

R U T C O R  
R E S E A R C H  
R E P O R T

EVERY STOCHASTIC GAME WITH  
PERFECT INFORMATION ADMITS A  
CANONICAL FORM

Endre Boros<sup>a</sup>      Vladimir Gurvich<sup>b</sup>  
Khaled Elbassioni<sup>c</sup>      Kazuhisa Makino<sup>d</sup>

RRR 9-2009, MAY 2009

RUTCOR  
Rutgers Center for  
Operations Research  
Rutgers University  
640 Bartholomew Road  
Piscataway, New Jersey  
08854-8003  
Telephone:      732-445-3804  
Telefax:      732-445-5472  
Email:      rrr@rutcor.rutgers.edu  
<http://rutcor.rutgers.edu/~rrr>

---

<sup>a</sup>RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ  
08854-8003; (boros@rutcor.rutgers.edu)

<sup>b</sup>RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ  
08854-8003; (gurvich@rutcor.rutgers.edu)

<sup>c</sup>Max-Planck-Institute for Informatics; Stuhlsatzenhausweg 85, 66123,  
Saarbruecken, Germany; (elbassio@mpi-sb.mpg.de)

<sup>d</sup>Graduate School of Information Science and Technology, University of  
Tokyo, Tokyo, 113-8656, Japan; (makino@mist.i.u-tokyo.ac.jp)

## RUTCOR RESEARCH REPORT

RRR 9-2009, MAY 2009

# EVERY STOCHASTIC GAME WITH PERFECT INFORMATION ADMITS A CANONICAL FORM

**Abstract.** We consider discounted and undiscounted stochastic games with perfect information in the form of a natural BWR-model with positions of three types:  $V_B$  Black,  $V_W$  White,  $V_R$  Random. These BWR-games lie in the complexity class  $NP \cap CoNP$  and contain the well-known cyclic games (when  $V_R$  is empty) and Markov decision processes (when  $V_B$  or  $V_W$  is empty). We show that the BWR-model is polynomial-time equivalent with the classical Gillette model, and, as follows from a recent result by Miltersen (2008), with simple stochastic games (so called Condon's games), as well.

Furthermore, we consider standard potential transformations  $r_x(v, u) = r(v, u) + x(v) - \beta x(u)$  of the local reward function  $r$ , where  $\beta \in [0, 1)$  is the discount factor and  $\beta = 1$  in the undiscounted case. As our main result, we show that every BWR-game can be reduced by such a transformation to a *canonical form* in which *locally* optimal strategies are *globally* optimal, and hence the value for every initial position and the optimal strategies of both players are obvious. Standardly, the optimal strategies are *uniformly* optimal (or *ergodic*, that is, do not depend on the initial position) and coincide with the optimal strategies of the original BWR-game; while the original values are transformed by a very simple formula:  $\mu_x(v) = \mu(v) + (1 - \beta)x(v)$ .

In the discounted case,  $\beta < 1$ , the transformed values are also ergodic and the corresponding potentials can be found in polynomial time. Yet, this time tends to infinity, as  $\beta \rightarrow 1^-$ .

---

**Acknowledgements:** This research was partially supported by DIMACS, a collaborative project of Rutgers University, Princeton University, AT&T Labs-Research, Bell Labs, NEC Laboratories America and Telcordia Technologies, as well as affiliate members Avaya Labs, HP Labs, IBM Research, Microsoft Research, Stevens Institute of Technology, Georgia Institute of Technology and Rensselaer Polytechnic Institute. DIMACS was founded as an NSF Science and Technology Center.

The second author also thankfully acknowledges the partial support provided by the Research Foundation and Center for Algorithmic Game Theory, University of Aarhus.

# 1 Introduction and sketch of the results

We consider the well-known class of two-player, zero-sum, turn-based, stochastic games with perfect information introduced in the classical paper of Gillette [Gil57]. In such a game, there are two players *White* or *maximizer* and *Black* or *minimizer*, which control a finite set of states  $V$ , partitioned into two sets, each controlled by one of the players. At each state  $u \in V$ , the controlling player chooses one of a finite set of possible actions, according to which a real-valued reward is paid by Black to White, and then a new state is reached with a certain probability which depends on the action. The play continues forever, and White's objective is to maximize her *limiting average payoff*, defined as

$$c = \liminf_{n \rightarrow \infty} \frac{\sum_{i=0}^n b_i}{n+1}, \quad (1)$$

where  $b_i$  is the reward paid to White at step  $i$  of the game. Similarly, the objective of Black is to minimize  $\limsup_{n \rightarrow \infty} \frac{\sum_{i=0}^n b_i}{n+1}$ . The fact that a *saddle point* exists in *pure positional strategies* was proved by Gillette [Gil57] and Liggett and Lippman [LL69]<sup>1</sup> by considering the *discounted* version, in which the payoff of White is discounted by a factor  $\beta^i$  at step  $i$ , giving the effective payoff:

$$a_\beta = (1 - \beta) \sum_{i=0}^{\infty} \beta^i b_i,$$

and then proceeding to the limit as the *discount factor*  $\beta \in [0, 1)$  goes to 1.

In this paper, we consider another class of games, which was first suggested in [GKK88], and recently considered under the name of *Stochastic Mean payoff games* in [CH08], and which will turn out to be equivalent with Gillette games with perfect information (GGPI). Each such game, which we call a *BWR-game*, is played by two players, White and Black, on an arc-weighted directed graph  $G = (V = V_B \cup V_W \cup V_R, E)$ , with given local rewards  $r : E \rightarrow \mathbb{R}$  and 3 types of vertices: black  $V_B$ , controlled by Black; white  $V_W$ , controlled by White; and random  $V_R$ , controlled by nature. When the play is at a white (black) vertex, White (resp., Black) selects a outgoing arc and Black pays White the reward on that arc. When the play is at a random vertex  $v$ , a vertex  $u$  is picked with specified probability  $p(v, u)$  and again Black pays White the value on the arc  $(v, u)$ . The play continues forever, and White aims to maximize (Black aims to minimize) the limiting average payoff, defined as in (1).

The special case when there are no random nodes, is known as *cyclic games* or *mean payoff games*, which were initially considered for complete bipartite digraphs in [Mou76b, Mou76a],

---

<sup>1</sup>Gillette proved the existence of pure equilibrium in the  $\beta$ -discounted case, and then for the undiscounted case by proceeding to the limit as  $\beta \rightarrow 1^-$ . However, his proof contained a technical error which was later fixed by Liggett and Lippman [LL69].

for all (not necessarily complete) bipartite digraphs in [EM79], and for arbitrary digraphs in [GKK88]. A further special case of this was considered extensively in the literature under the name of *parity games* [BV01a, BV01b, CJH04, Hal07, Jur98, JPZ06], and later generalized also to include random nodes in [CH08]. The game is reduced to the *minimum mean cycle problem* in case  $V_W = V_R = \emptyset$ , see for example [Kar78]. On the other hand, if one of the sets  $V_B$  or  $V_W$  is empty, we obtain a *Markov decision process*; see, for example, [MO70], and if both are empty  $V_B = V_W = \emptyset$ , we get a *weighted Markov chain*.

In the special case of a BWR-game when all rewards are zero except at a single node  $t$  called the terminal, at which there is a self-loop with reward 1, we obtain the so-called *simple stochastic games* (SSG), introduced by Condon [Con92]. In these games, the objective of White is to maximize the probability of reaching the terminal while Black wants to minimize this probability. SSG's have also been considered in several papers [Con92, Con93, GH08, Hal07].

It is easy to see that BWR-games are a special case of Gillette games. In fact, the opposite turns also out to be true, see Theorem 4 below. Less obvious is the following sequence of reductions due to Miltersen [GM08]:

$$\text{Undiscounted Gillette games} \subseteq_P \text{discounted Gillette games} \subseteq_P \text{SSG},$$

where  $\subseteq_P$  denotes polynomial-time reductions<sup>2</sup>. Thus, by a recent result of Halman [Hal07], all these games can be solved in randomized strongly subexponential time  $2^{O(\sqrt{n \log n})}$ , where  $n$  is the number of states or vertices; see section 6.4 for more details. Note that a number of pseudo-polynomial and subexponential algorithms already exists for mean payoff games [GKK88, KL93, Pis99, BV07, HBV04, Hal07, ZP96]; see also [DG06], and for parity games [JPZ06].

Given a BWR-game, we consider potential transformations  $x : V \rightarrow \mathbb{R}$ , assigning a real-value  $x(v)$  to each vertex  $v \in V$ , and transforming the local reward on each arc  $(v, u)$  into  $r_x(v, u) = r(v, u) + x(v) - x(u)$ . It is known that for cyclic games, there exists such a transformation such that, in the transformed game, the locally optimal strategies are globally optimal, and hence, the value and optimal strategies become trivial [GKK88]. An interesting question is whether such transformations also exist for the more general class of BWR- or Gillette games. Our main result is that such transformations do exist: in the transformed game, the equilibrium value  $\mu(v) = \mu_x(v)$  is given simply by the maximum local reward for  $v \in V_W$ , the minimum local reward for  $v \in V_B$ , and the average local reward for  $v \in V_R$ . In this case we say that the transformed game is in *canonical* form.

In the  $\beta$ -discounted case, the potential will be of the form  $r_x(v, u) = r(v, u) + x(v) - \beta x(u)$  [GKK88], while the original values are transformed by a very simple formula:  $\mu_x(v) =$

---

<sup>2</sup>Note that even though the reduction is not strong, it involves only numbers bounded polynomially in the bit length of the input.

$$\mu(v) + (1 - \beta)x(v).$$

**Theorem 1** *Any BWR-game or Gillette game, discounted or undiscounted, can be brought by a potential transformation to canonical form. For the  $\beta$ -discounted case, the corresponding vector of potentials can be determined by a polynomial-time algorithm for every fixed  $\beta$ .*

Theorem 1 not only gives a new constructive proof of the existence of a saddle point in pure strategies for all types of games mentioned above, but also a simple and efficient procedure when  $\beta < 1$  is fixed. However, the execution time asymptotically increases as  $(1 - \beta)^{-1}$ , as  $\beta \rightarrow 1^-$ . Theorem 1 also gives rise to a so-called *certifying algorithm* (see e.g. [KMMS03]), in the sense that, given an optimal pair of strategies, the vector of potentials provided by the algorithm can be used to verify optimality in *linear* time (otherwise verifying optimality requires solving a system of equations).

For the special case of Markov decision processes (when  $V_B$  or  $V_W$  is empty), the potentials mentioned in the theorem correspond to the *dual* variables in the standard linear programming formulation; see e.g. [MO70]<sup>3</sup>.

The special case when  $V_R = \emptyset$  was shown in [GKK88]. However, it is not clear how the algorithm given in [GKK88] can be generalized to the case with random nodes. Instead, our proof of Theorem 1 follows in the spirit of [Gil57, LL69] (see also [MO70], Chapter 4): First, we consider the discounted case; see Section 4, then, in Section 5, we reduce the undiscounted BWR-games to canonical form by just choosing  $\beta$  sufficiently close to 1. However, such approach requires exponential time, since one must choose  $\beta > 1 - \varepsilon/2^n$  to approximate the value of an undiscounted BWR-game with accuracy  $\varepsilon$ ; see Section 4 and Appendix C.4.

**Remark 1** *In the absence of random positions, the above approach becomes more efficient. To get the exact solution of an undiscounted BW-game with  $n = |V|$  positions and range of the reward function  $R$ , it is enough to solve the corresponding  $\beta$ -discounted game with any  $\beta > 1 - 1/(4n^3|R|)$  [ZP96]. Thus, Theorem 1 gives a new (simple) pseudo-polynomial (linear in  $R$ ) algorithm for the undiscounted BW-games; for other algorithms see [GKK88, ZP96, Pis99].*

Finally, we combine the above results with our reduction of the Gillette games to BWR-games to derive the existence of the canonical form for both the discounted and undiscounted Gillette games; see Section 6.

---

<sup>3</sup>In fact, one can use Theorem 1 to derive the dual LP-formulation for Markov decision processes.

## 2 Preliminaries

### 2.1 BWR-games in positional form

Let  $G = (V = V_W \cup V_B \cup V_R, E)$  be a directed graph (digraph) that may have loops and multiple arcs. We assume without any loss of generality that  $G$  has no terminal vertices, i.e., vertices of out-degree 0, (otherwise, one can add a loop to each terminal vertex.) There are two players: White, the maximizer, and Black, the minimizer. The vertices of  $G$  are called *positions*. They are partitioned into three sets  $V = V_B \cup V_W \cup V_R$  and called respectively *Black*, *White*, and *Random* positions.

An arc  $(v, u) \in E$  is called a *move* from position  $v$  to  $u$ . This move is chosen by a player, White if  $v \in V_W$  and Black if  $v \in V_B$ , or by chance if  $v \in V_R$ . In the latter case a probability  $p(v, u)$  is assigned to each arc  $(v, u) \in E$ , i.e.,  $0 \leq p(v, u) \leq 1$  for all  $v \in V_R, u \in V$  and  $\sum_{u | (v,u) \in E} p(v, u) = 1 \quad \forall v \in V_R$ . For convenience we will assume that  $p(v, u) > 0$  whenever  $(v, u) \in E$  and  $v \in V_R$ , and set  $p(v, u) = 0$  for  $(v, u) \notin E$ . Let us denote by  $P$  the obtained set of probability distributions for all  $v \in V_R$ . Furthermore, to each arc  $(v, u) \in E$  is assigned a real number  $r(v, u)$  called the *local reward* (cost or payoff) on  $(v, u)$ . It is assumed that Black pays and White gets  $r(v, u)$  whenever the move  $(v, u)$  is made in the game.

Finally, let us fix an initial position  $v_0 \in V$  or, more generally, an initial probability distribution  $p_0$  on  $V$  assuming that  $0 \leq p_0(v) \leq 1$  for each  $v \in V$  and  $\sum_{v \in V} p_0(v) = 1$ .

The game *in positional form* is then defined by the quadruple  $\mathcal{G} = (G, P, p_0, r)$ .

### 2.2 Effective limiting payoffs in weighted Markov chains

When  $V_B = V_W = \emptyset$ , and hence  $V = V_R$  consists only of random nodes, we obtain a *weighted Markov chain*. In this case  $P : V \times V \rightarrow [0, 1]$  is the probabilistic  $n \times n$  matrix (so-called *transition matrix*) whose entry  $p(v, u)$  is the probability of transition from  $v$  to  $u$  in one move, for every pair of positions  $v, u \in V$ . Then, it is obvious and well-known that for every integral  $i \geq 0$  matrix  $P^i : V \times V \rightarrow [0, 1]$  (the  $i$ -th power of  $P$ ) is the  $i$ -move transition matrix, whose entry  $p_i(v, u)$  is the probability of transition from  $v$  to  $u$  in exactly  $i$  moves, for every  $v, u \in V$ .

Let  $q_i(v, u)$  be the probability that that arc  $(v, u) \in E$  will be the  $(i+1)$ -st move, given the original distribution  $p_0$ , where  $i = 0, 1, 2, \dots$ , and denote by  $q_i : E \rightarrow [0, 1]$  the corresponding probabilistic  $|E|$ -vector. For convenience, we introduce  $|V| \times |E|$  vertex-arc transition matrix  $Q : V \times E \rightarrow [0, 1]$  whose entry  $q(\ell, (v, u))$  is equal to  $p(v, u)$  if  $\ell = v$  and 0 otherwise, for every  $\ell \in V$  and  $(v, u) \in E$ . Then, it is clear that  $q_i = p_0 P^i Q$ .

Let  $r$  be the  $|E|$ -dimensional vector of local rewards, and  $b_i$  denote the expected reward for the  $(i+1)$ -st move;  $i = 0, 1, 2, \dots$ , i.e.,  $b_i = \sum_{(v,u) \in E} q_i(v, u) r(v, u) = p_0 P^i Q r$ . Then the *effective payoff* of the *undiscounted* weighted Markov chain is defined to be the average

expected reward on the limit, i.e.,  $c = \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{i=0}^n b_i$ . It is well-known (see Appendix B.1) that this is equal to the limit as  $\beta \rightarrow 1^-$  of the  $\beta$ -discounted payoff, given by  $a_\beta = (1 - \beta) \sum_{i=0}^{\infty} \beta^i b_i$ :

$$c = \lim_{\beta \rightarrow 1^-} a_\beta = p_0 P^* Q r, \quad (2)$$

where  $P^*$  is the *limit Markov matrix* (see Appendix B.2 for more details).

### 2.3 BWR-games in normal form

Standardly, we define a strategy  $s_W \in S_w$  (respectively,  $s_B \in S_B$ ) as a mapping that assigns a move  $(v, u) \in E$  to each position  $v \in V_W$  (respectively,  $v \in V_B$ ). A pair of strategies  $s = (s_W, s_B)$  is called a *situation*. Given a BWR-game  $\mathcal{G} = (G, P, p_0, r)$  and situation  $s = (s_B, s_W)$ , let us define a weighted Markov chain  $\mathcal{G}_s = (G, P_s, p_0, r)$ . To do so, we just extend probability distributions  $P_s$  as follows:

$$\begin{aligned} P_s(v, u) &= 1 \text{ whenever } v \in V_W \text{ and } u = s_W(v) \text{ or } v \in V_B \text{ and } u = s_B(v); \\ P_s(v, u) &= 0 \text{ whenever } v \in V_W \text{ and } u \neq s_W(v) \text{ or } v \in V_B \text{ and } u \neq s_B(v). \end{aligned} \quad (3)$$

In other words, in every position  $v \in V_B \cup V_W$  we assign probability 1 to the (unique) move prescribed by the strategy  $s_B$  or  $s_W$  and probability 0 to every other move. For random position  $v \in V_R$  we assign  $P_s(v, u) = p(v, u) > 0$  if  $(v, u) \in E$  and  $P_s(v, u) = p(v, u) = 0$  otherwise, as before.

In the obtained weighted Markov chain  $\mathcal{G}_s = (G, P_s, p_0, r)$ , we define the discounted and undiscounted limiting (mean) effective payoffs  $a_s(\beta)$  and  $c_s = \lim_{\beta \rightarrow 1^-} a_s(\beta)$  as above. Thus, we obtain respectively a discounted and undiscounted matrix game  $A(\beta) : S_W \times S_B \rightarrow \mathbb{R}$ , and  $C : S_W \times S_B \rightarrow \mathbb{R}$ .

### 2.4 Solvability and ergodicity

It is known that every such game has a saddle point in pure strategies [Gil57, LL69]. In other words, both  $\beta$ -discounted and undiscounted BWR-games are solvable in pure positional strategies. Moreover, there are optimal strategies that do not depend on the original probability distribution  $p_0$ , so-called *ergodic* optimal strategies. In contrast, the value of the game can depend on  $p_0$ .

From now on, let us restrict our attention to unit probabilistic vectors  $p_0$ ; in other words, we will fix an initial position  $v_0 \in V$ . Then, a BWR-game is given by a quadruple  $\mathcal{G} = (G, P, v_0, r)$ , where  $G = (V = V_B \cup V_W \cup V_R, E)$ . The triplet  $\mathcal{G} = (G, P, r)$  is called a *not initialized BWR-game*. Furthermore,  $\mathcal{G}$  is called *ergodic* if the value  $\mu(v_0)$  of each corresponding BWR-game  $(G, P, v_0, r)$  is the same for all initial positions  $v_0 \in V$ .

## 2.5 Discounted and undiscounted potential transforms

Given a ( $\beta$ -discounted) game  $\mathcal{G} = (G, P, v_0, r)$ , let us introduce a mapping  $x : V \rightarrow \mathbb{R}$ , whose values  $x(v)$  will be called *potentials*, and define the transformed reward function  $r_x : E \rightarrow \mathbb{R}$  by formula:

$$r_x(v, u) = r(v, u) + x(v) - \beta x(u), \quad \text{where } (v, u) \in E \text{ and } 0 \leq \beta \leq 1. \quad (4)$$

This transform is called discounted if  $\beta < 1$  and undiscounted if  $\beta = 1$ .

We note that the two normal form matrices  $A_x(\beta)$  and  $A(\beta)$ , of the obtained game  $\mathcal{G}_x$  and the original game  $\mathcal{G}$ , are related in the following very simple way:

**Proposition 1** *In the  $\beta$ -discounted case we have:  $A_x(\beta) - A(\beta)$  is a constant matrix whose every entry is  $(1 - \beta)x(v_0)$ .*

In particular, their optimal (pure positional) strategies coincide, while the values are related by formula  $\mu_x(v_0) = \mu(v_0) + (1 - \beta)x(v_0)$  for every initial position  $v_0 \in V$ . Note that, in the undiscounted case,  $\beta = 1$ , the values also coincide:  $\mu_x(v_0) = \mu(v_0)$ .

## 3 Canonical forms

We will use the following notation through the paper: Given a function  $f : V \times V \rightarrow \mathbb{R}$ , we write  $\bar{M}[f(v, u)]$  to *symbolically* mean

$$\bar{M}[f(v, u)] \equiv \begin{cases} \max_{u|(v,u) \in E} f(v, u), & \text{for } v \in V_W, \\ \min_{u|(v,u) \in E} f(v, u), & \text{for } v \in V_B, \\ \sum_{u|(v,u) \in E} f(v, u) p(v, u), & \text{for } v \in V_R. \end{cases}$$

### 3.1 Ergodic canonical form

Given a (not initialized) BWR-game  $\mathcal{G} = (G, P, (v_0, )r)$ , let us define a mapping  $m : V \rightarrow \mathbb{R}$  as follows:

$$m(v) = \bar{M}[r(v, u)]. \quad (5)$$

A move  $(v, u) \in E$  in a position  $v \in V_W$  (respectively,  $v \in V_B$ ) is called *locally optimal* if it realizes the maximum (respectively, minimum) in (5). A strategy  $s_W$  of White (respectively,  $s_B$  of Black) is called *locally optimal* if it chooses a locally optimal move  $(v, u) \in E$  in every position  $v \in V_W$  (respectively,  $v \in V_B$ ).

**Definition 1** *We say that a (not initialized) BWR-game  $\mathcal{G}$  is in ergodic canonical form if function (5) is constant:  $m(v) \equiv m$  for some number  $m$ .*



**Proposition 2** *If a (not initialized) game is in ergodic canonical form then, for the  $\beta$ -discounted and undiscounted cases (i) every locally optimal strategy is optimal and (ii) the game is ergodic:  $m$  is its value for every initial position  $v_0 \in V$ .*

In Section 4 we will prove our main result for the discounted case.

**Theorem 2** *Given a (not initialized) BWR-game  $\mathcal{G}$  and discount factor  $\beta < 1$ , there is a discounted potential transformation (4) reducing  $\mathcal{G}$  to an ergodic canonical form. The corresponding vector of potentials can be determined by a polynomial algorithm for every fixed  $\beta$ .*

We remark that another *non potential-based* algorithm, pseudo-polynomial also in  $\beta$ , for solving  $\beta$ -discounted games, was obtained by Littman in his PhD thesis [Lit96].

### 3.2 Canonical form for undiscounted BWR-games

Given a (not initialized) undiscounted BWR-game  $\mathcal{G}$ , let us compute function  $m : V \rightarrow \mathbb{R}$  defined by formula (5) and then introduce another function  $M : V \rightarrow \mathbb{R}$  by the following similar formula:

$$M(v) = \bar{M}[m(u)]. \quad (6)$$

**Definition 2** *We say that a (not initialized) undiscounted BWR-game  $\mathcal{G}$  is in canonical form if (i)  $m(v) = M(v)$  for all  $v \in V$  and, moreover, (ii) for every  $v \in V_W \cup V_B$  there is a move  $(v, u) \in E$  such that  $m(v) = m(u) = r(v, u)$ , or in other words, move  $(v, u)$  is locally optimal and it respects the value of function  $m$ .*

**Proposition 3** *If a (not initialized) game is in canonical form then, for both the  $\beta$ -discounted and undiscounted cases (i) every locally optimal strategy is optimal and there are no others; (ii)  $m(v) = M(v)$  is the value of the game whenever  $v$  is its initial position.*

Property (ii) shows that the undiscounted BWR-game in canonical form can be not ergodic. In Section 5 we will prove our main result for the undiscounted case.

**Theorem 3** *For each (not initialized) undiscounted BWR-game  $\mathcal{G}$ , there is an undiscounted potential transform (4) reducing  $\mathcal{G}$  to canonical form.*

## 4 Polynomial pumping algorithm for $\beta$ -discounted BWR-games; proof of Theorem 2

Given a  $\beta$ -discounted BWR-game  $\mathcal{G} = (G, P, r)$ , let  $[r] = [r^-; r^+]$  denote the range of the local reward function  $r$ , that is,  $r^+ = \max(r(v, u) \mid (v, u) \in E)$  and  $r^- = \min(r(v, u) \mid (v, u) \in E)$ .

Similarly, let  $[m] = [m^-; m^+]$  denote the range of the function  $m$ , defined by formula (5). Furthermore, given a potential  $x : V \rightarrow \mathbb{R}$ , we consider the transformed game  $\mathcal{G}_x = (G, P, r_x)$  whose local reward function  $r_x$  is defined by (4). We will find a potential  $x$  such that function  $m_x : V \rightarrow \mathbb{R}$  is constant, that is,  $m_x^- = m_x^+$ , where  $m_x$  is defined by formula (5) in which  $r_x$  substitutes for  $r$ .

The following simple procedure reduces  $||[m_x]|| \stackrel{\text{def}}{=} m_x^+ - m_x^-$  to within an arbitrary accuracy  $\varepsilon$ .

---

### Algorithm 1 Pumping algorithm

---

**Require:** a  $\beta$ -discounted BWR-game  $\mathcal{G} = (G = (V, E), P, r)$ , an accuracy  $\varepsilon$ , and two parameters  $a, b \in [0, 1]$ .

**Ensure:** a potential  $x : V \rightarrow \mathbb{R}$  s.t.  $|m_x(v) - m_x(u)| \leq \varepsilon$  for all  $u, v \in V$

let  $x(v) = 0$  for all  $v \in V$

**while**  $||[m_x]|| \geq \varepsilon$  **do**

**for** all  $v$  s.t.  $x(v) \geq m_x^- + a||[m_x]||$  **do**

$x(v) := x(v) - b||[m_x]||$

**end for**

**end while**

**return**  $x$

---

**Lemma 1** *When run with  $a = b = \frac{1}{2}$ , the above procedure terminates in  $N = \frac{\log ||[r]|| - \log \varepsilon}{1 - \log(1 + \beta)}$  iterations.*

In Appendix C.3, we estimate the necessary accuracy at which we can guarantee that function  $m$  is constant. Namely, let us assume that (i)  $\beta = 1 - B'/B \in [0, 1)$  is a rational number; (ii) all local rewards, are integral in the range  $[-R, R]$ ; (iii) probabilities  $p(v, u)$ , for all arcs  $(v, u) \in E$  such that  $v \in V_R$ , are rational numbers with the least common denominator  $q$ . then it is enough to take  $\varepsilon = 1/n^2$  for the BW-case, and  $\varepsilon = (1/(qBB'))^{O(n^4)} \cdot (1/R)^{O(nh)}$  for the BWR-case, where  $h = |E|$ . Combining this with the bound in Lemma 1, we arrive at Theorem 2.

Let us remark, however, that the constant  $1 - \log(1 + \beta)$  in the running time tends to 0, as  $\beta \rightarrow 1^-$ . More precisely, if  $y = 1 - \beta \rightarrow 0^+$  then  $1 - \log(1 + \beta) = 1 - \log(2 - y) \sim y/(2 \ln 2)$ , and thus we obtain for the number of iterations  $N \sim 2 \ln 2 \frac{(\log R - \log \varepsilon)}{(1 - \beta)}$ . In Appendix C.4, we

recall an example of Condon [Con92] which shows that  $\beta > 1 - 2^{-n}$  might be needed for a “sufficiently good” approximation.

On the other hand, Miltersen [GM08] showed that it is enough to take  $\beta = 1 - ((n!)^2 2^{2n+3} M^{2n^2})^{-1}$ , so that the undiscounted values will be equal to the discounted ones. Thus, for the undiscounted BWR-games the limit transition  $\beta \rightarrow 1^-$  provides a finite but exponential in the worst case algorithm. Note, however, that this is not enough to prove Theorem 3, since the canonical form will contain a factor  $\beta < 1$  in it. In the next section, we overcome this problem by taking  $\beta$  to the limit.

## 5 Canonical form for the undiscounted BWR-games; proof of Theorem 3

In deriving Theorem 3 from Theorem 2, we face one difficulty: some of the potentials tend to  $\infty$  as  $\beta \rightarrow 1^-$ . We overcome this by modifying the potentials somehow, and then using a convergence result of Blackwell [Bla62]. However, this results in what we call *weak canonical form*, which satisfies all the conditions of a canonical form, except that the transformed rewards  $r_x(v, u)$  could be arbitrary, for some moves  $(v, u)$  such that  $v \in V_W \cup V_B$  and  $m(u) \neq m(v)$  (i.e condition (i) of Definition 2 may not hold). However, it is clear that in this case the local reward  $r_x(v, u)$  does not matter, since anyway, such a move cannot be locally optimal in an undiscounted game (in contrast with a discounted one). In fact, we can show easily that the existence of weak canonical form already implies Theorem 3.

**Proposition 4** *An undiscounted BWR-game in a weak canonical form is reduced to a canonical form by the potential transform  $x = \{x_v = Cm(v) \mid v \in V\}$  whenever  $C$  is negative and  $|C|$  is sufficiently large.*

Thus, in the following we show the existence of a weak canonical form.

From the results in Section 4, we know that, for any  $0 \leq \beta < 1$ , there exist  $m^\beta \in \mathbb{R}$  and  $x = x^\beta \in \mathbb{R}^V$  such that the function  $m_x : V \rightarrow \mathbb{R}$ , given by formula (5) with  $r$  substituted by  $r_x$  is constant:

$$m^\beta = m_x(v) = \bar{M}[r(v, u) + x^\beta(v) - \beta x^\beta(u)] \quad \text{for all } v \in V. \quad (7)$$

Furthermore, from Proposition 1, we know that the value of the game when started at vertex  $v \in V$  is

$$\mu^\beta(v) = m^\beta - (1 - \beta)x^\beta(v) = \bar{M}[r(v, u) + \beta(x^\beta(v) - x^\beta(u))]. \quad (8)$$

We first observe that (as in Theorem 5.1 in [ZP96] for the BW-case) the values  $\mu^\beta(v)$  satisfy the following equations:

$$\mu^\beta(v) = \bar{M}[(1 - \beta)r(v, u) + \beta\mu^\beta(u)], \quad (9)$$

for all  $v \in V$ . Indeed, using (8),

$$\begin{aligned}
\bar{M}[(1 - \beta)r(v, u) + \beta\mu^\beta(u)] &= \bar{M}[(1 - \beta)r(v, u) + \beta(m^\beta - (1 - \beta)x^\beta(u))] \\
&= (1 - \beta)\bar{M}[r(v, u) - \beta x^\beta(u)] + \beta m^\beta \\
&= (1 - \beta)\bar{M}[r(v, u) + \beta(x^\beta(v) - x^\beta(u))] + \beta(m^\beta - (1 - \beta)x^\beta(v)) \\
&= (1 - \beta)\mu^\beta(v) + \beta\mu^\beta(v) = \mu^\beta(v).
\end{aligned}$$

By Theorem 2, for each  $\beta \in [0, 1)$ , there exists an optimal situation  $s(\beta)$  in the  $\beta$ -discounted BWR-game, and potential  $x^\beta$  satisfying (7) and (8). Let us consider all such situations as  $\beta \rightarrow 1^-$ . Among this infinite sequence of situations, one situation  $s$  appears infinitely many times, since the total number of possible strategies is finite. Let us fix such a situation  $s$  and consider the corresponding infinite subsequence  $\{\beta_i\}_{i=0}^\infty$  for which  $s$  is optimal in the corresponding game. Then  $\lim_{i \rightarrow \infty} \beta_i = 1$  and (7), (8) and (9) hold for every  $\beta \in \{\beta_i\}_{i=0}^\infty$ . For  $v \in V$ , let  $\mu(v) = \lim_{i \rightarrow \infty} \mu^{\beta_i}(v)$  and note that this limit exists by (28). Furthermore, since (9) is satisfied for all  $\beta_i$ , it is also satisfied in the limit as  $i \rightarrow \infty$ , i.e.,  $\mu(v) = \bar{M}[\mu(u)]$ . In particular<sup>4</sup>

$$\mu(v) \geq \mu(u), \text{ for all } (v, u) \in E, v \in V_W \quad (10)$$

$$\mu(v) \leq \mu(u), \text{ for all } (v, u) \in E, v \in V_B \quad (11)$$

$$\mu(v) = \sum_{u|(v,u) \in E} \mu(u) p(v, u), v \in V_R, \quad (12)$$

and the extremum values in (10) and (11) are attained, i.e., for every  $v \in V_W \cup V_B$ , there is a  $u$  such that  $(v, u) \in E$  and  $\mu(u) = \mu(v)$ ; we call such subset of edges extremal edges and denote it by *Ext*.

Note that, in the *non-ergodic* case, as  $\beta \rightarrow 1^-$ , (8) implies that  $|x^\beta(v)| \rightarrow \infty$ , for some vertices  $v \in V$ ; otherwise all the values  $\mu(v)$  are equal to  $\lim_{\beta \rightarrow 1} m^\beta$ , independent of the starting position. We will modify the potentials, in this case, to guarantee that they become finite, without affecting the value of the game.

Cosnider any  $\beta \in \{\beta_i\}_{i=0}^\infty$ . From (8), we can express the potential at  $v \in V$  as follows

$$x^\beta(v) = \frac{m^\beta - \mu^\beta(v)}{1 - \beta}. \quad (13)$$

Define, for  $v \in V$ , the new potential:

$$y^\beta(v) = x^\beta(v) - \frac{m^\beta - \mu(v)}{1 - \beta} = \frac{\mu(v) - \mu^\beta}{1 - \beta}. \quad (14)$$

---

<sup>4</sup>This statement also follows from the results of Gillette [Gil57], since  $\mu(v)$  is the value of the game starting at  $v$ .

In particular, substituting  $y^\beta(v) - y^\beta(u) = x^\beta(v) - x^\beta(u) + \frac{\mu(v) - \mu(u)}{1 - \beta}$ , we have

$$\mu^\beta(v) = \bar{M}[r(v, u) + \beta(y^\beta(v) - y^\beta(u))], \quad (15)$$

for any  $v \in V_W \cup V_B$  and  $(v, u) \in Ext$ . Furthermore, for any  $v \in V_R$ , we have by (12),

$$\begin{aligned} \mu^\beta(v) &= \sum_{u|(v,u) \in E} p(v, u)(r(v, u) + \beta(x^\beta(v) - x^\beta(u))) \\ &= \sum_{u|(v,u) \in E} p(v, u)(r(v, u) + \beta(y^\beta(v) - y^\beta(u))) - \beta \sum_{u|(v,u) \in E} p(v, u) \frac{\mu(v) - \mu(u)}{1 - \beta} \\ &= \sum_{u|(v,u) \in E} p(v, u)(r(v, u) + \beta(y^\beta(v) - y^\beta(u))). \end{aligned} \quad (16)$$

Let  $P_s$  be the transition matrix obtained by extending  $P$  by setting the entries corresponding to  $s$  to 1,  $Q_s$  and  $P_s^*$  be the corresponding  $Q$ -matrix and limiting transition matrix, respectively. Recall that  $\mu^\beta = (1 - \beta) \sum_{i=0}^{\infty} \beta^i P_s^i Q_s r$ ,  $\mu = \lim_{\beta \rightarrow 1} \mu^\beta = P_s^* Q_s r$ .

Rewriting  $P_s^* Q_s r = (1 - \beta) \sum_{i=0}^{\infty} \beta^i P_s^* Q_s r$ , for any  $\beta \in (0, 1)$ , we obtain

$$\begin{aligned} y &= \lim_{i \rightarrow \infty} y^{\beta^i} = \lim_{\beta \rightarrow 1} y^\beta = \lim_{\beta \rightarrow 1} (1 - \beta)^{-1} (\mu - \mu^\beta) \\ &= \lim_{\beta \rightarrow 1} \sum_{i=0}^{\infty} \beta^i (P_s^* - P_s^i) Q_s r = -[(I - (P_s - P_s^*))^{-1} - P_s^*] Q_s r, \quad (\text{by (29)}). \end{aligned}$$

So  $y$  exists in the limit and it transforms the game to weak canonical form as seen from (15) and (16).

## 6 Gillette games

### 6.1 BWR- and Gillette games are equivalent

Recall that in a Gillette game with perfect information (GGPI)  $\mathcal{G} = (V_B, V_W, P, r)$ , we are given a finite set  $V = V_B \cup V_W$ :  $V_W$  controlled by White and  $V_B$  controlled by Black. At each state  $v$ , the controlling player has to decide among a given set of possible actions  $S(v)$ : if the player chooses action  $k \in S(v)$ , then Black pays White<sup>5</sup>  $r^k(v, u)$  and the game moves to state  $u$  with probability  $p^k(v, u)$ .

**Theorem 4** *BWR-games and Gillette games are polynomial-time equivalent.*

<sup>5</sup>It is common to assume w.l.o.g. that  $r^k(v, u) = r^k(v)$ , i.e., the local reward depends only on the starting state. It is easy to see that the more general case can be reduced to this by setting  $r^k(v) = \sum_{u \in V} p^k(v, u) r^k(v, u)$ .

## 6.2 Existence of canonical form for Gillette game: The undiscounted case

Given a potential transformation  $x : V \rightarrow \mathbb{R}$ , let us define a mapping  $m_x : V \rightarrow \mathbb{R}$  as follows

$$m_x(v) = \begin{cases} \max & (m_x^k(v) \mid k \in S(v)), & \text{for } v \in V_W, \\ \min & (m_x^k(v) \mid k \in S(v)), & \text{for } v \in V_B, \end{cases} \quad (17)$$

where  $m_x^k(v) = \sum_{(v,u) \in E} p^k(v,u)(r^k(v,u) + x(v) - x(u))$ . Similarly, we define the mapping  $M_x : V \rightarrow \mathbb{R}$

$$M_x(v) = \begin{cases} \max & (M_x^k(v) \mid k \in S(v)), & \text{for } v \in V_W, \\ \min & (M_x^k(v) \mid k \in S(v)), & \text{for } v \in V_B, \end{cases} \quad (18)$$

where  $M_x^k(v) = \sum_{(v,u) \in E} p^k(v,u)m_x^k(v)$ .

**Theorem 5 (Canonical form for GGPI)** *For any GGPI, there exists a potential transformation  $x : V \rightarrow \mathbb{R}$ , such that for all  $v \in V$ : (i)  $m_x(v) = M_x(v)$ ; and (ii) there exists a  $k \in S(v)$  such that  $m_x^k(v) = M_x(v) = m_x(v)$ .*

## 6.3 Existence of canonical form for Gillette game: The discounted case

Given  $\beta < 1$  and a potential transformation  $x : V \rightarrow \mathbb{R}$ , let us define a mapping  $m_x^\beta : V \rightarrow \mathbb{R}$  as follows

$$m_x^\beta(v) = \begin{cases} \max & (\sum_{(v,u) \in E} p^k(v,u)(r^k(v,u) + x(v) - \beta x(u)) \mid k \in S(v)), & \text{for } v \in V_W, \\ \min & (\sum_{(v,u) \in E} p^k(v,u)(r^k(v,u) + x(v) - \beta x(u)) \mid k \in S(v)), & \text{for } v \in V_B. \end{cases} \quad (19)$$

**Theorem 6 (Canonical form for discount GGPI)** *For any discount GGPI with discount factor  $\beta < 1$ , there exists a potential transformation  $x : V \rightarrow \mathbb{R}$  and a number  $m$  such that  $m_x^\beta(v) = m$ , for all  $v \in V$ , and such that value of the game at vertex  $v$  is given by  $\mu(v) = m - (1 - \beta)x(v)$ . Such a transformation can be found in polynomial time for every fixed  $\beta$ .*

## 6.4 BWR- and Gillette games are solvable in subexponential time

Zwick and Paterson [ZP96] observed that undiscounted BW-games are polynomial-time reducible to the discounted ones. In fact, it is enough to choose any  $\beta > 1 - 1/(4n^3R)$ , when rewards are integral with maximum absolute value  $R$ ; see [ZP96], Theorem 5.2. Furthermore, they showed that the discounted BW-games are polynomial-time reducible to Simple

Stochastic Games (SSG) (or Condon’s games); see [ZP96] Theorem 6.1. Miltersen [GM08] has recently modified their reduction to show that any discounted Gillette game can be represented as an SSG, where the probabilities are bilinear in  $\beta$ , the original transition probabilities, and original rewards. Furthermore, he also showed that any undiscounted Gillette game is reduced to a discounted one with  $\beta = 1 - ((n!)^2 2^{2n+3} M^{2n^2})^{-1}$ , when the rewards and transition probabilities are assumed to be rational with integral numerators and denominators of maximum absolute value  $M$ .

Halman [Hal07] showed that any SSG with  $m = |V_B| + |V_W|$  deterministic nodes can be solved in randomized *strongly* subexponential-time  $2^{O(\sqrt{n \log n})}$ . We observe further that the reduction in [GM08] can increase only the number of random nodes. Thus we obtain the following result.

**Theorem 7** *Any BWR-game or Gillette game on  $n$  vertices is solvable in strongly  $2^{O(\sqrt{n \log n})}$  expected time.*

## 7 Conclusions

It was shown that any BWR-game or Gillette game can be brought by a potential transformation into canonical form, in which optimal strategies become obvious. While our algorithm for finding such transformations is efficient for the discounted case, it is open whether such an algorithm exists for the undiscounted case. However, we believe that our framework will inspire new algorithms for solving this wide class of games. In particular, since BWR-games generalize SSG’s and for the latter no pseudo-polynomial algorithm is known yet, the existence of canonical form might be one direction to explore towards finding such an algorithm.

## Acknowledgment.

The second author is thankful to Peter Bro Miltersen for helpful discussions.

## References

- [Bla62] D. Blackwell. Discrete dynamic programming. *Ann. Math. Statist.*, 33:719–726, 1962.
- [BV01a] E. Beffara and S. Vorobyov. Adapting gurvich-karzanov-khachiyan’s algorithm for parity games: Implementation and experimentation. Technical Report 2001-020, Department of Information Technology, Uppsala University, available at: <https://www.it.uu.se/research/reports/#2001>, 2001.

- [BV01b] E. Beffara and S. Vorobyov. Is randomized gurvich-karzanov-khachiyan's algorithm for parity games polynomial? Technical Report 2001-025, Department of Information Technology, Uppsala University, available at: <https://www.it.uu.se/research/reports/#2001>, 2001.
- [BV07] H. Björklund and S. Vorobyov. A combinatorial strongly sub-exponential strategy improvement algorithm for mean payoff games. *Discrete Applied Mathematics*, 155(2):210–229, 2007.
- [CH08] K. Chatterjee and T. A. Henzinger. Reduction of stochastic parity to stochastic mean-payoff games. *Inf. Process. Lett.*, 106(1):1–7, 2008.
- [CJH04] K. Chatterjee, M. Jurdziński, and T. A. Henzinger. Quantitative stochastic parity games. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 121–130, Philadelphia, PA, USA, 2004. Society for Industrial and Applied Mathematics.
- [Con92] A. Condon. The complexity of stochastic games. *Information and Computation*, 96:203–224, 1992.
- [Con93] A. Condon. An algorithm for simple stochastic games. In *Advances in computational complexity theory, volume 13 of DIMACS series in discrete mathematics and theoretical computer science*, 1993.
- [DG06] V. Dhingra and S. Gaubert. How to solve large scale deterministic games with mean payoff by policy iteration. In *valuetools '06: Proceedings of the 1st international conference on Performance evaluation methodologies and tools*, page 12, New York, NY, USA, 2006. ACM.
- [EM79] A. Eherenfeucht and J. Mycielski. Positional strategies for mean payoff games. *International Journal of Game Theory*, 8:109–113, 1979.
- [Fil81] J. A. Filar. Ordered field property for stochastic games when the player who controls transitions changes from state to state. *J. of Optimization Theory and Applications*, 34(4):503–515, 1981.
- [Fly74] J. Flynn. Averaging vs. discounting in dynamic programming: a counterexample. *Annals of Statistics*, 2:411–413, 1974.
- [GH08] H. Gimbert and F. Horn. Simple stochastic games with few random vertices are easy to solve. In *FoSSaCS*, pages 5–19, 2008.



- [Gil57] D. Gillette. Stochastic games with zero-sum stop probabilities. In A.W. Tucker M. Dresher and P. Wolfe, editors, *Contribution to the Theory of Games III, in Annals of Mathematics Studies*, volume 39, pages 179–187. Princeton University Press, 1957.
- [GKK88] V. Gurvich, A. Karzanov, and L. Khachiyan. Cyclic games and an algorithm to find minimax cycle means in directed graphs. *USSR Computational Mathematics and Mathematical Physics*, 28:85–91, 1988.
- [GLS88] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer, New York, 1988.
- [GM08] V. Gurvich and P. B. Miltersen. On the computational complexity of solving stochastic mean-payoff games. *CoRR*, abs/0812.0486, 2008.
- [Hal07] N. Halman. Simple stochastic games, parity games, mean payoff games and discounted payoff games are all lp-type problems. *Algorithmica*, 49(1):37–50, 2007.
- [HBV04] S. Sandberg H. Björklund and S. Vorobyov. A combinatorial strongly subexponential strategy improvement algorithm for mean payoff games. DIMACS Technical Report 2004-05, DIMACS, Rutgers University, 2004.
- [HL31] G. H. Hardy and J. E. Littlewood. Notes on the theory of series (xvi): two tauberian theorems. *J. of London Mathematical Society*, 6:281–286, 1931.
- [How60] R. A. Howard. *Dynamic programming and Markov processes*. Technology press and Willey, New York, 1960.
- [JPZ06] M. Jurdziński, M. Paterson, and U. Zwick. A deterministic subexponential algorithm for solving parity games. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 117–123, New York, NY, USA, 2006. ACM.
- [Jur98] M. Jurdziński. Deciding the winner in parity games is in  $\text{up} \cap \text{co-up}$ . *Inf. Process. Lett.*, 68(3):119–124, 1998.
- [Kar78] R. M. Karp. A characterization of the minimum cycle mean in a digraph. *Discrete Math.*, 23:309–311, 1978.
- [KL93] A. V. Karzanov and V. N. Lebedev. Cyclical games with prohibition. *Mathematical Programming*, 60:277–293, 1993.

- [KMMS03] D. Kratsch, R. M. McConnell, K. Mehlhorn, and J. P. Spinrad. Certifying algorithms for recognizing interval graphs and permutation graphs. In *SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 158–167, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.
- [KS63] J. G. Kemeny and J. L. Snell. *Finite Markov chains*. Springer-Verlag, New York, 1963.
- [Lit96] M. L. Littman. *Algorithm for sequential decision making, CS-96-09*. PhD thesis, Dept. of Computer Science, Brown Univ., USA, 1996.
- [LL69] T. M. Liggett and S. A. Lippman. Stochastic games with perfect information and time-average payoff. *SIAM Review*, 4:604–607, 1969.
- [MO70] H. Mine and S. Osaki. *Markovian decision process*. American Elsevier Publishing Co., New York, 1970.
- [Mou76a] H. Moulin. Extension of two person zero sum games. *Journal of Mathematical Analysis and Application*, 5(2):490–507, 1976.
- [Mou76b] H. Moulin. Prolongement des jeux à deux joueurs de somme nulle. *Bull. Soc. Math. France, Memoire*, 45, 1976.
- [Pis99] N. N. Pisaruk. Mean cost cyclical games. *Mathematics of Operations Research*, 24(4):817–828, 1999.
- [SF92] R. Sznardner and J. A. Filar. Some comments on a theorem of hardy and littlewood. *J. of Optimization Theory and Applications*, 75(1):201–208, 1992.
- [Sha53] L. S. Shapley. Stochastic games. *Proc. Nat. Acad. Science, USA*, 39:1095–1100, 1953.
- [Ste75] M. Stern. *n stochastic games with limiting average payoff*. PhD thesis, Univ. of Illinois at Chicago, Chicago, USA, 1975.
- [ZP96] U. Zwick and M. Paterson. The complexity of mean payoff games on graphs. *Theoretical Computer Science*, 158(1-2):343–359, 1996.

# Appendix A Abel- and Cesaro-average; the Hardy and Littlewood Theorem

Given a sequence of real numbers  $\{b_i\}_{i=0}^{\infty}$  and *discount factor*  $\beta \in [0, 1)$ , let

$$a_k(\beta) = (1 - \beta) \sum_{i=0}^k \beta^i b_i \quad \text{and} \quad c_k = \frac{1}{k+1} \sum_{i=0}^k b_i \quad (20)$$

It is said that the sequence is *Cesaro-summable* to  $c$  or *Abel-summable* to  $a$  if, respectively,

$$\lim_{k \rightarrow \infty} c_k = \lim_{k \rightarrow \infty} \frac{1}{k+1} \sum_{i=0}^k b_i = c \quad \text{or} \quad \lim_{\beta \rightarrow 1^-} a(\beta) = \lim_{\beta \rightarrow 1^-} (1 - \beta) \sum_{i=0}^{\infty} \beta^i b_i = a \quad (21)$$

It is known that for every sequence the following inequalities hold

$$\liminf_{k \rightarrow \infty} c_k \leq \liminf_{\beta \rightarrow 1^-} a_k(\beta) \leq \limsup_{\beta \rightarrow 1^-} a_k(\beta) \leq \limsup_{k \rightarrow \infty} c_k \quad (22)$$

showing that Cesaro-summability implies Abel-summability to the same limit.

By the famous Hardy-Littlewood Theorem [HL31], for the *bounded* sequences the inverse claim holds too, that is, Abel-summability implies Cesaro-summability to the same limit. However, even  $(0, 1)$ -sequences may be not summable; as the following well-known example shows:

$$0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, \dots,$$

This sequence begins with 0; after which blocks of ones and zeros alternate and the  $i$ -th block consists of  $2^i$  elements,  $i = 0, 1, 2, \dots$ . It is easy to verify that

$$\liminf_{k \rightarrow \infty} c_k = \lim_{i \rightarrow \infty} c_{2^{2i}} = 1/3 < \limsup_{k \rightarrow \infty} c_k = \lim_{i \rightarrow \infty} c_{2^{2i+1}} = 2/3.$$

Hence, by the Hardy-Littlewood Theorem,

$$\liminf_{\beta \rightarrow 1^-} a_k(\beta) < \limsup_{\beta \rightarrow 1^-} a_k(\beta).$$

We refer the reader to the nice mini-survey [SF92] for more detail and the bibliography; see also [LL69, SF92].

The above concepts of Abel and Cesaro summation are widely applied in studying weighted Markov chains, Markov decision process, and stochastic games; see [Sha53, Gil57, Bla62, LL69] and very many later works. If a path (play) with local payoffs  $b_0, b_1, b_2, \dots$  appears then the limit effective payoff is defined by formulae (20) and (21), where  $a$  is the discounted payoff with the discount factor  $\beta$ , while  $c$  is the undiscounted (so-called limit mean) payoff.

## Appendix B: Basic facts about weighted Markov chains

### B.1 Effective limiting payoffs in weighted Markov chains

Let  $\mathcal{G} = (G, P, p_0, r)$  be a weighted Markov chain, where  $G = (V, E)$  is a digraph with  $n$  vertices and  $h$  arcs,  $|V| = n, |E| = h$ ; furthermore  $V = V_R$ , while  $V_B = V_W = \emptyset$ , and  $r$  is the  $h$ -vector of local rewards.

In this case  $p_0$  is a probabilistic  $n$ -vector  $p_0 : V \rightarrow [0, 1]$  and  $P : V \times V \rightarrow [0, 1]$  is the probabilistic  $n \times n$  matrix (so-called *transition matrix*) whose entry  $p(v, u)$  is the probability of transition from  $v$  to  $u$  in one move, for every pair of positions  $v, u \in V$ . Then, it is obvious and well-known that for every integral  $i \geq 0$  matrix  $P^i : V \times V \rightarrow [0, 1]$  (the  $i$ -th power of  $P$ ) is the  $i$ -move transition matrix, whose entry  $p_i(v, u)$  is the probability of transition from  $v$  to  $u$  in exactly  $i$  moves, for every  $v, u \in V$ .

Let  $q_i(v, u)$  be the probability that that arc  $(v, u) \in E$  will be the  $(i + 1)$ -st move, given the original distribution  $p_0$ , where  $i = 0, 1, 2, \dots$  and denote by  $q_i : E \rightarrow [0, 1]$  be the corresponding probabilistic  $h$ -vector. To compute  $q_i$ , we introduce  $n \times h$  vertex-arc transition matrix  $Q : V \times E \rightarrow [0, 1]$  whose entry  $q(\ell, (v, u))$  is equal to  $p(v, u)$  if  $\ell = v$  and 0 otherwise, for every  $\ell \in V$  and  $(v, u) \in E$ . Then, it is clear that  $q_i = p_0 P^i Q$ .

Let  $b_i$  denote the expected reward for the  $(i + 1)$ -st move;  $i = 0, 1, 2, \dots$

$$b_i = \sum_{(v,u) \in E} q_i(v, u) r(v, u) = p_0 P^i Q r. \quad (23)$$

The *effective payoff* of the *undiscounted* weighted Markov chain is defined to be the average expected reward on the limit, i.e.,

$$c = \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{i=0}^n b_i, \quad (24)$$

the so-called *Cesaro-sum*. The existence of such a limit follows from the well-known *Hardy-Littlewood Theorem* (see Appendix A for more details). Indeed, given a discount factor  $\beta \in [0, 1)$ , the  $\beta$ -discounted payoff is given by

$$(1 - \beta) \sum_{i=0}^{\infty} \beta^i b_i, \quad (25)$$

the so-called *Abel-sum*. It is not difficult to verify that sequence  $\{b_i\}_{i=0}^{\infty}$  is Abel-summable for every discount factor  $\beta \in [0, 1)$ . Indeed, the  $n \times n$ -matrix  $(I - \beta P)$  is non-singular, since it has dominating main diagonal and

$$a_\beta = (I - \beta P)^{-1} = \sum_{i=0}^{\infty} (\beta P)^i. \quad (26)$$

Moreover, there exists the limit  $n \times n$ -matrix

$$\lim_{\beta \rightarrow 1^-} (1 - \beta)(I - \beta P)^{-1} = P^* \quad (27)$$

see, for example, [How60, Bla62, MO70]. Thus, the following chain of equalities holds:

$$\begin{aligned} \lim_{\beta \rightarrow 1^-} a_\beta &= \lim_{\beta \rightarrow 1^-} (1 - \beta) \sum_{i=0}^{\infty} \beta^i p_0 P^i Qr = \\ p_0 \left[ \lim_{\beta \rightarrow 1^-} (1 - \beta) \sum_{i=0}^{\infty} (\beta P)^i \right] Qr &= p_0 \left[ \lim_{\beta \rightarrow 1^-} (1 - \beta)(I - \beta P)^{-1} \right] Qr = p_0 P^* Qr. \end{aligned} \quad (28)$$

In other words, the sequence  $\{b_i\}_{i=0}^{\infty}$  is Abel-summable to  $a = p_0 A Qr$  and, obviously, this sequence is bounded. Hence, it is also Cesaro-summable to the same limit, by the Hardy-Littlewood Theorem.

## B.2 Related results from theory of Markov chains

The above approach plays an important role in stochastic games and Markov decision processes. First, the result was obtained in the pioneering work of Gillette [Gil57] in 1957. Yet, his proof contained a flaw (an overstatement of the Hardy-Littlewood Theorem) that was pointed out and corrected in 1969 by Liggett and Lippman [LL69]. These matters were further clarified in [Fly74, Ste75, Fil81, SF92].

An independent proof, obtained by Blackwell [Bla62] in 1962, is based on the following classical results on Markov chains:

Given a  $n \times n$  transition matrix  $P$ , the Cesaro partial sums  $\frac{1}{k+1} \sum_{i=1}^k P^i$  converge, as  $k \rightarrow \infty$ , to the limit Markov matrix  $P^*$  such that:

- (i)  $PP^* = P^*P = P^*P^* = P^*$ ; (ii)  $\text{rank}(I - P) + \text{rank}P^* = n$ ;
- (iii) For each  $n$ -vector  $c$  system  $Px = x$ ,  $P^*x = c$  has a unique solution.
- (iv) matrix  $I - (P - P^*)$  is non-singular and

$$H(\beta) = \sum_{i=0}^{\infty} \beta^i (P^i - P^*) \rightarrow H = (I - (P - P^*))^{-1} - P^* \text{ as } \beta \rightarrow 1^-; \quad (29)$$

(v)

$$H(\beta)P^* = P^*H(\beta) = HP^* = P^*H = 0 \text{ and } (I - P)H = H(I - P) = I - P^*.$$

Claim (iv) (which will play very important role in our paper) is proved in 1962 by Blackwell, [Bla62], while for the remaining four claims, he cites the text-book in finite Markov chains by Kemeny and Snell [KS63] (that was published, in fact, one year later, in 1963).

### B.3 Limiting distribution

Let  $(G = (V, E), P)$  be a Markov chain, and let  $C_1, \dots, C_k \subseteq V$  be the vertex sets of the strongly connected components (classes) of  $G$ . For  $i \neq j$ , let us (standardly) write  $C_i \prec C_j$ , if and only if there is an arc  $(v, u) \in E$  such that  $v \in C_i$  and  $u \in C_j$ . The components  $C_i$ , such that there is no  $C_j$  with  $C_i \prec C_j$  are called the *absorbing* (or *recurrent*) classes, while the other components are called *transient* or *non-recurrent*. Let  $J = \{i : C_i \text{ is absorbing}\}$ ,  $A = \cup_{i \in J} C_i$ , and  $T = V \setminus A$ . For  $X, Y \subseteq V$ , a matrix  $H \subseteq \mathbb{R}^{V \times V}$ , a vector  $h \subseteq \mathbb{R}^V$ , we denote by  $H[X; Y]$  the submatrix of  $P$  induced by  $X$  as rows and  $Y$  as columns, and by  $h[X]$  the subvector of  $h$  induced by  $X$ . Let  $I = I[V; V]$  be the  $|V| \times |V|$  identity matrix,  $e = e[V]$  be the vector of all ones of dimension  $|V|$ . For simplicity, we drop the indices of  $I[\cdot, \cdot]$  and  $e[\cdot]$ , when they are understood from the context. Then  $P[C_i; C_j] = 0$  if  $C_j \prec C_i$ , and hence in particular,  $P[C_i; C_i]e = e$  for all  $i \in J$ , while  $P[T, T]e$  has at least one component of value strictly less than 1.

The following are well-known facts about  $P^i$  and the limiting distribution  $p_w = e_w P^*$ , when the initial distribution is the  $w$ th unit vector  $e_w$  of dimension  $|V|$  (see, e.g., [KS63]):

(L1)  $p_w[A] > 0$  and  $p_w[T] = 0$ ;

(L2)  $\lim_{i \rightarrow \infty} P^i[V; T] = 0$ ;

(L3)  $\text{rank}(I - P[C_i; C_i]) = |C_i| - 1$  for all  $i \in J$ ,  $\text{rank}(I - P[T; T]) = |T|$ , and  $(I - P[T; T])^{-1} = \sum_{i=0}^{\infty} P[T; T]^i$ ;

(L4) the absorption probabilities  $y_i \in [0, 1]^V$  into a class  $C_i$ ,  $i \in J$ , are given by the unique solution of the linear system:  $(I - P[T; T])y_i[T] = P[T; C_i]e$ ,  $y_i[C_i] = e$  and  $y_i[C_j] = 0$  for  $j \neq i$ ;

(L5) the limiting distribution  $p_w \in [0, 1]^V$  is given by the unique solution of the linear system:  $p_w[C_i](I - P[C_i; C_i]) = 0$ ,  $p_w[C_i]e = y_i(w)$ , for all  $i \in J$ , and  $p_w[T] = 0$ .

## Appendix C: Accuracy estimation

### C.1 Encoding length

Recall that [GLS88] the encoding length  $\langle x \rangle$  of a rational number  $x$  is given by  $\langle x \rangle \stackrel{\text{def}}{=} \langle y \rangle + \langle z \rangle$ , where  $y, z \in \mathbb{Z}$ ,  $x = y/z$  is the unique coprime representation of  $x$ , and for an integer  $k$ ,  $\langle k \rangle = 1 + \lceil \log_2(|k| + 1) \rceil$ . The encoding length of a matrix  $\langle A \rangle$  is the sum of the encoding lengths of its entries. We will make use of the following facts from [GLS88]:

(R1) For every rational number  $x \neq 0$ ,  $2^{-(x)+1} \leq |x| \leq 2^{(x)-1} - 1$ ;

(R2) For any rational numbers  $x_1, \dots, x_k$ ,  $\langle x_1 + \dots + x_k \rangle \leq 2(\langle x_1 \rangle + \dots + \langle x_k \rangle)$ ;

(R3) For a nonsingular matrix  $A \in \mathbb{Q}^{n \times n}$ ,  $\langle A^{-1} \rangle \leq 4n^2 \langle A \rangle$ ;

(R4) For two matrices  $A \in \mathbb{Q}^{m \times n}$  and  $B \in \mathbb{Q}^{n \times p}$ ,  $\langle AB \rangle \leq 2p \langle A \rangle + 2m \langle B \rangle$ .

**Proposition 5** *Let  $P$  be a rational  $n \times n$  transition matrix of a Markov chain and  $p_w$  be the limiting distribution of the chain starting from  $w$ . Then  $p_w$  is rational and*

$$\langle p_w \rangle = O(n^3) \langle P \rangle. \quad (30)$$

**Proof** We use the notation of the previous section. From (L4), we get that  $y_i[T] = (I - P[T; T])^{-1} P[T; C_i] e$  for any  $i \in J$ . Applying (R3), and repeatedly (R2) and (R4), we get that  $\langle y_i[T] \rangle \leq 10|T|^3 + 4|T|^2 \langle P[T; T] \rangle + |T| \langle P[T; C_i] \rangle$ . Now using (L4), we have  $p_w[C_i] = z_i A_i^{-1}$ , where  $z_i = (0, 0, \dots, 0, y_i(w))$  and  $A_i$  is the  $|C_i| \times |C_i|$  matrix whose first  $|C_i| - 1$  columns are any  $|C_i| - 1$  linearly independent columns of  $I - P[C_i; C_i]$  and the last column is  $e$ . Again, repeated use of (R2), (R3) and (R4) gives  $\langle p_w[C_i] \rangle \leq |C_i|^2 + |C_i| \langle y_i(w) \rangle + 4|C_i|^2 (3|C_i| + \langle P[C_i; C_i] \rangle)$ . Using the bound we obtained on  $\langle y_i[T] \rangle$ , an summing over all  $i \in J$ , we get the bound in (30).  $\square$

## C.2 An upper bound for the accuracy for undiscounted BWR-games

First let us consider the undiscounted BW-games assuming that the reward function  $r : E \rightarrow \mathbb{Z}$  takes integral values. Obviously, for each situation  $s = (s_W, s_B)$  the corresponding play consists of an initial part (a debut) that is followed by an infinitely repeated directed cycle  $C = C(s)$ . It is also clear that the corresponding effective limit Cesaro payoff  $c = c(s)$  does not depend on the debut and is equal to the mean local reward along  $C$ , that is,  $c = \sum_{e \in C} r(e) / |C|$ , where  $|C|$  is the number of edges (or vertices) in  $C$ . By definition, the denominator  $|C|$  cannot exceed  $n = |V|$ . Hence,  $|c' - c''| \geq n^{-2}$  for any two distinct limit payoffs  $c' = c(s')$  and  $c'' = c(s'')$ . Thus, accuracy  $n^{-2}$  would suffice, or in other words,  $\varepsilon \geq n^{-2}$ .

Now consider the undiscounted BWR-case. Again assume that all local rewards are integral in the range  $[-R, R]$ , and all probabilities  $p(v, u)$ , for  $(v, u) \in E$  and  $v \in V_R$ , are rational numbers with the least common denominator  $q$ .

Consider two situations  $s$  and  $s'$  such that  $c(s) \neq c(s')$ . By formula (28)

$$c(s) - c(s') = (p_s^* Q_s - p_{s'}^* Q_{s'}) r,$$

where  $p_s^* = p_0 P_s^*$  and  $p_{s'}^* = p_0 P_{s'}^*$  are the the limiting distributions and  $p_0$  is the unit vector corresponding to the starting vertex.

Applying (R2), (R3) and (R4), we get

$$\begin{aligned}\langle c(s) - c(s') \rangle &\leq 4(\langle p_s^* \rangle + \langle p_{s'}^* \rangle) + 8(\langle Q_s \rangle + \langle Q_{s'} \rangle) + 8n\langle r \rangle \\ &= O(n^3)(\langle P_s \rangle + \langle P_{s'} \rangle) + 8(\langle Q_s \rangle + \langle Q_{s'} \rangle) + 8n\langle r \rangle,\end{aligned}$$

where the last equation follows from (30). Using  $\langle P_s \rangle \leq 2n^2\langle q \rangle$  and  $\langle Q_s \rangle \leq 2nh\langle q \rangle$  for any situation  $s$ , and  $\langle r \rangle \leq h\langle R \rangle$ , where  $h = |E|$ , we conclude that

$$\langle c(s) - c(s') \rangle = O(n^5)\langle q \rangle + O(nh)\langle R \rangle.$$

Now using (R1) we get that

$$\begin{aligned}|c(s) - c(s')| &\geq \left(\frac{1}{2}\right)^{O(n^5)\langle q \rangle + O(nh)\langle R \rangle} \\ &\geq \left(\frac{1}{q}\right)^{O(n^5)} \cdot \left(\frac{1}{R}\right)^{O(nh)},\end{aligned}$$

which gives an upper bound on the required accuracy  $\varepsilon$  in this case.

### C.3 An upper bound for the accuracy for the discounted BWR-game

Consider the  $\beta$ -discounted BWR-games where we assume that: (i)  $\beta = 1 - B'/B \in [0, 1)$  is a rational number; (ii) all local rewards, as before, are integral in the range  $[-R, R]$ ; (iii) probabilities  $p(v, u)$ , for all arcs  $(v, u) \in E$  such that  $v \in V_R$ , are rational numbers with the least common denominator  $q$ . Fix an initial vertex  $v_0$  and denote by  $c_\beta(s)$  the value of the game when the play starts at  $v_0$  following strategy  $s$ .

Consider two situations  $s$  and  $s'$  such that  $a_\beta(s) \neq a_\beta(s')$ . By (23) and (26),

$$c_\beta(s) - c_\beta(s') = (1 - \beta)p_0[(I - \beta P_s)^{-1}Q_s - (I - \beta P_{s'})^{-1}Q_{s'}]r,$$

where  $p_0 = e_{v_0}$ .

Applying (R2), (R3) and (R4) repeatedly, we get

$$\begin{aligned}\langle c_\beta(s) - c_\beta(s') \rangle &\leq 2 + \langle \beta \rangle + 2n\langle p_0 \rangle + \langle (I - \beta P_s)^{-1} \rangle + \langle (I - \beta P_{s'})^{-1} \rangle + \langle Q_s \rangle + \langle Q_{s'} \rangle + 2n\langle r \rangle \\ &\leq 2 + \langle \beta \rangle + 2n\langle p_0 \rangle + 8n^2(4n + 2n^2\langle \beta \rangle + \langle P_s \rangle + \langle P_{s'} \rangle) + \langle Q_s \rangle + \langle Q_{s'} \rangle + 2n\langle r \rangle.\end{aligned}$$

Using the facts that  $\langle P_s \rangle \leq 2n^2\langle q \rangle$  and  $\langle Q_s \rangle \leq 2nh\langle q \rangle$ , for any situation  $s$ , and  $\langle r \rangle \leq h\langle R \rangle$ , where  $h = |E|$ , we conclude that<sup>6</sup>

$$\langle c_\beta(s) - c_\beta(s') \rangle = O(n^4)\langle \beta \rangle + O(n^4)\langle q \rangle + O(nh)\langle R \rangle.$$

---

<sup>6</sup>we remark that we made no attempt at getting the most tight upper bound



Now using (R1) we get that

$$\begin{aligned} |c_\beta(s) - c_\beta(s')| &\geq \left(\frac{1}{2}\right)^{O(n^4)\langle\beta\rangle + O(n^4)\langle q\rangle + O(nh)\langle R\rangle} \\ &\geq \left(\frac{1}{qBB'}\right)^{O(n^4)} \cdot \left(\frac{1}{R}\right)^{O(nh)}. \end{aligned}$$

This gives an upper bound on the necessary accuracy  $\varepsilon$ .

## C.4 Condon's example

For the following very simple discounted weighted Markov chain (R-game) with  $n + 1$  positions, we have to choose  $\beta \in [1 - \varepsilon/2^n, 1)$  to guarantee that the Cesaro- and Abel-average differ at most by  $\varepsilon$ .

Let  $V = \{v_0, v_1, \dots, v_n\}$ , where the last position  $v_n$  is absorbing; there is only one move in it, which is a loop with the local reward  $r(v_n, v_n) = 1$ . In every other position  $v_i \in V$  for  $i = 1, \dots, n - 1$ , there are two moves  $(v_i, v_{i+1})$  and  $(v_i, v_0)$ , both with the local reward 0 and probability  $1/2$ .

**Remark 2** *It is easy to verify that we reduce this chain to a canonical form choosing potentials  $x(v_i) = 2(2^n - 2^i)$  for  $i = 0, 1, \dots, n$ .*

In this example  $v_n$  is a unique absorbing position and hence, the Cesaro-sum  $c$  is equal to 1. In other words, almost all plays come to  $v_n$  “sooner or later, yet, they do it rather later than sooner”.

To count the Abel-average, let us denote by  $p_k$  the probability to be at  $v_n$  after  $k$  moves. Obviously,  $p_k = 0$  for  $k < n$  and  $p_k = 1/2^n$  for  $k = n$ ; furthermore,  $p_k \rightarrow 1$ , as  $k \rightarrow \infty$ . The following upper bound holds:

$$p_k \leq 1 - (1 - 2^{-n})^k = 1 - (1 - 2^{-n})^{2^n(k/2^n)} \sim 1 - e^{-k/2^n}$$

Furthermore, let  $\beta = 1 - 1/B$ , then  $1 - \beta = 1/B$  and

$$\beta^k = (1 - 1/B)^k = (1 - 1/B)^{B(k/B)} \sim e^{-k/B}.$$

Now, let us rewrite the inequality

$$|c - a(\beta)| \leq \varepsilon \text{ as } a(\beta) = (1 - \beta) \sum_{k=0}^{\infty} p_k \beta^k > 1 - \varepsilon.$$

Then, substituting  $p_k$ ,  $\beta^k$ , and  $1 - \beta = 1/B$  we obtain

$$\sum_{k=0}^{\infty} p_k \beta^k \sim \sum_{k=0}^{\infty} (1 - e^{-k/2^{-n}}) e^{-k/B} = \sum_{k=0}^{\infty} e^{-k/B} - \sum_{k=0}^{\infty} e^{-k(2^{-n}+1/B)} \geq B(1 - \varepsilon).$$

Let us recall that  $\sum_{k=0}^{\infty} z^k = (1 - z)^{-1}$  and rewrite this inequality as

$$(1 - e^{-B})^{-1} - (1 - e^{2^{-n}+1/B})^{-1} \geq B(1 - \varepsilon).$$

Finally, let us approximate  $e^y$  by  $1 + y$  and simplify the expression to get

$$B \geq 2^n(\varepsilon^{-1} - 1) \text{ or asymptotically just } B \geq 2^n/\varepsilon$$

**Remark 3** *In contrast, the undiscounted BW-games can be approximated by discounted BW-games very efficiently. The former are polynomially reduced to the latter, as it follows from the following observation by Zwick and Paterson [ZP96]: It is sufficient to choose  $\beta > 1 - \frac{1}{4n^3R}$  to get the exact solution of the corresponding undiscounted BW-game, where  $n = |V|$  is the number of positions and  $R = |[r]| = r^+ - r^-$  is the range of the reward function  $r$ .*

*Let us also recall that for the undiscounted BW-games the accuracy  $\varepsilon = 1/n^2$  is sufficient, while for the undiscounted BWR-games much smaller  $\varepsilon$  is needed.*

*In particular, Condon's example shows that  $\beta \rightarrow 1^-$  transition is pretty efficient for the BW-games but not for weighted Markov chains (R-games). In general, these two classes are "sort of opposite". For this reason, the BWR-games look more difficult than than BW-games and R-games.*

## Appendix D: Omitted proofs

**Proof of Proposition 1.** Let  $B$  be the  $|E| \times |V|$  matrix with entries defined as follows:  $b(e, v) = 1$  and  $b(e, u) = -\beta$  if  $e = (v, u)$  and  $v \neq u$ ,  $b(e, v) = 1 - \beta$  if  $e = (v, v)$ , and  $b(e, u) = 0$  otherwise. Then we can write (4) in matrix form as  $r_x = r + Bx$ . Fix an optimal situation  $s$ , and let  $P_s$  be the transition matrix obtained by extending  $P$  as in (3), and  $Q_s$  and  $P_s^*$  be the corresponding  $Q$ -matrix and limiting transition matrix, respectively. Note that  $Q_s B = I - \beta P_s$ .

Given two sequences  $\{b_i, x_i\}_{i=0}^{\infty}$ , and a discount factor  $\beta \in [0, 1)$ , let

$$a(\beta) = (1 - \beta) \sum_{i=0}^{\infty} \beta^i b_i \text{ and } a_x(\beta) = (1 - \beta) \sum_{i=0}^{\infty} \beta^i (b_i + x_i - \beta x_{i+1}); \quad (31)$$

it is obvious that both series are Abel-summable (or not) simultaneously and

$$a_x(\beta) = a(\beta) + (1 - \beta)x_0. \quad (32)$$

In particular,  $a_x(\beta) \rightarrow a(\beta)$ , as  $\beta \rightarrow 1^-$ . Moreover, the undiscounted transform respects the limit Cesaro sum:  $c_x = c$ .

Consider now an infinite play  $v_0, v_1, v_2, \dots$  (Of course, positions will repeat themselves, since game is finite.) Let us set  $b_i = \mathbb{E}[r(v_i, v_{i+1})] = p_0 P_s^i Q_s r$  as given by (23) and  $x_i = \mathbb{E}[x(v_i)] = p_0 P_s^i x$  for  $i = 0, 1, 2, \dots$ . Then,

$$b_i + x_i - \beta x_{i+1} = p_0 P_s^i [Q_s r + x - \beta P_s x] = p_0 P_s^i Q_s (r + Bx) = p_0 P_s^i Q_s r_x$$

and hence the statement follows by (31) and (32). □

**Proof of Proposition 2.** It is straightforward. Indeed, if White (Black) applies a locally optimal strategy then after every own move (s)he will get (pay)  $m$ , while for each move of the opponent the local reward will be at least (at most)  $m$ , and finally, for each random position the expected local reward is  $m$ . Thus, every locally optimal strategy of a player is optimal.

Furthermore, if both players choose their optimal strategies then the expected local reward  $b_i$  equals  $m$  for every step  $i$ . Hence, the  $\beta$ -discounted value  $(1 - \beta) \sum_{i=0}^{\infty} m \beta^i$  for each  $\beta \in [0, 1)$  and its limit, as  $\beta \rightarrow 1^-$ , which is the undiscounted value equals  $m$ , too. □

**Proof of Proposition 3.** We just repeat the arguments from the proof of Proposition 2. Yet, in part (ii)  $b_i \equiv m(v)$  can now depend on the initial position  $v$ . However, it does not depend on  $i$ , as before. □

**Proof of Proposition 4.** Let us consider potentials  $x(v) = Cm(v)$ , where  $C$  is a constant, and apply the undiscounted potential transform:

$$r_x(v, u) = r(v, u) + x(v) - x(u) = r(v, u) + C(m(v) - m(u)).$$

First, let us consider a random position  $v \in V_R$ . In this case, by definition,  $m(v) = \sum_u p(v, u)r(v, u)$  and for the transformed value  $m_x(v)$  we obtain

$$\begin{aligned} m_x(v) &= \sum_u p(v, u)r_x(v, u) = \sum_u p(v, u)[r(v, u) + C(m(v) - m(u))] \\ &= \sum_u p(v, u)r(v, u) + Cm(v) \sum_u p(v, u) - C \sum_u p(v, u)m(u) \\ &= (C + 1)m(v) - Cm(v). \end{aligned}$$

Furthermore,  $m(v) = M(v)$  for every game in weak canonical form. Hence, in this case  $m_x(v) = m(v)$ , that is, the value of  $m$  is kept by transform  $x$ .

Now let  $v \in V_W \cup V_B$ . Then  $r_x(v, u) = r(v, u) + C(m(v) - m(u))$ . In particular,  $r_x(v, u) = r(v, u)$  whenever  $m(v) = m(u)$ .

Given a game in weak canonical form, let us assume that  $m(v) \neq m(u)$ .

If  $v \in V_W$  (respectively,  $v \in V_B$ ) then  $m(v) > m(u)$  (respectively,  $m(v) < m(u)$ ), since already in a *weak* canonical form no player can improve  $m$ . Hence, if  $C$  is a very large negative constant then a very large negative (respectively, positive) local reward  $r_x(v, u)$  is assigned to each move of White (respectively, of Black) that does not respect the value of  $m$ . In other words, these not optimal moves become also *locally* unattractive.  $\square$

**Proof of Lemma 1.** We show that the range of  $m$  decreases in each iteration by a factor

$$c = \frac{1 + \beta}{2} = 1 - \frac{1 - \beta}{2}. \quad (33)$$

In fact,  $a, b$  will be chosen such that this is the maximum possible decrease in one iteration. Without loss of generality, we can assume that  $m$  is the unit interval,  $[m] = [0, 1]$ . Indeed, if  $[m]$  is just one point then the game is already in an ergodic canonical form and the problem is solved; otherwise, there is unique (bijective) linear map of  $[m_x]$  onto  $[0, 1]$ .

Given two parameters  $a, b \in [0, 1]$ , let us define potential  $x = x_{a,b}$  as follows:  $x(v) = -b$  for a vertex  $v \in V$  whenever  $m(v) \geq a$  and  $x(v) = 0$  otherwise. We will show that the optimal choice  $a = b = \frac{1}{2}$  results in  $[m_x] = [0, c]$ , where  $c$  is given by (33).

Indeed, it is easy to verify that

$$\begin{aligned} (1 - \beta)b \leq m(v) - m_x(v) \leq b \text{ if } m(v) \geq a, \text{ while} \\ 0 \leq m_x(v) - m(v) \leq \beta b \text{ if } m(v) < a. \end{aligned}$$

Clearly,  $a \geq b$  should hold, since otherwise  $m_x(v)$  could become negative for a vertex  $v$  such that  $m(v) = a$ .

On the other hand, we have to minimize  $c$  subject to  $m(v) \notin [c, 1]$  for all  $v \in V$ . Hence,  $c \geq a + \beta b$  and  $c \geq 1 - (1 - \beta)b$ . To optimize, we set  $a + \beta b = 1 - (1 - \beta)b$  which results in  $a + b = 1$ .

Finally, we have to minimize  $c = a + \beta b$  subject to  $0 \leq b \leq a$ ,  $a + b = 1$ , and  $0 \leq \beta \leq 1$ . Obviously, the optimal  $c$  is given by (33) when  $a = b = \frac{1}{2}$ .

Thus, in one iteration range  $[m]$  is reduced at least by the factor (33). Using  $[m] \subseteq [r]$ , we must have, after  $N$  iterations,

$$|[m_x]| \leq |[r]| \left( \frac{1 + \beta}{2} \right)^N \leq \varepsilon, \quad (34)$$

by our choice of  $N$ . □

**Proof of Theorem 4.** Obviously, a BWR-game  $\mathcal{G} = (G, P, r)$  is just a special case of Gillette game. Conversely, given a Gillette game  $\mathcal{G} = (V_W, V_B, P, r)$ , it can be reduced to a BWR-game  $\mathcal{G}' = (G' = (V' = V'_W \cup V'_B \cup V'_R, E'), P', r')$ , by setting  $V'_W = V_W$ ,  $V'_B = V_B$ , and introducing for every vertex  $v \in V$ , and every action  $k \in S(v)$ , a vertex  $v^k \in V'_R$ , and an arc  $(v, v^k)$  with reward  $r'(v, v^k) = 0$ . For every  $v, u \in V$ , and every action  $k \in S(v)$ , we introduce an arc  $(v^k, u) \in E'$  with reward  $r'(v^k, u) = r^k(v, u)$  and transition probability  $p'(v^k, u) = p^k(v, u)$ .

Finally, in the obtained BWR-game we introduce the discount factor  $\sqrt{\beta}$  whenever the original Gillette game was  $\beta$ -discounted. Since the local payoff is 0 for every odd move, the obtained to games are equivalent modulo some constant factor. More precisely, if  $\mu^\beta(v)$ ,  $\tilde{\mu}^\beta(v)$  are respectively the values of the Gillette's game and the BWR-games at vertex  $v$ , then

$$\begin{aligned} \tilde{\mu}^\beta &= (1 - \sqrt{\beta}) \sum_{i: i \text{ is odd}} (\sqrt{\beta})^i \tilde{r}_i = (1 - \sqrt{\beta}) \sum_{i: i \text{ is odd}} (\sqrt{\beta})^i \frac{r_{\lfloor i/2 \rfloor}}{\sqrt{\beta}} \\ &= \frac{1 - \beta}{1 + \sqrt{\beta}} \sum_{i=0}^{\infty} \beta^i r_i = \frac{\mu^\beta(v)}{1 + \sqrt{\beta}}, \end{aligned} \quad (35)$$

where  $r_i$  and  $\tilde{r}_i$  denote the expected local rewards at step  $i$  of the play in the Gillette and the BWR-games, respectively. Clearly, the undiscounted games are equivalent too, since they both are obtained by the same limit transition  $\beta \rightarrow 1^-$ . □

**Proof of Theorem 5.** We derive the result from Theorem 3. Consider the reduction used in the proof of Theorem 4. By Theorem 2, there exist a potential transformation  $\tilde{x} : V' \rightarrow \mathbb{R}$ , and a mapping  $\tilde{m} : V' \rightarrow \mathbb{R}$  such that

(i) for every  $v \in V_W = V'_W$ , there is a  $k \in S(v)$  such that

$$\tilde{x}(v) - \tilde{x}(v^k) = \tilde{m}(v) = \tilde{m}(v^k), \quad (36)$$

$$\tilde{x}(v) - \tilde{x}(v^{k'}) \leq \tilde{m}(v) \text{ for all } k' \in S \setminus \{k\}, \quad (37)$$

$$\tilde{m}(v^{k'}) \leq \tilde{m}(v) \text{ for all } k' \in S(v) \setminus \{k\} \quad (38)$$

(ii) similarly, for every  $v \in V_B = V'_B$ , there is a  $k \in S(v)$  such that  $\tilde{x}(v) - \tilde{x}(v^k) = \tilde{m}(v) = \tilde{m}(v^k)$ ,  $\tilde{x}(v) - \tilde{x}(v^{k'}) \geq \tilde{m}(v)$  and  $\tilde{m}(v^{k'}) \geq \tilde{m}(v)$  for all  $k' \in S(v) \setminus \{k\}$ ;

(iii) for every  $v \in V$  and  $k \in S(v)$ ,

$$\tilde{m}(v^k) = \sum_{u: (v^k, u) \in E'} p^k(v, u) (r^k(v, u) + \tilde{x}(v^k) - \tilde{x}(u)) \quad (39)$$

$$\tilde{m}(v^k) = \sum_{u: (v^k, u) \in E'} p^k(v, u) \tilde{m}(u). \quad (40)$$

We define  $m(v) = 2\tilde{m}(v)$  and  $x(v) = \tilde{x}(v)$ , for all  $v \in V$ . Consider, without loss of generality, a vertex  $v \in V_W$ . It follows from (36) and (40) that there is a  $k \in S(v)$  such that  $m(v) = \sum_{(v,u) \in E} p^k(v,u)m(u)$  and from (38) and (40) that  $m(v) \geq \sum_{(v,u) \in E} p^{k'}(v,u)m(u)$  for any  $k' \in S(v)$ . It follows also from (36) and (39) that  $\tilde{m}(v) = \sum_{(v,u) \in E} p^k(v,u)(r^k(v,u) + x(v) - x(u))$ , and from (37),(38), and (39) that  $m(v) \geq \tilde{x}(v) - \tilde{x}(v^{k'}) + \tilde{m}(v^{k'}) = \tilde{x}(v) - \tilde{x}(v^{k'}) + \sum_{u:(v^k,u) \in E'} p^{k'}(v,u)(r^{k'}(v,u) + \tilde{x}(v^{k'}) - \tilde{x}(u)) = \sum_{u:(v^k,u) \in E'} p^{k'}(v,u)(r^{k'}(v,u) + \tilde{x}(v) - \tilde{x}(u))$ , for any  $k' \in S_w(v)$ . The theorem follows.  $\square$

**Proof of Theorem 6.** We derive the result from Theorem 2. We use the same reduction as before but to a discount BWR-game with discount factor  $\sqrt{\beta}$ , and for every  $v, u \in V$  such that  $(v, u) \in E$ , and every action  $k \in S(v)$ , we introduce an arc  $(v^k, u) \in E'$  with reward  $r'(v^k, v) = r^k(v, u)/\sqrt{\beta}$ . By Theorem 2, there exist a potential transformation  $\tilde{x} : V' \rightarrow \mathbb{R}$  and a constant  $\tilde{m}$  such that

(i) for every  $v \in V_W = V'_W$ , there is a  $k \in S(v)$  such that

$$\tilde{x}(v) - \sqrt{\beta}\tilde{x}(v^k) = \tilde{m}, \quad (41)$$

$$\tilde{x}(v) - \sqrt{\beta}\tilde{x}(v^{k'}) \leq \tilde{m} \text{ for all } k' \in S(v) \setminus \{k\}, \quad (42)$$

$$(43)$$

(ii) similarly, for every  $v \in V_B = V'_B$ , there is a  $k \in S(v)$  such that  $\tilde{x}(v) - \sqrt{\beta}\tilde{x}(v^k) = \tilde{m}$ ,  $\tilde{x}(v) - \beta\tilde{x}(v^{k'}) \geq \tilde{m}$  for all  $k' \in S(v) \setminus \{k\}$ ;

(iii) for every  $v \in V$  and  $k \in S(v)$ ,

$$\tilde{m} = \sum_{u:(v^k,u) \in E'} p^k(v,u) \left( \frac{r^k(v,u)}{\sqrt{\beta}} + \tilde{x}(v^k) - \sqrt{\beta}\tilde{x}(u) \right); \quad (44)$$

(iv) the value of the game at vertex  $v \in V'$  is given by  $\tilde{\mu}(v) = \tilde{m} - (1 - \sqrt{\beta})\tilde{x}(v)$ .

Multiplying (44) by  $\sqrt{\beta}$  and adding to (41) (or (42)) shows the existence of the claimed constant  $m = (1 + \sqrt{\beta})\tilde{m}$  and the potential  $x(v) = \tilde{x}(v)$  for all  $v \in V$ . By (35), the value of the GGPI at vertex  $v \in V$  is given by

$$\mu^\beta(v) = (1 + \sqrt{\beta})\tilde{\mu}^\beta(v) = (1 + \sqrt{\beta})(\tilde{m} - (1 - \sqrt{\beta})\tilde{x}(v)) = m - (1 - \beta)x(v),$$

using (iv).  $\square$