Perfect Information Stochastic Priority Games

Hugo Gimbert, Wieslaw Zielonka

Perfect information stochastic priority games

Hugo Gimbert[1] and Wiesław Zielonka[2]
[1]LIX, École Polytechnique, Palaiseau, France
gimbert@lix.polytechnique.fr
[2]LIAFA, Université Paris 7 and CNRS, Paris, France
zielonka@liafa.jussieu.fr

Abstract

We introduce stochastic priority games — a new class of perfect information stochastic games. These games can take two different, but equivalent, forms. In stopping priority games a play can be stopped by the environment after a finite number of stages, however, infinite plays are also possible. In discounted priority games only infinite plays are possible and the payoff is a linear combination of the classical discount payoff and of a limit payoff evaluating the performance at infinity. Shapley games [12] and parity games [6] are special extreme cases of priority games.

1 Introduction

Recently de Alfaro, Henzinger and Majumdar[4] introduced a new variant of μ-calculus: discounted μ-calculus. As it is known since the seminal paper [6] of Emerson and Jutla μ-calculus is strongly related to parity games and this relationship is preserved even for stochastic games, [5]. In this context it is natural to ask if there is a class of games that corresponds to discounted μ-calculus of [4]. A partial answer to this question was given in [8], where an appropriate class of infinite discounted games was introduced. However, in [8], only deterministic systems were considered and much more challenging problem of stochastic games was left open. In the present paper we return to the problem but in the context perfect information stochastic games. The most basic and usually non-trivial question is if the games that we consider admit "simple" optimal strategies for both players. We give a positive answer, for all games presented in this paper both players have pure stationary optimal strategies. Since our games contain parity games as a very special case, our paper extends the result known for perfect information parity games [2, 10, 3, 14].

However, we have an objective which is larger than just transferring to stochastic games the results known for deterministic systems. Parity games are used (directly or through an associated logic) in verification. Conditions that are verified often do not depend on any finite prefix of the play (take as a typical example a simple condition like "A wins if we visit infinitely often some set X of states"). However, certainly all real systems have a finite life span thus we can ask what is the meaning of infinite games when they are used to examine such systems. Notice that the same

question arises in classical game theory [11]. The obvious answer is that the life span is finite but unknown or sufficiently long and thus infinite games are a convenient approximation of finite games. However, what finite games are approximated by parity games? Notice that for the games like mean-payoff games that are used in economics the answer is simple: infinite mean-payoff games approximate finite mean-payoff games of long or unknown duration. But we do not see any obvious candidate for "finite parity games". Suppose that C is a parity condition and f_C a payoff mapping associated with C, i.e. f_C maps to 1 (win) all infinite sequence of states that satisfy C and to 0 all "loosing" sequences. Now we can look for a sequence $f_n, n \in \mathbb{N}$, of payoff functions, such that each f_n, defined for state sequences of length n, gives a payoff for games of length n and such that for each infinite sequence $s_0 s_1 \dots$ of states $f_n(s_0 \dots s_{n-1}) \xrightarrow[n\to\infty]{} f_C(s_0 s_1 \dots)$. However, except for very special parity conditions C, such payoff mappings f_n do not exist, thus parity games cannot approximate finite games in the same way as infinite mean-payoff games approximate finite mean-payoff games.

Nevertheless, it turns out that parity games approximate finite games, however "finite" does not mean here that the number of steps is fixed, instead these games are finite in the sense that they stop with probability 1. In Section 4 we present a class of priority stopping games. In the simplest case, when the stopping probabilities are positive for all states, stopping games are stochastic games defined by Shapley [12]. However, we examine also stopping games for which stopping probabilities are positive only for some states. One of the results of this paper can be interpreted in the following way: parity games are a limit of stopping games when the stopping probabilities tend to 0 but all at the same time but rather one after another, in the order determined by priorities.

2 Arenas and perfect information games

Perfect information stochastic games are played by two players, tha we call player 1 and player 2. We assume that player $i \in \{1, 2\}$ controls a finite set S_i of states, S_1 and S_2 are disjoint and $S = S_1 \cup S_2$ is the set of all states.

With each state $s \in S$ is associated a finite non-empty set A_s of actions that are available at s and we set $A = \cup_{s\in S} A_s$ to be the set of all actions.

If the current state is $s \in S_i$ then player i controlling this state chooses an available action $a \in S_i$ and, with a probability $p(s'|s, a)$, the systems changes its state to $s' \in S$. Thus $p(\cdot|s, a), s \in S, a \in A_s$, are transition probabilities satisfying the usual conditions: $0 \le p(s'|s, a) \le 1$ and $\sum_{s'\in S} p(s'|s, a) = 1$.

Let $\mathcal{H}^\omega$ be the set of *histories*, i.e. the set of all infinite sequences $s_0 a_0 s_1 a_1 s_2 \dots$ alternating states and actions. Assuming that the sets S and A are equipped with the discrete topology, we equip $\mathcal{H}^\omega$ with the product topology, i.e. the smallest topology for which the mappings

$$\mathbf{S}_i : \mathcal{H}^\omega \to S, \qquad \mathbf{S}_i : \mathcal{H}^\omega \ni s_0 a_0 \dots s_i a_i \dots \mapsto s_i$$

and

$$\mathbf{A}_i : \mathcal{H}^\omega \to A, \qquad \mathbf{A}_i : \mathcal{H}^\omega \ni s_0 a_0 \dots s_i a_i \dots \mapsto a_i$$

are continuous. Thus $(\mathbf{S}_i)_{i\in\mathbb{N}}$ and $(\mathbf{A}_i)_{i\in\mathbb{N}}$, are stochastic processes on the probability space $(\mathcal{H}^\omega, \mathcal{B})$, where $\mathcal{B}$ is Borel σ-algebra generated by open subsets of $\mathcal{H}^\omega$.

The data consisting of the state sets S_1, S_2, available actions $(A_s)_{s\in S}$ and transition probabilities $p(\cdot, s, a)$ is an *arena* $\mathcal{A}$.

Let $u : \mathcal{H}^\omega \to \mathbb{R}$ be a bounded Borel measurable mapping. We interpret $u(h), h \in \mathcal{H}^\omega$, as the payoff obtained by player 1 from player 2 after an infinite play h.

A couple $(\mathcal{A}, u)$ consisting of an arena and a payoff mapping is a *perfect information stochastic game*.

Let $\mathcal{H}_i^+ = (SA)^* S_i, i \in \{1, 2\}$, be the set of finite non-empty histories terminating at a state controlled by player i. A *strategy* for player i is a family of conditional probabilities $\sigma(a|h_n)$ for all $h_n = s_0 a_0 \dots s_n \in \mathcal{H}_i^+$ and $a \in A_{a_n}$. Intuitively, $\sigma(a|s_0 a_0 \dots s_n)$ gives the probability that player i controlling the last state s_n chooses an (available) action a, while the sequence h_n describes the first n steps of the game. As usual $0 \le \sigma(a|s_0 a_0 \dots s_n) \le 1$ and $\sum_{a\in A_{s_n}} \sigma(a|s_0 a_0 \dots s_n) = 1$.

A strategy σ is said to be *pure* if for each finite history $h_n = s_0 a_0 \dots s_n \in \mathcal{H}_1^+$ there is an action $a \in A_{s_n}$ such that $\sigma(a|h_n) = 1$, i.e. no randomization is used to choose an action to execute. A strategy σ is *stationary* if for each finite history $h_n = s_0 a_0 \dots s_n \in \mathcal{H}_1^+$, $\sigma(\cdot|h_n) = \sigma(\cdot|s_n)$, i.e. the probability distribution used to choose actions depends only on the last state.

Notice that pure stationary strategies for player i can be identified with mappings $\sigma : S_i \to A$ such that $\sigma(s) \in A_s$ for $s \in S_i$.

In the sequel we shall use σ, possibly with subscripts or superscripts, to denote a strategy of player 1. On the other hand, τ will always denote a strategy of player 2.

Given and initial state s, strategies σ, τ of both players determine a unique probability measure $\mathbb{P}^s_{\sigma,\tau}$ on $(\mathcal{H}^\omega, \mathcal{B})$, [7].

The expectation corresponding to the probability measure $\mathbb{P}^s_{\sigma,\tau}$ is denoted $\mathbb{E}^s_{\sigma,\tau}$. Thus $\mathbb{E}^s_{\sigma,\tau}(u)$ gives the expected payoff obtained by player 1 from player 2 in the game $(\mathcal{A}, u)$ starting at state s when the players use strategies σ, τ respectively. If $\sup_\sigma \inf_\tau \mathbb{E}^s_{\sigma,\tau}(u) = \inf_\tau \sup_\sigma \mathbb{E}^s_{\sigma,\tau}(u)$ for each state s then the quantity appearing on both side of this equality is *the value of the game* (for initial state s) and is denoted $\mathrm{val}^s(u)$.

Strategies $\sigma^\sharp$ and $\tau^\sharp$ of players 1, 2 are *optimal* in the game $(\mathcal{A}, u)$ if for each state $s \in S$ and for all strategies $\sigma \in \Sigma$, $\tau \in \mathcal{T}$

$$\mathbb{E}^s_{\sigma,\tau^\sharp}[u] \le \mathbb{E}^s_{\sigma^\sharp,\tau^\sharp}[u] \le \mathbb{E}^s_{\sigma^\sharp,\tau}[u] \ .$$

If $\sigma^\sharp$ and $\tau^\sharp$ are optimal strategies then $\mathrm{val}^s(u) = \mathbb{E}^s_{\sigma^\sharp,\tau^\sharp}[u]$, i.e. the expected payoff obtained when both players use optimal strategies is equal to the value of the game.

3 Priority games

Starting from this moment we assume that each arena $\mathcal{A}$ is equipped with a *priority mapping*

$$\varphi : S \to \{1, \dots, k\} \tag{1}$$

from the set S of states to the set $\{1, \dots, k\}$ of (positive integer) *priorities*. The composition

$$\boldsymbol{\varphi}_n = \varphi \circ \mathbf{S}_n, \quad , n \in \mathbb{N} \ , \tag{2}$$

$\boldsymbol{\varphi}_n : \mathcal{H}^\omega \to \{1, \ldots, k\}$, gives therefore a stochastic process with values in $\{1, \ldots, k\}$. Then $\liminf_i \boldsymbol{\varphi}_i$ is a random variable

$$\mathcal{H}^\omega \ni h \mapsto \liminf_i \boldsymbol{\varphi}_i(h)$$

giving for each infinite history $h \in \mathcal{H}^\omega$ its priority which the smallest priority visited infinitely often in h (we assume that $\{1, \ldots, k\}$ is equipped with the usual order on integers and $\liminf$ is taken for this order). From this moment onward, we assume that there is a fixed a *reward mapping*

$$r : \{1, \ldots, k\} \to [0, 1] \tag{3}$$

from priorities to the interval $[0, 1]$.

The *priority payoff mapping* $u : \mathcal{H}^\omega \to [0, 1]$ is defined as

$$u(h) = r(\liminf_i \boldsymbol{\varphi}_i(h)), \quad h \in \mathcal{H}^\omega \ . \tag{4}$$

Thus, in the priority game $(\mathcal{A}, u)$, the payoff received by player 1 from player 2 is the reward corresponding to the minimal priority visited infinitely often. If r maps odd priorities to 1 and even priorities to 0 then we get a parity game.

4 Stopping priority games

In the sequel we assume that besides the priority and reward mappings (1) and (3) we have also a mapping

$$\lambda : \{1, \ldots, k\} \to [0, 1] \tag{5}$$

from priorities to the interval $[0, 1]$.

We modify the rules of the priority game of Section 3 in the following way.

Every time a state s is visited the game can stop with probability $1 - \lambda(\varphi(s))$, where $\varphi(s)$ is the priority of s. If the game stops at s then player 1 receives from player 2 the payoff $r(\varphi(s))$. If the game does not stop then the player controlling s chooses an action $a \in A_s$ and we go to a state t with probability $p(t|s, a)$. (Thus $p(t|s, a)$ should now be interpreted as the probability to go to t *under the condition that the games does not stop.*) The rules above determine the payoff in the case when the games stops at some state s. However, λ can be 1 for some states (priorities) and then it is possible to have also infinite plays with a positive probability. For such infinite plays the payoff is calculated as in priority games of the preceding section.

Let us note that if $\lambda(p) = 1$ for all priorities $p \in \{1, \ldots, k\}$ then actually we never stop and the game described above is the same as the priority game of the preceding section.

On the other hand, if $\lambda(p) < 1$ for all priorities p, i.e. the stopping probabilities are positive for all states, then the game will stop with probability 1. Shapley [12] proved that for such games both players have optimal stationary strategies. In fact Shapley considered general stochastic games while we limit ourselves to perfect information stochastic games and for such games the optimal strategies constructed in [12] are not only stationary but also pure.

Theorem 1 (Shapley 1953)**.** *If, for all priorities i, $\lambda(i) < 1$ then both players have pure stationary optimal strategies in the priority stopping game.*

Stopping games have an appealing intuitive interpretation but they are not consistent with the framework fixed in Section 2, where the probability space consisted of infinite histories only. This obstacle can be removed in the following way. For each priority $i \in \{1, \dots, k\}$ we create a new "stopping" state $i^\sharp$ that we add to the arena $\mathcal{A}$. The priority of $i^\sharp$ is set to i, $\varphi(i^\sharp) = i$. The set of newly created states is denoted $S^\sharp$. There is only one action available at each $i^\sharp$ and executing this action we return immediately to $i^\sharp$ with probability 1, it is impossible to leave a stopping state. Note also that since there is only one action available at $i^\sharp$ it does not matter which of the two players controls "stopping" states. For each non-stopping state $s \in S$ we modify the transition probabilities. Formally we define new transition probabilities $p^\sharp(\cdot|\cdot,\cdot)$ by setting, for $s, t \in S$, $a \in A_s$,

$$p^\sharp(t|s,a) = \lambda(\varphi(s)) \cdot p(t|s,a)$$

and

$$p^\sharp(i^\sharp|s,a) = \begin{cases} 1 - \lambda(\varphi(s)) & \text{if } i = \varphi(s), \\ 0 & \text{otherwise } . \end{cases}$$

Let us note by $\mathcal{A}^\sharp_\lambda$ the arena obtained from $\mathcal{A}$ in this way. It is worth noticing that, even if the set of finite histories of $\mathcal{A}^\sharp_\lambda$ strictly contains the set of finite histories of $\mathcal{A}$, we can identify the strategies in both arenas. In fact, given a strategy for arena $\mathcal{A}$ there is only one possible way to complete it to a strategy in $\mathcal{A}^\sharp_\lambda$ since for finite histories in $\mathcal{A}^\sharp_\lambda$ that end in a stopping state $i^\sharp$ any strategy chooses always the unique action available at $i^\sharp$. Clearly, playing a stopping priority game on $\mathcal{A}$ is the same as playing priority game on $\mathcal{A}^\sharp_\lambda$: stopping at state s in $\mathcal{A}$ yields the same payoff as an infinite history in $\mathcal{A}^\sharp_\lambda$ that loops at $i^\sharp$, where $i = \varphi(s)$.

5 Discounted priority games

The aim of this section is to introduce a new class of inifinite games that are equivalent to stopping priority games.

As previously, we suppose that arenas are equipped with a priority mapping (1) and that a reward mapping (3) is fixed.

On the other hand, the mapping λ of (5), although also present, has now another interpretation, it does not define stopping probabilities but it provides discount factors applied to one-step rewards.

Let

$$\boldsymbol{r}_i = r \circ \boldsymbol{\varphi}_i \quad \text{and} \quad \boldsymbol{\lambda}_i = \lambda \circ \boldsymbol{\varphi}_i, \quad i \in \mathbb{N} \ , \tag{6}$$

be stochastic processes giving respectively the reward and the discount factor at stage i. Then the payoff mapping $u_\lambda : \mathcal{H}^\omega \to \mathbb{R}$ of *discounted priority games* is defined as

$$u_\lambda = \sum_{i=0}^{\infty} \boldsymbol{\lambda}_0 \cdots \boldsymbol{\lambda}_{i-1}(1 - \boldsymbol{\lambda}_i)\boldsymbol{r}_i + (\prod_{i=0} \boldsymbol{\lambda}_i) \cdot r(\liminf_n \boldsymbol{\varphi}_n) \ . \tag{7}$$

Thus u_λ is composed of two parts, the *discount part*

$$u_\lambda^{\mathrm{disc}} = \sum_{i=0}^{\infty} \boldsymbol{\lambda}_0 \cdots \boldsymbol{\lambda}_{i-1}(1-\boldsymbol{\lambda}_i)\boldsymbol{r}_i \tag{8}$$

and the *limit part*

$$u_\lambda^{\lim} = (\prod_{i=0}^{\infty} \boldsymbol{\lambda}_i) \cdot r(\liminf_n \boldsymbol{\varphi}_n) \ . \tag{9}$$

Some remarks concerning this definition are in order. Let

$$T = \inf\{i \mid \boldsymbol{\lambda}_j = 1 \text{ for all } j \geq i\} \ . \tag{10}$$

Since, by convention, the infimum of the empty set is ∞, $\{T = \infty\}$ consists of of all infinite histories $h \in \mathcal{H}^\omega$ for which $\boldsymbol{\lambda}_i < 1$ for infinitely many i. Thus we can rewrite u_λ as:

$$u_\lambda = \sum_{i<T} \boldsymbol{\lambda}_0 \cdots \boldsymbol{\lambda}_{i-1}(1-\boldsymbol{\lambda}_i)\boldsymbol{r}_i + (\prod_{i<T} \boldsymbol{\lambda}_i) \cdot r(\liminf_n \boldsymbol{\varphi}_n) \ . \tag{11}$$

Moreover, if $T = \infty$ then the product $\prod_{i<T} \boldsymbol{\lambda}_i$, containing infinitely many factors smaller than 1, is equal to 0 and for such infinite histories the limit part $u_\lambda^{\lim}$ disappears while the discount part is (a sum of) an infinite series. The other extreme case is $T = 0$, i.e. when the discount factor is 1 for all visited states. Then it is the the discount part that disappears from 11 and the payoff is just $r(\liminf_n \boldsymbol{\varphi}_n)$, the same as for priority games of Section 3.

Let $(\mathcal{A}, u_\lambda)$ be a discounted priority game on $\mathcal{A}$. As explained in the preceding section, a stopping priority game on $\mathcal{A}$ with stopping probabilities given by means of λ can be identified with the priority game $(\mathcal{A}_\lambda^\sharp, u)$ on the transformed arena $\mathcal{A}_\lambda^\sharp$. As noted also in the preceding section, there is a natural correspondence allowing to identify strategies in both arenas. We shall note by $\mathbb{P}_{\sigma,\tau}^{\sharp s}$ the probability generated by strategies σ and τ on $\mathcal{A}_\lambda^\sharp$ and $\mathbb{P}_{\sigma,\tau}^{s}$ the similar probability generated by the same strategies on $\mathcal{A}$. The corresponding expectations are denoted $\mathbb{E}_{\sigma,\tau}^{\sharp s}$ and $\mathbb{E}_{\sigma,\tau}^{s}$. Having all this facts in mind, the following proposition shows that stopping priority games and discounted priority games are equivalent in the sense that the same strategies yield the same payoffs in both games:

Proposition 2. *For all strategies σ, τ of players 1, 2 and all states $s \in S$, $\mathbb{E}_{\sigma,\tau}^{\sharp s}[u] = \mathbb{E}_{\sigma,\tau}^{s}[u_\lambda]$.*

Proof. (sketch) Let $T = \inf\{i \mid \mathbf{S}_i \in S^\sharp\}$ be the first moment in the game $(\mathcal{A}_\lambda^\sharp, u)$ when we enter a stopping state. Direct calculations show that $\mathbb{P}_{\sigma,\tau}^{\sharp s}(\mathbf{S}_{i+1} = s_{i+1} | \mathbf{S}_0 = s_0, \ldots, \mathbf{S}_i = s_i) = \lambda(\varphi(s_i))\mathbb{P}_{\sigma,\tau}^{s}(\mathbf{S}_{i+1} = s_{i+1} | \mathbf{S}_0 = s_0, \ldots, \mathbf{S}_i = s_i)$ if all states $s_0, \ldots, s_i, s_{i+1}$ are not stopping. This can be used to show that[1] $\mathbb{E}_{\sigma,\tau}^{\sharp s}[u; T = \infty] = \mathbb{E}_{\sigma,\tau}^{s}[u_\lambda^{\lim}]$. On the other hand, $\mathbb{E}_{\sigma,\tau}^{\sharp s}[u; T = m] = \mathbb{E}_{\sigma,\tau}^{s}[\boldsymbol{\lambda}_0 \cdots \boldsymbol{\lambda}_{m-1}(1-\boldsymbol{\lambda}_m)\boldsymbol{r}_m]$, implying $\mathbb{E}_{\sigma,\tau}^{\sharp s}[u; T < \infty] = \mathbb{E}_{\sigma,\tau}^{s}[u_\lambda^{\mathrm{disc}}]$.

□

[1] By $\mathbb{E}_{\sigma,\tau}^{\sharp s}[u; A]$ we denote the integral of u over the set A.

We can note that in the special case when all discount factors are strictly smaller than 1 (i.e. all stopping probabilities are greater than 0) Proposition 2 reduces to a well-known folklore fact: stopping (Shapley) games[12] and discounted games are equivalent.

6 Limits of priority discounted games

The main aim of this section is to prove that discounted priority games $(\mathcal{A}, u_\lambda)$ admit pure stationary optimal strategies for both players. Of course, due to Shapley's theorem, we already know that this is true for discounted mappings λ such that $\lambda(i) < 1$ for all priorities i. Our proof will use in an essential way the concept of uniformly optimal strategies, which is of independent interest.

Let $\lambda_1, \ldots, \lambda_m$, $1 \leq m \leq k$, be a sequence of constants, all belonging to the right-open interval $[0, 1)$. Let λ be the following discount mapping:

$$\text{for all } i \in [1..k], \quad \lambda(i) = \begin{cases} \lambda_i & \text{if } i \leq m, \\ 1 & \text{if } i > m. \end{cases} \tag{12}$$

In the sequel we shall write $u^{(k)}_{\lambda_1,\ldots,\lambda_m}$ to denote the discounted priority payoff mapping u_λ, where λ is given by (12). (Note, however, that one should not confuse $\lambda_1, \lambda_2, \ldots$ which are used to denote real numbers from $[0, 1)$ with bold $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots$ that are used to denote a stochastic process (6)).

It is worth noticing that in fact we can limit ourselves to discounted priority payoff mappings of the form $u^{(k)}_{\lambda_1,\ldots,\lambda_m}$. Let us say that $\lambda : \{1, \ldots, k\} \to [0, 1]$ is regular if, for each i, $\lambda(i) = 1$ implies $\lambda(j) = 1$ for all $j > i$. Let λ be any discount mapping and let $\pi : \{1, \ldots, k\} \to \{1, \ldots, k\}$ the unique permutation of $\{1, \ldots, k\}$ such $\pi(i) < \pi(j)$ iff one of the following conditions holds: (A) $i < j$ and $\lambda(i) = \lambda(j) = 1$, (B) $i < j$ and $\lambda(i) < 1$ and $\lambda(j) < 1$, (C) $\lambda(i) < \lambda(j) = 1$. Define λ' by setting $\lambda'(\pi(i)) = \lambda(i)$. Then λ' is regular (because of (C)) and for each $h \in \mathcal{H}^\omega$, $u_\lambda(h) = u_{\lambda'}(h)$.

Which strategies are optimal in the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_m})$ usually depends heavily on the discount factors $\lambda_1, \ldots, \lambda_m$. But, in an important paper [1] Blackwell observed that in discounted Markov decision processes optimal strategies are independent of the discount factor if this factor is close to 1. This leads to the concept of uniformly optimal strategies:

Definition 3. *Let $\mathcal{A}$ be a finite arena. Let us fix values of the first $m-1$ discount factors $\lambda_1, \ldots, \lambda_{m-1} \in [0, 1)$. Strategies σ, τ for players 1, 2 are said to be* uniformly optimal *for $\lambda_1, \ldots, \lambda_{m-1}$ if there exists an $\epsilon > 0$ (that can depend on $\lambda_1, \ldots, \lambda_{m-1}$) such that σ, τ are optimal for all games $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m})$ with $1 - \epsilon < \lambda_m < 1$.*

Now we are prepared to announce the main result of the paper:

Theorem 4. *For each $m \in \{1, \ldots, k\}$ the games $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m})$ admit pure stationary uniformly optimal strategies for both players. Moreover, if $(\sigma^\sharp, \tau^\sharp)$ is a pair of such strategies then $\sigma^\sharp, \tau^\sharp$ are also optimal in the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1}})$.*

Proposition 5 below establishes the following chain of implications:
if $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_m})$ admits pure stationary optimal strategies then $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_m})$ admits

pure stationary *uniformly* optimal strategies which in turn implies that $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1}})$ admits pure stationary optimal strategies. Since, by Shapley's theorem, $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_k})$ has pure stationary optimal strategies, trivial backward induction on m will yields immediately Theorem 4.

Proposition 5. *Let $\mathcal{A}$ be a finite arena with states labelled by priorities from $\{1,\ldots,k\}$. Let $m \in \{1,\ldots,k\}$ and $\lambda_1,\ldots,\lambda_{m-1}$ be a sequence of discount factors for priorities $1,\ldots,m$, all belonging to the interval $[0,1)$. Suppose that the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_m})$ has pure stationary strategies for both players. Then the following conditions hold:*

(i) for both players there exist pure stationary uniformly optimal strategies in the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m})$,

(ii) there exists an $\epsilon > 0$ such that, for each pair of pure stationary strategies (σ,τ) for players 1 and 2, whenever σ and τ are optimal in the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m})$ for some $1-\epsilon < \lambda_m < 1$ then σ and τ optimal for all games $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m})$ with $1-\epsilon < \lambda_m < 1$, in particular σ and τ are uniformly optimal,

(iii) if σ, τ are pure stationary uniformly optimal strategies in the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_m})$ then they are optimal in the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1}})$,

(iv)

$$\lim_{\lambda_m \uparrow 1} \mathrm{val}_s(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_m}) = \mathrm{val}_s(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1}}) \ , \tag{13}$$

where $\mathrm{val}_s(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_m})$ is the value of the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_m})$ for an initial state s.

Lemma 6. *Suppose that $\lambda_1,\ldots,\lambda_k$, the discount factors for all priorities, are strictly smaller than 1. Let σ, τ be pure stationary strategies for players 1 and 2 in the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_k})$. Then the expectation $\mathbb{E}^s_{\sigma,\tau}[u^{(k)}_{\lambda_1,\ldots,\lambda_k}]$ is a rational function of $\lambda_1,\ldots,\lambda_n$ bounded on $[0,1)^k$.*

Proof. If we fix pure stationary strategies then we get a finite Markov chain with discounted evaluation. In this context the result is standard, at least for one discount factor, see for example [9]. For several discount factors the proof is identical and given in detail in Appendix C. □

The proof of the following lemma is given in Appendix B.

Lemma 7. *Let $f(x_1,\ldots,x_k)$ be a rational function well-defined and bounded on $[0,1)^k$. Then, for each $0 \leq m < k$, the iterated limit $\lim_{x_{m+1}\uparrow 1} \ldots \lim_{x_k \uparrow 1} f(x_1,\ldots,x_k)$ exists and is finite. Moreover, for every fixed $(x_1,\ldots,x_{m-1}) \in [0,1)^{m-1}$ there exists $\epsilon > 0$ such that the one-variable mapping*

$$x_m \mapsto \lim_{x_{m+1}\uparrow 1} \ldots \lim_{x_k \uparrow 1} f(x_1,\ldots,x_{m-1},x_m,x_{m+1},\ldots,x_k)$$

is rational on the interval $[1-\varepsilon, 1)$.

For any infinite history $h \in \mathcal{H}^\omega$ the value $u^{(k)}_{\lambda_1,\ldots,\lambda_m}(h)$ can be seen as a function of discount factors $\lambda_1, \ldots, \lambda_m$. It turns out that

Lemma 8. *For each $m \in \{1, \ldots, k\}$ and for each $h \in \mathcal{H}^\omega$,*

$$\lim_{\lambda_m \uparrow 1} u^{(k)}_{\lambda_1,\ldots,\lambda_m}(h) = u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1}}(h) \ . \tag{14}$$

The proof Lemma 8 can be found in Appendix A.

Proof of Proposition 5. Since the payoff mappings $u^{(k)}_{\lambda_1,\ldots,\lambda_{i+1}}$ are bounded and Borel-measurable, Lebesgue's dominated convergence theorem and Lemma 8 imply that for all strategies σ and τ for players 1 and 2

$$\lim_{\lambda_{i+1}\uparrow 1} \mathbb{E}^s_{\sigma,\tau}(u^{(k)}_{\lambda_1,\ldots,\lambda_{i+1}}) = \mathbb{E}^s_{\sigma,\tau}(\lim_{\lambda_{i+1}\uparrow 1} u^{(k)}_{\lambda_1,\ldots,\lambda_{i+1}}) = \mathbb{E}^s_{\sigma,\tau}(u^{(k)}_{\lambda_1,\ldots,\lambda_i}) \ . \tag{15}$$

Iterating we get

$$\lim_{\lambda_{m+1}\uparrow 1} \ldots \lim_{\lambda_k \uparrow 1} \mathbb{E}^s_{\sigma,\tau}(u^{(k)}_{\lambda_1,\ldots,\lambda_k}) = \mathbb{E}^s_{\sigma,\tau}(\lim_{\lambda_{m+1}\uparrow 1} \ldots \lim_{\lambda_k \uparrow 1} u^{(k)}_{\lambda_1,\ldots,\lambda_k}) = \mathbb{E}^s_{\sigma,\tau}(u^{(k)}_{\lambda_1,\ldots,\lambda_m}) \ . \tag{16}$$

Suppose now that strategies σ and τ are pure stationary.Then, by Lemma 6, the mapping

$$[0,1)^k \ni (\lambda_1, \ldots, \lambda_k) \mapsto \mathbb{E}^s_{\sigma,\tau}(u^{(k)}_{\lambda_1,\ldots,\lambda_k})$$

is rational and bounded. Lemma 7 applied to the left hand side of (16) allows us to deduce that, for fixed $\lambda_1, \ldots, \lambda_{m-1}$, the mapping

$$(0,1) \ni \lambda_m \mapsto \mathbb{E}^s_{\sigma,\tau}(u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m}) \tag{17}$$

is a rational mapping (of λ_m) for λ_m sufficiently close to 1.

For pure stationary strategies σ and $\sigma^\sharp$ for player 1 and τ, $\tau^\sharp$ for player 2 and fixed discount factors $\lambda_1, \ldots, \lambda_{m-1}$ we consider the mapping

$$[0,1) \ni \lambda_m \mapsto \Phi_{\sigma^\sharp,\tau^\sharp,\sigma,\tau}(\lambda_m) := \mathbb{E}^s_{\sigma^\sharp,\tau^\sharp}(u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m}) - \mathbb{E}^s_{\sigma,\tau}(u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m}) \ .$$

As a difference of rational mappings, all mappings $\Phi_{\sigma^\sharp,\tau^\sharp,\sigma,\tau}$ are rational for λ_m sufficiently close to 1. Since rational mappings are continuous and have finitely many zeros, for each $\Phi_{\sigma^\sharp,\tau^\sharp,\sigma,\tau}$ we can find $\epsilon > 0$ such that $\Phi_{\sigma^\sharp,\tau^\sharp,\sigma,\tau}$ does not change the sign for $1 - \epsilon < \lambda_m < 1$, i.e.

$$\begin{aligned} &\forall \lambda_m \in (1-\epsilon, 1), \\ &\quad \Phi_{\sigma^\sharp,\tau^\sharp,\sigma,\tau}(\lambda_m) \geq 0, \quad \text{or} \quad \Phi_{\sigma^\sharp,\tau^\sharp,\sigma,\tau}(\lambda_m) = 0, \quad \text{or} \quad \Phi_{\sigma^\sharp,\tau^\sharp,\sigma,\tau}(\lambda_m) \leq 0 \ . \end{aligned} \tag{18}$$

Moreover, since there is only a finite number of pure stationary strategies, we can choose in (18) the same ϵ for all mappings $\Phi_{\sigma^\sharp,\tau^\sharp,\sigma,\tau}$, where $\sigma, \sigma^\sharp$ range over pure stationary strategies of player 1 while $\tau, \tau^\sharp$ range over pure stationary strategies of player 2.

Suppose that $\sigma^\sharp$, $\tau^\sharp$ are optimal pure stationary strategies in the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m})$ for some $\lambda_m \in (1-\epsilon, 1)$. This means that for all strategies σ, τ for both players

$$\mathbb{E}^s_{\sigma,\tau^\sharp}(u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m}) \leq \mathbb{E}^s_{\sigma^\sharp,\tau^\sharp}(u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m}) \leq \mathbb{E}^s_{\sigma^\sharp,\tau}(u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m}) \ . \qquad (19)$$

For pure stationary strategies σ, τ, Eq. (19) is equivalent with $\Phi_{\sigma^\sharp,\tau^\sharp,\sigma,\tau^\sharp}(\lambda_m) \geq 0$ and $\Phi_{\sigma^\sharp,\tau^\sharp,\sigma^\sharp,\tau}(\lambda_m) \leq 0$. However, if these two inequalities are satisfied for some λ_m in $(1-\epsilon, 1)$ then they are satisfied for all such λ_m, i.e. (19) holds for all λ_m in $(1-\epsilon, 1)$ for all all pure stationary strategies σ, τ. (Thus, intuitively, we have proved that $\sigma^\sharp$ and $\tau^\sharp$ are optimal for all λ_m in $(1-\epsilon, 1)$ but only if we restrict ourselves to the class of pure stationary strategies.)

But we have assumed that for each λ_m the game $(\mathcal{A}, u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1},\lambda_m})$ has optimal pure stationary strategies (now we take into account all strategies), and under this assumption it is straightforward to prove that if (19) holds for all pure stationary strategies σ, τ then it holds for all strategies σ, τ, i.e. $\sigma^\sharp$ and $\tau^\sharp$ are optimal in the class of all strategies and for all $\lambda_m \in (1-\epsilon, 1)$. In this way we have proved conditions (i) and (ii) of Proposition 5.

Applying the limit $\lambda_m \uparrow 1$ to (19) and taking into account (15) we get

$$\mathbb{E}^s_{\sigma,\tau^\sharp}(u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1-1},\lambda_{m-1}}) \leq \mathbb{E}^s_{\sigma^\sharp,\tau^\sharp}(u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1-1},\lambda_{m-1}}) \leq \mathbb{E}^s_{\sigma^\sharp,\tau}(u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1e-1},\lambda_{m-1e}}) \ ,$$

which proves condition (iii) of the thesis. It is obvious that this implies also (iv). □

References

[1] D. Blackwell. Discrete dynamic programming. *Annals of Mathematical Statistics*, 33:719–726, 1962.

[2] K. Chatterejee, M. Jurdziński, and T.A. Henzinger. Quantitative stochastic parity games. In *Proceedings of the 15th Annual Symposium on Discrete Algorithms SODA*, pages 114–123, 2004.

[3] L. de Alfaro. *Formal Verification of Probabilistic Systems*. PhD thesis, Stanford University, december 1997.

[4] L. de Alfaro, T. A. Henzinger, and Rupak Majumdar. Discounting the future in systems theory. In *ICALP 2003*, volume 2719 of *LNCS*, pages 1022–1037. Springer, 2003.

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

[6] E.A. Emerson and C. Jutla. Tree automata, μ-calculus and determinacy. In *FOCS'91*, pages 368–377. IEEE Computer Society Press, 1991.

[7] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.

[8] H. Gimbert and W. Zielonka. Deterministic priority mean-payoff games as limits of discounted games. In *ICALP 2006*, volume 4052, part II of *LNCS*, pages 312–323. Springer, 2006.

[9] A. Hordijk and A.A. Yushkevich. Blackwell optimality. In E.A. Feinberg and A. Schwartz, editors, *Handbook of Markov Decision Processes*, chapter 8. Kluwer, 2002.

[10] A.K. McIver and C.C. Morgan. Games, probability and the quantitative μ-calculus qmu. In *Proc. LPAR*, volume 2514 of *LNAI*, pages 292–310. Springer, 2002. full version `arxiv.org/abs/cs.LO/0309024`.

[11] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. The MIT Press, 2002.

[12] L. S. Shapley. Stochastic games. *Proceedings Nat. Acad. of Science USA*, 39:1095–1100, 1953.

[13] D. W. Stroock. *An Introduction to Markov Processes*. Springer, 2005.

[14] W. Zielonka. Perfect-information stochastic parity games. In *FOSSACS 2004*, volume 2987 of *LNCS*, pages 499–513. Springer, 2004.

A Appendix

This appendix is devoted to the proof of Lemma 8.

Lemma 9. *Let (a_i) be a sequence of real numbers such that $\lim_{i\to\infty} a_i = 0$. Let*

$$f(\lambda) = (1-\lambda)\sum_{i=0}^{\infty} \lambda^i a_i, \quad \lambda \in [0,1)$$

Then $\lim_{\lambda\uparrow 1} f(\lambda) = 0$.

Proof. Take any $\epsilon > 0$. Since a_i tend to 0 there exists k such that $|a_i| < \epsilon/2$ for all $i > k$. Thus

$$|f(\lambda)| \leq (1-\lambda)\sum_{i=0}^{k} \lambda^i |a_i| + (1-\lambda)\sum_{i=k+1}^{\infty} \lambda^i(\epsilon/2) = (1-\lambda)A + \epsilon/2 \ ,$$

where $A = \max\{a_i \mid 0 \leq i \leq k\}$. For λ sufficiently close to 1, $(1-\lambda)A < \epsilon/2$. Thus $|f(\lambda)| < \epsilon$ for λ close to 1 and since ϵ can be chosen arbitrarily small we get the thesis. □

Let us recall Lemma 8:

Lemma. *For each $m \in \{1,\ldots,k\}$ and for each $h \in \mathcal{H}^\omega$,*

$$\lim_{\lambda_m\uparrow 1} u^{(k)}_{\lambda_1,\ldots,\lambda_m}(h) = u^{(k)}_{\lambda_1,\ldots,\lambda_{m-1}}(h) \ . \tag{20}$$

Proof. Let $u^{(k)}_{\lambda_1,\ldots,\lambda_m} = u^{\mathrm{disc}}_{\lambda_1,\ldots,\lambda_m} + u^{\lim}_{\lambda_1,\ldots,\lambda_m}$ be the decomposition of $u^{(k)}_{\lambda_1,\ldots,\lambda_m}$ onto the discount and limit parts. Let $\lambda : \{1,\ldots,k\} \to [0,1]$, $\lambda^\star : \{1,\ldots,k\} \to [0,1]$, be discount factor mappings such that for each priority $i \in \{1,\ldots,k\}$,

$$\lambda(i) = \begin{cases} \lambda_i & \text{if } i \le m, \\ 1 & \text{if } i > m, \end{cases}$$

$\lambda^\star(i) = \lambda(i)$ for $i \neq m$ and $\lambda^\star(i) = 1$ for $i = m$. As usually, $\boldsymbol{\lambda}_i = \lambda \circ \boldsymbol{\varphi}_i$ and $\boldsymbol{\lambda}^\star_i = \lambda^\star \circ \boldsymbol{\varphi}_i$ are the corresponding stochastic processes.

We examine three cases:

Case 1: $m < \liminf_i \boldsymbol{\varphi}_i(h)$.

In this case, all priorities appearing infinitely often in the sequence $\boldsymbol{\varphi}_i(h), i = 0, 1, \ldots$ have the corresponding discount factors equal to 1. Thus $T(h) = \min\{j \mid \boldsymbol{\lambda}_l(h) = 1 \quad \text{for all } l \ge j\}$ is finite. Then, cf. (11),

$$u^{\mathrm{disc}}_{\lambda_1,\ldots,\lambda_m}(h) = \sum_{0\le l<T(h)} \boldsymbol{\lambda}_0(h)\cdots\boldsymbol{\lambda}_{l-1}(h)(1-\boldsymbol{\lambda}_l(h))\boldsymbol{r}_l(h) \xrightarrow[\lambda_m\uparrow 1]{} \sum_{0\le l<T(h)} \boldsymbol{\lambda}^\star_0(h)\cdots\boldsymbol{\lambda}^\star_{l-1}(h)(1-\boldsymbol{\lambda}^\star_l(h))\boldsymbol{r}_l(h) = u^{\mathrm{disc}}_{\lambda_1,\ldots,\lambda_{m-1}}(h) \ , \quad (21)$$

since $u^{\mathrm{disc}}_{\lambda_1,\ldots,\lambda_m}(h)$ is just a polynomial of variables $\lambda_1,\ldots,\lambda_m$.

Similarly, $\prod_{l=0}^{\infty} \boldsymbol{\lambda}_l(h) = \prod_{0\le l=0}^{T(h)} \boldsymbol{\lambda}_l(h)$ tends to $\prod_{0\le l=0}^{T(h)} \boldsymbol{\lambda}^\star_l(h)$ with $\lambda_m \uparrow 1$, implying

$$\lim_{\lambda_m\uparrow 1} u^{\lim}_{\lambda_1,\ldots,\lambda_m}(h) = u^{\lim}_{\lambda_1,\ldots,\lambda_{m-1}}(h) \ .$$

This and (21) yield (20).

Case 2: $m = \liminf_i \boldsymbol{\varphi}_i(h)$.

Since for infinitely many i, $\boldsymbol{\lambda}_i(h) = \lambda_m < 1$, we have $\prod_{i=0}^{\infty} \boldsymbol{\lambda}_i(h) = 0$, and then $u^{\lim}_{\lambda_1,\ldots,\lambda_m}(h) = 0$.

Let

$$T_0(h) := \max_j\{\boldsymbol{\varphi}_j(h) < m\}$$

be the last moment when a priority strictly smaller than m appears in the sequence $\boldsymbol{\varphi}_i(h), i \in \mathbb{N}$, of visited priorities. Notice that $T_0(h) < \infty$ since the priorities appearing infinitely often in $\boldsymbol{\varphi}_i(h), i \in \mathbb{N}$, are greater or equal m. We have

$$\sum_{0\le l\le T_0(h)} \boldsymbol{\lambda}_0(h)\cdots\boldsymbol{\lambda}_{l-1}(h)(1-\boldsymbol{\lambda}_l(h))\boldsymbol{r}_l(h) \xrightarrow[\lambda_m\uparrow 1]{} \sum_{0\le l\le T_0(h)} \boldsymbol{\lambda}^\star_0(h)\cdots\boldsymbol{\lambda}^\star_{l-1}(h)(1-\boldsymbol{\lambda}^\star_l(h))\boldsymbol{r}_l(h) = \sum_{l=0}^{\infty} \boldsymbol{\lambda}^\star_0(h)\cdots\boldsymbol{\lambda}^\star_{l-1}(h)(1-\boldsymbol{\lambda}^\star_l(h))\boldsymbol{r}_l(h) = u^{\mathrm{disc}}_{\lambda_1,\ldots,\lambda_{m-1}}(h) \ , \quad (22)$$

because $\boldsymbol{\varphi}_l(h) \ge m$ for $l > T_0(h)$, implying $\boldsymbol{\lambda}^\star_i(h) = 1$ for all $l > T_0(h)$. We define by induction:

$$T_{i+1}(h) = \min\{j \mid j > T_i(h) \quad \text{and} \quad \boldsymbol{\varphi}_j(h) = m\}, \quad i = 1, 2, \ldots \ .$$

Intuitively, starting from the moment $T_0(h)$ we count the moments when we visit priority m, and then, for $i \geq 1$, $T_i(h)$ gives the moment of the i-th such visit. We have

$$\sum_{l=T_0(h)+1}^{\infty} \boldsymbol{\lambda}_0(h)\cdots\boldsymbol{\lambda}_{l-1}(h)(1-\boldsymbol{\lambda}_l(h))\boldsymbol{r}_l(h) =$$

$$\boldsymbol{\lambda}_0(h)\cdots\boldsymbol{\lambda}_{T_0(h)}\cdot \sum_{l=T_0(h)+1}^{\infty} \boldsymbol{\lambda}_{T_0(h)+1}(h)\ldots\boldsymbol{\lambda}_{l-1}(h)(1-\boldsymbol{\lambda}_l(h))\boldsymbol{r}_l(h) =$$

$$(\prod_{j=0}^{T_0(h)} \boldsymbol{\lambda}_j(h))\cdot[(1-\boldsymbol{\lambda}_{T_1(h)})\boldsymbol{r}_{T_1(h)}+\boldsymbol{\lambda}_{T_1(h)}(1-\boldsymbol{\lambda}_{T_2(h)})\boldsymbol{r}_{T_2(h)}+\boldsymbol{\lambda}_{T_1(h)}\boldsymbol{\lambda}_{T_2(h)}(1-\boldsymbol{\lambda}_{T_3(h)})\boldsymbol{r}_{T_3(h)}+\ldots] \tag{23}$$

where the last equality follows from the fact that, for each $l > T_0(h)$, if $l \notin \{T_1(h), T_h(h), \ldots\}$ then the priority $\boldsymbol{\varphi}_l(h)$ is strictly greater than m and the corresponding discount factor $\boldsymbol{\lambda}_l(h)$ is equal to 1. On the other hand, $\boldsymbol{\lambda}_{T_l(h)} = \lambda_m$ and $\boldsymbol{r}_{T_l(h)} = r(m)$ for all $l = 1, 2, \ldots$. Thus (23) can be written as

$$(\prod_{j=0}^{T_0(h)} \boldsymbol{\lambda}_j(h))\cdot\sum_{l=0}^{\infty}(\lambda_m)^l(1-\lambda_m)r(m) = (\prod_{j=0}^{T_0(h)} \boldsymbol{\lambda}_j(h))\cdot r(m) \xrightarrow[\lambda_m\uparrow 1]{}$$

$$(\prod_{j=0}^{T_0(h)} \boldsymbol{\lambda}^\star_j(h))r(m) = (\prod_{j=0}^{\infty} \boldsymbol{\lambda}^\star_j(h))r(\liminf_i \boldsymbol{\varphi}_i(h)) = u^{\lim}_{\lambda_1,\ldots,\lambda_{m-1}}(h) \ .$$

The limit above and (22) show that

$$\lim_{\lambda_m\uparrow 1} u^{\mathrm{disc}}_{\lambda_1,\ldots,\lambda_m}(h) = u^{\mathrm{disc}}_{\lambda_1,\ldots,\lambda_{m-1}}(h) + u^{\lim}_{\lambda_1,\ldots,\lambda_{m-1}}(h) \ .$$

Case 3: $m > \liminf_i \boldsymbol{\varphi}_i(h)$.
As in the preceding case $u^{\lim}_{\lambda_1,\ldots,\lambda_m}(h) = 0$. Since $m-1 \geq \liminf_i \boldsymbol{\varphi}_i(h)$ also $u^{\lim}_{\lambda_1,\ldots,\lambda_{m-1}}(h) = 0$. Thus it suffices to show that

$$\lim_{\lambda_m\uparrow 1} u^{\mathrm{disc}}_{\lambda_1,\ldots,\lambda_m}(h) = u^{\mathrm{disc}}_{\lambda_1,\ldots,\lambda_{m-1}}(h) \ . \tag{24}$$

For a subset Z of $\mathbb{N}$ let us define

$$f_Z(\lambda_1,\ldots,\lambda_m) = \sum_{i\in Z}(1-\boldsymbol{\lambda}_i(h))\boldsymbol{\lambda}_0(h)\cdots\boldsymbol{\lambda}_{i-1}(h)\boldsymbol{r}_i(h)$$

and consider $f_X(\lambda_1,\ldots,\lambda_m)$ and $f_Y(\lambda_1,\ldots,\lambda_m)$, where

$$X = \{i \mid \boldsymbol{\varphi}_i(h) = m\} \quad \text{and} \quad Y = \mathbb{N}\setminus X \ . \tag{25}$$

We show that

$$\lim_{\lambda_m\uparrow 1} f_X(\lambda_1,\ldots,\lambda_m) = 0 \ . \tag{26}$$

This is obvious if X is finite since $\boldsymbol{\lambda}_i(h) = \lambda_m$ for all $i \in X$ and then $f_X(\lambda_1, \ldots, \lambda_m) = (1-\lambda_m) r(m) \sum_{i \in X} \boldsymbol{\lambda}_0(h) \ldots \boldsymbol{\lambda}_{i-1}(h) \xrightarrow[\lambda_m \uparrow 1]{} 0$.

Suppose that X is infinite. Define a process T_i: $T_0(h) = -1$, $T_{i+1}(h) = \min\{j \mid j > T_i(h) \quad \text{and} \quad \boldsymbol{\varphi}_j(h) = m\}$. Thus $T_i(h), i = 1, 2, \ldots$, gives the time of the i-th visit to a state with priority m. Set $p(h) = \liminf_i \boldsymbol{\varphi}_i(h)$ and define another process[2]:

$$W_i(h) = \sum_{j=0}^{T_i(h)-1} \mathbf{1}_{\{\boldsymbol{\varphi}_j(h) = p(h)\}} \ .$$

Thus $W_i(h)$ gives the number states with priority $p(h)$ that were visited prior to the moment $T_i(h)$. Notice that, for all $i \geq 1$, $\boldsymbol{\lambda}_0(h) \ldots \boldsymbol{\lambda}_{T_i(h)-1}$ contains $i-1$ factors λ_m and $W_i(h)$ factors $\lambda_{p(h)}$ (and possibly other discount factors) whence $\boldsymbol{\lambda}_0(h) \ldots \boldsymbol{\lambda}_{T_i(h)-1} \leq (\lambda_m)^{i-1} (\lambda_{p(h)})^{W_i(h)}$ implying

$$f_X(\lambda_1, \ldots, \lambda_m) = (1-\lambda_m) r(m) \sum_{i=0}^{\infty} \boldsymbol{\lambda}_0(h) \ldots \boldsymbol{\lambda}_{T_i(h)-1}(h) \leq$$
$$(1-\lambda_m) r(m) \sum_{i=0}^{\infty} (\lambda_{p(h)})^{W_{i+1}(h)} (\lambda_m)^{i-1}$$

Now notice that $\lim_{i \to \infty} W_i(h) = \infty$ since $p(h)$ is visited infinitely often in h. Since $p(h) < m$, we have $\lambda_{p(h)} < 1$ and $\lim_{i \to \infty} (\lambda_{p(h)})^{W_{i+1}(h)} = 0$. Thus Lemma 9 applies and we deduce that (26) holds.

Now let us examine $f_Y(\lambda_1, \ldots, \lambda_m)$. Note that

$$f_Y(\lambda_1, \ldots, \lambda_{m-1}, 1) = \sum_{j \in Y} \boldsymbol{\lambda}^\star_0(h) \cdots \boldsymbol{\lambda}^\star_{j-1}(h)(1 - \boldsymbol{\lambda}^\star_j(h)) \boldsymbol{r}_j(h) =$$
$$\sum_{j=0}^{\infty} \boldsymbol{\lambda}^\star_0(h) \cdots \boldsymbol{\lambda}^\star_{j-1}(h)(1 - \boldsymbol{\lambda}^\star_j(h)) \boldsymbol{r}_j(h) = u^{\mathrm{disc}}_{\lambda_1, \ldots, \lambda_{m-1}}(h) \ ,$$

where the second equality follows from the fact that $\boldsymbol{\lambda}^\star_j(h) = 1$ for $j \in X$. Then

$$\lim_{\lambda_m \uparrow 1} f_Y(\lambda_1, \ldots, \lambda_m) = f_Y(\lambda_1, \ldots, \lambda_{m-1}, 1)$$

follows directly from the well-know Abel's theorem for power series[3]. This fact and (26) yield (24). □

B Appendix

This section is devoted to the proof of Lemma 7.

[2] We use the usual notation, $\mathbf{1}_A$ is the indicator function of an event A, $\mathbf{1}_A(h) = 1$ if $\mathcal{H}^\omega \ni h \in A$ and $\mathbf{1}_A(h) = 0$ otherwise.

[3] Abel's theorem states that for any convergent series $\sum_{i=0}^{\infty} a_i$ of real or complex numbers $\lim_{z \uparrow 1} \sum_{i=0}^{\infty} a_i z^i = \sum_{i=0}^{\infty} a_i$.

For a polynomial $f(x) = \sum_{i=0}^{n} a_i x^i$ we define the order of f

$$\mathrm{ord}(f) = \min\{i \mid a_i \neq 0\} \ . \tag{27}$$

Since min of the empty set is ∞ the order of the zero polynomial is ∞.

The proof of the following elementary observation is left to the reader:

Lemma 10. *Let $f(x) = \sum_{i=0}^{n} a_i x^i$ and $g(x) = \sum_{i=0}^{m} b_i x^i$ be non-zero polynomials with real coefficients such that the rational function $h(x) = \frac{f(x)}{g(x)}$ is bounded on the interval $(0,1)$ (in particular $g \neq 0$). Then*

(1) $\mathrm{ord}(g) \leq \mathrm{ord}(f)$ *and*

(2)

$$\lim_{x \downarrow 0} h(x) = \begin{cases} 0 & \text{if } \mathrm{ord}(g) < \mathrm{ord}(f), \\ \frac{a_k}{b_k} & \text{if } \mathrm{ord}(f) = \mathrm{ord}(g) = k. \end{cases}$$

Proof. Let $f(x) = \sum_{i=m}^{k} a_i x^i$, $g(x) = \sum_{i=p}^{n} b_i x^i$, where $k = \mathrm{ord}(f)$ and $p = \mathrm{ord}(g)$. Then $\frac{f(x)}{g(x)} = x^{m-p} \frac{\sum_{i=m}^{k} a_i x^{i-m}}{\sum_{i=p}^{n} b_i x^{i-p}}$ tends, with $x \downarrow 0$, to (A) 0 whenever $m > p$, (B) $\frac{a_m}{b_p}$ whenever $m = p$, (C) ∞ or $-\infty$, depending on the sign of $\frac{a_m}{b_p}$, whenever $m < k$. Moreover, in the last case $\frac{f(x)}{g(x)}$ is not bounded in the neighborhood of 0. □

For two vectors $(i_1, \ldots, i_n), (j_1, \ldots, j_n) \in \mathbb{N}^n$ of non-negative integers we write $(i_1, \ldots, i_n) \prec (j_1, \ldots, j_n)$ if $(i_1, \ldots, i_n) \neq (j_1, \ldots, j_n)$ and $i_k < j_k$, where $k = \max\{1 \leq l \leq n \mid i_l \neq j_l\}$. Note that $\prec$ is a (strict) total order relation over $\mathbb{N}^n$. The non-strict version of $\prec$ will be denoted $\preceq$.

Let

$$f(x_1, \ldots, x_n) = \sum_{i_1=0}^{k_1} \ldots \sum_{i_n=0}^{k_n} a_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n} \tag{28}$$

be a non-zero multivariate polynomial with real coefficients. We extend the order definition (27) to such polynomials by defining $\mathrm{ord}_{\prec}(f) \in \mathbb{N}^n$ to be the vector $(i_1, \ldots, i_n)$ such that $a_{i_1 \ldots i_n} \neq 0$ and $(i_1, \ldots, i_n) \preceq (j_1, \ldots, j_n)$ for all $(j_1, \ldots, j_n)$ with $a_{j_1 \ldots j_n} \neq 0$. Moreover, we shall write $a_{\mathrm{ord}_{\prec}(f)}$ to denote the coefficient $a_{i_1 \ldots i_n}$, where $(i_1, \ldots, i_n) = \mathrm{ord}_{\prec}(f)$.

As usually, the degree of a monomial $x_1^{i_1} \cdots x_n^{i_n}$ is defined as $\deg(x_1^{i_1} \cdots x_n^{i_n}) = i_1 + \cdots + i_n$ while the degree $\deg(f)$ of a polynomial $f(x_1, \ldots, x_n)$ of (28) is the maximum of the degrees over all monomials with non-zero coefficients $a_{i_1 \ldots i_n}$.

Lemma 11. *Let*

$$f(x_1, \ldots, x_n) = \sum_{i_1=0}^{k_1} \ldots \sum_{i_n=0}^{k_n} a_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n} \tag{29}$$

and

$$g(x_1, \ldots, x_n) = \sum_{i_1=0}^{l_1} \ldots \sum_{i_n=0}^{l_n} b_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n} \tag{30}$$

be non-zero multivariate polynomials such that the rational function $h(x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)}$ is bounded on $(0,1)^n$. Then

(C1) $\mathrm{ord}_\prec(g) \preceq \mathrm{ord}_\prec(f)$,

(C2)

$$\lim_{x_1\downarrow 0}\dots\lim_{x_n\downarrow 0} h(x_1,\dots,x_n) = \begin{cases} 0 & \text{if } \mathrm{ord}_\prec(g) \prec \mathrm{ord}_\prec(f), \\ \frac{a_{i_1\dots i_n}}{b_{i_1\dots i_n}} & \text{if } \mathrm{ord}_\prec(g) = \mathrm{ord}_\prec(f) = (i_1,\dots,i_n), \end{cases}$$

(C3) there exists $\epsilon > 0$ *such that the mapping*

$$x_1 \mapsto h_1(x_1) := \lim_{x_2\downarrow 0}\dots\lim_{x_n\downarrow 0} h(x_1,x_2,\dots,x_n)$$

is rational on the interval $(0,\epsilon)$.

Proof. For an integer p we define a morphism

$$\eta_p : \mathbb{R}[x_1,\dots,x_n] \longrightarrow \mathbb{R}[x]$$

from the ring of n-variable polynomials into the ring of one-variable polynomials by setting $\eta_p(a) = a$ for $a \in \mathbb{R}$ and $\eta_p(x_i) = x^{p^{i-1}}$. Thus for a monomial $x_1^{i_1}\dots x_n^{i_n}$ we have $\eta_p(x_1^{i_1}\dots x_n^{i_n}) = x^{i_1+i_2*p+i_3*p^2+\dots+i_n*p^{n-1}}$ and the image of a polynomial $f(x_1,\dots,x_n)$ of the form (29) is a one-variable polynomial $\eta_p(f)(x) = \sum_{i_1=0}^{k_1}\dots\sum_{i_n=0}^{k_n} a_{i_1\dots i_n}\eta_p(x_1^{i_1}\dots x_n^{i_n})$.

Now note that for any two monomials $x_1^{i_1}\dots x_n^{i_n}$ and $x_1^{j_1}\dots x_n^{j_n}$ and each p such that $i_1+\dots+i_n \le p$ and $j_1+\dots+j_n \le p$ we have $(i_1,\dots,i_n) \prec (j_1,\dots,j_n)$ if and only if $\deg(\eta_p(x_1^{i_1}\dots x_n^{i_n})) = i_1+\dots+i_n*p^{n-1} < j_1+\dots+j_n*p^{n-1} = \deg(\eta_p(x_1^{j_1}\dots x_n^{j_n}))$.

Therefore, for f,g as in (29) and (30), taking $p = \max\{\deg(f),\deg(g)\}+1$, we have $\mathrm{ord}_\prec(f) \prec \mathrm{ord}_\prec(g)$ iff $\mathrm{ord}(\eta_p(f)) < \mathrm{ord}(\eta_p(g))$.

Finally note that if $\frac{f}{g}$ is bounded on $(0,1)^n$ then also the rational one-variable function $\frac{\eta_p(f)}{\eta_p(g)}$ is bounded on $(0,1)$.

The last two remarks and Lemma 10 imply that condition (C1) of Lemma 11 holds.

We shall now prove conditions (C2) and (C3) by induction on the number n of variables. If $n = 1$ then (C2) is given by Lemma 10 while (C3) is void.

Thus suppose that (C2) holds for $n-1$ variables.

Defining one-variable polynomials

$$f_{i_2\dots i_n}(x_1) = \sum_{i_1=0}^{k_1} a_{i_1 i_2\dots i_n} x_1^{i_1}, \quad 0 \le i_2 \le k_2,\dots,0 \le i_n \le k_n \tag{31}$$

and

$$g_{j_2\dots j_n}(x_1) = \sum_{j_1=0}^{l_1} b_{j_1 j_2\dots j_n} x_1^{j_1}, \quad 0 \le j_2 \le l_2,\dots,0 \le j_n \le l_n, \tag{32}$$

we can rewrite f and g as

$$f(x_1,\dots,x_n) = \sum_{i_2=0}^{k_2}\cdots\sum_{i_n=0}^{k_n} (f_{i_2\dots i_n}(x_1))\cdot x_2^{i_2}\cdots x_n^{i_n} \tag{33}$$

and

$$g(x_1,\dots,x_n)=\sum_{j_2=0}^{l_2}\cdots\sum_{j_n=0}^{l_n}(g_{j_2\dots j_n}(x_1))\cdot x_2^{j_2}\cdots x_n^{j_n}\ . \tag{34}$$

For a fixed value of $a\in(0,1)$ we consider polynomials f^a and g^a of $n-1$ variables $x_2,\dots,x_n$ defined as:

$$(x_2,\dots,x_n)\mapsto f^a(x_2,\dots,x_n):=f(a,x_2,\dots,x_n)\ ,$$
$$(x_2,\dots,x_n)\mapsto g^a(x_2,\dots,x_n):=g(a,x_2,\dots,x_n)\ .$$

(Thus here a is considered as a parameter, for different values of a we have different polynomials f^a and g^a.)

Thus

$$f^a(x_2,\dots,x_n)=\sum_{i_2=0}^{k_2}\cdots\sum_{i_n=0}^{k_n}f_{i_2\dots i_n}(a)\cdot x_2^{i_2}\cdots x_n^{i_n}$$

and

$$g^a(x_2,\dots,x_n)=\sum_{j_2=0}^{l_2}\cdots\sum_{j_n=0}^{l_n}g_{j_2\dots j_n}(a)\cdot x_2^{j_2}\cdots x_n^{j_n}\ .$$

The order $\mathrm{ord}_\prec(f^a)$ of the polynomials $f^a(x_2,\dots,x_n)$ can vary with the value of the parameter a, depending on whether a is a zero of polynomials $f_{i_2\dots i_n}(x_1)$. A similar remark is valid for g^a.

Let us define

$$A_f=\{(i_2,\dots,i_n)\mid f_{i_2\dots i_n}\not\equiv 0\}\ ,$$
$$A_g=\{(j_2,\dots,j_n)\mid g_{j_2\dots j_n}\not\equiv 0\}\ ,$$

where the notation $h\not\equiv 0$ means that h is not a zero-polynomial. (This should not be confused with $h(x_1,\dots,x_n)\neq 0$ which means that the value of h is different from 0 for a given argument $(x_1,\dots,x_n)$.)

Now since one-variable polynomials have a finite number of zeros and since the sets A_f and A_g are finite, there exists $\epsilon>0$ such that all the polynomials $f_{i_2\dots i_n}$, $(i_2,\dots,i_n)\in A_f$, and $g_{j_2\dots j_n}$, $(j_2,\dots,j_n)\in A_g$, have no zeros on the interval $(0,\epsilon)$. This means that if the parameter $x_1=a$ is in the interval $(0,\epsilon)$ then $\mathrm{ord}_\prec(f^a)$ and $\mathrm{ord}_\prec(g^a)$ do not depend on the value a and in fact we have

$$\mathrm{ord}_\prec(f^a)=\min_\prec A_f\quad\text{and}\quad\mathrm{ord}_\prec(g^a)=\min_\prec A_g\ ,$$

where $\min_\prec$ means that the minimum is taken the order $\prec$ over $\mathbb{N}^{n-1}$.

Thus suppose that $a\in(0,\epsilon)$. By (C1) applied to the rational mapping $(x_2,\dots,x_n)\mapsto\frac{f^a(x_2,\dots,x_n)}{g^a(x_2,\dots,x_n)}$ we obtain that

(A) either $\mathrm{ord}_\prec(g^a)\prec\mathrm{ord}_\prec(f^a)$ and then $\lim_{x_2\downarrow 0}\dots\lim_{x_n\downarrow 0}\frac{f^a(x_2,\dots,x_n)}{g^a(x_2,\dots,x_n)}=0$

(B) or $\mathrm{ord}_\prec(g^a)=\mathrm{ord}_\prec(f^a)=(m_2,\dots,m_n)$ and then $\lim_{x_2\downarrow 0}\dots\lim_{x_n\downarrow 0}\frac{f^a(x_2,\dots,x_n)}{g^a(x_2,\dots,x_n)}=\frac{f_{m_2\dots m_n}(a)}{g_{m_2\dots m_n}(a)}$.

Since $\lim_{x_2\downarrow 0}\ldots\lim_{x_n\downarrow 0}\frac{f(a,x_2,\ldots,x_n)}{g(a,x_2,\ldots,x_n)}=\lim_{x_2\downarrow 0}\ldots\lim_{x_n\downarrow 0}\frac{f^a(x_2,\ldots,x_n)}{g^a(x_2,\ldots,x_n)}$, in (A) as well as in (B) we get that (C3) holds, i.e. this iterated limit is a rational function of $x_1=a$ whenever x_1 smaller than ϵ.

Again suppose that $a\in(0,\epsilon)$ and $\mathrm{ord}_\prec(f^a)=(m_2,\ldots,m_n)$. Then $\mathrm{ord}_\prec(f)=(m_1,m_2,\ldots,m_n)$, where $m_1=\mathrm{ord}(f_{m_2\ldots m_n}(x_1))$. The order $\mathrm{ord}_\prec(g)$ can be obtained i a similar way.

This implies that one of the following cases holds:

- either $\mathrm{ord}_\prec(g^a)\prec\mathrm{ord}_\prec(f^a)$, which implies that $\mathrm{ord}_\prec(g)\prec\mathrm{ord}_\prec(f)$ and then, by (A),
$$\lim_{x_1\downarrow 0}\lim_{x_2\downarrow 0}\ldots\lim_{x_n\downarrow 0}\frac{f(x_1,x_2,\ldots,x_n)}{g(x_1,x_2,\ldots,x_n)}=\lim_{x_1\downarrow 0}0=0\ ,$$
- or $\mathrm{ord}_\prec(g^a)=\mathrm{ord}_\prec(f^a)=(m_2,\ldots,m_n)$. Let $p_1=\mathrm{ord}_\prec(f_{m_2\ldots m_n})$ and $q_1=\mathrm{ord}_\prec(g_{m_2\ldots m_n})$. Then $\mathrm{ord}_\prec(f)=(p_1,m_2,\ldots,m_n)$ and $\mathrm{ord}_\prec(g)=(q_1,m_2,\ldots,m_n)$ and
$$\lim_{x_1\downarrow 0}\lim_{x_2\downarrow 0}\ldots\lim_{x_n\downarrow 0}\frac{f(x_1,x_2,\ldots,x_n)}{g(x_1,x_2,\ldots,x_n)}=$$
$$\lim_{a\downarrow 0}\frac{f_{m_2\ldots m_n}(a)}{g_{m_2\ldots m_n}(a)}=\begin{cases}0 & \text{if } q_1<p_1,\\ \frac{a_{m_1m_2\ldots m_n}}{b_{m_1m_2\ldots m_n}} & \text{if } p_1=q_1=m_1.\end{cases}$$

This ends the proof of (C2). □

Lemma 7 follows immediately from Lemma 11 since, for a rational function h, $\lim_{x_m\uparrow 1}\ldots\lim_{x_k\uparrow 1}h(x_1,\ldots,x_{m-1},x_m,\ldots,x_k)=\lim_{x_m\downarrow 0}\ldots\lim_{x_k\downarrow 0}h(x_1,\ldots,x_{m-1},1-x_m,\ldots,1-x_k)$.

C Appendix

Proof of Lemma 6

Proof. For each state s set $\lambda(s):=\lambda_i$ and $r(s):=r(i)$, where i the priority of s (i.e. $\lambda(s)$ and $r(s)$ are discount factor and reward associates with s). Let

$$M^\lambda_{\sigma,\tau}[s,s']=\begin{cases}\lambda(s)\cdot p(s'|s,\sigma(s)) & \text{if } s\in S_1,\\ \lambda(s)\cdot p(s'|s,\tau(s)) & \text{if } s\in S_2.\end{cases}$$

be a square matrix indexed by states:and a column vector R^λ:

$$\text{for } s\in S,\quad (R^\lambda)[s]=(1-\lambda(s))r(s)\ .$$

Direct verification shows that the s-th entry of the vector $(\sum_{i=0}^\infty(M^\lambda_{\sigma,\tau})^i)\cdot R^\lambda$ is equal to $\mathbb{E}^s_{\sigma,\tau}[\sum_{i=0}^\infty(1-\boldsymbol{\lambda}_i)\boldsymbol{\lambda}_0\cdots\boldsymbol{\lambda}_{i-1}\boldsymbol{r}_i]=\mathbb{E}^s_{\sigma,\tau}[u_\lambda]$ (the limit part of u_λ is 0 in this case). By a standard technique, cf. [13], it can be shown that the matrix $I-M^\lambda_{\sigma,\tau}$ is invertible and

$$(I-M^\lambda_{\sigma,\tau})^{-1}=\sum_{i=0}^\infty(M^\lambda_{\sigma,\tau})^i\ . \tag{35}$$

Since the entries of $I - M^{\lambda}_{\sigma,\tau}$ are polynomial, Cramer's rule from linear algebra show that the elements of the inverse matrix are rational, which ends the proof. The boundedness is immediate since $|\sum_{i=0}^{\infty}(1-\boldsymbol{\lambda}_i)\boldsymbol{\lambda}_0\cdots\boldsymbol{\lambda}_{i-1}\boldsymbol{r}_i| \leq \max_s r(s)$. □