PERTURBATIONS OF MARKOV CHAINS WITH APPLICATIONS TO STOCHASTIC GAMES

EILON SOLAN
Northwestern University
Evanston, USA

Tel Aviv University
Tel Aviv, Israel

Abstract. In this chapter we will review several topics that are used extensively 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 presented here appeared in [17],[16], and a few appeared in [12],[13].

The first main issue is Markov chains where the transition rule is a Puiseux probability distribution. We define the notion of communicating sets and construct 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 the exit distribution 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, where $\Delta(K)$ is the set of probability distributions over K.

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

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

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

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 $k \in K$.

In words, the process eventually reaches an absorbing state with probability 1.

We define the *arrival time* to state l by

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

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

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

By A.1-A.2, e_B is finite a.s. Let $Q^l_{k,p}(B) = \mathbf{P}_{k,p}(k_{e_B} = l)$ be the probability that the first state outside B that the process visits is l. Clearly, this probability depends on the initial state. We denote by $Q_{k,p}(B) = (Q^l_{k,p}(B))_{l \in K}$ the *exit distribution* from B. Since e_B is finite a.s., this is a probability distribution. It will be used only when $k \in B$.

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 $l \in K \setminus \{k\}$ with $[k \to l] \in g$;
- g has no cycle; that is, there are no positive integers 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).

This definition implies 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.

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

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

The following lemma relates the exit distribution from B to the weights of all B-graphs.

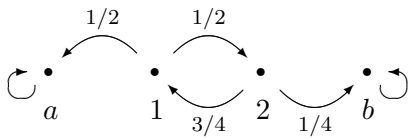
Lemma 1 (Freidlin and Wentzell [6], Lemma 6.3.3) *If $k \in B$ and $l \notin B$,*

$$Q^l_{k,p}(B) = \frac{\sum_{g \in G_B(k \to l)} p(g)}{\sum_{g \in G_B} p(g)}.$$

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

Lemma 1 implies that $Q^l_{k,p}(B)$ is continuous as a function of the transition rule p, for every fixed B, $k \in B$ and $l \notin B$.

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, $A = \{a, b\}$ and the process reaches an absorbing state in finite time a.s. Graphically, the Markov chain looks as follows.



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

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, that $Q^a_{1,p}(B) = 4/5$, while $Q^a_{2,p}(B) = 3/5$.

2. Puiseux Markov Chains

Puiseux series were introduced to the study of stochastic games by Bewley and Kohlberg [3]. 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 were used by Mertens and Neyman [7] to prove the existence of the uniform value in zero-sum games, by Solan [12] and Solan and Vieille [14] for n-player stochastic games, and by Coulomb [4] and Rosenberg et al. [10],[9] to prove the existence of the uniform min-max value in stochastic games with imperfect monitoring. Puiseux series were used in other fields as well (see, e.g., Eaves and Rothblum [5]).

All the definitions and results we have stated in Section 1 do not rely on 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 functions $\hat{f} : (0, 1) \to \mathbf{R}$ that have a representation

$$\hat{f}_\epsilon = \sum_{i=L}^{\infty} a_i \epsilon^{i/M} \tag{1}$$

in an open neighborhood of 0, for some integer L and positive integer M. As a rule, Puiseux functions are denoted with a hat. The *valuation* of a Puiseux function f with representation (1) 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

$$\hat{f}_0 = \lim_{\epsilon \to 0} \hat{f}_\epsilon = \begin{cases} 0 & \mathrm{w}(\hat{f}) > 0 \\ a_0 & \mathrm{w}(\hat{f}) = 0 \end{cases} . \tag{2}$$

It is easy to verify that

$$\mathrm{w}(\hat{f}\hat{g}) = \mathrm{w}(\hat{f}) + \mathrm{w}(\hat{g}), \tag{3}$$

$$\lim_{\epsilon \to 0} \frac{\hat{f}_\epsilon}{\hat{g}_\epsilon} = 0 \text{ whenever } \mathrm{w}(\hat{f}) > \mathrm{w}(\hat{g}), \text{ and} \tag{4}$$

$$\lim_{\epsilon \to 0} \hat{f}_\epsilon \text{ is finite implies that } \mathrm{w}(\hat{f}) \geq 0. \tag{5}$$

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 (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 zeroes in any neighborhood of 0, then it is the zero function. In particular, if a Puiseux function is not zero, then it has no zeroes in a neighborhood of 0. Therefore, in a neighborhood of 0, the collection of recurrent sets 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 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 (3) 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 by Lemma 1 that for every Puiseux transition rule $\hat{p}$, $Q^l_{k,\hat{p}}(B)$ is a Puiseux function. By (2) and (5), the limit $\lim_{\epsilon \to 0} Q_{k,\hat{p}_\epsilon}(B)$ exists, and is a probability distribution.

Define $G_B^{\min}$ to be the collection of all B-graphs $g \in G_B$ that have minimal valuation among all B-graphs in G_B. Set $G_B^{\min}(k \to l) = G_B(k \to l) \cap G_B^{\min}$. This set may be empty. By (4) it follows that if $k \in B$ then

$$\lim_{\epsilon \to 0} Q^l_{k,\hat{p}_\epsilon}(B) = \lim_{\epsilon \to 0} \frac{\sum_{g \in G_B^{\min}(k \to l)} \hat{p}_\epsilon(g)}{\sum_{g \in G_B^{\min}} \hat{p}_\epsilon(g)}, \tag{6}$$

where the sum over an empty set is 0.

3. Communicating Sets

Bather [2] 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 [11] defined another notion of communication in Markov decision processes. A set B in a Markov decision process is *strongly communicating* if it is recurrent under some transition rule.

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

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 decision processes, this definition coincides with the one given by Ross and Varadarajan [11]. We then study the relation between the two definitions.

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

Definition 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 $+\infty$. This implies the following.

Lemma 2 *If B is communicating w.r.t. $\hat{p}$, then B is closed under $\hat{p}_0$.*

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 $k \in C \setminus B$, then

$$\lim_{\epsilon \to 0} \sum_{l \in B} Q^l_{k,\hat{p}_\epsilon}(C \setminus B) = 1. \tag{7}$$

Define a hierarchy (or a partial order) on $\mathcal{C}(\hat{p})$ by set inclusion. Definition 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 [11], and a different type of hierarchy is used in Avṣar and Baykal-Gürsoy [1].

Let B and C be communicating sets w.r.t. $\hat{p}$. B is a *child* of C if B is a strict subset of C and 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 hierarchy as a forest).

Definition 1 implies the following.

Lemma 3 *If B is communicating w.r.t. $\hat{p}$ then $\lim_{\epsilon\to 0} Q_{k,\hat{p}_\epsilon}(B)$ is independent of k, provided $k \in B$.*

For every $B \in \mathcal{C}(\hat{p})$, the limit $Q^*_{\hat{p}}(B) = \lim_{\epsilon\to 0} Q_{k,\hat{p}_\epsilon}(B)$, which is independent of $k \in B$, 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 $(\widetilde{K}, q)$ as follows.

- The state space is $\widetilde{K} = \{d_1, \ldots, d_L\} \cup (K \setminus \cup_{l=1}^{L} D_l)$, where $d_1, \ldots, d_L$ are L distinct symbols.
- 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 $k' \notin \cup_l D_l$.
  - $q(d_l, d_{l'}) = \sum_{k' \in D_{l'}} Q^{*,k'}_{\hat{p}}(D_l)$.

In words, we replace each maximal communicating subset D_l of C by a single state d_l. 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. (7) implies the following.

Lemma 4 *Under the above notations, C is recurrent in $(\widetilde{K}, q)$.*

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 2 *A* Puiseux strategy *for player 1 is a function $\hat{\alpha} : (0,1) \times S \to \Delta(A)$ such that for every $z \in S$ and every $a \in A$, $\hat{\alpha}^a_z$ is a Puiseux function.*

Observe 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 3 *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. *$\|\alpha - \alpha'\|_\infty < \delta$ and $\|\beta - \beta'\|_\infty < \delta$.*
2. *C is closed under (α', β'); that is, $p(C \mid z', \alpha', \beta') = 1$ for every $z' \in C$.*
3. *$\mathbf{P}_{z',\alpha',\beta'}(z_n = z \text{ for some } n \geq 1) = 1$ for every $z' \in C$.*

Observe that if C is weakly communicating w.r.t. (α, β), then it is closed under (α, β).

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

Lemma 5 *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 $\delta > 0$ and $z \in C$.

Let $g \in G^{\min}_{C\setminus\{z\}}$. By (7) and (6), all states $z' \in C \setminus \{z\}$ lead to z in g.

For each $[z' \to z''] \in g$ choose an action pair $(a_{z'}, b_{z'})$ that minimizes $\mathrm{w}(\hat{p}(z', a, b))$ among all action pairs (a, b) such that $\hat{p}(z'' \mid z', a, b) > 0$. Define a stationary profile in C by

$$\alpha'(z') = (1 - \delta)\hat{\alpha}_0(z') + \delta a_{z'}, \text{ and } \beta'(z') = (1 - \delta)\hat{\beta}_0(z') + \delta b_{z'}.$$

In particular, (1) of Definition 3 holds.

The choice of $(a_{z'}, b_{z'})$ implies that (2) of Definition 3 holds. Indeed, otherwise there would be $z' \in C\setminus\{z\}$ and $z^\star \notin C$ such that $p(z^\star \mid z', \alpha'_{z'}, \beta'_{z'}) > 0$.

Define a B-graph g' by replacing the unique edge that leaves z' in g by the edge $[z' \to z^\star]$. Then $\mathrm{w}(g') \leq \mathrm{w}(g)$, and therefore $g' \in G^{\min}_{C\setminus\{z\}}$. By (6) this contradicts the fact that $Q^{*,z}_{\hat{\alpha},\hat{\beta}}(C \setminus \{z\}) = 1$.

Since all states in $C \setminus \{z\}$ lead to z under g, (3) of Definition 3 holds. ∎

The following example shows that the two notions of communication 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.

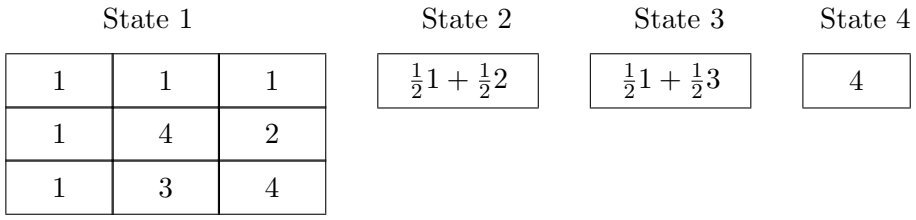


Figure 2

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.

Having established the relation between communication (w.r.t. Puiseux strategies) and weak communication (w.r.t. stationary strategies), we shall 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, using statistical tests and threats of punishment, 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*:

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

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. The set of all *exit distributions* is

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

$\mathcal{Q}^C(\alpha,\beta)$ is the set of all exit distributions from C that can be generated if the players at every stage play mainly (α,β), and perturb to other actions with low 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 $P_l \in \mathcal{Q}_j^C(\alpha,\beta)$ for $l \in L_j$. This representation is not necessarily unique, but this fact will not cause any difficulty.

Let $Q = (Q[z])_{z\in S}$ be an exit distribution from C, and let $\gamma \in (\mathbf{R}^2)^S$ 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, $\mathbf{E}_Q[\gamma] = \sum_z Q[z]\gamma_z$, $\mathbf{E}[\gamma \mid z,\alpha_z,\beta_z] = \mathbf{E}_{p(\cdot|z,\alpha_z,\beta_z)}[\gamma]$, and $v^i = (v_z^i)_{z\in S}$ is the min-max value of player i (see [8]).

Definition 4 *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 two bounded stopping times P_δ^1, P_δ^2 such that for every initial state $z \in C$ the following conditions hold.*

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

In this definition, $(\sigma_\delta,\tau_\delta)$ should be thought of as strategies that support the exit distribution Q, and (P_δ^1, P_δ^2) 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 desired exit distribution. Condition 2 says that the probability of false detection of deviation is small, whereas conditions 3 and 4 ensure that no player can 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 [18] for two-player absorbing games (see [15]).

In the sequel we prove several conditions which imply that some exit distributions are controllable. The exit distribution induced by the strategies we construct is only approximately Q, rather than equal to Q. By slightly changing the construction (at the cost of higher complexity) one can ensure 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 times, since we do not specify what is the exact δ that should be taken.

Lemma 6 *Let C be a weak communicating set w.r.t. (α,β), let $\gamma \in (\mathbf{R}^2)^S$ be a payoff vector, and let $Q = \sum_{l\in L}\eta_l P_l$ be an exit distribution. Assume that the following conditions hold.*

1. *$\gamma_z^i \geq v_z^i$ and $\gamma_z = \mathbf{E}_Q[\gamma]$ for every $i = 1,2$ and $z \in C$.*
2. *$\mathbf{E}_{P_l}[\gamma^1] = \mathbf{E}_Q[\gamma^1]$ for every $l \in L_1$.*
3. *$\mathbf{E}_{P_l}[\gamma^2] = \mathbf{E}_Q[\gamma^2]$ for every $l \in L_2$.*
4. *For every $z \in C$ and every $a \in A$, $\mathbf{E}[v^1(\cdot) \mid z, a, \beta_z] \leq \mathbf{E}_Q[\gamma^1]$.*
5. *For every $z \in C$ and every $b \in B$, $\mathbf{E}[v^2(\cdot) \mid z, \alpha_z, b] \leq \mathbf{E}_Q[\gamma^2]$.*

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) $\| (\alpha^z,\beta^z) - (\alpha,\beta) \| < \delta^\star$, 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) is defined as follows. In a cyclic manner do the following for each exit distribution 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, $P_l = p(\cdot \mid z, a, \beta_z)$), play $((1-\delta)\alpha_z + \delta a, \beta_z)$.
   (b) If $l \in L_2$ (that is, $P_l = p(\cdot \mid z, \alpha_z, b)$), play $(\alpha_z, (1-\delta)\beta_z + \delta b)$.
   (c) If $l \in L_3$ (that is, $P_l = p(\cdot \mid z, a, b)$), 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 β_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 to

b) approximately in the specified 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 [15].

If $\delta^\star$ and ϵ are sufficiently small, test (d) can be employed effectively, since exiting C occurs after $O(1/\delta^\star)$ stages, whereas each player should perturb with probability $O(\sqrt{\delta^\star})$. Hence, until exiting occurs, each player should perturb $O(1/\sqrt{\delta^\star})$ times, which is enough for an effective 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:

e) 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 also holds for general n-player games. It is used in Solan [12] for three-player absorbing games, and in Solan and Vieille [14] for n-player stochastic games.

The two players in the conditions of Lemma 6 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 7 *Let $C \in \mathcal{D}(\alpha, \beta)$ be a weak communicating set, let $\gamma \in (\mathbf{R}^2)^S$ be a payoff vector, and let Q be an exit distribution from C. Assume that*

1) $\gamma_z^i \geq v_z^i$ and $\gamma_z = \mathbf{E}_Q[\gamma]$ for every $i = 1, 2$ and $z \in C$.
2) For every $z \in C$ and $a \in A$, $\mathbf{E}[v^1 \mid z, a, \beta_z] \leq \mathbf{E}_Q[\gamma^1]$.
3) For every $z \in C$ and $b \in B$, $\mathbf{E}[v^2 \mid z, \alpha_z, b] \leq \mathbf{E}_Q[\gamma^2]$.
4) There exists a representation $Q = \sum_{m=1}^M \eta_m Q_m$ such that for every $m = 1, \ldots, M$:

(a) $\mathbf{E}_{Q_m}[\gamma^1] = \mathbf{E}_Q[\gamma^1]$.

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

(c) There exists 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. Consider the following construction. Player 1 chooses $m \in \{1,\dots,M\}$, according to the probability distribution $\eta = (\eta_m)_{m=1}^M$. Using the action a_m in state z_m (condition (4.c)) player 1 signals his choice to player 2. Once m is known to both players, they implement an exit from F_m according to Q_m (condition (4.b)). By condition (4.a) player 1's payoff is independent of the chosen m. 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 the stopping times P^1, P^2 more formally. Let $\delta > 0$ be sufficiently small.

1) Player 1 chooses $m^\star \in \{1,\dots,M\}$. Each m is chosen with probability η_m, independently 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 this fact 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 β_{z_m}. Player 1 plays α_{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 (with the same choice that was made at step (2.a)), or until player 1 has 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 leaves C, we are done. Otherwise, the players go back to step (1).

If the players follow (σ,τ), then in each round of step 2, if $m^\star = m$ a signal is sent to player 2 with probability $(1-\delta)^{1/\delta^4} < \delta$. Moreover, 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 $\mathbf{E}_Q[\gamma]$.

The stopping times are defined as in the proof of Lemma 6, with the following addition.

f) 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's expected payoff is $\mathbf{E}_Q[\gamma^1]$, regardless of the m he chooses, 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, conditional on player 1 not having played any action a_m in step (2.b), is approximately $\mathbf{E}_Q[\gamma^2]$.
– 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 once $m^\star$ is revealed. ∎

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 $\mathbf{E}_{P_l}[\gamma^1] = \mathbf{E}_Q[\gamma^1]$ for all 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 try to use just once each exit P_l with probability $\delta\nu_l$, as was done in the proof of Lemma 6. 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 $P_l = p(\cdot \mid z, a, \beta_z)$.
   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 is 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 low probability, the overall expected continuation payoff of player 2 is close to $\mathbf{E}_Q[\gamma^2]$, so he cannot profit by deviating either.

Remark 2 More generally, if Q_m satisfies the conditions of Lemma 6 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 7, but when the counter has the value m, they follow steps (1)-(3) in the proof of Lemma 6 once for each exit.

It can be verified that if the players follow this strategy profile then the exit distribution is approximately $\mathbf{E}_Q[\gamma]$. The statistical tests employed in the proof of Lemma 6 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 6 w.r.t. C and $\mathbf{E}_{Q_m}[\gamma]$ for $m = 1, 2$, then condition (4.c) is not needed at all.

Instead of alternately signaling to player 2 whether $m^\star = 1$ or $m^\star = 2$, player 1 first signals to player 2 whether $m^\star = 1$, and, if no signal is sent, both players 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 that ensures that an exit distribution is controllable, which follows from Lemma 7 and the last three remarks. This condition is used in Vieille's [16] proof of existence of equilibrium in two-player non-zero-sum stochastic games.

Lemma 8 (Vieille [16]) *Let $C \in \mathcal{D}(\alpha, \beta)$ be a weak communicating set, let $\gamma \in (\mathbf{R}^2)^S$ be a payoff vector, and let $Q = \sum_{l \in L} \nu_l P_l$ be an exit distribution. Assume that the following conditions hold.*

1) *$\gamma_z^i \geq v_z^i$ and $\gamma_z = \mathbf{E}_Q[\gamma]$ for every $i = 1, 2$ and $z \in C$.*
2) *$\mathbf{E}_{P_l}[\gamma^1] = \mathbf{E}_Q[\gamma^1]$ for every $l \in L_1$.*
3) *For every $z \in C$ and every $a \in A$, $\mathbf{E}[v^1 \mid z, a, \beta_z] \leq \mathbf{E}_Q[\gamma^1]$.*
4) *For every $z \in C$ and every $b \in B$, $\mathbf{E}[v^2 \mid z, \alpha_z, b] \leq \mathbf{E}_Q[\gamma^2]$.*
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 $L_2^0 = \{l \in L_2 \mid \mathbf{E}_{P_l}[\gamma^2] = \mathbf{E}_Q[\gamma^2]\}$ and, for every $m \geq 1$,*
   (a) $\mathbf{E}_{P_l}[\gamma^2] = \mathbf{E}_{Q_m}[\gamma^2]$ 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 $\mathbf{E}[\gamma^2 \mid z, \alpha_z, b] \leq \mathbf{E}_{Q_m}[\gamma^2]$.
   (c) $\mathbf{E}_{Q_m}[\gamma^1] = \mathbf{E}_Q[\gamma^1]$.
   (d) $\mathbf{E}_{Q_m}[\gamma^2] \geq \max_{z \in F_m} v_z^2$.
   (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. $\mathbf{E}_{Q_m}[\gamma]$. Indeed, by (5.a) Q_m is supported by unilateral exits of player 2, and player 2 receives the same continuation payoff using any one 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 exist 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 6 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 $\mathbf{E}_{Q'_1}[\gamma^1] = \mathbf{E}_Q[\gamma^1]$, it follows that $\mathbf{E}_{Q'_2}[\gamma^1] = \mathbf{E}_Q[\gamma^1]$; 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. Avşar, Z.M. and Baykal-Gürsoy, M. (1999) A decomposition approach for undiscounted two-person zero-sum stochastic games, *Mathematical 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. Coulomb, J.M. (2002) Stochastic games without perfect monitoring, mimeo.
5. Eaves, B.C. and Rothblum, U.G. (1989) A theory on extending algorithms for parametric problems, *Mathematics of Operations Research* **14**, 502-533.
6. Freidlin, M. and Wentzell, A. (1984) *Random Perturbations of Dynamical Systems*, Springer-Verlag, Berlin.
7. Mertens, J.F. and Neyman, A. (1981) Stochastic games, *International Journal of Game Theory* **10**, 53–66.
8. Neyman, A. (2003) Stochastic games: Existence of the minmax, in A. Neyman and S. Sorin (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 11, pp. 173–193.
9. Rosenberg, D., Solan, E. and Vieille, N. (2002) Stochastic games with imperfect monitoring, Discussion Paper 1341, The Center for Mathematical Studies in Economics and Management Sciences, Northwestern University.
10. Rosenberg, D., Solan, E. and Vieille, N. (2002) On the maxmin value of stochastic games with imperfect monitoring, Discussion Paper 1344, The Center for Mathematical Studies in Economics and Management Sciences, Northwestern University.

11. 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.
12. Solan, E. (1999) Three-person absorbing games, *Mathematics of Operations Research* **24**, 669–698.
13. Solan, E. (2000) Stochastic games with two non-absorbing states, *Israel Journal of Mathematics* **119**, 29–54.
14. Solan, E. and Vieille, N. (2002), Correlated equilibrium in stochastic games, *Games and Economic Behavior* **38**, 362–399.
15. Thuijsman, F. (2003) Repeated games with absorbing states, in A. Neyman and S. Sorin (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 13, pp. 205–213.
16. Vieille, N. (2000) Equilibrium in two-person stochastic games II: The case of recursive games, *Israel Journal of Mathematics* **119**, 93–126.
17. Vieille, N. (2000) Small perturbations and stochastic games, *Israel Journal of Mathematics* **119**, 127–142.
18. Vrieze, O.J. and Thuijsman, F. (1989) On equilibria in repeated games with absorbing states, *International Journal of Game Theory* **18**, 293–310.