Commutative Stochastic Games

Xavier Venel

May 7, 2010

Abstract

We are interested in stochastic games with finite sets of actions where the transitions commute. The Big Match and more generally absorbing games can be formulated in this model. When there is only one player, we show that the existence of a uniform value in pure strategies implies the existence of 0-optimal strategies. For stochastic games we prove the existence of the uniform value when the set of states is finite and players observe past actions but not the state. They reduce to a specific class of zero-sum stochastic games on $\mathbb{R}^n$ which we solve by using the theorem of Mertens Neyman [MN81]. The same proof extends to the non zero-sum case if we use the result of Vieille [Vie00a][Vie00b].

MSC 2000 subject classification: 91A15, 91A05 .

Key-words: Stochastic Game, Commutation, Uniform Value, Markov Decision Process.

1 Introduction

We are interested in two-players zero-sum stochastic games where the transitions commute. In this model, given a sequence of decisions, the order of the decisions is irrelevant to know the reached state. The exploitation of a mineral resource such as oil or gold is an example of an economic problem fitting this assumption. It is enough to remember how much of the resource has been exploited in the past to define the remaining quantity. Another example is a competition between firms which have to sell some stocks. The state variable is the vector of stocks of all firms and at each stage the firms decide how much they want to sell. The rewards depend on the quantities sold on the market and the state depends on the different past decisions and not on their order.

In the standard cases of Markov decision processes (MDP) where the set of states and the sets of actions are finite, Blackwell [Bla62] proved the existence of the uniform value. This result was extended to MDP with partial observation by Rosenberg, Solan and Vieille [RSV02] and Renault [Ren09] gives sufficient conditions to the existence of a uniform value if the set of states is precompact. In Theorem 1, we state that in commutative MDP when the uniform value in pure strategies exists there exists a strategy which guarantees exactly the value. Moreover under topological assumptions similar to Renault [Ren09], it is possible to build such a strategy without randomization. This result applies especially to MDP in the dark where the transitions commute.

For stochastic game the existence of the uniform value in the finite case was proven by Mertens and Neyman [MN81] when the players observe everything. If the players do not observe past actions the uniform value may fail to exist. When the set of states is precompact Renault [Ren07] gives an existence result for a two players game on a compact subset of a normed vector space where one player controls the transition. The case of two-players stochastic games with a compact set of states is still open. In this paper we are interested in a model where the set of states is a compact subset of $\mathbb{R}^n$ and the transitions are applications which are non expansive for the norm $\|.\|_1$. This is satisfied for example if the set of states is a simplex of probabilities over a finite set and the transitions are defined via a Markov chain. Theorem 2 states that under the assumption of commutation there exists a uniform value and we deduce the existence of the uniform value for commutative stochastic games with finite sets of states and actions where the players observe past actions but not the state. This signalling structure is based on the study of repeated games with symmetric incomplete information. In these models Kohlberg and Zamir[Koh74] and Forges [For82] proved the existence of a uniform value. Neymann and Sorin extended their results to the non zero-sum case [NS98] and Geitner [Gei02] to a model with stochastic games.

In section 2 we introduce classic definitions on stochastic games and the formal definition of commutation. In section 3 we state the results. The section 4 is dedicated to the proof of Theorem 1 on one-player game and the section 5 focus on the proof of Theorem 2 on stochastic games. In the last section we give some extensions.

2 The Model

2.1 Definition

If Z is a non empty set, we denote by $\Delta_f(Z)$ the set of probabilities on Z with finite support. When Z is finite, we denote it by $\Delta(Z)$.

We will consider a model of zero-sum stochastic game $\Gamma(p_1) = (Z, I, J, q, r, p_1)$ given by: a non empty set of states Z, two finite non empty sets of actions I and J, a transition function $q : Z \times I \times J \to \Delta_f(Z)$, a reward function $r : Z \times I \times J \to [0, 1]$ and an initial probability distribution $p_1 \in \Delta_f(Z)$. We say that the transition function q is deterministic when the image is a Dirac measure.

The interpretation is the following. An initial state z_1 is chosen according to p_1 and announced to the players. At each stage $n \geqslant 1$, player 1 and player 2 choose simultaneously an action, $i_n \in I$ and $j_n \in J$. player 1 receives the payoff $r(z_n, i_n, j_n)$, player 2 receives the opposite $-r(z_n, i_n, j_n)$ and the game moves to a new state z_{n+1} chosen according to the probability distribution $q(.|z_n, i_n, j_n)$. Then both players observe the couple of actions (i_n, j_n), the state z_{n+1} and the stage goes to $n + 1$.

Hence at stage n the set of past histories for both players is $H_n = (Z \times I \times J)^{n-1} \times Z$. A strategy for player 1 is an element $(\sigma_n)_{n\geqslant 1}$, where for each n, σ_n is a mapping from H_n to $\Delta(I)$ giving the random action played by player 1 at stage n given the past history. A strategy for player 2 is an element $\tau = (\tau_n)_{n\geqslant 1}$, where for each n, τ_n is a mapping from H_n to $\Delta(J)$. Denote by Σ and $\mathcal{T}$ their respective sets of strategies. If a strategy is such that for each integer n the image is a Dirac measure, the strategy is said to be pure.

Fix a strategy profile (σ, τ) and an initial probability p_1, it induces a probability distribution on the set of finite histories of length n, $(Z \times I \times J)^{n-1} \times Z$ for all integer n. It is standard that this probability distribution can be uniquely extended to the set of infinite plays $(Z \times I \times J)^{\infty}$. For each positive N, we define the average expected payoff for player 1 after N stages

$$\gamma_N(p_1, \sigma, \tau) = E_{p_1,\sigma,\tau}\left(\frac{1}{N}\sum_{n=1}^{N} r(z_n, i_n, j_n)\right)$$

We define also the average payoff between two stages M and N.

$$\gamma_{M,N}(p_1, \sigma, \tau) = E_{p_1,\sigma,\tau}\left(\frac{1}{N-M+1}\sum_{n=M}^{N} r(z_n, i_n, j_n)\right)$$

To study the infinite game $\Gamma(p_1)$ we focus on the notion of uniform value and of ϵ-optimal strategies.

Definition 1. *Let v be a real number,*

- *player 1 can guarantee v in $\Gamma(p_1)$ if for all $\epsilon > 0$ there exists a strategy σ of player 1 and $N \in \mathbb{N}$, such that*

$$\forall n \geqslant N \ \forall \tau \in \mathcal{T} \ \gamma_n(p_1, \sigma, \tau) \geqslant v - \epsilon$$

  We say that such a strategy σ guarantees $v - \epsilon$ in $\Gamma(p_1)$.

- *player 2 can guarantee v in $\Gamma(p_1)$ if for all $\epsilon > 0$ there exists a strategy τ of player 2 and $N \in \mathbb{N}$, such that*

$$\forall n \geqslant N \ \forall \sigma \in \Sigma \ \gamma_n(p_1, \sigma, \tau) \leqslant v + \epsilon$$

  We say that such a strategy τ guarantees $v + \epsilon$ in $\Gamma(p_1)$.

- *If both players can guarantee v, v is the uniform value of the game and we denote it by $v(p_1)$.*

When the uniform value exists, given $\epsilon \geqslant 0$,

- *a strategy σ of player 1 is ϵ-optimal if*

$$\liminf_n \inf_{\tau \in \mathcal{T}} \gamma_n(p_1, \sigma, \tau) \geqslant v(p_1) - \epsilon$$

- *a strategy τ of player 2 is ϵ-optimal if*

$$\limsup_n \sup_{\sigma \in \Sigma} \gamma_n(p_1, \sigma, \tau) \leqslant v(p_1) + \epsilon$$

Notice that the uniform value exists if and only if there exists v such that players 1 and 2 have for each $\epsilon > 0$ strategies which guarantee respectively $v - \epsilon$ and $v + \epsilon$. We denote by max min the maximum of the values that player 1 can guarantee and min max the minimum of the values that player 2 can guarantee. It is easy to see that the game has a uniform value if both are equal.

We will also be interested in another type of game $\Gamma^{sb}(z_1) = (K, I, J, q, r, p_1, \Sigma_{sb}, \mathcal{T}_{sb})$ where the set of strategies of player 1 is restricted to $\Sigma_{sb} \subset \Sigma$ and the set of strategies of player 2 is restricted to $\mathcal{T}_{sb} \subset \mathcal{T}$. A strategy $\sigma \in \Sigma$ is in Σ_{sb} if for all $h_n = (z_1, i_1, j_1, z_2, ..., j_{n-1}, z_n)$, $h'_n = (z'_1, i'_1, j'_1, z'_2, ..., j'_{n-1}, z'_n) \in (Z \times I \times J)^{n-1} \times Z$ such that $i_l = i'_l$ and $j_l = j'_l$ for all $l \in \{1, .., n-1\}$ then $\sigma_n(h_n) = \sigma_n(h'_n)$. $\mathcal{T}_{sb}$ is defined similarly. The interpretation is that the players at each stage observe past actions but not the state. When K is finite, this game is equivalent to a game $\Gamma(p_1) = (Z, I, J, q, r, p_1)$ where $Z = \Delta(K) \subset \mathbb{R}^{\sharp K}$, I and J are the same, q and r are the linear extension of q and r to Z and p_1 is the Dirac mass in z_1. So this game reduces to the previous model with a deterministic transition non expansive for the norm 1 of $\mathbb{R}^{\sharp K}$.

2.2 Commutation assumption

Let $\Gamma = (Z, I, J, q, r)$ be a stochastic game, we define the applications q and r on $\Delta_f(Z)$ by their linear extensions.

Definition 2. *The transition function q commutes on Z if for all $z \in Z$, for all $i, i' \in I$and $j, j' \in J$,*

$$q(q(z, i', j'), i, j) = q(q(z, i, j), i', j')$$

It means that the state is the same if the couple of actions (i, j) is played before (i', j') or if (i, j) is played after (i', j'). If no player can influence the trajectory, the commutation assumption is automatically fulfilled.

Example 1. *Let Z be the circle of center ω and $\theta : I \times J \to \Delta_f([0, 2\pi])$. The transition function q is defined by $q(z, i, j) = \sum_\rho \theta(\rho) r(\rho, z)$ where $r(\rho, .)$ is the rotation of angle ρ and center ω.*

Definition 3. *The transition function q weakly commutes on Z if for all $z \in Z$, for all $i, i' \in I$and $j, j' \in J$, there exists $i'' \in I$ and $j'' \in J$,*

$$q(q(z, i', j'), i, j) = q(q(z, i, j), i'', j'')$$

The second definition is a weaker assumption. The interpretation is that if a couple of action is played on a trajectory, this couple could have been played before. In this definition the two couples of actions do not play symmetric roles as in the first one.

Example 2. *Let $Z = \mathbb{N}^2$ be the set of states, $I = \{(-1, 0), (0, 4), (1, 0)\}$, $J = \{(2, 0), (0, 1)\}$ the sets of actions and $q(z, i, j) = z + i + j$ the transition function. By commutation of the addition the transitions commute. Consider $Z' = Z$, $I' = I$ and $J' = J \cup \{\alpha\}$ with the same transition as before if $(i', j') \in I \times J$ and $q(z, ., \alpha) = (0, 0)$. This new game is not commutative but still is weakly commutative.*

Furthermore the classical class of absorbing games introduced by Kohlberg [Koh74] can be viewed as a subclass of commutative games. Recall that an absorbing game is a stochastic game $\Gamma = (\{\alpha\} \cup Z, I, J, q, r)$ where for each $z \in Z$, z is absorbing and the payoff in z does not depend on the actions. The state α is the only one where the players have an influence on the payoff and on the trajectory.

Proposition 1. *Let Γ be an absorbing game, there exists a commutative game Γ' such that the set of states of Γ is included in the set of states of Γ' and for all these states for all $n \in N$, $v_n(z) = v'_n(z)$. Moreover if there exists a strategy σ' which guarantees w in $\Gamma'(z)$ then there exists a strategy σ which guarantees w in $\Gamma(z)$.*

Let $\Gamma = (\{\alpha\} \cup Z, I, J, q, r)$ be an absorbing game and let build a commutative game $\Gamma' = (Z', I', J', q', r')$. We assume that I and J are disjoints. We define $Z' = Z \cup \{\alpha, \omega, z_i, z_j, z_{i,j} \mid \forall i \in I\ \forall j \in J\}$ and $I' = I$, $J' = J$. We have added $1 + \sharp I + \sharp J + \sharp (I \times J)$ new states to the game. In Γ' the states in Z are absorbing and with the same payoff as in Γ. Since they are absorbing the commutation assumption is satisfied and their values are the same in both games. Moreover we will build the transition such that there is no transition which leads to them so we will forget them for the rest of the proof.

We define $g(i,j) = 1 - q(\alpha, i, j)(\alpha)$ the probability of absorption if the couple of actions (i,j) is played and $q(\alpha, i, j|Z)$ the conditional probability on Z if there has been absorption. Let the payoff function be defined by,

$$\begin{aligned} \forall i, i' \in I,\ j, j' \in J \qquad r'(\alpha, i, j) &= r(\alpha, i, j) \\ r'(z_{i',j'}, i, j) &= E_{q(\alpha, i', j'|Z)}(r(z)) \\ r'(z_{i'}, i, j) &= 1 \\ r'(z_{j'}, i, j) &= 0 \\ r'(\omega, i, j) &= 1/2 \end{aligned}$$

and the transition by $q'(z, i, j) = (1 - g(i,j))\delta_z + g(i,j)\delta_{s(z,i,j)}$ for all $z \in Z'$, $i \in I$ and $j \in J$ with s given by the following formula :

$$\begin{aligned} \forall i, i' \in I,\ j, j' \in J \qquad s(\alpha, i, j) &= z_{i,j} \\ s(z_{i',j'}, i, j) &= \begin{cases} z_{i',j'} & \text{if } i = i' \text{ and } j = j' \\ z_{i'} & \text{if } i = i' \text{ and } j \neq j' \\ z_{j'} & \text{if } i \neq i' \text{ and } j = j' \\ \omega & \text{if } i \neq i' \text{ and } j \neq j' \end{cases} \\ s(z_{i'}, i, j) &= \begin{cases} z_{i'} & \text{if } i = i' \\ \omega & \text{if } i \neq i' \end{cases} \\ s(z_{j'}, i, j) &= \begin{cases} z_{j'} & \text{if } j = j' \\ \omega & \text{if } j \neq j' \end{cases} \\ s(\omega, i, j) &= \omega \end{aligned}$$

Let $(i', j') \in I \times J$, the transition is designed such that $z_{i',j'}$ is invariant by the couple (i', j'). Moreover if player 1 deviates the play stays in $z_{i',j'}$ with probability $1 - g(i', j')$ and goes to $z_{j'}$ with probability $g(i', j')$. The state $z_{j'}$ is controlled by player 2 and the payoff is 0 so it is a punishing state for player 1. The situation is symmetric for player 2 and if both deviate the trajectory absorbed in ω.

Let show that the function s commutes. Since w is absorbing there is nothing to check in this state. Let $i' \in I$, the game $\Gamma(z_{i'})$ is controlled by player 1 and his actions come down to two actions i' and something else. When the same action is played twice the assumption is automatically satisfied so it is enough to check the cases where he plays i' at one stage and $i \neq i'$ at the other. Whenever this action is played there is absorption in ω so the commutation assumption is fulfilled for all states $z_{i'}$, $i' \in I$ and similarly for the states $z_{j'}$, $j' \in J$.

Let $(i', j') \in I \times J$, in the state $z_{i',j'}$ the situation reduces as before for each player to two actions $(i', other)$ and $(j', other)$. We check the different cases. Assume that one of the couples is $(other, other)$. When it is played first the stage goes directly to w. Otherwise the state after one step is in $\{z_{i'}, z_{j'}, z_{i',j'}, w\}$ and the couple is still of the form $(other, other)$ and leads to w so commutation is fulfilled. If one couple is (i', j') then on $\{z_{i'}, z_{j'}, z_{i',j'}, w\}$, the transition is the identity and it commutes with everything. There is left to check if the couples are of one of the following forms $(i', other)(i', other)$, $(other, j')(other, j')$, $(i', other)(other, j')$. In the two first the situation is the same as if the same couple of action is played twice. In the last one the transitions lead to w trough $k_{i'}$ if the order is $(i', other)$,$(other, j')$ and through $k_{j'}$ when the order is $(other, j')$,$(i', other)$. Thus s commutes.

We deduce that q commutes. Let $z \in Z$, $i, i' \in I$ and $j, j' \in J$ then

$$\begin{aligned} q(q(z,i,j),i',j') = {} & (1-g(i,j)(1-g(i',j'))\delta_z + (1-g(i,j)g(i',j')\delta_{s(z,i,j)} \\ & + (1-g(i',j'))g(i,j)\delta_{s(z,i',j')} + g(i,j)g(i',j')\delta_{s(s(z,i,j),i',j')} \end{aligned}$$

The same computation if the actions are played in the other order leads to a symmetric result except for the last term where appears $s(s(z,i',j'),i,j)$. So q is a commutative transition.

Now we prove that the value is the same in α. First we compute the value of the game Γ' in the different states. The state ω is absorbing so $v(\omega) = 1/2$. For all i' in I, the state is controlled by player 1 and the action i' guarantees him to stay in $z_{i'}$ so $v(z_{i'}) = 1$. Similarly for all $j' \in J$, $v(z_{j'}) = 0$ and following $v(z_{i',j'}) = E_{q(\alpha,i',j'|Z)}(r(z,i,j))$. By replacing all these states by their values, the situation in α is the same as in $\Gamma(\alpha)$. So the value is the same in both games and a strategy which guarantees w in $\Gamma'(\alpha)$, guarantees w in $\Gamma(\alpha)$. □

3 Results

3.1 Markov Decision Process

A MDP is a one player stochastic game. Formally with the previous notations it is a stochastic game where J is a singleton. Thus we denote a MDP Γ by (Z, I, q, r).

Theorem 1. *Let $\Gamma = (Z, I, q, r)$ be a MDP such that q is deterministic and weakly commutative.*

- *If for all $p \in \Delta_f(Z)$, $\Gamma(p)$ has a uniform value in pure strategies then for all $p \in \Delta_f(Z)$ there exists a 0-optimal strategy.*
- *Moreover if Z is a precompact metric space, q is non expansive and r is uniformly continuous then there exists a 0-optimal pure strategy.*

Note that Renault [Ren09] proves that the topological assumptions of the second part are sufficient to ensure the existence of the uniform value in pure strategies. Exemple 3 shows that there may exists a value in pure strategies without a 0-optimal pure strategy.

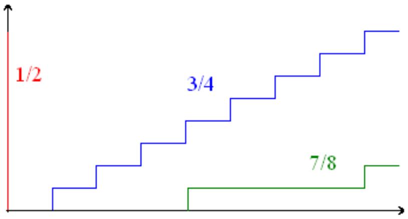

Figure 1: No pure 0-optimal strategy

Example 3. *The set of states is $\mathbb{N} \times \mathbb{N}$ and there are only two actions R and T. R increments the first coordinate and T the second one.*

$$q((x, y), R) = (x + 1, y)$$
$$q((x, y), T) = (x, y + 1)$$

Let $\epsilon_l = \frac{1}{2^l}$, for each $l \geqslant 1$ we define $w_l = \sum_{m=1}^{l} \left(3^{m-1} - 1\right)$ and the payoff by:

$$r\left(w_l, 0\right) = 1 - \epsilon_l$$
$$r\left(x, y\right) = 1 - \epsilon_l \textit{ if } \ x \in \left[w_l + (y-1)\left(3^{l-1} - 1\right), w_l + y\left(3^{l-1} - 1\right)\right]$$

For each $l \in \mathbb{N}$, there is a play induced by a pure strategy where the payoff is $1 - \epsilon_l$ at each stage. So the uniform value is equal to 1 and can be guaranted with pure strategies. But these plays move away one from each others too quickly. A 0-optimal pure strategy has to jump from one path to another but if a play leaves the path l at stage n it needs to visit more than n states with payoff 0 to reach the path $l + 1$. So there exists no 0-optimal strategy.

As stated before if we consider $\Gamma^{sb} = (K, I, q, r, \Sigma_{sb})$ a stochastic game with restricted strategies, Γ^{sb} is equivalent to a game Γ on the set of states $Z = \Delta(K)$ which is compact and with a deterministic transition function non expansive for the norm 1. Thus we can apply the Theorem 1 and deduce the following corollary.

Corollary 1. *Let $\Gamma^{sb} = (K, I, q, r, \Sigma_{sb})$ be a commutative MDP with a finite set of states K and a finite set of actions I where the player does not observe the state. For all $p_1 \in \Delta(K)$, $\Gamma^{sb}(p_1)$ has a uniform value and there exists a* 0*-optimal pure strategy.*

Rosenberg, Solan and Vieille asked the question of the existence of a 0-optimal strategy in MDP with signals. Our assumption ensures that there exists such a strategy. The following example due to Renault shows it is not true in general and therefore there exist games which cannot be transformed in order to fulfill the commutation assumption.

Example 4. *Define an MDP with no signals as follows. Let $Z = \{\alpha, \beta, 0, 1\}$, and $I = \{w, g\}$. The payoff is* 0 *except in state* 1 *where it is* 1*. The state* 0 *and*

1 *are absorbing and in the other states the transition rule q is given by*

$$\begin{aligned} q(\alpha, w) &= 1/2\delta_\alpha + 1/2\delta_\beta \\ q(\beta, w) &= \delta_\beta \\ q(\alpha, g) &= \delta_0 \\ q(\beta, g) &= \delta_1 \end{aligned}$$

An ϵ-optimal strategy in $\Gamma(\alpha)$ is to play the action w until the probability to be in β is high enough then to play g. So the uniform value starting from α is 1 but there exists no 0-optimal strategy.

3.2 Stochastic games

For two-player stochastic games the commutation does not imply the existence of 0-optimal strategies. Indeed the Big Match introduced by Gillette [Gil57] is an absorbing game without 0-optimal strategy and with deterministic transitions. Thus by Proposition 1 there exists a commutative stochastic game with deterministic transitions with the same value and in this game player 1 has no 0-optimal strategy.

Theorem 2. *Let $\Gamma(p_1) = (Z, I, J, q, r, p_1)$ be a stochastic game such that Z is a compact set of $\mathbb{R}^m$, I and J are finite sets, q is commutative deterministic non expansive for $\|.\|_1$ and r is continuous. The stochastic game $\Gamma(p_1)$ has a uniform value.*

And we can apply this result to the game Γ^{sb} where the players do not observe the state but observe past actions and prove the existence of the uniform value.

Corollary 2. *Let $\Gamma^{sb} = (K, I, J, q, r, \Sigma_{sb}, \mathcal{T}_{sb})$ be a commutative MDP with a finite set of states K and finite sets of actions I and J where the players do not observe the state. For all $p_1 \in \Delta(K)$, $\Gamma^{sb}(p_1)$ has a uniform value.*

Example 5. *Let $K = \mathbb{Z}/m\mathbb{Z}$ and define $q : K \times I \times J \to \Delta(K)$ by $q(k, i, j) = \delta_k \oplus f(i, j)$ where $f : I \times J \to \Delta(K)$ and δ_k is the Dirac measure on k. q is the law of the sum of two independent random variables of law δ_k and $f(i, j)$.*

The addition of random independent variables is a commutative and associative operation, therefore q commutes on K. For example let $m = 3$, $I = \{T, B\}$, $J = \{L, R\}$ and the function f given by

$$\begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} T \\ B \end{array} & \begin{pmatrix} 1/2\delta_1 + 1/2\delta_2 & \delta_1 \\ \delta_1 & \delta_0 \end{pmatrix} \end{array}$$

If the players play (T, L) then the new state is one of the other states with equal probability. If the players play (B, R) the state does not change. And otherwise the state goes to the next state.

4 Markov Decision Process

We focus in this section on the proof of Theorem 1. Let $\Gamma = (Z, I, J, q, r)$ be a game where the transitions are deterministic and commute. By simplicity we rename the actions $\{1, ..., I\}$. Assume that for all $p \in \Delta_f(Z)$ there exists a uniform value in pure strategies, we prove that for all $z_1 \in Z$, $\Gamma(z_1) = \Gamma(\delta_{z_1})$ has a 0-optimal strategy. It implies immediately the result for an initial probability $p_1 \in \Delta_f(Z)$. Since there is only one player and the transitions are deterministic, given an initial distribution δ_{z_1} and a pure strategy we can build a sequence of actions with the same distribution on the set of histories. So we restrict in this section to sequence of actions. Lehrer and Sorin [LS92] show that the value is always non increasing. Let show that the commutation assumption implies it is constant.

By weak commutation the state after n actions $h_n = i_1, ..., i_n$ is the same as after an ordered sequence of actions: $M(h, 1, n)$ times action 1, $M(h, 2, n)$ times action 2,..., $M(h, I, n)$ times action I. Take the lexicographic order on the set $\{1, ..., I\}^n$. The weak commutation ensures that if the sequence is not ordered there exists a transformation which leads to a smaller element. So by iteration the process converge to a minimal element which is well ordered. We denote this sequence by a vector $M(h, n)$ in $\mathbb{R}^{\sharp I}$ called the ordered representation of the strategy h_n. Remark that $M(h, n)$ is not unique.

Lemma 1. *Let h and h' such that $M(h, n) < M(h', n')$ then there exists some actions $w_{1,..,n'-n}$ such that the state is the same after $(h_{1,..,n}, w_{1,..,n'-n})$ and $h_{1,..,n'}$*

Proof: Let h and h' be two infinite histories such that there exists n and n' with $M(h, n) < M(h', n')$. Starting from the ordered representation $M(h', n')$, by commutation we can reject to the end the actions which should not be played in $M(h, n)$. We obtain a sequence of actions such that the state at stage n is the same than after $M(h, n)$ and at stage n' the same than after $M(h', n')$. □
Under the assumption of the lemma it is possible to complete the history h to reach the path taken by the history h'.

Lemma 2. *Let $z \in Z$ and ϵ a positive number there exists an ϵ-optimal strategy such that the value is non decreasing on the trajectory.*

Proof: Let $z_1 \in Z$, $(\epsilon_l)_{l \in \mathbb{N}}$ a decreasing sequence of positive numbers which converges to 0 and for each $l \in \mathbb{N}$, h_l an ϵ_l-optimal strategy in $\Gamma(z_1)$. Given h_l, we denote $M(l, n) = M(h_l, n)$. Let define $M(l) \in (\mathbb{N} \cup \{+\infty\})^I$ by iteration. Define φ_1 an extraction such that $M(l, 1, \varphi_1(n))$ converges to the inferior limit of the sequence $(M(l, 1, n))_{n \in \mathbb{N}}$. Given φ_k define ψ_{k+1} such that $M(l, k+1, \varphi_k(\psi_{k+1}(n)))$ converges to the inferior limit of the sequence $(M(l, k+1, \varphi_k(n)))_{n \in \mathbb{N}}$ and define $\varphi_{k+1} = \varphi_k \circ \psi_{k+1}$. It represents the total number of times the actions can be considered simultaneously on an infinite history.

The number of actions is finite so we can define $\psi : \mathbb{N} \to \mathbb{N}$ an extraction such that for all $i \in I$, $M(\psi(l), i)$ is increasing in l. If an action is played a finite number of times in the strategy $\sigma_{\psi(l)}$, then this action is played more times in each $\sigma_{\psi(l')}$ for $l' \geqslant l$. If it is played infinitely often in the strategy $\sigma_{\psi(l)}$ then this action is also played an infinity of times in all $\sigma_{\psi(l')}$ for $l' \geqslant l$.

Let $l \in \mathbb{N}$, we prove that the value is non decreasing along the trajectory $\sigma_{\psi(l)}$. Let $m \in \mathbb{N}$ and $z = z_m(z_1, \sigma_{\psi(l)})$ a point on the trajectory. By definition of the inferior limit, there exists $\widetilde{m} \geqslant m$ such that the ordered representation of $z' = z_{\widetilde{m}}(z_1, \sigma_{\psi(l)})$ is smaller than $M(\sigma_{\psi(l)})$.

$$\forall i \in \{1, ..., I\}\ M(\psi(l), i) \geqslant M(\psi(l), i, \widetilde{m})$$

Let $l' \geqslant l$ and let show that there exists m' such that

$$\forall i \in \{1, ..., I\}\ M(\psi(l'), i, m') \geqslant M(\psi(l), i, \widetilde{m})$$

and we will be able to continue the history $\psi(l)$ from $\widetilde{m}$ to join the path $\psi(l')$.

By definition of $M(\psi(l'))$, there exists $m' \in \mathbb{N}$ such that if $M(\sigma_{\psi(l')}, i)$ is infinity then $M(\psi(l'), i, m') \geqslant M(\psi(l), i, \widetilde{m})$ and if $M(\sigma_{\psi(l')}, i)$ is finite then $M(\psi(l'), m') = M(\psi(l'))$.

$$M(\psi(l'), i, m') = M(\psi(l'), i) \geqslant M(\psi(l), i) \geqslant M(\psi(l), i, \widetilde{m})$$

So for any $l' \geqslant l$, there exists, by Lemma 1, actions which allow to reach the trajectory followed by $\psi(l')$ from stage m. Define σ' the strategy which follows first $\sigma_{\psi(l)}$ for $\widetilde{m}$ stages, then jumps from $\sigma_{\psi(l)}$ to $\sigma_{\psi(l')}$ and finally follows $\sigma_{\psi(l')}$ from m'. By construction the trajectory from stage m' is the same as $\sigma_{\psi(l')}$ so this strategy guarantees $v(z_1) - 2\epsilon_{\psi(l')}$. Moreover at stage m the state is z so the value of the game $\Gamma(z)$ is greater than $v(z_1) - 3\epsilon_{\psi(l')}$. This is true for all $l' \geqslant l$, so the value in z is equal to $v(z_1)$. Finally the result is independent of the integer m so the value is non decreasing on the trajectory. □

Existence of a mixted 0-optimal strategy:

We define our 0-optimal strategy by concatenation of strategies given by the Lemma 2. Given a stopping time u and two strategies σ, σ' we define $\sigma u \sigma'$ as follows: play σ until u, then switch to σ' (and forget the history up to u). Formally, for every $n \in N$ and every $h_n = (z_1, i_1, j_1, ..., z_n)$, $(\sigma u \sigma')(h_n) = \sigma(h_n)$ if $u(h_n) > n$ and $(\sigma u \sigma')(h_n) = \sigma'(h_n^u)$ if $u(h_n) > n$ where $h_n^u = (z_u, i_u, j_u ..., z_n)$.

Let $z_1 \in Z$ and $(\epsilon_l)_{l \in \mathbb{N}}$ a decreasing sequence converging to 0. For each $z \in Z$ and integer l we denote by $\sigma_l(z)$ an ϵ_l-optimal strategy in $\Gamma(z_1)$ such that the value is constant on the trajectory and $N(l, z)$ an integer such that

$$\forall n \geqslant N(l, z)\ \gamma_n(z, \sigma_l(z)) \geqslant v(z) - \epsilon_l$$

Define recursively a sequence $(T_j)_{j \in \mathbb{N}}$ of finite sets of stages. Let $T_0^1 = 0$ and assume that the set T_j exists. We denote $t_{j+1} = \left[\frac{1}{\epsilon_{j+1}}\right] + 1$ and define the next set T_{j+1} by:

$$\begin{aligned}
T_{j+1}^1 &= T_j^{t_j} + N(j, z_{T_j^{t_j}}) + \frac{1}{\epsilon_j} T_j^{t_j} \\
T_{j+1}^2 &= T_{j+1}^1 + N(j+1, z_{T_{j+1}^1}) \\
&\ldots\ldots \\
T_{j+1}^{t_{j+1}} &= T_{j+1}^{t_{j+1}-1} + N(j+1, z_{T_{j+1}^{t_{j+1}-1}})
\end{aligned}$$

For each set T_j we define a stopping times u_j such that for all $m \in \{1, ..., t_j\}$, $P(u_j = T_j^m) = \frac{1}{t_j}$. τ_j is a random variable on a finite state and by construction of t_j, $P(u_j = T_j^m) \leqslant \epsilon_j$. Denote by

$$\sigma_j^*(z_1) = \sigma_0(z_1) u_1 \sigma_1(z_{u_1}) .... u_j \sigma_j(z_{u_j})$$

and the strategy which coincides with σ_j^* on the set $\{n \leqslant u_{j+1}\}$.

$$\sigma^*(z_1) = \sigma_0(z_1)u_1\sigma_1(z_{u_1})....u_j\sigma_j(z_{u_j})...$$

Let prove that σ^* is a 0-optimal strategy. Let $n \geqslant T_{j+1}^1$ we show first that the strategy $\sigma_j^*(z_1)$ is $2\epsilon_j$-optimal and more precisely that for each realisation of the stopping times u_j the payoff is $2\epsilon_j$-optimal. Indeed by construction $\sigma_j^* = \sigma_{j-1}^* u_j \sigma_j(z_{u_j})$ and for all realisation of u_j, $n - u_j \geqslant N(j, z_{u_j})$. Let $m \in \{1, ..., t_j\}$ then

$$\begin{aligned}
\gamma_n(z_1, \sigma_{j-1}^* T_j^m \sigma_j) &= E\left[\frac{T_j^m}{n}\gamma_{T_j^m}(z_1, \sigma_j^*) + \frac{n - T_j^m}{n}\gamma_{T_j^m+1,n}(z_1, \sigma_j^*)\right] \\
&\geqslant E\left[\gamma_{T_j^m+1,n}(z_1, \sigma_j^*) - \frac{T_j^m}{n}\right] \\
&\geqslant E\left[\gamma_{n-T_j^m}(z_{T_j^m}, \sigma_j(z_{T_j^m}))\right] - E\left[\frac{T_j^m}{n}\right] \\
&\geqslant E\left[v(z_{T_j^m}) - \epsilon_j\right] - \frac{T_j^{t_j}}{T_{j+1}^1} \\
&\geqslant v(z_1) - 2\epsilon_j
\end{aligned}$$

The strategy we are interesting in when $n \geqslant T_{j+1}$ is σ_{j+1}^*. Let prove that $\sigma_{j+1}^*(z_1)$ is $3\epsilon_j$-optimal. Since $\sigma_{j+1}^* = \sigma_j^* u_{j+1}\sigma_{j+1}(z_{u_{j+1}})$ both strategies are the same until the realisation of u_{j+1}.

$$\begin{aligned}
&\gamma_n(z_1, \sigma_{j+1}^*) \\
&= E\left[\left(\frac{u_{j+1}}{n}\gamma_{u_{j+1}}(z_1, \sigma_{j+1}^*) + \frac{n - u_{j+1}}{n}\gamma_{u_j+1,n}(z_1, \sigma_{j+1}^*)\right)\mathbb{1}_{u_{j+1} \leqslant n} + \left(\gamma_n(z_1, \sigma_{j+1}^*)\right)\mathbb{1}_{u_{j+1} > n}\right] \\
&= E\left[\sum_{t \in T_{j+1}}\left(\left(\frac{t}{n}\gamma_t(z_1, \sigma_j^*) + \frac{n-t}{n}\gamma_{t+1,n}(z_1, \sigma_{j+1}^*)\right)\mathbb{1}_{u_{j+1}=t \leqslant n} + \left(\gamma_n(z_1, \sigma_j^*)\right)\mathbb{1}_{u_{j+1}=t>n}\right)\right]
\end{aligned}$$

Let $n \in [T_{j+1}^1, T_{j+2}^1]$, by definition of the elements of T_{j+1} there exists a unique $m \in \{1, ..., t_{j+1}\}$ such that $n \in [T_{j+1}^m, T_{j+1}^{m+1}]$. Moreover for all $l \leqslant m$,

$$n - T_{j+1}^l \geqslant N(j+1, z_{T_{j+1}^l})$$

so in the previous decomposition there are three cases : the first block before m, the block m and the rest. Let $l < m$ then

$$\begin{aligned}
&\frac{T_{j+1}^l}{n}\gamma_{T_{j+1}^l}(z_1, \sigma_j^*) + \frac{n - T_{j+1}^l}{n}\gamma_{T_{j+1}^l+1,n}(z_1, \sigma_{j+1}^*) \\
&\geqslant \frac{T_{j+1}^l}{n}(v(z_1) - 2\epsilon_j) + \frac{n - T_{j+1}^l}{n}(v(z_{u_{j+1}}) - \epsilon_{j+1}) \\
&\geqslant v(z_1) - 2\epsilon_j
\end{aligned}$$

We have showed that σ_i^* is $2\epsilon_j$-optimal and we have played the strategy σ_{j+1} for enough time in order to be ϵ_{j+1}-optimal in the game starting in $z_{u_{j+1}}$ for all realisation of the stopping time. Let $l > m$ then both strategies σ_{j+1}^* and σ_j^* are equal so σ_{j+1}^* guarantees $v(z_1) - 2\epsilon_j$ on this event. In the last case we do not control the value but we know that the probability of the event $\{u_j = m\}$ is less than ϵ_j by construction. So we can conclude by

$$\begin{aligned}\gamma_n(z_1, \sigma^*_{j+1}) &\geqslant P(u_j \leqslant m)(v(z_1) - 2\epsilon_j) + P(u_j \geqslant m+1)(v(z_1) - 2\epsilon_j)\\ &\geqslant v(z_1) - 2\epsilon_j - P(u_j = m)\\ &\geqslant v(z_1) - 3\epsilon_j\end{aligned}$$

The strategy σ^* coincides with σ^*_j between the stages T^1_{j+1} and T^1_{j+1} so it guarantees $v(z_1) - 3\epsilon_j$ on this period. This is true for all integers j so the strategy σ^* is a 0-optimal strategy. □

Existence of a pure 0-optimal strategy in the precompact case: Let show first that we can assume Z is compact without loss of generality. Since Z is a precompact metric space, q non expansive and r uniformly continuous. The game has a uniform value in pure strategies by Renault [Ren09]. Let $\hat{Z}$ be the Cauchy completion of Z. We can extend q and r to the adherence of Z in $\hat{Z}$ which is $\hat{Z}$. It defines a game $\hat{\Gamma}$ on a compact set with a non expansive transition and a reward function uniformly continuous. By Renault [Ren09] this game has also a uniform value. Moreover if z_1 is an initial point in Z, the trajectory in $\hat{\Gamma}$ from z_1 stays in Z so both values are equal and a 0-optimal strategy in $\hat{\Gamma}$ is well defined in Γ. So in the rest of the proof we assume Z compact.

Let $z_1 \in Z$ and $(\epsilon_l)_{l\in\mathbb{N}}$ a decreasing sequence of positive numbers which converges to 0. For each $z \in Z$ and $l \in \mathbb{N}$, we denote by $\sigma_l(z)$ an ϵ_l-optimal strategy in $\Gamma(z)$ such that the value is constant and by $N(l, z)$ an integer such that

$$\forall n \geqslant N(l, z)\ \gamma_n(z, \sigma_l(z)) \geqslant v(z) - \epsilon_l$$

Since r is uniformly continuous and q is non expansive, there exists a sequence $(\eta_l)_{l\in\mathbb{N}}$ such that

$$\forall z, z' \in Z\ d(z, z') \leqslant \eta_l\ \forall\sigma,\ \forall n \in \mathbb{N}\ |\gamma_n(z, \sigma) - \gamma_n(z', \sigma)| \leqslant \epsilon_n$$

Let $z^1 = z_1$ and given $(z^j)_{j\leqslant l}$ we define z^{l+1} as an adherence point of the trajectory $(z^l, \sigma_l(z^l))$. Since the value is constant on the trajectory $(z^l, \sigma_l(z_l))$, the uniform value in z^{l+1} is equal to $v(z_1)$.

To construct our 0-optimal strategy we will first split each trajectory $\sigma_j(z^j)$ in block by recurrence and then concatenate this block. For each $j \in \mathbb{N}$, if $(n^j_k)_{k\in\mathbb{N}}$ is an increasing sequence of integers we denote w^j_0 the actions before stage n^j_0 and for all $k \in \mathbb{N}$, w^j_k the $\sharp w^j_k$ actions between the indices n^j_{k-1} and n^j_k. Assume that $(n^j_k)_{j\leqslant(l-1),k\in\mathbb{N}}$ are defined and let build the sequence for $j = l$. We define $L_l = \sum_{j\leqslant l-1,k\leqslant l} \sharp w^j_k$ and denote $(z^l_n)_{n\in\mathbb{N}}$ the sequence of state of $(z_l, \sigma_l(z_l))$. We do not care about the first indices and the first condition is on n^l_{l+1}. We choose n^l_{l+1} such that the mean payoff before this stage is good.

$$n^l_{l+1} \geqslant N(l, z^l)$$

Moreover we choose the length in order to be much bigger than some blocks taken on the trajectories $(\sigma_j(z^j))_{j\leqslant l}$ which will be played before,

$$\frac{L_l}{n^l_{l+1}} \leqslant \epsilon_l$$

and such that at the beginning of the l block of this decomposition the state is near the adherence point

$$d(z^l_{n^l_k}, z^{l+1}) \leqslant \frac{\eta_k}{k}$$

Finally we assume that this block is long respect to some blocks taken on the trajectories $(\sigma_j(z^j))_{j \leqslant l}$ which will be played just after and the time to be optimal if we start to play in $\Gamma(z^{l+1})$.

$$\frac{N(l+1, z^{l+1}) + \sum_{j=0}^{l-1} \sharp w^j_{l+1}}{n^l_{l+1}} \leqslant \epsilon_l$$

Let define the strategy σ^* by blocks. The idea is to follow on the block l first the strategy $\sigma_l(z^l)$ then some actions to be sure that the state at the beginning of the next block is near z^{l+1}. A sequence of actions is said to satisfy the property $H(l)$ if at stage L_l the state is the same as after the following strategy: the n^1_l first actions of $\sigma_1(z^1)$,..., and the n^{l-1}_l first actions of the strategy $\sigma_{l-1}(z^{l-1})$. Assume that the first $l-1$ blocks are built and satisfy $H(l)$. On the block l follow $\sigma_l(z^l)$ for n^l_{l+1} stages then play w' given by Lemma 1 such that the new sequence fulfills $H(l+1)$. The size of w' is $\sum_{j=0}^{l-1} \sharp w^j_{l+1}$.

Let denote by $(z_k)_{k \in \mathbb{N}}$ the sequence of states when σ^* is played and let prove that the mean payoff converges to $v(z_1)$. We focus on the state at stage L_l which is the beginning of the block l. Since it satisfies the property H_l, the state is the same as after the sequence of actions $(w^0_k)_{k \leqslant l}$,..., $(w^{l-1}_k)_{k \leqslant l}$. The application q is non expansive so we have by an immediate recurrence

$$d(z_{L_l}, z^l) \leqslant \eta_l$$

If we consider the game of length $L_l + n^l_{l+1}$, the strategy played is optimal for n^l_{l+1} stages in $\Gamma(z^l)$ and we control the distance between z_{L_l} and z^l. So we have

$$\begin{aligned}
\gamma_{L_l + n^l_{l+1}}(z_1, \sigma^*) &= \frac{L_l}{L_l + n^l_{l+1}} \gamma_{L_l}(z, \sigma^*) + \frac{n^l_{l+1}}{L_l + n^l_{l+1}} \gamma_{L_l+1, L_l + n^l_{l+1}}(z, \sigma^*) \\
&\geqslant \frac{n^l_{l+1}}{L_l + n^l_{l+1}} \gamma_{L_l+1, L_l + n^l_{l+1}}(z, \sigma^*) \\
&\geqslant \gamma_{L_l+1, L_l + n^l_{l+1}}(z, \sigma^*) - \frac{L_l}{L_l + n^l_{l+1}} \\
&\geqslant \gamma_{L_l+1, L_l + n^l_{l+1}}(z, \sigma^*) - \epsilon_l \\
&\geqslant \gamma_{n^l_{l+1}}(z_{L_l}, \sigma_l(z^l)) - \epsilon_l \\
&\geqslant \gamma_{n^l_{l+1}}(z^l, \sigma_l(z^l)) - 2\epsilon_l \\
&\geqslant v(z_1) - 3\epsilon_l
\end{aligned}$$

If $n \in [L_l + n^{l+1}_l, L_{l+1} + N(l+1, z_{l+1})]$ the number of stages is close to the previous cases since $n - L_l - n^l_{l+1} \leqslant N(l+1, z^{l+1}) + \sum_{j=0}^{l-1} \sharp w^j_{l+1}$ so

$$\begin{aligned}
\gamma_n(z_1,\sigma^*) &= \frac{L_l + n^l_{l+1}}{n}\gamma_{L_l+n^l_{l+1}}(z,\sigma^*) + \frac{n - L_l - n^l_{l+1}}{n}\gamma_{L_l+n^l_{l+1}+1,n}(z,\sigma^*) \\
&\geqslant \frac{L_l + n^l_{l+1}}{n}\gamma_{L_l+n^l_{l+1}}(z,\sigma^*) \\
&\geqslant \gamma_{L_l+n^l_{l+1}}(z,\sigma^*) - \frac{n - L_l - n^l_{l+1}}{n} \\
&\geqslant v(z_1) - 3\epsilon_l - \frac{n - L_l - n^l_{l+1}}{n^l_{l+1}} \\
&\geqslant v(z_1) - 4\epsilon_l
\end{aligned}$$

And finally if $n \in [L_{l+1} + N(l+1, z_{l+1}), L_{l+1} + n^{l+2}_{l+1}]$ we concatenate a strategy good until L_{l+1} and a strategy good from stage L_{l+1}.

$$\begin{aligned}
\gamma_n(z_1,\sigma^*) &= \frac{L_{l+1}}{n}\gamma_{L_{l+1}}(z,\sigma^*) + \frac{n - L_{l+1}}{n}\gamma_{L_{l+1}+1,n}(z,\sigma^*) \\
&\geqslant \frac{L_{l+1}}{n}(v(z_1) - 4\epsilon_l) + \frac{n - L_{l+1}}{n}(v(z_1) - 4\epsilon_{l+1}) \\
&\geqslant v(z_1) - 4\epsilon_l
\end{aligned}$$

So the mean payoff converges to $v(z_1)$ and we have built a 0-optimal strategy without randomization. □

5 Uniform value in stochastic games

This section is dedicated to the proof of Theorem 2. We focus on the case where the initial probability is a Dirac mass. The general result is an immediate consequence. The application q is deterministic, so we can define the trajectory along a sequence of actions and we denote by $q_{i,j}$ the operator from Z to Z defined by $q_{i,j}(z) = q(i,j,z)$. Let $n \in \mathbb{N}$ and $h = (i_1, j_1, ..., i_n, j_n) \in (I \times J)^n$, for all integer $s \leqslant n$, we denote $z_{s+1}(h) = q_{i_s,j_s}...q_{i_1,j_1}z_1 = \prod_{l=1}^{s} q_{i_l,j_l}z_1$. To study the system, we introduce the orbits of z by several families of actions and with prescribed number of stages. On one hand if $S \subset I \times J$ and $l \in \mathbb{N}^*$, let $\Lambda^+_S(z,l) = \{z_n(h), h \in S^n,\ n \geqslant l\}$. It is the set of points reached along a path of at least l stages with actions in S. On another hand $\Lambda^-_S(z,l) = \{z_n(h), h \in S^n,\ n \leqslant l\}$ is the set of points reached in less than l stages. Notice that $\Lambda^+_S(z,1)$ is the set of points reached with actions in S without restriction on the number of stages.

The operator $q_{i,j}$ is a non expansive mapping on $\mathbb{R}^n$ for $\|.\|_1$ so it has a specific ergodic behaviour. We can deduce from Sine [Sin90] the following lemma

Lemma 3. *Let M be an operator from $Z \subset \mathbb{R}^m$ to Z non expansive for $\|.\|_1$ then there exists an integer $L \leqslant \varphi(m)$ and a family of operators $B_0,\cdots, B_{L-1}$ such that*

$$\forall l \in \{0, ..., L-1\} \lim_{n\to+\infty} M^{nL+l} = B_l.$$

A classic example is the case where M is the transition of a Markov chain on a finite set. If λ is a complex eigenvalue of M then $|\lambda| \leqslant 1$ since the application is non expansive. Moreover the theorem of Perron-Frobenius ensures that if $|\lambda| = 1$ then there exists $l \leqslant m$ such that $\lambda^l = 1$. The integer L is a common multiple of such l and for example we can take $\varphi(m) = m!$.

For each $z \in Z$ we separate the couples of actions in two different groups. On one hand the actions which from z come back to z and on the other hand the rest.

Definition 4. *Let $z \in Z$, the couple $(i, j) \in I \times J$ is cyclic in z if there exists $t \leqslant \varphi(m)$ such that $q_{i,j}^t z = z$. We denote by $S(z)$ the set of cyclic actions in z and $\overline{S}(z)$ its complementary.*

Lemma 4. *If $z \in Z$ and $z' \in \Lambda^+_{I \times J}(z, 1)$ then $S(z)$ is included in $S(z')$.*

Proof : Indeed fix $z' \in \Lambda^+_{I \times J}(p, 1)$, there exists a sequence $(i_1, j_1, ..., i_n, j_n) \in (I \times J)^n$ such that $z' = \prod_{l=1..n} q_{i_l,j_l} z$. Let $(i^*, j^*) \in S(z)$ and $d \in \mathbb{N}$ such that $q^d_{i^*,j^*} z = z$ then

$$q^d_{i^*,j^*} z' = \prod_{l=1..n} q_{i_l,j_l} q^d_{i^*,j^*} z = z'.\square$$

Example 6. *Let $Z = \Delta(\mathbb{Z}/2\mathbb{Z})$, $z_1 = (1, 0)$, $I = \{1\}$, $J = \{1\}$ and $A = A(1, 1) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$. Then $S(z_1)$ is empty and $S(Az_1) = \{(1, 1)\}$.*

Thus the mapping S is increasing for the inclusion order along a trajectory and we can prove our result by induction. Given the game Γ, we show that the uniform value exists by induction on the cardinality of $\overline{S}(z)$ for all initial points $z \in Z$. The induction hypothesis is the following,

$$\Psi_k : \{ \text{ if } z \in Z \text{ and } \sharp\overline{S}(z) \leqslant k \text{ then } \Gamma(z) \text{ has a uniform value } \}.$$

We will see that the initial step Ψ_0 is true since the use of cyclic actions leads to a finite number of states. For the recurrence step we define an auxiliary game. In the auxiliary game if the trajectory in $\Gamma(z)$ goes near to a point z' where the induction hypothesis is satisfied absorption occurs with payoff the uniform value in z'.

Proposition 2. *Ψ_0 is true.*

Lemma 5. *The number of stages needed to reach all points reached by iteration of actions in $S(z)$ is finite.*

First we prove the proposition. If there are only cyclic actions, the number of states reached during the game is finite by the lemma. The game is formally a stochastic game with a finite set of states and finite sets of actions with classical sets of strategies. Therefore it has a uniform value by the result of Mertens Neymann [MN81].

Proof of the lemma: Let $z \in Z$, we prove that we can restrict to $(\varphi(m) - 1)\sharp S(z)$ stages.

$$\Lambda^+_{S(z)}(z, 1) = \Lambda^-_{S(z)}(z, (\varphi(m) - 1)\sharp S(z))$$

By definition the set of points reached in less than $(\varphi(m)-1)\sharp S(z)$ is included in the set of points reached without limitation on the number of stages. We show the other inclusion by contradiction. Assume there exists z^* which is not reached in $(\varphi(m)-1)\sharp S(z)$ stages.

Let $n^* = \inf_{n\in\mathbb{N}^*}\{n,\ \exists h=(i_l,j_l)_{l=1..n}\in(I\times J)^n,\ z_n(h)=z^*\}$ the minimum number of stages needed to reach α. It is well defined and strictly superior to $(\varphi(m)-1)\sharp S(z)$ by definition of z^*.

$$\begin{aligned}
&\sum_{(i,j)\in S(z)} \sharp\{l,(i_l,j_l)=(i,j)\} = n^* \\
\Leftrightarrow\ &\exists(i^*,j^*)\in S(z)\ \sharp\{l,(i_l,j_l)=(i^*,j^*)\} \geqslant \frac{n^*}{\sharp S(z)} \\
\Leftrightarrow\ &\exists(i^*,j^*)\in S(z)\ \sharp\{l,(i_l,j_l)=(i^*,j^*)\} \geqslant \varphi(m)
\end{aligned}$$

So an action is repeated more than $\varphi(m)$ times. By definition there exists $d^*\leqslant\varphi(m)$ such that $q_{i,j}^{d^*}z=z$. Hence denote by h' the sequence of actions deduced from h by deleting d^* times the action (i^*,j^*). By commutation $z_{n^*-d}(h')=z^*$ which contradicts the definition of n^* and concludes the proof. □

Focus now on the step of the induction. Let $k\in\mathbb{N}$ such that Ψ_k is true and $z_1\in Z$ such that $\sharp\overline{S}(z_1)=k+1$. Since we are studying a game on a compact set with a non expansive transition function we have the following lemma which allows us to use for all $z\in Z$ an ϵ-optimal strategy in $\Gamma(z)$ in the games beginning in the neighbourhood of z.

Lemma 6. *Given $\epsilon>0$ there exists $\eta>0$ such that if player 1 guarantees w in $\Gamma(z')$ then for all z, such that $\|z-z'\|\leqslant\eta$, he guarantees $w-\epsilon$ in $\Gamma(z)$*

Given $\epsilon>0$, for all $(i,j)\in I\times J$ the application $r(.,i,j)$ is uniformly continuous. Moreover there exists a finite number of applications so there exists $\eta>0$ such that

$$\forall z,\ z'\in Z\ \|z-z'\|\leqslant\eta,\ \forall(i,j)\in(I\times J)\ |r(z,i,j)-r(z',i,j)|\leqslant\epsilon.$$

For all $\sigma\in\Sigma$, we can define a strategy σ^* which does not depend on the state such that for all $\tau\in\mathcal{T}$ the probability on the histories under (z,σ,τ) and (z,σ^*,τ) are the same. It plays as if the game was $\Gamma(z)$ for all the initial point. Let $\tau=(j_1,...,j_n)$ be a sequence of actions of player 2. Denote by h_n the trajectory given by (z,σ^*,τ) and h'_n the trajectory given by (z',σ^*,τ). For all $(i,j)\in I\times J$, q is a non expansive application so for all $n\in\mathbb{N}$, $\|z_n-z'_n\|\leqslant\|z-z'\|\leqslant\eta$ and

$$\begin{aligned}
|\gamma_n(z,\sigma^*,\tau)-\gamma_n(z',\sigma^*,\tau)| &\leqslant \frac{1}{N}\sum_{n=1}^{N}|r(z_n,i_n,j_n)-r(z'_n,i_n,j_n)| \\
&\leqslant \epsilon
\end{aligned}$$

If player 1 guarantees w in $\Gamma(z')$ then he guarantees $w-\epsilon$ in the game $\Gamma(z)$. □

Let $\epsilon > 0$ and η given by Lemma 6 and define Ξ a correspondence which gives for each point the points in the neighbourhood where we know the uniform value exists by the induction hypothesis.

$$\Xi(z) = \{z' \in Z \text{ such that } \sharp\overline{S}(z') \leqslant n \text{ and } \|z - z'\|_1 \leqslant \eta\}.$$

Denote by Φ the set of points reachable from z_1 where $\Xi(z)$ is empty.

$$\Phi = \{z \in \Lambda^+_{I\times J}(z_1, 1),\ \forall z' \in B_1(z, \eta),\ \sharp\overline{S}(z') \geqslant k + 1\}.$$

Proposition 3. *The set Φ is finite.*

Lemma 7. *Let $z \in Z$, $(i, j) \in \overline{S}(z)$ and $\epsilon > 0$ then there exists an integer u such that*

$$\forall f \in \Lambda^+_{(i,j)}(z, u)\ \exists f' \in Z\ \|f - f'\|_1 \leqslant \epsilon \textit{ and } \sharp S(f') > \sharp S(z).$$

When iterating a non cyclic action the trajectory converges to a periodic orbit where we can apply the recurrence hypothesis.

Example 7. *Let $Z = \Delta(\mathbb{Z}/2\mathbb{Z})$, $z_1 = (1, 0)$, $I = \{1\}$, $J = \{1\}$ and $A = A(1, 1) = \begin{pmatrix} 1/4 & 3/4 \\ 3/4 & 1/4 \end{pmatrix}$. Then for all $n \in \mathbb{N}$, $S(A^n z_0)$ is empty but $A^n z_0$ converges to $z_\infty = (1/2, 1/2)$ and $S(z_\infty) = \{(1, 1)\}$.*

Proof of the lemma : Let $z \in Z$, $(i, j) \in \overline{S}(z)$ a couple of action and ϵ a positive real. By the Lemma 3 applied to $M = q_{i,j}$, there exists an integer L and operators B_0,..., B_{L-1} such that

$$\forall l \in \{0, ..., L - 1\} \lim_{n\to+\infty} M^{nL+l} = B_l$$

Let $y = B_0 z$ then the sequence of iterated converges to the family $(M^l y)_{l=0..L-1}$. There exists an integer u such that $\forall n \geqslant u$, $\|M^{nL} z - y\|_1 \leqslant \epsilon$ and as M is non expansive for the norm 1, $\|M^{nL+l} z - yM^l\|_1 \leqslant \epsilon$. We denote $u' = u(L+1)$ and

$$\forall x \in \Lambda^+_{(i,j)}(z, u')\ \exists x' \in \{M^l z,\ l = 0, \ldots, L - 1\}\ \|x - x'\|_1 \leqslant \epsilon$$

The trajectory converges to a finite set of points.

Furthermore $S(z)$ is included in $S(y)$. Let $(i', j') \in S(z)$ and d an integer such that $q^d_{i',j'} z = z$ then

$$\begin{aligned} q^d_{i',j'} y = q^d_{i',j'} B_0 z &= \lim_n q^d_{i',j'} M^{nL} z \\ &= \lim_n M^{nL} q^d_{i',j'} z = \lim_n M^{nL} z = y \end{aligned}$$

And $(i, j) \in S(y)$ by construction of y: the integer L is smaller than $\varphi(m)$ and

$$M^L y = M^L B_0 z = \lim_n M^L M^{nL} z = \lim_n M^{(n+1)L} z = z$$

Under our assumption (i, j) is in $S(y)$ but not in $S(z)$ so $\sharp S(y) > \sharp S(z)$ and since S is increasing along the trajectory we proved that for all $l \in \{0, ..., L-1\}$, $\sharp S(yM^l) > \sharp S(z)$. $\square$

Proof of the proposition : Let $H = \{h \in (I \times J)^{\mathbb{N}} | \; \exists n \geqslant 1, z_n(h) \in \Phi\}$, the set of possible histories associated to elements of Φ. For all $z \in \Phi$, we denote $n^*(z) = \inf\{n | \; \exists h \in (I \times J)^n \; z_n(h) = z\}$, the least number of stages necessary to reach z. Let show that the Lemmas 5 and 7 imply that the set $F = \{n^*(z) | z \in \Phi\}$ is finite. By the Lemma 7 there exists an integer u such that for all couples (i,j) in $\overline{S}(z_1)$.

$$\forall f \in \Lambda^+_{(i,j)}(p,u) \; \exists f' \in \mathbb{R}^m \; \|f - f'\|_1 \leqslant \epsilon \text{ and } \sharp\overline{S}(f') < k+1.$$

We prove that $N = \max(u, \varphi(m))\sharp(I \times J)$ is a superior bound of F. If $n^* \geqslant N$ and h is an associated history to n^*, then one action (i^*, j^*) is repeated more than $\max(u, \varphi(m))$ times. As in the proof of Proposition 2 either this action is in $S(z_1)$ and the history can be shortened which is absurd respect to the definition of n^* or it is in $\overline{S}(z_1)$ and by construction there exists $f' \in Z$ such that

$$\begin{aligned} \|q^u_{i^*,j^*} z_1 - f'\|_1 &\leqslant \epsilon \\ \sharp\overline{S}(f') &< k+1 \end{aligned}$$

But all the transitions are non-expansive and by Lemma 4, S is increasing along the orbits. Therefore p is not in Φ which is absurd. Thus F is finite and since at each stage there exists a finite number of actions, Φ is finite. □

We define ξ a selection of Ξ on the element of Φ where Ξ is non empty and an auxiliary game $\dot{\Gamma}(\epsilon, z_1)$ as the following: the initial state is z_1, the set of actions are I et J and the transition and reward functions are given by :

$$\dot{q}(z,i,j) = \begin{cases} z & \text{if } \Xi(z) \text{ non empty} \\ q_{i,j} z & \text{otherwise} \end{cases}$$
$$\dot{r}(z,i,j) = \begin{cases} v(\xi(z)) & \text{if } \Xi(z) \text{ non empty} \\ r(z,i,j) & \text{otherwise} \end{cases}$$

The sets of strategy for player 1 and 2 are the same as in the game Γ.

Proposition 4. $\dot{\Gamma}(\epsilon, z_1)$ *has a uniform value.*

All the points are in Φ so this game is formally a stochastic game with a finite set of states and finite sets of actions. So $\dot{\Gamma}(\epsilon, z_1)$ has a uniform value by the theorem of Mertens Neymann [MN81]). □

Moreover the value of the auxiliary game is a good approximation of what the players can guarantee in $\Gamma(z_1)$.

Proposition 5. *If player* 1 *can guarantee* w *in* $\dot{\Gamma}(\epsilon, z_1)$ *then he can guarantee* $w - 3\epsilon$ *in* $\Gamma(z_1)$.

Proof of the proposition: By assumption, there exists $\dot{\sigma}$ a strategy of player 1 in $\dot{\Gamma}(\epsilon, z_1)$ and a stage $\dot{N}$ such that

$$\forall n \geqslant \dot{N} \; \forall \dot{\tau} \; \dot{\gamma}_n(z_1, \dot{\sigma}, \dot{\tau}) \geqslant w - \epsilon.$$

Moreover for each couple of points $(z, \xi(z))$ if $\xi(z)$ exists, we denote $\sigma^{\xi,z}$ the strategy given by Lemma 6 and we have

$$\exists N(z) \; \forall n \geqslant N(z) \; \forall \tau \; \gamma_n(z, \sigma^{\xi,z}, \tau) \geqslant v(\xi(z)) - 2\epsilon.$$

Let θ be the mapping from $(Z \times I \times J)^{\mathbb{N}}$ to $\mathbb{N}^*$ given by

$$\theta(h) = \inf_{n \in \mathbb{N}^*} \{n |\ \Xi(z_t(h)) \neq \emptyset\}.$$

and σ the following strategy in the game $\Gamma(z_1)$:

$$\sigma_n(h) = \begin{cases} \dot{\sigma}_n(h) & \text{if } n \leqslant \theta(h) \\ \sigma^{\xi, z_{\theta(h)}(h)}_{n-\theta(h)} & \text{if } n > \theta(h) \end{cases}$$

This strategy plays as in $\dot{\Gamma}(\epsilon, z_1)$ until a point z' where the process would have absorbed then play optimally as if the remaining game started from point $\xi(z')$.

Let show that this strategy guarantees $w - 3\epsilon$. If τ is a strategy of player 2, we define the sequence of random variables $\widetilde{z}_t$ of the state at stage t and a stopping time $\widetilde{\theta}$

$$\widetilde{\theta} = \inf_{n \in \mathbb{N}^*} \{n | \Xi(\widetilde{z}_n) \neq \emptyset\}.$$

Let M be a superior bound of $\{N(z),\ z \in \Phi\}$, N^* an integer such that $\frac{M}{N^*} \leqslant \epsilon$ and $N \in \mathbb{N}$ greater than N^*.

$$\begin{aligned} \gamma_N(z, \sigma, \tau) &= \frac{1}{N} E_{z,\sigma,\tau} \left(\sum_{n=1}^{\widetilde{\theta}} r(\widetilde{z}_n, \widetilde{\imath}_n, \widetilde{\jmath}_n) + \sum_{n=\widetilde{\theta}+1}^{N} r(\widetilde{z}_n, \widetilde{\imath}_n, \widetilde{\jmath}_n) \right) \\ &= \frac{1}{N} E_{z,\sigma,\tau} \left(\psi(\widetilde{z}_n, \widetilde{\imath}, \widetilde{\jmath}) \right) \\ &= \frac{1}{N} E_{z,\sigma,\tau} \left(\psi(\widetilde{z}_n, \widetilde{\imath}, \widetilde{\jmath}) \mathbb{1}_{N-\widetilde{\theta} \geqslant \overline{N}} + \psi(\widetilde{z}_n, \widetilde{\imath}, \widetilde{\jmath}) \mathbb{1}_{N-\widetilde{\theta} < \overline{N}} \right) \end{aligned}$$

We study both parts separately. In the first one, the condition on $\widetilde{\theta}$ ensures that the number of stages after the change of strategy is long enough to play well. In the second one, we do not control the payoff but the number of stages after $\widetilde{\theta}$ is small respect to N.

$$\begin{aligned} A &= \frac{1}{N} E_{z,\sigma,\tau} \left(\psi(\widetilde{z}_n, \widetilde{\imath}, \widetilde{\jmath}) \mathbb{1}_{N-\widetilde{\theta} < \overline{N}} \right) \\ &= E_{z,\sigma,\tau} \left(\frac{1}{N} \left(\sum_{n=1}^{\widetilde{\theta}} r(\widetilde{z}_n, \widetilde{\imath}_n, \widetilde{\jmath}_n) + \sum_{n=\widetilde{\theta}+1}^{N} r(\widetilde{z}_n, \widetilde{\imath}_n, \widetilde{\jmath}_n) \right) \mathbb{1}_{N-\widetilde{\theta} < \overline{N}} \right) \\ &\geqslant E_{z,\sigma,\tau} \left(\frac{1}{N} \left(\sum_{n=1}^{\widetilde{\theta}} r(\widetilde{z}_n, \widetilde{\imath}_n, \widetilde{\jmath}_n) + v(\xi(\widetilde{z}_{\widetilde{\theta}}))(N - \widetilde{\theta}) - 2\overline{N} \right) \mathbb{1}_{N-\widetilde{\theta} < \overline{N}} \right) \\ &\geqslant E_{z,\sigma,\tau} \left(\frac{1}{N} \left(\sum_{n=1}^{\widetilde{\theta}} r(\widetilde{z}_n, \widetilde{\imath}_n, \widetilde{\jmath}_n) + v(\xi(\widetilde{z}_{\widetilde{\theta}}))(N - \widetilde{\theta}) \right) \mathbb{1}_{N-\widetilde{\theta} < \overline{N}} - 2\epsilon \mathbb{1}_{N-\widetilde{\theta} < \overline{N}} \right) \end{aligned}$$

Focus now on the first part. By assumption we have that $\|\widetilde{z}_{\widetilde{\theta}} - \xi(\widetilde{z}_{\widetilde{\theta}})\| \leqslant \eta$ and $N - \widetilde{\theta} \geqslant N(z)$. If σ^{h_n} and τ^{h_n} are the strategies induced by σ and τ after $\widetilde{\theta}$

given h_n:

$$\begin{aligned} E_{z,\sigma,\tau}\left(\sum_{n=\widetilde{\theta}+1}^{N} r(\widetilde{z}_n,\widetilde{\imath}_n,\widetilde{\jmath}_n)\right) &= E_{z,\sigma,\tau}\left(\gamma_{N-\widetilde{\theta}}(\widetilde{z}_{\widetilde{\theta}},\sigma^{h_n},\tau^{h_n})(N-\widetilde{\theta})\right) \\ &\geqslant E_{z,\sigma,\tau}\left(\left(v(\xi(\widetilde{z}_{\widetilde{\theta}}))-2\epsilon\right)(N-\widetilde{\theta})\right) \end{aligned}$$

Therefore the contribution of the second term can be transformed into

$$\begin{aligned} B &= \frac{1}{N}E_{z,\sigma,\tau}(\psi(\widetilde{z}_n,\widetilde{\imath},\widetilde{\jmath})\mathbb{1}_{N-\widetilde{\theta}\geqslant\overline{N}}) \\ &= E_{z,\sigma,\tau}\left(\frac{1}{N}\left(\sum_{n=1}^{\widetilde{\theta}} r(\widetilde{z}_n,\widetilde{\imath}_n,\widetilde{\jmath}_n)+\sum_{n=\widetilde{\theta}+1}^{N} r(\widetilde{z}_n,\widetilde{\imath}_n,\widetilde{\jmath}_n)\right)\mathbb{1}_{N-\widetilde{\theta}\geqslant\overline{N}}\right) \\ &\geqslant E_{z,\sigma,\tau}\left(\frac{1}{N}\left(\sum_{n=1}^{\widetilde{\theta}} r(\widetilde{z}_n,\widetilde{\imath}_n,\widetilde{\jmath}_n)+v(\xi(\widetilde{z}_{\widetilde{\theta}}))(N-\widetilde{\theta})-2\epsilon(N-\widetilde{\theta})\right)\mathbb{1}_{N-\widetilde{\theta}\geqslant\overline{N}}\right) \\ &\geqslant E_{z,\sigma,\tau}\left(\frac{1}{N}\left(\sum_{n=1}^{\widetilde{\theta}} r(\widetilde{z}_n,\widetilde{\imath}_n,\widetilde{\jmath}_n)+v(\xi(\widetilde{z}_{\widetilde{\theta}}))(N-\widetilde{\theta})\right)\mathbb{1}_{N-\widetilde{\theta}\geqslant\overline{N}}-2\epsilon\mathbb{1}_{N-\widetilde{\theta}\geqslant\overline{N}}\right) \end{aligned}$$

So the summation of this two inferior bounds gives the result.

$$\gamma_N(z,\sigma,\tau)\geqslant\dot{\gamma}_N(z,\dot{\sigma},\tau)-2\epsilon\geqslant w-3\epsilon. \ \square$$

To conclude our proof, denote by $v(\epsilon)$ the value of the game $\dot{\Gamma}(z_1,\epsilon)$. By the previous proposition player 1 can guarantee $v(\epsilon)-3\epsilon$ for all ϵ in $\Gamma(z_1)$. So he can guarantee the superior limit when ϵ converges to 0 and hence $\limsup_{\epsilon\to 0}(v(\epsilon))$. The same demonstration proves that player 2 can guarantee the inferior limit. Therefore

$$\limsup_{\epsilon} v(\epsilon)\leqslant \max\min\leqslant\min\max\leqslant\liminf_{\epsilon} v(\epsilon)$$

Since the inferior limit is inferior to the superior limit, the max min and the min max are equal and $\Gamma(z_1)$ has a uniform value. The induction hypothesis is proven at the next step and the proof of the theorem is finished. For all $z\in Z$, the game $\Gamma(z)$ has a uniform value.

5.1 Extensions

The proof of the Theorem 2 can be extended by switching some of the lemmas with more general results. First of all the result of Sine [Sin90] applies to more general norms than the norm $\|.\|_1$.

Definition 5. *A norm on $\mathbb{R}^n$ is polyhedral if the unit ball has a finite number of extreme points.*

For example the norm $\|.\|_1$ and the sup norm are polyhedral norms but not the euclidean norm. We can deduce the following lemma and the proof of Theorem 2 leads to the same theorem with a polyhedral norm.

Lemma 8. *Let $N(.)$ be a polyhedral norm and $K \subset \mathbb{R}^m$ a compact. There exists $\varphi(N,m)$ such that for all mapping T non expansive for N, there exists $t \leqslant \varphi(N,m)$ such that $(T^{tn}_{n\in\mathbb{N}}$ converge.*

The problem if the transition is non expansive for the norm $\|.\|_2$ like the example on the circle is still open since the norm $\|.\|_2$ is not a polyhedral norm.

We can also change the result by replacing the theorem from Mertens Neyman [MN81] with other existence results. First Vieille [Vie00a][Vie00b] proves the existence of an equilibrium payoff in every two-players stochastic games. So our proof adapted to the non zero-sum case leads to the result :

Theorem 3. *Let $\Gamma = (Z, I, J, q, r_1, r_2)$ be a two-players non zero-sum stochastic game such that Z is a compact set of $\mathbb{R}^m$, I and J are finite sets, q is commutative deterministic non expansive for $\|.\|_1$ and r_1 and r_2 are continuous. The stochastic game $\Gamma(p_1)$ has an equilibrium payoff.*

Secondly there exist some specific classes of n-players stochastic games where the existence of an equilibrium has been proven. For example Flesch, Schoenmakers and Vrieze [FSV08][FSV09] prove the existence of an equilibrium for m-players stochastic games where each player controls a finite Markov chain and the payoffs depend on the m states and the m actions at stage n. Note that the commutation assumption here is reduced to a condition player by player. Under the same assumption as in Theorem 2 and 3, there exists an equilibrium payoff.

Lastly there are some open questions. In Theorem 1 and 2 we restrict to deterministic transitions and it allows us to study stochastic games where the players don't observe the state. The more general model where the players monitor past actions and have a signal on the state is linked to models with probabilistic transitions on the state of beliefs. Thus it is interesting to find a good assumption of commutation with probabilistic transitions. Another problem is to adapt the proof of Theorem 2 to the weakly commutation context. Some arguments still hold but the recurrence assumption is not pertinent any more.

References

[Bla62] David Blackwell. Discrete dynamic programming. *Ann. Math. Statist.*, 33:719–726, 1962.

[For82] F. Forges. Infinitely repeated games of incomplete information: Symmetric case with random signals. *International Journal of Game Theory*, 11(3):203–213, 1982.

[FSV08] J. Flesch, G. Schoenmakers, and K. Vrieze. Stochastic games on a product state space. *Mathematics of Operations Research*, 33(2):403–420, 2008.

[FSV09] J. Flesch, G. Schoenmakers, and K. Vrieze. Stochastic games on a product state space: The periodic case. *International Journal of Game Theory*, 38(2):263–289, 2009.

[Gei02] J. Geitner. Note Equilibrium payoffs in stochastic games of incomplete information: the general symmetric case. *International Journal of Game Theory*, 30(3):449–452, 2002.

[Gil57] D. Gillette. Stochastic games with zero stop probabilities. *Ann. Math. Stud*, 39:178–187, 1957.

[Koh74] E. Kohlberg. Repeated games with absorbing states. *The Annals of Statistics*, 2(4):724–738, 1974.

[LS92] E. Lehrer and S. Sorin. A uniform Tauberian theorem in dynamic programming. *Mathematics of Operations Research*, pages 303–307, 1992.

[MN81] J.-F. Mertens and A. Neyman. Stochastic games. *Internat. J. Game Theory*, 10(2):53–66, 1981.

[NS98] Abraham Neyman and Sylvain Sorin. Equilibria in repeated games of incomplete information: the general symmetric case. *Internat. J. Game Theory*, 27(2):201–210, 1998.

[Ren07] J. Renault. The value of Repeated Games with an informed controller. *arXiv:0803.3345v2, preprint*, 2007.

[Ren09] J. Renault. Uniform value in dynnamic programming. *arXiv:0803.2758v2, to appear JEMS*, 2009.

[RSV02] Dinah Rosenberg, Eilon Solan, and Nicolas Vieille. Blackwell optimality in Markov decision processes with partial observation. *Ann. Statist.*, 30(4):1178–1193, 2002.

[Sin90] Robert Sine. A nonlinear Perron-Frobenius theorem. *Proc. Amer. Math. Soc.*, 109(2):331–336, 1990.

[Vie00a] N. Vieille. Two-player stochastic games II: The case of recursive games. *Israel Journal of Mathematics*, 119(1):93–126, 2000.

[Vie00b] Nicolas Vieille. Two player stochastic games. I. A reduction. *Israel J. Math.*, 119:55–91, 2000.