Journal of Dynamics and Games
©American Institute of Mathematical Sciences
Volume **2**, Number **1**, January **2015**

doi:10.3934/jdg.2015.2.103

pp. **103–115**

REVERSIBILITY AND OSCILLATIONS IN ZERO-SUM DISCOUNTED STOCHASTIC GAMES

Sylvain Sorin

Sorbonne Universités, UPMC Univ Paris 06
Institut de Mathématiques de Jussieu-Paris Rive Gauche
UMR 7586, CNRS, Univ Paris Diderot, Sorbonne Paris Cité, F-75005
Paris, France

Guillaume Vigeral

Université Paris-Dauphine, CEREMADE
Place du Maréchal De Lattre de Tassigny
75775 Paris cedex 16, France

(Communicated by Rida Laraki)

Abstract. We show that by coupling two well-behaved exit-time problems one can construct two-person zero-sum dynamic games having oscillating discounted values. This unifies and generalizes recent examples of stochastic games with finite state space, due to Vigeral (2013) and Ziliotto (2013).

1. **Introduction.** 1) We first consider zero-sum games in discrete time where the purpose is to control the law of a stopping time of exit. For each evaluation of the stream of outcomes (defined by a probability distribution on the positive integers $n = 1, 2, ...$), value and optimal strategies are well defined. In particular for a given discount factor $\lambda \in]0, 1]$ optimal stationary strategies define an inertia rate Q_λ: the expected normalized time before exit.

When two such configurations (1 and 2) are coupled: the exit of each one being the starting point of the other, this induces a dynamic game. Under optimal play the state will move from one configuration to the other in a way depending on the previous rates $Q^i_\lambda, i = 1, 2$. The main observation is that the discounted value is a function of the ratio $\frac{Q^1_\lambda}{Q^2_\lambda}$ that can oscillate as λ goes to 0, when both inertia rates converge to 0.

2) This construction reveals a common structure in two recent "counter-examples" by Vigeral [12] and Ziliotto [13] dealing with two-person zero-sum stochastic games with finite state space: compact action spaces and standard signalling in the first case, finite action spaces and signals on the state space in the second. In both cases it was proved that the family of discounted values does not converge.

2010 *Mathematics Subject Classification.* Primary: 91A15, 91A05; Secondary: 91A25.

Key words and phrases. Zero-sum games, stochastic games, dynamic games, oscillations, reversibility.

This was co-funded by the European Union under the 7th Framework Program "FP7-PEOPLE-2010-ITN", grant agreement number 264735-SADCO. The second author was partially supported by the French Agence Nationale de la Recherche (ANR) "ANR JEUDY: ANR-10-BLAN 0112" and "ANR GAGA: ANR-13-JS01-0004-01".

2. **A basic model.** A *configuration* P is defined by a general two-person zero-sum repeated game form (see [6], Chapter IV) in discrete time on a state space Ω, with a specific starting state $\overline{\omega}$ and an exit state $\omega^* \neq \overline{\omega}$. We will restrict ourselves to the three following frameworks :

A) Finite games with signals : the set of actions, states and signals are finite.

B) Games with a finite state space, compact action sets, continuous transition function, and standard signaling (observation of the moves and state by both players).

C) Games with a countable state space, finite action sets, and standard signaling. In that case we also assume that for any state and couple of actions the law of the transition has a finite support.

Let S be the stopping time of exit of $\overline{\Omega} = \Omega \setminus \{\omega^*\}$:

$$S = \min\{n \in I\!N; \omega_n = \omega^*\}$$

where ω_n is the state at stage n.

Each couple of strategies (σ, τ) of the players specifies, with the parameters of the game (initial state, transition function on states and signals), the law of S. For each evaluation $\theta = \{\theta_n; n = 1, 2, ..., \ \theta_n \geq 0, \sum_n \theta_n = 1\}$, let $d_\theta(\sigma, \tau)$ be the expected (normalized) duration spent in $\overline{\Omega}$:

$$d_\theta(\sigma, \tau) = E_{\sigma,\tau}[\sum_{n=1}^{S-1} \theta_n].$$

For each real parameters $\alpha < \beta$, consider the game $\Gamma_{\alpha,\beta}$ with payoff α in any state in $\overline{\Omega}$ and with absorbing payoff β in ω^*. Then for any evaluation θ, player 1 (the maximizer) minimizes $d_\theta(\sigma, \tau)$ since the payoff $\gamma_\theta(\sigma, \tau)$ is given by:

$$\gamma_\theta(\sigma, \tau) = \alpha d_\theta(\sigma, \tau) + \beta(1 - d_\theta(\sigma, \tau)).$$

In particular, one has:

Lemma 2.1. *For any $\alpha < \beta$ and evaluation θ the game $\Gamma_{\alpha,\beta}$ has a value v_θ and*

$$v_\theta = \alpha Q_\theta + \beta(1 - Q_\theta)$$

with $Q_\theta = \sup_\tau \inf_\sigma d_\theta(\sigma, \tau) = \inf_\sigma \sup_\tau d_\theta(\sigma, \tau)$, called the inertia rate.

Proof. The only thing to prove is the existence of v_θ. In case A) this follows from [6], Chapter IV. In case B), Proposition 5.3 and 5.4 in [9] are easily extended to any evaluation with finite length, then to the general case by approximation. Finally, in case C), under evaluations of finite length the hypotheses imply that only a finite number of states can be visited so we are back to case B). □

Here are 3 examples corresponding to a Markov Chain (0 player), a Dynamic Programming Problem (1 player) and a Stochastic Game (2 players).

In all cases $\Omega = \{\overline{\omega}, \omega^*, \omega^-\}$ and ω^- is an absorbing state. For the sake of readability we will not draw the transition probabilities from one state to itself.

2.1. **0 player.** The configuration is given is Figure 1, where a (resp. b, $1-(a+b)$) is the probability to go from $\overline{\omega}$ to ω^* (resp. to ω^-, $\overline{\omega}$) with $a, b, a+b \in [0, 1]$, and a $*$ stands for an absorbing payoff.

2.2. **1 player.** See Figure 2 ; the action set is $X = [0, 1]$ and the transition probability is given by $a(x)$ from $\overline{\omega}$ to ω^* and $b(x)$ from $\overline{\omega}$ to ω^-, where a and b are two continuous function from $[0, 1]$ to $[0, 1]$ with $a + b \in [0, 1]$.

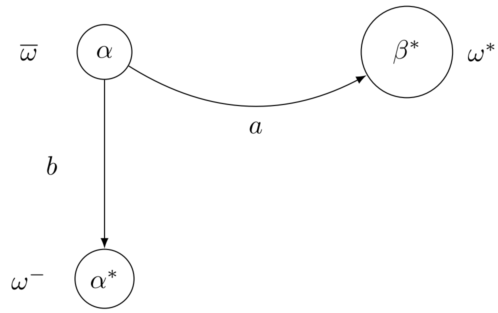

FIGURE 1.

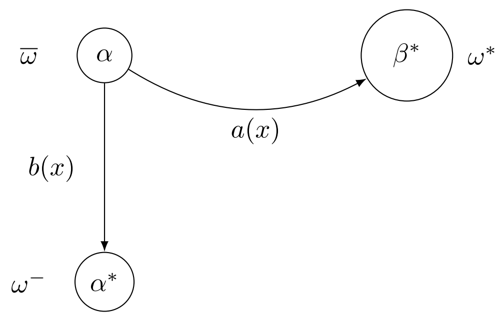

FIGURE 2.

2.3. **2 players.** In state ϖ the players have two actions and the transitions are given by:

	Stay	Quit
Stay	ϖ	ω^*
Quit	ω^*	ω^-

Let x (resp. y) be the probability on Stay and $a(x,y) = x(1-y) + y(1-x)$, $b(x,y) = xy$. The mixed extension gives the *configuration* in Figure 3. Of course one can define such a configuration for any maps a and b from $[0,1]^2$ to $[0,1]$ with $a + b \leq 1$.

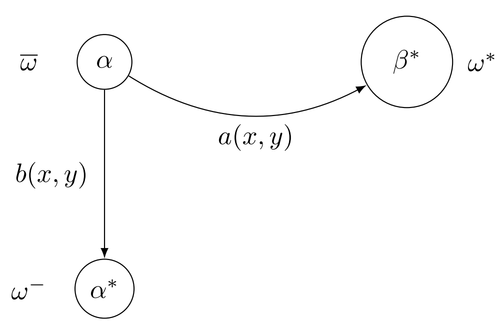

FIGURE 3.

Consider the λ-discounted case P_λ where $\theta_n = \lambda(1-\lambda)^{n-1}$. Let $r_\lambda(x,y)$ be the expected payoff induced by the stationary strategies σ and τ corresponding to x and y.

Lemma 2.2.
$$r_\lambda(x,y) = \frac{(\lambda + (1-\lambda)b(x,y)) \times \alpha + (1-\lambda)a(x,y) \times \beta}{\lambda + (1-\lambda)(a(x,y)+b(x,y))}$$

and
$$d_\theta(\sigma,\tau) = q_\lambda(x,y) = \frac{(\lambda + (1-\lambda)b(x,y))}{\lambda + (1-\lambda)(a(x,y)+b(x,y))}.$$

Proof. By stationarity:
$$r_\lambda(x,y) = \lambda \times \alpha + (1-\lambda)[a(x,y) \times \beta + b(x,y) \times \alpha + (1-a(x,y)-b(x,y)) \times r_\lambda(x,y)].$$

In particular letting:
$$q_\lambda(x,y) = \frac{(\lambda + (1-\lambda)b(x,y))}{\lambda + (1-\lambda)(a(x,y)+b(x,y))} \tag{1}$$

one has:
$$r_\lambda(x,y) = q_\lambda(x,y) \times \alpha + (1 - q_\lambda(x,y)) \times \beta.$$

□

3. **Reversibility.** Consider now a two person zero-sum dynamic game G on $\overline{\Omega}^1 \cup \overline{\Omega}^2$ generated by two dual configurations P^1 and P^2 of the previous type, which are coupled in the following sense: the exit state from P^1, ω^{*1} is the starting state $\overline{\omega}^2$ in P^2 and reciprocally. In addition the exit events are known by the players : any transition from $\overline{\Omega}^i$ to $\overline{\omega}^{-i}$ is observed by both. Finally the payoff is $\alpha^1 = -1$ on $\overline{\Omega}^1$ and $\alpha^2 = 1$ on $\overline{\Omega}^2$.

We thus obtain a reversible game (it is possible to go from $\overline{\Omega}^1$ to $\overline{\Omega}^2$ and vice versa) in which player 1 minimize (reps. maximize) the expected time spent in $\overline{\Omega}^1$ (resp. in $\overline{\Omega}^2$).

3.1. **Two examples.**

3.1.1. *Two configurations with one player in each.* There are four states $\Omega = \{\overline{\omega}^1, \overline{\omega}^2, \omega^-, \omega^+\}$.

Both ω^+ and ω^- are absorbing states with constant payoff $+1$ and -1, respectively.

The payoff in state $\overline{\omega}^i$ is also constant and equals to -1 for $i = 1$ and to $+1$ for $i = 2$. The action set for player 1 in $\overline{\omega}^1$ is $X = [0,1]$ and the transition probability is given by $a^1(x)$ from $\overline{\omega}^1$ to $\overline{\omega}^2$ and $b^1(x)$ from $\overline{\omega}^1$ to ω^-, where a^1 and b^1 are two continuous function from $[0,1]$ to $[0,1]$.

Similarly the action set for player 2 in $\overline{\omega}^2$ is $Y = [0,1]$ and $a^2(y)$ is the transition probability from $\overline{\omega}^2$ to $\overline{\omega}^1$ and $b^2(y)$ from $\overline{\omega}^2$ to ω^+. This corresponds to the coupling of configuration 2.2 and its dual, see Figure 4.

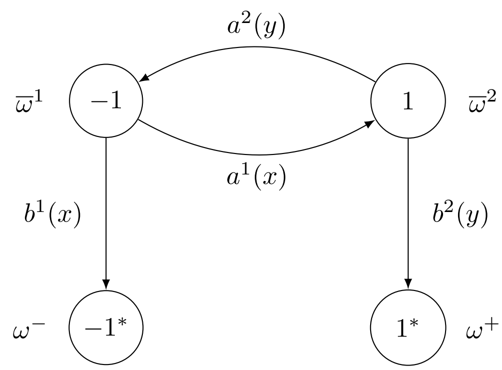

FIGURE 4.

3.1.2. *Two configurations with 2 players.* There are two absorbing states with payoff 1 and -1. In the two other states (ϖ^1 and ϖ^2) the payoff is constant and the transitions are given by the following matrices (compare to Bewley and Kohlberg [1]):

ϖ^2	Stay	Quit
Stay	1	$\underset{\rightarrow}{1}$
Quit	$\underset{\rightarrow}{1}$	1^*

ϖ^1	Stay	Quit
Stay	-1	$\underset{\leftarrow}{-1}$
Quit	$\underset{\leftarrow}{-1}$	-1^*

where an arrow means a transition to the other state.

The mixed extension corresponds to the coupling of configuration 2.3 and its dual, with $a(x, y) = x + y - 2xy$ and $b(x, y) = xy$, see Figure 5.

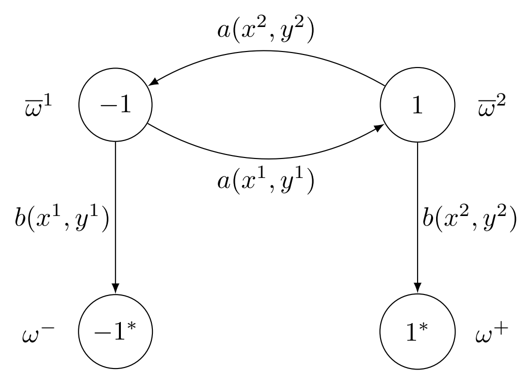

FIGURE 5.

3.2. **The discounted framework.** For each $\lambda \in]0, 1]$ the coupling between the two configurations defines a discounted game G_λ with value v_λ satisfying:

$$v_\lambda(\varpi^1) = v^1_\lambda \in [-1, 1[, \qquad v_\lambda(\varpi^2) = v^2_\lambda \in]-1, 1].$$

In particular, starting from state ϖ^1 the game is value-equivalent to the configuration where the exit state is ϖ^2 with absorbing payoff $v_\lambda(\varpi^2)$ (by observation of the exit event and stationarity of the discounted evaluation), which thus corresponds to the payoff $\beta_1 > \alpha_1$ in the configuration P^1 of the previous section 2.

Hence one obtains, using Lemma 2.1, where Q_λ stands for Q_θ when θ is the λ-discounted evaluation, that $\{v^i_\lambda\}$ is a solution of the next system of equations:

Proposition 1.

$$\begin{aligned} v^1_\lambda &= Q^1_\lambda \times (-1) + (1 - Q^1_\lambda) \times v^2_\lambda \\ v^2_\lambda &= Q^2_\lambda \times (+1) + (1 - Q^2_\lambda) \times v^1_\lambda. \end{aligned}$$

It follows that:

Corollary 1.

$$\begin{aligned} v^1_\lambda &= \frac{Q^2_\lambda - Q^1_\lambda - Q^1_\lambda Q^2_\lambda}{Q^1_\lambda + Q^2_\lambda - Q^1_\lambda Q^2_\lambda} \\ v^2_\lambda &= \frac{Q^2_\lambda - Q^1_\lambda + Q^1_\lambda Q^2_\lambda}{Q^1_\lambda + Q^2_\lambda - Q^1_\lambda Q^2_\lambda}. \end{aligned}$$

Comments:

1) In the framework of section 2.2 and 2.3 one has $Q_\lambda = \min_x \max_y q_\lambda(x, y) = \max_y \min_x q_\lambda(x, y)$ where q_λ appears in (1).

2) As λ goes to 0, Q_λ converges to 0 in the model of section 2.2, as soon as $\limsup \frac{a(x)}{b(x)} = +\infty$, as x goes to 0.

Note also that by (1), x_λ minimizes $q_\lambda(x)$ iff it minimizes

$$\rho_\lambda(x) = \frac{\lambda + (1-\lambda)b(x)}{a(x)} = \frac{1-\lambda}{1/q_\lambda(x) - 1} \tag{2}$$

and then $Q_\lambda \sim \rho_\lambda(x_\lambda)$ as soon as they both tend to 0.

3) Assuming that both Q^i_λ go to 0, the asymptotic behavior of v^1_λ depends upon the evolution of the ratio $\frac{Q^1_\lambda}{Q^2_\lambda}$. In fact one has:

$$v^1_\lambda \sim v^2_\lambda \sim \frac{1 - \frac{Q^1_\lambda}{Q^2_\lambda}}{1 + \frac{Q^1_\lambda}{Q^2_\lambda}}.$$

4) In particular one obtains:

Theorem 3.1. *Assume that $Q^i_\lambda, i = 1, 2$ go to 0 as λ goes to 0 and that $\frac{Q^1_\lambda}{Q^2_\lambda}$ has more than one accumulation point, then v^i_λ does not converge.*

More precisely it is enough that $Q^i_\lambda \sim \lambda^r f^i(\lambda)$ for some $r > 0$, with $0 < A \le f^i \le B$ and that exactly one of the $f^i(\lambda)$ does not converge as λ goes to 0, to obtain the result.

The next sections 4 and 5 will describe several models generating such probabilities Q^i_λ, with f^i converging or not.

We will use the terminology *regular or oscillating configurations.*

The above result implies that by coupling any two of these configurations (of the same order of magnitude r) where one is oscillating, one can generate a game for which the family of discounted values does not converge, see Section 6.

4. **Some regular configurations of order $\frac{1}{2}$.** We give here three examples of regular configurations of order $\frac{1}{2}$.

4.1. **A regular configuration with 0 players and countable state space.** Consider a random walk on $\mathbf{N} \cup \{-1\}$ and exit state -1. In any other state $m \in \mathbf{N}$ the transition is $\frac{1}{2}\delta_{m-1} + \frac{1}{2}\delta_{m+1}$. The starting state is 0. Denote by s_n the probability that exit occurs at stage n ; it is well known (Theorem 5b p. 164 in [3]) that the generating function of S is given by $F(z) = \frac{1-\sqrt{1-z^2}}{z}$. Hence,

$$\begin{aligned} Q_\lambda &= \sum_{n=1}^{+\infty} s_n \sum_{t=1}^{n} \lambda(1-\lambda)^{i-1} \\ &= \sum_{n=1}^{+\infty} s_n(1-(1-\lambda)^n) \\ &= F(1) - F(1-\lambda) \\ &= \frac{\sqrt{2\lambda - \lambda^2} - \lambda}{1-\lambda} \\ &\sim \sqrt{2\lambda}. \end{aligned}$$

4.2. **A regular configuration with one player, finitely many states, compact action space and continuous transition.** Consider example 2.2.

Take $a(x) = x$ and $b(x) = x^2$. Then $Q_\lambda = \min_x \{\frac{\lambda + (1-\lambda)x^2}{\lambda + (1-\lambda)x^2 + (1-\lambda)x}\}$ and a first order condition gives $x_\lambda = \sqrt{\frac{\lambda}{1-\lambda}}$ hence $Q_\lambda \sim 2\sqrt{\lambda}$.

4.3. **A regular configuration with two players and finitely many states and actions.** Consider example 2.3.

It is straightforward [12] to compute that in Γ_λ the optimal strategy for each player is $x_\lambda = y_\lambda = \frac{\sqrt{\lambda}}{1+\sqrt{\lambda}}$. Hence:

$$\begin{aligned} Q_\lambda &= \frac{\lambda + (1-\lambda)x_\lambda y_\lambda}{\lambda + (1-\lambda)(x_\lambda + y_\lambda - x_\lambda y_\lambda)} \\ &\sim \sqrt{\lambda}. \end{aligned}$$

4.4. **Remarks.** These configurations are in a certain sense minimal ones. Any configuration with one player, with finitely many states and actions and full observation is, by Blackwell optimality, asymptotically equivalent to a finite Markov chain. And in any such chain,

- either with positive probability the state never exits $\overline{\Omega}$, and Q_λ is of order 0.
- or at each stage, given no prior exit, an exit occurs in the next m stages with probability at least p, where m and $p > 0$ are fixed. This implies that Q_λ is of order 1.

5. Some oscillating configurations of order $\frac{1}{2}$.

5.1. **Example 4.2. perturbed.** Recall that the choice of $a(x) = x$ and $b(x) = x^2$ leads to $Q_\lambda \sim 2\sqrt{\lambda}$.

To get oscillations one can choose $b = x^2$ and $a(x) = xf(x)$ with $f(x)$ bounded away from 0, oscillating and such that $f'(x) = o(1/x)$. For example, $f(x) = 2 + \sin(ln(-lnx))$.

Proposition 2. *For this choice of transition functions one has:*

$$Q_\lambda \sim \frac{2\sqrt{\lambda}}{f(\sqrt{\lambda})}.$$

Proof. Using Comment 2 after Corollary 1, the first order condition in (2) gives:

$$\frac{\lambda}{1-\lambda} = \frac{x^2(f(x) - xf'(x))}{f(x) + xf'(x)}$$

which leads to:

$$x_\lambda \sim \sqrt{\lambda}.$$

By the mean value theorem and since $f'(x) = o(1/x)$,

$$\frac{\|f(x_\lambda) - f(\sqrt{\lambda})\|}{\|x_\lambda - \sqrt{\lambda}\|} = o\left(\frac{1}{\sqrt{\lambda}}\right)$$

hence $f(x_\lambda) \sim f(\sqrt{\lambda})$ and:

$$Q_\lambda \sim \frac{2\sqrt{\lambda}}{f(\sqrt{\lambda})}.$$

In particular $\frac{Q_\lambda}{\sqrt{\lambda}}$ has no limit. □

5.2. **Example 4.3. perturbed.** Let $s \in C^1(]0, \frac{1}{16}], \mathbb{R})$ such that s and $x \to xs'(x)$ are both bounded by $\frac{1}{16}$. Consider a configuration as in Figure 3 but for perturbed functions a and b:

$$\begin{aligned}
a(x,y) &= \frac{(\sqrt{x}+\sqrt{y})(1-\sqrt{x}+s(x))(1-\sqrt{y}+s(y)}{2(1-x)(1-y)(1-f_2(x,y))} \\
b(x,y) &= \frac{\sqrt{xy}\left[(1-\sqrt{x})(1-\sqrt{y}) + f_1(x,y) - \sqrt{xy}f_2(x,y)\right]}{(1-x)(1-y)(1-f_2(x,y))}
\end{aligned}$$

where

$$f_1(x,y) = \begin{cases} \frac{\sqrt{x}s(x)-\sqrt{y}s(y)}{\sqrt{x}-\sqrt{y}} & \text{if } x \neq y \\ 2xs'(x) + s(x) & \text{if } x = y \end{cases}$$

and

$$f_2(x,y) = \begin{cases} \frac{\sqrt{y}s(x)-\sqrt{x}s(y)}{\sqrt{x}-\sqrt{y}} & \text{if } x \neq y \\ 2xs'(x) - s(x) & \text{if } x = y. \end{cases}$$

Then a and b are continous (Lemma 12 and Lemma 10 in [12]), the game with payoff q_λ has a value in pure strategies and $x_\lambda = y_\lambda = \lambda$ are optimal [12]. Hence:

$$\begin{aligned}
Q_\lambda &= \frac{\frac{\lambda}{1-\lambda} + b(\lambda,\lambda)}{\frac{\lambda}{1-\lambda} + b(\lambda,\lambda) + a(\lambda,\lambda)} \\
&\sim \frac{\lambda + \frac{\lambda(1+s(\lambda)+2\lambda s'(\lambda))}{1+s(\lambda)-2\lambda s'(\lambda)}}{\frac{2\sqrt{\lambda}(1+s(\lambda))^2}{2(1+s(\lambda)-2\lambda s'(\lambda))}} \\
&\sim \frac{\lambda(1+s(\lambda)-2\lambda s'(\lambda)+1+s(\lambda)+2\lambda s'(\lambda))}{\sqrt{\lambda}(1+s(\lambda))^2} \\
&\sim \frac{2\sqrt{\lambda}}{1+s(\lambda)}.
\end{aligned}$$

The configuration is thus oscillating for $s(x) = \frac{\sin \ln x}{16}$ for example.

Next we recall 4 models that appear in Ziliotto [13] (in which the divergence of v_λ was proven) and we compute the corresponding Q_λ.

5.3. **Countable action space.** Consider again the Example 2.2 but assume now that the action space X is a countable subset of $[0,1]$: $X = \{0\} \cup \{\frac{1}{2^n}, n \in \mathbb{N}^*\}$. The transition are still given by $(a(x), b(x)) = (x, x^2)$.

Proposition 3. *For this configuration $Q_\lambda/\sqrt{\lambda}$ oscillates on a sequence $\{\lambda_m\}$ of discount factors like $\lambda_m = \frac{1}{2^m}$.*

Proof. Use Comment 2 and recall by (2) that $\rho_\lambda(x) = \frac{\lambda}{x} + (1-\lambda)x$ which is strictly convex.

For $\lambda = \frac{1}{4^n}$, $\rho_\lambda(\frac{1}{2^n}) \sim 2\sqrt{\lambda}$ while $\rho_\lambda(\frac{1}{2^{n+1}}) \sim \rho_\lambda(\frac{1}{2^{n-1}}) \sim 5/2\sqrt{\lambda}$, hence $Q_\lambda \sim 2\sqrt{\lambda}$.

On the other hand, for $\lambda^2 = \frac{1}{4^n}\frac{1}{4^{n+1}}$ one obtains:

$$\begin{aligned}\rho_\lambda(\frac{1}{2^n}) &\sim (\frac{1}{2} \times \frac{1}{4^n} + \frac{1}{4^n})2^n \\ &\sim \frac{3\sqrt{2}}{2}\sqrt{\lambda}\end{aligned}$$

and similarly:

$$\begin{aligned}\rho_\lambda(\frac{1}{2^{n+1}}) &\sim (\frac{1}{2} \times \frac{1}{4^n} + \frac{1}{4^{n+1}})2^{n+1} \\ &\sim \frac{3\sqrt{2}}{2}\sqrt{\lambda}.\end{aligned}$$

Finally one checks that $\rho_\lambda(\frac{1}{2^{n+m}}) \geq \frac{3\sqrt{2}}{2}\sqrt{\lambda}$ for $m = -n, ..., -1$ and $m \geq 2$. Thus for this specific λ, $Q_\lambda(x) \sim \frac{3\sqrt{2}}{2}\sqrt{\lambda}$.

It follows that $Q_\lambda/\sqrt{\lambda}$ oscillates between 2 and $\frac{3\sqrt{2}}{2}$ on a sequence $\{\lambda_m\}$ of discount factors like $\lambda_m = \frac{1}{2^m}$. □

Note that this result is conceptually similar to example 5.1.

5.4. **Countable state space.** We consider here a configuration which is the dual of the previous one: finite action space and countable state space.

The state space is a countable family of probabilities $y = (y^A, y^B)$ on two positions A and B with $y_n = (\frac{1}{2^n}, 1 - \frac{1}{2^n}), n = 0, 1, ...$, and two absorbing states A^* and B^*.

The player has two actions: *Stay* or *Quit*. Consider state y_n. Under *Quit* an absorbing state is reached: A^* with probability y_n^A and B^* with probability y_n^B. Under *Stay* the state evolves from y_n to y_{n+1} with probability 1/2 and to $y_0 = (1, 0)$ with probability 1/2.

The player is informed upon the state, the starting state is y_0 and the exit state is B^*.

A strategy of the player can be identified with a stopping time corresponding to the first state y_n when he chooses *Quit*.

Let T_n be the random time corresponding to the first occurrence of y_n (under *Stay*) and μ_n the associated strategy: *Quit* (for the first time) at y_n. This strategy plays the role of $x = 1/2^n$ in the previous example.

Proposition 4. *Under μ_n the λ-discounted normalized duration before B^* is*

$$q_\lambda(n) = 1 - \frac{(1-\lambda^2)\left(1-\frac{1}{2^n}\right)}{1+2^{n+1}\lambda(1-\lambda)^{-n}-\lambda}.$$

Proof. Lemma 2.5 in Ziliotto [13] gives

$$\mathsf{E}[(1-\lambda)^{T_n}] = \frac{1-\lambda^2}{1+2^{n+1}\lambda(1-\lambda)^{-n}-\lambda}$$

and

$$q_\lambda(n) = 1 + (\frac{1}{2^n} - 1)\mathsf{E}[(1-\lambda)^{T_n}].$$

□

Proposition 5. *The configuration is irregular : $\frac{Q_\lambda}{\sqrt{\lambda}}$ oscillates between two positive values.*

Proof. With our notations, Ziliotto's Lemma 2.8 [13] states that

$$q_\lambda\left(-\frac{\ln\lambda + \ln 2 + 2\ln c}{2\ln 2}\right) \sim (c+c^{-1})\sqrt{2\lambda}$$

hence,

$$\frac{Q_\lambda}{\sqrt{2\lambda}} \sim \min\left\{c + c^{-1} \text{ such that } -\frac{\ln\lambda + \ln 2 + 2\ln c}{2\ln 2} \in \mathbf{N}\right\}.$$

When $-\frac{\ln\lambda+\ln 2}{2\ln 2}$ is an integer, one can take $c = 1$ which gives $Q_\lambda \sim 2\sqrt{2\lambda}$. Whereas when $-\frac{\ln\lambda+\ln 2}{2\ln 2}$ is an integer plus one half, the best choice is $c = \sqrt{2}$, leading to $Q_\lambda \sim 3\sqrt{\lambda}$. □

5.5. **A MDP with signals.** The next configuration corresponds to a Markov decision process with 2 states: A and B, 2 absorbing states A^* and B^* and with signals on the state. The player has 2 actions: *Stay* or *Quit*. The transitions are as follows:

A	$\frac{1}{2};\ell$	$\frac{1}{2};r$
Stay	A	$(\frac{1}{2}A + \frac{1}{2}B)$
Quit	A^*	A^*

B	$\frac{1}{2};\ell$	$\frac{1}{2};r$
Stay	A	B
Quit	B^*	B^*

Hence the transition is random: with probability 1/2 of type ℓ and probability 1/2 of type r. The player is not informed on the state reached but get only the signal ℓ or r.

The natural "auxiliary state" space is then the beliefs of the player on (A, B) and one can check [13] that the model is equivalent to the previous one, starting from A and where the exit state is B^*. In fact under Stay, ℓ occurs with probability 1/2 and the new parameter is $y_0 = (1, 0)$. On the other hand, after r the belief goes from y_n to y_{n+1}.

Again this configuration generates an oscillating Q_λ of the order of $\sqrt{\lambda}$. The crucial point here is that the belief evolves in a countable subset of $\Delta(\Omega)$, and in a reversible way.

5.6. **A game in the dark.** A next transformation is to introduce two players and to generate the random variable $\frac{1}{2}(\ell) + \frac{1}{2}(r)$ in the above configuration by a process induced by the moves of the players.

This leads to the original framework of the game defined by Ziliotto [13]: action and state spaces are finite and the only information of the players is the initial state and the sequence of moves along the play.

Player 1 has three moves: Stay1, Stay2 and Quit, and player 2 has 2 moves: Left and Right. The transitions are as follows:

A	$Left$	$Right$
Stay1	A	$(\frac{1}{2}A + \frac{1}{2}B)$
Stay2	$(\frac{1}{2}A + \frac{1}{2}B)$	A
Quit	A^*	A^*

B	$Left$	$Right$
Stay1	A	B
Stay2	B	A
Quit	B^*	B^*

By playing $(1/2, 1/2, 0)$ (resp. $(1/2, 1/2)$) player 1 (resp. player 2) can mimick the previous distribution on (ℓ, r) where ℓ corresponds to the event "the moves are on the main diagonal". It follows that this behavior is consistent with optimal strategies hence the induced distribution on plays is like in the previous example 5.5.

6. **Combinatorics.** In order to obtain oscillations for the discounted values of a stochastic game, it is enough to consider the coupled dynamics generated by a regular and an oscillating configuration, both of order $\frac{1}{2}$.

6.1. **Example 4.2 + Example 5.1.** Combining these two configurations yields a coupling of two one-person decision problems, hence a compact stochastic game with perfect information and no asymptotic value. Explicitly the game is given for example by Figure 6.

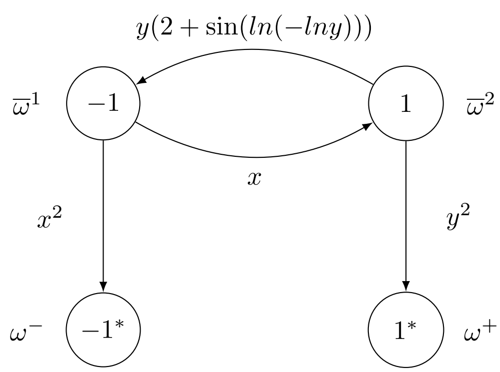

FIGURE 6.

Remark that the transition functions can be taken as smooth as one wants.

6.2. **Examples 4.2 + Example 5.3.** With this combination one recovers exactly an example of Ziliotto (see section 4.2 in [13]) which is also a stochastic game with perfect information and no asymptotic value. The main difference is that in that case the action space of player 1 is countable instead of being an interval.

6.3. **Example 4.3 + Example 5.1.** Combining these two configurations yields a stochastic game with finite action space for player 2 and no asymptotic value. Here also the transition functions can be taken as smooth as one wants.

6.4. **Example 5.2 + 5.2.** By coupling Example 5.2 with a similar configuration controlled by the other player, one recovers exactly the family of counterexamples in [12]. Note than in this case both configurations are oscillating, but with a different phase so the ratio does not converge.

6.5. **Example 5.4 + 5.4, 5.5 + 5.5 and 5.6 + 5.6.** Three examples of Ziliotto ([13], sections 2.1 2.2 and 4.1) are combinations of either 5.4, 5.5 or 5.6 with a similar configuration. In those cases both configurations are oscillating of order $\frac{1}{2}$ but one is oscillating twice as fast as the other hence the oscillations of v_λ in the combined game.

6.6. **Example 4.1 + 5.4.** This gives a MDP with a countable number of states (and only 2 actions) in which v_λ does not converge. Observe that one can compactify the state space in such a way that both the payoff and transition functions are continuous.

7. Comparison and conclusion.

7.1. **Irreversibility.** The above analysis shows that oscillations in the inertia rate and reversibility allows for non convergence of the discounted values.

These two properties seem to be also necessary. In fact, Sorin and Vigeral [11] prove the convergence of the discounted values for stochastic games with finite state space, continuous action space, continuous payoffs and transitions, in the two following framework: absorbing games (see also [4, 5, 8]) and recursive games (see also [10]). These two classes corresponds to the "irreversible" case where once one leaves a state, it cannot be reached again.

A similar property holds for incomplete information games where the irreversible aspect is due to the martingale property of the sequence of states (beliefs).

7.2. **Oscillations.** Remark that any oscillating configuration of Section 5 leads, under optimal play, to an almost immediate exit. Hence, by itself, any such configuration leads to a regular asymptotic behavior. It is only the "resonance" between two configurations that yields asymptotic issues.

7.3. **Semi-algebraic.** For stochastic games with finitely many states and full monitoring, all the examples of the previous section lack a semi-algebraic structure: either transition functions oscillate infinitely often or a set of actions has infinitely many connected components. While the existence of an asymptotic value with semi-algebraic parameters in the case of either perfect information or finitely many actions on one side holds [2], it is not known in full generality. In particular, an interesting question is to determine whether there exists a configuration with semi-algebraic parameters such that Q_λ is not semi-algebraic.

7.4. **Related issues.** The stationarity of the model is crucial here. However it is possible to construct similar examples in which $\lim v_n$ does not exist. The idea grounds on a lemma of Neyman [7] giving sufficient conditions for the two sequences v_n and v_{λ_n} for $\lambda_n = \frac{1}{n}$, to have the same asymptotic behavior as n tends to infinity. See [12] for specific details in the framework of sections 5.1 and 5.2 and [13] in the framework of sections 5.3-5.6.

REFERENCES

[1] T. Bewley and E. Kohlberg, On stochastic games with stationary optimal strategies, *Mathematics of Operations Research*, **3** (1978), 104–125.

[2] J. Bolte, S. Gaubert and G. Vigeral, Definable zero-sum stochastic games, *Mathematics of Operation Research*, **40** (2015), 171–191.

[3] G. Grimmett and D. Stirzaker, *Probability and Random Processes,* Oxford University Press, 2001.

[4] R. Laraki, Explicit formulas for repeated games with absorbing states, *International Journal of Game Theory*, **39** (2010), 53–69.

[5] J.-F. Mertens, A. Neyman and D. Rosenberg, Absorbing games with compact action spaces, *Mathematics of Operation Research*, **34** (2009), 257–262.

[6] J.-F. Mertens, S. Sorin and S. Zamir, *Repeated Games,* Cambridge University Press, 2015.

[7] A. Neyman, Stochastic games and nonexpansive maps, in *Stochastic Games and Applications* (eds. A. Neyman and S. Sorin), NATO Sci. Ser. C Math. Phys. Sci., 570, Kluwer Academic Publishers, Dordrecht, 2003, 397–415.

[8] D. Rosenberg and S. Sorin, An operator approach to zero-sum repeated games, *Israel Journal of Mathematics*,**121** (2001), 221–246.

[9] S. Sorin, *A First Course on Zero-SumRepeated Games*, Springer-Verlag, 2002.

[10] S. Sorin, The operator approach to zero-sum stochastic games, in *Stochastic Games and Applications* (eds. A. Neyman and S. Sorin), NATO Sci. Ser. C Math. Phys. Sci., 570, Kluwer Academic Publishers, Dordrecht, 2003, 417–426.

[11] S. Sorin and G. Vigeral, Existence of the limit value of two person zero-sum discounted repeated games via comparison theorems, *Journal of Opimization Theory and Applications*, **157** (2013), 564–576.

[12] G. Vigeral, A zero-sum stochastic game with compact action sets and no asymptotic value, *Dynamic Games and Applications*, **3** (2013), 172–186.

[13] B. Ziliotto, Zero-sum repeated games: Counterexamples to the existence of the asymptotic value and the conjecture maxmin=lim v_n, to appear in *Annals of Probability*, 2013. Available from: `https://hal.archives-ouvertes.fr/hal-00824039`.

Received November 2014; revised January 2015.

E-mail address: sylvain.sorin@imj-prg.fr
E-mail address: vigeral@ceremade.dauphine.fr