5

Turn-Based Stochastic Games

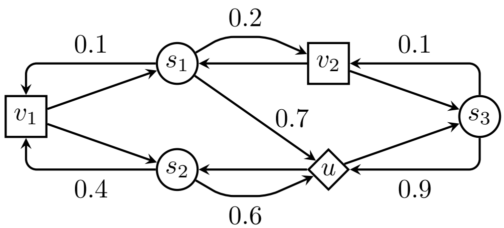
Antonín Kučera
Masaryk University

Abstract

In this chapter, we give a taxonomy of winning objectives in stochastic turn-based games, discuss their basic properties, and present an overview of the existing results. Special attention is devoted to games with infinitely many vertices.

5.1 Introduction

Turn-based stochastic games are infinitely long sequential games of perfect information played by two 'ordinary' players and a random player. Such games are also known as *simple stochastic games* and are a special type of (simultaneous-choice) *stochastic games* considered by Shapley [1953]. Intuitively, a turn-based stochastic game is a directed graph with a finite or countably infinite set of vertices, where each vertex 'belongs' either to player $\Box$, player $\Diamond$, or the random player $\bigcirc$. An example of a turn-based stochastic game with finitely many vertices is given below.



The outgoing transitions of stochastic vertices (drawn as circles) are selected randomly according to fixed probability distributions. In the other vertices (boxes and diamonds), the outgoing transitions are selected by the respective player according to his *strategy*, which may be *randomised* and *history-dependent*. Thus, every pair of strategies for both players determines a *play* of the game, which is a Markov chain obtained by applying the strategies to the original game. The aim of player $\Box$ is to maximise the expected *payoff* associated to runs in plays, or to play so that a certain *property* is satisfied. Player $\Diamond$ usually (but not necessarily) aims at the opposite.

In computer science, turn-based stochastic games are used as a natural model for discrete systems where the behaviour in each state is either controllable, adversarial, or stochastic. The main question is whether there is a suitable controller (strategy of player $\Box$) such that the system satisfies a certain property no matter what the environment and unpredictable users do. For a successful implementation of a controller, it is also important what kind of information about the computational history is required and whether the controller needs to randomise. This is the main source of motivation for considering the abstract problems presented in the next sections.

Since real-world computational systems are usually very large and complex, they can be analysed only indirectly by constructing a simplified *formal model*. A formal model of a given system is an abstract computational device which faithfully reflects the important behavioural aspects of the system. For purposes of formal modeling, the expressive power of *finite-state* devices is often insufficient, and some kind of unbounded data storage (such as counters, channels, stacks, queues, etc.) is needed to obtain a sufficiently precise model. Hence, in the computer science context, the study of turn-based stochastic games is not limited just to finite-state games, but also includes certain classes of *infinite-state* games which correspond to various types of computational devices such as pushdown automata, channel systems, vector addition systems, or process calculi.

5.1.1 Preliminaries

We start by recalling some notation and basic concepts that are necessary for understanding the material presented in subsequent sections.

Words, paths, and runs

In this chapter, the set of all real numbers is denoted by $\mathbb{R}$, and we also use the standard way of writing intervals of real numbers (e.g., $(0, 1]$ abbreviates $\{x \in \mathbb{R} \mid 0 < x \leq 1\}$).

Let M be a finite or countably infinite alphabet. A ***word*** over M is a finite or infinite sequence of elements of M. The empty word is denoted by ε, and the set of all finite words over M is denoted by M^*. Sometimes we also use M^+ to denote the set $M^* \setminus \{\varepsilon\}$. The length of a given word w is denoted by $len(w)$, where $len(\varepsilon) = 0$ and the length of an infinite word is ∞. The individual letters in w are denoted by $w(0), w(1), \ldots$, and for every infinite word w and every $i \geq 0$ we use w_i to denote the infinite word $w(i), w(i{+}1), \ldots$

A ***transition system*** is a pair $\mathcal{T} = (S, \rightarrow)$, where S is a finite or countably infinite set of ***states*** and $\rightarrow \subseteq S \times S$ a ***transition relation*** such that for every $s \in S$ there is at least one outgoing transition (i.e., a transition of the form $s \rightarrow t$). We say that $\mathcal{T}$ is ***finitely branching*** if every $s \in S$ has only finitely many outgoing transitions. A ***path*** in $\mathcal{T}$ is a finite or infinite word w over S such that $w(i) \rightarrow w(i{+}1)$ for every $0 \leq i < len(w)$. A ***run*** is an infinite path. The sets of all finite paths and all runs in $\mathcal{T}$ are denoted by $Fpath(\mathcal{T})$ and $Run(\mathcal{T})$, respectively. Similarly, for a given $w \in Fpath(\mathcal{T})$, we use $Fpath(\mathcal{T}, w)$ and $Run(\mathcal{T}, w)$ to denote the sets of all finite paths and all runs that start with w, respectively. When $\mathcal{T}$ is clear from the context, it is usually omitted (for example, we write just Run instead of $Run(\mathcal{T})$).

Probability spaces

Let A be a finite or countably infinite set. A ***probability distribution*** on A is a function $\mu : A \rightarrow [0, 1]$ such that $\sum_{a \in A} \mu(a) = 1$. A distribution μ is ***rational*** if $\mu(a)$ is rational for every $a \in A$, ***positive*** if $\mu(a) > 0$ for every $a \in A$, and ***Dirac*** if $\mu(a) = 1$ for some $a \in A$. A Dirac distribution μ where $\mu(a) = 1$ is also denoted by μ_a or just a.

Let Ω be a set of ***elementary events***. A ***σ-field*** over Ω is a set $\mathcal{F} \subseteq 2^\Omega$ that includes Ω and is closed under complement and countable union. A ***measurable space*** is a pair $(\Omega, \mathcal{F})$ where $\mathcal{F}$ is a σ-field over Ω. An ***$\mathcal{F}$-measurable function*** over $(\Omega, \mathcal{F})$ is a function $f : \Omega \rightarrow \mathbb{R}$ such that $f^{-1}(I) \in \mathcal{F}$ for every interval I in $\mathbb{R}$.

Example 5.1 Let $\mathcal{T} = (S, \rightarrow)$ be a transition system. Let $\mathcal{B}$ be the least σ-field over Run containing all ***basic cylinders*** $Run(w)$ where $w \in Fpath$ (i.e., $\mathcal{B}$ is the Borel σ-field generated by open sets in the Cantor topology on Run). Then $(Run, \mathcal{B})$ is a measurable space, and the elements of $\mathcal{B}$ are called ***Borel*** sets of runs.

The Borel σ-field $\mathcal{B}$ contains many interesting elements. For example, let $s, t \in S$ and let $Reach(s, t)$ be the set of all runs initiated in s which visit t. Obviously, $Reach(s, t)$ is the union of all basic cylinders $Run(w)$ where $w \in Fpath(s)$ and w visits t, and hence $Reach(s, t) \in \mathcal{B}$. Similarly, one can show that the set of all runs initiated in s that visit t infinitely often is Borel.

Actually, most of the 'interesting' sets of runs are Borel, although there exist also subsets of Run that are not in $\mathcal{B}$.

Let $A \in \mathcal{B}$, and let $f : Run \to \{0,1\}$ be a function which to a given $w \in Run$ assigns either 1 or 0, depending on whether $w \in A$ or not, respectively. Then f is $\mathcal{B}$-measurable, because for every interval I in $\mathbb{R}$ we have that $f^{-1}(I)$ is equal either to Run, A, $Run \setminus A$, or $\emptyset$, depending on whether $I \cap \{0,1\}$ is equal to $\{0,1\}$, $\{1\}$, $\{0\}$, or $\emptyset$, respectively. □

A ***probability measure*** over a measurable space $(\Omega, \mathcal{F})$ is a function $\mathcal{P} : \mathcal{F} \to [0,1]$ such that, for each countable collection $\{A_i\}_{i\in I}$ of pairwise disjoint elements of $\mathcal{F}$, we have that $\mathcal{P}(\bigcup_{i\in I} A_i) = \sum_{i\in I} \mathcal{P}(A_i)$, and moreover $\mathcal{P}(\Omega) = 1$. A ***probability space*** is a triple $(\Omega, \mathcal{F}, \mathcal{P})$ where $(\Omega, \mathcal{F})$ is a measurable space and $\mathcal{P}$ is a probability measure over $(\Omega, \mathcal{F})$.

A ***random variable*** over a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ is an $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}$. The ***expected value of*** X, denoted by $\mathbb{E}(X)$, is defined as the Lebesgue integral $\int_\Omega X \, d\mathcal{P}$. A random variable X is ***discrete*** if there is a finite or countably infinite subset N of $\mathbb{R}$ such that $\mathcal{P}(X^{-1}(N)) = 1$. The expected value of a discrete random variable X is equal to $\sum_{n\in N} n \cdot \mathcal{P}(X{=}n)$, where $X{=}n$ denotes the set $X^{-1}(\{n\})$.

Markov chains

A ***Markov chain*** is a tuple $\mathcal{M} = (S, \to, Prob, \mu)$ where $(S, \to)$ is a transition system, $Prob$ is a function which to each $s \in S$ assigns a positive probability distribution over the outgoing transitions of s, and $\mu : S \to [0,1]$ is an *initial* probability distribution. We write $s \xrightarrow{x} t$ to indicate that $s \to t$ and $Prob(s)((s,t)) = x$.

Consider the measurable space $(Run, \mathcal{B})$ introduced in Example 5.1, i.e., $\mathcal{B}$ is the least σ-field containing all $Run(w)$ such that $w \in Fpath$. By Carathéodory's extension theorem (see, e.g., Billingsley [1995], Kemeny et al. [1976]), there is a unique probability measure $\mathcal{P}_\mu$ over $(Run, \mathcal{B})$ such that for every basic cylinder $Run(w)$ we have that $\mathcal{P}_\mu(Run(w))$ is defined in the 'natural' way, i.e., by multiplying $\mu(w(0))$ with the probabilities of all transitions in w. Thus, we obtain the probability space $(Run, \mathcal{B}, \mathcal{P}_\mu)$.

Example 5.2 Consider the finite-state Markov chain $\mathcal{M}$ of Figure 5.1 with the initial distribution μ_s (recall that μ_s is a Dirac distribution where $\mu_s(s) = 1$). Let $Reach(s,t)$ be the set of all runs initiated in s which visit t. Obviously, $Reach(s,t) = \bigcup_{i=1}^{\infty} Reach^i(s,t)$, where $Reach^i(s,t)$ consists of all $w \in Reach(s,t)$ that visit t for the first time after exactly i transitions. Further, every $Reach^i(s,t)$ can be partitioned into 2^{i-1} pairwise disjoint basic cylinders according to the first $i-1$ transitions. Each of these cylinders has

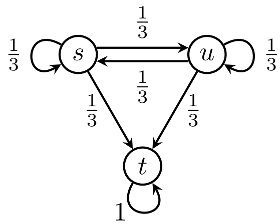


Figure 5.1 A finite-state Markov chain $\mathcal{M}$

the probability $(\frac{1}{3})^i$, hence $\mathcal{P}(Reach^i(s,t)) = (\frac{2}{3})^{i-1} \cdot \frac{1}{3}$. Since $Reach^i(s,t) \cap Reach^j(s,t) = \emptyset$ whenever $i \neq j$, we obtain

$$\mathcal{P}(Reach(s,t)) = \mathcal{P}\left(\bigcup_{i=1}^{\infty} Reach^i(s,t)\right) = \sum_{i=1}^{\infty} \mathcal{P}(Reach^i(s,t)) = \sum_{i=1}^{\infty} \left(\frac{2}{3}\right)^{i-1} \cdot \frac{1}{3} = 1$$

as expected. Also note that $Run(s) \setminus Reach(t)$ is uncountable but its probability is zero. $\square$

Turn-based stochastic games and Markov decision processes

A ***(turn-based) stochastic game*** is a tuple $G = (V, \mapsto, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ where $(V, \mapsto)$ is a transition system, $(V_\Box, V_\Diamond, V_\bigcirc)$ is a partition of V, and $Prob$ is a function which to each $v \in V_\bigcirc$ assigns a positive probability distribution on the set of its outgoing transitions. We say that G is ***finitely branching*** if $(V, \mapsto)$ is finitely branching. A ***Markov decision process (MDP)*** is a stochastic game where $V_\Diamond = \emptyset$ or $V_\Box = \emptyset$. Note that Markov chains can be seen as stochastic games where $V_\Box = V_\Diamond = \emptyset$.

A stochastic game is played by two players, $\Box$ and $\Diamond$, who select transitions in the vertices of $V_\Box$ and $V_\Diamond$, respectively. Let $\odot \in \{\Box, \Diamond\}$. A ***strategy*** for player $\odot$ is a function which to each $wv \in V^* V_\odot$ assigns a probability distribution on the set of outgoing transitions of v. The sets of all strategies for player $\Box$ and player $\Diamond$ are denoted by Σ_G and Π_G (or just by Σ and Π if G is understood), respectively. We say that a strategy τ is ***memoryless (M)*** if $\tau(wv)$ depends just on the last vertex v, and ***deterministic (D)*** if $\tau(wv)$ is Dirac for all wv. Strategies that are not necessarily memoryless are called ***history-dependent (H)***, and strategies that are not necessarily deterministic are called ***randomised (R)***. A special type of history-dependent strategies are ***finite-memory (F)*** strategies. A strategy τ of player $\odot$ is a finite-memory strategy if there is a finite set $C = \{c_1, \ldots, c_n\}$ of *colours*, a *colouring* $\nu : V \to C$, and a deterministic finite-state automaton M over the alphabet C such that for every $wv \in V^* V_\odot$ we have that $\tau(wv)$ depends only on v and the control state entered by M after reading the (unique) word $\nu(wv)$ obtained by replacing each vertex in wv with its associated colour. ***Infinite-memory*** strategies are strategies which are not finite-memory. To sum up, we consider the following six *types* of strategies: MD, MR, FD, FR, HD, and HR, where XD $\subseteq$ XR for every X $\in \{\mathrm{M}, \mathrm{F}, \mathrm{H}\}$ and MY $\subseteq$ FY $\subseteq$ HY for every Y $\in \{\mathrm{D}, \mathrm{R}\}$. The sets of all XY strategies of player $\Box$ and player $\Diamond$ are denoted by Σ^{XY} and Π^{XY}, respectively (note that $\Sigma = \Sigma^{\mathrm{HR}}$ and $\Pi = \Pi^{\mathrm{HR}}$).

Every pair of strategies $(\sigma, \pi) \in \Sigma \times \Pi$ and every ***initial*** probability distribution $\mu : V \to [0,1]$ determine a unique ***play*** of the game G, which is a Markov chain $G_\mu^{(\sigma,\pi)}$ where V^+ is the set of states, and $wu \xrightarrow{x} wuu'$ iff $u \mapsto u'$, $x > 0$, and one of the following conditions holds:

- $u \in V_\Box$ and $\sigma(wu)$ assigns x to $u \mapsto u'$;
- $u \in V_\Diamond$ and $\pi(wu)$ assigns x to $u \mapsto u'$;
- $u \in V_\bigcirc$ and $u \overset{x}{\mapsto} u'$.

The initial distribution of $G_\mu^{(\sigma,\pi)}$ assigns $\mu(v)$ to all $v \in V$, and zero to the other states. Note that every run w in $G_\mu^{(\sigma,\pi)}$ naturally determines a unique run w_G in G, where $w_G(i)$ is the last vertex of $w(i)$ for every $i \geq 0$.

5.2 Winning objectives in stochastic games

In this section we give a taxonomy of winning objectives in turn-based stochastic games and formulate the main problems of the associated algorithmic analysis. For the rest of this section, we fix a turn-based stochastic game $G = (V, \mapsto, (V_\Box, V_\Diamond, V_\bigcirc), \mathit{Prob})$.

Let $\mathit{plays}(G)$ be the set of all plays of G (i.e., $\mathit{plays}(G)$ consists of all Markov chains of the form $G_\mu^{(\sigma,\pi)}$). For every $\odot \in \{\Box, \Diamond\}$, let $\mathit{yield}_\odot : \mathit{plays}(G) \to \mathbb{R}$ be a function which to every play of G assigns the ***yield*** of player $\odot$.

Remark In the standard terminology of game theory, the yield of a given player under a given strategy profile is usually called a ***payoff***. However, in the context of turn-based stochastic games, the word 'payoff' usually refers to a function $f_\odot : \mathit{Run}(G) \to \mathbb{R}$ whose *expected value* is to be maximised by player $\odot$ (see Section 5.2.1 for more details). $\Box$

The objective of each player is to maximise his yield. For turn-based stochastic games, most of the studied yield functions are ***zero-sum***, i.e., for every play $G_\mu^{(\sigma,\pi)}$ of G we have that

$$\mathit{yield}_\Box(G_\mu^{(\sigma,\pi)}) + \mathit{yield}_\Diamond(G_\mu^{(\sigma,\pi)}) = 0.$$

For zero-sum games, it suffices to specify just $\mathit{yield}_\Box$, which is denoted simply by *yield*. Then, the objective of player $\Box$ and player $\Diamond$ is to maximise and minimise the value of *yield*, respectively.

There are two major classes of zero-sum turn-based stochastic games, which can be characterised as follows:

(1) ***Games with Borel measurable payoffs.*** Every $\mathcal{B}$-measurable function over $\mathit{Run}(G)$ determines a unique random variable over the runs of a given play of G. The yield assigned to the play is the expected value of this random variable.

(2) ***Win–lose games.*** A win–lose yield is either 1 or 0 for every play of G, depending of whether the play satisfies a certain property or not, respectively. Typically, the property is encoded as a formula of some probabilistic temporal logic.

These two classes of games are formally introduced and discussed in the next subsections.

Let us note that there are also results about *non-zero-sum* turned-based stochastic games, where the objectives of the (two or more) players are not necessarily conflicting. The main question is the existence and effective computability of Nash equilibria for various classes of players' objectives. The problem has been considered also in the more general setting of simultaneous-choice stochastic games. We refer to Secchi and Sudderth [2001], Chatterjee et al. [2004c, 2006] and Ummels and Wojtczak [2009] for more details.

5.2.1 Games with Borel measurable payoffs

A ***payoff*** is a bounded[1] $\mathcal{B}$-measurable function $f : \mathit{Run}(G) \to \mathbb{R}$. Observe that for every pair of strategies $(\sigma, \pi) \in \Sigma \times \Pi$ and every initial probability distribution μ on V, the function $f_\mu^{\sigma,\pi} : \mathit{Run}(G_\mu^{(\sigma,\pi)}) \to \mathbb{R}$ defined by $f_\mu^{\sigma,\pi}(w) = f(w_G)$ is a random variable. Thus, every payoff f determines the associated yield defined by

$$\mathit{yield}_f(G_\mu^{(\sigma,\pi)}) = \mathbb{E}(f_\mu^{\sigma,\pi}).$$

As observed by Maitra and Sudderth [1998], the determinacy result for Blackwell games achieved by Martin [1998] implies that for every vertex $v \in V$ we have the following:

$$\sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathbb{E}(f_v^{\sigma,\pi}) = \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \mathbb{E}(f_v^{\sigma,\pi}). \tag{5.1}$$

[1] A real-valued function f is *bounded* if there is $b \in \mathbb{R}$ such that $-b \le f(x) \le b$ for every x in the domain of f.

Hence, every vertex $v \in V$ has its ***f*-value**, denoted by $val_f(v, G)$ (or just by $val_f(v)$ if G is understood), which is defined by Equality (5.1). Note that this result holds without any additional assumptions about the game G. In particular, G may have infinitely many vertices and some (or all) of them may have infinitely many outgoing transitions.

Remark It is worth noting that the result presented by Maitra and Sudderth [1998] is even more general and holds also for ***concurrent*** stochastic games (where both players make their choice simultaneously) with an *arbitrary* (not necessarily countable) action and state spaces. The only requirement is that f is a bounded $\mathcal{B}$-measurable function and both players choose their actions at random according to finitely additive probability measures on the power sets of their respective action sets. This can be used, e.g., to show that vertices in various types of *timed* stochastic games have an f-value. $\square$

An important subclass of payoff functions are ***qualitative*** payoffs which simply classify each run as good or bad according to some criteria. The good runs are assigned 1, and the bad ones are assigned 0. General (not necessarily qualitative) payoffs are also called ***quantitative*** payoffs.

Note that a qualitative payoff f is just a characteristic function of some Borel set $B_f \subseteq Run(G)$. For every pair of strategies $(\sigma, \pi) \in \Sigma \times \Pi$ and every initial distribution μ we have that

$$\mathbb{E}(f_\mu^{\sigma,\pi}) = \mathcal{P}_\mu(\{w \in Run(G_\mu^{(\sigma,\pi)}) \mid w_G \in B_f\}).$$

Hence, player $\Box$ and player $\Diamond$ in fact try to maximise and minimise the *probability* of all runs in B_f, respectively.

Observe that Equality (5.1) does not guarantee the existence of ***optimal*** strategies for player $\Box$ and player $\Diamond$ which would achieve the yield $val_f(v)$ or better against every strategy of the opponent. As we shall see, optimal strategies do *not* exist in general, but they may exist in some restricted cases. On the other hand, Equality (5.1) *does* imply the existence of ε-optimal strategies for an arbitrarily small $\varepsilon > 0$.

Definition 5.3 Let $\varepsilon \geq 0$. A strategy $\hat{\sigma} \in \Sigma$ is an ***ε-optimal maximising strategy*** in a vertex $v \in V$ if $\inf_{\pi \in \Pi} \mathbb{E}(f_v^{\hat{\sigma},\pi}) \geq val_f(v) - \varepsilon$. Similarly, a strategy $\hat{\pi} \in \Pi$ is an ***ε-optimal minimising strategy*** in a vertex $v \in V$ if $\sup_{\sigma \in \Sigma} \mathbb{E}(f_v^{\sigma,\hat{\pi}}) \leq val_f(v) + \varepsilon$. Strategies that are 0-optimal are ***optimal***.

Qualitative payoffs

Let $C = \{c_1, \ldots, c_n\}$ be a finite set of ***colours***, and $\nu : V \to 2^C$ a ***valuation***. An important and well-studied class of qualitative payoffs are characteristic functions of ω-***regular*** subsets of $Run(G)$. The membership in an ω-regular set of runs is determined by one of the ***acceptance conditions*** listed below. These conditions correspond to acceptance criteria of finite-state automata over infinite words (see Section 5.2.2 for more details).

- ***Reachability and safety.*** A run $w \in Run(G)$ satisfies the *reachability* condition determined by a colour $c \in C$ if $c \in \nu(w(i))$ for some $i \geq 0$. The *safety* condition determined by c is dual, i.e., $c \notin \nu(w(i))$ for all $i \geq 0$.
- ***Büchi and co-Büchi.*** A run $w \in Run(G)$ satisfies the *Büchi* condition determined by a colour $c \in C$ if $c \in \nu(w(i))$ for infinitely many $i \geq 0$. The *co-Büchi* condition is dual, i.e., there are only finitely many $i \geq 0$ such that $c \in \nu(w(i))$.
- ***Rabin, Rabin-chain, and Street.*** Let $Pairs = \{(c_1, d_1), \ldots, (c_m, d_m)\}$ be a finite set of pairs of colours. A run $w \in Run(G)$ satisfies the *Rabin* condition determined by $Pairs$ if there is $(c, d) \in Pairs$ such that w satisfies the Büchi condition determined by d and the co-Büchi condition determined by c. The *Street* condition determined by $Pairs$ is dual to Rabin, i.e., for every $(c, d) \in Pairs$ we have that w satisfies the co-Büchi condition determined by d or the Büchi condition determined by c.

  For a given colour c, let $V(c)$ be the set of all $v \in V$ such that $c \in \nu(v)$. The *Rabin-chain* (or *parity*) condition is a special case of the Rabin condition where $Pairs$ and ν satisfy $V(c_1) \subset V(d_1) \subset \cdots \subset V(c_m) \subset V(d_m)$.
- ***Muller.*** Let $M \subseteq 2^C$ be a set of subsets of colours. A run $w \in Run(G)$ satisfies the *Muller* condition determined by M if the set of all $c \in C$ such w satisfies the Büchi condition determined by c is an element of M.

Let us note that ω-regular sets of runs are relatively simple in the sense that they are contained in the first two levels of the Borel hierarchy (the sets of runs satisfying the reachability and safety conditions are in the first level).

Quantitative payoffs

Quantitative payoff functions can capture more complicated properties of runs that are particularly useful in performance and dependability analysis of stochastic systems.

Let $r : V \to \mathbb{R}$ be a bounded function which to every vertex v assigns the ***reward*** $r(v)$, which can be intuitively interpreted as a price paid by player $\Diamond$ to player $\Box$ when visiting v. The limit properties of rewards that are paid along a given run are captured by the following payoff functions:

- ***Limit-average payoff*** (also ***mean-payoff***) assigns to each $w \in Run(G)$

the average reward per vertex visited along w. For every $n \geq 1$, let $avg^n(w) = \frac{1}{n}\sum_{i=0}^{n-1} r(w(i))$. Since $\lim_{n\to\infty} avg^n(w)$ may not exist for a given run w, the mean-payoff function appears in two flavours, defined as follows:

$$avg\ sup(w) = \limsup_{n\to\infty} avg^n(w) \qquad avg\ inf(w) = \liminf_{n\to\infty} avg^n(w).$$

Let us note that if the players play in a sufficiently 'weird' way, it may happen that $avg\ sup(w) \neq avg\ inf(w)$ for all runs w in a given play, even if the underlying game G has just two vertices.

Mean-payoff functions were introduced by Gillette [1957]. For *finite-state* games, it is sometimes stipulated that the aim of player $\Box$ is to maximise the expectation of $avg\ sup$ and the aim of player $\Diamond$ is to minimise the expectation of $avg\ inf$. For finite-state games, we have that

$$\sup_{\sigma\in\Sigma}\inf_{\pi\in\Pi}\mathbb{E}(avg\ sup_v^{\sigma,\pi}) = \inf_{\pi\in\Pi}\sup_{\sigma\in\Sigma}\mathbb{E}(avg\ inf_v^{\sigma,\pi}) \tag{5.2}$$

and there are MD strategies $\hat{\sigma} \in \Sigma$ and $\hat{\pi} \in \Pi$ such that $\inf_{\pi\in\Pi}\mathbb{E}(avg\ sup_v^{\hat{\sigma},\pi})$ and $\sup_{\sigma\in\Sigma}\mathbb{E}(avg\ inf_v^{\sigma,\hat{\pi}})$ are equal to the value defined by Equality (5.2) (see Gillette [1957], Liggett and Lippman [1969]).

Note that Equality (5.2) does not follow from Equality (5.1) and is *invalid* for infinite-state games. To see this, consider an arbitrary sequence $(a_i)_{i=0}^{\infty}$ such that $a_i \in \{0,1\}$ and

$$A_{\text{inf}} = \liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} a_i < \limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} a_i = A_{\text{sup}}.$$

This sequence can be encoded as a game with countably many vertices $v_0, v_1, \ldots$ where $v_i \mapsto v_{i+1}$ and $r(v_i) = a_i$ for all $i \geq 0$. Obviously, for all $(\sigma,\pi) \in \Sigma\times\Pi$ we have that

$$\mathbb{E}(avg\ inf_{v_0}^{\sigma,\pi}) = A_{\text{inf}} < A_{\text{sup}} = \mathbb{E}(avg\ sup_{v_0}^{\sigma,\pi})$$

which means that Equality (5.2) does not hold. Also observe that $val_{avg\ inf}(v_0) \neq val_{avg\ sup}(v_0)$.

- ***Discounted payoff*** assigns to each $w \in Run(G)$ the sum of discounted rewards

$$\sum_{i=0}^{\infty} d^i \cdot r(w(i))$$

where $0 < d < 1$ is a *discount factor*. Discounting formally captures the natural idea that the far-away future is not as important as the near future.

Observe that the above series converges absolutely which is mathematically convenient.

- ***Weighted reachability*** payoff assigns to every run w either 0 (if w does not visit a target vertex) or the reward of the first target vertex visited by w. Here, it is usually assumed that r is positive. One can also consider a ***discounted reachability*** payoff which is a variant of discounted payoff where $r(v)$ is either 1 or 0 depending on whether v is a target vertex or not.
- ***lim-max*** and ***lim-min*** payoffs, which assigns to each $w \in Run(G)$ either the maximal or the minimal reward which appears infinitely often along w. More precisely, we assume that r takes only finitely many values, and define $\textit{lim-max}(w)$ and $\textit{lim-min}(w)$ as the max and min of the set $\{x \in \mathbb{R} \mid r(w(i)) = x \text{ for infinitely many } i \geq 0\}$, respectively.

The presented list of Borel measurable payoffs contains only selected conceptual representatives and it is surely not exhaustive.

The problems of interest

For a given class of turn-based stochastic games $\mathcal{G}$ and a given class of payoff functions F, we are interested in answering the following basic questions:

(1) *Do optimal strategies exist for all $G \in \mathcal{G}$ and $f \in F$?*
(2) *What is the type of optimal and ε-optimal strategies?*
(3) *Can we compute/approximate $val_f(v)$?*
(4) *Can we compute optimal and ε-optimal strategies?*

If the answer to Question (1) is negative, we also wonder whether the existence of an optimal strategy in a given vertex is decidable and what is the complexity of this problem.

Optimal strategies (if they exist) may require memory and/or randomisation. A full answer to Question (2) should provide an optimal upper bound on the size of the required memory. It can also happen that optimal strategies require memory *or* randomisation, but not necessarily both.

Question (3) can also be reformulated as a decision problem. For a given rational constant ϱ, we ask whether $val_f(v)$ is bounded by ϱ (from above or below). In particular, if f is a qualitative payoff, it is often important to know whether $val_f(v)$ is positive or equal to one, and this special *qualitative* variant of the problem tends to be computationally easier. If $\mathcal{G}$ is a class of infinite-state games, $val_f(v)$ can be irrational even if all transition probabilities are rational and f is reachability payoff (a simple example is presented in

Section 5.3). In such cases, all we can hope for is an efficient algorithm which approximates $val_f(v)$ up to an arbitrarily small given precision.

Question (4) also requires special attention in the case of infinite-state games, because then even MD strategies do not necessarily admit a finite description. Since the vertices of infinite-state games typically carry some algebraic structure, optimal strategies may depend just on some finite information gathered by analysing the structure of vertices visited along a run in a play.

The existing results

In this section we give a short summary of the existing results about turn-based stochastic games with Borel measurable payoffs. We start with *finite-state* games.

In general, optimal strategies in finite-state games with the Borel measurable payoff functions introduced in the previous sections exist, do not need to randomise, and are either memoryless or finite-memory. In some cases, memory can be 'traded' for randomness (see Chatterjee et al. [2004a]). The values are rational (assuming that transition probabilities in games are rational) and computable. The main techniques for establishing these results are the following:

- ***Strategy improvement.*** This technique was originally developed for general stochastic games (see Hoffman and Karp [1966]). An initial strategy for one of the players is successively improved by switching it at positions at which the current choices are not locally optimal.
- ***Value iteration.*** The tuple of all values is computed/approximated by iterating a suitable functional on some initial vector. For example, the Bellman functional Γ defined in Section 5.3.1 can be used to compute/approximate the values in games with reachability payoffs.
- ***Convex optimisations.*** These methods are particularly useful for MDPs (see Puterman [1994], Filar and Vrieze [1996]). For example, the tuple of values in MDPs with reachability payoffs is computable in polynomial time by a simple linear program (see below).
- ***Graph-theoretic methods.*** Typically, these methods are used to design efficient (polynomial-time) algorithms deciding whether the value of a given vertex is equal to one (see, e.g., Chatterjee et al. [1994]).

The individual payoff functions are discussed in greater detail below.

Reachability. Turn-based stochastic games with reachability payoffs were first considered by Condon [1992] where it was shown that the problem of whether $val(v) > \frac{1}{2}$ for a given vertex v is in $\mathbf{NP} \cap \mathbf{coNP}$. It was also observed that both players have optimal MD strategies. The algorithm proposed by Condon [1992] requires transformation of the original game into a ***stopping*** game where ***local optimality equations*** admit a unique solution (i.e., the functional Γ defined in Section 5.3.1 has a unique fixed-point) which is a tuple of rational numbers of linear size. The tuple can then be guessed and verified (which leads to the $\mathbf{NP} \cap \mathbf{coNP}$ upper bound), or computed by a quadratic program with linear constraints. This algorithm is exponential even if the number of random vertices is fixed. Randomised algorithms with sub-exponential expected running time were proposed by Halman [2007] and Ludwig [1995]. A deterministic algorithm for computing the values and optimal strategies with $\mathcal{O}(|V_\bigcirc|! \cdot (|V| \cdot |{\mapsto}| + |p|))$ running time, where $|p|$ is the maximum bit length of a transition probability, was recently proposed by Gimbert and Horn [2008]. This algorithm is *polynomial* for every fixed number of random vertices.

Let us note that the exact complexity of the problem of whether $val(v) > \frac{1}{2}$ remains unsettled, despite substantial effort of the community. Since the problem belongs to $\mathbf{NP} \cap \mathbf{coNP}$, it is not likely to be $\mathbf{NP}$ or $\mathbf{coNP}$ complete. At the same time, it is not known to be in $\mathbf{P}$. On the other hand, the qualitative variant of the problem (i.e., the question whether $val(v) = 1$) is solvable in polynomial time (this follows, e.g., from a more general result about Büchi objectives achieved by de Alfaro and Henzinger [2000]). For MDPs, the values and optimal strategies are computable in polynomial time (both in the maximising and minimising subcase) by linear programming. Given a MDP $G = (V, \mapsto, (V_\Box, V_\bigcirc), \mathit{Prob})$ where $V = \{v_1, \ldots, v_n\}$ and v_n is the only target vertex, the tuple of all $val(v_i)$ is computable by the following program (a correctness proof can be found in, e.g., Filar and Vrieze [1996]):

$$
\begin{array}{ll}
\multicolumn{2}{l}{\textbf{minimise } y_1 + \cdots + y_n} \\
\multicolumn{2}{l}{\text{subject to}} \\
\quad y_n = 1 & \\
\quad y_i \geq y_j & \text{for all } v_i \mapsto v_j \text{ where } v_i \in V_\Box \text{ and } i < n \\
\quad y_i = \sum_{v_i \stackrel{x}{\mapsto} v_j} x \cdot y_j & \text{for all } v_i \in V_\bigcirc,\ i < n \\
\quad y_i \geq 0 & \text{for all } i \leq n.
\end{array}
$$

An optimal maximising strategy can be constructed in polynomial time even naively (i.e., for each $v \in V_\Box$ we successively identify a transition $v \mapsto v'$ such that the tuple of values does not change when the other outgoing transitions of v are removed).

(co-)Büchi. In turn-based stochastic games with Büchi and co-Büchi payoffs, both players have optimal MD strategies, the problem of whether $val(v) \geq \varrho$ for a given rational $\varrho \in [0, 1]$ is in $\mathbf{NP} \cap \mathbf{coNP}$, and the problem

of whether $val(v) = 1$ is **P**-complete (see Chatterjee et al. [2004b], de Alfaro and Henzinger [2000]).

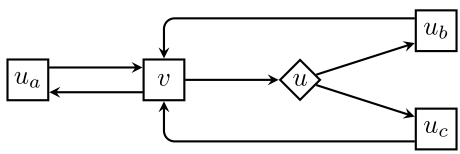
Rabin-chain (parity). In turn-based stochastic games with Rabin-chain (parity) payoffs, both players still have optimal MD strategies (see McIver and Morgan [2002], Chatterjee et al. [2004b]). The problem of whether $val(v) \geq \varrho$ for a given rational $\varrho \in [0, 1]$ is in **NP** $\cap$ **coNP**, and this is currently the best upper bound also for the problem of whether $val(v) = 1$ (see Chatterjee et al. [2004b]).

Rabin and Street. In turn-based stochastic games with Rabin payoffs, player $\Box$ has an optimal MD strategy (see Chatterjee et al. [2005]). This does not hold for player $\Diamond$, as demonstrated by the following simple example:

Consider the Rabin condition $\{(a, b), (b, a)\}$, where $\nu(u_a) = \{a\}$, $\nu(u_b) = \{b\}$, and $\nu(v) = \emptyset$. Obviously, $val(v) = 0$, but an optimal minimising strategy must ensure that both u_a and u_b are visited infinitely often, which is not achievable by a MD strategy.

Consequently, the problem of whether $val(v) \geq \varrho$ is in **NP** for Rabin payoffs. Since the problem of whether $val(v) = 1$ is **NP**-hard (see Emerson and Jutla [1988]), both problems are **NP**-complete. Since the Street acceptance condition is dual to Rabin, this also implies **coNP**-completeness of the two problems for Street payoffs.

Muller. In turn-based stochastic games with Muller payoffs, both players have optimal FD strategies, and the memory cannot be traded for randomness (i.e., the players do not necessarily have MR optimal strategies). To see this, consider the following game, where $\nu(u_a) = \{a\}$, $\nu(v) = \nu(u) = \emptyset$, $\nu(u_b) = \{b\}$, $\nu(u_c) = \{c\}$, and the Muller condition is $\{\{b\}, \{a, c\}, \{a, b, c\}\}$ (the example is taken from Chatterjee et al. [2004a]):



It is easy to check that $val(v) = 1$, and player $\Box$ has a FD optimal strategy which in every state wv selects either $v \mapsto u_a$ or $v \mapsto u$, depending on whether the last vertex of w is u_c or not, respectively. It is also easy to see that player $\Box$ does not have a MR optimal strategy.

For Muller payoffs, the problem of whether $val(v) \geq \varrho$ is

PSPACE-complete, and the same holds for the problem of whether $val(v) = 1$ (see Chatterjee [2007], Hunter and Dawar [2005]).

Mean-payoff and discounted payoff. In mean-payoff and discounted payoff turn-based stochastic games, both players have optimal MD strategies (see Gillette [1957], Liggett and Lippman [1969]), and the problem of whether $val(v) \geq \varrho$ is in $\mathbf{NP} \cap \mathbf{coNP}$. We refer to Filar and Vrieze [1996], Neyman and Sorin [2003] for more comprehensive expositions of algorithms for mean-payoff and discounted payoff turn-based stochastic games.

Basic properties of the other quantitative payoffs (in particular, *lim-min* and *lim-max* payoffs) are carefully discussed in Chatterjee et al. [2009].

Finally, we give a brief overview of the existing results about *infinite-state* turn-based stochastic games. There are basically three types of such games studied in the literature.

- ***Recursive stochastic games***, also known as ***stochastic BPA games***. These are games defined over stateless pushdown automata or (equivalently) 1-exit recursive state machines. Roughly speaking, a stochastic BPA game is a finite system of rules of the form $X \hookrightarrow \alpha$, where X is a stack symbol and α is a (possibly empty) sequence of stack symbols. The finitely many stack symbols are split into three disjoint subsets of symbols that 'belong' to player $\Box$, player $\Diamond$, or the virtual random player. A ***configuration*** is a finite sequence of stack symbols. The leftmost symbol of a given configuration is rewritten according to some rule, which is selected by the respective player (for every stochastic stack symbol Y, there is a fixed probability distribution over the rules of the form $Y \hookrightarrow \beta$).

  A ***termination*** objective is a special type of reachability objective where the only target vertex is ε (i.e., the empty stack). BPA MDPs and stochastic BPA games with termination objectives were studied by Etessami and Yannakakis [2005, 2006]. Some of these results were generalised to *reachability* objectives by Brázdil et al. [2008] and Brázdil et al. [2009a]. BPA MDPs and stochastic BPA games with *positive rewards* were studied by Etessami et al. [2008]. Here, each rule is assigned some fixed reward $r > 0$ which is collected whenever the rule is executed, and the objective of player $\Box$ is to maximise the expected total reward (which can be infinite).
- ***Stochastic games with lossy channels.*** A lossy channel system is a finite-state automaton equipped with a finite number of unbounded but unreliable (lossy) channels. A transition may change a control state and read/write from/to a channel. Since the channels are lossy, an arbitrary number of messages may be lost from the channels before and after each transition. A probabilistic variant of lossy channel systems defines a proba-

bilistic model for message losses. Usually, it is assumed that each individual message is lost independently with probability $\lambda > 0$ in every step. A stochastic game with lossy channels (SGLC) is obtained by splitting the control states into two disjoint subsets that are controlled by player $\Box$ and player $\Diamond$, respectively. However, message losses still occur randomly.

In Baier et al. [2006], it was shown that SGLC with qualitative reachability objectives are decidable. MDPs with lossy channels and various ω-regular objectives were examined by Baier et al. [2007]. SGLC with Büchi objectives were studied recently by Abdulla et al. [2008].

- ***One-counter stochastic games.*** These are stochastic games generated by one-counter machines, i.e., finite-state automata equipped with an unbounded counter which can store non-negative integers. The set of control states is split into three disjoint subsets controlled by player $\Box$, player $\Diamond$, or the random player, who are responsible for selecting one of the available transitions. One-counter MDPs with various reachability objectives were recently studied by Brázdil et al. [2010].

5.2.2 Win–lose games

Another important class of zero-sum stochastic turned-based games are *win–lose games* where the objective of player $\Box$ is to satisfy some property which is either valid or invalid for every play of G (the associated yield assigns 1 to the plays where the property is valid, and 0 to the other plays). An important subclass of such properties are ***temporal objectives*** that can be encoded as formulae of suitable temporal logics.

Let φ be a formula which is either valid or invalid in every state of every play of G. We say that a strategy $\sigma \in \Sigma$ is ***φ-winning in** v* if for every strategy $\pi \in \Pi$ we have that the state v of $G_v^{(\sigma,\pi)}$ satisfies φ. Similarly, a strategy $\pi \in \Pi$ is *$\neg\varphi$-winning in v* if for every $\sigma \in \Sigma$ we have that the state v of $G_v^{(\sigma,\pi)}$ does *not* satisfy φ. We say that G (with φ) is ***determined*** if for every $v \in V$ either player $\Box$ has a φ-winning strategy in v or player $\Diamond$ has a $\neg\varphi$-winning strategy in v.

Temporal logics can be classified as ***linear-time*** or *branching-time*, depending on whether they 'ignore' the branching structure of transition systems or not, respectively (see, e.g., Emerson [1991]). The syntax of these logics is built upon a countable set $Ap = \{a, b, c, \ldots\}$ of *atomic propositions*. A *valuation* is a function $\nu : V \to 2^{Ap}$ which to every vertex v assigns the set $\nu(v) \subseteq Ap$ of all atomic propositions that are valid in v. Note that ν can be naturally extended to the states of a play of G (for every state $wv \in V^*V$

of a play $G_\mu^{(\sigma,\pi)}$ we put $\nu(wv) = \nu(v)$). For the rest of this section, we fix a valuation ν.

Linear-time logics

Linear-time logics specify properties of runs in transition systems. For a given linear-time formula ψ, the associated temporal property is specified by a constraint on the probability of all runs that satisfy ψ. This constraint is written as a ***probabilistic operator*** $\mathbf{P}^{\succ\varrho}$, where $\succ \in \{>, \geq\}$ and $\varrho \in [0,1]$ is a rational constant. Thus, we obtain a ***linear-time objective*** $\mathbf{P}^{\succ\varrho}\psi$ whose intuitive meaning is 'the probability of all runs satisfying ψ is $\succ$-related to ϱ'. An important subclass of linear-time objectives are ***qualitative linear-time objectives*** where the constant ϱ is either 0 or 1.

An example of a widely used linear-time logic is LTL, introduced in Pnueli [1977]. The syntax of LTL formulae is specified by the following abstract syntax equation:

$$\psi \quad ::= \quad \mathbf{tt} \quad | \quad a \quad | \quad \neg\psi \quad | \quad \psi_1 \wedge \psi_2 \quad | \quad \mathbf{X}\psi \quad | \quad \psi_1 \mathbf{U} \psi_2$$

Here a ranges over Ap. Note that the set $Ap(\psi)$ of all atomic propositions that appear in a given LTL formula ψ is finite. Every LTL formula ψ determines its associated *ω-language* L_ψ consisting of all infinite words u over the alphabet $2^{Ap(\psi)}$ such that $u \models \psi$, where the relation $\models$ is defined inductively as follows (recall that the symbol u_i, where $i \geq 0$, denotes the infinite word $u(i), u(i+1), \ldots$):

$$
\begin{array}{lll}
u \models \mathbf{tt} & & \\
u \models a & \text{iff} & a \in u(0) \\
u \models \neg\psi & \text{iff} & u \not\models \psi \\
u \models \psi_1 \wedge \psi_2 & \text{iff} & u \models \psi_1 \text{ and } u \models \psi_2 \\
u \models \mathbf{X}\psi & \text{iff} & u_1 \models \psi \\
u \models \psi_1 \mathbf{U} \psi_2 & \text{iff} & u_j \models \psi_2 \text{ for some } j \geq 0 \text{ and } u_i \models \psi_1 \text{ for all } 0 \leq i < j.
\end{array}
$$

For a given run w of G (or a play of G), we put $w \models^\nu \psi$ iff $\hat{w} \models \psi$, where $\hat{w}$ is the infinite word defined by $\hat{w}(i) = \nu(w(i)) \cap Ap(\psi)$. In the following we also use $\mathbf{F}\psi$ and $\mathbf{G}\psi$ as abbreviations for $\mathbf{tt}\mathbf{U}\psi$ and $\neg\mathbf{F}\neg\psi$, respectively.

Another important formalism for specifying properties of runs in transition systems are finite-state automata over infinite words with various acceptance criteria, such as Büchi, Rabin, Street, Muller, etc. We refer to Thomas [1991] for a more detailed overview of the results about finite-state automata over infinite words. Let M be such an automaton with an input alphabet $2^{Ap(M)}$, where $Ap(M)$ is a finite subset of Ap. Then M can also be understood as a 'formula' interpreted over the runs of G (or a play of G) by stipulating that $w \models^{\nu} M$ iff $\hat{w}$ is accepted by M, where $\hat{w}$ is the infinite word defined by $\hat{w}(i) = \nu(w(i)) \cap Ap(M)$. Let us note that every LTL formula ψ can be effectively translated into an equivalent finite-state automaton M_ψ which accepts the language L_ψ. If the acceptance condition is Rabin-chain (or more powerful), the automaton M_ψ can be effectively transformed into an equivalent *deterministic* automaton with the same acceptance condition. In general, the cost of translating ψ into M_ψ is at least exponential. On the other hand, there are properties expressible by finite-state automata that cannot be encoded as LTL formulae. We refer to Thomas [1991] for more details.

There is a close connection between linear-time objectives and the ω-regular payoffs introduced in the previous section. Since this connection has several subtle aspects that often lead to confusion, it is worth an explicit discussion. Let $M = (Q, 2^{Ap(M)}, \rightarrow, q_0, Acc)$ be a *deterministic* finite-state automaton, where Q is a finite set of control states, $2^{Ap(M)}$ is the input alphabet, $\rightarrow \subseteq Q \times 2^{Ap(M)} \times Q$ is a total transition function, and Acc some acceptance criterion. Then we can construct a synchronous product of the game G and M, which is a stochastic game $G \times M$ where $V \times Q$ is the set of vertices partitioned into $(V_\Box \times Q, V_\Diamond \times Q, V_\bigcirc \times Q)$ and $(v, q) \mapsto (v', q')$ iff $v \mapsto v'$ and $q \xrightarrow{\mathcal{A}} q'$ where $\mathcal{A} = \nu(q) \cap Ap(M)$. Since M is deterministic and the transition function of M is total, the probability assignment is just inherited from G (i.e., $(v, q) \overset{x}{\mapsto} (v', q')$ only if $v \overset{x}{\mapsto} v'$). Further, the acceptance criterion of M is translated into the corresponding ω-regular payoff over the runs of $G \times M$ in the natural way. Note that $G \times M$ is constructed so that M just observes the runs of G and the constructed ω-regular payoff just reflects the accepting/rejecting verdict of this observation. Thus, we can reduce the questions about the existence and effective constructibility of a $(\mathbf{P}^{\succ \varrho} M)$-winning strategy for player $\Box$ in a vertex v of G to the questions about the value and effective constructibility of an optimal maximising strategy in the vertex (v, q_0) of $G \times M$. However, this reduction does not always work completely smoothly, particularly for infinite-state games. Some of the reasons are mentioned below.

- For infinite-state games, the product $G \times M$ is not necessarily definable in the same formalism as the original game G. Fortunately, most of the studied formalisms correspond to abstract computational devices equipped with a finite-state control, which can also encode the structure of M. However, this does not necessarily work if the finite-state control is trivial (i.e., it has just one or a fixed number of control states) or if it is required to satisfy some special conditions.

- For infinite-state games, the reduction to ω-regular payoffs described above can be problematic also because optimal maximising/minimising strategies in infinite-state games with ω-regular payoffs (even reachability payoffs) do not necessarily exist. For example, even if we somehow check that the value of (v, q_0) in $G \times M$ is 1, this does yet mean that player $\Box$ has a $(\mathbf{P}^{=1}M)$-winning strategy in v.
- For finite-state games, the two problems discussed above usually disappear. However, there are still some issues related to complexity. In particular, the results about the type of optimal strategies in $G \times M$ do not carry over to G. For example, assume that we are given a linear-time objective $\mathbf{P}^{=1}M$ where M is a deterministic Rabin-chain automaton. If G has finitely many states, then $G \times M$ is also finite-state and hence we can rely on the results presented by McIver and Morgan [2002] and Chatterjee et al. [2004b] and conclude that the value of (v, q_0) is computable in time polynomial in the size of $G \times M$ and there is an optimal maximising MD strategy σ computable in polynomial time. From this we can deduce that the existence of a $(\mathbf{P}^{=1}M)$-winning strategy for player $\Box$ in v is decidable in polynomial time. However, since the optimal MD strategy σ may depend *both* on the current vertex of G and the current state of M, we cannot conclude that if player $\Box$ has some $(\mathbf{P}^{=1}M)$-winning strategy in v, then he also has an MD $(\mathbf{P}^{=1}M)$-winning strategy in v (still, the strategy σ can be translated into a FD $(\mathbf{P}^{=1}M)$-winning strategy which simulates the execution of M on the history of a play).

To sum up, linear-time objectives *are* closely related to ω-regular payoffs, but the associated problems cannot be seen as 'equivalent' in general.

Branching-time logics

Branching-time logics such as CTL, CTL*, or ECTL* (see, e.g., Emerson [1991]) allow explicit existential/universal quantification over runs. Thus, one can express that a given *path formula* holds for some/all runs initiated in a given state.

In the probabilistic setting, the existential/universal path quantifiers are replaced with the *probabilistic operator* $\mathbf{P}^{\succ\varrho}$ introduced in the previous section. In this way, every (non-probabilistic) branching-time logic determines its probabilistic counterpart. The probabilistic variants of CTL, CTL*, and ECTL* are denoted by PCTL, PCTL*, and PECTL*, respectively (see Hansson and Jonsson [1994]).

The syntax of PCTL* *path* and *state* formulae is defined by the following equations:

$$\begin{array}{lllllllllll} \psi & ::= & \varphi & \mid & \neg\psi & \mid & \psi_1 \wedge \psi_2 & \mid & \mathbf{X}\psi & \mid & \psi_1 \mathbf{U} \psi_2 \\ \varphi & ::= & \mathbf{tt} & \mid & a & \mid & \neg\varphi & \mid & \varphi_1 \wedge \varphi_2 & \mid & \mathbf{P}^{\succ\varrho}\psi \end{array}$$

Note that all LTL formulae are PCTL* path formulae. In the case of PCTL, the syntax of path formulae is restricted to $\psi ::= \varphi \mid \mathbf{X}\psi \mid \psi_1\mathbf{U}\psi_2$. Since the expressive power of LTL is strictly smaller than that of finite-state automata over infinite words (see the previous section), the logic CTL* can be further enriched by allowing arbitrary *automata connectives* in path formulae (see Wolper [1981]). The resulting logic is known as *extended CTL* (ECTL*)*, and its probabilistic variant as PECTL*.

Let $G_\mu^{(\sigma,\pi)}$ be a play of G. For every run $w \in Run(G_\mu^{(\sigma,\pi)})$ and every path formula ψ we define the relation $w \models^\nu \psi$ in the same way as in the previous section, where $w \models^\nu \varphi$ iff $w(0) \models^\nu \varphi$ (see below). For every state s of $G_\mu^{(\sigma,\pi)}$ and every state formula φ we define the relation $s \models^\nu \varphi$ inductively as follows:

$$\begin{array}{ll} s \models^\nu \mathbf{tt} & \\ s \models^\nu a & \text{iff } \ a \in \nu(s) \\ s \models^\nu \neg\varphi & \text{iff } \ s \not\models \varphi \\ s \models^\nu \varphi_1 \wedge \varphi_2 & \text{iff } \ s \models^\nu \varphi_1 \text{ and } s \models^\nu \varphi_2 \\ s \models^\nu \mathbf{P}^{\succ\varrho}\psi & \text{iff } \ \mathcal{P}_s(\{w \in Run(s) \mid w \models^\nu \psi\}) \succ \varrho \end{array}$$

A state formula φ is *qualitative* if each occurrence of the probabilistic operator in φ is of the form $\mathbf{P}^{\succ 0}$ or $\mathbf{P}^{\succ 1}$. General (not necessarily qualitative) state formulae are also called *quantitative*.

A ***branching-time objective*** is a state formula φ of a branching-time probabilistic temporal logic. Important subclasses of branching-time objectives are PCTL, PCTL*, and PECTL* objectives.

Let us note that state formulae of branching-time probabilistic logics are sometimes interpreted directly on vertices of Markov decision processes (see Bianco and de Alfaro [1995]) and stochastic games (see de Alfaro and Majumdar [2004]). Path formulae are interpreted over the runs of G in the same way as above, and all state formulae except for $\mathbf{P}^{\succ\varrho}\psi$ are also interpreted in the natural way. Due to the presence of non-determinism in G, it is not possible to measure the probability of runs in G, and the probabilistic operator $\mathbf{P}^{\succ\varrho}$ has a different meaning defined as follows:

$$v \models^\nu \mathbf{P}^{\succ\varrho}\psi \quad \text{iff} \quad \exists\sigma \in \Sigma \;\; \forall\pi \in \Pi : \mathcal{P}_v(\{w \in Run(G_v^{(\sigma,\pi)}, v) \mid w \models^\nu \psi\}) \succ \varrho.$$

If G is a Markov decision process, then $v \models^\nu \mathbf{P}^{\succ\varrho}\psi$ iff for every strategy τ

of the only player we have that

$$\mathcal{P}_v(\{w \in Run(G_v^{(\tau)}, v) \mid w \models^\nu \psi\}) \succ \varrho. \tag{5.3}$$

For finite-state MDPs, condition (5.3) is equivalent to

$$\inf_\tau\{\mathcal{P}_v(\{w \in Run(G_v^{(\tau)}, v) \mid w \models^\nu \psi\})\} \succ \varrho \tag{5.4}$$

which is exactly the semantics proposed by Bianco and de Alfaro [1995]. For general (infinite-state) MDPs, conditions (5.3) and (5.4) are not equivalent.

At first glance, one might be tempted to think that $v \models^\nu \psi$ iff player $\Box$ has a φ-winning strategy in v. A straightforward induction on the structure of φ reveals that '$\Rightarrow$' holds, but the opposite direction is *invalid*. To see this, consider the formula $\varphi \equiv (\mathbf{P}^{>0}\mathbf{F}a) \vee (\mathbf{P}^{>0}\mathbf{F}b)$ and the vertex v of the following game, where $\nu(u_a) = \{a\}$ and $\nu(u_b) = \{b\}$:

Intuitively, the formula φ says 'a state satisfying a is reachable, or a state satisfying b is reachable'. Note that player $\Box$ has a φ-winning strategy in v (in fact, every strategy of player $\Box$ is φ-winning). However, $v \not\models^\nu \mathbf{P}^{>0}\mathbf{F}a$, because player $\Diamond$ has a strategy which makes the vertex u_a unreachable. Similarly, $v \not\models^\nu \mathbf{P}^{>0}\mathbf{F}b$, and hence $v \not\models^\nu \varphi$. This means that the *model-checking* problem for stochastic games and formulae of probabilistic branching-time logics (i.e., the question of whether $v \models^\nu \varphi$) is *different* from the problem of deciding the winner in stochastic games with branching-time objectives. As we shall see, this difference is substantial.

The problems of interest

Let $\mathcal{G}$ be a class of turn-based stochastic games and Φ a class of temporal objectives. The most important questions about the games of $\mathcal{G}$ and the objectives of Φ include the following:

(1) *Are all games of $\mathcal{G}$ determined for all objectives of Φ?*
(2) *What is the type of winning strategies if they exist?*
(3) *Who wins in a given vertex?*
(4) *Can we compute winning strategies?*

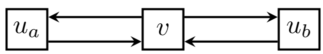
As we shall see in Section 5.3.2, stochastic games with temporal objectives are *not* necessarily determined (even for linear-time objectives). This means that 'nobody' is an eligible answer to Question (3). Since randomisation and memory can help the players to win, Question (3) can be refined into *'Does player $\Box$ (or player $\Diamond$) have a winning strategy of type XY in a given vertex*

v?' This problem can be decidable even if the existence of *some (i.e., HR)* winning strategy in v is undecidable.

The existing results

Finite-state stochastic games with linear-time objectives are rarely studied explicitly because most of the results can be deduced from the corresponding results about stochastic games with ω-regular payoffs (see Section 5.2.1). For infinite-state stochastic games, the relationship between linear-time objectives and ω-regular payoffs is more subtle. For example, even for reachability payoffs, the question of whether $val(v) = 1$ is *not* the same as the question of whether player $\Box$ has a $\mathbf{P}^{=1}\mathbf{F}t$-winning strategy in v, where the atomic proposition t is satisfied exactly in the target vertices (the details are given in Section 5.3).

Finite-state turn-based stochastic games with branching-time objectives were first considered by Baier et al. [2004], where it was shown that winning strategies for PCTL objectives may require memory and/or randomisation. Consider the following game, where $\nu(u_a) = \{a\}$, $\nu(u_b) = \{b\}$, and $\nu(v) = \emptyset$.



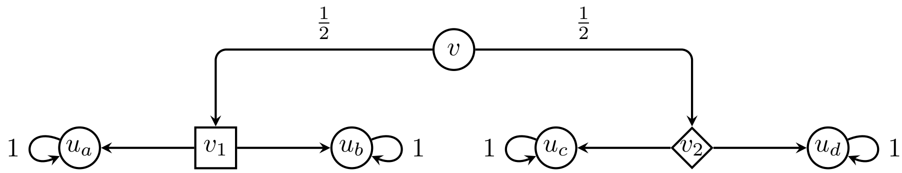
Let

- $\varphi_1 \equiv (\mathbf{P}^{=1}\mathbf{X}a) \wedge (\mathbf{P}^{=1}\mathbf{F}b)$
- $\varphi_2 \equiv (\mathbf{P}^{>0}\mathbf{X}a) \wedge (\mathbf{P}^{>0}\mathbf{X}b)$
- $\varphi_3 \equiv (\mathbf{P}^{>0}\mathbf{X}a) \wedge (\mathbf{P}^{>0}\mathbf{X}b) \wedge (\mathbf{P}^{=1}\mathbf{F}(\mathbf{P}^{=1}\mathbf{G}a))$.

Obviously, player $\Box$ has a φ_i-winning strategy for every $i \in \{1, 2, 3\}$, and each such strategy must inevitably use memory for φ_1, randomisation for φ_2, and both memory and randomisation for φ_3.

In Baier et al. [2004], it was also shown that for PCTL objectives, the problem of whether player $\Box$ has a MD φ-winning strategy in a given vertex of a given finite-state MDP is **NP**-complete. MR strategies for PCTL objectives were considered by Kučera and Stražovský [2008], where it was shown that the existence of a φ-winning MR strategy in a given vertex is in **EXPTIME** for finite-state turn-based stochastic games, and in **PSPACE** for finite-state MDPs.

In Brázdil et al. [2006], it was noted that turn-based stochastic games with PCTL objectives are *not* determined (for any strategy type). To see this, consider the following game, where $\nu(u_a) = \{a\}$, $\nu(u_b) = \{b\}$, $\nu(u_c) = \{c\}$, $\nu(u_d) = \{d\}$, and $\nu(v) = \nu(v_1) = \nu(v_2) = \emptyset$.

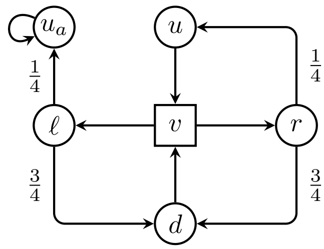


Let

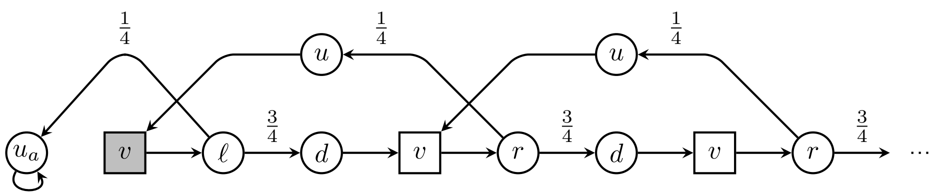
$$\varphi \equiv \big(\mathbf{P}^{=1}\mathbf{F}(a \vee c)\big) \ \vee \ \big(\mathbf{P}^{=1}\mathbf{F}(b \vee d)\big) \ \vee \ \big((\mathbf{P}^{>0}\mathbf{F}c) \wedge (\mathbf{P}^{>0}\mathbf{F}d)\big)\,.$$

Assume that player $\Box$ has a φ-winning strategy σ in v. The strategy σ cannot randomise at v_1, because then both of the subformulae $\mathbf{P}^{=1}\mathbf{F}(a \vee c)$ and $\mathbf{P}^{=1}\mathbf{F}(b \vee d)$ become invalid in v, and player $\Diamond$ can always falsify the subformula $(\mathbf{P}^{>0}\mathbf{F}c) \wedge (\mathbf{P}^{>0}\mathbf{F}d)$. Hence, the strategy σ must choose one of the transitions $v_1 \mapsto u_a$, $v_1 \mapsto u_b$ with probability one. For each of these choices, player $\Diamond$ can falsify the formula φ, which means that σ is not φ-winning. Similarly, one can show that player $\Diamond$ does not have a $\neg\varphi$-winning strategy in v (in particular, note that player $\Diamond$ cannot randomise at v_2, because this would make the subformula $(\mathbf{P}^{>0}\mathbf{F}c) \wedge (\mathbf{P}^{>0}\mathbf{F}d)$ valid).

In Brázdil et al. [2006], it was also shown that for PCTL objectives, the existence of a φ-winning MD strategy for player $\Box$ in a given vertex of a finite-state stochastic turn-based game is $\Sigma_2 = \mathbf{NP}^{\mathbf{NP}}$-complete, which complements the aforementioned result for MDPs. Further, it was shown that the existence of a φ-winning HR (or HD) strategy in a given vertex of a finite-state MDP is *highly undecidable* (i.e., beyond the arithmetical hierarchy). The proof works even for a fixed *quantitative* PCTL formula ξ. The use of a non-qualitative probability constraint in ξ is in fact *unavoidable* –as it was shown later by Brázdil et al. [2008], the existence of a φ-winning HR (or HD) strategy in finite-state MDPs with *qualitative* PCTL and PECTL* objectives is **EXPTIME**-complete and **2-EXPTIME**-complete, respectively. It is worth noting that these algorithms are actually *polynomial* for every fixed qualitative PCTL or PECTL* formula. A HR (or HD) φ-winning strategy for player $\Box$ may require infinite memory, but it can always be implemented by an effectively constructible *one-counter automaton* which reads the history of a play. To get some intuition, consider the following game, where $\nu(u_a) = \{a\}$ and $\nu(y) = \emptyset$ for all vertices y different from u_a.



Let $\varphi \equiv \mathbf{P}^{>0}\mathbf{G}(\neg a \wedge (\mathbf{P}^{>0}\mathbf{F}a))$, and let σ be a HD strategy which in every wv selects either $v \mapsto \ell$ or $v \mapsto r$, depending on whether $\#_d(w) - \#_u(w) \leq 0$ or not, respectively. Here $\#_d(w)$ and $\#_u(w)$ denote the number of occurrences of d and u in w, respectively. Obviously, the strategy σ can be implemented by a one-counter automaton. The play $G_v^{(\sigma)}$ initiated in v closely resembles a one-way infinite random walk where the probability of going right is $\frac{3}{4}$ and the probability of going left is $\frac{1}{4}$. More precisely, the play $G_v^{(\sigma)}$ corresponds to the unfolding of the following infinite-state Markov chain (the initial state is grey):



A standard calculation shows that the probability of all $w \in \mathit{Run}(G_v^{(\sigma)})$ initiated in v such that w visits a state satisfying a is equal to $\frac{1}{3}$. Note that for every $w \in \mathit{Run}(G_v^{(\sigma)})$ initiated in v which does *not* visit a state satisfying a we have that $w(i) \models^\nu \neg a \wedge (\mathbf{P}^{>0}\mathbf{F}a)$ for every $i \geq 0$. Since the probability of all such runs is $\frac{2}{3}$, we obtain that the formula φ is valid in the state v of $G_v^{(\sigma)}$. On the other hand, there is no finite-memory φ-winning strategy $\hat{\sigma}$ in v, because then the play $G_v^{(\hat{\sigma})}$ corresponds to an unfolding of a finite-state Markov chain, and the formula φ does not have a finite-state model (see, e.g., Brázdil et al. [2008]).

The memory requirements of φ-winning strategies for various fragments of qualitative branching-time logics were analysed by Brázdil and Forejt [2007] and Forejt [2009]. The decidability/complexity of the existence of HR (or HD) φ-winning strategies in turn-based stochastic games with qualitative branching-time objectives is still open.

5.3 Reachability objectives in games with finitely and infinitely many vertices

As we have already mentioned, the properties of stochastic games with finitely and infinitely many vertices are different in many respects. To illustrate this, we examine the properties of turn-based stochastic games with *reachability* objectives in greater detail. Most of the negative results presented in this section are valid also for the other objectives introduced in Section 5.2.

For the rest of this section, we fix a turn-based stochastic game $G = (V, \mapsto, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ and a set $T \subseteq V$ of ***target*** vertices. The set of all $w \in Run(G)$ which visit a vertex of T is denoted by $Reach(T)$. Further, for every pair of strategies $(\sigma, \pi) \in \Sigma \times \Pi$ and every initial vertex v, we use $\mathcal{P}_v^{(\sigma,\pi)}(Reach(T))$ to denote the probability $\mathcal{P}_v(\{w \in Run(G_v^{(\sigma,\pi)}) \mid w_G \in Reach(T)\})$.

5.3.1 The existence of a value revisited

Recall that every vertex v of G has a value $val(v)$ defined by Equality 5.1 which now takes the following simple form:

$$\sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathcal{P}_v^{(\sigma,\pi)}(Reach(T)) = \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \mathcal{P}_v^{(\sigma,\pi)}(Reach(T)). \tag{5.5}$$

A direct proof of Equality (5.5) is actually simple and instructive. Consider the following (Bellman) functional $\Gamma : [0,1]^{|V|} \to [0,1]^{|V|}$ defined as follows:

$$\Gamma(\alpha)(v) = \begin{cases} 1 & \text{if } v \in T; \\ \sup\,\{\alpha(v') \mid v \mapsto v'\} & \text{if } v \notin T \text{ and } v \in V_\Box; \\ \inf\,\{\alpha(v') \mid v \mapsto v'\} & \text{if } v \notin T \text{ and } v \in V_\Diamond; \\ \sum_{v \stackrel{x}{\mapsto} v'} x \cdot \alpha(v') & \text{if } v \notin T \text{ and } v \in V_\bigcirc. \end{cases}$$

Since Γ is a monotonic function over a complete lattice $([0,1]^{|V|}, \sqsubseteq)$, where $\sqsubseteq$ is a component-wise ordering, we can apply the Knaster–Tarski Theorem (see Tarski [1955]) and conclude that Γ has the least fixed-point $\mu\Gamma$. Observe that for every $v \in V$ we have that

$$\mu\Gamma(v) \le \sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathcal{P}_v^{(\sigma,\pi)}(Reach(T)) \le \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \mathcal{P}_v^{(\sigma,\pi)}(Reach(T)).$$

The second inequality follows directly from definitions and it is actually valid for *arbitrary* Borel measurable payoffs. The first inequality is obtained by demonstrating that the tuple of all $\sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathcal{P}_v^{(\sigma,\pi)}(Reach(T))$ is a fixed-point of Γ, which is also straightforward. So, it remains to show that the inequality

$$\mu\Gamma(v) \leq \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \mathcal{P}_v^{(\sigma,\pi)}(Reach(T)) \tag{5.6}$$

cannot be strict.

Let us first assume that every vertex $u \in V_\Diamond$ has a ***locally optimal*** outgoing transition $u \overset{a}{\mapsto} u'$ where $\mu\Gamma(u) = \mu\Gamma(u')$ (in particular, note that if u has finitely many outgoing transitions, some of them must be locally optimal because $\mu\Gamma$ is a fixed-point of Γ). Now consider a MD strategy $\hat{\pi}$ which in every $u \in V_\Diamond$ selects some (fixed) locally optimal outgoing transition of u. One can easily show that $\mu\Gamma(v) = \sup_{\sigma\in\Sigma} \mathcal{P}_v^{(\sigma,\hat{\pi})}(Reach(T))$ for every $v \in V$, which implies that Inequality (5.6) is an equality. Further, observe that $\hat{\pi}$ is an optimal minimising strategy in every vertex of G. Thus, we obtain the following:

Proposition 5.4 *If every $u \in V_\Diamond$ has a locally optimal outgoing transition, then there is a MD strategy of player $\Diamond$ which is optimal minimising in every vertex of G.*

In the general case when the vertices of $V_\Diamond$ do not necessarily have locally optimal outgoing transitions (this can of course happen only if G is infinitely branching), Inequality (5.6) is proven as follows. We show that for every $\varepsilon > 0$ and every $v \in V$ there is a HD strategy $\hat{\pi}_\varepsilon \in \Pi$ such that

$$\sup_{\sigma\in\Sigma} \mathcal{P}_v^{(\sigma,\hat{\pi}_\varepsilon)}(Reach(T)) \leq \mu\Gamma(v) + \varepsilon.$$

This implies that Inequality (5.6) cannot be strict. Intuitively, in a given state $wu \in V^*V_\Diamond$, the strategy $\hat{\pi}_\varepsilon$ selects a transition $u \mapsto u'$ whose ***error*** $\mu\Gamma(u) - \mu\Gamma(u')$ is 'sufficiently small'. Observe that the error can be made arbitrarily small because $\mu\Gamma$ is a fixed point of Γ. The strategy $\hat{\pi}_\varepsilon$ selects transitions with progressively smaller and smaller error so that the 'total error' $\mathcal{P}_v^{(\sigma,\hat{\pi}_\varepsilon)}(Reach(T)) - \mu\Gamma(v)$ stays bounded by ε no matter what player $\Box$ does. A detailed proof can be found in, e.g., Brázdil et al. [2009a].

If G is finitely branching, then Γ is not only monotonic but also ***continuous***, i.e., $\Gamma(\bigvee_{i=0}^{\infty} \vec{y}_i) = \bigvee_{i=0}^{\infty} \Gamma(\vec{y}_i)$ for every infinite non-decreasing chain $\vec{y}_1 \sqsubseteq \vec{y}_2 \sqsubseteq \vec{y}_3 \sqsubseteq \cdots$ in $([0,1]^{|V|}, \sqsubseteq)$. By the Kleene fixed-point theorem, we have that $\mu\Gamma = \bigvee_{i=0}^{\infty} \Gamma^i(\vec{0})$, where $\vec{0}$ is the vector of zeros. For every $n \geq 1$, let $Reach^n(G)$ be the set of all $w \in Run(G)$ such that $w(i) \in T$ for some $0 \leq i < n$. A straightforward induction on n reveals that

$$\Gamma^n(\vec{0})(v) \;=\; \sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathcal{P}_v^{(\sigma,\pi)}(Reach^n(T)) \;=\; \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \mathcal{P}_v^{(\sigma,\pi)}(Reach^n(T)).$$

Further, for every $n \geq 1$ we define HD strategies $\sigma_n \in \Sigma$ and $\pi_n \in \Pi$ as follows:

- The strategies σ_1 and π_1 are defined arbitrarily.
- For all $n \geq 2$ and $wv \in V^* V_\Box$ such that $len(wv) < n$, the strategy σ_n selects a transition $v \mapsto v'$ such that $\Gamma^k(\vec{0})(v') = \max\{\Gamma^k(\vec{0})(v'') \mid v \mapsto v''\}$ where $k = n - len(wv)$.
- For all $n \geq 2$ and $wv \in V^* V_\Diamond$ such that $len(wv) < n$, the strategy π_n selects a transition $v \mapsto v'$ such that $\Gamma^k(\vec{0})(v') = \min\{\Gamma^k(\vec{0})(v'') \mid v \mapsto v''\}$ where $k = n - len(wv)$.

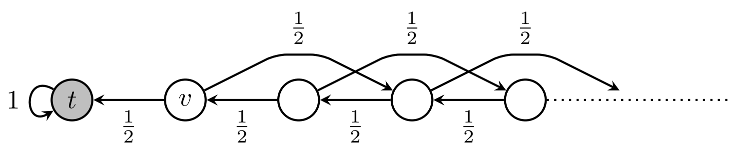
It is easy to prove that for every $n \geq 1$

$$\Gamma^n(\vec{0})(v) = \inf_{\pi \in \Pi} \mathcal{P}_v^{(\sigma_n, \pi)}(Reach^n(T)) = \sup_{\sigma \in \Sigma} \mathcal{P}_v^{(\sigma, \pi_n)}(Reach^n(T)).$$

A direct corollary to these observations is the following:

Proposition 5.5 *If G is finitely branching, then for all $v \in V$ and $\varepsilon > 0$ there are $n \geq 0$ and a HD strategy $\hat{\sigma} \in \Sigma$ such that $\mathcal{P}_v^{(\hat{\sigma}, \pi)}(Reach^n(T)) \geq val(v) - \varepsilon$ for every $\pi \in \Pi$.*

Finally, let us note that the values in infinite-state games can be irrational, even if all transition probabilities are equal to $\frac{1}{2}$. To see this, consider the following Markov chain $\mathcal{M}$, where t is the only target vertex.



Obviously, $val(v)$ is equal to the probability of all $w \in Run(v)$ which visit t, where μ_v is the initial probability distribution. By inspecting the structure of $\mathcal{M}$, it is easy to see that $val(v)$ has to satisfy the equation $x = \frac{1}{2} + \frac{1}{2}x^3$. Actually, $val(v)$ is the *least* solution of this equation in the interval $[0, 1]$, which is $\frac{\sqrt{5}-1}{2}$ (the 'golden ratio').

5.3.2 Optimal strategies and determinacy

In this section we classify the conditions under which optimal strategies exist, analyse the type of optimal strategies, and resolve the determinacy of games with $\mathbf{P}^{\succ \varrho}\mathbf{F}t$ objectives.

The properties of optimal minimising strategies are summarised in the next proposition.

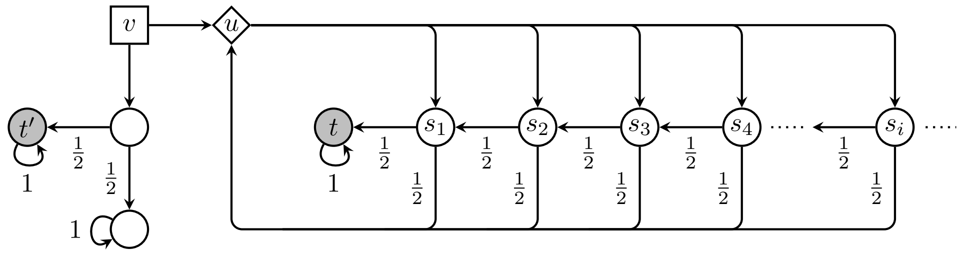


Figure 5.2 An optimal minimising strategy does not necessarily exist.

Proposition 5.6 *Let G be a stochastic game with a reachability objective associated to a set of target vertices T. Let v be a vertex of G. Then*

(a) an optimal minimising strategy in v does not necessarily exist; and if it exists, it may require infinite memory;

(b) an ε-optimal minimising strategy in v may require infinite memory for every fixed $\varepsilon \in (0, 1)$;

(c) if G is finitely branching, then there is a MD strategy which is optimal minimising in every vertex of G.

A counterexample for claims (a) and (b) is given in Figure 5.2, where t, t' are the only target vertices. Observe that $val(u) = 0$, but there is no optimal minimising strategy in u. Further, observe that for each fixed $\varepsilon \in (0, 1)$, every ε-optimal minimising strategy π in u must employ infinitely many transitions of the form $u \mapsto s_i$ (otherwise, the target vertex t would be inevitably reached with probability 1). Hence, π requires infinite memory. Finally, note that $val(v) = \frac{1}{2}$ and there is an optimal minimising strategy π' in v which requires infinite memory (the strategy π' must ensure that the probability of reaching a target vertex from u is at most $\frac{1}{2}$, because then player $\Box$ does not gain anything if she uses the transition $v \mapsto u$; hence, π' requires infinite memory). Claim (c) follows directly from Proposition 5.4.

The properties of optimal maximising strategies are remarkably different from those of optimal minimising strategies. The most notable (and perhaps somewhat surprising) difference is that an optimal maximising strategy may require *infinite memory*, even in finitely branching games.

Proposition 5.7 *Let G be a stochastic game with a reachability objective associated to a set of target vertices T. Let v be a vertex of G. Then*

(a) an optimal maximising strategy in v does not necessarily exist even if G is finitely branching;

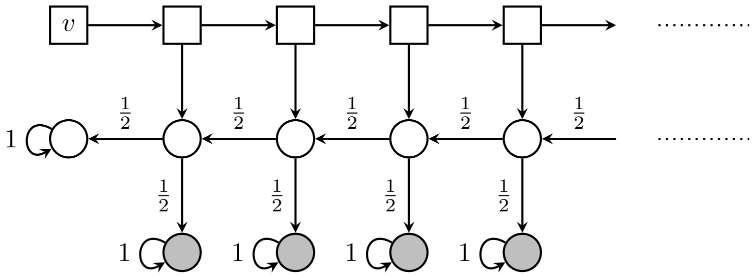


Figure 5.3 An optimal maximising strategy does not necessarily exist

(b) an optimal maximising strategy in v (if it exists) may require infinite memory even if G is finitely branching;

(c) if G has finitely many vertices, then there is a MD strategy which is optimal maximising in every vertex of G.

A counterexample for claim (a) is given in Figure 5.3 (target vertices are grey). Observe that $val(v) = 1$ but there is no optimal maximising strategy in v. An example demonstrating that an optimal maximising strategy may require infinite memory (originally due to Brožek [2009]) is given in Figure 5.4. The outgoing transitions of the vertex $\hat{v}$ are the same as the outgoing transitions of the vertex v in Figure 5.3 and they are not shown in the picture. Observe that $val(\hat{v}) = 1$ and hence $val(e_i) = 1$ for all $i \geq 1$. Further, we have that $val(s_i) = 1 - (\frac{1}{2})^i$ for all $i \geq 1$, and hence also $val(d_i) = 1 - (\frac{1}{2})^i$ for all $i \geq 1$. From this we get $val(v) = \frac{2}{3}$. Also observe that player $\Box$ has a HD optimal maximising strategy σ which simply ensures that player $\Diamond$ cannot gain anything by using transitions $d_i \mapsto e_i$. That is, whenever the vertex $\hat{v}$ is visited, the strategy σ finds a d_i stored in the history of a play, and starts to behave as an ε-optimal maximising strategy in $\hat{v}$, where $\varepsilon < (\frac{1}{2})^i$. Thus, σ achieves the result $\frac{2}{3}$ or better against every strategy of player $\Diamond$. However, for every finite-memory strategy $\hat{\sigma}$ of player $\Box$ there is a fixed constant $P^{\hat{\sigma}} < 1$ such that $\mathcal{P}_{\hat{v}}^{(\hat{\sigma},\pi)}(Reach(T)) \leq P^{\hat{\sigma}}$ for every $\pi \in \Pi$. Since $P^{\hat{\sigma}} < 1$, there surely exists $j \geq 1$ such that $P^{\hat{\sigma}} < 1 - (\frac{1}{2})^k$ for all $k > j$. Now let $\hat{\pi}$ be a MD strategy of player $\Diamond$ which in d_i selects either the transition $d_i \mapsto e_i$ or $d_i \mapsto s_i$, depending on whether $i \geq j$ or not, respectively. Now one can easily check that $\mathcal{P}_v^{(\hat{\sigma},\hat{\pi})}(Reach(T)) < \frac{2}{3}$, which means that $\hat{\sigma}$ is not an optimal maximising strategy in v.

Let us note that Claim (b) of Proposition 5.7 does *not* hold for MDPs. By applying Theorem 7.2.11 of Puterman [1994], we can conclude that if there is *some* optimal maximising strategy in a vertex v of a (possibly infinitely

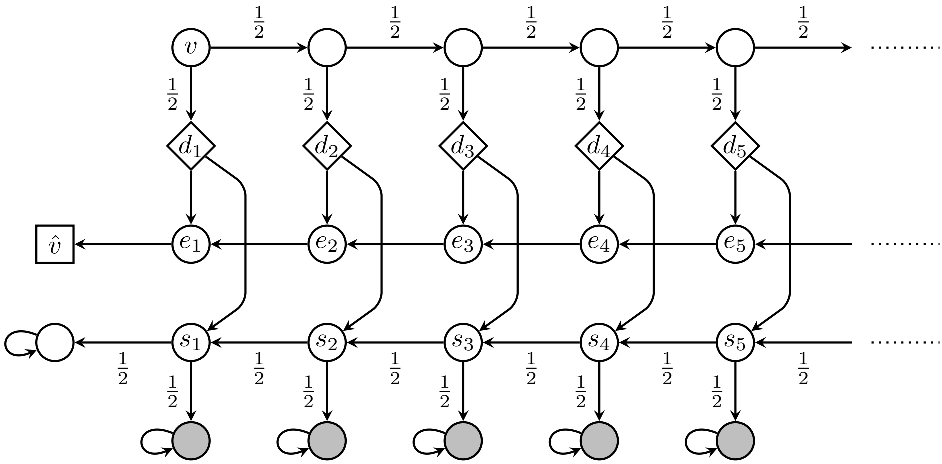
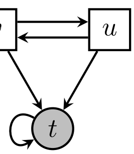


Figure 5.4 An optimal maximising strategy may require infinite memory

branching) MDP G, then there is also an MD optimal maximising strategy in v.

Claim (c) is not as trivial as it might seem. A naive idea of constructing an optimal maximising MD strategy just by selecting *some* value-maximising transition in every vertex does not work. To see this, consider the following MDP, where t is the only target vertex:



Obviously, $val(v) = val(u) = val(t) = 1$, but the MD strategy which selects the transitions $v \mapsto u$ and $u \mapsto v$ in the vertices v and u, respectively, is not optimal maximising. Nevertheless, it is not hard to show that for every vertex of $v \in V_\square$ there is a transition $v \mapsto v'$ such that the other outgoing transitions of v can be safely removed without influencing the value in any vertex. This result actually holds for all finitely branching stochastic games. Claim (c) then follows immediately because if $V_\square$ is finite, then we can successively fix such a transition in *every* $v \in V_\square$.

Proposition 5.8 *Let $G = (V, \mapsto, (V_\square, V_\Diamond, V_\bigcirc), Prob)$ be a finitely branching stochastic game with a reachability objective associated to a set $T \subseteq V$. For every $v \in V_\square$ there is a transition $v \mapsto v'$ such that the value of all $u \in V$*

remains unchanged when all outgoing transitions of v except for $v \mapsto v'$ are deleted from G.

Proof Let $v \in V_\Box$. If $v \in T$, the transition $v \mapsto v'$ can be chosen arbitrarily. Now assume that $v \notin T$. For every strategy $\tau \in \Sigma \cup \Pi$ we define the (unique) strategy $\tau[v]$ such that $\tau[v](w) = \tau(w')$, where w' is the shortest suffix of w which is either equal to w or starts with v. Intuitively, $\tau[v]$ behaves identically to τ until the vertex v is revisited. Then, $\tau[v]$ 'forgets' the history and behaves as if the play just started in v.

Let us define two auxiliary Borel sets of runs $\neg T \,\mathbf{U}\, v$ and $\neg v \,\mathbf{U}\, T$ where

- $\neg T \,\mathbf{U}\, v$ consists of all $w \in \mathit{Run}(G)$ such that $w(j) = v$ for some $j > 0$ and $w(i) \notin T$ for all $1 \leq i < j$;
- $\neg v \,\mathbf{U}\, T$ consists of all $w \in \mathit{Run}(G)$ such that $w(j) \in T$ for some $j > 0$ and $w(i) \neq v$ for all $1 \leq i < j$.

Observe that for all $(\sigma, \pi) \in \Sigma \times \Pi$ such that $\mathcal{P}_v^{(\sigma,\pi)}(\neg T \,\mathbf{U}\, v) < 1$ we have that $\mathcal{P}_v^{(\sigma[v],\pi[v])}(\mathit{Reach}(T))$ is equal to

$$\sum_{i=0}^{\infty} \left(\mathcal{P}_v^{(\sigma,\pi)}(\neg T \,\mathbf{U}\, v)\right)^i \cdot \mathcal{P}_v^{(\sigma,\pi)}(\neg v \,\mathbf{U}\, T) = \frac{\mathcal{P}_v^{(\sigma,\pi)}(\neg v \,\mathbf{U}\, T)}{1 - \mathcal{P}_v^{(\sigma,\pi)}(\neg T \,\mathbf{U}\, v)}.$$

For the moment, assume the following equality:

$$\sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathcal{P}_v^{(\sigma,\pi)}(\mathit{Reach}(T)) = \sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi^{\mathrm{MD}}} \mathcal{P}_v^{(\sigma[v],\pi)}(\mathit{Reach}(T)). \tag{5.7}$$

Note that for every $\pi \in \Pi^{\mathrm{MD}}$ we have that $\pi = \pi[v]$. For every $\sigma \in \Sigma$ and every transition $v \mapsto v'$, let $\sigma_{v\mapsto v'}$ be the strategy which agrees with σ on all arguments except for v where $\sigma_{v\mapsto v'}(v)$ selects the transition $v \mapsto v'$ with probability 1. It is easy to check that for every $\sigma \in \Sigma$ there must be some σ-*good* transition $v \mapsto v'$ satisfying

$$\inf_{\pi\in\Pi^{\mathrm{MD}}} \mathcal{P}_v^{(\sigma[v],\pi)}(\mathit{Reach}(T)) \leq \inf_{\pi\in\Pi^{\mathrm{MD}}} \mathcal{P}_v^{(\sigma_{v\mapsto v'}[v],\pi)}(\mathit{Reach}(T)).$$

For every $i \geq 1$, let us fix a strategy $\sigma_i \in \Sigma$ such that

$$\inf_{\pi\in\Pi^{\mathrm{MD}}} \mathcal{P}_v^{(\sigma_i[v],\pi)}(\mathit{Reach}(T)) \geq \mathit{val}(v) - \frac{1}{2^i}.$$

Since G is finitely branching, there is a transition $v \mapsto v'$ which is σ_i-good for infinitely many i's, and hence the value of v (and therefore also the value of the other vertices of V) does not change if all outgoing transitions of v except for $v \mapsto v'$ are deleted from G.

So, it remains to prove Equality (5.7). We start with the '$\geq$' direction. Let

$\hat{\pi}$ be the MD optimal minimising strategy which exists by Proposition 5.6. Then

$$\begin{aligned}\sup_{\sigma\in\Sigma}\inf_{\pi\in\Pi}\mathcal{P}_v^{(\sigma,\pi)}(\mathit{Reach}(T)) &= \sup_{\sigma\in\Sigma}\mathcal{P}_v^{(\sigma,\hat{\pi})}(\mathit{Reach}(T))\\ &\geq \sup_{\sigma\in\Sigma}\mathcal{P}_v^{(\sigma[v],\hat{\pi})}(\mathit{Reach}(T))\\ &\geq \sup_{\sigma\in\Sigma}\inf_{\pi\in\Pi^{\mathrm{MD}}}\mathcal{P}_v^{(\sigma[v],\pi)}(\mathit{Reach}(T)).\end{aligned}$$

Now assume that the '$\leq$' direction of Equality (5.7) does not hold. Then there is some $\varepsilon > 0$ such that

(1) there is an ε-optimal maximising strategy $\hat{\sigma}$ in v;
(2) for every $\sigma \in \Sigma$ there is $\pi \in \Pi^{\mathrm{MD}}$ s.t. $\mathcal{P}_v^{(\sigma[v],\pi)}(\mathit{Reach}(T)) \leq \mathit{val}(v) - 2\varepsilon$.

Note that condition (2) implies that for every $\sigma \in \Sigma$ there is $\pi \in \Pi^{\mathrm{MD}}$ such that either $\mathcal{P}_v^{(\sigma,\pi)}(\neg T\,\mathbf{U}\,v) = 1$, or $\mathcal{P}_v^{(\sigma,\pi)}(\neg T\,\mathbf{U}\,v) < 1$ and

$$\frac{\mathcal{P}_v^{(\sigma,\pi)}(\neg v\,\mathbf{U}\,T)}{1-\mathcal{P}_v^{(\sigma,\pi)}(\neg T\,\mathbf{U}\,v)} \leq \mathit{val}(v) - 2\varepsilon.$$

Now consider the strategy $\hat{\sigma}$ of condition (1) and a play initiated in v. Using condition (2) repeatedly, we obtain a strategy $\hat{\pi} \in \Pi$ such that whenever a state of the form wv is visited in the play $G_v^{(\hat{\sigma},\hat{\pi})}$, then either $\mathcal{P}_{wv}^{(\hat{\sigma},\hat{\pi})}(\neg T\,\mathbf{U}\,v) = 1$, or $\mathcal{P}_{wv}^{(\hat{\sigma},\hat{\pi})}(\neg T\,\mathbf{U}\,v) < 1$ and

$$\frac{\mathcal{P}_{wv}^{(\hat{\sigma},\hat{\pi})}(\neg v\,\mathbf{U}\,T)}{1-\mathcal{P}_{wv}^{(\hat{\sigma},\hat{\pi})}(\neg T\,\mathbf{U}\,v)} \leq \mathit{val}(v) - 2\varepsilon.$$

From this we obtain $\mathcal{P}_v^{(\hat{\sigma},\hat{\pi})}(\mathit{Reach}(T)) \leq \mathit{val}(v) - 2\varepsilon$ which is a contradiction. $\square$

Finally, let us consider a temporal objective $\mathbf{P}^{\succ\varrho}\mathbf{F}t$ where the atomic proposition t is valid exactly in the target vertices of T. The next proposition (taken from Brázdil et al. [2009a]) answers the associated determinacy question. Again, the answer seems somewhat unexpected.

Proposition 5.9 *Let $G = (V, \mapsto, (V_\Box, V_\Diamond, V_\bigcirc), \mathit{Prob})$ be a stochastic game with a temporal objective $\mathbf{P}^{\succ\varrho}\mathbf{F}t$ associated to a subset of target vertices $T \subseteq V$. Then G is* not *necessarily determined. However, if G is finitely branching, then it is determined.*

Proof A counterexample for the first part of Proposition 5.9 is easy to construct. Let G_u and G_v be the games of Figure 5.2 and Figure 5.3, where

u and v are the initial vertices of G_u and G_v, and T_u and T_v are the sets of target vertices of G_u and G_v, respectively. Consider a game G obtained by taking the disjoint union of G_u and G_v extended with a fresh stochastic vertex s with two outgoing transitions $s \xrightarrow{0.5} u$ and $s \xrightarrow{0.5} v$. The set of target vertices of G is $T_u \cup T_v$, and the initial vertex is s. Since $val(u) = 0$ and $val(v) = 1$, we obtain that $val(s) = \frac{1}{2}$. Now consider the temporal objective $\mathbf{P}^{\geq 0.5}\mathbf{F}t$. First, assume that player $\Box$ has a winning strategy $\hat{\sigma}$ in s. Since player $\Box$ has no optimal maximising strategy in v, there is a constant $P^{\hat{\sigma}} < \frac{1}{2}$ such that $\mathcal{P}_s^{(\hat{\sigma},\pi)}(Reach(T_v)) \leq P^{\hat{\sigma}}$ for every $\pi \in \Pi$. Since $val(u) = 0$, there is a strategy $\hat{\pi}$ of player $\Diamond$ such that $\mathcal{P}_s^{(\sigma,\hat{\pi})}(Reach(T_u)) < \frac{1}{2} - P^{\hat{\sigma}}$ for every $\sigma \in \Sigma$. Hence, $\mathcal{P}_s^{(\hat{\sigma},\hat{\pi})}(Reach(T_u \cup T_v)) < \frac{1}{2}$ which contradicts the assumption that $\hat{\sigma}$ is a winning strategy for player $\Box$. Similarly, one can show that there is no winning strategy for player $\Diamond$ which would achieve the negated objective $\mathbf{P}^{<0.5}\mathbf{F}t$ against every strategy of player $\Box$.

Now let us assume that G is finitely branching, and let us fix a vertex v of G. For technical convenience, assume that every target vertex t has only one outgoing transition $t \mapsto t$. The second part of Proposition 5.9 is not completely trivial, because player $\Box$ does not necessarily have an optimal maximising strategy in v even if G is finitely branching. Observe that if $\varrho > val(v)$, then player $\Box$ has a $(\mathbf{P}^{\succ\varrho}\mathbf{F}t)$-winning strategy in v (he may use, e.g., an ε-optimal maximising strategy where $\varepsilon = (\varrho - val(v))/2$). Similarly, if $\varrho < val(v)$, then player $\Diamond$ has a $(\neg\mathbf{P}^{\succ\varrho}\mathbf{F}t)$-winning strategy in v. Now assume that $\varrho = val(v)$. Obviously, it suffices to show that if player $\Diamond$ does *not* have a $(\neg\mathbf{P}^{\succ\varrho}\mathbf{F}t)$-winning strategy in v, then player $\Box$ has a $(\mathbf{P}^{\succ\varrho}\mathbf{F}t)$-winning strategy in v. This means to show that

$$\forall \pi \in \Pi \;\; \exists \sigma \in \Sigma \;:\; \mathcal{P}_v^{\sigma,\pi}(Reach(T)) \succ \varrho \tag{5.8}$$

implies

$$\exists \sigma \in \Sigma \; \forall \pi \in \Pi \;:\; \mathcal{P}_v^{\sigma,\pi}(Reach(T)) \succ \varrho. \tag{5.9}$$

Observe that if $\succ$ is $>$, then (5.8) does not hold because player $\Diamond$ has an optimal minimising strategy by Proposition 5.6. For the constraint ≥ 0, the statement is trivial. Hence, it suffices to consider the case when $\succ$ is $\geq$ and $\varrho = val(v) > 0$. Assume that (5.8) holds. We say that a vertex $u \in V$ is *good* if

$$\forall \pi \in \Pi \;\; \exists \sigma \in \Sigma \;:\; \mathcal{P}_u^{\sigma,\pi}(Reach(T)) \geq val(u). \tag{5.10}$$

Further, we say that a transition $u \mapsto u'$ of G is *optimal* if either $u \in V_{\bigcirc}$, or $u \in V_{\Box} \cup V_{\Diamond}$ and $val(u) = val(u')$. Observe that for every $u \in V_{\Box} \cup V_{\Diamond}$ there is at least one optimal transition $u \mapsto u'$, because G is finitely branching.

Further, note that if $u \in V_\Box$ is a good vertex, then there is at least one optimal $u \mapsto u'$ where u' is good. Similarly, if $u \in V_\Diamond$ is good then for every optimal transition $u \mapsto u'$ we have that u' is good, and if $u \in V_\bigcirc$ is good and $u \mapsto u'$ then u' is good. Hence, we can define a game $\bar{G}$, where the set of vertices $\bar{V}$ consists of all good vertices of G, and for all $u, u' \in \bar{V}$ we have that (u, u') is a transition of $\bar{G}$ iff $u \mapsto u'$ is an optimal transition of G. The transition probabilities in $\bar{G}$ are the same as in G. Now we prove the following three claims:

(a) For every $u \in \bar{V}$ we have that $val(u, \bar{G}) = val(u, G)$.
(b) $\exists \bar{\sigma} \in \Sigma_{\bar{G}} \quad \forall \bar{\pi} \in \Pi_{\bar{G}} : \mathcal{P}_v^{\bar{\sigma},\bar{\pi}}(Reach(T, \bar{G})) \geq val(v, \bar{G}) = \varrho$.
(c) $\exists \sigma \in \Sigma_G \quad \forall \pi \in \Pi_G : \mathcal{P}_v^{\sigma,\pi}(Reach(T, G)) \geq \varrho$.

Note that Claim (c) is exactly (5.9). We start by proving Claim (a). Let $u \in \bar{V}$. Due to Proposition 5.6, there is a MD strategy $\pi \in \Pi_G$ which is optimal minimising in every vertex of G (particularly in u) and selects only the optimal transitions. Hence, the strategy π can also be used in the restricted game $\bar{G}$ and thus we obtain $val(u, \bar{G}) \leq val(u, G)$. Now suppose that $val(u, \bar{G}) < val(u, G)$. By applying Proposition 5.6 to $\bar{G}$, there is an optimal minimising MD strategy $\bar{\pi} \in \Pi_{\bar{G}}$. Further, for every vertex t of G which is not good there is a strategy $\pi_t \in \Pi_G$ such that for every $\sigma \in \Sigma_G$ we have that $\mathcal{P}_t^{\sigma,\pi_t}(Reach(T, G)) < val(u, G)$ (this follows immediately from (5.10)). Now consider a strategy $\pi' \in \Pi_G$ which for every play of G initiated in u behaves in the following way:

- As long as player $\Box$ uses only the transitions of G that are preserved in $\bar{G}$, the strategy π' behaves exactly like the strategy $\bar{\pi}$.
- When player $\Box$ uses a transition $r \mapsto r'$ which is not a transition in $\bar{G}$ for the first time, then the strategy π' starts to behave either like the optimal minimising strategy π or the strategy $\pi_{r'}$, depending on whether r' is good or not (observe that if r' is good, then $val(r', G) < val(r, G)$ because $r \mapsto r'$ is not a transition of $\bar{G}$).

Now it is easy to check that for every $\sigma \in \Sigma_G$ we have that $\mathcal{P}_u^{\sigma,\pi'}(Reach(T, G)) < val(u, G)$, which contradicts the assumption that u is good.

Now we prove Claim (b). Due to Proposition 5.5, for every $u \in \bar{V}$ we can fix a strategy $\bar{\sigma}_u \in \Sigma_{\bar{G}}$ and $n_u \geq 1$ such that for every $\bar{\pi} \in \Pi_{\bar{G}}$ we have that $\mathcal{P}_u^{\bar{\sigma}_u,\bar{\pi}}(Reach^{n_u}(T, \bar{G})) > val(u, \bar{G})/2$. For every $k \geq 0$, let $B(k)$ be the set of all vertices u reachable from v in $\bar{G}$ via a path of length exactly k which does not visit T. Observe that $B(k)$ is finite because $\bar{G}$ is finitely branching. Further, for every $i \geq 0$ we define a bound m_i inductively as follows: $m_0 = 1$, and $m_{i+1} = m_i + \max\{n_u \mid u \in B(m_i)\}$. Now we define a strategy $\bar{\sigma} \in \Sigma_{\bar{G}}$ which turns out to be $(\mathbf{P}^{\geq\varrho}\mathbf{F}t)$-winning in the vertex v of $\bar{G}$. For every $w \in \bar{V}^*\bar{V}_\Box$ such that $m_i \leq |w| < m_{i+1}$ we put $\bar{\sigma}(w) = \bar{\sigma}_u(uw_2)$, where $w = w_1uw_2$, $|w_1| = m_i - 1$ and $u \in \bar{V}$. Now it is easy to check that for every $i \geq 1$ and every strategy $\bar{\pi} \in \Pi_{\bar{G}}$ we have that $\mathcal{P}_v^{\bar{\sigma},\bar{\pi}}(\mathit{Reach}^{m_i}(T,\bar{G})) > (1 - \frac{1}{2^i})\varrho$. This means that the strategy $\bar{\sigma}$ is $(\mathbf{P}^{\geq\varrho}\mathbf{F}t)$-winning in v.

It remains to prove Claim (c). Consider a strategy $\sigma \in \Sigma_G$ which for every play of G initiated in v behaves as follows:

- As long as player $\Diamond$ uses only the optimal transitions, the strategy σ behaves exactly like the strategy $\bar{\sigma}$.
- When player $\Diamond$ uses a non-optimal transition $r \mapsto r'$ for the first time, the strategy σ starts to behave like an ε-optimal maximising strategy in r', where $\varepsilon = (\mathit{val}(r',G) - \mathit{val}(r,G))/2$. Note that since $r \mapsto r'$ is not optimal, we have that $\mathit{val}(r',G) > \mathit{val}(r,G)$.

It is easy to check that σ is $(\mathbf{P}^{\geq\varrho}\mathbf{F}t)$-winning in v. □

5.4 Some directions of future research

There are many challenging open problems and emerging lines of research in the area of stochastic games. Some of them have already been mentioned in the previous sections. We close by listing a few attractive topics (the presented list is of course far from being complete).

- ***Infinite-state games.*** The existing results about infinite-state games concern mainly games and MDPs generated by pushdown automata, lossy channel systems, or one-counter automata (see Section 5.2 for a more detailed summary). As indicated in Section 5.3, even in the setting of simple reachability objectives, many questions become subtle and require special attention. There is a plethora of automata-theoretic models with specific advantages, and the corresponding games can have specific objectives relevant to the chosen model. When compared to finite-state games, this field of research appears unexplored and offers many open problems.
- ***Games with non-conflicting objectives.*** It has been argued that non-zero-sum stochastic games are also relevant for purposes of formal verification of computer systems (see, e.g., Chatterjee et al. [2004c]). In this case, the main problem is the existence and computability of Nash equilibria (see Nash [1950]). Depending on the concrete objectives of the players, a Nash equilibrium may or may not exist, and there can be several

equilibrium points. Some existing literature about non-zero-sum stochastic games is mentioned in Section 5.2. The current knowledge is still limited.

- ***Games with time.*** The modelling power of continuous-time stochastic models such as continuous-time (semi)Markov chains (see, e.g., Norris [1998], Ross [1996]) or the real-time probabilistic processes of Alur et al. [1991] can be naturally extended by the element of choice. Thus, we obtain various types of continuous-time stochastic games. Stochastic games and MDPs over continuous-time Markov chains were studied by Baier et al. [2005], Neuhäußer et al. [2009], Brázdil et al. [2009b] and Rabe and Schewe [2010]. In this context, it makes sense to consider various types of strategies that measure or ignore the elapsed time, and study specific types of objectives that can be expressed by, e.g., the timed automata of Alur and Dill [1994].

The above discussed concepts are to a large extent orthogonal and can be combined almost arbitrarily. Thus, one can model very complex systems of time, chance, and choice. Many of the fundamental results are still waiting to be discovered.

Acknowledgements: I thank Václav Brožek and Tomáš Brázdil for reading a preliminary draft of this chapter. The work has been supported by the Czech Science Foundation, grant No. P202/10/1469.

References

P. Abdulla, N. Henda, L. de Alfaro, R. Mayr, and S. Sandberg. Stochastic games with lossy channels. In *Proceedings of FoSSaCS 2008*, volume 4962 of *Lecture Notes in Computer Science*, pages 35–49. Springer, 2005.

R. Alur and D. Dill. A theory of timed automata. *Theoretical Computer Science*, 126(2):183–235, 1994. Fundamental Study.

R. Alur, C. Courcoubetis, and D. Dill. Model-checking for probabilistic real-time systems. In *Proceedings of ICALP'91*, volume 510 of *Lecture Notes in Computer Science*, pages 115–136. Springer, 1991.

C. Baier, M. Größer, M. Leucker, B. Bollig, and F. Ciesinski. Controller synthesis for probabilistic systems. In *Proceedings of IFIP TCS'2004*, pages 493–506. Kluwer, 2004.

C. Baier, H. Hermanns, J.-P. Katoen, and B. Haverkort. Efficient computation of time-bounded reachability probabilities in uniform continuous-time Markov decision processes. *Theoretical Computer Science*, 345:2–26, 2005.

C. Baier, N. Bertrand, and P. Schnoebelen. On computing fixpoints in well-structured regular model checking, with applications to lossy channel systems. In *Proceedings of LPAR 2006*, volume 4246 of *Lecture Notes in Computer Science*, pages 347–361. Springer, 2006.

C. Baier, N. Bertrand, and P. Schnoebelen. Verifying nondeterministic probabilistic channel systems against ω-regular linear-time properties. *ACM Transactions on Computational Logic*, 9(1), 2007.

A. Bianco and L. de Alfaro. Model checking of probabilistic and nondeterministic systems. In *Proceedings of FST&TCS'95*, volume 1026 of *Lecture Notes in Computer Science*, pages 499–513. Springer, 1995.

P. Billingsley. *Probability and Measure.* Wiley, Hoboken, New Jersey, 1995.

T. Brázdil and V. Forejt. Strategy synthesis for Markov decision processes and branching-time logics. In *Proceedings of CONCUR 2007*, volume 4703 of *Lecture Notes in Computer Science*, pages 428–444. Springer, 2007.

T. Brázdil, V. Brožek, V. Forejt, and A. Kučera. Stochastic games with branching-time winning objectives. In *Proceedings of LICS 2006*, pages 349–358. IEEE Computer Society Press, 2006.

T. Brázdil, V. Brožek, V. Forejt, and A. Kučera. Reachability in recursive Markov decision processes. *Information and Computation*, 206(5):520–537, 2008.

T. Brázdil, V. Forejt, J. Křetínský, and A. Kučera. The satisfiability problem for probabilistic CTL. In *Proceedings of LICS 2008*, pages 391–402. IEEE Computer Society Press, 2008.

T. Brázdil, V. Forejt, and A. Kučera. Controller synthesis and verification for Markov decision processes with qualitative branching time objectives. In *Proceedings of ICALP 2008, Part II*, volume 5126 of *Lecture Notes in Computer Science*, pages 148–159. Springer, 2008.

T. Brázdil, V. Brožek, A. Kučera, and J. Obdržálek. Qualitative reachability in stochastic BPA games. In *Proceedings of STACS 2009*, volume 3 of *Leibniz International Proceedings in Informatics*, pages 207–218. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2009a. A full version is available at arXiv:1003.0118 [cs.GT].

T. Brázdil, V. Forejt, J. Krčál, J. Křetínský, and A. Kučera. Continuous-time stochastic games with time-bounded reachability. In *Proceedings of FST&TCS 2009*, volume 4 of *Leibniz International Proceedings in Informatics*, pages 61–72. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2009b.

T. Brázdil, V. Brožek, K. Etessami, A. Kučera, and D. Wojtczak. One-counter Markov decision processes. In *Proceedings of SODA 2010*, pages 863–874. SIAM, 2010.

V. Brožek. *Basic Model Checking Problems for Stochastic Games.* PhD thesis, Masaryk University, Faculty of Informatics, 2009.

K. Chatterjee. *Stochastic ω-regular Games.* PhD thesis, University of California, Berkeley, 2007.

K. Chatterjee, M. Jurdziński, and T. Henzinger. Simple stochastic parity games. In *Proceedings of CSL'93*, volume 832 of *Lecture Notes in Computer Science*, pages 100–113. Springer, 1994.

K. Chatterjee, L. de Alfaro, and T. Henzinger. Trading memory for randomness. In *Proceedings of 2nd Int. Conf. on Quantitative Evaluation of Systems (QEST'04)*, pages 206–217. IEEE Computer Society Press, 2004a.

K. Chatterjee, M. Jurdziński, and T. Henzinger. Quantitative stochastic parity games. In *Proceedings of SODA 2004*, pages 121–130. SIAM, 2004b.

K. Chatterjee, R. Majumdar, and M. Jurdziński. On Nash equilibria in stochastic games. In *Proceedings of CSL 2004*, volume 3210 of *Lecture Notes in Computer Science*, pages 26–40. Springer, 2004c.

K. Chatterjee, L. de Alfaro, and T. Henzinger. The complexity of stochastic Rabin and Streett games. In *Proceedings of ICALP 2005*, volume 3580 of *Lecture Notes in Computer Science*, pages 878–890. Springer, 2005.

K. Chatterjee, T. Henzinger, and M. Jurdziński. Games with secure equilibria. *Theoretical Computer Science*, 365(1–2):67–82, 2006.

K. Chatterjee, L. Doyen, and T. Henzinger. A survey of stochastic games with limsup and liminf objectives. In *Proceedings of ICALP 2009*, volume 5556 of *Lecture Notes in Computer Science*, pages 1–15. Springer, 2009.

A. Condon. The complexity of stochastic games. *Information and Computation*, 96 (2):203–224, 1992.

L. de Alfaro and T. Henzinger. Concurrent omega-regular games. In *Proceedings of LICS 2000*, pages 141–154. IEEE Computer Society Press, 2000.

L. de Alfaro and R. Majumdar. Quantitative solution of omega-regular games. *Journal of Computer and System Sciences*, 68:374–397, 2004.

E. Emerson. Temporal and modal logic. *Handbook of Theoretical Computer Science*, B:995–1072, 1991.

E. Emerson and C. Jutla. The complexity of tree automata and logics of programs. In *Proceedings of FOCS'88*, pages 328–337. IEEE Computer Society Press, 1988.

K. Etessami and M. Yannakakis. Recursive Markov decision processes and recursive stochastic games. In *Proceedings of ICALP 2005*, volume 3580 of *Lecture Notes in Computer Science*,pages 891–903. Springer, 2005.

K. Etessami and M. Yannakakis. Efficient qualitative analysis of classes of recursive Markov decision processes and simple stochastic games. In *Proceedings of STACS 2006*, volume 3884 of *Lecture Notes in Computer Science*, pages 634–645. Springer, 2006.

K. Etessami, D. Wojtczak, and M. Yannakakis. Recursive stochastic games with positive rewards. In *Proceedings of ICALP 2008, Part I*, volume 5125 of *Lecture Notes in Computer Science*, pages 711–723. Springer, 2008.

J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, Berlin, 1996.

V. Forejt. *Controller Synthesis for Markov Decision Processes with Branching-Time Objectives*. PhD thesis, Masaryk University, Faculty of Informatics, 2009.

G. Gillette. Stochastic games with zero stop probabilities. *Contributions to the Theory of Games, vol III*, pages 179–187, 1957.

H. Gimbert and F. Horn. Simple stochastic games with few random vertices are easy to solve. In *Proceedings of FoSSaCS 2008*, volume 4962 of *Lecture Notes in Computer Science*, pages 5–19. Springer, 2005.

N. Halman. Simple stochastic games, parity games, mean payoff games and discounted payoff games are all LP-type problems. *Algorithmica*, 49(1):37–50, 2007.

H. Hansson and B. Jonsson. A logic for reasoning about time and reliability. *Formal Aspects of Computing*, 6:512–535, 1994.

A. Hoffman and R. Karp. On nonterminating stochastic games. *Management Science*, 12:359–370, 1966.

P. Hunter and A. Dawar. Complexity bounds for regular games. In *Proceedings of MFCS 2005*, volume 3618 of *Lecture Notes in Computer Science*, pages 495–506. Springer, 2005.

J. Kemeny, J. Snell, and A. Knapp. *Denumerable Markov Chains*. Springer, 1976.

A. Kučera and O. Stražovský. On the controller synthesis for finite-state Markov decision processes. *Fundamenta Informaticae*, 82(1–2):141–153, 2008.

T. Liggett and S. Lippman. Stochastic games with perfect information and time average payoff. *SIAM Review*, 11(4):604–607, 1969.

W. Ludwig. A subexponential randomized algorithm for the simple stochastic game problem. *Information and Computation*, 117(1):151–155, 1995.

A. Maitra and W. Sudderth. Finitely additive stochastic games with Borel measurable payoffs. *International Journal of Game Theory*, 27:257–267, 1998.

D. Martin. The determinacy of Blackwell games. *Journal of Symbolic Logic*, 63(4): 1565–1581, 1998.

A. McIver and C. Morgan. Games, probability, and the quantitative μ-calculus. In *Proceedings of LPAR 2002*, volume 2514 of *Lecture Notes in Computer Science*, pages 292–310. Springer, 2002.

J. Nash. Equilibrium points in N-person games. *Proceedings of the National Academy of Sciences*, 36:48–49, 1950.

M. Neuhäußer, M. Stoelinga, and J.-P. Katoen. Delayed nondeterminism in continuous-time Markov decision processes. In *Proceedings of FoSSaCS 2009*, volume 5504 of *Lecture Notes in Computer Science*, pages 364–379. Springer, 2009.

A. Neyman and S. Sorin. *Stochastic Games and Applications*. Kluwer, Dordrecht, 2003.

J. Norris. *Markov Chains*. Cambridge University Press, Cambridge, 1998.

A. Pnueli. The temporal logic of programs. In *Proceedings of 18th Annual Symposium on Foundations of Computer Science*, pages 46–57. IEEE Computer Society Press, 1977.

M. Puterman. *Markov Decision Processes*. Wiley, Hoboken, New Jersey, 1994.

M. Rabe and S. Schewe. Optimal time-abstract schedulers for CTMDPs and Markov games. In *Eighth Workshop on Quantitative Aspects of Programming Languages*, 2010.

S. Ross. *Stochastic Processes*. Wiley, Hoboken, New Jersey, 1996.

P. Secchi and W. Sudderth. Stay-in-a-set games. *International Journal of Game Theory*, 30:479–490, 2001.

L. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39: 1095–1100, 1953.

A. Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285–309, 1955.

W. Thomas. Automata on infinite objects. *Handbook of Theoretical Computer Science*, B:135–192, Elsevier, Amsterdam, 1991.

M. Ummels and D. Wojtczak. Decision problems for Nash equilibria in stochastic games. In *Proceedings of CSL 2009*, volume 5771 of *Lecture Notes in Computer Science*, pages 515–529. Springer, 2009.

P. Wolper. Temporal logic can be more expressive. In *Proceedings of 22nd Annual Symposium on Foundations of Computer Science*, pages 340–348. IEEE Computer Society Press, 1981.