Deterministic priority mean-payoff games as limits of discounted games

Hugo Gimbert, Wieslaw Zielonka

Deterministic priority mean-payoff games as limits of discounted games*

Hugo Gimbert[1] and Wiesław Zielonka[2]

[1] Instytut Informatyki, Warsaw University, Poland
[2] Université Paris 7 and CNRS, LIAFA, case 7014
2, place Jussieu, 75251 Paris Cedex 05, France

Abstract. Inspired by the paper of de Alfaro, Henzinger and Majumdar [2] about discounted μ-calculus we show new surprising links between parity games and different classes of discounted games.

1 Introduction

One of the major results in the theory of stochastic games states that the value of mean-payoff games is the limit of the values of discounted games [7, 4]. Recently de Alfaro, Henzinger and Majumdar [2] presented results that seem to indicate that it is possible to obtain parity games as an appropriate limit of multi-discounted games. In fact, the authors of [2] use the language of the μ-calculus rather than games, but as the links between μ-calculus and parity games are well-known since the advent [3] it is natural to wonder how discounted μ-calculus from [2] can be reflected in games.

Suppose that $\mathcal{A}$ is our arena with each vertex belonging to one of the two players 0 and 1. If the current state s belongs to player P then he chooses an outgoing edge (s, s') and the system moves to the target state s'. Suppose that the states are labeled by priorities from the finite set $\mathbf{D} = \{0, \ldots, k-1\}$. Inspecting thoroughly the formulas of the discounted μ-calculus from [2] it is not too difficult to discover that it corresponds to the following games. Let us associate with each priority $d \in \mathbf{D}$ a discount factor λ_d from the interval $(0; 1)$. Let $d_0 d_1 d_2 \ldots$ be an infinite sequence of priorities visited during the play. Then we calculate the payoff obtained by player 0 from player 1 using the formula

$$\sum_{i=0}^{\infty} \lambda_{d_0} \cdots \lambda_{d_{i-1}} (1 - \lambda_{d_i}) r_i \tag{1}$$

when

$$r_d = \begin{cases} 0 & \text{if the priority } d \text{ is odd} \\ 1 & \text{if the priority } d \text{ is even} \end{cases}$$

* This research was supported by European Research Training Network: Games and Automata for Synthesis and Validation.

The fact that such games have values and optimal strategies, and in the case of perfect information stochastic games even positional optimal strategies, is known since the seminal paper of Shapley [9]. The results of [2] indicate that

$$\lim_{\lambda_0 \uparrow 1} \dots \lim_{\lambda_{k-1} \uparrow 1} \mathrm{val}_\lambda(s) = \mathrm{val}(s) \tag{2}$$

where $\mathrm{val}_\lambda(s)$ is the value of the multi-discounted game with the payoff (1) for the initial state s and $\mathrm{val}(s)$ is the value of the parity game for the initial state s (more precisely we should take in this case the following version of the parity games: player 0 wins 1 if the smallest priority visited infinitely often is even, otherwise he wins 0).

The first point to note is that if we are in the realm of games rather than μ-calculus then it is completely artificial to limit the numbers r_i appearing in (1) to 0 and 1, it would be much more natural to consider the games with any real valued r_i (and this is of course the point of view adopted by Shapley [9]). Thus now we assume that states are labeled rather by pairs $(d, r) \in \mathbf{D} \times \mathbb{R}$ composed of a priority d and a real number r. If during an infinite play we visit the sequence $(d_0, r_0), (d_1, r_1), \dots$ of labels then we can still calculate the payment obtained by player 0 from player 1 using the formula (1). What about the equation (2) in this case? Does there exists a game that replaces the parity game and such that its value can be put on the right hand side of the equality (2)? As it turns out the things go correctly and such games, *priority mean-payoff games*, exist. In fact priority mean-payoff games were previously introduced in [5] where it was proved that they admit optimal positional strategies. In this paper we show that their values are related to the values of multi-discounted games, generalizing [3]the result of [2].

The formula (2) has a rather limited interest, we would prefer to find a link not only between the game values but also between their optimal strategies. To this end in Section 4 we introduce a new family of discounted games: priority discounted games. They have a considerable advantage over multi-discounted games: their values depend on only one parameter, i.e. to find the limits of their values we do need to use iterated limits. And, what is more important, it is possible to carry out to this framework the concept of Blackwell optimality [6]: for all values of the discount factor sufficiently close to 0, the optimal strategies in priority-discounted games are also optimal for priority mean-payoff games. Note that since the parity games are just a very special subclass of priority mean-payoff games this result establishes a rather unexpected property of parity games. Can it be used in practice to calculate optimal strategies for parity games? This is the main open problem.

[3] This is not really exact, since [2] examines the μ-calculus corresponding to perfect information stochastic systems while in our paper we limit ourselves to deterministic games. The full generalization to the stochastic case remains to be done.

2 Games

An *arena* is a tuple $\mathcal{A} = (S_0, S_1, A, \Re)$, where S_0 and S_1 are the sets of *states* controlled by player 0 and player 1 respectively, A is the set of *actions* and $\Re$ is the set of *rewards*.

By $S = S_0 \cup S_1$ we denote the set of all states. Then $A \subseteq S \times \Re \times S$, i.e. each action $a = (s', r, s'') \in A$ is a triple composed of the *source state* $\mathrm{source}(a) = s'$, the *target state* $\mathrm{target}(a) = s''$ and a reward $r = \mathrm{reward}(a) \in \Re$.

An action a is *available* at state s if $a \in A_s$, where A_s denotes the set of actions with source s.

We consider only arenas where the sets of states and actions are finite and such that for each state s the set A_s of available actions is non-empty.

A *path* in arena $\mathcal{A}$ is a finite or infinite sequence $p = a_0 a_1 a_2 \dots$ of actions such that $\forall i, \mathrm{target}(a_i) = \mathrm{source}(a_{i+1})$. The source of the first action a_0 is the source, $\mathrm{source}(p)$, of the path p. If p is finite then the target of the last action is the target, $\mathrm{target}(p)$, of p.

It is convenient to assume that for each state s there is an empty path $\mathbf{1}_s$ with the source and the target s.

Two players 0 and 1 play on $\mathcal{A}$ in the following way. If the current state s is controlled by player $P \in \{0, 1\}$, i.e. $s \in S_P$, then player P chooses an action $a \in A_s$ available at s, this action is executed and the system goes to the state $\mathrm{target}(a)$.

Starting from an initial state s, the infinite sequence of consecutive moves of both players yields an infinite sequence $p = a_0 a_1 \dots$ of executed actions such that $\mathrm{source}(p) = s$. Such sequences are called *plays*, thus plays in this game are just infinite paths in the underlying arena $\mathcal{A}$.

We shall also use the term "a finite play" as a synonym of "a finite path" but "play" without any qualifier will always denote an infinite play.

An infinite sequence $r_0 r_1 r_2 \dots$ of rewards is *finitely generated* if there exists a finite subset $\Re'$ of $\Re$ such that all elements of this sequence belong to $\Re'$. The set of all infinite finitely generated sequences of $\Re$ is denoted $\Re^\omega$.

By $\Re^*$ we denote the set of all finite sequences of $\Re$ and we set $\Re^\infty = \Re^* \cup \Re^\omega$.

Each path $p = a_0 a_1 \dots$ yields a sequence of rewards

$$\mathrm{reward}(p) = \mathrm{reward}(a_0)\,\mathrm{reward}(a_1)\dots \quad . \tag{3}$$

Note that since our arenas are finite, if p is an infinite path then $\mathrm{reward}(p)$ is finitely generated.

A *utility mapping*

$$u : \Re^\omega \to \mathbb{R} \tag{4}$$

maps each finitely generated infinite reward sequence $x \in \Re^\omega$ to a real number $u(x) \in \mathbb{R}$. The interpretation is that at the end of a play p player 0 receives from player 1 the *payoff* $u(\mathrm{reward}(p))$ (if $u(\mathrm{reward}(p)) < 0$ then it is rather player 1 that receives from player 0 the amount $|u(\mathrm{reward}(p))|$).

A *game* $(\mathcal{A}, u)$ is couple composed of an arena and a utility mapping.

A strategy of a player P is his plan of action that tells him which action to take when the game is at a state $s \in S_P$. The choice of the action can depend on the whole past sequence of moves. Thus a *strategy* for player 0 is a mapping

$$\sigma : \{p \mid p \text{ a finite play with } \mathrm{target}(p) \in S_0\} \longrightarrow A \tag{5}$$

such that for each finite play p with $s = \mathrm{target}(p) \in S_0$, $\sigma(p) \in A_s$.

Strategy σ of player 0 is said to be *positional* if for every state $s \in S_0$ and every finite play p such that $\mathrm{target}(p) = s$, $\sigma(p) = \sigma(\mathbf{1}_s)$. Thus the action chosen by a positional strategy depends only on the current state, previously visited states and executed actions are irrelevant. To simplify the notation it is convenient to view a positional strategy as a mapping

$$\sigma : S_0 \to A \tag{6}$$

such that $\sigma(s) \in A_s$.

A finite or infinite play $p = a_0 a_1 \dots$ is said to be *consistent* with a strategy $\sigma \in \Sigma$ if for each $i \in \mathbb{N}$ such that $\mathrm{target}(a_{i-1}) = \mathrm{source}(a_i) \in S_0$, we have $a_i = \sigma(a_0 \dots a_{i-1})$. Moreover, if $s = \mathrm{source}(a_0) \in S_0$ then we require that $a_0 = \sigma(\mathbf{1}_s)$.

Strategies, positional strategies and consistent plays are defined in the analogous way for player 1 with S_1 replacing S_0.

In the sequel Σ and $\mathcal{T}$ will stand for the set of strategies for player 0 and player 1, Σ_p and $\mathcal{T}_p$ are the corresponding subsets of positional strategies and finally σ and τ, possibly with subscripts or superscripts, will denote the elements of Σ and $\mathcal{T}$.

Given a pair of strategies $\sigma \in \Sigma$ and $\tau \in \mathcal{T}$, there exists a unique infinite play in arena $\mathcal{A}$, denoted $p(s, \sigma, \tau)$, consistent with σ and τ and such that $s = \mathrm{source}(p(s, \sigma, \tau))$. The corresponding sequence of rewards $\mathrm{reward}(p(s, \sigma, \tau))$ will be denoted $r(s, \sigma, \tau)$.

Definition 1. *Strategies $\sigma^\sharp \in \Sigma$ and $\tau^\sharp \in \mathcal{T}$ are* optimal *in the game $(\mathcal{A}, u)$ if*

$$\forall s \in S, \forall \sigma \in \Sigma, \forall \tau \in \mathcal{T},$$
$$u(r(s, \sigma, \tau^\sharp)) \le u(r(s, \sigma^\sharp, \tau^\sharp)) \le u(r(s, \sigma^\sharp, \tau)) \ . \tag{7}$$

We say that a utility mapping u admits positional optimal strategies *if for all games $(\mathcal{A}, u)$ over* finite arenas *there exist positional optimal strategies for both players.*

Thus if both strategies are optimal the players do not have any incentive to change them unilaterally: player 0 cannot increase his gain by switching to another strategy σ while player 1 cannot decrease his loses by switching to τ.

Note that zero-sum games, where the gain of one player is equal to the loss of his adversary, satisfy the exchangeability property for optimal strategies: for any two pairs of optimal strategies $(\sigma^\sharp, \tau^\sharp)$ and $(\sigma^\star, \tau^\star)$, the pairs $(\sigma^\star, \tau^\sharp)$ and $(\sigma^\sharp, \tau^\star)$ are also optimal and, moreover, $u(r(s, \sigma^\sharp, \tau^\sharp)) = u(r(s, \sigma^\star, \tau^\star))$, i.e. the value of the expression $u(r(s, \sigma^\sharp, \tau^\sharp))$ is independent of the choice of the optimal strategies — this is *the value of the game $(\mathcal{A}, u)$ at state s.*

Lemma 2. *Let u be a utility mapping admitting optimal positional strategies for both players.*

(A) Suppose that $\sigma \in \Sigma$ is any strategy while $\tau^\sharp \in \mathcal{T}_p$ is positional. Then there exists a positional strategy $\sigma^\sharp \in \Sigma_p$ such that

$$\forall s \in S, \quad u(r(s, \sigma, \tau^\sharp)) \leq u(r(s, \sigma^\sharp, \tau^\sharp)) \ . \tag{8}$$

(B) Similarly, if $\tau \in \mathcal{T}$ is any strategy and $\sigma^\sharp \in \Sigma_p$ a positional strategy then there exists a positional strategy $\tau^\sharp \in \mathcal{T}_p$ such that

$$\forall s \in S, \quad u(r(s, \sigma^\sharp, \tau^\sharp)) \leq u(r(s, \sigma^\sharp, \tau)) \ .$$

Proof. We prove (A), the proof of (B) is similar. Take any strategies $\sigma \in \Sigma$ and $\tau^\sharp \in \mathcal{T}_p$. Let $\mathcal{A}'$ be a subarena of $\mathcal{A}$ obtained by restricting the actions of player 1 to the actions given by the strategy $\tau^\sharp$, i.e. in $\mathcal{A}'$ the only possible strategy for player 1 is the strategy $\tau^\sharp$. The actions of player 0 are not restricted, i.e. in $\mathcal{A}'$ player 0 has the same available actions as in $\mathcal{A}$. Since $\tau^\sharp$ is positional $\mathcal{A}'$ is a well-defined finite arena and by the assumption concerning u there exists an optimal positional strategy $\sigma^\sharp$ for player 0 in $\mathcal{A}'$; obviously $\tau^\sharp$ is the optimal positional strategy for player 1 in $\mathcal{A}'$. This implies that (8) holds in $\mathcal{A}'$ and therefore also in $\mathcal{A}$. □

Lemma 3. *Suppose that the utility mapping u admits optimal positional strategies. Suppose $\sigma^\sharp \in \Sigma_p$ and $\tau^\sharp \in \mathcal{T}_p$ are positional strategies such that*

$$\forall s \in S, \forall \sigma \in \Sigma_p, \forall \tau \in \mathcal{T}_p, \\ u(r(s, \sigma, \tau^\sharp)) \leq u(r(s, \sigma^\sharp, \tau^\sharp) \leq u(r(s, \sigma^\sharp, \tau)) \ , \tag{9}$$

i.e. $\sigma^\sharp$ and $\tau^\sharp$ are optimal in the class of positional strategies. Then $\sigma^\sharp$ and $\tau^\sharp$ are optimal.

Proof. Suppose that

$$\exists \tau \in \mathcal{T}, \quad u(r(s, \sigma^\sharp, \tau)) < u(r(s, \sigma^\sharp, \tau^\sharp)) \ . \tag{10}$$

By Lemma 2 there exists a positional strategy $\tau^\star \in \mathcal{T}_p$ such that $u(r(s, \sigma^\sharp, \tau^\star)) \leq u(r(s, \sigma^\sharp, \tau)) < u(r(s, \sigma^\sharp, \tau^\sharp))$, contradicting (9). Thus $\forall \tau \in \mathcal{T}, u(r(s, \sigma^\sharp, \tau^\sharp)) \leq u(r(s, \sigma^\sharp, \tau))$. The left hand side of (7) can be proved in the similar way. □

3 Priority mean-payoff games as the limit of multi-discounted games

In the sequel of this paper we fix the set of rewards to be

$$\Re = \mathbf{D} \times \mathbb{R} \ , \tag{11}$$

where $\mathbf{D} = \{d \in \mathbb{N} \mid 0 \leq d < k\}$ is fixed finite set of *priorities*. we shall note by $|\mathbf{D}| = k$ the cardinality of $\mathbf{D}$.

3.1 Multi-discounted games

A *discount mapping*

$$\lambda : \mathbf{D} \longrightarrow [0, 1)$$

associates with each priority a real number from the interval $[0, 1)$. The value of λ for a priority $d \in \mathbf{D}$, noted λ_d, is called the discount factor of d.

Given a discount mapping λ we define *multi-discounted utility mapping* u_λ. It is convenient to define u_λ uniformly for infinite as well as for finite reward sequences $t = (d_0, r_0), (d_1, r_1), \ldots \in \Re^\infty$:

$$\begin{aligned} u_\lambda(t) &= (1 - \lambda_{d_0})r_0 + \lambda_{d_0}(1 - \lambda_{d_1})r_1 + \lambda_{d_0}\lambda_{d_1}(1 - \lambda_{d_2})r_2 + \ldots \\ &= \sum_{0 \le i < |t|} \lambda_{d_0} \ldots \lambda_{d_{i-1}}(1 - \lambda_{d_i})r_i \ , \end{aligned} \tag{12}$$

where $|t|$ is the length of t if t is finite and ∞ otherwise.

By an obvious adaptation of the proof of Shapley [9] one can obtain the following theorem which in fact holds even for a more general class of perfect information stochastic games:

Theorem 4 (Shapley). *For each discount mapping $\lambda : \mathbf{D} \to [0; 1)$, the multi-discounted utility mapping u_λ admits optimal positional strategies for both players. In particular each game $(\mathcal{A}, u_\lambda)$ has a value $\mathrm{val}_\lambda(s)$ for every initial state s.*

3.2 Priority mean-payoff games

Definition 5. *The* priority *of an infinite reward sequence $t = (d_0, r_0), (d_1, r_1), \ldots \in \Re^\omega$ is the minimal priority occurring infinitely often in t:*

$$\mathrm{priority}(t) = \liminf_{i \to \infty} d_i \ . \tag{13}$$

For any reward sequence $t = (d_0, r_0), (d_1, r_1), \ldots \in \Re^\infty$ and $d \in \mathbb{N}$ let

$$\Pi_d(t) = \{i \in \mathbb{N} \mid 0 \le i < |t| \quad \text{and} \quad d_i = d\}, \tag{14}$$

be the sequence consisting of the indices for which the priority is equal d in t.

Definition 6. *Let $t = (d_0, r_0), (d_1, r_1), \ldots \in \Re^\omega$ be an infinite reward sequence and let $\Pi_d(t) = i_0, i_1, \ldots$ the sequence consisting of the indices i for which $d_i =$ priority(t). Then*

$$\mu(t) = \liminf_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} r_{i_j}$$

defines priority mean-payoff utility *mapping $\mu : \Re^\omega \to \mathbb{R}$.*

Thus, intuitively, to calculate $\mu(t)$ we first use the priorities to choose an appropriate subsequence $t' = (d_{i_0}, r_{i_0}), (d_{i_1}, r_{i_1}), (d_{i_1}, r_{i_1}), \ldots$ of t consisting of rewards such that priority$(t) = d_{i_0} = d_{i_1} = d_{i_1} = \ldots$ and next we apply the usual mean-payoff to the corresponding subsequence $r_{i_0} r_{i_1} r_{i_2} \ldots$ of rewards.

The following result is proved in [5]:

Theorem 7. *The priority mean-payoff utility μ admits optimal positional strategies for both players.*

The *value of the priority mean-payoff game* for an initial state s will be noted $\mathrm{val}(s)$.

The following theorem connects multi-discount and priority mean-payoff games:

Theorem 8. *Let $\mathbf{D} = \{0, \ldots, k-1\}$ be the set of priorities. Then for each initial state s*

$$\lim_{\lambda_0 \uparrow 1} \lim_{\lambda_1 \uparrow 1} \ldots \lim_{\lambda_{k-1} \uparrow 1} \mathrm{val}_\lambda(s) = \mathrm{val}(s) \ , \tag{15}$$

i.e. the value of the priority mean-payoff game is the (iterated) limit of the value of the multi-discounted game ($\lambda_i \uparrow 1$ means that λ_i tends to 1 from below).

The order in which the limits are taken in (15) does matter and is related to the fact that in (13) we have chosen the minimal priority appearing infinitely often as the priority of an infinite sequence of rewards. Let us note that in the particular case when there is only one priority, $|\mathbf{D}| = 1$, Theorem 8 holds in the much larger setting of stochastic games, this is a seminal result of Mertens and Neyman [7]. We skip the proof of Theorem 8 since it is too long to be given here. In fact Theorem 8 will not be used in the sequel and, in our opinion, the subsequent Section 4 contains much more interesting results which are provided with complete proofs.

4 Priority-discounted games

In Section 3.2 we have established that the value of the priority mean-payoff game is an iterated limit of the multi-discounted game. However, iterated limits are cumbersome so a natural question is if we cannot replace them by a single limit.

Another weakness of multi-discounted games is that they are related to parity mean-payoff games only by their values but not by their optimal strategies.

In this section we introduce the class of priority-discounted games which behave much in this respect.

Let us take $\beta \in (0; 1]$ and, for $d \in \mathbf{D}$, set

$$\lambda_d(\beta) = 1 - \beta^d \ . \tag{16}$$

A priority-discounted game is a multi-discounted game in which the discount factor associated with some priority $d \in \mathbf{D}$ is $\lambda_d(\beta)$. Let $t = (d_0, r_0), (d_1, r_1), \ldots \in$

$\Re^\infty$. Then, putting (16) into (12), we get the definition of the *priority-discounted* utility mapping:

$$\begin{aligned} u^\beta(t) &= \beta^{d_0} r_0 + (1-\beta^{d_0})\beta^{d_1} r_1 + (1-\beta^{d_0})(1-\beta^{d_1})\beta^{d_2} r_2 + \dots \\ &= \sum_{0 \le i < |t|} (1-\beta^{d_0})(1-\beta^{d_1}) \dots (1-\beta^{d_{i-1}})\beta^{d_i} r_i \ . \end{aligned} \tag{17}$$

Let us note that $\lambda_d(\beta) \uparrow 1$ iff $\beta \downarrow 0$. The following theorem is analogous to Theorem 8.

Theorem 9. *Let $\mathcal{A}$ be a finite arena. Then*

(i) For every finite arena $\mathcal{A}$ and for all $\beta \in (0;1]$ both players have optimal positional strategies in the single discounted game $(\mathcal{A}, u^\beta)$.

(ii) Let $\mathrm{val}^\beta(s)$ be the value of the single discounted game $(\mathcal{A}, u^\beta)$ for an initial state s. Then

$$\lim_{\beta \downarrow 0} \mathrm{val}^\beta(s) = \mathrm{val}(s) \ , \tag{18}$$

where $\mathrm{val}(s)$ is the value of the priority mean-payoff game.

Proof. (i) obviously is just a special case of Theorem 4. The proof of (ii) will be given at the end of Section 4.1. □

4.1 Blackwell optimality

The concept known as Blackwell optimality was introduced in [1]. A readable modern presentation can be found in [6]. Roughly speaking, a policy of a Markov decision process with the discounted reward criterion is Blackwell optimal if it is optimal for all discount factors sufficiently close to 1. It turns out that such policies are also automatically optimal for mean-payoff games, hence Blackwell optimality is stronger than the classical concept of optimality in mean-payoff games.

We adapt here the concept of Blackwell optimality to two-person priority-discounted games. We show that corresponding Blackwell optimal strategies exist and that they are optimal for priority mean-payoff games.

Let us fix a finite arena $\mathcal{A}$. Strategies $(\sigma^\sharp, \tau^\sharp) \in \Sigma \times \mathcal{T}$ are *β-optimal* if they are optimal in the priority-discounted game $(\mathcal{A}, u^\beta)$ with the discount factor β.

Definition 10. *Strategies $(\sigma^\sharp, \tau^\sharp) \in \Sigma \times \mathcal{T}$ are Blackwell optimal if they are β-optimal for all values β in an interval $0 < \beta < \beta_0$ for some constant $\beta_0 > 0$.*

The following two lemmas will be useful for establishing the existence of Blackwell optimal strategies in priority-discounted games, stated in Theorem 13. In those Lemmas, we consider the different discounted-priority games obtained when β tends to 0. In Lemma 11 we fix some finite play and describe the asymptotic behavior of the values of this play when β tends to 0. In Lemma 12 we consider the case of ultimately periodic plays.

Lemma 11. *Let $y = (d_0, r_0) \dots (d_n, r_n) \in \Re^*$ be a finite sequence of rewards, $a = \min\{d_0, \dots, d_n\}$ and $I = \{i \mid 0 \le i \le n \quad \text{and} \quad d_i = a\}$. Then*

$$\lim_{\beta \downarrow 0} \frac{u^\beta(y)}{1 - (1-\beta^{d_0}) \cdots (1-\beta^{d_n})} = \frac{1}{|I|} \sum_{i \in I} r_i \ ,$$

where $|I|$ denotes the cardinality of I.

Proof. This is just an elementary exercise: $u^\beta(y) = \beta^{d_0} r_0 + (1-\beta^{d_0})\beta^{d_1} r_1 + (1-\beta^{d_0})(1-\beta^{d_1})\beta^{d_2} r_2 + \ldots + (1-\beta^{d_0})(1-\beta^{d_1}) \cdots (1-\beta^{d_{n-1}})\beta^{d_n} r_n = (\sum_{i \in I} r_i)\beta^a + p(\beta)$, where $p(\beta)$ is a polynomial with all monomials having degree $> a$. Similarly, $g(\beta) = 1 - (1-\beta^{d_0}) \cdots (1-\beta^{d_n}) = |I|\beta^a + q(\beta)$, where $q(\beta)$ is a sum of monomials of degree $> a$. Thus

$$u^\beta(\beta)/g(\beta) = ((\sum_{i \in I} r_i) + p(\beta)/\beta^a)/(|I| + q(\beta)/\beta^a) \xrightarrow{\beta \to 0} (\sum_{i \in I} r_i)/|I| \ .$$

□

Lemma 12. *Given an initial state s and positional strategies $(\sigma, \tau) \in \Sigma_p \times \mathcal{T}_p$,*

(i) the function $\beta \mapsto u^\beta(r(s, \sigma, \tau)))$, defined for $0 < \beta < 1$, is a rational function[4] of β.
(ii) $\lim_{\beta \to 0} u^\beta(r(s, \sigma, \tau))) = \mu(r(s, \sigma, \tau)))$, where μ is the priority mean-payoff utility, see Definition 6.

Proof. (i) Since σ and τ are positional, the play $p(s, \sigma, \tau)$ and the resulting sequence $r(s, \sigma, \tau)$ of rewards are ultimately periodic. Thus, for some $x, y \in \Re^*$, $r(s, \sigma, \tau) = xyyy \ldots = xy^\omega$. Then (17) yields

$$\begin{aligned} u^\beta(xy^\omega) = & \\ & u^\beta(x) + (1-\beta^{d_0}) \cdots (1-\beta^{d_l}) u^\beta(y) \sum_{i=0}^{\infty} \left[(1-\beta^{d_{l+1}}) \cdots (1-\beta^{d_m})\right]^i \\ & = u^\beta(x) + \frac{(1-\beta^{d_0}) \cdots (1-\beta^{d_l})}{1 - (1-\beta^{d_{l+1}}) \cdots (1-\beta^{d_m})} \ u^\beta(y), \quad (19) \end{aligned}$$

where $d_0, \ldots, d_l$ is the sequence of priorities of x and $y = (r_{l+1}, d_{l+1}), \ldots, (r_m, d_m)$.

Since x and y are finite $u^\beta(x)$ and $u^\beta(y)$ are just polynomials of β.

(ii) It suffices to note that in (19), if $\beta \to 0$ then $u^\beta(x)$ tends to 0 while $(1 - \beta^{d_0}) \cdots (1 - \beta^{d_l})$ tends to 1. Thus this result is an immediate consequence of Lemma 11. □

We can now state our main result about Blackwell optimality in priority-discounted games.

[4] a quotient of two polynomials

Theorem 13. *For each finite arena $\mathcal{A}$ there exist Blackwell optimal positional strategies.*

Proof. The proof follows very closely the proof given in [6] for Markov decision processes.

Since $\mathcal{A}$ is finite, the set $\Sigma_p \times \mathcal{T}_p$ of pairs of positional strategies is finite. Thus there exists a pair $(\sigma^\sharp, \tau^\sharp) \in \Sigma_p \times \mathcal{T}_p$ of positional β-optimal strategies for all $\beta = \beta_n$, where (β_n) is some sequence such that $\beta_n \downarrow 0$. We claim that $(\sigma^\sharp, \tau^\sharp)$ are Blackwell optimal.

Suppose the contrary. Then there exists a state s and a sequence γ_n tending to 0 with $n \to \infty$ such that

(i) either there exists a sequence $\sigma_n^\star$ of strategies such that $u^{\gamma_n}(r(s, \sigma^\sharp, \tau^\sharp)) < u^{\gamma_n}(r(s, \sigma_n^\star, \tau^\sharp))$,
(ii) or there exists a sequence $\tau_n^\star$ of strategies such that $u^{\gamma_n}(r(s, \sigma^\sharp, \tau_n^\star)) < u^{\gamma_n}(r(s, \sigma^\sharp, \tau^\sharp))$.

Due to Lemma 2, the strategies $\sigma_n^\star$ and $\tau_n^\star$ can be chosen positional and since the number of positional strategies is finite, taking a subsequence if necessary, we can fix one strategy $\sigma^\star$ and one strategy $\tau^\star$ for all n.

Thus we have obtained that

(1) either there exist a state s, a positional strategy $\sigma^\star \in \Sigma_p$ and a sequence (γ_n), $\gamma_n \downarrow 0$, such that for all n

$$u^\beta(r(s, \sigma^\sharp, \tau^\sharp)) < u^\beta(r(s, \sigma^\star, \tau^\sharp)) \quad \text{for all } \beta = \gamma_1, \gamma_2, \dots \ , \tag{20}$$

(2) or there exist a state s, a positional strategy $\tau^\star \in \mathcal{T}_p$ and a sequence (γ_n), $\gamma_n \downarrow 0$, such that for all n

$$u^\beta(r(s, \sigma^\sharp, \tau^\star)) < u^\beta(r(s, \sigma^\sharp, \tau^\sharp)) \quad \text{for all } \beta = \gamma_1, \gamma_2, \dots \ . \tag{21}$$

Suppose that (20) holds.

The choice of $(\sigma^\sharp, \tau^\sharp)$ guarantees that

$$u^\beta(r(s, \sigma^\star, \tau^\sharp)) \le u^\beta(r(s, \sigma^\sharp, \tau^\sharp)) \quad \text{for all } \beta = \beta_1, \beta_2, \dots \ . \tag{22}$$

Consider the function

$$f(\beta) = u^\beta(r(s, \sigma^\star, \tau^\sharp)) - u^\beta(r(s, \sigma^\sharp, \tau^\sharp)) \ . \tag{23}$$

By Lemma 12, $f(\beta)$ coincides for $0 < \beta < 1$ with a rational function of the variable β. But from (20) and (22) we can deduce that when $\beta \downarrow 0$ then $f(\beta)$ takes infinitely many times the value 0. This is possible for a rational function only if it is identical to 0, contradicting (20). In a similar way we can prove that (21) entails a contradiction. These contradictions show that $\sigma^\sharp$ and $\tau^\sharp$ are Blackwell optimal. □

Now that we know that Blackwell optimal positional strategies exist we are ready to show that they are also optimal for priority mean-payoff games:

Theorem 14. *If $(\sigma^\sharp, \tau^\sharp)$ are Blackwell optimal positional strategies then they are also optimal for the priority mean-payoff game.*

Proof. Suppose the contrary, i.e. that $(\sigma^\sharp, \tau^\sharp)$ is not a pair of optimal strategies for the priority mean-payoff game. This means that there exists a state s such that either

$$\mu(r(s, \sigma^\sharp, \tau^\sharp)) < \mu(r(s, \sigma, \tau^\sharp)) \tag{24}$$

for some strategy σ or

$$\mu(r(s, \sigma^\sharp, \tau)) < \mu(r(s, \sigma^\sharp, \tau^\sharp)) \tag{25}$$

for some strategy τ. Since priority mean-payoff games have optimal positional strategies, by Lemma 2, we can assume without loss of generality that σ and τ are positional. Suppose that (24) holds. By Lemma 12 (B)

$$\lim_{\beta \downarrow 0} u^\beta(r(s, \sigma^\sharp, \tau^\sharp)) = \mu(r(s, \sigma^\sharp, \tau^\sharp)) < \mu(r(s, \sigma, \tau^\sharp)) = \lim_{\beta \downarrow 0} u^\beta(r(s, \sigma, \tau^\sharp)) \ . \tag{26}$$

However inequality (26) implies that there exists $0 < \beta_0$ such that

$$\forall \beta < \beta_0, \quad u^\beta(r(s, \sigma^\sharp, \tau^\sharp)) < u^\beta(r(s, \sigma, \tau^\sharp)) \ ,$$

in contradiction with the Blackwell optimality of $(\sigma^\sharp, \tau^\sharp)$. Similar reasoning shows that also (25) is in contradiction with the Blackwell optimality of $(\sigma^\sharp, \tau^\sharp)$. □

The proof of Theorem 9 (ii) is a direct consequence of Theorems 14 and 13 and Lemma 12.

4.2 Open questions

The most interesting open question is if we can, given an arena $\mathcal{A}$, find an estimate of the constant β_0 from the definition of Blackwell optimal strategies. If this were possible and β_0 were not too small then we could try to find optimal strategies for priority mean-payoff games (and therefore also parity games) by solving priority-discounted games. And for solving priority-discounted games we can adapt policy improvement algorithms developed for discounted games [8].

References

1. D. Blackwell. Discrete dynamic programming. *Annals of Mathematical Statistics*, 33:719–726, 1962.
2. 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.
3. E.A. Emerson and C. Jutla. Tree automata, μ-calculus and determinacy. In *FOCS'91*, pages 368–377. IEEE Computer Society Press, 1991.
4. J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.

5. H. Gimbert and W. Zielonka. Games where you can play optimally without any memory. In *CONCUR 2005*, volume 3653 of *LNCS*, pages 428–442. Springer, 2005.
6. 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.
7. J.F. Mertens and A. Neyman. Stochastic games. *International Journal of Games Theory*, 10:53–56, 1981.
8. T.E.S Raghavan. Finite-step algorithms for single-controller and perfect information stochastic games. In A. Neyman and S. Sorin, editors, *Stochastic Games and Applications*, volume 570 of *NATO Science Series C, Mathematical and Physical Sciences*, pages 227–251. Kluwer Academic Publishers, 2003.
9. L. S. Shapley. Stochastic games. *Proceedings Nat. Acad. of Science USA*, 39:1095–1100, 1953.