BLACKWELL OPTIMAL STRATEGIES IN PRIORITY MEAN-PAYOFF GAMES

HUGO GIMBERT
Laboratoire Bordelais de Recherche en Informatique, CNRS and Université Bordeaux I
351, cours de la Libération, 33405 Talence cedex, France
hugo.gimbert@labri.fr

WIESLAW ZIELONKA
Laboratoire d’Informatique Algorithmique: Fondements et Applications
Université Paris Diderot - Paris 7, Case 7014, 75205 Paris Cedex 13, France
wieslaw.zielonka@liafa.jussieu.fr

Received 8 December 2010
Accepted 20 May 2011
Communicated by Angelo Montanari

We examine perfect information stochastic mean-payoff games – a class of games containing as special sub-classes the usual mean-payoff games and parity games. We show that deterministic memoryless strategies that are optimal for discounted games with state-dependent discount factors close to 1 are optimal for priority mean-payoff games establishing a strong link between these two classes.

Keywords: Two-player games; Blackwell optimality; parity games; mean-payoff games.

1. Introduction

One of the recurring themes in the theory of stochastic games and Markov decision processes is the interplay between discounted games and mean-payoff games. This culminates in the seminal paper of Mertens and Neyman [16] showing that concurrent mean-payoff stochastic games have a value and this value is the limit of the values of discounted games when the discount factor tends to 1 (concurrent games are games where at each state both players chose actions simultaneously and independently). Note however that optimal strategies in both games are very different. As shown by Shapley [18] concurrent discounted stochastic games admit memoryless optimal strategies. On the other hand concurrent mean-payoff games have only $\varepsilon$-optimal strategies and players may need unbounded memory, a famous example is given in [2].

The connections between discounted and mean-payoff games become much tighter for Markov decision processes, i.e. for one-player perfect information stochastic games. As discovered by Blackwell [1], if the discount factor is close to 1 then
optimal memoryless deterministic strategies in discounted one-player games are also optimal for mean-payoff one-player games (but not the other way round). Thus both discounted and mean-payoff one-player games are closely related not only by their values but also through their optimal strategies. Blackwell’s result extends easily to two-player perfect information stochastic games (games where the players move in turns rather than simultaneously).

What happens if instead of mean-payoff games we consider parity games – a class of games more directly relevant to computer science [12]? In particular, are parity games related to discounted games?

The first insight that there is a link between parity games and discounted games is due to de Alfaro at al. [4]. It turns out that parity games are related to multi-discounted games where each state has its own discount factor (in contrast with the usual discounted games where all states have the same discount factor).

Like in the classical theory of stochastic games, we examine what happens when all discount factors of a multi-discounted game tend simultaneously to 1, however it is essential that they tend to 1 with different rates. The idea is that controlling suitably the relative convergence rates of different discount factors in the limit we obtain a parity game. Note that if we have several state dependent discount factors $\lambda_1, \ldots, \lambda_k$ then there are two possibilities to approach 1:

- We can study the iterated limit $\lim_{\lambda_1 \to 1} \cdots \lim_{\lambda_k \to 1}$ when discount factors tend to 1 one after another (i.e. first we go to 1 with the discount factor $\lambda_k$ associated with some group of states, when the limit is reached then we go to 1 with the next discount factor $\lambda_{k-1}$ and so on).
- Another possibility is to examine a simultaneous limit when all factors go to 1 at the same time but with very different rates, this will be made precise in Section 4.

The first approach is easier to handle but it leads to weaker results, in particular we lose the links between optimal strategies in discounted games and optimal strategies in parity games.

In this paper we adopt the second approach, we study the effect of all discount factors tending to 1 at the same time. We have begun our analysis of relations between multi-discounted and parity games in [7,8], where we limited ourselves to deterministic games. Already this preliminary work revealed that the natural framework for such a study goes far beyond the parity games. The reason is that in general the games that we obtain as the limit of multi-discounted games are neither parity games nor mean-payoff games but a natural generalization of both, we call this new class of games priority mean-payoff games. Priority mean-payoff games combine in a special way priorities used in parity games and mean-payoffs. Moreover parity games as well as mean-payoff games are special subclasses of priority mean-payoff games.

The main contribution of the paper is the following. We consider multi-discounted games where each state $s$ is endowed with a discount factor $\lambda_t(s) \in (0,1)$. All these discount factors depend on a common real parameter $t < 1$ such that
lim_{t \to 1} \lambda_t(s) = 1 \text{ for all states } s. \text{ We examine what happens at the limit when } t \text{ approaches } 1. \text{ The relative convergence rates of discount factors for two different states } x \text{ and } y \text{ is captured by the value of the limit } \lim_{t \to 1} \frac{1 - \lambda_t(x)}{1 - \lambda_t(y)} \text{ (intuitively if the limit is } 0 \text{ then } \lambda_t(x) \text{ tends to } 1 \text{ much faster than } \lambda_t(y)).$

With each multi-discounted game we associate a priority mean-payoff game, various parameters of the associated priority mean-payoff game such the priorities of states, the weights of states, depend on the relative convergence rates of discount factors.

We prove three major facts:

- With } t \text{ tending to } 1 \text{ the value of the multi-discounted game tends to the value of the associated priority mean-payoff game.}
- Under suitable conditions concerning the parametrization } t \mapsto \lambda_t(s), \text{ for parameter } t \text{ sufficiently close to } 1 \text{ the optimal memoryless deterministic strategies for the multi-discounted game stabilize, i.e. they do not change when } t \text{ approaches } 1. \text{ Let us call such optimal strategies that do not depend on } t \text{ for } t \text{ close to } 1 \text{ Blackwell optimal.}
- Under the same conditions as above, Blackwell optimal strategies are optimal for the associated priority mean-payoff game.

We think that these results are interesting for several reasons.
First of all establishing a very strong link between two apparently different classes of games, multi-discounted and priority mean-payoff, has its own intrinsic interest.

Discounted games were thoroughly studied in the past and our result shows that algorithms for such games can, in principle, be used to solve parity games (admittedly all depends on how much the parameter } t \text{ should be close to } 1 \text{ to get Blackwell optimal strategies and this problem remains open).}

The other reason is related to the concept of Blackwell optimality. Blackwell invented Blackwell optimality because he was not satisfied with the notion of optimal strategies for mean-payoff Markov Decision Processes (one-player games), he claimed that the usual concept of optimal strategies for mean-payoff games is not selective enough and as a remedy he proposed Blackwell optimal strategies (each Blackwell optimal strategy is optimal for the mean-payoff game but the converse does not hold).

One can argue that we need also a stronger concept of optimality for parity games and that the usual optimality concept for parity games is too weak.

We can proceed in two different ways. Either we refine the payoff mapping for parity games, this mapping takes only two values 0 and 1 (lose or win) and the aim would be to replace it by another payoff } f \text{ with a larger range of values. The new payoff } f \text{ should be consistent with the parity payoff in the sense that } f(h_1) < f(h_2) \text{ for all infinite histories such that the parity payoff of } h_1 \text{ is } 0 \text{ and the parity payoff of } h_2 \text{ is } 1.
The other possibility is to restrict the notion of optimal strategies. Blackwell optimality can be used for this purpose since Blackwell optimal strategies are optimal for the associated priority mean-payoff game but the converse is not true, i.e. requiring Blackwell optimality restrains the class of optimal strategies.

The paper is organized as follows. Section 2 introduces the general framework of perfect information stochastic games. In Section 3 we define (multi-)discounted games. The existence of Blackwell optimal strategies for multi-discounted games is proved in Section 4. In Section 5 we introduce the class of priority mean-payoff games. In Section 6 we prove that for Markov chains the expected priority mean-payoff is the limit of discounted payoffs. In Section 7 we prove the main result of the paper stating that under appropriate conditions Blackwell optimal strategies for multi-discounted games are optimal for the associated priority mean-payoff games.

In Section 8 contains final remarks and examples showing that for priority mean-payoff games the Blackwell optimality is more restrictive than the usual optimality.

The present paper is largely based on two conference papers [10, 9], but it contains also new results and most the proofs are reworked.

2. Perfect Information Stochastic Games

Notation and terminology. In this paper \( \mathbb{N} = \{1, 2, \ldots\} \), \( \mathbb{Z}_+ = \mathbb{N} \cup \{0\} \), and \( \mathbb{R}_+ = [0, \infty) \).

For each finite set \( X \), \( \mathcal{M}_1(X) \) is the set of probability measures over \( X \), i.e. it is the set of mappings \( p : X \to [0, 1] \) such that \( \sum_{x \in X} p(x) = 1 \).

We assume that the reader is familiar with basic probability theory and with discrete stochastic processes, the first few chapters of [13] are amply sufficient. In fact what will be needed is rather the familiarity with probabilistic terminology, all necessary facts are provided.

In our framework a game without players (a 0-player game) is just a discrete-time finite state Markov chain so some familiarity with such chains will be helpful, the reader can consult [17, 21, 13]. Since no other more general Markov chains are used in the paper the term Markov chain will always mean discrete-time finite state Markov chain.

2.1. Arenas

Two players Max and Min are playing an infinite game on an arena

\[ \mathcal{A} = (\mathcal{S}, \mathcal{S}_{\text{Max}}, \mathcal{S}_{\text{Min}}, \mathcal{A}, (\mathcal{A}(s))_{s \in \mathcal{S}}, \delta), \]

where a finite set of states \( \mathcal{S} \) is partitioned in two sets, the set \( \mathcal{S}_{\text{Max}} \) of states controlled by player Max and the set \( \mathcal{S}_{\text{Min}} \) of states controlled by player Min and for each state \( s \in \mathcal{S} \) there is a non-empty finite set \( \mathcal{A}(s) \) of actions available in \( s \), \( \mathcal{A} = \bigcup_{s \in \mathcal{S}} \mathcal{A}(s) \). If at stage \( i \in \mathbb{Z}_+ \) the game is in a state \( s_i \in \mathcal{S} \) then the player controlling \( s_i \) chooses an action from \( \mathcal{A}(s_i) \) and a new state \( s_{i+1} \) is chosen with probability specified by the transition mapping \( \delta \). Transition mapping \( \delta \) maps each
pair \((s, a)\), where \(s \in S\) and \(a \in A(s)\), to an element of \(M_1(S)\). Intuitively, if in a state \(s\) an action \(a\) is executed then \(\delta(s, a)(t)\) gives the probability that at the next stage the game is in state \(t\). To simplify the notation we shall write \(\delta(s, a, t)\) rather than \(\delta(s, a)(t)\).

It is convenient to extend \(\delta(s, a, t)\) to actions \(a\) that are not in \(A(s)\), in this case we set \(\delta(s, a, t) = 0\) for all states \(t\).

Throughout the paper we assume that all arenas are finite, i.e. the sets of states and actions are finite.

An arena is said to be a one-player arena controlled by player Max if the set \(S_{\text{Min}}\) of states controlled by player Min is empty.

One-player arenas controlled by player Min are defined similarly.

### 2.2. Payoffs

Let \(A\) be a fixed arena. A infinite history in \(A\) is an infinite sequence \(h = s_0a_1s_1a_2 \ldots\) interleaving states and actions, the set of infinite histories is denoted \(\mathcal{H}^\omega = (SA)^\omega\).

A finite history \(h = s_0a_1s_1a_2 \ldots s_n\) is a finite sequence interleaving states and actions starting and ending in a state, \(\mathcal{H}^* = (SA)^*S\) will stand for the set of finite histories.

In the sequel “history” without any attribute will be used as a synonym of “infinite history”.

After an infinite history player Max receives a payoff from player Min. The objectives of the players are opposite, the goal of Max is to maximize the payoff while player Min wants to minimize the payoff. In fact, since in our framework the players play in stochastic environment, what they try to maximize/minimize is rather the expectation of the payoff. However the payoff expectation is well-defined only if we impose appropriate measurability conditions.

The following presentation is standard and can be found for example in [20].

We assume that \(A\) and \(S\) are endowed with the discrete topology: all subsets of \(A\) and \(S\) are open.

Let \(S_i : \mathcal{H}^\omega \to S, i \in \mathbb{Z}_+\), be the mappings such that, for \(h = s_0a_1s_1a_2 \ldots\), \(S_i(h) = s_i\). Similarly, \(A_i : \mathcal{H}^\omega \to A, i \in \mathbb{N}\), are the mappings such that \(A_i(h) = a_i\). We assume that \(\mathcal{H}^\omega\) is endowed with product topology. Let us recall that the product topology is the smallest topology for which the mappings \(S_i\) and \(A_i\) are continuous, i.e. it is the smallest topology for which the sets \(\{h \in \mathcal{H}^\omega | S_i(h) = s\}\) and \(\{h \in \mathcal{H}^\omega | A_i(h) = a\}\) are open for all \(i\) and all \(s \in S\) and \(a \in A\).

By \(\mathcal{B}(\mathcal{H}^\omega)\) we note the Borel \(\sigma\)-field on \(\mathcal{H}^\omega\), i.e. the smallest \(\sigma\)-field containing all open sets. Elements of \(\mathcal{B}(\mathcal{H}^\omega)\) are called events. The set of all probability measures on \((\mathcal{H}^\omega, \mathcal{B}(\mathcal{H}^\omega))\) is denoted \(M_1(\mathcal{H}^\omega)\).

In the sequel we assume that the set \(\mathbb{R}\) of real numbers is also equipped with the \(\sigma\)-field \(\mathcal{B}(\mathbb{R})\) of Borel sets.

A payoff function is any bounded \(\mathcal{B}(\mathcal{H}^\omega)/\mathcal{B}(\mathbb{R})\)-measurable mapping

\[
u : \mathcal{H}^\omega \to \mathbb{R}
\]

from infinite histories to real numbers.
Measurability and boundedness ensure that for each probability measure \( P \in \mathcal{M}_1(\mathcal{H}^\omega) \) the payoff expectation \( E(u) \) is always well-defined and finite.

A game is a couple \( \Gamma = (\mathcal{A}, u) \) made of an arena and a payoff function.

Usually we are not interested in a particular game but rather in a class of games. Each class of games considered in this paper has the following structure. Let \( C \) be a set, the elements of \( C \) are called colors. We consider arenas endowed with labeling \( \varphi : S \to C \). Thus each history \( h = s_0a_1s_1a_2s_2 \ldots \) gives rise to an infinite sequence of colors, \( \varphi(h) := \varphi(s_0)\varphi(s_1)\varphi(s_2) \ldots \). We assume that \( C \) is endowed with the discrete topology and the set \( C^\omega \) of infinite color sequences is equipped with the product topology. By \( \mathcal{B}(C^\omega) \) we denote the Borel \( \sigma \)-field generated by this topology. Note that \( \varphi : \mathcal{H}^\omega \to C^\omega \) defined above is \( \mathcal{B}(\mathcal{H}^\omega)/\mathcal{B}(C^\omega) \)-measurable (in fact it is even continuous). Then a class of games is defined by specifying a bounded \( \mathcal{B}(C^\omega)/\mathcal{B}(\mathbb{R}) \)-measurable mapping \( u_C : C^\omega \to \mathbb{R} \) (a payoff mapping for infinite color sequences).

Clearly the composition \( u := u_C \circ \varphi \) defines a payoff mapping on infinite histories.

A one-player game or a Markov Decision Process (MDP) is a game on a one-player arena.

A zero-player game is a game such that each state has only one available action. For such games the action executed at stage \( i \) is determined by the state visited at stage \( i \) and we can as well forget actions and consider only the stochastic process \( (S_i)_{i \in \mathbb{Z}^+} \). Clearly such a process forms a Markov chain. When we formulate or prove some facts concerning a Markov chain \( (S_i)_{i \in \mathbb{Z}^+} \) then we assume that the corresponding probability space is equipped with measure \( P_s \) such that \( P_s(S_0 = s') = 1 \) if \( s = s' \) and \( E_s \) will denote the corresponding expectation.

2.3. Strategies

Playing a game players use strategies.

A strategy for player Max is a mapping \( \sigma : (\mathcal{SA})^*S_{\text{Max}} \to \mathcal{M}_1(\mathcal{A}) \) from the set \( (\mathcal{SA})^*S_{\text{Max}} \) of finite histories with the last state controlled by player Max into actions such that for every finite history \( h = s_0a_0s_1a_1 \ldots s_n \) with \( s_n \in S_{\text{Max}} \) and for all \( a \in \mathcal{A} \), if \( \sigma(h)(a) > 0 \) then \( a \in \mathcal{A}(s_n) \) (only available actions can be chosen).

Strategies for player Min are defined similarly and denoted \( \tau \).

A strategy profile is a pair \( (\sigma, \tau) \) of strategies of players Max and Min. Each strategy profile defines a mapping

\[
\sigma \cup \tau : \mathcal{H}^* \to \mathcal{M}_1(\mathcal{A})
\]

such that, for each finite history \( h \), \( (\sigma \cup \tau)(h) \) is equal either to \( \sigma(h) \) if \( h \) terminates in a state controlled by Max or to \( \sigma(h) \) if \( h \) terminates in a state controlled by Min. Intuitively, \( (\sigma \cup \tau)(h)(a) \) gives the probability of executing action \( a \) after the history \( h \) if players play using strategies \( \sigma \) and \( \tau \).

Certain types of strategies are of particular interest. A strategy is deterministic (or pure) if it chooses actions in a deterministic way, and it is memoryless (or stationary, Markovian) if it depends only on the current state. Formally:
Definition 1. A strategy $\sigma$ of player $i \in \{\text{Min}, \text{Max}\}$ is said to be:

- deterministic if, for each finite history $h$ terminating in a state controlled by player $i$, there exists only one action $a$ such that $\sigma(h)(a) > 0$,
- memoryless if, for each finite history $h$ terminating in a state $s$ controlled by player $i$, $\sigma(h) = \sigma(s)$.

For each finite history $h = s_0a_1s_1 \ldots s_n$ a cone generated by $h$ is the event

$$Q(h) = \{S_0 = s_0, A_1 = a_1, S_1 = s_1, \ldots, S_n = s_n\}. \quad (1)$$

Note that the cones generate the $\sigma$-field $\mathcal{B}(\mathcal{H}^\omega)$ of Borel sets.

The initial state $s \in S$, player's strategy profile $(\sigma, \tau)$ and transition probabilities $\delta$ yield a probability measure $\mathbb{P}^{\sigma, \tau}_s$ on cones; for $Q(h)$ as in (1) we set

$$\mathbb{P}^{\sigma, \tau}_s(Q(h)) = I_{\{s_0 = s\}} \cdot (\sigma \cup \tau)(s_0)(a_1) \cdot \delta(s_0, a_1, s_1) \cdot (\sigma \cup \tau)(s_0a_1s_1)(a_2) \cdot \delta(s_1, a_2, s_3) \cdots \cdot (\sigma \cup \tau)(s_0a_1s_1 \ldots s_{n-1})(a_n) \cdot \delta(s_{n-1}, a_n, s_n), \quad (2)$$

where $I_{\{s_0 = s\}}$ is the indicator function of the event $\{s_0 = s\}$.

Ionescu Tulcea's theorem [19] implies that there exists a unique probability measure $\mathbb{P}^{\sigma, \tau}_s \in \mathcal{M}_1(\mathcal{H}^\omega)$ on $(\mathcal{H}^\omega, \mathcal{B}(\mathcal{H}^\omega))$ satisfying (2).

Given a payoff mapping $u$, the expected value of $u$ under $\mathbb{P}^{\sigma, \tau}_s$ is denoted $\mathbb{E}^{\sigma, \tau}_s \{u\}$.

2.4. Memoryless deterministic strategy profile and Markov chains

Very often we will consider memoryless deterministic strategy profiles $(\sigma, \tau)$ (i.e. with both $\sigma$ and $\tau$ memoryless deterministic). Fixing such a strategy profile $(\sigma, \tau)$ we have $A_i = (\sigma \cup \tau)(S_i)$, i.e. the action executed at stage $i$ is completely determined by $S_i$. Moreover, the conditional probability $P^{\sigma, \tau}_s(S_{i+1} = s_{i+1} | S_0 = s_0, \ldots, S_i = s_i) \equiv \delta(s_i, (\sigma \cup \tau)(s_0 \ldots s_i), s_{i+1}) = \delta(s_i, (\sigma \cup \tau)(s_i), s_{i+1}) = P^{\sigma, \tau}_s(S_{i+1} = s_{i+1} | S_i = s_i)$ depends only on states $s_i$ and $s_{i+1}$ but not on the stage $i$. In other words, if $(\sigma, \tau)$ is a memoryless deterministic strategy profile then $(S_i)_{i \in \mathbb{Z}_+}$ is a Markov chain defined on the probability space $(\mathbb{P}^{\sigma, \tau}_s, \mathcal{H}^\omega, \mathcal{B}(\mathcal{H}^\omega))$ (recall again that in this paper Markov chain means a discrete-time finite state Markov chain).

2.5. Optimal strategies

When players Max and Min play a game $(A, u)$ then they want respectively maximize/minimize the expected payoff $\mathbb{E}^{\sigma, \tau}_s \{u\}$. For any game we have always

$$\underline{\text{val}}_s(u) := \sup_{\sigma} \inf_{\tau} \mathbb{E}^{\sigma, \tau}_s \{u\} \leq \inf_{\tau} \sup_{\sigma} \mathbb{E}^{\sigma, \tau}_s \{u\} =: \overline{\text{val}}_s(u)$$

and if the left-hand and the right-hand sides are equal then the state $s$ is said to have the value $\text{val}_s(u) = \underline{\text{val}}_s(u) = \overline{\text{val}}_s(u)$. 

}\end{quote}
A strategy $\sigma^\#$ of player Max is optimal if it guarantees him the expected payoff of at least $\text{val}\{u\}$ against any strategy of player Min, i.e. $\sigma^\#$ is optimal if $\inf_{\tau} E_{\sigma^\#,\tau} u \geq \text{val}\{u\}$.

Similarly, a strategy $\tau^\#$ of player Min is optimal if it guarantees him that his loss will not exceed $\text{val}\{u\}$ against any strategy of player Max, i.e. $\tau^\#$ is optimal if $\sup_{\sigma} E_{\sigma,\tau^\#} u \leq \text{val}\{u\}$.

Clearly if one of the players has an optimal strategy then the game has value but the converse does not hold, i.e. the existence of the value does not imply the existence of optimal strategies (and the existence of an optimal strategy for one player does not guarantee the existence of optimal strategy for the other player). Martin’s theorem [15] on determinacy of Blackwell games ensures that every state has value but this result will not be used in the paper.

We will say that $(\sigma^\#, \tau^\#)$ is an optimal strategy profile if $\sigma^\#$ and $\tau^\#$ are optimal. The following fact will be used frequently: $(\sigma^\#, \tau^\#)$ is an optimal strategy profile if and only if for each state $s$ and all strategies $\sigma, \tau$ of Max and Min

$$E_{\sigma^\#, \tau^\#} u \leq E_{\sigma^\#} u \leq E_{\sigma^\#, \tau^\#} u.$$  

3. Discounted Games

Arenas for discounted games are equipped with two mappings

$$\lambda : S \to [0, 1) \text{ and } r : S \to \mathbb{R}.$$  

The discount mapping $\lambda$ maps each state $s$ to a discount factor $\lambda(s) \in [0, 1)$ and the reward mapping $r$ maps each state $s$ to a real valued reward $r(s)$.

The payoff

$$u_{r, \lambda} : H^\omega \to \mathbb{R}$$  

for discounted games is calculated in the following way:

$$u_{r, \lambda} = (1 - \lambda(S_0))r(S_0) + \lambda(S_0)(1 - \lambda(S_1))r(S_1)$$  

$$+ \lambda(S_0)\lambda(S_1)(1 - \lambda(S_2))r(S_2) + \ldots$$  

$$= \sum_{i=0}^{\infty} \lambda(S_0)\ldots\lambda(S_{i-1})(1 - \lambda(S_i))r(S_i).$$  

(3)

Usually when discounted games are considered it is assumed that there is only one discount factor, i.e. that there exists $\lambda \in [0, 1)$ such that $\lambda(s) = \lambda$ for all $s \in S$. But in this paper it is essential that the discount factors depend on the state.

Shapley [18] proved* that

*In fact, Shapley considered a much larger class of stochastic games with simultaneous moves. For these games he proved that both players have memoryless optimal strategies. For perfect information games his proof yields optimal strategies that are not only memoryless but also deterministic.
Theorem 2 (Shapley) Discounted games \((A, u_r, \lambda)\) over finite arenas admit optimal deterministic memoryless strategies for both players.

4. Blackwell Optimality

We will consider what happens with game values and with optimal strategies when the discount factors tend to 1. The novelty in comparison with the traditional approach [5,1] is that we examine what happens when the discount factors of different states tend to 1 simultaneously but with different rates.

A rational discount parametrization is a family of mappings \(\lambda_t = (\lambda_t(s))_{s \in S}\), such that for each state \(s\),

(R1) \(t \mapsto \lambda_t(s)\) is a rational\(^b\) function of \(t\),
(R2) there exists \(0 < \varepsilon < 1\) such that \(\lambda_t(s) \in [0,1)\) for all \(t \in [1-\varepsilon,1)\) (since the set of states is finite we can choose the same \(\varepsilon\) for all states),
(R3) \(\lim_{t \to 1} \lambda_t(s) = 1\).

A typical example of a rational discount parametrization is the canonical rational discount parametrization defined in the following way. For each state \(s\) we fix a positive integer \(\pi(s) \in \mathbb{N}\) called the priority of \(s\) and a positive real number \(w(s) \in (0,\infty)\) called the weight of \(s\). The canonical rational discount parametrization is defined as

\[
\lambda_t(s) = 1 - w(s)(1-t)^{\pi(s)}, \quad \text{for } s \in S, t \in \mathbb{R}.
\] (4)

Clearly this parametrization satisfies (R1)-(R3).

Suppose that \(\lambda_t\) is a rational discount parametrization. By Theorem 2 we know that for each fixed \(t\) the discounted game with payoff \(u_r, \lambda_t\) has optimal memoryless deterministic strategies, but obviously such strategies depend on \(t\).

A deterministic memoryless strategy \(\sigma^t\) of player Max is Blackwell optimal for a rational discount parametrization \(\lambda_t\) if there exists \(0 < \varepsilon < 1\) such that \(\sigma^t\) is optimal for Max for all payoffs \(u_r, \lambda_t\) with \(1-\varepsilon < t < 1\). Replacing Max by Min we obtain the definition of Blackwell optimality for player Min.

Thus Blackwell optimal strategies are just strategies optimal for all parameters \(t\) sufficiently close to 1.

Theorem 3 (Blackwell optimality) Let us fix an arena \(A\) of a discounted game and a rational discount parametrization \(\lambda_t\) for \(A\). Then

(1) both players have Blackwell optimal strategies and
(2) there exists \(0 < \varepsilon < 1\) such that for each state \(s\) the mapping \((1-\varepsilon,1) \ni t \mapsto \text{val}_s(u_r, \lambda_t) \in \mathbb{R}\) is rational.

Blackwell [1] has proved the existence of Blackwell optimal strategies only for one-player games and for the uniform parametrization \(\lambda_t(s) = t\) for all states \(s\).

\(^b\)Rational in the sense that \(\lambda_t(s)\) is a quotient of two polynomials of \(t\).
The generalization for two-player games with any rational parametrization is rather straightforward, in particular in this paper we will adapt the one-player proof of [14].

We begin with the following lemma which will also be useful in the next section.

**Lemma 4.** Let \((S_t)_{t \in \mathbb{Z}}\) be a Markov chain.

If \(t \mapsto \lambda_t\) is a rational discount parametrization then, for each state \(s\) and for \(t\) sufficiently close to 1, \(t \mapsto E_s\{u_{r,\lambda_t}\}\) is a rational function of \(t\).

**Proof.** The proof is standard and follows the one of [14]. The set \(\mathbb{R}^{S \times S}\) of functions from \(S \times S\) into the real numbers can be seen as the set of square real valued matrices with rows and columns indexed by \(S\) and endowed with natural matrix addition and scalar multiplication \(\mathbb{R}^{S \times S}\) is a vector space. Matrix multiplication defines a product on \(\mathbb{R}^{S \times S}\), for \(M, N \in \mathbb{R}^{S \times S}\), \(MN\) is an element \(U\) of \(\mathbb{R}^{S \times S}\) with entries 

\[U[s', s''] = \sum_{s''} M[s', s] N[s, s''].\]

We endow \(\mathbb{R}^{S \times S}\) with a norm, for \(M \in \mathbb{R}^{S \times S}\), 

\[\|M\| = \max_{s' \in S} \sum_{s''} |M[s', s'']|\]

It can easily be shown that \(\|MN\| \leq \|M\| \cdot \|N\|\) for \(M, N \in \mathbb{R}^{S \times S}\) and \(\mathbb{R}^{S \times S}\) is a complete metric space for the metric induced by the norm \(\|\cdot\|\), see Section 3.2.1 of [21] for proofs.

On the other hand, we consider the vector space \(\mathbb{R}^S\) of functions from \(S\) into \(\mathbb{R}\), elements of \(\mathbb{R}^S\) can be seen as column vectors indexed by states. Of course if \(M \in \mathbb{R}^{S \times S}\) and \(v \in \mathbb{R}^S\) then \(Mv \in \mathbb{R}^S\), where 

\[(Mv)[s] = \sum_{s' \in S} M[s, s']v[s']\]

for \(s \in S\).

We equip \(\mathbb{R}^S\) with a norm, for \(v \in \mathbb{R}^S\), \(\|v\|_\infty = \max_{s \in S} |v[s]|\). The norms on \(\mathbb{R}^{S \times S}\) and \(\mathbb{R}^S\) are compatible in the sense that we have \(\|Mv\|_\infty \leq \|M\| \cdot \|v\|_\infty\).

Let \(\delta(s', s'')\) be transition probabilities of the Markov chain \((S_t)\). In the sequel \(M\) will denote the element of \(\mathbb{R}^{S \times S}\) defined in the following way

\[M[s', s''] = \lambda_t(s') \delta(s', s''), \quad \text{for } s', s'' \in S.\]

Let \(I \in \mathbb{R}^{S \times S}\) be the identity matrix, i.e. \(I[s', s'']\) is 1 if \(s' = s''\) and 0 otherwise.

We shall show that for \(t\) close to 1 the matrix \((I - M)\) is invertible and

\[(I - M)^{-1} = \sum_{i=0}^{\infty} M^i.\]

First we show that the series on the right-hand side of (5) converges.

Let \(q_t = \max_{s \in S} \lambda_t(s)\). Then \(\|M\| \leq q_t\) and, for \(k < l,\)

\[\|M^l - M^k\| \leq \sum_{i=k}^{l-1} \|M^i\| \leq \sum_{i=k}^{l} (q_t)^i = \frac{(q_t)^k - (q_t)^{l+1}}{1 - q_t} \xrightarrow{k, l \to \infty} 0\]

since, by the definition of a rational discount parametrization, \(0 < q_t < 1\) for \(t\) sufficiently close to 1. Thus the series \(\sum_{i=0}^{\infty} M^i\) satisfies the Cauchy condition and convergence follows from the completeness of the norm \(\|\cdot\|\). Now it suffices to note that

\[(I - M) \sum_{i=0}^{k} M^i = I = -M^{k+1}\]
and $\|M^{k+1}\| \leq \|M\|^{k+1} \leq q_k^{k+1} \xrightarrow{k \to \infty} 0$, which implies that $(I - M) \cdot \sum_{i=0}^{\infty} M^i - I = 0$, yielding (5).

Now note that

$$\mathbb{E}_s \{ u_{r, \lambda} \} = \mathbb{E}_s \left\{ \sum_{i=0}^{\infty} \lambda_i(S_0) \cdots \lambda_i(S_{i-1}) (1 - \lambda_i(S_i)) r(S_i) \right\}$$

$$= \lim_{k \to \infty} \mathbb{E}_s \left\{ \sum_{i=0}^{k} \lambda_i(S_0) \cdots \lambda_i(S_{i-1}) (1 - \lambda_i(S_i)) r(S_i) \right\}, \quad (6)$$

where the second equality follows from the Lebesgue dominated convergence theorem. (To apply dominated convergence we should show that the sequence of functions is bounded by an integrable function, here and everywhere in the paper this will be trivial, we always apply dominated convergence to function sequences bounded by a constant, for example the sequence appearing in (6) is bounded by $\max_{s \in \mathcal{S}} |r(s)|$.)

Let $v$ be an element of $\mathbb{R}^S$ such that

$$v[s] = (1 - \lambda_i(s)) r(s), \quad \text{for } s \in \mathcal{S}.$$ 

Elementary induction on $i$ shows that, for $s, s' \in \mathcal{S}$,

$$\mathbb{E}_s \{ \lambda_i(S_0) \cdots \lambda_i(S_{i-1}) | S_i = s' \} = M^i[s, s'],$$

i.e. the entry $[s, s']$ of the $i$-th power of $M$ is the expectation of $\lambda_i(S_0) \cdots \lambda_i(S_{i-1})$ under the condition that $S_0 = s$ and $S_i = s'$. This yields

$$(M^i v)[s] = \sum_{s' \in \mathcal{S}} M^i[s, s'] \cdot v[s']$$

$$= \sum_{s' \in \mathcal{S}} \mathbb{E}_s \{ \lambda_i(S_0) \cdots \lambda_i(S_{i-1}) | S_i = s' \} \cdot (1 - \lambda_i(s')) r(s')$$

$$= \mathbb{E}_s \{ \lambda_i(S_0) \cdots \lambda_i(S_{i-1}) (1 - \lambda_i(S_i)) r(S_i) \}.$$ \quad (7)

Taking the sum from $i = 0$ to $k$ on both sides of (7) and next the limit with $k$ tending to infinity, using (6) and (5), we get

$$\mathbb{E}_s \{ u_{r, \lambda} \} = ((I - M)^{-1} v)[s].$$

But the elements of the matrix $I - M$ are rational functions of $t$, thus Cramer’s rule for matrix inversion shows that the elements of $(I - M)^{-1}$ are also rational functions of $t$, and since the elements of $v$ are also rational functions we can see that $\mathbb{E}_s \{ u_{r, \lambda} \}$ is a rational function of $t$. \hfill \Box

**Proof of Theorem 3.** Since discounted games admit optimal deterministic memoryless strategies, Lemma 4 shows that (2) is a consequence of (1).

We prove (1) as follows.

Let $X$ be the set of all tuples $(q, \sigma, \tau, \sigma', \tau')$, where $q$ is a state, $\sigma, \sigma'$ are deterministic memoryless strategies for player Max and $\tau, \tau'$ are deterministic memoryless
strategies for player Min. Note that for finite arenas $X$ is finite. Let $\lambda_t$ be a rational discount parametrization and let $0 < \varepsilon < 1$ be such that $\lambda_t(s) \in (0, 1)$ for all states $s$ and all $t \in (1 - \varepsilon, 1)$.

For each $(q, \sigma, \tau, \sigma', \tau') \in X$ we consider the function $\Phi_{q,\sigma,\tau,\sigma',\tau'} : (1 - \varepsilon, 1) \to \mathbb{R}$ defined by:

$$
t \mapsto \Phi_{q,\sigma,\tau,\sigma',\tau'}(t) = \mathbb{E}_{\sigma,\tau}^{\sigma'} \{ u_{r,\lambda_t} \} - \mathbb{E}_{\sigma'}^{\sigma'} \{ u_{r,\lambda_t} \}.
$$

According to Lemma 4, $\Phi_{q,\sigma,\tau,\sigma',\tau'}(t)$ is a rational function of $t$ for $t$ sufficiently close to 1. Since a rational function can change sign (cross the $x$-axis) only finitely many times there exists $\varepsilon_1 = \varepsilon_1(q, \sigma, \tau, \sigma', \tau') > 0$ such that the sign of $\Phi_{q,\sigma,\tau,\sigma',\tau'}(t)$ does not change in the interval $(1 - \varepsilon_1, 1)$. Let $\varepsilon_2 = \min\{\varepsilon\} \cup \{\varepsilon_1(q, \sigma, \tau, \sigma', \tau') : (q, \sigma, \tau, \sigma', \tau') \in X\}$.

Since $X$ is finite, $\varepsilon_2$ is strictly positive as the minimum of a finite number of positive numbers.

Let us take any $t \in (1 - \varepsilon_2, 1)$. Let $\sigma^1, \tau^1$ be optimal deterministic memoryless strategies in the discounted game $(A, u_{r,\lambda_t})$ (Theorem 2). Then, in particular, we have

$$
\mathbb{E}_{\sigma}^{\sigma^1} \{ u_{r,\lambda_t} \} \leq \mathbb{E}_{\tau}^{\tau^1} \{ u_{r,\lambda_t} \} \leq \mathbb{E}_{\sigma}^{\tau^1} \{ u_{r,\lambda_t} \}
$$

for all deterministic memoryless strategies $\sigma, \tau$. We can rewrite (8) as

$$
\Phi_{q,\sigma^1,\tau^1,\sigma,\tau}(t) \geq 0 \quad \text{and} \quad \Phi_{q,\sigma^1,\tau^1,\sigma,\tau}(t) \geq 0.
$$

However, functions $\Phi_{q,\sigma^1,\tau^1,\sigma,\tau}(t)$ and $\Phi_{q,\sigma^1,\tau^1,\sigma,\tau}(t)$ do not change sign in $(1 - \varepsilon_2, 1)$. Therefore (8) holds for all $t \in (1 - \varepsilon_2, 1)$. Finally Theorem 2 implies that if (8) holds for all deterministic memoryless strategies $\sigma$ and $\tau$ (with fixed deterministic memoryless $\sigma^1$ and $\tau^1$) then it holds for all strategies $\sigma, \tau$. □

5. Priority Mean-Payoff Games

In mean-payoff games the players try to optimize (maximize/minimize) the mean value of the payoff received at each stage. In such games the reward mapping

$$
r : S \to \mathbb{R}
$$

gives, for each state $s$, the payoff received by player Max when $s$ is visited. The mean-payoff is defined as the limit of the means of stage payments: $u_s = \limsup_k \left( \frac{1}{k^t} \sum_{t=0}^{k} r(S_t) \right)$, where we take limsup rather than the simple limit since the latter may not exist. We slightly generalize mean-payoff games by equipping arenas with a new mapping

$$
w : S \to \mathbb{R}_+^+
$$

associating with each state $s$ a strictly positive real number $w(s)$, the weight of $s$. We can interpret $w(s)$ as the amount of time spent in state $s$ upon each visit to
s. In this setting $r(s)$ should be seen as the payoff per time unit when $s$ is visited, thus the weighted mean payoff received by player Max is

$$u_{r,w} = \limsup_k \frac{\sum_{i=0}^{k} w(S_i) r(S_i)}{\sum_{i=0}^{k} w(S_i)}.$$

(9)

As a final ingredient we add to the arena a priority mapping

$$\pi : S \rightarrow \mathbb{N}$$

assigning to each state $s$ a positive integer priority $\pi(s)$.

We define the priority mapping $\pi_* : \mathcal{H}^\omega \rightarrow \mathbb{N}$ on $\mathcal{H}^\omega$ by setting

$$\pi_* = \liminf_i \pi(S_i).$$

Thus the priority of a history $h = s_0a_0s_1a_1s_2a_2\ldots$ is the smallest priority appearing infinitely often in the sequence $\pi(s_0)\pi(s_1)\pi(s_2)\ldots$ of priorities visited in $h$.

By $1_{\{\pi(S_i) = \pi_*\}}$ we denote the indicator function of the event $\{\pi(S_i) = \pi_*\}$, i.e. $1_{\{\pi(S_i) = \pi_*\}}(h)$ is equal to 1 for histories $h = s_0a_1s_1a_2\ldots$ such that $s_i = \pi_*(h)$ and is 0 otherwise.

We define on $\mathcal{H}^\omega$ the real valued random process

$$u^{k}_{r,w,\pi} = \frac{\sum_{i=0}^{k-1} 1_{\{\pi(S_i) = \pi_*\}} \cdot w(S_i) \cdot r(S_i)}{\sum_{i=0}^{k-1} 1_{\{\pi(S_i) = \pi_*\}} \cdot w(S_i)}.$$  

(10)

The priority mean-payoff mapping $u_{r,w,\pi}$ is defined as

$$u_{r,w,\pi} = \limsup_k u^{k}_{r,w,\pi}.$$  

(11)

For a given history $h = s_0a_1s_1a_2s_3\ldots$ Eq.(10) yields $u^{k}_{r,w,\pi}(h) = \frac{\sum_{i=0}^{k-1} 1_{\{s_i = \pi_*(h)\}} w(s_i) r(s_i)}{\sum_{i=0}^{k-1} 1_{\{s_i = \pi_*(h)\}} w(s_i)}$. We can interpret $1_{\{s_i = \pi_*(h)\}} w(s_i)$ as the modified weight of the state $s_i$ visited in stage $i$. This modified weight is either equal to $w(s_i)$ if $\pi(s_i) = \pi_*(h)$ or is equal to 0 otherwise (in other words we shrink to 0 the weights of states with priority different from $\pi_*(h)$).

Let us note that the denominator $1_{\{s_i = \pi_*(h)\}} w(s_i)$ is different from 0 for $k$ large enough, in fact it tends to infinity since $1_{\{s_i = \pi_*(h)\}}(h) = 1$ for infinitely many $i$. For small $k$ the numerator and the denominator can be equal to 0 and then it is convenient to assume that the indefinite value 0/0 is equal to $-\infty$.

For any history $h = s_0a_1s_1a_2s_2\ldots$ and any $i \in \mathbb{N}$ let $\pi_i(h) = s_{w_i}s_{w_i} \ldots$ be a sub-sequence (sub-word) of $h$ obtained by removing from $h$ all actions and all states with priority different from $i$. This operation allows to give an alternative definition of the priority mean-payoff (11),

$$u_{r,w,\pi}(h) = u_{r,w}(\pi_i(h)), \text{ where } i = \pi_*(h),$$

(12)

i.e. we apply the weighted mean-payoff (9) to the sub-word $\pi_i(h)$ where $i$ is the priority of $h$. 

Blackwell Optimal Strategies 699
In the sequel the couple \((w, \pi)\) consisting of weight mapping and priority mappings will be called a weighted priority system.

Let us note that priority mean-payoff games are a vast generalization of parity games. In fact parity games correspond to a very particular case of priority mean-payoff games, we recover the usual parity games if we set, for each state \(s\), \(w(s) = 1\) and \(r(s) = 1\) if \(\pi(s)\) is even and \(r(s) = 0\) if \(\pi(s)\) is odd. Let us note also that priority mean-payoff games are very different from mean-payoff parity games examined in [3], the latter are obtained by superimposing parity and mean-payoff games.

**Theorem 5.** Priority mean-payoff games over finite arenas admit optimal deterministic memoryless strategies for both players.

We begin by proving Theorem 5 for one-player games (Markov Decision Processes). To this end we will use the following result due to the first author [6]

**Theorem 6.** Under the notation of Section 2.2, we consider a class of games with payoff mapping \(u_C : C^* \rightarrow \mathbb{R}\), where \(C\) is a set of colors. Suppose that \(u_C\) is prefix independent, i.e. for all \(x \in C^*, y \in C^*, u_C(xy) = u_C(y)\) and that \(u_C\) is sub-mixing, i.e. for each infinite sequence of nonempty words \(x_1, y_1, x_2, y_2, \ldots \in C^*\), we have \(u_C(x_1y_1x_2y_2\ldots) \leq \max\{u_C(x_1x_2\ldots), u_C(y_1y_2\ldots)\}\).

Then, for each one-player game \((A, u_C)\) played on an arena controlled by player Max, this player has an optimal memoryless deterministic strategy.

**Lemma 7.** One-player priority mean-payoff games with unique player Max have optimal memoryless deterministic strategies.

**Proof.** That the priority mean-payoff function is prefix independent is evident.

We first prove the sub-mixing property for weighted mean-payoff of (9). Set \(u_{r,w}^k = \sum_{i=0}^k w(S_i)r(S_i)/\sum_{i=0}^k w(S_i)\). Thus \(\lim \sup_k u_{r,w}^k = u_{r,w}\).

Let \(h \in S^*\) be an infinite sequence of states and let \(h = x_1y_1x_2y_2\ldots\) be a factorization of \(h\), where \(x_i\) and \(y_i\) are finite nonempty sequence of \(S\). Let \(S_i, i \in \mathbb{Z}_+\), be the \(i\)-th state of \(h\). We set \(U_x = \{i \mid S_i\) belongs to one of the factors \(x_j\}\), \(U_y = \{i \mid S_i\) belongs to one of the factors \(y_j\}\), \(U_x(k) = U_x \cap \{0, 1, \ldots, k\}\) and \(U_y(k) = U_y \cap \{0, 1, \ldots, k\}\). Then we have

\[
u_{r,w}^k = \frac{\sum_{i=0}^k w(S_i)r(S_i)}{\sum_{i=0}^k w(S_i)} = \frac{\sum_{i \in U_x(k)} w(S_i)r(S_i)}{\sum_{i \in U_x(k)} w(S_i)} + \frac{\sum_{i \in U_y(k)} w(S_i)r(S_i)}{\sum_{i \in U_y(k)} w(S_i)} = \frac{\sum_{i \in U_x(k)} w(S_i)}{\sum_{i \in U_y(k)} w(S_i)} \frac{\sum_{i \in U_y(k)} w(S_i)r(S_i)}{\sum_{i \in U_y(k)} w(S_i)}.
\]

(13)
implies that \( \lim \) follows directly from (13) we deduce that

\[
\alpha_k = \frac{\sum_{i \in U_k} w(S_i)}{\sum_{i \in S} w(S_i)} \quad \text{and} \quad \beta_k = \frac{\sum_{i \in V_k} w(S_i)}{\sum_{i \in S} w(S_i)},
\]

we can see that \( \alpha_k + \beta_k = 1 \) and \( \alpha_k, \beta_k \geq 0 \), thus from (13) we deduce that

\[
u_{r,w}^k \leq \max \left\{ \frac{\sum_{i \in U_k} w(S_i)r(S_i)}{\sum_{i \in U_k} w(S_i)}, \frac{\sum_{i \in V_k} w(S_i)r(S_i)}{\sum_{i \in V_k} w(S_i)} \right\}
\]

(14)

But \( \nu_{r,w} = \lim \sup \nu_{r,w}^k \) and

\[
u_{r,w}(x) = \lim \sup \frac{\sum_{i \in U_k} w(S_i)r(S_i)}{\sum_{i \in U_k} w(S_i)}, \quad \nu_{r,w}(y) = \lim \sup \frac{\sum_{i \in V_k} w(S_i)r(S_i)}{\sum_{i \in V_k} w(S_i)},
\]

where \( x = x_1x_2 \ldots \) and \( y = y_1y_2 \ldots \). These facts and (14) imply that \( \nu_{r,w}(h) \leq \max(\nu_{r,w}(x), \nu_{r,w}(y)) \), i.e. \( \nu_{r,w} \) is sub-mixing.

This and (12) will be used to show that the priority mean-payoff is sub-mixing. Let us take again \( h = x_1y_1x_2y_2 \ldots, x = x_1x_2 \ldots \) and \( y = y_1y_2 \ldots \). Clearly \( \pi_\ast(h) \leq \pi_\ast(h) \leq \pi_\ast(y) \) and at least one of these inequalities is in fact the equality (if a priority appears infinitely often in \( h \) then it appears infinitely often either in \( x \) or in \( y \)). Thus there are essentially two cases to examine.

Let \( i = \pi_\ast(h) = \pi_\ast(x) \) and \( \pi_\ast(h) < \pi_\ast(y) \). Then \( \pi_i(h) \) is a finite word, thus \( \pi_i(h) \) and \( \pi_i(x) \) differ only on a finite prefix implying that \( \nu_{r,w}(\pi_i(h)) = \nu_{r,w}(\pi_i(x)) \) implying, by (12), that \( \nu_{r,w,\pi}(h) = \nu_{r,w,\pi}(x) \). Thus \( \nu_{r,w,\pi}(h) \leq \max(\nu_{r,w,\pi}(x), \nu_{r,w,\pi}(y)) \).

The other case is when \( i = \pi_\ast(h) = \pi_\ast(x) = \pi_\ast(y) \). But then \( \pi_i(h) \) is just an interleaving (shuffle) of two infinite words \( \pi_i(x) \) and \( \pi_i(y) \) and then the fact \( \nu_{r,w,\pi} \) is sub-mixing follows directly from (12) since \( \nu_{r,w} \) is sub-mixing.

For a state \( s \in S \) and \( n \in \mathbb{N} \) let \( V_s(n) = \sum_{k=0}^n 1_{\{s_k = s\}} \) be the number of visits to \( s \) up to \( n \).

We will use the following result [17]:

Theorem 8 (Ergodic theorem for Markov chains) Let \((S_i)_{i \in \mathbb{Z}^+} \) be an irreducible finite state Markov chain. Then, with probability 1, \( \frac{V_s(n)}{n} \) tends to \( \frac{1}{T} \) as \( n \rightarrow \infty \), where \( T = \min \{ i > 0 \mid S_i = s \} \) is the return time to \( s \).

Since for each Markov chain with probability one the set of states visited infinitely often forms a closed communicating class, Theorem 8 implies that \( v \lim_{n} \frac{V_s(n)}{n} \) exists almost surely for each Markov chain and each state \( s \).

Corollary 9. Let \((S_i)_{i \in \mathbb{Z}^+} \) be a finite Markov chain with state space \( S \). Let \( r, w \) and \( \pi \) be reward, weight and priority functions defined on \( S \). Then almost surely

\[
\lim_{k} \nu_{r,w,\pi}^k = \frac{\sum_{s \in S, \pi(s) = \pi_\ast} w(s)r(s)V_s}{\sum_{s \in S, \pi(s) = \pi_\ast} w(s)V_s},
\]

where \( V_s = \lim_{n} \frac{V_s(n)}{n} \), in particular the limit on the left-hand side exists almost surely.
Proof. We have
\[
\sum_{i=0}^{k} \mathbf{1}_{\{\pi(S_i) = \pi_s\}} \cdot w(S_i) = \sum_{s \in S} \mathbf{1}_{\{\pi(S_i) = \pi_s\}} \cdot w(S_i) \cdot r(S_i) \cdot V_s(k).
\]
Dividing the numerator and denominator of the right-hand side by \(k\) and taking the limit we get the required result.

Lemma 10. One-player priority mean-payoff games with the unique player Min have optimal memoryless deterministic strategies.

Proof. Define a new payoff mapping:
\[
u^*_{r,w,\pi} = \liminf_k \ u^k_{r,w,\pi},
\]
i.e. \(u^*_{r,w,\pi}\) is just like priority mean-payoff (11) with \(\limsup\) replaced by \(\liminf\). Player Min wants to minimize the expectation of \(u_{r,w,\pi}\). Note however that
\[
\text{inf}_r \ E^r_s \{u_{r,w,\pi}\} = - \text{sup}_r \ E^r_s \{-u_{r,w,\pi}\}
\]
and
\[
-u_{r,w,\pi} = \liminf_k \ \sum_{i=0}^{k} \mathbf{1}_{\{\pi(S_i) = \pi_s\}} \cdot w(S_i) \cdot (-r(S_i)) = u^*_{-r,w,\pi}.
\]
Thus a strategy \(\tau\) minimizes the expected payoff \(u_{r,w,\pi}\) if and only if it maximizes the expected payoff \(u^*_{-r,w,\pi}\). Thus instead of a one-player game with player Min and the payoff \(u_{r,w,\pi}\) we can consider a one-player game with player Max and the payoff \(u^*_{-r,w,\pi}\). Unfortunately we cannot use Theorem 6 to prove that Max has an optimal deterministic memoryless strategy for one-player games with payoff \(u^*_{-r,w,\pi}\) since this payoff is not sub-mixing (in fact if we replace \(\limsup\) by \(\liminf\) in (9) then the resulting payoff is not sub-mixing).

Suppose however that \(\tau\) is a memoryless deterministic strategy maximizing the expected payoff \(E\{u_{-r,w,\pi}\}\), such a strategy exists by Lemma 9. By Corollary 9 in the resulting Markov chain \(u_{-r,w,\pi} = u^*_{-r,w,\pi}\) since \(\liminf\) and \(\limsup\) can be replaced by \(\lim\). Thus the same strategy \(\tau\) maximizes the expected payoff \(E\{u^*_{-r,w,\pi}\}\).

Proof of Theorem 5. Theorem 5 is a direct consequence of Lemmas 7 and 10 and the following theorem proved in [11]:

Theorem 11. Under the notation of Section 2.2 we consider the class of games with a payoff mapping \(u_C : C^\omega \to \mathbb{R}\), where \(C\) a set of colors. Suppose all one-player games in the class admit optimal deterministic memoryless strategies (we should consider here all one-player games with player Max as well as all one-player games with player Min). Then all two-player games in the class have also optimal deterministic memoryless strategies for both players.
6. Priority Mean-Payoff as a Limit of Discounted Payoff for Markov Chains

Definition 12. We say that mappings \((a, 1) \ni t \mapsto \lambda_t(s) \in (0, 1), s \in S, a < 1,\) form a regular discount parametrization if

- for each state \(s, \lim_{t \to 1} \lambda_t(s) = 1\) and
- for all states \(x, y \in S\), there exists a limit

\[
w_{x,y} = \lim_{t \to 1} \frac{1 - \lambda_t(x)}{1 - \lambda_t(y)}
\]

(we do not assume that this limit is always finite, we admit the possibility of \(w_{x,y} = \infty\) or equivalently \(w_{y,x} = 0\)).

With a given regular discount parametrization we associate the priority and weight mappings. First we define a precedence relation \(\preceq\) on states: \(y \preceq x\) if \(w_{x,y} < \infty\), where \(x, y \in S\). Note that \(\preceq\) is a total, reflexive and transitive binary relation on \(S\). We write \(x \approx y\) if \(x \preceq y\) and \(y \preceq x\) (which holds iff \(w_{x,y}\) is different from 0 and from \(\infty\)). The priority mapping \(\pi : S \to \mathbb{N}\) is any mapping such that \(\pi(x) = \pi(y)\) if \(x \approx y\) and \(\pi(y) < \pi(x)\) if \(w_{x,y} = 0\). Thus \(\pi(y) < \pi(x)\) iff \(1 - \lambda_0(x)\) tends to 0 much faster than \(1 - \lambda_0(y)\). On the other hand, \(\pi(y) = \pi(x)\) iff \(1 - \lambda_0(x)\) and \(1 - \lambda_0(y)\) tend to 0 with comparable rates. Looking at (3) we can see that, intuitively, if \(x\) and \(y\) are two states visited infinitely often and such that \(\pi(y) < \pi(x)\) then the fact that \(w_{x,y}\) is close to 0 means that the visits to \(y\) contribute much more to the overall discounted payoff than the visits to \(x\).

Finally, as the weight mapping we take any mapping \(S \ni s \mapsto w(s) \in \mathbb{R}_+\) such that for all \(x, y \in S\) with \(x \approx y\) we have \(w_{x,y} = \frac{w(x)}{w(y)}\).

We call \((w, \pi)\) satisfying the conditions above a weighted priority system associated with parametrization \(\lambda_t(s)\). This system is not uniquely defined. For priorities only the order is important, the exact priority values are not important. Similarly all the weights of the states in the same \(\approx\) equivalence class can always be multiplied by a positive constant.

In the sequel we will need the following [17]:

Theorem 13 (Strong Markov Property) Let \((S_i)_{i \in \mathbb{Z}_+}\) be a Markov chain with transition probabilities \(\delta(s, s')\) and let \(T\) be a stopping time. Then, conditional on \(T < \infty\) and \(S_T = s\), \((S_{T+i})_{i \in \mathbb{Z}_+}\) is a Markov chain with the same transition probabilities \(\delta\) and with initial state \(s\) and this chain is independent of \(S_0, \ldots, S_{T-1}\).

Theorem 14. Let \(r : S \to \mathbb{R}\) be a reward mapping and \(\lambda_t\) a regular discount parametrization. Then, for each finite state Markov chain \((S_i)_{i \in \mathbb{Z}_+}\),

\[
\lim_{t \to 1} \mathbb{E}_s\{u_{r,\lambda_t}\} = \mathbb{E}_s\{u_{r,w,\pi}\},
\]

where \((w, \pi)\) is a weighted priority system associated with \(\lambda_t\).
Proof. We first prove the theorem for irreducible Markov chains \((S_i)_{i \in \mathbb{Z}_+}\) such that initial state \(S_0\) has minimal priority, i.e.

\[
\lim_{t \to 1} \frac{1 - \lambda_t(x)}{1 - \lambda_t(S_0)} = w_{x,S_0} < \infty \quad \text{for all states } x.
\]

Let \(T\) be the return time to state \(S_0\), i.e. \(T = \min\{k > 0 \mid S_k = S_0\}\). Note that, by irreducibility of \((S_i)_{i \in \mathbb{Z}_+}\), \(T < \infty\) almost surely. Define the random variable

\[
u = \sum_{i=0}^{T-1} \lambda_t(S_0) \cdots \lambda_t(S_{i-1})(1 - \lambda_t(S_i))r(S_i).
\]

Theorem 13 implies that \(\mathbb{E}\{\sum_{t=0}^{\infty} \lambda_t(S_T) \cdots \lambda_t(S_{T-1})(1 - \lambda_t(S_i))r(S_i)\} = \mathbb{E}\{\nu_{r,\lambda_t}\} \) (since \((S_i)\) and \((S_{T+i})\) have the same distribution and the same initial state) and \(\mathbb{E}\{\nu(S_0) \cdots \nu_t(S_{T-1}) \cdots \nu_0(S_0)\} = \mathbb{E}\{\nu(S_0) \cdots \nu_t(S_{T-1})\cdot \mathbb{E}\{\sum_{t=0}^{\infty} \lambda_t(S_T) \cdots \lambda_t(S_{T-1})(1 - \lambda_t(S_i))r(S_i)\}\) (by independence of the past).

Using these equalities we obtain

\[
\mathbb{E}\{u_{r,\lambda_t}\} = \mathbb{E}\left\{\sum_{i=0}^{\infty} \lambda_t(S_0) \cdots \lambda_t(S_{i-1})(1 - \lambda_t(S_i))r(S_i)\right\}
= \mathbb{E}\left\{\nu_{r,\lambda_t} + \lambda_t(S_0) \cdots \lambda_t(S_{T-1}) \sum_{i=T}^{\infty} \lambda_t(S_T) \cdots \lambda_t(S_{T-1})(1 - \lambda_t(S_i))r(S_i)\right\}
= \mathbb{E}\left\{\nu_{r,\lambda_t} + \nu(S_0) \cdots \nu_t(S_{T-1})\right\} \cdot \mathbb{E}\{u_{r,\lambda_t}\}.
\]

From the equation above we get

\[
\mathbb{E}\{u_{r,\lambda_t}\} = \frac{\mathbb{E}\{\nu_{r,\lambda_t}\}}{1 - \mathbb{E}\{\nu(S_0) \cdots \nu_t(S_{T-1})\}} = \frac{\mathbb{E}\{\nu_{r,\lambda_t}\}}{\frac{1 - \lambda_t(S_0) \cdots \lambda_t(S_{T-1})}{\mu_t(S_0)}},
\]

where for each state \(y\) we set \(\mu_t(y) = 1 - \lambda_t(y)\).

We show that

\[
\lim_{t \to 1} \frac{1 - \lambda_t(S_0) \cdots \lambda_t(S_{T-1})}{\mu_t(S_0)} = \frac{1}{w(S_0)} \sum_{0 \leq i < T} 1_{(\pi(S_i) = \pi(S_0))} \cdot w(S_i).
\]

We have

\[
1 - \frac{\lambda_t(S_0) \cdots \lambda_t(S_{T-1})}{\mu_t(S_0)} = \frac{1 - (1 - \mu_t(S_0)) \cdots (1 - \mu_t(S_{T-1}))}{\mu_t(S_0)}
= \sum_{0 \leq i < T} \mu_t(S_i) \sum_{0 \leq j < i < T} \mu_t(S_j) \mu_t(S_i) + \sum_{0 \leq i < j < k < T} \mu_t(S_i) \mu_t(S_j) \mu_t(S_k) + \cdots
+ (-1)^T \mu_t(S_0) \cdots \mu_t(S_{T-1}).
\]
By (18) the first sum on the right-hand side tends to \( \sum_{0 \leq i < T} w_{S_i, S_0} \) when \( t \to 1 \). All the other sums tend to 0 since, again by (18), \( \frac{\mu_t(S_i)}{\mu_t(S_0)} \) tend to \( w_{S_i, S_0} \) and \( \mu_t(S_i) \) tend to 0. Now it suffices to observe that for \( S_0 \) of minimal priority, \( w_{S_i, S_0} = 0 \) if \( \pi(S_0) < \pi(S_i) \) and \( w_{S_i, S_0} = \frac{\mu_t(S_i)}{\mu_t(S_0)} \) if \( \pi(S_i) = \pi(S_0) \). This terminates the proof of (22).

Now we show that
\[
\lim_{t \to 1} \frac{u^T_{r, \lambda_i}}{\mu_t(S_0)} = \frac{1}{w(S_0)} \sum_{0 \leq i < T} 1_{\{\pi(S_i) = \pi(S_0)\}} w(S_i)r(S_i). \tag{23}
\]
Indeed we have
\[
\sum_{0 \leq i < T} \frac{\lambda_t(S_0) \cdots \lambda_t(S_{i-1})(1 - \lambda_t(S_i))r(S_i)}{\mu_t(S_0)}
= \sum_{0 \leq i < T} \frac{\lambda_t(S_0) \cdots \lambda_t(S_{i-1}) \cdot \frac{\mu_t(S_i)}{\mu_t(S_0)} \cdot r(S_i)}{t} \to \sum_{0 \leq i < T} w_{S_i, S_0}r(S_i)
= \frac{1}{w(S_0)} \sum_{0 \leq i < T} 1_{\{\pi(S_i) = \pi(S_0)\}} w(S_i)r(S_i)
\]
since all \( \lambda_t(S_i) \) tend to 1 and \( w_{S_i, S_0} = 0 \) if \( \pi(S_i) \neq \pi(x) \) and \( w_{S_i, S_0} = \frac{\mu_t(S_i)}{\mu_t(S_0)} \) if \( \pi(S_i) = \pi(S_0) \).

Since the right-hand side of (22) is different from 0, the random variables used on the left-hand side of (22) and (23) are bounded, dominated convergence theorem and (21), (22), and (23) yield
\[
\lim_{t \to 1} \mathbb{E}\{u_{r, \lambda_i}\} = \mathbb{E}\left\{\frac{\sum_{0 \leq i < T} 1_{\{\pi(S_i) = \pi(S_0)\}} w(S_i)r(S_i)}{\sum_{0 \leq i < T} 1_{\{\pi(S_i) = \pi(S_0)\}} w(S_i)}\right\}. \tag{24}
\]

Let \( W = \{s \in S \mid \pi(s) = \pi(S_0)\} \) be the set of all states with minimal priority and, for each state \( s \), let \( V_s(n) = \sum_{i=0}^{n-1} 1_{\{S_i = s\}} \) be the random variable giving the number of visits to \( s \) before \( n \).

Then we can rewrite (24) as
\[
\lim_{t \to 1} \mathbb{E}\{u_{r, \lambda_i}\} = \frac{\sum_{s \in W} \mathbb{E}\{V_s(T)\} \cdot w(s) \cdot r(s)}{\sum_{s \in W} \mathbb{E}\{V_s(T)\} \cdot w(s)}. \tag{25}
\]

Let \( T_i, i \in \mathbb{Z}_+ \), be the sequence of return times to \( S_0 \), \( T_0 = 0, T_{i+1} = \min\{k > T_i \mid S_k = S_0\} \). Since \( (S_i)_{i \in \mathbb{Z}_+} \) is irreducible we have almost surely \( T_i < \infty \) for all \( i \), implying \( \pi(S_0) = \pi_\infty \) almost surely.

By the repeated application of Theorem 13 the random variables \( H_k = S_{T_k}, \ldots, S_{T_{k+1}-1} \) giving the sequence of states between two consecutive visits to \( S_0 \) are independent identically distributed. Note that \( V_s(T_{k+1}) - V_s(T_k) \) giving the number of visits to \( s \) during \( T_k, \ldots, T_{k+1} - 1 \) is a function of \( H_k \). Thus the strong law of large numbers implies that, for each state \( s \), \( \lim_{k \to \infty} \frac{V_s(T_{k+1}) - V_s(T_k)}{k} = \lim_{k \to \infty} \frac{\sum_{0 \leq i < k} (V_s(T_{i+1}) - V_s(T_i))}{k} = \mathbb{E}\{V_s(T)\} \) almost surely.
Similar reasoning shows that \( \lim_{k \to \infty} \frac{E_k}{k} = \mathbb{E}\{T_1\} = \mathbb{E}\{T\} \) almost surely. Thus we have \( \lim_{k \to \infty} \frac{V_k}{k} = \mathbb{E}\{V(T)\} / \mathbb{E}\{T\} \) almost surely. But \( \frac{V_k}{k} \) is just a sub-sequence of the almost surely convergent sequence \( \frac{V(n)}{n} \) (Theorem 8), i.e. both random sequences have the same limit.

Dividing the numerator and denominator of (25) by \( \mathbb{E}\{T\} \) and using the observation above we get
\[
\lim_{t \to 1} \mathbb{E}_x \{u_{r,\lambda_i}\} = \lim_{k \to \infty} \frac{\sum_{s \in W} V_k(s)}{\sum_{s \in W} V_k(s)} \cdot w(s) \cdot r(s)
\]
where the limit on the right-hand side is almost surely \( \mathbb{E}\{u_{r,w,x}\} \), see Corollary 9. This ends the proof for irreducible Markov chains with initial state of minimal priority.

Now let us suppose that \( (S_i)_{i \in \mathbb{Z}_+} \) is any Markov chain. Its set of states can be then partitioned as \( U \cup R_1 \cup \ldots \cup R_k \), where \( R_i \) are closed communicating classes and \( U \) is the set of transient states. For each class \( R_i \) let \( X_i \) be the set of states that have minimal priority in \( R_i \), \( X_i = \{ x \in R_i \mid \pi(x) \leq \pi(s) \text{ for all } s \in R_i \} \). Almost surely the Markov chain will hit the set \( X := \bigcup_{i=1}^k X_i \), let \( T = \min\{i \geq 0 \mid S_i \in X\} \) be the moment of the first visit to \( X \).

Proceeding like in (20) but taking the conditional expectation rather than the expectation we get for \( x \in X \)
\[
\mathbb{E}_x \{u_{r,\lambda_i}\} = \mathbb{E}_x \{u_{r,\lambda_i}\} + \mathbb{E}_x \{\lambda_0(S_0) \ldots \lambda_{T}(S_{T-1})\} \mathbb{E}_x \{u_{r,\lambda_i}\}.
\]
But \( \lim_{t \to 1} \mathbb{E}_x \{u_{r,\lambda_i}\} = 0 \) and \( \lim_{t \to 1} \mathbb{E}_x \{\lambda_0(S_0) \ldots \lambda_{T}(S_{T-1})\} = 1 \) since the functions under integral tend to 0 and 1 respectively (use the dominated convergence). Thus \( \lim_{t \to 1} \mathbb{E}_x \{u_{r,\lambda_i}\} = \lim_{t \to 1} \mathbb{E}_x \{u_{r,\lambda_i}\} = \mathbb{E}_x \{u_{r,w,x}\} \), where the last equality follows from the first part of the proof and from the fact that the Markov chain starting in a state of \( X \) is recurrent. On the other hand \( \mathbb{E}_x \{u_{r,\lambda_i}\} = \sum_{x \in X} P_x \{S_T = x\} \mathbb{E}_x \{u_{r,\lambda_i}\} \mathbb{E}_x \{u_{r,\lambda_i}\} = \sum_{x \in X} P_x \{S_T = x\} \mathbb{E}_x \{u_{r,\lambda_i}\} = \mathbb{E}_x \{u_{r,w,x}\} \).

7. Priority Mean-Payoff Games as Limits of Discounted Games – Two-Player Case

Let us fix an arena \( A \) and a reward mapping \( r \) on \( A \). Let \( t \mapsto \lambda_t(s) \) be a regular discount parametrization. We say that memoryless deterministic strategy \( \sigma^d \) of Max is limit optimal if for each \( 0 < \varepsilon \) there exists \( t \in (1 - \varepsilon, 1) \) such that \( \sigma^d \) is optimal in the discounted game with payoff \( u_{r,\lambda_i} \). Limit optimal strategies of player Min are defined in a similar way.

Note that limit optimal strategies exist. This follows from the fact that in discounted games both players have optimal memoryless deterministic strategies and for a given game there is only a finite number of deterministic memoryless strategies.

For a memoryless deterministic strategy \( \sigma^d \) of Max let \( O(\sigma^d) \) be the set of \( t < 1 \) such that \( \sigma^d \) is optimal for \( \lambda_t \).
Note that $\sigma^\sharp$ is limit optimal if and only if $\sup O(\sigma^\sharp) = 1$. The identical equivalence holds also for player Min. To compare limit optimality with Blackwell optimality let us note that $\sigma^\sharp$ is Blackwell optimal if $O(\sigma^\sharp)$ contains an interval $(1-\varepsilon, 1)$ for some positive $\varepsilon$. Clearly Blackwell optimality implies limit optimality.

Note also that even if $\sup O(\sigma^\sharp) = 1$ and $\sup O(\tau^\sharp) = 1$ for a strategy profile $(\sigma^\sharp, \tau^\sharp)$ this does not imply that $\sup(O(\sigma^\sharp) \cap O(\tau^\sharp)) = 1$, it may happen that there exists $\varepsilon > 0$ such that $\sigma^\sharp$ and $\tau^\sharp$ are never simultaneously optimal for any $t \in (1-\varepsilon, 1)$ even if they are both limit optimal.

The following theorem shows that priority mean-payoff games can be seen as limits of discounted games. What is interesting is that not only the values of the discounted games tend to the value of a priority mean-payoff game but also the strategies limit optimal for discounted games are optimal in the corresponding priority mean-payoff game.

**Theorem 15.** Let $A$ be an arena and let $t \mapsto \lambda_t$ a regular discount parametrization for $A$. Let $(w, \pi)$ be a weighted priority system associated with $\lambda_t$. Then

(a) each memoryless deterministic strategy limit optimal for payoff $u_{r,\lambda_t}$ is optimal for the priority mean-payoff game with payoff $u_{r,w,\pi}$,

(b) if $\sigma^\sharp$ and $\tau^\sharp$ are limit optimal for the discounted game then

$$\lim_{t \uparrow 1} \mathbb{E}^{\sigma^\sharp,\tau^\sharp}_{s} \{u_{r,\lambda_t}\} = \text{val}_s(r_{r,\pi}),$$

where $\text{val}_s(r_{r,\pi})$ is the value of the game $(A, u_{r,w,\pi})$,

(c) for each state $s$, $\lim_{t \uparrow 1} \text{val}_s(r_{r,\lambda_t}) = \text{val}_s(r_{r,w,\pi})$.

**Proof.** (a) Suppose that $\sigma^\sharp$ is limit optimal for player Max, $\sup O(\sigma^\sharp) = 1$. For each $t \in O(\sigma^\sharp)$ there exists a deterministic memoryless strategy for Min that is optimal for the corresponding discounted game, this strategy depends on $t$. However player Min has a finite number memoryless deterministic strategies, this implies that there exists a $X \subset O(\sigma^\sharp)$ with $\sup X = 1$ and a strategy $\tau^\sharp$ for Min such that $(\sigma^\sharp, \tau^\sharp)$ is an optimal strategy profile for all $\lambda_t$ with $t \in X$. Therefore

$$\mathbb{E}^{\sigma^\sharp,\tau^\sharp}_{s} \{u_{r,\lambda_t}\} \leq \mathbb{E}^{\sigma^\sharp,\tau^\sharp}_{s} \{u_{r,\lambda_t}\} \leq \mathbb{E}^{\tau^\sharp}_{s} \{u_{r,\lambda_t}\}$$

for all $t \in X$ and all deterministic memoryless strategies $\sigma, \tau$ of Max and Min. By Theorem 14, with $t \uparrow 1$ over $X$, we get

$$\mathbb{E}^{\sigma^\sharp,\tau^\sharp}_{s} \{u_{r,w,\pi}\} \leq \mathbb{E}^{\sigma^\sharp,\tau^\sharp}_{s} \{u_{r,w,\pi}\} \leq \mathbb{E}^{\sigma^\sharp,\tau^\sharp}_{s} \{u_{r,w,\pi}\},$$

which shows that $\sigma^\sharp$ and $\tau^\sharp$ are optimal in the class of deterministic memoryless strategies for the priority mean-payoff game. But Theorem 5 implies that for priority mean-payoff games strategies optimal in the class of deterministic memoryless strategies are optimal also when all strategies are allowed. For limit optimal strategies of player Min the proof is identical.

(b) From (a) it follows that $\mathbb{E}^{\sigma^\sharp,\tau^\sharp}_{s} \{u_{r,w,\pi}\} = \text{val}_s(r_{r,w,\pi})$. But, by Theorem 14, $\lim_{t \uparrow 1} \mathbb{E}^{\sigma^\sharp,\tau^\sharp}_{s} \{u_{r,\lambda_t}\} = \mathbb{E}^{\sigma^\sharp,\tau^\sharp}_{s} \{u_{r,w,\pi}\}$.

(c) Suppose that (c) does not hold. Then there exists $\varepsilon > 0$ and a sequence $t_n < 0$ with $\lim_n t_n = 1$ such that $|\text{val}_s(u_{r,\lambda_{t_n}}) - \text{val}_s(u_{r,w,\pi})| > \varepsilon$. Since players Max and
Min have finitely many deterministic memoryless strategies, there exists an infinite subsequence \( t_{nm} \) of \( t_n \) and memoryless deterministic strategies \( \sigma^\#, \tau^\# \) optimal for all \( t_{nm} \). But, by (b), \( \text{val}(u_r, \lambda_{t_{nm}}) = E_{\sigma^\#, \tau^\#}^{t_{nm}} \{ u_r, \lambda_{t_{nm}} \} \xrightarrow{m \to \infty} \text{val}(u_r, w, \pi) \), a contradiction.

If instead of regular discount parametrization we consider rational discount parametrization of Section 4 then we obtain stronger results.

First we shall show that if \( \lambda_t \) is a rational discount parametrization then

\[
\lambda_t(s) = 1 - g_s(t)(1 - t)^\pi(s),
\]

where \( t \mapsto g_s(t) \) is a rational function such that \( g_s(1) > 0 \).

Indeed the fact that \( \lim_{t \to 1} (1 - \lambda_t(s)) = 0 \) implies that, for each state \( s \), the rational function \( t \mapsto 1 - \lambda_t(s) \) factorizes as \( g_s(t)(1 - t)^\pi(s) \), where \( \pi(s) \in \mathbb{N} \) and \( t \mapsto g_s(t) \) is a rational function such that \( g_s(1) \neq 0 \). Moreover since \( 1 - \lambda_t(s) \) is positive for \( t \in (1 - \varepsilon, 1) \), \( g_s(t) \) is also positive in the same interval and by continuity \( g_s(1) > 0 \).

Clearly (26) implies that a rational discount parametrization is regular.

If \( t \mapsto \lambda_t(s) \) is a rational parametrization for each state \( s \), take \( \pi(s) \) from (26) as the priority of \( s \) and \( w(s) := g_s(1) \) as the weight of \( s \). We say that \( (w, \pi) \) defined in this way is the weighted priority system associated with the rational discount parametrization \( \lambda_t \). Note that the weighted priority system associated with a rational discount parametrization is just a special case of a weighted priority system that we have associated with any regular parametrization. Thus Theorem 15 holds for rational discount parametrizations. However we can prove more in this case:

**Corollary 16.** Let \( \lambda_t \) be a rational discount parametrization for \( \mathcal{A} \) and let \( (w, \pi) \) be the associated weighted priority system. Then all Blackwell optimal strategies for \( (\mathcal{A}, u_r, \lambda_t) \) are optimal for \( (\mathcal{A}, u_r, w, \pi) \).

**Proof.** By Theorem 15 since Blackwell optimal strategies are limit optimal.

Clearly it is Corollary 16 in conjunction with Theorem 3 that constitute the backbone of the paper and the concept of Blackwell optimality is much more interesting than that regular optimality.

However rational discount parametrizations are rather special and a natural question is what happens if we take parametrizations that are not rational. And Theorem 15 shows what can be proved with weaker assumptions concerning discount parametrization. One can ask if Blackwell optimal strategies exist for regular discount parametrizations (without rationality assumption). Without some additional assumptions on \( \lambda_t \) this seems little probable. However if \( \lambda_t \) is a regular discount parametrization such that the mapping \( t \mapsto \lambda_t \) is continuous on some interval \( (1 - \varepsilon, 1) \) then the existence of Blackwell optimal strategies seems to be much more
plausible. Note for example that Lemma 4 remains valid if we replace “rational” by “continuous” (in the statement and in the proof).

8. Final Remarks

One can wonder what we gain by linking discounted games with priority mean-payoff games. We can always pretend that the fact that we can relate two apparently different games has an intrinsic interest. However while the priority mean-payoff games contain and generalize the parity games and the latter have many applications in computer science [12], discounted games with many different discount factors and with payoff given by a rather obscure formula (3) look neither natural nor appealing (contrary to discounted games with one discount factor which are widely studied in game theory). In the seminal paper [18] Shapley gives another, much more attractive, interpretation of (3) in terms of stopping games. For each state $s$, $1 - \lambda(s)$ is interpreted as the probability that the nature will stop the game when $s$ is visited. Since $0 \leq \lambda(s) < 1$ for all $s \in S$, the stopping probabilities are strictly positive which implies that the game will eventually stop with probability 1 after a finite number of steps. If the game stops in $s$ then player Max receives from player Min the payment $r(s)$. Let us stress that player Max receives the payoff only once, when the game stops and the amount received is determined solely by the state in which the game ends. If the game does not stop in $s$ at stage $i$ then there is no payment at this stage and the player controlling state $s$ chooses an action to execute. Note that $\lambda(s_0) \ldots \lambda(s_{i-1})(1 - \lambda(s_i))$ is the probability that the game has not stopped in any of the states $s_0, \ldots, s_{i-1}$ but it does stop in the state $s_i$. Since this event results in the payment $r(s_i)$, (3) represents in this interpretation the payoff expectation for an infinite play $s_0 a_0 s_1 a_1 s_2 a_2 \ldots$. For two states $x$ and $y$ such that $\lim_{t \to 1} \frac{1 - \lambda_t(x)}{1 - \lambda_t(y)} = 0$ and for all $t$ close to 1, the probability to stop at $x$ in the stopping game becomes negligible comparing to the probability of stopping at $y$. This is reflected by the fact that in the associated priority mean-payoff game the priority of $y$ is smaller than the priority of $x$ (recall that it is precisely the states with smaller priorities that are important for priority mean-payoff games). On the other hand if $\lim_{t \to 1} \frac{1 - \lambda_t(x)}{1 - \lambda_t(y)} = \frac{w(x)}{w(y)}$ is finite and different from 0 then, intuitively, for $t$ close to 0 the probabilities to stop in $x$ and in $y$ are comparable and in the limit this relative probabilities are captured by the weights of both states.

The concept of Blackwell optimality can be used in order to refine the notion of optimal strategies. From Corollary 16 we know that Blackwell optimal strategies are optimal for priority mean-payoff games, the converse is not true (an example for parity games will follow). Already Blackwell noted that Blackwell optimal strategies allow to distinguish between two histories that otherwise appear identical (yield the same payoff). His example deals with mean-payoff games and discounted games with a unique discount factor. Consider two histories, $h_1$ begins with a very long but finite sequence of rewards 1 followed by an infinite sequence of rewards 0. The mean payoff for $h_1$ is 0, the initial sequence of 1 does not count in the limit. Consider now the
history \( h_2 \) consisting of an infinite sequence of rewards 0. Here also the mean payoff is 0. Even if the mean-payoff is the same for both histories from the point of view of Max the first history is much better than the second one, conversely Min would prefer \( h_2 \) to \( h_1 \). That the players should distinguish between \( h_1 \) and \( h_2 \) becomes even more evident if we take into account that infinite games do not exist in real life (all interactions always span over a finite time, no system has in infinite life time). Discounted payoffs are different for \( h_1 \) and for \( h_2 \), even if discount factor is very close to 1 the discounted payoff for \( h_1 \) is positive and the that of \( h_2 \) is 0 thus Blackwell optimal strategies can distinguish between \( h_1 \) and \( h_2 \).

Also for parity games the traditional notion of optimal strategies seems to be not sufficiently selective. Parity games have purely qualitative character (only binary payoffs 0 and 1), it would be interesting to have a finer graduation of payoffs and to add some “quantitative” properties. On the other hand we do not want to lose the most useful property of parity games, the existence of optimal memoryless deterministic strategies. There is some ongoing effort to define “quantitative” parity games and Blackwell optimality can be seen as a step in this direction. possibilities.

Consider the game where for even priorities the reward is 1 and for odd priorities the reward is 0, \( r(i) = i \mod 2 \) for states with priority \( i \). The discount factor for states with priority \( i \) is \( \lambda_t(i) = 1 - (1 - t)^i \) (the canonical rational parametrization). Consider two infinite priority histories \( h_1 = (12)\omega \) and \( h_2 = 1\omega \). In both histories the minimal priority visited infinitely often is 1 resulting in payoff 1 for parity games. However discounted payoff with discount factors described above is different for both histories, for \( h_2 \) the discount payoff is still 1 but for \( h_1 \) it is smaller than 1, thus for Blackwell optimality the two histories are not equivalent. One can argue that this is because the frequency of the minimal priority is not the same for both histories, in \( h_1 \) the frequency of the minimal priority is 1/2, in \( h_2 \) it is 1. However Blackwell optimality is more subtle. Consider the priority history \( h_3 = (21)\omega \) with the reward and discount factors as above. Again the parity payoff is 1 for \( h_1 \) and \( h_3 \) and moreover \( h_1 \) and \( h_3 \) have the same frequency of 1. However the discounted payoff is different for \( h_1 \) and \( h_3 \), it is greater for \( h_1 \) for all \( t \) sufficiently close to 1. Thus Blackwell optimality distinguishes between \( h_1 \) and \( h_3 \) while the frequency analysis cannot take them apart. The informal intuition is that \( h_1 \) is preferred to \( h_3 \) by player Max since the better priority 1 arrives earlier in \( h_1 \) than in \( h_3 \). Of course these examples give just a very rough and inexact idea how Blackwell optimality can distinguish histories, the only exact reasoning is in terms of stopping games where a lower priority of a state translates into a much higher probability to stop in such a state.

References

http://hal.archives-ouvertes.fr/hal-00140144/fr/.
http://hal.archives-ouvertes.fr/hal-00438359/en/, 2010.