Uniform value in Dynamic Programming

Jérôme Renault*

14 janvier 2007

Abstract

We consider dynamic programming problems and give sufficient conditions for the existence of the uniform value, which implies the existence of ε-Blackwell optimal plays. As a consequence, we obtain the existence of the uniform value when the state space is compact metric, payoffs are continuous and the transition correspondence is non expansive. We also apply our results to Markov Decision Processes and obtain a few generalizations.

Key words. Markov Decision Processes, Blackwell optimality, average payoffs, bounded payoffs, precompact state space, non expansive correspondence.

1 Introduction

We consider a Dynamic Programming Problem with infinite time horizon, or equivalently a deterministic Markov Decision Process (MDP hereafter). We assume that payoffs are bounded and denote, for each n, the value of the n-stage problem with average payoffs by v_n. The uniform value of the MDP exists iif (v_n) converges to some limit v, and for each $\varepsilon > 0$ there exists a play giving a payoff not lower than $v - \varepsilon$ in any sufficiently long n-stage problem. So when the uniform value exists, a decision maker can play ε-optimally simultaneously in any long enough problem. This notion is linked to the average cost criterion, see Araposthathis *et al.*, 1993, or Hernández-Lerma and Lasserre, 1996, Chapter 5. We show that it is slightly stronger than : the existence of a limit for the discounted values, together with the existence of ε-Blackwell optimal plays, i.e. plays which are ε-optimal in any discounted problem with low enough discount factor (see Rosenberg *al.*, 2002).

In our setup, Monderer and Sorin (1993), and Lehrer and Monderer (1994) showed that the uniform convergence of $(v_n)_n$ (or/and of the discounted value (v_λ)) was not enough to ensure the existence of the uniform value. In the context

*Ceremade, Université Paris Dauphine, Place du Maréchal de Lattre, 75776 Paris Cedex 16, France. email:renault@ceremade.dauphine.fr

of zero-sum stochastic games, Mertens and Neyman (1981) provided sufficient conditions, different from ours, on the discounted values to ensure the existence of the uniform value.

We define here, for every m and n, a value $w_{m,n}$ as the supremum payoff the decision maker can achieve when his payoff is defined as the minimum, for t in $\{1,..,n\}$, of his average rewards computed between stages $m+1$ and $m+t$. Our main result, theorem 3.6, states that if the set of states is a precompact metric space and the family $(w_{m,n})$ is uniformly equicontinuous, then the uniform value exists. Moreover, it is equal to $\sup_{m\geq 0}\inf_{n\geq 1} w_{m,n}(z) = \inf_{n\geq 1}\sup_{m\geq 0} w_{m,n}(z)$, and other relations are obtained. This result is stated in section 3, and proved in the next section. Section 5 contains comments about stationary ε-optimal plays, ε-Blackwell optimality, examples and counterexamples, and also the corollary of theorem 3.6 announced in the abstract. We finally consider probabilistic MDPs in section 6 and show : 1) in a usual probabilistic MDP with finite set of states and arbitrary set of actions, the uniform value exists, and 2) if the decision maker can randomly select his actions, the same result also holds when there is imperfect observation of the state.

This work was motivated by the study of a particular class of repeated games generalizing those introduced in Renault, 2006. Theorem 3.6 can also be used to prove the existence of the uniform value in a specific class of stochastic games, which leads to the existence of the value in general repeated games with an informed controller. This is done in a companion paper (see Renault, 2007).

2 Model

We consider, as in Lehrer Sorin 1992, Monderer Sorin 1993 and Lehrer Monderer 1994 :

- a non empty set Z,
- a correspondence F from Z to Z with non empty values,
- a mapping r from Z to $[0,1]$,
- and an element z_0 in Z.

Z is called the set of states, F is the transition correspondence, r is the reward (or payoff) function, and z_0 is called the initial state. The interpretation is the following. The initial state is z_0. A decision maker (also called player) first has to select a new state z_1 in $F(z_0)$, and is rewarded by $r(z_1)$. Then he has to choose z_2 in $F(z_1)$, has a payoff of $r(z_2)$, etc... The decision maker is interested in maximizing his "long run average payoffs", i.e. quantities $\frac{1}{t}(r(z_1)+r(z_2)+...+r(z_t))$ for t large.

(Z,F,r,z_0) is called a deterministic Markov decision problem (MDP, for short), or a dynamic programming problem.

Notice that it is assumed that r takes values in $[0,1]$, so compared with a model with rewards in $\mathbb{R}$ we are essentially making the hypothesis that rewards

are bounded. From now on we fix $\Gamma = (Z, F, r)$, and for every state z_0 we denote by $\Gamma(z_0) = (Z, F, r, z_0)$ the corresponding MDP with initial state z_0.

For z_0 in Z, a play at z_0 is a sequence $s = (z_1, ..., z_t, ...) \in Z^\infty$ such that : $\forall t \geq 1, z_t \in F(z_{t-1})$. We denote by $S(z_0)$ the set of plays at z_0, and by $S = \cup_{z_0 \in Z} S(z_0)$ the set of all plays.

For $n \geq 1$ and $s = (z_t)_{t \geq 1} \in S$, we define the average payoff of s up to stage n by :

$$\gamma_n(s) = \frac{1}{n} \sum_{t=1}^{n} r(z_t).$$

For $n \geq 1$ and $z \in Z$, the n-stage value of $\Gamma(z) = (Z, F, r, z)$ is :

$$v_n(z) = \sup_{s \in S(z)} \gamma_n(s).$$

Definition 2.1. *Let z be in Z.*

$$\textit{The liminf value of } \Gamma(z) \textit{ is : } v^-(z) = \liminf_n v_n(z).$$

$$\textit{The limsup value of } \Gamma(z) \textit{ is : } v^+(z) = \limsup_n v_n(z).$$

We say that the decision maker can guarantee, or secure, the payoff x in $\Gamma(z)$ if there exists a play s at z such that $\liminf_n \gamma_n(s) \geq x$.

The lower long-run average value is defined by :

$$\begin{aligned} \underline{v}(z) &= \sup\{x \in I\!R, \text{the decision maker can guarantee } x \text{ in } \Gamma(z)\} \\ &= \sup_{s \in S(z)} \left(\liminf_n \gamma_n(s) \right). \end{aligned}$$

Claim 2.2.

$$\underline{v}(z) \leq v^-(z) \leq v^+(z).$$

Definition 2.3.

The MDP $\Gamma(z)$ has a limit value if $v^-(z) = v^+(z)$.

The MDP $\Gamma(z)$ has a uniform value if $\underline{v}(z) = v^+(z)$.

When the limit value exists, we denote it by $v(z) = v^-(z) = v^+(z)$. For $\varepsilon \geq 0$, a play s in $S(z)$ such that $\liminf_n \gamma_n(s) \geq v(z) - \varepsilon$ is then called an ε-optimal play for $\Gamma(z)$.

We clearly have :

Claim 2.4. *$\Gamma(z)$ has a uniform value if and only if $\Gamma(z)$ has a limit value $v(z)$ and for every $\varepsilon > 0$ there exists an ε-optimal play for $\Gamma(z)$.*

Remark 2.5. *A related notion is the following (see Hernández-Lerma and Lasserre, 1996, definition 5.2.3. p. 79). A play s in $S(z)$ is "strong Average-Cost optimal in the sense of Flynn" if $\lim_n(\gamma_n(s) - v_n(z)) = 0$. Notice that $(v_n(z))$ is not assumed to converge here. A 0-optimal play for $\Gamma(z)$ satisfies this optimality condition, but in general ε-optimal plays do not.*

Remark 2.6. *Discounted payoffs.*

Other type of evaluations are used. For $\lambda \in (0,1]$, we can define the λ-discounted payoff of a play $s = (z_t)_t$ by : $\gamma_\lambda(s) = \sum_{t=1}^{\infty} \lambda(1-\lambda)^{t-1} r(z_t)$. And the λ-discounted value of $\Gamma(z)$ is $v_\lambda(z) = \sup_{s \in S(z)} \gamma_\lambda(s)$.

A play s at z_0 is said to be Blackwell optimal in $\Gamma(z_0)$ if there exists $\lambda_0 > 0$ such that for all $\lambda \in (0, \lambda_0)$, $\gamma_\lambda(s) \geq v_\lambda(z_0)$. Blackwell optimality has been extensively studied after the seminal work of Blackwell (1962) who prove the existence of such plays in the context of probabilistic MDP with finite sets of states and actions, see subsection 6.1 . A survey can be found in Hordijk and Yushkevich, 2002. In general Blackwell optimal plays do not exist. Say that a play s at z_0 is ε-Blackwell optimal in $\Gamma(z_0)$ if there exists $\lambda_0 > 0$ such that for all $\lambda \in (0, \lambda_0)$, $\gamma_\lambda(s) \geq v_\lambda(z_0) - \varepsilon$.

We will show in remark 5.9 that : 1) if $\Gamma(z)$ has a uniform value $v(z)$, then $(v_\lambda(z))_\lambda$ converges to $v(z)$ as λ goes to zero, and ε-Blackwell optimal plays exist for each positive ε. And 2) the converse is false. Consequently, the notion of uniform value is (slightly) stronger than the existence of a limit for v_λ and ε-Blackwell optimal plays.

3 Main result

We will give in the sequel sufficient conditions for the existence of the uniform value. We start with general notations and lemmas.

Definition 3.1. *For $s = (z_t)_{t \geq 1}$ in S, $m \geq 0$ and $n \geq 1$, we put :*

$$\gamma_{m,n}(s) = \frac{1}{n} \sum_{t=1}^{n} r(z_{m+t}),$$

and

$$\nu_{m,n}(s) = \min\{\gamma_{m,t}(s), t \in \{1, ..., n\}\}.$$

We have $\nu_{m,n}(s) \leq \gamma_{m,n}(s)$, and $\gamma_{0,n}(s) = \gamma_n(s)$. We also put $\nu_n(s) = \nu_{0,n}(s) = \min\{\gamma_t(s), t \in \{1, ..., n\}\}$.

Definition 3.2. *For z in Z, $m \geq 0$, and $n \geq 1$, we put :*

$$v_{m,n}(z) = \sup_{s \in S(z)} \gamma_{m,n}(s),$$

and

$$w_{m,n}(z) = \sup_{s \in S(z)} \nu_{m,n}(s).$$

We have $v_{0,n}(z) = v_n(s)$, and we also put $w_n(z) = w_{0,n}(z)$. $v_{m,n}$ corresponds to the case where the decision maker first makes m moves in order to reach a "good initial state", then plays n moves for payoffs. $w_{m,n}$ corresponds to the case where the decision maker first makes m moves in order to reach a "good initial state", but then his payoff only is the minimum of his next n average rewards (as if some

adversary trying to minimize the rewards was then able to choose the length of the remaining game). Of course we have $w_{m,n+1} \leq w_{m,n} \leq v_{m,n}$ and, since r takes values in $[0,1]$,

$$nv_n \leq (m+n)v_{m+n} \leq nv_n + m. \quad \text{and} \quad nv_{m,n} \leq (m+n)v_{m+n} \leq nv_{m,n} + m. \quad (1)$$

The next lemmas are true without any further assumption on the MDP. In some sense, they show that the quantities $w_{m,n}$ are not that low.

Lemma 3.3. $\forall k \geq 1, \forall n \geq 1, \forall m \geq 0, \forall z \in Z$,

$$v_{m,n}(z) \leq \sup_{l \geq 0} w_{l,k}(z) + \frac{k-1}{n}.$$

Proof : Fix k, n, m and z. Put $A = \sup_{l \geq 0} w_{l,k}(z)$, and consider $\varepsilon > 0$.

By definition of $v_{m,n}(z)$, there exists a play s at z such that $\gamma_{m,n}(s) \geq v_{m,n}(z) - \varepsilon$. For any $i \geq m$, we have that : $\min\{\gamma_{i,t}(s), t \in \{1, ..., k\}\} = \nu_{i,k}(s) \leq w_{i,k}(z) \leq A$. So we know that for every $i \geq m$, there exists $t(i) \in \{1, ..., k\}$ s.t. $\gamma_{i,t(i)}(s) \leq A$.

Define now by induction $i_1 = m$, $i_2 = i_1 + t(i_1)$,..., $i_q = i_{q-1} + t(i_{q-1})$, where q is such that $i_q \leq n < i_q + t(i_q)$. We have $n\gamma_{m,n}(s) \leq \sum_{p=1}^{q-1} t(i_p)A + (n - i_q)1 \leq nA + k - 1$, so $\gamma_{m,n}(s) \leq A + \frac{k-1}{n}$. □

Lemma 3.4. *For every state z in Z,*

$$v^+(z) \leq \inf_{n \geq 1} \sup_{m \geq 0} w_{m,n}(z) = \inf_{n \geq 1} \sup_{m \geq 0} v_{m,n}(z).$$

Notice that it is not true in general that $\sup_{m \geq 0} \inf_{n \geq 1} w_{m,n}(z) = \sup_{m \geq 0} \inf_{n \geq 1} v_{m,n}(z)$, as example 5.5 will show later.

Proof of lemma 3.4 : Using lemma 3.3 with $m = 0$ and arbitrary positive k, we can obtain $\limsup_n v_n(z) \leq \sup_{l \geq 0} w_{l,k}(z)$. So $v^+(z) \leq \inf_{n \geq 1} \sup_{m \geq 0} w_{m,n}(z)$. We always have $w_{m,n}(z) \leq v_{m,n}(z)$, so clearly $\inf_{n \geq 1} \sup_{m \geq 0} w_{m,n}(z) \leq \inf_{n \geq 1} \sup_{m \geq 0} v_{m,n}(z)$. Finally, lemma 3.3 gives : $\forall k \geq 1$, $\forall n \geq 1$, $\forall m \geq 0$, $v_{m,nk}(z) \leq \sup_{l \geq 0} w_{l,k}(z) + \frac{1}{n}$, so $\sup_m v_{m,nk}(z) \leq \sup_{l \geq 0} w_{l,k}(z) + \frac{1}{n}$. So $\inf_n \sup_m v_{m,n}(z) \leq \inf_n \sup_m v_{m,nk}(z) \leq \sup_{l \geq 0} w_{l,k}(z)$, and this holds for every positive k. □

Definition 3.5. *We define*

$$v^*(z) = \inf_{n \geq 1} \sup_{m \geq 0} w_{m,n}(z) = \inf_{n \geq 1} \sup_{m \geq 0} v_{m,n}(z).$$

We now have the following chain of inequalities :

$$\underline{v}(z) \leq v^-(z) \leq v^+(z) \leq v^*(z), \quad (2)$$

and we now state our main result.

Theorem 3.6. *Let Z be a non empty set, F be a correspondence from Z to Z with non empty values, and r be a mapping from Z to $[0,1]$. Assume that Z is endowed with a distance d such that :*

a) (Z,d) is a precompact metric space, and

b) $(w_{m,n})_{m\geq 0,n\geq 1}$ is a uniformly equicontinuous family of mappings from Z to $[0,1]$.

Then for every initial state z in Z, the MDP $\Gamma(z)=(Z,F,r,z)$ has a uniform value which is :

$$v^*(z)=\underline{v}(z)=v^-(z)=v^+(z)=\sup_{m\geq 0}\inf_{n\geq 1}w_{m,n}(z)=\sup_{m\geq 0}\inf_{n\geq 1}v_{m,n}(z).$$

And the sequence $(v_n)_n$ uniformly converges to v^*.

Notice that if Z is finite, we can consider d such that $d(z,z')=1$ if $z\neq z'$, so theorem 3.6 gives the well known result : in the finite case, the uniform value exists.

4 Proof of theorem 3.6

In the sequel we will prove theorem 3.6. We assume that Z, F and r satisfy the hypotheses of the theorem. The proof requires several steps. We start with preliminaries.

Step 0. Completing Z.

(Z,d) may not be compact, but the completion of (Z,d) is a compact metric space. We simply denote by $(\overline{Z},d)$ a compact metric space extending (Z,d), and such that the closure of Z in $\overline{Z}$ is $\overline{Z}$. In the sequel if C is a subset of $\overline{Z}$, we will denote by $\overline{C}$ its closure in $\overline{Z}$.

Step 1. Iterating F.

Given two correspondences G and H from Z to Z, the composition $G\circ H$ is defined as the correspondence from Z to Z such that : $\forall z\in Z$, $G\circ H(z)=\{z"\in Z,\exists z'\in H(z),z"\in G(z')\}$. We now inductively define a sequence of correspondences $(F^n)_n$ from Z to Z, with $F^0(z)=\{z\}$ for every state z, and $\forall n\in I\!N$, $F^{n+1}=F^n\circ F$.

$F^n(z)$ represents the set of states that the decision maker can reach in n stages if the initial state is z. It is easily shown by induction on m that :

$$\forall m\geq 0,\forall n\geq 1,\forall z\in Z,\quad w_{m,n}(z)=\sup_{y\in F^m(z)}w_n(y). \tag{3}$$

We also define, for every initial state z : $\overline{F}^n(z)=\overline{F^n(z)}$, $K^n(z)=\bigcup_{m=0}^n\overline{F}^n(z)$, and $K^\infty(z)=\overline{\bigcup_{n=0}^\infty K^n(z)}=\overline{\bigcup_{n=0}^\infty F^n(z)}$. The set $K^\infty(z)$ is the closure of the set of states that the decision maker, starting from z, can reach in a finite number of stages.

Since $(K^n(z))_n$ is a sequence of compact subsets of the metric compact space $\overline{Z}$, it is easy to see, and fundamental for us, that $(K^n(z))_n$ converges, for the Haussdorff topology, to $K^\infty(z)$, so :

$$\forall \varepsilon > 0, \forall z \in Z, \exists n \in I\!N, \forall x \in K^\infty(z), \exists y \in K^n(z) \text{ s.t. } d(x,y) \leq \varepsilon. \qquad (4)$$

Step 2. Modulus of continuity.

By assumption b), we know that : $\forall \varepsilon > 0, \exists \alpha > 0, \forall z \in Z, \forall z' \in Z, \forall m \geq 0, \forall n \geq 1, \ \ (d(z,z') \leq \alpha \Longrightarrow |w_{m,n}(z) - w_{m,n}(z')| \leq \varepsilon)$. We define, for every $\alpha \geq 0$, $\hat{\varepsilon}(\alpha) = \sup_{m,n\, z,z' \text{s.t.} d(z,z') \leq \alpha} |w_{m,n}(z) - w_{m,n}(z')|$. The function $\hat{\varepsilon}$ is called a modulus of continuity for the family $(w_{m,n})_{m\geq 0, n \geq 1}$. $\hat{\varepsilon}(0) = 0$, $\hat{\varepsilon}$ is non decreasing and continuous at 0. We have for all $m \geq 0$ and $n \geq 1$: $\forall z \in Z$, $\forall z' \in Z$, $|w_{m,n}(z) - w_{m,n}(z')| \leq \hat{\varepsilon}(d(z,z'))$.

In the sequel, we will say that a mapping f from Z to the reals is $\hat{\varepsilon}$ uniformly continuous if f admits $\hat{\varepsilon}$ as a modulus of continuity, i.e. if : $\forall z \in Z$, $\forall z' \in Z$, $|f(z) - f(z')| \leq \hat{\varepsilon}(d(z,z'))$.

Recall that for each state z, $v^*(z) = \inf_{n\geq 1} \sup_{m \geq 0} w_{m,n}(z)$. Since $w_{m,n}(z) \leq w_{m,n}(z') + \hat{\varepsilon}(d(z,z'))$ for all m, n, z, z', we obtain that v^* itself is $\hat{\varepsilon}$ uniformly continuous.

Step 3. Convergence of $(v_n(z))_n$.

3.a. Here we will show that :

$$\forall \varepsilon > 0, \forall z \in Z, \exists M \geq 0, \forall n \geq 1, \exists m \leq M \text{ s.t. } w_{m,n}(z) \geq v^*(z) - \varepsilon.$$

Fix $\varepsilon > 0$ and z in Z. There exists $\alpha > 0$ such that $\hat{\varepsilon}(2\alpha) \leq \varepsilon$, and by property (4) there exists M in $I\!N$ such that : $\forall x \in K^\infty(z) \exists y \in K^M(z)$ s.t. $d(x,y) \leq \alpha$. Take any positive n.

There exists $m(n)$ such that $w_{m(n),n}(z) \geq \sup_m w_{m,n}(z) - \varepsilon$. So by equation (3) and the definition of v^*, we have : $\sup_{y \in F^{m(n)}(z)} w_n(y) = w_{m(n),n}(z) \geq v^*(z) - \varepsilon$. So one can find y_n in $F^{m(n)}(z)$ s.t. $w_n(y_n) \geq v^*(z) - 2\varepsilon$. Since $y_n \in K^\infty(z)$, there exists y'_n in $K^M(z)$ such that $d(y_n, y'_n) \leq \alpha$. Using the definition of $K^M(z)$, one can find $m' \in \{0, ..., M\}$ and y''_n in $F^{m'}(z)$ such that $d(y''_n, y_n) \leq 2\alpha$. So $|w_n(y''_n) - w_n(y_n)| \leq \hat{\varepsilon}(2\alpha) \leq \varepsilon$, and consequently $w_n(y''_n) \geq v^*(z) - 3\varepsilon$. This concludes step 3.a.

3.b. Fix $\varepsilon > 0$ and z in Z. Step 3.a. defines some $M \geq 0$ such that $\forall n \geq 1, \exists m \leq M$ s.t. $w_{m,n}(z) \geq v^*(z) - \varepsilon$. Consider some m in $\{0, ..., M\}$such that $w_{m,n}(z) \geq v^*(z) - \varepsilon$ is true for infinitely many n's. Since $w_{m,n+1} \leq w_{m,n}$, $w_{m,n}(z) \geq v^*(z) - \varepsilon$ is true for every n. We have improved step 3.a. and obtained :

$$\forall \varepsilon > 0, \forall z \in Z, \exists m \geq 0, \forall n \geq 1, \ w_{m,n}(z) \geq v^*(z) - \varepsilon. \qquad (5)$$

ε, z and m being fixed, we have for every n, using inequalities (1) : $(n+m)v_{n+m}(z) \geq n v_{m,n}(z) \geq n w_{m,n}(z) \geq n(v^*(z) - \varepsilon)$. Dividing by $n+m$ and taking

the liminf as n goes to infinity gives that $v^-(z) \geq v^*(z) - \varepsilon$. This is true for every positive ε, so $v^-(z) \geq v^*(z)$. Inequalities (2) give :

$$v^-(z) = v^+(z) = v^*(z), \text{ so } (v_n(z))_n \text{ converges to } v^*(z).$$

3.c. Property (5) gives : $\forall z \in Z, \forall \varepsilon > 0, \sup_m \inf_n w_{m,n}(z) \geq v^*(z) - \varepsilon$. So for every initial state z, $\sup_m \inf_n w_{m,n}(z) \geq v^*(z)$. We have : $v^*(z) \leq \sup_m \inf_n w_{m,n}(z) \leq \sup_m \inf_n v_{m,n}(z) \leq \inf_n \sup_m v_{m,n}(z)$, and the last quantity is by definition $v^*(z)$. So we also obtain :

$$v^*(z) = \sup_m \inf_n w_{m,n}(z) = \sup_m \inf_n v_{m,n}(z).$$

Step 4. Uniform convergence of $(v_n)_n$.

4.a. Write, for each state z and $n \geq 1$: $f_n(z) = \sup_{m \geq 0} w_{m,n}(z)$. We have : $v^*(z) = \inf_n f_n(z)$. $(f_n)_n$ is non increasing, and simply converges to v^*. Each f_n is $\hat{\varepsilon}$ uniformly continuous, and Z is precompact, so $(f_n)_n$ uniformly converges to v^*.

As a consequence we get :

$$\forall \varepsilon > 0, \exists n_0, \forall z \in Z, \ \sup_{m \geq 0} w_{m,n_0}(z) \leq v^*(z) + \varepsilon.$$

By lemma 3.3, we obtain :

$$\forall \varepsilon > 0, \exists n_0, \forall z \in Z, \forall m \geq 0, \forall n \geq 1, v_{m,n}(z) \leq v^*(z) + \varepsilon + \frac{n_0 - 1}{n}.$$

Considering $n_1 \geq n_0/\varepsilon$ gives :

$$\forall \varepsilon > 0, \exists n_1, \forall z \in Z, \forall n \geq n_1, v_n(z) \leq \sup_{m \geq 0} v_{m,n}(z) \leq v^*(z) + 2\varepsilon \qquad (6)$$

4.b. Write now, for each state z and $m \geq 0$: $g_m(z) = \sup_{m' \leq m} \inf_n w_{m',n}(z)$. $(g_m)_m$ is non decreasing and we have : $v^*(z) = \sup_m \inf_n w_{m,n}(z) = \lim_{m \to \infty} g_m(z)$. Each g_m is $\hat{\varepsilon}$ uniformly continuous and Z is precompact, so $(g_m)_m$ uniformly converges to v^*.

As a consequence we obtain :

$$\forall \varepsilon > 0, \exists M \geq 0, \forall z \in Z, \exists m \leq M, \ \inf_{n \geq 1} w_{m,n}(z) \geq v^*(z) - \varepsilon. \qquad (7)$$

Fix $\varepsilon > 0$, and consider M given above. Consider $N \geq M/\varepsilon$. Then $\forall z \in Z$, $\forall n \geq N$, $\exists m \leq M$ s.t. $w_{m,n}(z) \geq v^*(z) - \varepsilon$. But $v_n(z) \geq v_{m,n}(z) - m/n$ by (1), so we obtain $v_n(z) \geq v_{m,n}(z) - \varepsilon \geq v^*(z) - 2\varepsilon$. We have shown :

$$\forall \varepsilon > 0, \exists N, \forall z \in Z, \forall n \geq N, v_n(z) \geq v^*(z) - 2\varepsilon. \qquad (8)$$

Properties (6) and (8) show that $(v_n)_n$ uniformly converges to v^*.

Step 5. Uniform value.

By claim 2.4 to prove that $\Gamma(z)$ has a uniform value it remains to show that ε-optimal plays exist for every $\varepsilon > 0$. We start with a lemma.

Lemma 4.1. $\forall \varepsilon > 0, \exists M \geq 0, \exists K \geq 1, \forall z \in Z, \exists m \leq M, \forall n \geq K, \exists s = (z_t)_{t\geq 1} \in S(z)$ *such that :*

$$\nu_{m,n}(s) \geq v^*(z) - \varepsilon/2, \text{ and } v^*(z_{m+n}) \geq v^*(z) - \varepsilon.$$

This lemma has the same flavor as Proposition 2 in Rosenberg *et al.* (2002), and as Proposition 2 in Lehrer Sorin (1992). If we want to construct ε- optimal plays, for every large n we have to construct a play which : 1) gives good average payoffs if one stops the play at *any* large stage before n, and 2) after n stages, leaves the player with a good "target" payoff. This explains the importance of the quantities $\nu_{m,n}$ which have led to the definition of the mappings $w_{m,n}$.

Proof of lemma 4.1 : Fix $\varepsilon > 0$. Take M given by property (7). Take K given by (6) such that : $\forall z \in Z$, $\forall n \geq K$, $v_n(z) \leq \sup_m v_{m,n}(z) \leq v^*(z) + \varepsilon$.

Fix an initial state z in Z. Consider m given by (7), and $n \geq K$. We have to find $s = (z_t)_{t\geq 1} \in S(z)$ such that : $\nu_{m,n}(s) \geq v^*(z) - \varepsilon/2$, and $v^*(z_{m+n}) \geq v^*(z) - \varepsilon$.

We have $w_{m,n'}(z) \geq v^*(z) - \varepsilon$ for every $n' \geq 1$, so $w_{m,2n}(z) \geq v^*(z) - \varepsilon$, and we consider $s = (z_1, ..., z_t, ...) \in S(z)$ which is ε-optimal for $w_{m,2n}(z)$, in the sense that $\nu_{m,2n}(s) \geq w_{m,2n}(z) - \varepsilon$. We have :

$$\nu_{m,n}(s) \geq \nu_{m,2n}(s) \geq w_{m,2n}(z) - \varepsilon \geq v^*(z) - 2\varepsilon.$$

Write : $X = \gamma_{m,n}(s)$ and $Y = \gamma_{m+n,n}(s)$.

$$s \quad \underbrace{z_1 \qquad z_m} \quad \overbrace{z_{m+1} \qquad\qquad z_{m+n}}^{X} \quad \overbrace{z_{m+n+1} \qquad\qquad z_{m+2n}}^{Y}$$

Since $\nu_{m,2n}(s) \geq v^*(z) - 2\varepsilon$, we have $X \geq v^*(z) - 2\varepsilon$, and $(X + Y)/2 = \gamma_{m,2n}(s) \geq v^*(z) - 2\varepsilon$. Since $n \geq K$, we also have $X \leq v_{m,n}(z) \leq v^*(z) + \varepsilon$. And $n \geq K$ also gives $v_n(z_{m+n}) \leq v^*(z_{m+n}) + \varepsilon$, so $v^*(z_{m+n}) \geq v_n(z_{m+n}) - \varepsilon \geq Y - \varepsilon$.

$Y/2 = (X + Y)/2 - X/2$ gives $Y/2 \geq (v^*(z) - 5\varepsilon)/2$. So $Y \geq v^*(z) - 5\varepsilon$, and finally $v^*(z_{m+n}) \geq v^*(z) - 6\varepsilon$. □

Proposition 4.2. *For every state z and $\varepsilon > 0$ there exists an ε-optimal play in $\Gamma(z)$.*

Proof : Fix $\alpha > 0$.

For every $i \geq 1$, put $\varepsilon_i = \frac{\alpha}{2^i}$. Define $M_i = M(\varepsilon_i)$ and $K_i = K(\varepsilon_i)$ given by lemma 4.1 for ε_i. Define also n_i as the integer part of $1 + \text{Max}\{K_i, \frac{M_{i+1}}{\alpha}\}$, so that simply $n_i \geq K_i$ and $n_i \geq \frac{M_{i+1}}{\alpha}$.

We have : $\forall i \geq 1, \forall z \in Z, \exists m(z,i) \leq M_i, \exists s = (z_t)_{t\geq 1} \in S(z)$, s.t.

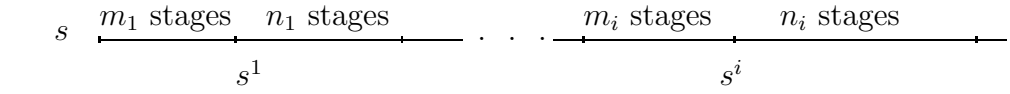
$$\nu_{m(z,i),n_i}(s) \geq v^*(z) - \frac{\alpha}{2^{i+1}} \text{ and } v^*(z_{m(z,i)+n_i}) \geq v^*(z) - \frac{\alpha}{2^i}.$$

We now fix the initial state z in Z, and for simplicity write v^* for $v^*(z)$. If $\alpha \geq v^*$ it is clear that α-optimal plays at $\Gamma(z)$ exist, so we assume $v^* - \alpha > 0$. We define a sequence $(z^i, m_i, s^i)_{i\geq 1}$ by induction :

• first put $z^1 = z$, $m_1 = m(z^1, 1) \leq M_1$, and pick $s^1 = (z_t^1)_{t\geq 1}$ in $S(z^1)$ such that $\nu_{m_1,n_1}(s^1) \geq v^*(z^1) - \frac{\alpha}{2^2}$, and $v^*(z^1_{m_1+n_1}) \geq v^*(z^1) - \frac{\alpha}{2}$.

• for $i \geq 2$, put $z^i = z^{i-1}_{m_{i-1}+n_{i-1}}$, $m_i = m(z^i, i) \leq M_i$, and pick $s^i = (z_t^i)_{t\geq 1} \in S(z^i)$ such that $\nu_{m_i,n_i}(s^i) \geq v^*(z^i) - \frac{\alpha}{2^{i+1}}$ and $v^*(z^i_{m_i+n_i}) \geq v^*(z^i) - \frac{\alpha}{2^i}$.

Consider finally $s = (z_1^1, ..., z^1_{m_1+n_1}, z_1^2, ..., z^2_{m_2+n_2}, ..., z_1^i, ..., z^i_{m_i+n_i}, z_1^{i+1}, ...)$. s is a play at z, and is defined by blocks : first s^1 is followed for $m_1 + n_1$ stages, then s^2 is followed for $m_2 + n_2$ stages, etc... Since $z^i = z^{i-1}_{m_{i-1}+n_{i-1}}$ for each i, s is a play at z. For each i we have $n_i \geq M_{i+1}/\alpha \geq m_{i+1}/\alpha$, so the "$n_i$ subblock" is much longer than the "m_{i+1} subblock".



For each $i \geq 1$, we have $v^*(z^i) \geq v^*(z^{i-1}) - \frac{\alpha}{2^{i-1}}$. So $v^*(z^i) \geq -\frac{\alpha}{2^{i-1}} - \frac{\alpha}{2^{i-2}} ... - \frac{\alpha}{2} + v^*(z^1) \geq v^* - \alpha + \frac{\alpha}{2^i}$. So $\nu_{m_i,n_i}(s^i) \geq v^* - \alpha$.

Let now T be large.

First assume that $T = m_1 + n_1 + ... + m_{i-1} + n_{i-1} + r$, for some positive i and r in $\{0, ..., m_i\}$. We have :

$$\begin{aligned}
\gamma_T(s) &= \frac{T - m_1}{T} \frac{1}{T - m_1} \sum_{t=1}^{T} g(s_t) \\
&\geq \frac{T - m_1}{T} \frac{1}{T - m_1} \sum_{t=m_1+1}^{T} g(s_t) \\
&\geq \frac{T - m_1}{T} \frac{1}{T - m_1} \left(\sum_{j=1}^{i-1} n_j \right) (v^* - \alpha)
\end{aligned}$$

But $T - m_1 \leq n_1 + m_2 + ... + n_{i-1} + m_i \leq (1 + \alpha) \left(\sum_{j=1}^{i-1} n_j \right)$, so

$$\gamma_T(s) \geq \frac{T - m_1}{T(1 + \alpha)} (v^* - \alpha).$$

And the right hand-side converges to $(v^* - \alpha)/(1 + \alpha)$ as T goes to infinity.

Assume now that $T = m_1 + n_1 + ... + m_{i-1} + n_{i-1} + m_i + r$, for some positive i and r in $\{0, ..., n_i\}$. The previous computation shows that : $\sum_{t=1}^{m_1+n_1+...+m_i} g(s_t) \geq \frac{n_1+...+m_i}{(1+\alpha)}(v^* - \alpha)$. Since $\nu_{m_i,n_i}(s^i) \geq v^* - \alpha$, we also have $\sum_{t=m_1+n_1+...+m_i+1}^{T} g(s_t) \geq r(v^* - \alpha)$. Consequently :

$$\begin{aligned}
T\gamma_T(s) &\geq (T - m_1 - r) \frac{v^* - \alpha}{1 + \alpha} + r(v^* - \alpha), \\
&\geq T \frac{v^* - \alpha}{1 + \alpha} - m_1 \frac{v^* - \alpha}{1 + \alpha} + r \frac{\alpha(v^* - \alpha)}{1 + \alpha}, \\
\gamma_T(s) &\geq \frac{v^* - \alpha}{1 + \alpha} - \frac{m_1}{T} \frac{(v^* - \alpha)}{1 + \alpha}.
\end{aligned}$$

So we obtain $\liminf_T \gamma_T(s) \geq (v^* - \alpha)/(1+\alpha) = v^* - \frac{\alpha}{1+\alpha}(1+v^*)$. We have proved the existence of a $\alpha(1+v^*)$ optimal play in $\Gamma(z)$ for every positive α, so this concludes the proofs of proposition 4.2 and consequently, of theorem 3.6.

5 Comments and consequences

Remark 5.1. *When the uniform value exists, ε-optimal play can be chosen stationary.*

A play $s = (z_t)_{t\geq 1}$ in S is said to be stationary at z_0 if there exists a mapping f from Z to Z such that for every positive t, $z_t = f(z_{t-1})$. We now show that if $\Gamma(z)$ has a uniform value, then stationary ε-optimal plays for $\Gamma(z)$ exist. We make no assumption on the MDP, and proceed as in the proof of theorem 2 in Rosenberg *et al.*, 2002.

Fix the initial state z. Consider $\varepsilon > 0$, a play $s = (z_t)_{t\geq 1}$ in $S(z)$, and T_0 such that $\forall T \geq T_0$, $\gamma_T(s) \geq v(z) - \varepsilon$.

Case 1 : Assume that there exist t_1 and t_2 such that $z_{t_1} = z_{t_2}$ and the average payoff between t_1 and t_2 is good in the sense that : $\gamma_{t_1,t_2}(s) \geq v(z) - 2\varepsilon$. It is then possible to repeat the cycle between t_1 and t_2 and obtain the existence of a stationary ("cyclic") 2ε-optimal play in $\Gamma(z)$.

Case 2 : Assume that there exists z' in Z such that $\{t \geq 0, z_t = z'\}$ is infinite : the play goes through z' infinitely often. Then necessarily case 1 holds.

Case 3 : Assume finally that case 1 does not hold. For every state z', the play s goes through z' a finite number of times, and the average payoff between two stages when z' occurs (whenever these stages exist) is low.

We "shorten" s as much as possible. Put : $y_0 = z_0$, $i_1 = \max\{t \geq 0, z_t = z_0\}$, $y_1 = z_{i_1+1}$, $i_2 = \max\{t \geq 0, z_t = y_1\}$, and by induction for each k, $y_k = z_{i_k+1}$ and $i_{k+1} = \max\{t \geq 0, z_t = y_k\}$, so that $z_{i_{k+1}} = y_k = z_{i_k+1}$. The play $s' = (y_t)_{t\geq 0}$ can be played at z. Since all y_t are distinct, it is a stationary play at z. Regarding payoffs, going from s to s' we removed average payoffs of the type $\gamma_{t_1,t_2}(s)$, where $z_{t_1} = z_{t_2}$. Since we are not in case 1, each of these payoffs is less than $v(z) - 2\varepsilon$, so going from s to s' we increased the average payoffs and we have : $\forall T \geq T_0$, $\gamma_T(s') \geq v(z) - \varepsilon$. s' is an ε-optimal play at z, and this concludes the proof.

Notice that we did *not* obtain the existence of a mapping f from Z to Z such that for every initial state z, the play $(f^t(z))_{t\geq 1}$ (where f^t is f iterated t times) is ε-optimal at z. In our proof, the mapping f depends on the initial state. □

The hypotheses of theorem 3.6 depend on the auxiliary functions $(w_{m,n})$. We now apply this theorem to obtain an existence result with hypotheses directly expressed in terms of the basic data (Z, F, r). We will use the following definitions. Let (Z, d) be a metric space.

Definition 5.2. *A correspondence F from Z to itself is non expansive if :*

$$\forall z \in Z, \forall z' \in Z, \forall z_1 \in F(z), \exists z_1' \in F(z') \text{ s.t. } d(z_1, z_1') \leq d(z, z').$$

If A and B are compact subsets of Z, we denote by $d(A, B)$ the Hausdorff distance between A and $B : d(A, B) = \text{Max}\{\sup_{a\in A}\inf_{b\in B}d(a, b),\ \sup_{b\in B}\inf_{a\in A}d(a, b)\}$. Assume for example now that F is a correspondence from Z to Z with compact values. Then F is non expansive if and only if it is 1-Lipschitz : $d(F(z), F(z')) \leq d(z, z')$ for all (z, z') in Z^2.

It is surprising to us that the following result did not already appear in the literature.

Theorem 5.3. *Let Z be a non empty set, F be a correspondence from Z to Z with non empty values, and r be a mapping from Z to $[0, 1]$. Assume that Z is endowed with a distance d such that :*

a) (Z, d) is a compact metric space,

b) r is continuous,

c) F is non expansive.

Then we have the conclusions of theorem 3.6. For every initial state z in Z, the MDP $\Gamma(z) = (Z, F, r, z)$ has a uniform value which is :

$$v^*(z) = \underline{v}(z) = v^-(z) = v^+(z) = \sup_{m\geq 0}\inf_{n\geq 1}w_{m,n}(z) = \sup_{m\geq 0}\inf_{n\geq 1}v_{m,n}(z).$$

And the sequence $(v_n)_n$ uniformly converges to v^*.

Proof of theorem 5.3 : We assume that the assumptions of theorem 5.3 hold.

Consider z and z' in Z, and a play $s = (z_t)_{t\geq 1}$ in $S(z)$. We have $z_1 \in F(z)$, and F is non expansive, so there exists $z_1' \in F(z')$ such that $d(z_1, z_1') \leq d(z, z')$. $z_2 \in F(z_1)$, so there exists z_2' in $F(z_1')$ such that $d(z_2, z_2') \leq d(z_1, z_1') \leq d(z, z')$, etc... By induction it is easy to construct a play $(z_t')_t$ in $S(z')$ such that for each t, $d(z_t, z_t') \leq d(z, z')$. We have obtained :

$$\forall (z, z') \in Z^2, \forall s = (z_t)_{t\geq 1} \in S(z), \exists s' = (z_t')_{t\geq 1} \in S(z') \text{ s.t. } \forall t \geq 1, d(z_t, z_t') \leq d(z, z').$$

We now consider payoffs. Z being compact, r is indeed uniformly continuous. We consider the modulus of continuity $\hat{\varepsilon}$ given by : for each $\alpha \geq 0$, $\hat{\varepsilon}(\alpha) = \max_{z,z' \text{s.t.} d(z,z')\leq\alpha}|r(z) - r(z')|$. So $|r(z) - r(z')| \leq \hat{\varepsilon}(d(z, z'))$ for each pair of states z, z', and $\hat{\varepsilon}$ is continuous at 0. Using the previous construction, we obtain that for z and z' in Z, and for s in $S(z)$, there exists s' in $S(z)$ such that : $\forall m \geq 0, \forall n \geq 1$, $|\gamma_{m,n}(s) - \gamma_{m,n}(s')| \leq \hat{\varepsilon}(d(z, z'))$ and $|\nu_{m,n}(s) - \nu_{m,n}(s')| \leq \hat{\varepsilon}(d(z, z'))$. Consequently : $\forall (z, z') \in Z^2, \forall m \geq 0, \forall n \geq 1$,

$$|v_{m,n}(z) - v_{m,n}(z')| \leq \hat{\varepsilon}(d(z, z')) \text{ and } |w_{m,n}(z) - w_{m,n}(z')| \leq \hat{\varepsilon}(d(z, z')).$$

In particular, the family $(w_{m,n})_{m\geq 0,n\geq 1}$ is uniformly continuous. Applying theorem 3.6 gives the result. □

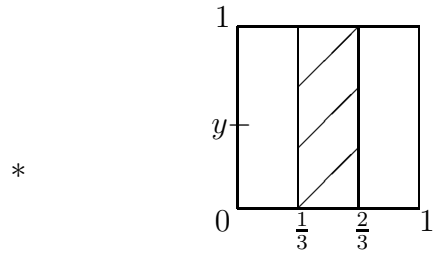
Remark 5.4. *The proof shows that the conclusions of theorem 5.3 also hold if : a′) (Z, d) is a precompact metric space, b′) r is uniformly continuous, and c) F is non expansive.*

We now present illustrating examples.

Example 5.5.

This example may be seen as an adaptation to the compact setup of the example of section 1.4.1 in Sorin (2002). It first illustrates the importance of condition c) (F non expansive) in the hypotheses of theorem 5.3. It also shows that in general one may have : $\sup_{m\geq 0}\inf_{n\geq 1} w_{m,n}(z) \neq \sup_{m\geq 0}\inf_{n\geq 1} v_{m,n}(z)$.

Define the set of states Z as the unit square $[0,1]^2$ plus some isolated point $*$. The transition is defined by $F(*) = \{(0, y), y \in [0,1]\}$, and for (x, y) in $[0,1]^2$, $F(x, y) = \{(\text{Min}\{1, x + y\}, y)\}$. The initial state being $*$, the interpretation is the following. The decision maker only has one decision to make, he has to choose at the first stage a point $(0, y)$, with $y \in [0,1]$. Then the play is determined, and the state evolves horizontally (the second coordinate remains forever y) with arithmetic progression until it reaches the line $x = 1$. y also represents the speed chosen by the decision maker : if $y = 0$, then the state will remain $(0, 0)$ forever. If $y > 0$, the state will evolve horizontally with speed y until reaching the point $(1, y)$.



Let now the reward function r be such that for every $(x, y) \in [0,1]^2$, $r(x, y) = 1$ if $x \in [1/3, 2/3]$, and $r(x, y) = 0$ if $x \notin [1/4, 3/4]$. The payoff is low when x takes extreme values, so intuitively the decision maker would like to maximize the number of stages where the first coordinate of the state is "not too far" from $1/2$.

Endow for example $[0,1]^2$ with the distance d induced by the norm $\|.\|_1$ of $\mathbb{R}^2$, and put $d(*, (x, y)) = 1$ for every x and y in $[0,1]$. (Z, d) is a compact metric space, and it is possible to define $r(x, y)$ for $x \in [1/4, 1/3) \cup (2/3, 3/4]$ so that r is continuous, and even Lipschitz, on Z. One can check that F is 2-Lipschitz, i.e. we have $d(F(z), F(z')) \leq 2d(z, z')$ for each z, z'.

For each $n \geq 2$, we have $v_n(*) \geq 1/2$ because the decision maker can reach the line $x = 2/3$ in exactly n stages by choosing initially $(0, \frac{2}{3(n-1)})$. But for each play s at $*$, we have $\lim_n \gamma_n(s) = 0$, so $\underline{v}(*) = 0$. The uniform value does not exist for $\Gamma(*)$. This shows the importance of condition c) of theorem 5.3. Although F is very smooth, it is not non expansive, so one can not apply theorem 5.3 here.

As a byproduct, we obtain that there is no distance on Z compatible with the Euclidean topology which makes the correspondence F non expansive.

We now show that $\sup_{m\geq 0}\inf_{n\geq 1}w_{m,n}(*) < \sup_{m\geq 0}\inf_{n\geq 1}v_{m,n}(*)$. For every $n>0$, we have $nv_{1,n}(*) = (n+1)v_{n+1}(*)$, because the payoff of stage 1 is necessarily zero. So $\sup_{m\geq 0}\inf_{n\geq 1}v_{m,n}(*) \geq \inf_{n\geq 1}v_{1,n}(*) \geq 1/2$. Fix now $m\geq 0$, and $\varepsilon > 0$. Take n larger than $\frac{3m}{\varepsilon}$, and consider a play $s=(z_t)_{t\geq 1}$ in $S(*)$ such that $\nu_{m,n}(s) > 0$. By definition of $\nu_{m,n}$, we have $\gamma_{m,1}(s) > 0$, so the first coordinate of z_{m+1} is in $[1/4, 3/4]$. If we denote by y the second coordinate of z_1, the first coordinate of z_{m+1} is $m\,y$, so $m\,y \geq 1/4$. But this implies that $4m\,y \geq 1$, so at any stage greater than $4m$ the payoff is zero. Consequently $n\gamma_{m,n}(s) \leq 3m$, and $\gamma_{m,n}(s) \leq \varepsilon$. $\nu_{m,n}(s) \leq \varepsilon$, and this holds for any play s. So $\sup_{m\geq 0}\inf_{n\geq 1}w_{m,n}(*) = 0$.

Remark 5.6. *On the continuity of the payoffs.*

It is assumed in theorem 5.3 that the mapping r is continuous, whereas there is no direct assumption on r in theorem 3.6. The next example shows that, unsurprisingly, assumption b) of theorem 5.3 is necessary. Let Z be the set of complex numbers with modulus 1, and let F be such that $F(\exp(i\theta)) = \{\exp(i(\theta+1))\}$ for every real θ. Z is compact and F is non expansive. Assume that r is such that $\frac{1}{T}\sum_{t=1}^{T} r(\exp(it))$ diverges when T goes to infinity[1] The sequence $(v_n(0))_n$ diverges, hence the limit value of $\Gamma(0)$ does not exist.

Example 5.7. *0-optimal strategies may not exist*

The following example shows that 0-optimal strategies may not exist, even when the assumptions of theorem 5.3 hold and F has compact values. It is the deterministic adaptation of example 1.4.4. in Sorin (2002). Define Z as the simplex $\{z=(p^a,p^b,p^c)\in I\!R^3_+, p^a+p^b+p^c=1\}$. The payoff is $r(p^a,p^b,p^c) = p^b - p^c$, and the transition is defined by : $F(p^a,p^b,p^c) = \{((1-\alpha-\alpha^2)p^a, p^b+\alpha p^a, p^c+\alpha^2 p^a), \alpha \in [0,1/2]\}$. The initial state is $z_0=(1,0,0)$. Notice that along any path, the second coordinate and the third coordinate are non decreasing.

The probabilistic interpretation is the following : there are 3 points a, b and c, and the initial point is a. The payoff is 0 at a, it is +1 at b, and -1 at c. At point a, the decision maker has to choose $\alpha \in [0,1/2]$: then b is reached with probability α, c is reached with probability α^2, and the play stays in a with the remaining probability $1-\alpha-\alpha^2$. When b (resp. c) is reached, the play stays at b (resp. c) forever. So the decision maker starting at point a wants to reach b and to avoid c.

Back to our deterministic setup, we use norm $\|.\|_1$ and obtain that Z is compact, F is non expansive and r is continuous. So theorem 5.3 gives that the uniform value exists.

[1]For example we may have, for every positive integer t, that $r(\exp(it)) = 1$ if there exists an integer l such that $4^l \leq t < 2\times 4^l$, and that $r(\exp(it)) = 0$ otherwise.

Fix ε in $(0, 1/2)$. The decision maker can choose at each stage the same probability ε, i.e. he can choose at each state $z_t = (p_t^a, p_t^b, p_t^c)$ the next z_{t+1} as $((1-\varepsilon-\varepsilon^2)p^a, p^b + \varepsilon p^a, p^c + \varepsilon^2 p^a)$. A simple analysis shows that this sequence of states $s = (z_t)_t$ converges to $(0, \frac{1}{1+\varepsilon}, \frac{\varepsilon}{1+\varepsilon})$. So $\liminf_t \gamma_t(s) = \frac{1-\varepsilon}{1+\varepsilon}$. Finally we obtain that the uniform value at z_0 is 1.

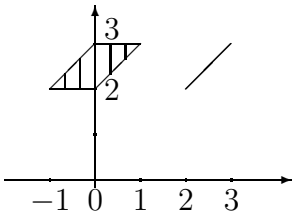
But as soon as the decision maker chooses a positive α at point a, he has a positive probability to be stuck forever with a payoff of -1, so it is clear that no 0-optimal strategy exist here.

Remark 5.8. *(Non) existence of continuous stationary ε-optimal strategies*

Assume that the hypotheses of theorem 5.3 are satisfied. Assume also that Z is a subset of a Banach space and F has closed and convex values, so that F admits a continuous selection (by Michael's theorem).

The uniform value exists, and by remark 5.1 we know that ε-optimal plays can be chosen to be stationary. So if we fix an initial state z, we can find a mapping f from Z to Z such that the play $(f^t(z))_{t\geq 1}$ is ε-optimal at z. Can f be chosen as a continuous selection of Γ ?

A stronger result would be the existence of a continuous f such that *for every initial state* z, the play $(f^t(z))_{t\geq 1}$ is ε-optimal at z. However this existence is not guaranteed, as the following example shows. Define $Z = [-1, 1] \cup [2, 3]$, with the usual distance. Put $F(z) = [2, z+3]$ if $z \in [-1, 0]$, $F(z) = [z+2, 3]$ if $z \in [0, 1]$, and $F(z) = \{z\}$ if $z \in [2, 3]$. Consider the payoff $r(z) = |z - 5/2|$ for each z.



The hypotheses of theorem 5.3 are satisfied. The states in $[2, 3]$ correspond to final ("absorbing" states), and $v(z) = |z - 5/2|$ if $z \in [2, 3]$. If the initial state z is in $[-1, 1]$, one can always choose the final state to be 2 or 3, so that $v(z) = 1/2$. Take now any continuous selection f of Γ. Necessarily $f(-1) = 2$ and $f(1) = 3$, so there exists z in $(-1, 1)$ such that $f(z) = 5/2$. But then the play $s = (f^t(z))_{t\geq 1}$ gives a null payoff at every stage, and for $\varepsilon \in (0, 1/2)$ is not ε-optimal at z.

Remark 5.9. *Discounted payoffs, proofs.*

We prove here the results announced in remark 2.6 about discounted payoffs. Proceeding similarly as in definition 2.3 and claim 2.4 , we say that $\Gamma(z)$ has a d-uniform value if : $(v_\lambda(z))_\lambda$ has a limit $v(z)$ when λ goes to zero, and for every $\varepsilon > 0$, there exists a play s at z such that $\liminf_{\lambda\to 0} \gamma_\lambda(s) \geq v(z) - \varepsilon$. Whereas the definition of uniform value fits Cesaro summations, the definition of d-uniform value fits Abel summations.

An Abel mean can be written as an infinite convex combination of Cesaro means, and it is possible to show that $\limsup_{\lambda\to 0} v_\lambda(z) \leq \limsup_{n\to\infty} v_n(z)$ (see Sorin, 2002, proposition 5.30). A simple example were $\lim_{\lambda\to 0} v_\lambda(z)$ and $\lim_{n\to\infty} v_n(z)$ both exist and differ is the example in section 1.4.1 in Sorin, 2002, which was first present in Lehrer and Sorin 1992. In this 1992 paper, a strong link between the two approaches is also proved : the uniform convergence of (v_n) (resp. (v_λ)) to a mapping v as n goes to infinity (resp. as λ goes to zero) implies the uniform convergence of (v_λ) (resp. (v_n)) to the same v.

Given a sequence $(a_t)_{t\geq 1}$ of nonnegative real numbers, we denote for each $n \geq 1$ and $\lambda \in (0,1]$, by $\bar{a}_n$ the Cesaro mean $\frac{1}{n}\sum_{t=1}^{n} a_t$, and by $\bar{a}_\lambda$ the Abel mean $\sum_{t=1}^{\infty} \lambda(1-\lambda)^{t-1} a_t$. We have the following Abelian theorem (see Araposthakis *et al.*, 1993, Lippman 1969, or Sznajder and Filar, 1992) :

$$\limsup_{n\to\infty} \bar{a}_n \geq \limsup_{\lambda\to 0} \bar{a}_\lambda \geq \liminf_{\lambda\to 0} \bar{a}_\lambda \geq \liminf_{n\to\infty} \bar{a}_n.$$

And, surprisingly enough, the convergence of $\bar{a}_\lambda$, as λ goes to zero, implies the convergence of $\bar{a}_n$, as n goes to infinity, to the same limit (Hardy and Littlewood Theorem, see Lippman 1969).

Lemma 5.10. *If $\Gamma(z)$ has a uniform value $v(z)$, then $\Gamma(z)$ has a d-uniform value which is also $v(z)$.*

Proof : Assume that $\Gamma(z)$ has a uniform value $v(z)$. Then for every $\varepsilon > 0$, there exists a play s at z such that $\liminf_{\lambda\to 0} \gamma_\lambda(s) \geq \liminf_{n\to\infty} \gamma_n(s) \geq v(z) - \varepsilon$. So $\liminf_{\lambda\to 0} v_\lambda(z) \geq v(z)$. But one always has $\limsup_n v_n(z) \geq \limsup_\lambda v_\lambda(z)$. So $v_\lambda(z) \longrightarrow_{\lambda\to 0} v(z)$, and there is a d-uniform value. □

We now give a counter-example to the converse of lemma 5.10. Liggett and Lippman, 1969, showed how to construct a sequence $(a_t)_{t\geq 1}$ with values in $\{0,1\}$ such that $a^* := \limsup_{\lambda\to 0} \bar{a}_\lambda < \limsup_{n\to\infty} \bar{a}_n$. Let[2] us define $Z = I\!N$ and $z_0 = 0$. The transition satisfies : $F(0) = \{0,1\}$, and $F(t) = \{t+1\}$ is a singleton for each positive t. The reward function is defined par $r(0) = a^*$, and for each $t \geq 1$, $r(t) = a_t$. A play in $S(z_0)$ can be identified with the number of positive stages spent in state 0 : there is the play $s(\infty)$ which always remains in state 0, and for each $k \geq 0$ the play $s(k) = (s_t(k))_{t\geq 1}$ which leaves state 0 after stage k, i.e. $s_t(k) = 0$ for $t \leq k$, and $s_t(k) = t - k$ otherwise.

For every λ in $(0,1]$, $\gamma_\lambda(s(\infty)) = a^*$, $\gamma_\lambda(s(0)) = \bar{a}_\lambda$, and for each k, $\gamma_\lambda(s(k))$ is a convex combination between $\gamma_\lambda(s(\infty))$ and $\gamma_\lambda(s(0))$, so $v_\lambda(z_0) = \max\{a^*, \bar{a}_\lambda\}$. So $v_\lambda(z_0)$ converges to a^* as λ goes to zero. Since $s(\infty)$ guarantees a^* in every game, $\Gamma(z_0)$ has a d-uniform value.

For each $n \geq 1$, $v_n(z_0) \geq \gamma_n(s(0)) = \bar{a}_n$, so $\limsup_n v_n(z_0) \geq \limsup_{n\to\infty} \bar{a}_n$. But for every play s at z_0, $\liminf_n \gamma_n(s) \leq \max\{a^*, \liminf_n \bar{a}_n\} = a^*$. The decision maker can guarantee nothing more than a^*, so he can not guarantee $\limsup_n v_n(z_0)$, and $\Gamma(z_0)$ has no uniform value.

[2]We proceed similarly as in Flynn (1974), who showed that a Blackwell optimal play need not be optimal with respect to "Derman's average cost criterion".

6 Applications to probabilistic MDPs

We start with a simple case.

6.1 Probabilistic MDPs with a finite set of states.

Consider a finite set of states K, with an initial probability p_0 on K, a non empty set of actions A, a transition function q from $K \times A$ to the set $\Delta(K)$ of probability distributions on K, and a reward function g from $K \times A$ to $[0,1]$.

This probabilistic MDP is played as follows. An initial state k_1 in K is selected according to p_0 and told to the decision maker, then he selects a_1 in A and receives a payoff of $g(k_1, a_1)$. A new state k_2 is selected according to $q(k_1, a_1)$ and told to the decision maker, etc... A strategy of the decision maker is then a sequence $\sigma = (\sigma_t)_{t\geq 1}$, where for each t, $\sigma_t : (K \times A)^{t-1} \times K \longrightarrow A$ defines the action to be played at stage t. Considering expected average payoffs in the first n stages, the definition of the n-stage value $v_n(p_0)$ naturally adapts to this case. And the notions of limit value and uniform value also adapt here. Write $\Psi(p_0)$ for this probabilistic MDP.

We define an auxiliary deterministic MDP $\Gamma(z_0)$. We view $\Delta(K)$ as the set of vectors $p = (p^k)_k$ in $\mathbb{R}^K_+$ such that $\sum_k p^k = 1$. We introduce :

- a new set of states $Z = \Delta(K) \times [0,1]$,
- a new initial state $z_0 = (p_0, 0)$,
- a new payoff function $r : Z \longrightarrow [0,1]$ such that $r(p,y) = y$ for all (p,y) in Z,
- a transition correspondence F from Z to Z such that for every $z = (p,y)$ in Z,

$$F(z) = \left\{ \left(\sum_{k\in K} p^k q(k,a_k), \sum_{k\in K} p^k g(k,a_k) \right), a_k \in A \; \forall k \in K \right\}.$$

Notice that $F((p,y))$ does not depend on y, hence the value functions in $\Gamma(z)$ only depend on the first component of z. It is easy to see that the value functions of Γ and Ψ are linked as follows : $\forall z = (p,y) \in Z, \forall n \geq 1, \;\; v_n(z) = v_n(p)$. Moreover, anything that can be guaranteed by the decision maker in $\Gamma(p,0)$ can also be guaranteed in $\Psi(p)$. So if we prove that the auxiliary MDP $\Gamma(p_0,0)$ has a uniform value, then $(v_n(p_0))_n$ has a limit that can be guaranteed, up to every $\varepsilon > 0$, in $\Gamma(p_0,0)$, hence also in $\Psi(p_0)$. And we obtain the existence of the uniform value for $\Psi(p_0)$.

It is convenient to set $d((p,y),(p',y')) = \max\{\|p-p'\|_1, |y-y'|\}$. Z is compact and r is continuous. F may have non compact values, but is non expansive so that we can apply theorem 5.3. Consequently, for each p_0, $\Psi(p_0)$ has a uniform value, and we have obtained the following result.

Theorem 6.1. *Any probabilistic MDP with finite set of states has a uniform value.*

We could not find theorem 6.1 in the literature. The case where A is finite is well known since the seminal work of Blackwell (1962), who showed the existence of Blackwell optimal plays. If A is compact and both q and g are continuous in a, the uniform value exists by Corollary 5.26 p. 110 in Sorin, 2002. In this case, more properties on (ε)-optimal strategies have been obtained.

6.2 Probabilistic MDPs with partial observation.

We now consider a more general model where after each stage, the decision maker does not perfectly observe the state. We still have a finite set of states K, an initial probability p_0 on K, a non empty set of actions A, but we also have a non empty set of signals S. The transition q now goes from $K \times A$ to $\Delta_f(S \times K)$, the set of probabilities with finite support on $S \times K$, and the reward function g still goes from $K \times A$ to $[0, 1]$.

This MDP $\Psi(p_0)$ is played by a decision maker knowing K, p_0, A, S, q and g and the following description. An initial state k_1 in K is selected according to p_0 and is not told to the decision maker. At every stage t the decision maker selects an action $a_t \in A$, and has a (unobserved) payoff $g(k_t, a_t)$. Then a pair (s_t, k_{t+1}) is selected according to $q(k_t, a_t)$, and s_t is told to the decision maker. The new state is k_{t+1}, and the play goes to stage $t + 1$.

The existence of the uniform value was proved in Rosenberg *et al.* in the case where A and S are finite sets[3]. We show here how to apply theorem 3.6 to this setup, and generalize the mentioned result of Rosenberg *et al.* to the case of arbitrary sets of actions and signals.

A pure strategy of the decision maker is then a sequence $\sigma = (\sigma_t)_{t\geq 1}$, where for each t, $\sigma_t : (A \times S)^{t-1} \longrightarrow A$ defines the action to be played at stage t. More general strategies are behavioral strategies, which are sequences $\sigma = (\sigma_t)_{t\geq 1}$, where for each t, $\sigma_t : (A \times S)^{t-1} \longrightarrow \Delta_f(A)$ and $\Delta_f(A)$ is the set of probabilities with finite support on A. In $\Psi(p_0)$ we assume that players use behavior strategies. Any strategy induces, together with p_0, a probability distribution over $(K \times A \times S)^\infty$, and we can define expected average payoffs and n-stage values $v_n(p_0)$. These n-stage values can be obtained with pure strategies. However, one has to be careful when dealing with an infinite number of stages : in general it may not be true that something which can be guaranteed by the decision maker in $\Psi(p_0)$, i.e.. with behavior strategies, can also be guaranteed by the decision maker with pure strategies. We will prove here the existence of the uniform value in $\Psi(p_0)$, and thus obtain :

Theorem 6.2. *If the set of states is finite, a probabilistic MDP with partial observation, played with behavioral strategies, has a uniform value.*

[3]These authors also considered the case of a compact action set, with some continuity on g and q, see comment 5 p. 1192.

Proof : As in the previous model, we view $\Delta(K)$ as the set of vectors $p = (p^k)_k$ in $\mathbb{R}^K_+$ such that $\sum_k p^k = 1$. We write $X = \Delta(K)$, and use $\|.\|_1$ on X. Assume that the state of some stage has been selected according to p in X and the decision maker plays some action a in A. This defines a probability on the future belief of the decision maker on the state of the next stage. It is a probability with finite support because we have a belief in X for each possible signal S, and we denote this probability on X by $\hat{q}(p, a)$. To introduce a deterministic MDP we need a larger space than X.

We define $\Delta(X)$ as the set of Borel probabilities over X, and endow $\Delta(X)$ with the weak-* topology. $\Delta(X)$ is now compact and the set $\Delta_f(X)$ of probabilities on X with finite support is a dense subset of $\Delta(X)$ (see for example, Malliavin, 1995, p.99). Moreover, the topology on $\Delta(X)$ can be metrized by the (Fortet-Mourier-)Wasserstein distance, defined by :

$$\forall u \in \Delta(X), \forall v \in \Delta(X), \quad d(u, v) = \sup_{f \in E_1} |u(f) - v(f)|,$$

where : E_1 is the set of 1-Lipschitz functions from X to $\mathbb{R}$, and $u(f) = \int_{p \in X} f(p) du(p)$. One can check that this distance also has the nice following properties :[4]

1) for p and q in X, the distance between the Dirac measures δ_p and δ_q is $\|p - q\|_1$.

2) For every continuous mapping from X to the reals, let us denote by $\tilde{f}$ the affine extension of f to $\Delta(X)$. We have $\tilde{f}(u) = u(f)$ for each u. Then for each $C \geq 0$, we obtain the equivalence : f is C-Lipschitz if and only if $\tilde{f}$ is C-Lipschitz.

We will need to consider a whole class of value functions. Let $\theta = \sum_{t \geq 1} \theta_t \delta_t$ be in $\Delta_f(\mathbb{N}^*)$, i.e. θ is a probability with finite support over positive integers. For p in X and any behavior strategy σ, we define the payoff : $\gamma^p_{[\theta]}(\sigma) = \mathbb{E}_{\mathbb{P}_{p,\sigma}} \left(\sum_{t=1}^{\infty} \theta_t \, g(k_t, a_t) \right)$, and the value : $v_{[\theta]}(p) = \sup_\sigma \gamma^p_{[\theta]}(\sigma)$. If $\theta = 1/n \sum_{t=1}^n \delta_t$, $v_{[\theta]}(p)$ is nothing but $v_n(p)$. $v_{[\theta]}$ is a 1-Lipschitz function so its affine extension $\tilde{v}_{[\theta]}$ also is. A standard recursive formula can be written : if we write θ^+ for the law of $t^* - 1$ given that t^* (selected according to θ) is greater than 1, we get for each θ and p : $v_{[\theta]}(p) = \sup_{a \in A} \left(\theta_1 \sum_k p^k g(k, a) + (1 - \theta_1) \tilde{v}_{[\theta^+]}(\hat{q}(p, a)) \right)$.

We now define a deterministic MDP $\Gamma(z_0)$. An element u in $\Delta_f(X)$ is written $u = \sum_{p \in X} u(p) \delta_p$, and similarly an element v in $\Delta_f(A)$ is written $v = \sum_{a \in A} v(a) \delta_a$. Notice that if $p \neq q$, then $1/2\, \delta_p + 1/2\, \delta_q$ is different from $\delta_{1/2\, p + 1/2\, q}$. We introduce :

- a new set of states $Z = \Delta_f(X) \times [0, 1]$,
- a new initial state $z_0 = (\delta_{p_0}, 0)$,
- a new payoff function $r : Z \longrightarrow [0, 1]$ such that $r(u, y) = y$ for all (u, y) in Z,

[4]Notice that if $d(k, k') = 2$ for any distinct states in K, then $\sup_{f : K \to \mathbb{R}, 1-Lip} |\sum_k p^k f(k) - \sum_k q^k f(k)| = \|p - q\|_1$ for every p and q in $\Delta(K)$.

• a transition correspondence F from Z to Z such that for every $z=(u,y)$ in Z :

$$F(z)=\{(H(u,f),R(u,f)),f:X\longrightarrow\Delta_f(A)\},$$

where $H(u,f)=\sum_{p\in X}u(p)\left(\sum_{a\in A}f(p)(a)\delta_{\hat{q}(p,a)}\right)\in\Delta_f(X)$,
and $R(u,f)=\sum_{p\in X}u(p)\left(\sum_{k\in K,a\in A}p^k f(p)(a)g(k,a)\right)$.

$\Gamma(z_0)$ is a well defined deterministic MDP. $F(u,y)$ does not depend on y, so the value functions in $\Gamma(z)$ only depend on the first coordinate of z. For every $\theta=\sum_{t\geq 1}\theta_t\delta_t$ in $\Delta_f(I\!N^*)$ and play $s=(z_t)_{t\geq 1}$, we define the payoff $\gamma_{[\theta]}(s)=\sum_{t=1}^{\infty}\theta_t r(z_t)$, and the value : $v_{[\theta]}(z)=\sup_{s\in S(z)}\gamma_{[\theta]}(s)$. If $\theta=1/n\ \sum_{t=m+1}^{n}\delta_t$, $\gamma_{[\theta]}(s)$ is nothing but $\gamma_{m,n}(s)$, and $v_{[\theta]}(z)$ is nothing but $v_{m,n}(z)$, see definitions 3.1 and 3.2. $\gamma_{[t]}(s)$ is just the payoff of stage t, i.e. $r(z_t)$. The recursive formula now is : $v_{[\theta]}((u,y))=\sup_{f:X\longrightarrow\Delta_f(A)}(\theta_1 R(u,f)+(1-\theta_1)v_{[\theta^+]}(H(u,f),0))$, and the supremum can be taken on deterministic mappings $f:X\longrightarrow A$. Consequently, the value functions of the MDPs are linked as follows : $\forall z=(u,y)\in Z, v_{[\theta]}(z)=\tilde{v}_{[\theta]}(u)$. Moreover, anything which can be guaranteed by the decision maker in $\Gamma(z_0)$ can be guaranteed in the original MDP $\Psi(p_0)$. So the existence of the uniform value in $\Gamma(z_0)$ will imply the existence of the uniform value in $\Psi(p_0)$.

We put $d((u,y),(u',y'))=\max\{d(u,u'),|y-y'|\}$. Since $\Delta_f(X)$ is dense in $\Delta(X)$ for the Wasserstein distance, Z is a precompact metric space. By theorem 3.6, if we show that the family $(w_{m,n})_{m\geq 0,n\geq 1}$ is uniformly equicontinuous, we will be done. Notice already that since $\tilde{v}_{[\theta]}$ is a 1-Lipschitz function of u, $v_{[\theta]}$ is a 1-Lipschitz function of z.

Fix now z in Z, $m\geq 0$ and $n\geq 1$. We define an auxiliary zero-sum game $\mathcal{A}(m,n,z)$ as follows : player 1's strategy set is $S(z)$, player 2's strategy set is $\Delta(\{1,...,n\})$, and the payoff for player 1 is given by : $l(s,\theta)=\sum_{t=1}^{n}\theta_t\gamma_{m,t}(s)$. We will apply a minmax theorem to $\mathcal{A}(m,n,z)$, in order to obtain : $\sup_s\inf_\theta l(s,\theta)=\inf_\theta\sup_s l(s,\theta)$. We can already notice that $\sup_s\inf_\theta l(s,\theta)=\sup_{s\in S(z)}\inf_{t\in\{1,...,n\}}\gamma_{m,t}(s)=w_{m,n}(z)$. $\Delta(\{1,...,n\})$ is convex compact and l is affine continuous in θ. We will show that $S(z)$ is a convex subset of Z, and first prove that F is an affine correspondence.

Lemma 6.3. *For every z' and z'' in Z, and $\lambda\in[0,1]$, $F(\lambda z'+(1-\lambda)z'')=\lambda F(z')+(1-\lambda)F(z'')$.*

Proof : Write $z'=(u',y')$, $z''=(u'',y'')$ and $z=(u,y)=\lambda z'+(1-\lambda)z''$. We have $u(p)=\lambda u'(p)+(1-\lambda)u''(p)$ for each p. It is easy to see that $F(z)\subset\lambda F(z')+(1-\lambda)F(z'')$, so we just prove the reverse inclusion. Let $z'_1=(H(u',f'),R(u',f'))$ be in $F(z')$ and $z''_1=(H(u'',f''),R(u'',f''))$ be in $F(z'')$, with f' and f'' mappings from X to $\Delta_f(A)$. Using here the convexity of $\Delta_f(A)$, we simply define for each p in X, $f(p)=\frac{\lambda u'(p)}{u(p)}f'(p)+\frac{(1-\lambda)u''(p)}{u(p)}f''(p)$. We have for each p, $R(\delta_p,f)=\frac{\lambda u'(p)}{u(p)}R(\delta_p,f')+\frac{(1-\lambda)u''(p)}{u(p)}R(\delta_p,f'')$. So $R(u,f)=\lambda R(u',f')+(1-\lambda)R(u'',f'')$.

Similarly the transitions satisfy : $H(u,f) = \lambda H(u',f') + (1-\lambda)H(u'',f'')$. And we obtain that $\lambda z_1' + (1-\lambda)z_1'' = (H(u,f), R(u,f)) \in F(z)$. □

As a consequence, the graph of F is convex, and this implies the convexity of the sets of plays. So we have obtained the following result.

Corollary 6.4. *The set of plays $S(z)$ is a convex subset of Z^∞.*

Looking at the definition of the payoff function r, we now obtain that l is affine in s. Consequently, we can apply a standard minmax theorem (see e.g. Sorin 2002 proposition A8 p.157) to obtain the existence of the value in $\mathcal{A}(m,n,z)$. So $w_{m,n}(z) = \inf_{\theta\in\Delta(\{1,...,n\})}\sup_{s\in S(z)} \sum_{t=1}^n \theta_t \gamma_{m,t}(s)$. But $\sup_{s\in S(z)} \sum_{t=1}^n \theta_t \gamma_{m,t}(s)$ is equal to $v_{[\theta^{m,n}]}(z)$, where $\theta^{m,n}$ is the probability on $\{1,...,m+n\}$ such that $\theta_s^{m,n} = 0$ if $s \leq m$, and $\theta_s^{m,n} = \sum_{t=s-m}^n \frac{\theta_t}{t}$ if $m < s \leq n+m$. The precise value of $\theta^{m,n}$ does not matter much, but the point is to write : $w_{m,n}(z) = \inf_{\theta\in\Delta(\{1,...,n\})} v_{[\theta^{m,n}]}(z)$. So $w_{m,n}$ is 1-Lipschitz as an infimum of 1-Lipschitz mappings. The family $(w_{m,n})_{m,n}$ is uniformly equicontinuous, and the proof of theorem 6.2 is complete. □

Remark 6.5. *The following question, mentioned in Rosenberg et al., is natural and still open. Does there exist pure ε-optimal strategies ?*

References.

Araposthathis A., Borkar V., Fernández-Gaucherand E., Ghosh M. and S. Marcus (1993) : Discrete-time controlled Markov Processes with average cost criterion : a survey. SIAM Journal of Control and Optimization, 31, 282-344.

Blackwell, D. (1962) : Discrete dynamic programming. The Annals of Mathematical Statistics, 33, 719-726.

Hernández-Lerma, O. and J.B. Lasserre (1996) : Discrete-Time Markov Control Processes. Basic Optimality Criteria. Chapter 5 : Long-Run Average-Cost Problems. Applications of Mathematics, Springer.

Hordijk, A. and A. Yushkevich (2002) : Blackwell Optimality. Handbook of Markov Decision Processes, Chapter 8, 231-268. Kluwer Academic Publishers.

Filar, A. and Sznajder, R. (1992) : Some comments on a theorem of Hardy and Littlewood. Journal of Optimization Theory and Applications, 75, 201-208.

Flynn, J. (1974) : Averaging vs. Discounting in Dynamic Programming : a Counterexample. The Annals of Statistics, 2, 411-413.

Lehrer, E. and D. Monderer (1994) : Discounting versus Averaging in Dynamic Programming. Games and Economic Behavior, 6, 97-113.

Lehrer, E. and S. Sorin (1992) : A uniform Tauberian Theorem in Dynamic Programming. Mathematics of Operations Research, 17, 303-307.

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

Lippman, S. (1969) : Criterion Equivalence in Discrete Dynamic Programming. Operations Research 17, 920-923.

Malliavin, P. (1995) : Integration and probability. Springer-Verlag, New-York.

Mertens, J-F. and A. Neyman (1981) : Stochastic games. International Journal of Game Theory, 1, 39-64.

Monderer, D. and S. Sorin (1993) : Asymptotic properties in Dynamic Programming. International Journal of Game Theory, 22, 1-11.

Renault, J. (2006) : The value of Markov chain games with lack of information on one side. Mathematics of Operations Research, 31, 490-512.

Renault, J. (2007) : The value of Repeated Games with an informed controller. Preprint, Ceremade : http ://www.ceremade.dauphine.fr/preprints/CMD/2007-2.pdf

Rosenberg, D., Solan, E. and N. Vieille (2002) : Blackwell Optimality in Markov Decision Processes with Partial Observation. The Annals of Statisitics, 30, 1178-1193.

Sorin, S. (2002) : A first course on Zero-Sum repeated games. Mathématiques et Applications, SMAI, Springer.