set Theory.* Graduate Texts in Mathematics. Springer New York, 1995. *2 citations in pages 410 and 411*

[KEM06] Antonín Kucera, Javier Esparza, and Richard Mayr. Model checking probabilistic pushdown automata. *Logical Methods in Computer Science*, 2(1), 2006. *4 citations in pages 85, 317, 326 and 330*

[KFG19] Anurag Koul, Alan Fern, and Sam Greydanus. Learning finite state representations of recurrent policy networks. In *Proceedings of the*

7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6–9, 2019. OpenReview.net, 2019. *Cited in page 16*

[KMS$^+$20] Stefan Kiefer, Richard Mayr, Mahsa Shirmohammadi, Patrick Totzke, and Dominik Wojtczak. How to play in infinite MDPs (invited talk). In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *Proceedings of the 47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, Saarbrücken, Germany, July 8–11, 2020*, volume 168 of *LIPIcs*, pages 3:1–3:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. *3 citations in pages 6, 307 and 308*

[KMST24] Stefan Kiefer, Richard Mayr, Mahsa Shirmohammadi, and Patrick Totzke. Strategy complexity of reachability in countable stochastic 2-player games. *Dynamic Games and Applications*, 2024. *Cited in page 6*

[Kön27] Dénes König. Über eine Schlussweise aus dem Endlichen ins Unendliche. *Acta Sci. Math.(Szeged)*, 3(2-3):121–130, 1927. *Cited in page 115*

[Koz24] Alexander Kozachinskiy. Infinite separation between general and chromatic memory. In José A. Soto and Andreas Wiese, editors, *LATIN 2024: Theoretical Informatics - 16th Latin American Symposium, Puerto Varas, Chile, March 18-22, 2024, Proceedings, Part II*, volume 14579 of *Lecture Notes in Computer Science*, pages 114–128. Springer, 2024. *Cited in page 16*

[Kuh53] Harold W Kuhn. Extensive games and the problem of information. *Contributions to the Theory of Games*, 2(28):193–216, 1953. *3 citations in pages 7, 55 and 176*

[Lan02] Serge Lang. *Algebra*, volume 211 of *Graduate Texts in Mathematics*. Springer, revised third edition edition, 2002. *Cited in page 377*

[LP18] Stéphane Le Roux and Arno Pauly. Extending finite-memory determinacy to multi-player games. *Information and Computation*, 261:676–694, 2018. *Cited in page 16*

[Mai24] James C. A. Main. Arena-independent memory bounds for Nash equilibria in reachability games. In Olaf Beyersdorff, Mamadou Moustapha Kanté, Orna Kupferman, and Daniel Lokshtanov, editors, *41st International Symposium on Theoretical Aspects of Computer Science, STACS 2024, March 12-14, 2024, Clermont-Ferrand, France*, volume 289 of *LIPIcs*, pages 50:1–50:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024. *5 citations in pages 10, 14, 62, 89 and 99*

[Maz01] René Mazala. Infinite games. In Erich Grädel, Wolfgang Thomas, and Thomas Wilke, editors, *Automata, Logics, and Infinite Games: A Guide to Current Research [outcome of a Dagstuhl seminar, February 2001]*, volume 2500 of *Lecture Notes in Computer Science*, pages 23–42. Springer, 2001. *2 citations in pages 66 and 106*

[Mea55] George H. Mealy. A method for synthesizing sequential circuits. *The Bell System Technical Journal*, 34(5):1045–1079, 1955. *Cited in page 5*

[MPR20] Benjamin Monmege, Julie Parreaux, and Pierre-Alain Reynier. Reaching your goal optimally by playing at random with no memory. In Igor Konnov and Laura Kovács, editors, *Proceedings of the 31st International Conference on Concurrency Theory, CONCUR 2020, Vienna, Austria, September 1–4, 2020*, volume 171 of *LIPIcs*, pages 26:1–26:21. Schloss Dagstuhl –Leibniz-Zentrum für Informatik, 2020. *4 citations in pages 75, 97, 167 and 393*

[MR22] James C. A. Main and Mickael Randour. Different strokes in randomised strategies: Revisiting Kuhn's theorem under finite-memory assumptions. In Bartek Klin, Slawomir Lasota, and Anca Muscholl, editors, *Proceedings of the 33rd International Conference on Concurrency Theory, CONCUR 2022, Warsaw,* *Poland, September 12–16, 2022*, volume 243 of *LIPIcs*, pages 22:1–22:18. Schloss Dagstuhl –Leibniz-Zentrum für Informatik, 2022. *2 citations in pages 14 and 163*

[MR24] James C. A. Main and Mickael Randour. Different strokes in randomised strategies: Revisiting Kuhn's theorem under finite-memory assumptions. *Information and Computation*, 301:105229, 2024. *4 citations in pages 10, 14, 68 and 163*

[MR25] James C. A. Main and Mickael Randour. Mixing any cocktail with limited ingredients: On the structure of payoff sets in multi-objective MDPs and its impact on randomised strategies. *CoRR*, abs/2502.18296, 2025. *4 citations in pages 11, 14, 75 and 227*

[MRS21] James C. A. Main, Mickael Randour, and Jeremy Sproston. Time flies when looking out of the window: Timed games with window parity objectives. In Serge Haddad and Daniele Varacca, editors, *Proceedings of the 32nd International Conference on Concurrency Theory, CONCUR 2021, Virtual Conference, August 24–27, 2021*, volume 203 of *LIPIcs*, pages 25:1–25:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. *Cited in page 14*

[MRS22] James C. A. Main, Mickael Randour, and Jeremy Sproston. Timed games with bounded window parity objectives. In Sergiy Bogomolov and David Parker, editors, *Proceedings of the 20th International Conference on Formal Modeling and Analysis of Timed Systems, FORMATS 2022, Warsaw, Poland, September 13–15, 2022*, volume 13465 of *Lecture Notes in Computer Science*, pages 165–182. Springer, 2022. *Cited in page 14*

[Mun97] James R. Munkres. *Topology: A First Course.* Prentice Hall International, 1997. *3 citations in pages 400, 405 and 409*

[Nas50] John F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950. *2 citations in pages 3 and 63*

[OR94] Martin J. Osborne and Ariel Rubinstein. *A course in game theory.* The MIT Press, 1994.
8 citations in pages 2, 4, 7, 53, 64, 68, 164 and 175

[Orn69] Donald Ornstein. On the existence of stationary optimal strategies. *Proceedings of the American Mathematical Society*, 20(2):563–569, 1969. *3 citations in pages 6, 307 and 308*

[OW14] Joël Ouaknine and James Worrell. Positivity problems for low-order linear recurrence sequences. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5–7, 2014*, pages 366–379. SIAM, 2014.
2 citations in pages 12 and 83

[Pap94] Christos H. Papadimitriou. *Computational complexity.* Addison-Wesley, 1994. *Cited in page 58*

[PB24] Jakob Piribauer and Christel Baier. Positivity-hardness results on Markov decision processes. *TheoretiCS*, 3, 2024.
3 citations in pages 12, 83 and 86

[Pnu77] Amir Pnueli. The temporal logic of programs. In *Proceedings of the 18th Annual Symposium on Foundations of Computer Science, FOCS 1977, Providence, Rhode Island, USA, October 31 – November 1, 1977*, pages 46–57. IEEE Computer Society, 1977.
Cited in page 1

[QK21] Tim Quatmann and Joost-Pieter Katoen. Multi-objective optimization of long-run average and total rewards. In Jan Friso Groote and Kim Guldstrand Larsen, editors, *Proceedings (Part I) of the 27th International Conference on Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2021, Held as Part of ETAPS 2021, Luxemburg City, Luxemburg, March 27–April 1, 2021*, volume 12651 of *Lecture Notes in Computer Science*, pages 230–249. Springer, 2021. *2 citations in pages 76 and 229*

[QS82] Jean-Pierre Queille and Joseph Sifakis. Specification and verification of concurrent systems in CESAR. In Mariangiola Dezani-Ciancaglini and Ugo Montanari, editors, *International Symposium on Programming, 5th Colloquium, Torino, Italy, April 6-8, 1982, Proceedings*, volume 137 of *Lecture Notes in Computer Science*, pages 337–351. Springer, 1982. *Cited in page 2*

[Qua23] Tim Quatmann. *Verification of multi-objective Markov models.* PhD thesis, RWTH Aachen University, Germany, 2023. *2 citations in pages 229 and 237*

[Rab69] Michael O. Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, 1969. *Cited in page 2*

[RCDH07] Jean-François Raskin, Krishnendu Chatterjee, Laurent Doyen, and Thomas A. Henzinger. Algorithms for omega-regular games with imperfect information. *Logical Methods in Computer Science*, 3(3), 2007. *Cited in page 392*

[Roc70] R. Tyrrell Rockafellar. *Convex Analysis.* Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1970. *5 citations in pages 25, 26, 27, 248 and 255*

[RRS17] Mickael Randour, Jean-François Raskin, and Ocan Sankur. Percentile queries in multi-dimensional Markov decision processes. *Formal methods in system design*, 50(2-3):207–248, 2017. *8 citations in pages 6, 15, 62, 68, 76, 80, 81 and 204*

[RW89] Peter J. Ramadge and Walter Murray Wonham. The control of discrete event systems. *Proceedings of the IEEE*, 77(1):81–98, 1989. *Cited in page 2*

[Sav70] Walter J. Savitch. Relationships between nondeterministic and deterministic tape complexities. *Journal of Computer and System Sciences*, 4(2):177–192, 1970. *2 citations in pages 367 and 370*

[SB18] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction.* MIT Press, 2018. *3 citations in pages 4, 16 and 392*

[Sel65] Reinhard Selten. Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit: Teil I: Bestimmung des dynamischen Preisgleichgewichts. *Zeitschrift für die gesamte Staatswissenschaft*, 121(2):301–324, 1965. *Cited in page 393*

[SFM24] Guruprerana Shabadi, Nathanaël Fijalkow, and Théo Matricon. Theoretical foundations for programmatic reinforcement learning. *CoRR*, abs/2402.11650, 2024. *Cited in page 16*

[Sha53] Lloyd S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953. *Cited in page 80*

[Sip96] Michael Sipser. *Introduction to the Theory of Computation.* International Thomson Publishing, 1st edition, 1996. *Cited in page 58*

[Tiw92] Prasoon Tiwari. A problem that is easier to solve on the unit-cost algebraic RAM. *Journal of Complexity*, 8(4):393–397, 1992. *4 citations in pages 86, 372, 375 and 380*

[Umm06] Michael Ummels. Rational behaviour and strategy construction in infinite multiplayer games. In S. Arun-Kumar and Naveen Garg, editors, *Proceedings of the 26th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FST TCS 2006, Kolkata, India, December 13–15, 2006*, volume 4337 of *Lecture Notes in Computer Science*, pages 212–223. Springer, 2006. *2 citations in pages 63 and 121*

[Umm08] Michael Ummels. The complexity of Nash equilibria in infinite multiplayer games. In Roberto M. Amadio, editor, *Proceedings of the 11th International Conference on Foundations of Software Science and Computational Structures, FoSSaCS 2008, Held as Part of ETAPS 2008, Budapest, Hungary, March 29 – April 6, 2008*, volume 4962 of *Lecture Notes in Computer Science*, pages 20–34. Springer, 2008. *2 citations in pages 64 and 92*

[UW11] Michael Ummels and Dominik Wojtczak. The complexity of Nash equilibria in limit-average games. In Joost-Pieter Katoen and Barbara König, editors, *Proceedings of the 22nd International Conference on Concurrency Theory, CONCUR 2011, Aachen, Germany, September 6–9, 2011*, volume 6901 of *Lecture Notes in Computer Science*, pages 482–496. Springer, 2011. *Cited in page 92*

[vM44] John von Neumann and Oskar Morgenstern. Theory of games and economic behavior. In *Theory of games and economic behavior*. Princeton university press, 1944. *Cited in page 2*

[von28] John von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928. *Cited in page 4*

[VW86] Moshe Y. Vardi and Pierre Wolper. An automata-theoretic approach to automatic program verification (preliminary report). In *Proceedings of the Symposium on Logic in Computer Science (LICS'86), Cambridge, Massachusetts, USA, June 16–18, 1986*, pages 332–344. IEEE Computer Society, 1986. *Cited in page 2*

Index

Table of notations

Set-theoretic symbols

$\mathbb{N}$	set of natural numbers (non-negative integers)
$\mathbb{N}_{>0}$	set of positive natural numbers
$\bar{\mathbb{N}}$	set $\mathbb{N} \cup \{+\infty\}$
$\bar{\mathbb{N}}_{>0}$	set $\mathbb{N}_{>0} \cup \{+\infty\}$
$\mathbb{Q}$	set of rational numbers
$\mathbb{R}$	set of real numbers
$\bar{\mathbb{R}}$	extended real line $\mathbb{R} \cup \{-\infty, +\infty\}$
$\mathbb{C}$	set of complex numbers
$\llbracket n, n' \rrbracket$	set of natural numbers between $n, n' \in \bar{\mathbb{N}}$
$\llbracket n \rrbracket$	shorthand for $\llbracket 0, n \rrbracket$, where $n \in \bar{\mathbb{N}}$
$[x, y]$	closed interval (of $\bar{\mathbb{R}}$)
$]x, y[$	open interval (of $\bar{\mathbb{R}}$)
$\mathbb{1}_A$	indicator of a set A
$f^{-1}(B)$	inverse image of set B by a function f
$f^{-1}(b)$	inverse image of a singleton $\{b\}$ by f
$\mathsf{Im}(f)$	image of a function f
$\lvert A \rvert$	cardinality of a set A
A^*	set of finite words over a set A
A^+	set of non-empty finite words over a set A
A^ω	set of infinite words over a set A
ε	empty word

Probability symbols

$\mathcal{D}(A)$	set of distributions over a countable set A
$\mathsf{supp}(\mu)$	support of a discrete distribution μ
$\mathcal{D}(B, \mathcal{F})$	set of distributions of a measurable space $(B, \mathcal{F})$

Topology notation

$(X, \mathcal{T})$	topological space
$\mathsf{cl}(D)$	closure of a set D
$\mathsf{int}(D)$	interior of a set D
$\mathsf{bd}(D)$	boundary of a set D

Vector space notation

$\mathbf{v}$, $\mathbf{w}$	vectors
v_j, w_j	vector components
α, β, α_j, β_j	scalars
$\mathbf{1}_d$, $\mathbf{1}$	d-dimensional vector where all components are 1
$\mathbf{0}_d$, $\mathbf{0}$	d-dimensional vector where all components are 0
$\langle \mathbf{v}, \mathbf{w} \rangle$	scalar product of vectors $\mathbf{v}$ and $\mathbf{w}$
$\|\mathbf{v}\|_2$	Euclidean norm of vector $\mathbf{v}$
$\mathsf{ker}(L)$	kernel of linear map L
x^*, y^*	linear forms
$\mathsf{aff}(D)$	affine span of a set $D \subseteq \mathbb{R}^d$
$\mathsf{ri}(D)$	relative interior of a set $D \subseteq \mathbb{R}^d$
$\leq_{\mathsf{lex}}$, $<_{\mathsf{lex}}$	non-strict and strict lexicographic order over $\mathbb{R}^d$
$\mathsf{down}(D)$	downward-closure of a set $D \subseteq \mathbb{R}^d$
$[\mathbf{v}, \mathbf{w}]$, $]\mathbf{v}, \mathbf{w}[$	closed/open segment from $\mathbf{v} \in \mathbb{R}^d$ to $\mathbf{w} \in \mathbb{R}^d$
$\mathsf{conv}(D)$	convex hull of a set $D \subseteq \mathbb{R}^d$
$\mathsf{extr}(D)$	set of extreme points of a convex set $D \subseteq \mathbb{R}^d$

Arenas and Markov decision processes

General notation

n	number of players (in multi-player arenas)

$\mathcal{P}_i$	player i
$\mathcal{A}$	arena
S	state space of an arena
$s, t \in S$	states of an arena
δ	transition function of an arena
$\mathsf{Plays}(\mathcal{A})$	set of plays of $\mathcal{A}$
$\pi \in \mathsf{Plays}(\mathcal{A})$	play
$\mathsf{Hist}(\mathcal{A})$	set of histories of $\mathcal{A}$
$h \in \mathsf{Hist}(\mathcal{A})$	history
w	prefix of a history ending in an action (profile)
$\mathsf{first}(\pi)$, $\mathsf{first}(h)$	first state of a play π or a history h
$\mathsf{last}(h)$	last state of a history h
$h_1 \cdot h_2$	concatenation of two histories h_1, h_2
$h \cdot \pi$	concatenation of a history h and a play π
$\mathsf{Cyl}_{\mathcal{A}}(h)$, $\mathsf{Cyl}(h)$	cylinder of a history h
$\mathsf{Cyl}_{\mathcal{A}}(\mathcal{H})$, $\mathsf{Cyl}(\mathcal{H})$	union of cylinders of histories in $\mathcal{H} \subseteq \mathsf{Hist}(\mathcal{A})$

Concurrent arenas

$A^{(i)}$	action space of $\mathcal{P}_i$ in a concurrent arena
$a^{(i)} \in A^{(i)}$	action of $\mathcal{P}_i$ in a concurrent arena
$\bar{A} = A^{(1)} \times \ldots \times A^{(n)}$	set of action profiles in a concurrent arena
$\bar{a} = (a^{(1)}, \ldots, a^{(n)}) \in \bar{A}$	action profile in a concurrent arena
$\mathcal{A} = (S, (A^{(i)})_{i \in [\![1,n]\!]}, \delta)$	n-player concurrent arena
$\mathcal{A} = (S, A^{(1)}, A^{(2)}, \delta)$	two-player concurrent arena
$\pi = s_0\bar{a}_0s_1\ldots$	play of a concurrent arena
$\pi_{\leq \ell}$	prefix $s_0\bar{a}_0s_1\ldots\bar{a}_{\ell-1}s_\ell$ of a play $\pi = s_0\bar{a}_0s_1\ldots$
$\pi_{\geq \ell}$	suffix $s_\ell\bar{a}_\ell s_{\ell+1}\ldots$ of a play $\pi = s_0\bar{a}_0s_1\ldots$
$h = s_0\bar{a}_0s_1\ldots\bar{a}_{r-1}s_r$	history of a concurrent arena

Turn-based arenas and Markov decision processes

S_i	set of states controlled by $\mathcal{P}_i$
A	action space of an MDP or a turn-based arena

$a \in A$	action of an MDP or a turn-based arena
$\mathcal{M} = (S, A, \delta)$	Markov decision process (MDP)
$\mathcal{A} = ((S_i)_{i \in \llbracket 1,n \rrbracket}, A, \delta)$	n-player turn-based arena
$\mathcal{A} = (S_1, S_2, A, \delta)$	two-player turn-based arena
$\pi = s_0 a_0 s_1 \dots$	play of a turn-based arena
$\pi_{\leq \ell}$	prefix $s_0 a_0 s_1 \dots a_{\ell-1} s_\ell$ of a play $\pi = s_0 a_0 s_1 \dots$
$\pi_{\geq \ell}$	suffix $s_\ell a_\ell s_{\ell+1} \dots$ of a play $\pi = s_0 a_0 s_1 \dots$
$h = s_0 a_0 s_1 \dots a_{r-1} s_r$	history of a turn-based arena
$\mathsf{Hist}_i(\mathcal{A})$	set of histories ending in a state of $\mathcal{P}_i$

Markov chains

$\mathcal{C} = (S, \delta)$	Markov chain
$\pi = s_0 s_1 \dots$	play of a Markov chain
$h = s_0 s_1 \dots s_r$	history of a Markov chain

Strategies

σ_i, τ_i	(pure or behavioural) strategies of $\mathcal{P}_i$
$\sigma = (\sigma_1, \dots, \sigma_n)$	(pure or behavioural) strategy profile
$\sigma = (\sigma_i, \sigma_{-i})$	strategy profile, highlighting the strategy of $\mathcal{P}_i$
$\Sigma^i(\mathcal{A})$	set of strategies of $\mathcal{P}_i$ in an arena $\mathcal{A}$
$\Sigma^i_{\mathsf{pure}}(\mathcal{A})$	set of pure strategies of $\mathcal{P}_i$ in an arena $\mathcal{A}$
σ, τ	strategies of an MDP
$\Sigma(\mathcal{M})$	set of strategies of an MDP
$\Sigma_{\mathsf{pure}}(\mathcal{M})$	set of pure strategies of an MDP
Σ	subset of strategies of an MDP
$\mathbb{P}^{\sigma}_{\mathcal{A},s}$, $\mathbb{P}^{\sigma}_{s}$	measure induced by strategy (profile) σ from s
$\mathbb{P}_{\mathcal{C},s}$, $\mathbb{P}_s$	measure over plays of a Markov chain $\mathcal{C}$ from s
$\mathsf{Out}_{\mathcal{A}}(\sigma, s)$	outcome of a pure profile σ with $\mathcal{A}$ deterministic
μ_i	mixed strategy of $\mathcal{P}_i$
$\mu = (\mu_1, \dots, \mu_n)$	mixed strategy profile
μ	mixed strategy of an MDP

Mealy machines

$\mathfrak{M}$, $\mathfrak{N}$	Mealy machine
M, N	state space of a Mealy machine
m, n	memory states
μ_{init}, ν_{init}	initial distribution
m_{init}, n_{init}	initial memory state
$\mathsf{nxt}_{\mathfrak{M}}$, $\mathsf{nxt}_{\mathfrak{N}}$	next-move function
$\mathsf{up}_{\mathfrak{M}}$, $\mathsf{up}_{\mathfrak{N}}$	update function
$(M, \mu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$, $(N, \nu_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{N}}, \mathsf{up}_{\mathfrak{N}})$	Mealy machine tuple
$(M, m_{\mathsf{init}}, \mathsf{nxt}_{\mathfrak{M}}, \mathsf{up}_{\mathfrak{M}})$	tuple with deterministic initialisation
μ_w, ν_w	distribution over memory states after w occurs
$\widehat{\mathsf{up}_{\mathfrak{M}}}$	iterated deterministic update function

Objectives and payoffs

Ω	objective
f	payoff or cost function
$\mathbb{E}_s^{\sigma}(f)$	expectation of f when following σ from s
$\mathsf{Reach}(T)$	reachability objective for target $T \subseteq S$
$\mathsf{Reach}(t)$	reachability objective for target $\{t\}$
$\mathsf{Safe}(U)$	safety objective for unsafe set $U \subseteq S$
$\mathsf{Safe}(t)$	safety objective for unsafe set $\{t\}$
$\mathsf{Büchi}(T)$	Büchi objective for target $T \subseteq S$
$\mathsf{Büchi}(t)$	Büchi objective for target $\{t\}$
$\mathsf{coBüchi}(U)$	co-Büchi objective for unsafe set $U \subseteq S$
$\mathsf{coBüchi}(t)$	co-Büchi objective for unsafe set $\{t\}$
$w\colon S \times \bar{A} \to \mathbb{R}$	weight function
$\mathsf{DSum}_w^{\lambda}$	discounted-sum payoff (weight w, discount λ)
TRew_w	total-reward payoff (weight w)
SPath_w^T	shortest-path cost (weight w, target T)
$\theta \in \bar{\mathbb{R}}$	threshold to be ensured

Games

$\mathcal{G} = (\mathcal{A}, (f_i)_{i \in \llbracket 1,n \rrbracket})$	n-player game on $\mathcal{A}$ with payoffs
$\mathcal{G} = (\mathcal{A}, (\Omega_i)_{i \in \llbracket 1,n \rrbracket})$	n-player game on $\mathcal{A}$ with objectives
$\mathcal{G} = (\mathcal{A}, f)$	two-player zero-sum game with a payoff/cost
$\mathcal{G} = (\mathcal{A}, \Omega)$	two-player zero-sum game with an objective
$\mathsf{Val}_{\mathcal{G}}(s)$	value of state s in zero-sum game $\mathcal{G}$

Imperfect information

$\mathcal{Z}_i$	observation space of $\mathcal{P}_i$
Obs_i	observation function of $\mathcal{P}_i$
$\mathfrak{P}$	arena with imperfect information
$(\mathcal{A}, (\mathcal{Z}_i, \mathsf{Obs}_i)_{i \in \llbracket 1,n \rrbracket})$	tuple for an arena with imperfect information

One-counter MDPs

Q	state space of an OC-MDP
$q,\ p,\ t \in Q$	states of an OC-MDP
$\mathcal{Q} = (Q, A, \delta, w)$	one-counter MDP (OC-MDP)
$k \in \mathbb{N}$	counter value in an OC-MDP
$s = (q, k)$	configuration of an OC-MDP
$B \in \bar{\mathbb{N}}_{>0}$	counter upper bound
$\mathcal{M}^{\leq B}(\mathcal{Q})$	MDP over configurations induced by $\mathcal{Q}$
$\delta^{\leq B}$	transition function of $\mathcal{M}^{\leq B}(\mathcal{Q})$
$\mathsf{Reach}(T)$	state-reachability objective for target $T \subseteq Q$
$\mathsf{Term}(T)$	selective termination objective for target $T \subseteq Q$

Notation for the construction of Nash equilibria (Part II)

$W_1(\Omega)$	winning region of $\mathcal{P}_1$ in a zero-sum game $(\mathcal{A}, \Omega)$
$W_2(\mathsf{Plays}(\mathcal{A}) \setminus \Omega)$	winning region of $\mathcal{P}_2$ in a zero-sum game $(\mathcal{A}, \Omega)$
$\mathcal{A}_i$	derivative of $\mathcal{A}$ for a coalition game against $\mathcal{P}_i$
$\mathcal{G}_i$	coalition game against $\mathcal{P}_i$
$W_i(\Omega_i)$	winning region of $\mathcal{P}_i$ in a coalition game $(\mathcal{A}_i, \Omega_i)$
$\mathsf{Val}^i_{\mathcal{G}}(s)$	value of state s in a coalition game $\mathcal{G}_i$

sg	segment of a play
$\mathcal{S}$	segment decomposition of a play
$\mathsf{VisPl}^{\mathcal{G}}(\pi)$	players whose targets are visited (Reach, SPath)
$\mathsf{VisPos}^{\mathcal{G}}(\pi)$	positions of target visits in π (Reach, SPath)

Classification of randomised strategies (Part III)

XYZ Mealy machine	X: initialisation; Y: outputs; Z: updates
XYZ strategies	strategy induced by an XYZ Mealy machine
Σ_h^i	set of pure strategies of $\mathcal{P}_i$ consistent with h
$\mathcal{C}_{n,i}^{FP}$	fin. arenas with perfect recall for $\mathcal{P}_i$
$\mathcal{C}_{n,i}^{IP}$,	inf. arenas with perfect recall for $\mathcal{P}_i$
$\mathcal{C}_{n,i}^{FI}$	fin. arenas with or without perfect recall for $\mathcal{P}_i$
$\mathcal{C}_{n,i}^{II}$	inf. arenas with or without perfect recall for $\mathcal{P}_i$

Multi-objective MDPs (Part IV)

d	number of payoff functions
$\bar{f} = (f_j)_{j \in [\![1,d]\!]}$	multi-dimensional payoff
$f^+ = \max(f, 0)$	non-negative part of a one-dimensional payoff f
$f^- = \max(-f, 0)$	non-positive part of a one-dimensional payoff f
$\mathbf{q}$, $\mathbf{p}$	expected payoff vectors or achievable vectors
q_j, p_j	components of expected payoff vectors
$\mathsf{Pay}_s(\bar{f})$	expectations of $\bar{f}$ from s for all strategies
$\mathsf{Ach}_s(\bar{f})$	achievable vectors
$\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$	expectations of $\bar{f}$ from s for all pure strategies
$\mathsf{Ach}_s^{\mathsf{pure}}(\bar{f})$	purely achievable vectors
$\mathsf{Pay}_s^{\Sigma}(\bar{f})$	expectations of $\bar{f}$ from s for all $\sigma \in \Sigma$
$\mathsf{Ach}_s^{\Sigma}(\bar{f})$	achievable vectors witnessed by some $\sigma \in \Sigma$
$\mathsf{dist}_{\mathsf{proba}}$	metric over distributions in Chapter 15.1
$\mathcal{E} = (E, A_{\mathcal{E}})$	end-component of an MDP

Interval strategies in OC-MDPs (Part V)

General notation

$\mathcal{I}$, $\mathcal{J}$, $\mathcal{K}$	interval partition of set of counter values
$I \in \mathcal{I}$	interval of $\bar{\mathbb{N}}$
b^+, b^-	upper and lower bounds of an interval
OEIS	open-ended interval strategy
CIS	cyclic interval strategy
ρ	period of an interval partition or a CIS

Compressed Markov chains

$\mathcal{C}^\sigma_\mathcal{I}(\mathcal{Q})$, $\mathcal{C}^\sigma_\mathcal{I}$	compressed Markov chain for σ on $\mathcal{Q}$ w.r.t. $\mathcal{I}$
$S_\mathcal{I}$	state space of $\mathcal{C}^\sigma_\mathcal{I}$
$S^\perp_\mathcal{I}$	absorbing states of $\mathcal{C}^\sigma_\mathcal{I}$
$\delta^\sigma_\mathcal{I}$	transition function of $\mathcal{C}^\sigma_\mathcal{I}$
β, β_I	denotes $\log_2(\lvert I\rvert + 1)$ for some $I \in \mathcal{I}$
α	natural number smaller than β or β_I
$\mathcal{H}_{\mathsf{succ}}(s, s')$	histories from s to s' with no other $\mathcal{C}^\sigma_\mathcal{I}$-successor
$\bar{h}$	history of $\mathcal{C}^\sigma_\mathcal{I}$
$\mathcal{R}^\sigma_\mathcal{J} = (R_\mathcal{J}, \delta^\sigma_\mathcal{J})$	one-counter Markov chain for $\mathcal{C}^\sigma_\mathcal{I}$ if σ is a CIS

Transition probabilities in compressed Markov chains

We refer the reader to the text prefacing Theorem 18.6 (unbounded case, Page 326) and Theorem 18.9 (bounded case, Page 330) for a more precise description of the notation described below. In the following, *induced MC* refers to the induced Markov chain from which we derive a compressed Markov chain.

$\langle q \searrow p\rangle$	variable in the unbounded case
$\mathcal{H}_\alpha((q, k) \nearrow p)$	sets of histories for bounded case
$[(q, k) \nearrow p]_\alpha$	probability of $\mathcal{H}_\alpha((q, k) \nearrow p)$ in induced MC
$\langle (q, k) \nearrow p\rangle_\alpha$	variable for probability of $\mathcal{H}_\alpha((q, k) \nearrow p)$

$\mathcal{H}_\alpha((q,k) \searrow p)$	sets of histories for bounded case
$[(q,k) \searrow p]_\alpha$	probability of $\mathcal{H}_\alpha((q,k) \searrow p)$ in induced MC
$\langle (q,k) \searrow p \rangle_\alpha$	variable for probability of $\mathcal{H}_\alpha((q,k) \searrow p)$

Verification and realisability (logical formulae)

boldface letter, e.g., $\mathbf{z}$	vector or set of variables
starred variable, e.g., $\mathbf{z}^\star$	valuation of the variable (vector)
$\mathbf{y}$, y_s	variables for probabilities of the objective
$z^I_{q,a}$, $\mathbf{z}^I$, $\mathbf{z}$	variables for strategy probabilities
$\tau_{\mathbf{z}}$	interval strategy parameterised by $\mathbf{z}$
$\mathcal{C}_{\mathcal{I}}^{\tau_{\mathbf{z}}}$, $\mathcal{C}_{\mathcal{I}}^{\mathbf{z}}$	compressed Markov chain for $\tau_{\mathbf{z}}$
$\delta_{\mathcal{I}}^{\tau_{\mathbf{z}}}$, $\delta_{\mathcal{I}}^{\mathbf{z}}$	transition function of $\mathcal{C}_{\mathcal{I}}^{\mathbf{z}}$
$\mathbf{x}$, $x_{s,s'}$	variables for probabilities in $\mathcal{C}_{\mathcal{I}}^{\mathbf{z}}$ (OEIS)
$\Phi_\delta^{\mathcal{I}}(\mathbf{x},\mathbf{z})$	formula for transition probabilities of $\mathcal{C}_{\mathcal{I}}^{\mathbf{z}}$ (OEIS)
$\Phi_\Omega^{\mathcal{I}}(\mathbf{x},\mathbf{y})$	formula for objective probabilities in $\mathcal{C}_{\mathcal{I}}^{\mathbf{z}}$ (OEIS)
$\mathcal{R}_{\mathcal{J}}^{\tau_{\mathbf{z}}}$, $\mathcal{R}_{\mathcal{J}}^{\mathbf{z}}$	one-counter Markov chain inducing $\mathcal{C}_{\mathcal{I}}^{\mathbf{z}}$ (CIS)
$\delta_{\mathcal{J}}^{\tau_{\mathbf{z}}}$	transition function of $\mathcal{R}_{\mathcal{J}}^{\mathbf{z}}$
$\mathcal{C}_{\mathcal{K}}(\mathcal{R}_{\mathcal{J}}^{\mathbf{z}})$	compression of $\mathcal{C}^{\leq\infty}(\mathcal{R}_{\mathcal{J}}^{\mathbf{z}})$ with respect to $\mathcal{K}$
$S_{\mathcal{K}}(R_{\mathcal{J}})$	state space of $\mathcal{C}_{\mathcal{K}}(\mathcal{R}_{\mathcal{J}}^{\mathbf{z}})$
$\bar{s}$	element of $S_{\mathcal{K}}(R_{\mathcal{J}})$
$\delta_{\mathcal{K}}[\mathcal{R}_{\mathcal{J}}^{\mathbf{z}}]$	transition function of $\mathcal{C}_{\mathcal{K}}(\mathcal{R}_{\mathcal{J}}^{\mathbf{z}})$
$\mathbf{v}$, $v_{s,s',u}$	variables for probabilities in $\mathcal{R}_{\mathcal{J}}^{\mathbf{z}}$
$\mathbf{x}$, $x_{\bar{s},\bar{s}'}$	variables for probabilities in $\mathcal{C}_{\mathcal{K}}(\mathcal{R}_{\mathcal{J}}^{\mathbf{z}})$
$\Psi_\delta^{\mathcal{J}}(\mathbf{v},\mathbf{z})$	formula for transition probabilities in $\mathcal{R}_{\mathcal{J}}^{\mathbf{z}}$
$\Phi_\delta^{\mathcal{K}}(\mathbf{x},\mathbf{v})$	formula for probabilities in $\mathcal{C}_{\mathcal{K}}(\mathcal{R}_{\mathcal{J}}^{\mathbf{z}})$ (CIS)
$\Phi_\Omega^{\mathcal{K}}(\mathbf{x},\mathbf{y})$	formula for objective probabilities in $\mathcal{C}_{\mathcal{K}}(\mathcal{R}_{\mathcal{J}}^{\mathbf{z}})$
$\Phi_\sigma^{\mathcal{I},\mathcal{I}',\mathcal{B}}(\mathbf{z})$	formula for strategy probabilities (bounded)
$\Phi_\sigma^{\mathcal{I},\mathcal{I}'}(\mathbf{z})$	formula for strategy probabilities (OEIS)
$\Phi_\sigma^{\mathcal{J},\mathcal{J}'}(\mathbf{z})$	formula for strategy probabilities (CIS)

Square-root-sum hardness

$x_1, \dots, x_n$, y	inputs to the square-root-sum problem
$\mathbf{x} = (x_1, \dots, x_n)$	vector of square-root-sum inputs other than y

m	$\max_{i \in [\![1,n]\!]} x_i$ in a square-root-sum instance
$\mathcal{Q}_{\mathbf{x}}$	one-counter Markov chain used in our reduction
ε_B	error on termination probability in $\mathcal{C}^{\leq B}(\mathcal{Q}_{\mathbf{x}})$

NP-hardness

V	(finite) set of vertices
$v \in V$	vertex
$E \subseteq V \times V$	set of edges
$G = (V, E)$	directed graph