Perturbations of Markov Chains with Applications to Stochastic Games

Eilon Solan[*]

June 14, 2000

In this lecture we will review several topics that are extensively used in the study of n-player stochastic games. These tools were used in the proof of several results on non zero-sum stochastic games.

Most of the results that are presented here appeared in Vieille (1997a,b), and some appeared in Solan (1998, 1999).

The first main issue is Markov chains where the transition rule is a Puiseux probability distribution. We define the notion of communicating sets and induce a hierarchy on the collection of these sets. We then relate these concepts to stochastic games, and show several conditions that enable the players to control exit distributions from communicating sets.

1 Markov Chains

A *Markov chain* is a pair (K, p) where K is a finite set of states, and $p : K \to \Delta(K)$ is a transition rule. (As usual, $\Delta(K)$ stands for the set of probability distributions over K.)

The transition rule p together with an initial state k define a process on the states. Denote by k_n the state of the process at stage n, $n = 1, 2, \ldots$. Let $\mathbf{P}_{k_1,p}$ be the probability distribution induced by p and the initial state k_1 over the space of infinite histories.

A subset $C \subseteq K$ is *ergodic* if for every $k \in C$

[*] Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston IL 60208. e-mail: e-solan@nwu.edu. I thank Jerzi Filar, Sylvain Sorin and Nicolas Vieille for their comments.

1. $\sum_{k' \in C} p(k, k') = 1$.
2. For every $k' \in C$, $\mathbf{P}_{k_1,p}(k_n = k' \text{ for some } n \geq 1) = 1$.

Let $A = A(p) = \{k \in K \mid p(k,k) = 1\}$ be the set of absorbing states. In this section we consider only transition rules that satisfy the following two assumptions:

A.1 $A \neq \emptyset$.

A.2 $\mathbf{P}_{k,p}(\exists n \geq 1 \text{ s.t. } k_n \in A) = 1$ for every initial state $k \in K$.

We define the *arrival time* by

$$r_l = \inf\{n > 1 \mid k_n = l\}$$

where an infimum over an empty set is infinity. For a subset $B \subseteq K \setminus A$ and a state $k_1 \in B$ we define the *exit time* from B by

$$e_B = \inf\{n \geq 1 \mid k_n \notin B\}.$$

By A.2, e_B is finite a.s. Let $Q_p^l(B) = \mathbf{P}_{k_1,p}(k_{e_B} = l)$ be the probability that the first state outside B the process visits is l. Clearly this probability depends on the initial state. We denote by $Q_p(B) = (Q_p^l(B))_{l \in K}$ the exit distribution from B. Since e_B is finite a.s., this is indeed a probability distribution.

A *B-graph* is a set of pairs $g = \{[k \to l] \mid k \in B, l \in K\}$ such that

- For each $k \in B$ there is a unique $l \in B$ with $[k \to l] \in g$.
- g has no cycle; that is, there are no positive integer J and $k_1, \ldots, k_J \in B$ with $[k_j \to k_{j+1}] \in g$ for every $j = 1, \ldots, J$ (addition modulo J).

It is clear that for every $k \in B$ there exists a *unique* $l \notin B$ such that $[k \to k_1], [k_1 \to k_2], \ldots, [k_J \to l] \in g$ for some J and $k_1, \ldots, k_J$. In such a case we say that k *leads* to l in g.

We denote by G_B the set of all B-graphs, and by $G_B(k \to l)$ all the B-graphs in which k leads to l.

Example 1: $K = \{1, 2, a, b\}$, $p(a,a) = p(b,b) = 1$, $p(1,2) = p(1,a) = 1/2$ and $p(2,1) = 1 - p(2,b) = 3/4$. Thus, $A(p) = \{a, b\}$ and the process reaches an absorbing state in finite time p-a.s. Graphically, the Markov chain looks as follows.

Figure 1

Take $B = \{1, 2\}$. Then there are three B-graphs: $g_1 = \{[1, 2], [2, b]\}$, $g_2 = \{[1, a], [2, 1]\}$ and $g_3 = \{[1, a], [2, b]\}$. $G_B(1 \to a) = \{g_2, g_3\}$, $G_B(1 \to b) = \{g_1\}$, $G_B(2 \to b) = \{g_1, g_3\}$ and $G_B(2 \to a) = \{g_2\}$.

The *weight* of g w.r.t. p is

$$p(g) = \prod_{[k \to l] \in g} p(k, l).$$

Lemma 1.1 (Freidlin and Wentzell (1984)) *If $k_1 \in B$ and $l \notin B$,*

$$Q_p^l(B) = \frac{\sum_{g \in G_B(k_1 \to l)} p(g)}{\sum_{g \in G_B} p(g)}.$$

Assumptions A.1-A.2 imply that the denominator is positive.

Lemma 1.1 implies that $Q_p^l(B)$ is continuous as a function of the transition rule p.

Example 1 (continued): It is easy to verify that $p(g_1) = p(g_3) = 1/8$ and $p(g_2) = 3/8$. One can now calculate, using Lemma 1.1, that if $k_1 = 1$ then $Q_p^a(B) = 4/5$, while if $k_1 = 2$ then $Q_p^a(B) = 3/5$.

2 Puiseux Markov Chains

Puiseux series were introduced to the study of stochastic games by Bewley and Kohlberg (1976). Since Puiseux series form a real closed field, they proved to be a useful tool in analyzing asymptotic properties of discounted stochastic games. The asymptotic properties where used by Mertens and

Neyman (1981) to prove the existence of the undiscounted value in zero sum games, and by Solan (1999) and Solan and Vieille (1998) for n-player stochastic games. Puiseux series were used in other fields as well (see, e.g., Eaves and Rothblum (1989)).

All the definitions and results we have stated in section 1 do not use the fact that the field over which the transition rule is defined is the field of real numbers. Consider now the field $\mathcal{F}$ of *Puiseux functions*; that is, all function $\hat{f} : (0,1) \to \mathbf{R}$ that have an expansion

$$\hat{f}_\epsilon = \sum_{i=L}^{\infty} a_i \epsilon^{i/M}$$

for some integer L and positive integer M in an open neighborhood of 0. As a rule, Puiseux functions are denoted with a hat. The *valuation* of a Puiseux function is defined by $\mathrm{w}(\hat{f}) = \min\{i \mid a_i \neq 0\}/M$.

For every Puiseux function $\hat{f}$ with $\mathrm{w}(\hat{f}) \geq 0$ define

$$\hat{f}_0 = \lim_{\epsilon \to 0} \hat{f}_\epsilon = \begin{cases} 0 & \mathrm{w}(\hat{f}) > 0 \\ a_0 & \mathrm{w}(\hat{f}) = 0 \end{cases}$$

It is easy to verify that

$$\mathrm{w}(\hat{f}\hat{g}) = \mathrm{w}(\hat{f}) + \mathrm{w}(\hat{g}), \tag{1}$$

and that

$$\lim_{\epsilon \to 0} (\hat{f}_\epsilon / \hat{g}_\epsilon) = 0 \text{ whenever } \mathrm{w}(\hat{f}) > \mathrm{w}(\hat{g}). \tag{2}$$

A *Puiseux transition rule* is a function $\hat{p} : K \times K \to \mathcal{F}$ such that (i) for every $k, l \in K$, $\hat{p}(k,l)$ is a non-negative Puiseux function and $\sum_{l \in K} \hat{p}(k,l) = 1$, and (ii) for every $\epsilon \in (0,1)$, $\hat{p}_\epsilon(\cdot,\cdot)$ is a transition rule. A *Puiseux Markov chain* is a pair $(K, \hat{p})$ where K is a finite set, and $\hat{p} : K \times K \to \mathcal{F}$ is a Puiseux transition rule. Note that the valuation of $\hat{p}(k,l)$ is non-negative for every $k, l \in K$.

An important property of Puiseux functions is that if a Puiseux function has infinitely many 0's in any neighborhood of 0, then it is the zero function. In particular, if a Puiseux function is not zero, then it is non-zero in a neighborhood of 0. Therefore, in a neighborhood of 0, the ergodic structure of a Puiseux Markov chain (and the collection of absorbing states) is independent of ϵ.

In the sequel we will consider Puiseux transition rules $\hat{p}$ such that for every ϵ sufficiently small, $\hat{p}_\epsilon$ satisfies assumptions A.1 and A.2.

The weight of a B-graph is a Puiseux function $\hat{p}(g) = \prod_{[k\to l]\in g} \hat{p}(k,l)$. From (1) it follows that $\mathrm{w}(\hat{p}(g)) = \sum_{[k\to l]\in g} \mathrm{w}(\hat{p}(k,l))$.

Since Puiseux functions form a field, it follows that for every Puiseux transition rule $\hat{p}$, $Q^l_{\hat{p}}(B)$ is a Puiseux function. In particular, the limit $\lim_{\epsilon\to 0} Q_{\hat{p}_\epsilon}(B)$ exists, and is a probability distribution.

Define $G^{\min}_B(k \to l)$ to be the collection of all B-graphs $g \in G_B(k \to l)$ that have the minimal valuation. By (2) it follows that if $k_1 \in B$ then

$$\lim_{\epsilon\to 0} Q^l_{\hat{p}_\epsilon}(B) = \lim_{\epsilon\to 0} \frac{\sum_{g\in G^{\min}_B(k_1\to l)} \hat{p}_\epsilon(g)}{\sum_{g\in G^{\min}_B} \hat{p}_\epsilon(g)}. \tag{3}$$

3 Communicating Sets

Bather (1973) introduced the notion of communicating sets to the theory of Markov chains: a set B is *communicating* if for every $k,l \in B$, l is accessible from k (that is, $\mathbf{P}_{k,p}(r_l < +\infty) > 0$). A communicating set B is *closed* if whenever $k \in B$ and l is accessible from k, $l \in B$ as well.

Ross and Varadarajan (1991) defined another notion of communication. A set B in a Markov decision process is *strongly communicating* if it is recurrent under some transition rule.

Avṣar and Baykal-Gürsoy (1999) generalized the definition of strongly communicating sets to stochastic games. However, contrary to their claim (compare their Lemma 1 and Example 2 below), under their definition, two strongly communicating sets may have non-trivial intersection.

In the present section we generalize Bather's definition of communicating sets to Puiseux Markov chains. In the next section we provide another definition of communicating sets for stochastic games. When reduced to Markov chains, this definition coincides with that given by Ross and Varadarajan. We then study the relation between the two definitions.

Let $(K, \hat{p})$ be a Puiseux Markov chain.

Definition 3.1 *A set $B \subseteq K \setminus A$ is* communicating w.r.t. $\hat{p}$ *if for every $k, k' \in B$*

$$\lim_{\epsilon\to 0} \mathbf{P}_{k,\hat{p}_\epsilon}(e_B < r_{k'}) = 0.$$

That is, the probability that the process leaves B before it reaches any state in B goes to 0. Equivalently, as $\epsilon \to 0$, the number of times the process visits any state in B before leaving B increases to ∞.

We denote by $\mathcal{C}(\hat{p})$ the collection of all communicating sets w.r.t. $\hat{p}$. Note that if $C \in \mathcal{C}(\hat{p})$ is communicating, if $B \subset C$ and if $k_1 \in C \setminus B$ then

$$\lim_{\epsilon \to 0} \sum_{l \in B} Q^l_{\hat{p}_\epsilon}(C \setminus B) = 1. \tag{4}$$

Define a hierarchy (or a partial order) on $\mathcal{C}(\hat{p})$ by set inclusion. Definition 3.1 implies that two communicating sets are either disjoint or one is a subset of the other. Hence the directed graph of this partial order is a forest (a collection of disjoint trees). A similar hierarchy was already studied by Ross and Varadarajan (1991), and a different type of hierarchy is used in Avşar and Baykal-Gürsoy (1999).

Let B and C be communicating sets w.r.t. $\hat{p}$. B is a *child* of C if there is no communicating set D that satisfies $B \subset D \subset C$. Equivalently, B is a child of C if it is its child in the corresponding tree (when we represent the partial order as a forest).

Definition 3.1 implies the following.

LEMMA 3.2 *If B is communicating w.r.t. $\hat{p}$ then $\lim_{\epsilon \to 0} Q_{\hat{p}_\epsilon}(B)$ is independent of k_1, provided $k_1 \in B$.*

For every $B \in \mathcal{C}(\hat{p})$, the limit $Q^*_{\hat{p}}(B) = \lim_{\epsilon \to 0} Q_{\hat{p}_\epsilon}(B)$, which is independent of $k_1 \in B$, is the *exit distribution from B* (w.r.t. $\hat{p}$).

Let C be a communicating set, and let $D_1, \ldots, D_L$ be the children of C. Define a new Markov chain as follows:

- The state space is $\{d_1, \ldots, d_L\} \cup (K \setminus \cup_l D_l)$.
- The transition q is given as follows:
  - $q(k, k') = \hat{p}_0(k, k')$ for $k, k' \notin \cup_l D_l$.
  - $q(k, d_l) = \sum_{k' \in D_l} \hat{p}_0(k, k')$ for $k \notin \cup_l D_l$.
  - $q(d_l, k') = Q^{*,k'}_{\hat{p}}(D_l)$ for $k' \notin \cup_l D_l$.
  - $q(d_l, d_{l'}) = \sum_{k' \in D_{l'}} Q^{*,k'}_{\hat{p}}(D_l)$.

Thus, we replace each maximal communicating subset of C by a single state. Transitions from those new states are given by the exit distribution, whereas transitions from states that are not in any communicating set (transient states) are given by the limit probability distribution $\hat{p}_0$.

Eq. (4) implies the following.

LEMMA 3.3 *C is ergodic in (K, q).*

4 Stochastic Games

From now on we concentrate on stochastic games, and we study when an exit distribution from a communicating set can be controlled by the two players.

Let (S, A, B, r, p) be a two-player stochastic game.

We denote by $\mathbf{P}_{z,\sigma,\tau}$ the probability distribution over the space of infinite histories induced by the initial state z and the strategy pair (σ, τ), and by $\mathbf{E}_{z,\sigma,\tau}$ the corresponding expectation operator.

DEFINITION 4.1 *A* Puiseux strategy *for player 1 is a function $\hat{\alpha} : (0,1) \times S \to \Delta(A)$ such that for every $z \in S$, $\hat{\alpha}_\epsilon(z)$ is a Puiseux probability distribution. Puiseux strategies for player 2 are defined analogously.*

Note that for every $\epsilon \in (0,1)$, $\hat{\alpha}_\epsilon$ is a stationary strategy of player 1.

Any pair of Puiseux strategies $(\hat{\alpha}, \hat{\beta})$ defines a Markov chain over S with Puiseux transition rule $\hat{q}$:

$$\hat{q}(z, z') = \sum_{a,b} \hat{\alpha}^a(z) \hat{\beta}^b(z) p(z' | z, a, b).$$

In particular, with every pair of Puiseux strategies $(\hat{\alpha}, \hat{\beta})$ we can associate the collection of communicating sets $\mathcal{C}(\hat{\alpha}, \hat{\beta})$ and the corresponding hierarchy.

For every $C \in \mathcal{C}(\hat{\alpha}, \hat{\beta})$ we denote by $Q^*_{\hat{\alpha},\hat{\beta}}(C)$ the exit distribution from C in the corresponding Puiseux Markov chain.

A weaker definition of communication in stochastic games is the following.

DEFINITION 4.2 *Let (α, β) be a pair of stationary strategies, and $C \subset S$. C is* weakly communicating *w.r.t. (α, β) if for every $z \in C$ and every $\delta > 0$ there exists a pair of stationary strategies (α', β') such that*

1. *$\| (\alpha', \beta') - (\alpha, \beta) \| < \delta$.*

2. *C is stable under (α', β'); that is, $p(C \mid z', \alpha', \beta') = 1$ for every $z' \in C$.*

3. *$\mathbf{P}_{z',\alpha',\beta'}(z_n = z$ for some $n \geq 1) = 1$ for every $z' \in C$.*

We denote by $\mathcal{D}(\alpha, \beta)$ the set of weakly communicating sets w.r.t. (α, β).

Lemma 4.3 *Let $(\hat{\alpha}, \hat{\beta})$ be a pair of Puiseux strategies, and let $(\hat{\alpha}_0, \hat{\beta}_0)$ be the limit stationary strategy profile. Then*

$$\mathcal{C}(\hat{\alpha}, \hat{\beta}) \subseteq \mathcal{D}(\hat{\alpha}_0, \hat{\beta}_0).$$

Proof: Let $C \in \mathcal{C}(\hat{\alpha}, \hat{\beta})$. We will prove that $C \in \mathcal{D}(\hat{\alpha}_0, \hat{\beta}_0)$.

Fix $z \in C$. Let $g \in G^{\min}_{C \setminus \{z\}}(z)$. For each $[z' \to z''] \in g$ choose an action pair $(a_{z'}, b_{z'})$ that minimize $\mathrm{w}(\hat{p}(z', a, b))$. Define a stationary profile in C by

$$\begin{aligned} \alpha'(z') &= \frac{1}{2}\hat{\alpha}_0(z') + \frac{1}{2}a_{z'} \\ \beta'(z') &= \frac{1}{2}\hat{\beta}_0(z') + \frac{1}{2}b_{z'}. \end{aligned}$$

By (4) if C is stable under (α', β') and the players follow (α', β') then the play reaches z in finite time a.s.

Recall that C is stable under $(\hat{\alpha}_0, \hat{\beta}_0)$. Assume to the contrary that C is not stable under (α', β'). Then there exists z' such that either $p(C \mid z', a_{z'}, b_{z'}) < 1$, or $p(C \mid z', \hat{\alpha}_0(z'), b_{z'}) < 1$, or $p(C \mid z', a_{z'}, \hat{\beta}_0(z')) < 1$. Let $z^\star \notin C$ be a state that can be reached with positive probability under (α', β').

Define a B-graph g' by replacing the unique edge that leaves z' in g by the edge $[z' \to z^\star]$. Then $\mathrm{w}(g') \leq \mathrm{w}(g)$, which contradicts that fact that $Q^{*,z}(C \setminus \{z\}) = 1$. ■

The following example shows that the two notions are not equivalent.

Example 2:

Consider a game with 4 states. States 2 and 3 are dummy states, where each player has a single action, and the transition in each of these two states is: with probability 1/2 remain at the same state and with probability 1/2 move to state 1. State 4 is absorbing. In state 1 both players have 3 actions and transitions are deterministic. Graphically, transitions are as follows:

State 1

1	1	1
1	4	2
1	3	4

State 2

$\frac{1}{2}1 + \frac{1}{2}2$

State 3

$\frac{1}{2}1 + \frac{1}{2}3$

State 4

4

Denote by $\mathcal{D}(T, L)$ the set of weak communicating sets w.r.t. the pure strategy profile where the players play the Top-Left entry in state 1. One can verify that $\mathcal{D}(T, L) = \{\{1\}, \{1, 2\}, \{1, 3\}, \{1, 2, 3\}\}$. However, it is easy to see that $\{1, 2, 3\}$ is not communicating w.r.t. any Puiseux strategy.

After we established the relation between communication (w.r.t. Puiseux strategies) and weak communication (w.r.t. stationary strategies), we deal only with the latter.

5 Controlling Exits from a Communicating Set

In this section we will see how players can control the behavior of each other in a weak communicating set, and how such control can be used to induce a specific exit distribution from this set.

Let (α, β) be a stationary strategy pair and let $C \in \mathcal{D}(\alpha, \beta)$ be a weak communicating set. We define three types of *elementary exit distributions*:

$$\begin{aligned} \mathcal{Q}_1^C(\alpha, \beta) &= \{p(\cdot \mid z, a, \beta(z)), \text{ where } z \in C \text{ and } p(C \mid z, a, \beta(z)) < 1\}, \\ \mathcal{Q}_2^C(\alpha, \beta) &= \{p(\cdot \mid z, \alpha(z), b), \text{ where } z \in C \text{ and } p(C \mid z, \alpha(z), b) < 1\}, \\ \mathcal{Q}_3^C(\alpha, \beta) &= \{p(\cdot \mid z, a, b), \text{ where } z \in C, p(C \mid z, a, \beta(z)) = p(C \mid z, \alpha(z), b) = 1 \\ &\quad \text{ and } p(C \mid z, a, b) < 1\}. \end{aligned}$$

The first set corresponds to unilateral exits of player 1, the second to unilateral exits of player 2, and the third to joint exits. Note that an exit can give positive probability to a state in C. Define

$$\mathcal{Q}^C(\alpha, \beta) = \text{co}\{\mathcal{Q}_1^C(\alpha, \beta) \cup \mathcal{Q}_2^C(\alpha, \beta) \cup \mathcal{Q}_3^C(\alpha, \beta)\}.$$

$\mathcal{Q}^C(\alpha, \beta)$ is the set of all exit probability distributions that can be generated if the players play at every stage mainly (α, β), and perturb to other actions with small probability.

Whenever $Q \in \mathcal{Q}^C(\alpha, \beta)$, we can represent

$$Q = \sum_{l \in L_1} \eta_l P_l + \sum_{l \in L_2} \eta_l P_l + \sum_{l \in L_3} \eta_l P_l$$

where $P_l \in \mathcal{Q}_B^j(\alpha, \beta)$ for $l \in L_j$. This representation is not necessarily unique, but it will not cause difficulties.

Let $C \in \mathcal{D}(\alpha, \beta)$ be a weak communicating set w.r.t. (α, β), $Q = (Q[z])_{z \in Z}$ an exit distribution from C and $\gamma \in (\mathbf{R}^2)^S$ be a payoff vector. γ should be thought of as a continuation payoff once the game leaves C, and Q is the exit distribution we would like to ensure.

In the sequel, $Q\gamma = \sum_z Q[z]\gamma_z$ and $v^i = (v^i_z)_{z \in S}$ is the min-max value of player i (see chapter @ (Neyman)).

Definition 5.1 *Q is a* controllable exit distribution *from C (w.r.t. γ) if for every $\delta > 0$ there exist a strategy pair $(\sigma_\delta, \tau_\delta)$ and bounded stopping times P^1_δ, P^2_δ such that for every initial state $z \in C$*

1. $\mathbf{P}_{z,\sigma_\delta,\tau_\delta}(e_C < \infty) = 1$ *and* $\mathbf{P}_{z,\sigma_\delta,\tau_\delta}(z_{e_C} = z') = Q[z']$ *for every* $z' \in S$.
2. $\mathbf{P}_{z,\sigma_\delta,\tau_\delta}(\min\{P^1_\delta, P^2_\delta\} \le e_C) < \delta$.
3. *For every* σ, $\mathbf{E}_{z,\sigma,\tau_\delta}\left(\gamma^1(z_{e_C})1_{e_C < P^1_\delta} + v^1(z_{P^1_\delta})1_{e_C \ge P^1_\delta}\right) \le Q\gamma^1 + \delta$.
4. *For every* τ, $\mathbf{E}_{z,\sigma_\delta,\tau}\left(\gamma^2(z_{e_C})1_{e_C < P^2_\delta} + v^2(z_{P^2_\delta})1_{e_C \ge P^2_\delta}\right) \le Q\gamma^2 + \delta$.

In this definition, $(\sigma_\delta, \tau_\delta)$ should be thought of as strategies of the players that support the exit distribution Q, and (P^1_δ, P^2_δ) are two statistical tests that check for deviations. Condition 1 says that if the players follow $(\sigma_\delta, \tau_\delta)$ then the game will eventually leave C with the correct exit distribution. Condition 2 says that the probability of false detection of deviation is small, whereas conditions 3 and 4 ensure that no player will benefit more than δ by a deviation that is followed by a min-max punishment once detected.

A simple control mechanism was used by Vrieze and Thuijsman (1989) for two-player absorbing games (see chapter @ (Thuijsman)).

In the sequel we prove several conditions that imply that some exit distribution is controllable. The exit distribution that is induced by the strategies that we construct is only approximately Q, rather than equal to Q. By slightly changing the construction (at the cost of simplicity) one can make sure that the exit distribution is equal to Q. In any case, for our purposes, it is sufficient to have the exit distribution arbitrarily close to Q.

In our construction, we omit the subscript δ from the strategies and stopping rules, since we do not specify what is the exact δ that should be taken.

LEMMA 5.2 *Let C be a weak communicating set w.r.t. (α, β), $\gamma \in (\mathbf{R}^2)^S$ a payoff vector and $Q = \sum_{l \in L} \eta_l P_l$ an exit distribution. If*

1. *$\gamma \geq v$ and $\gamma_z = Q\gamma$ for every $z \in C$.*

2. *$P_l \gamma^1 = Q\gamma^1$ for every $l \in L_1$.*

3. *$P_l \gamma^2 = Q\gamma^2$ for every $l \in L_2$.*

4. *For every $z \in C$ and $a \in A$, $p(\cdot \mid z, a, \beta_z) v^1(\cdot) \leq Q\gamma^1$.*

5. *For every $z \in C$ and $b \in B$, $p(\cdot \mid z, \alpha_z, b) v^2(\cdot) \leq Q\gamma^2$.*

then Q is a controllable exit distribution from C w.r.t. γ.

Sketch of Proof: Fix $\delta^\star, \epsilon > 0$ sufficiently small.

By the definition of weak communication, for every $z \in C$ there exists a stationary strategy pair (α^z, β^z) that satisfies (i) $\| (\alpha^z, \beta^z) - (\alpha, \beta) \| < \delta$, and (ii) if the players follow (α^z, β^z), the game leaves C with probability 0, and reaches the state z with probability 1 in finite time (provided the initial state is in C).

The strategy pair (σ, τ) of the players is defined as follows. In a cyclic manner do the following for each exit P_l.

1. Denote by z the state at which the exit P_l occurs. Play (α^z, β^z) until the game reaches z.

2. Denote $\delta = \delta^\star \eta_l$.

   (a) If $l \in L_1$ (that is, $P_l = (z, a, \beta(z))$), play $((1-\delta)\alpha(z) + \delta a, \beta(z))$.

   (b) If $l \in L_2$ (that is, $P_l = (z, \alpha(z), b)$), play $(\alpha(z), (1-\delta)\beta(z) + \delta b)$.

   (c) If $l \in L_3$ (that is, $P_l = (z, a, b)$), play $((1-\sqrt{\delta})\alpha(z) + \sqrt{\delta} a, (1-\sqrt{\delta})\beta(z) + \sqrt{\delta} b)$.

3. Continue cyclically to the next exit.

Define the stopping times P^1 and P^2 as follows:

a) If player 1 (*resp.* player 2) plays an action which is not compatible with σ (*resp.* τ), P^1 (*resp.* P^2) is stopped.

b) For every $l \in L_1$, consider all stages where the game has been in step (2) for that l, and check whether the distribution of the realized actions of player 2 in those stages is approximately $\beta(z)$ (where z is the state at which P_l occurs). If the answer is negative (that is, the difference between the distribution of the realized actions and β in the supremum norm is larger than ϵ), P^2 is stopped.
This test is done only if the number of times the play was in step (2) for that exit is sufficiently large, so that the probability of false detection of deviation is small.

c) A similar test is done for player 1 for every $l \in L_2$.

d) For every $l \in L_3$, consider all stages where the play has been in step (2) for that l, and check whether the opponent perturbed to a (or b) approximately in the correct frequency. That is, whether the ratio between $\sqrt{\delta}$ and the number of times the realized action of player 1 (*resp.* player 2) was a (*resp.* b) is in $(1-\epsilon, 1+\epsilon)$.
This test is done only if the number of times the play was in step (2) for that exit is sufficiently large, so that the probability of false detection of deviation is small.

We have already seen how to implement test (b) in the proof of Vrieze and Thuijsman for two-player non-absorbing games (see chapter @ (Thuijsman)).

If $\delta^\star$ and ϵ are sufficiently small, the third test can be done effectively, since exiting C occurs after $O(1/\delta^\star)$ stages, whereas each player perturbs in frequency $O(\sqrt{\delta^\star})$. Hence until exiting occurs, each player should perturb at least $O(1/\sqrt{\delta^\star})$ times, which is enough for an efficient statistical test.

One last possible deviation that we should take care of is, what happens if all exits are unilateral exits of some player, and that player has an incentive never to leave C. To deal with such a deviation, we choose $t^\star$ sufficiently large such that under (σ, τ) exiting from C occurs before stage $t^\star$ with high probability, and we add the following constraint to P^1 and P^2:

d) P^1 and P^2 are bounded by $t^\star$.

Thus, there is no profitable deviation, and therefore Q is a controllable exit distribution from C w.r.t. γ, and the lemma is proved. ■

This lemma holds also for general n-player games. It was used in Solan (1999) for 3-player absorbing games.

The two players in the conditions of Lemma 5.2 are symmetric. We will now see a more sophisticated mechanism to control exits from a weak communicating set, where the players are not symmetric.

LEMMA 5.3 *Let $C \in \mathcal{D}(\alpha, \beta)$ be a weak communicating set, $\gamma \in (\mathbf{R}^2)^S$ be a payoff vector and Q be an exit distribution from C. If*

1) $\gamma \geq v$ and $\gamma_z = Q\gamma$ for every $z \in C$.

2) For every $z \in C$ and $a \in A$, $p(\cdot \mid z, a, \beta(z))v^1 \leq Q\gamma^1$.

3) For every $z \in C$ and $b \in B$, $p(\cdot \mid z, \alpha(z), b)v^2 \leq Q\gamma^2$.

4) There exists a representation $Q = \sum_{m=1}^M \eta_m Q_m$ such that for every $m = 1, \ldots, M$:

(a) $Q_m\gamma^1 = Q\gamma^1$.

(b) There exists $F_m \in \mathcal{D}(\alpha, \beta)$ such that Q_m is a controllable exit distribution from F_m w.r.t. γ.

(c) There exits a state $z_m \in F_m$ and an action $a_m \in A$ of player 1 such that $p(C \mid z_m, a_m, \beta(z_m)) = 1$ and $p(F_m \mid z_m, a_m, \beta(z_m)) < 1$.

Then Q is a controllable exit distribution from C w.r.t. γ.

Proof: Note that player 1 is indifferent between using any Q_m to exit C (condition (4.a)). Thus, he can choose an $m \in \{1, \ldots, M\}$, according to the probability distribution $\eta = (\eta_m)_{m=1}^M$. Using the action a_m in state z_m (condition (4.c)) he can signal his choice to player 2. Once m is known to both players, they can implement an exit from F_m according to Q_m (condition (4.b)). Conditions (1), (2) and (3) ensure that no deviation is profitable.

However, exiting F_m does not necessarily mean exiting C. If the game remains in C, the players start from the beginning: player 1 chooses a new m, signals it to player 2 and so on.

We shall now define the strategies (σ, τ) and stopping times P^1, P^2 more formally. Let $\delta > 0$ be sufficiently small.

1) Player 1 chooses $m^\star \in \{1, \ldots, M\}$. Each m is chosen with probability η_m, independent of the past play.

The players set $m = 1$, and do as follows.

2) (a) If $m^\star = m$, player 1 chooses whether to signal that $m^\star = m$ during the coming phase (with probability δ), or whether not to signal (with probability $1 - \delta$).

   (b) The players play the stationary strategy $(\alpha^{z_m}, \beta^{z_m})$ until the game reaches z_m.

   (c) In z_m, player 2 plays the mixed action $\beta(z_m)$. Player 1 plays $\alpha(z_m)$ if $m^\star = m$ and he chose to signal that fact to player 2, and $(1 - \delta)\alpha(z_m) + \delta a_m$ otherwise.

   The players repeat steps (2.b)-(2.c) $1/\delta^4$ times, or until player 1 played a_m in z_m for the first time, whichever occurs first.

3) If player 1 played the action a_m in step (2.b), the players increase cyclically m by 1, and go back to step (2).

4) Otherwise, the players continue with the strategy pair (σ_m, τ_m) that supports Q_m as a controllable exit distribution from F_m w.r.t. γ, until the game leaves F_m.

5) If by leaving F_m the game also left C, we are done. Otherwise, the players go back to step (1).

If the players follow (σ, τ), then

i) In each round of step 2, if $m^\star = m$ then a signal is sent to player 2 with probability $(1 - \delta)^{1/\delta^4} < \delta$.

ii) In $1/\delta^2$ repetitions of steps 1-3, the probability that in (2.a) player 1 ever chooses to signal to player 2 is $1 - (1 - \delta)^{1/\delta^2} > 1 - \delta$, and the probability that player 1 will *not* play the action a_m when $m \neq m^\star$ is $1 - (1 - (1 - \delta)^{1/\delta^4})^{1/\delta^2} < \delta$.

It follows that the expected continuation payoff is approximately $Q\gamma$.

The stopping times are defined as in the proof of Lemma 5.2, with the following addition.

e) Whenever the play is in step (4), the players use the stopping times that support Q_m as a controllable exit distribution from F_m w.r.t. γ, disregarding the history up to the stage where they started to follow (σ_m, τ_m).

Let us verify that no player can profit too much by deviating.

- Since player 1 is indifferent between choosing any m (his expected continuation payoff is $Q\gamma^1$), he cannot profit by deviating in the lottery stage.
- Since player 1 reveals the signal to player 2 each time with probability δ, the expected continuation payoff, conditioned that player 1 did not play any action a_m in step (2.b) yet, is approximately $Q\gamma^2$.
- Once player 2 is notified of $m^\star$, the game is in $F_{m^\star}$. Since $Q_{m^\star}$ is controllable, there is no profitable deviation.
- Conditions (2) and (3) ensure that detectable deviations are not profitable.

■

Remark 1: Note that if $Q_m = \sum_{l=1}^{L} \nu_l P_l$ is supported by unilateral exits (P_l) of player 1, and $P_l\gamma^1 = Q\gamma^1$ for all of these exits, then condition (4.c) for this m is not needed. Indeed, instead of signaling whether $m^\star$ is equal to m or not, the players will just try to use once each exit P_l with probability $\delta\nu_l$, as was done in the proof of Lemma 5.2. Thus, when the counter in step (2) points to that set, we replace step (2) with the following:

2) Set $l = 1$ and do the following.

   a) Denote $P_l = (z, a, \beta(z))$.

   b) Play the stationary strategy (α^z, β^z) until the game reaches z.

   c) Play $((1-\delta\nu_l)\alpha(z) + \delta\nu_l a, \beta(z))$.

   d) If a was played in (c), we are done. Otherwise, increase l by one, and go back to (a). If $l = L$, continue to the next m.

Since player 1 is indifferent between his unilateral exits, he cannot profit by deviating. Since any exit is used with small probability, the overall expected continuation payoff of player 2 is close to $Q\gamma^2$, so he cannot profit by deviating either.

Remark 2: More generally, if Q_m satisfies the conditions of Lemma 5.2 w.r.t. C and γ, then condition (4.c) is not needed for this m. $m^\star$ will be chosen by player 1 from the set $\{1, \ldots, M\} \setminus \{m\}$, with the normalized probability distribution. The players play as in the proof of Lemma 5.3, but when the counter has the value m, they follow steps (1)-(3) in the proof of Lemma 5.2 once for each exit.

It can be verified that if the players follow this strategy profile then the exit distribution is approximately $Q\gamma$. The statistical tests that we have employed in the proof of Lemma 5.2 can be employed here to deter players from deviating.

Remark 3: If (i) $M = 2$, (ii) $F_1 = F_2 = C$, (iii) Q_1 is supported by unilateral exits of player 1 and (iv) Q_m satisfies the conditions of Lemma 5.2 w.r.t. C and $Q_m\gamma$ for $m = 1, 2$, then condition (4.c) is not needed altogether.

Instead of alternately signaling player 2 whether $m^\star = 1$ or $m^\star = 2$, player 1 will first signal to player 2 whether $m^\star = 1$, and, if no signal was sent, both players will continue as if $m^\star = 2$.

The way to signal whether $m^\star = 1$ is, as in Remark 1, for player 1 to use one of the unilateral exits that support Q_1.

We now state another condition for an exit distribution to be controllable, that follows from Lemma 5.3 and the last three remarks. This condition is used in Vieille's (1997b) proof of existence of equilibrium in two-player non zero-sum stochastic games.

Lemma 5.4 (Vieille, 1997b) *Let $C \in \mathcal{D}(\alpha, \beta)$ be a weak communicating set, $\gamma \in (\mathbf{R}^2)^S$ a payoff vector and $Q = \sum_{l \in L} \nu_l P_l$ an exit distribution. If*

1) $\gamma \geq v$ and $\gamma_s = Q\gamma$ for every $z \in C$.

2) $P_l\gamma^1 = Q\gamma^1$ for every $l \in L_1$.

3) For every $z \in C$ and every $a \in A$, $p(\cdot \mid z, a, \beta(z))v^1 \leq Q\gamma^1$.

4) For every $z \in C$ and every $b \in B$, $p(\cdot \mid z, \alpha(z), b)v^2 \leq Q\gamma^2$.

5) There exists a partition $(L_2^0, \ldots, L_2^M)$ of L_2 and weak communicating subsets $F_1, \ldots, F_M \in \mathcal{D}(\alpha, \beta)$ of C such that $L_2^0 = \{l \in L_2 \mid P_l\gamma^2 = Q\gamma^2\}$ and, for every $m \geq 1$

(a) $P_l\gamma^2 = Q_m\gamma^2$ for every $l \in L_2^m$, where $Q_m = \sum_{l \in L_2^m} \frac{\nu_l}{\sum_{l \in L_2^m} \nu_l} P_l$.

(b) For every $z \in F_m$ and $b \in B$,

- *If $p(F_m \mid z, \alpha(z), b) < 1$ then $p(C \mid z, \alpha(z), b) < 1$.*
- *If $p(C \mid z, \alpha(z), b) < 1$ then $p(\cdot \mid z, \alpha(z), b)\gamma^2 \leq Q_m\gamma^2$.*

(c) $Q_m\gamma^1 = Q\gamma^1$.

(d) $Q_m\gamma^2 \geq \max_{z \in F_m} v_z^2$.

(e) For every $l \in L_2^m$, the state in which P_l occurs is in F_m.

then Q is a controllable exit distribution from C w.r.t. γ.

Sketch of Proof: First we note that the conditions imply that for every m, Q_m is a controllable exit distribution from F_m w.r.t. $Q_m\gamma$. Indeed, by (5.a) Q_m is supported by unilateral exits of player 2, and player 2 receives the same continuation payoff using any of them. By conditions (1) and (5.b) player 2 does not have a profitable deviation, and by (2), (3) and (5.c) player 1 does not have profitable deviations.

Second, define

$$Q' = \frac{\sum_{l \in L_1 \cup L_3 \cup L_2^0} \nu_l P_l}{\sum_{l \in L_1 \cup L_3 \cup L_2^0} \nu_l}.$$

Then Q is a convex combination of Q' and $(Q_m)_{m=1}^M$. If for every m there exists a state z_m and an action a_m such that $p(C \mid z_m, a_m, \beta) = 1$ while $p(F_m \mid z_m, a_m, \beta) < 1$, it follows by Lemma 4.3 and Remark 2 that Q is a controllable exit distribution from C w.r.t. γ.

Otherwise, one can show that $L_3 = \emptyset$ and player 2 is indifferent between his exits (that is, either $M = 0$, or $M = 1$, $L_2^0 = \emptyset$ and $F_1 = C$). If $M = 0$ we are done, since then the conditions of Lemma 5.2 are satisfied.

If $M = 1$ and $L_2^0 = \emptyset$, then $Q_2' = \frac{\sum_{l \in L_1} \nu_l P_l}{\sum_{l \in L_1} \nu_l}$ is an exit distribution from C that is supported by unilateral exits of player 1, $Q_1' = \frac{\sum_{l \in L_2} \nu_l P_l}{\sum_{l \in L_2} \nu_l}$ is an exit distribution from C that is supported by unilateral exits of player 2, and player 2 is indifferent between his exits. Since $L_3 = \emptyset$, Q is a convex combination of Q_1' and Q_2'.

Since $Q'_1\gamma^1 = Q\gamma^1$, it follows that $Q'_2\gamma^1 = Q\gamma^1$, hence Q'_2 is a controllable exit distribution from C w.r.t. γ. By Remark 3 it follows that Q is a controllable exit distribution from C w.r.t. γ. ■

References

[1] Avṣar Z.M. and Baykal-Gürsoy M. (1999) A Decomposition Approach for Undiscounted Two-Person Zero-Sum Stochastic Games, *Methematical Methods in Operations Research*, **3**, 483-500

[2] Bather J. (1973) Optimal Decision Procedures for Finite Markov Chains. Part III: General Convex Systems, *Advances in Applied Probability*, **5**, 541-553

[3] Bewley T. and Kohlberg E. (1976) The Asymptotic Theory of Stochastic Games, *Mathematics of Operations Research*, **1**, 197-208

[4] Eaves B.C. and Rothblum U.G. (1989) A Theory on Extending Algorithms for Parametric Problems, *Mathematics of Operations Research*, **14**, 502-533

[5] Freidlin M. and Wentzell A. (1984) *Random Perturbations of Dynamical Systems*. Springer-Verlag, Berlin.

[6] Mertens J.F. and Neyman A. (1981) Stochastic Games, *International Journal of Game theory*, **10**, 53-66

[7] Ross K.W. and Varadarajan R. (1991) Multichain Markov Decision Processes with a Sample Path Constraint: A Decomposition Approach, *Mathematics of Operations Research*, **16**, 195-207

[8] Solan E. (1999) Three-Person Absorbing Games, *Mathematics of Operations Research*, **24**(3)

[9] Solan E. (1998) Stochastic Games with 2 Non-Absorbing States, *Israel Journal of Mathematics*, to appear

[10] Solan E. and Vieille N. (1998) Correlated Equilibrium in Stochastic Games, *Games and Economic Behavior*, to appear

[11] Vieille N. (1997a) Large Deviations and Stochastic Games, *Israel Journal of Mathematics*, to appear

[12] Vieille N. (1997b) Equilibrium in 2-Person Stochastic Games II: The Case of Recursive Games, *Israel Journal of Mathematics*, to appear

[13] Vrieze O.J. and Thuijsman F. (1989) On Equilibria in Repeated Games With Absorbing States, *International Journal of Game Theory*, **18**, 293-310