Discussion Paper 1226

Correlated Equilibrium in Stochastic Games

by

Eilon Solan
MEDS Department, Northwestern University
J.L. Kellogg Graduate School of Management
Evanston, IL 60208-2001

Nicolas Vieille
CEREMADE
Université Paris
75 016 Paris, France

September 18, 1998

http://www.kellogg.nwu.edu/research/math

Correlated Equilibrium in Stochastic Games

Eilon Solan[†] and Nicolas Vieille[‡]

September 18, 1998

Abstract

We study the existence of correlated equilibrium payoff in stochastic games. The correlation devices that we use are either autonomous (they base their choice of signal on previous signals, but not on previous states or actions) or stationary (their choice is independent of any data, and is drawn according to the same probability distribution at every stage). We prove that any n-player stochastic game admits an autonomous correlated equilibrium payoff, and obtain a stronger result for recursive games. When the game is positive and recursive, a stationary correlated equilibrium payoff exists.

---

† MEDS Department, Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Rd., Evanston, IL 60208.
‡ CEREMADE, Université Paris 9-Dauphine, Place de Lattre de Tassigny, 75 016 Paris, France.

The first author thanks the hospitality of the Center for Rationality and Interactive Decision Theory and of the Laboratoire d'Econométrie de l'Ecole Polytechnique, and the financial support of the Institute of Mathematics at the Hebrew University of Jerusalem.

1 Introduction

A stochastic game is played in stages. At every stage the game is in some state of the world, and each player, given the whole history (including the current state), chooses an action in his action space. The action combination that was chosen by all players, together with the current state, determine the daily payoff that each player receives and the probability distribution according to which the new state of the game is chosen.

Stochastic games were introduced by Shapley (1953) who proved the existence of the discounted value in two-player zero-sum games, as well as the existence of stationary optimal strategies. This result was generalized for discounted equilibria in n-player games by Fink (1964).

Existence of the (undiscounted) value in two-player zero-sum stochastic games was proved by Mertens and Neyman (1981). It is well known (see, e.g., Blackwell and Ferguson (1968)) that optimal strategies when using the undiscounted evaluation, as well as stationary ϵ-optimal strategies, need not exist.

Vieille (1997a) proved that if every two-player positive recursive game that satisfies some property has an equilibrium payoff then every two-player stochastic game has an equilibrium payoff. Existence of equilibrium payoffs for this class of games was proven by Vieille (1997c).

Recursive games have been introduced by Everett (1957). These are stochastic games, with finitely many states and actions, where the payoff for the players in non-absorbing states is zero (a state is absorbing if the probability to leave it, whatever the players play, is 0). A recursive game is positive if the payoff for the players in absorbing states is positive.

Everett proved the existence of the value for two-player, zero-sum recursive games, and the existence of stationary ϵ-optimal strategies. A simpler proof was obtained by Thuijsman and Vrieze (1992). Recently, Rosenberg and Vieille (1997) extended the value existence result to recursive games with more general state spaces (and obtained some results on incomplete information recursive games).

For stochastic games with more than two players, very little is known (see Solan (1997), Solan and Vieille (1998)) and proving, or disproving, the existence of equilibrium payoffs seems a daunting task, even for the simplest games. We study here the existence of *correlated* equilibrium payoffs in n-player stochastic games.

Correlation devices were introduced by Aumann (1974, 1987). A correlation device chooses for every player a private signal before the start of play, and sends to each player the signal chosen for him. Each player can base his choice of an action on the private signal that he has received.

For multi-stage games, various generalizations of correlation devices have been introduced (Fudenberg and Tirole (1991)). (i) The most general device receives at every stage some private message from each player, and chooses for each player a private signal for that stage (*communication device*, Forges (1986, 1988), Myerson (1986), Mertens (1994)). (ii) The most restricted device chooses, as in the case of one-shot games, one private signal before the beginning of the game (*correlation device*, Forges (1986)). (iii) In between, there are devices that choose private signals at every stage, but base their choice only on the current state (*weak correlation devices*, Nowak (1991)) or only on previous signals (*autonomous correlation devices*, Forges (1986)).

Solan (1998) proved that every feasible and individually rational payoff in a stochastic game is a correlated equilibrium payoff where the correlation device chooses at every stage a signal that depends on the previous signal, as well as on the sequence of the states that the play has visited.

In the present paper we study two types of correlation devices: (i) stationary devices, that choose at every stage a signal according to the same probability distribution, independent of any data, and (ii) autonomous devices, that base their choice of new signal on the previous signal, but not on any other information.

We prove three results: (a) Every stochastic game has a correlated equilibrium, using an autonomous correlation device. The equilibrium path is sustained using threat strategies, in which players punish a deviator by his max-min value. This means that players need to correlate also in the punishment phase. However, punishment occurs only if a player disobeys the recommendation of the device. (b) If the game is recursive, then the equilibrium path can be sustained using the min-max value. In particular, players need not correlate their actions in the punishment phase. (c) If the game is positive recursive, then the correlation device can be taken to be stationary, and deviators are punished by their min-max value.

The proofs utilize various methods that appeared in the literature. They are divided into two steps. First we construct a "good" strategy profile; meaning, a profile that yields all players a high payoff, and no player can profit by a unilateral deviation that is followed by an indefinite punishment

(where in (a) punishment is given by the max-min value, and in (b) and (c) by the min-max value). Second we follow Solan (1998) and define a correlation device that *mimics* that profile: the device chooses for each player an action according to the probability distribution given by the profile, and recommends each player to play that action. In order to make deviations non-profitable, the device reveals to all players what were his recommendations in the previous stage. This way, a deviation is detected immediately, and can be punished. In particular, the device that we construct is not canonical (Forges (1988)).

The construction of the "good" strategy profile uses the method of Mertens and Neyman (1981) for (a), and a variant of the method of Vieille (1997c) for (b) and (c).

The paper is arranged as follows. In Section 2 we introduce the model and the main results. After some preliminaries in Section 3 we provide in Section 4 several sufficient conditions for existence of correlated equilibrium payoff. Finally, in Section 5 we prove that each sufficient condition is satisfied in a corresponding class of stochastic games.

2 The Model and the Main Results

A *stochastic game* G is given by (i) a finite set of players N, (ii) a finite set of states S, (iii) for every player $i \in N$, a finite set of available actions A^i. We denote $A = \times_{i \in N} A^i$. (iv) A transition rule $q : S \times A \to \Delta(S)$, where $\Delta(S)$ is the space of all probability distributions over S, and (v) a daily payoff function $r : S \times A \to \mathbf{R}^N$. We assume w.l.o.g. that $|r| \leq 1$.

The game lasts for infinitely many stages. The initial state s_1 is given. In stage n, the current state s_n is announced to the players. Each player i chooses an action $a_n^i \in A^i$; the action combination $a_n = (a_n^i)_{i \in N}$ is publicly announced. s_{n+1} is drawn according to $q(\cdot|s_n, a_n)$ and the game proceeds to stage $n+1$.

A state s in a stochastic game is *absorbing* if $q(s|s, a) = 1$ for every $a \in A$. We denote by $S^\star$ the subset of absorbing states.

The game is *recursive* if $r^i(s, \cdot) = 0$ for every non-absorbing state $s \in S \backslash S^\star$ and every player $i \in N$. It is *positive* if $r^i(s, \cdot) > 0$ for every absorbing state $s \in S^\star$ and every player $i \in N$.

DEFINITION 2.1 *An* autonomous correlation device *is a pair* $\mathcal{D} = ((M^i)_{i \in N}, (\mathcal{D}_n)_{n \in \mathbf{N}})$,

where (i) M^i is a finite set of signals for player i, and (ii) $\mathcal{D}_n : M^{n-1} \to \Delta(M)$, where $M = \times_{i \in N} M^i$. The device is stationary *if $\mathcal{D}_n$ is a constant function which is independent of n.*

If the functions $\mathcal{D}_n$ depend also on previous states that the game has visited - that is $\mathcal{D}_n : (S \times M)^{n-1} \times S \to \Delta(M)$ - then the device is an *extensive-form correlation device.*

Given a correlation device $\mathcal{D}$ we define an extended game $G(\mathcal{D})$. The game $G(\mathcal{D})$ is played exactly as the game G, but at the beginning of each stage n, a signal combination $m_n = (m_n^i)_{i \in N}$ is drawn according to $\mathcal{D}_n(m_1, \ldots, m_{n-1})$, and each player i is informed of m_n^i. Then, each player may base his choice of a_n^i also on previous signals $m_1^i, \ldots, m_n^i$ that he received.

At stage n, player i observes an element of $H_n^i(M) = (S \times M^i \times A)^{n-1} \times S \times M^i$. Therefore, a *strategy for player* i in $G(\mathcal{D})$ is a function $\sigma^i : H^i(M) \to \Delta(A^i)$, where $H^i(M) = \cup_{n \in \mathbf{N}} H_n^i(M)$. We denote by $\Sigma^i(M)$ the set of all strategies of player i in $G(\mathcal{D})$.

A *profile* $\sigma = (\sigma^i)_{i \in N}$ is a vector of strategies, one for each player. We denote $\Sigma(M) = \times_{i \in N} \Sigma^i(M)$, the space of all profiles in $G(\mathcal{D})$. Σ denotes the space of all profiles that are independent of the messages. We identify Σ with the space of profiles in the game G. Stationary strategies of player i are strategies that depend only on the current state, and not on previous signals, states or actions. Thus, a stationary strategy of player i can be identified with an element $x^i = (x_s^i)_{s \in S} \in (\Delta(A^i))^S$, with the understanding that x_s^i is the lottery used by player i to select his action in state s. We denote by X^i the set of stationary strategies of player i.

The set of *finite histories* is denoted by

$$H(M) = \cup_{n \in \mathbf{N}} (S \times M \times A)^{n-1} \times S \times M.$$

This is the set of all histories which are observed by an observer, who observes the signals that are received by *all* the players. The set of all *infinite histories* is denoted by

$$H_\infty(M) = (S \times M \times A)^{\mathbf{N}}.$$

We endow this space with the σ-algebra generated by all the finite cylinders.

Every correlation device $\mathcal{D}$, every profile $\sigma \in \Sigma(M)$ and every initial state $s \in S$ induce a probability measure over $H_\infty(M)$; that is, the proba-

bility measure induced by σ and $\mathcal{D}$, given the initial state is s. We denote expectation w.r.t. this measure by $\mathbf{E}_{\mathcal{D},s,\sigma}$.

Let $r_n^i(\mathcal{D},s,\sigma)$ be the expected payoff that player i receives at stage n given the initial state is s and the players follow the profile σ in $G(\mathcal{D})$. Formally

$$r_n^i(\mathcal{D},s,\sigma) = \mathbf{E}_{\mathcal{D},s,\sigma} r^i(s_n,a_n).$$

Define the expected payoff during the first n stages by:

$$\gamma_n^i(\mathcal{D},s,\sigma) = \frac{1}{n}\sum_{j=1}^{n} r_j^i(\mathcal{D},s,\sigma).$$

DEFINITION 2.2 *A payoff vector $\gamma \in \mathbf{R}^{S\times N}$ is an* autonomous (*resp.* stationary) correlated equilibrium payoff *if for every $\epsilon > 0$ there exists an autonomous (resp. stationary) correlation device $\mathcal{D}$, a profile σ in $G(\mathcal{D})$ and a finite horizon $n_0 \in \mathbf{N}$ such that for every $n \geq n_0$, every player $i \in N$, every strategy $\sigma_\star^i \in \Sigma^i(M)$ of player i and every initial state $s \in S$*

$$\gamma_s^i + \epsilon \geq \gamma_n^i(\mathcal{D},s,\sigma) \geq \gamma_s^i - \epsilon \geq \gamma_n^i(\mathcal{D},s,\sigma^{-i},\sigma_\star^i) - 2\epsilon,$$

where $\sigma^{-i} = (\sigma^j)_{j\neq i}$.

Note that for every $\epsilon > 0$ a different correlation device may be used.

The main result of the paper is:

THEOREM 2.3 *Every stochastic game possesses an autonomous correlated equilibrium payoff.*

The equilibrium path is sustained by threat of punishment by the max-min value, and a player is punished only if he disobeys the recommendation of the device.

If the game is recursive, then one has a stronger result:

THEOREM 2.4 *If the game is recursive, then there is an autonomous correlated equilibrium payoff, where the equilibrium path is sustained using threat of punishment by the min-max value.*

In particular, in recursive games, correlation is needed only on the equilibrium path, and not in the punishment phase.

When the game is positive, one can find a correlated equilibrium payoff where the device is stationary:

THEOREM 2.5 *Every positive recursive game possesses a stationary correlated equilibrium payoff, where the correlation device is independent of* ϵ.

In this case, however, the profile that is used by the players depends on ϵ.

3 Preliminaries

The mixed extension of q to $s \times \times_{i \in N} \Delta(A^i)$ is still denoted by q. For every finite set K and probability distribution $\mu \in \Delta(K)$, $\mu[k]$ is the probability of k under μ. For any probability measure μ and real valued function u defined over S, we denote $\mu u = \sum_{s \in S} \mu[s] u(s)$ the expectation of u under μ. In particular, for every $x \in X$ and every $s \in S$, $q_{s,x} u = \sum_{s' \in S} q(s'|s,x) u(s')$ is the expectation of u under $q(\cdot|s,x)$.

Denote $H_n = (S \times A)^{n-1} \times S$ the space of all histories of length n in G, and $H = \cup_{n \in \mathbf{N}} H_n$ the space of all finite histories. $H_\infty = (S \times A)^{\mathbf{N}}$ is the space of all infinite histories.

Any $a^i \in A^i$ is identified with the element of $\Delta(A^i)$ which assigns probability 1 to a^i. For any subset $L \subseteq N$, we denote $A^L = \times_{i \in L} A^i$ and $A^{-L} = \times_{i \notin L} A^i$.

We denote by v^i the punishment level of player i. It will sometimes be the max-min value, and sometimes the min-max value.

Every history $h \in H$ and every profile τ in G induce a probability measure $\mathbf{P}_{h,\tau}$ over H_∞ - the measure induced by τ in the subgame starting after the history h.

In general, the symbol τ denotes a profile in G, whereas σ denotes a profile in an extended game $G(\mathcal{D})$.

3.1 Communicating Sets

Let $x \in X$ be a stationary profile.

The profile x' is a *perturbation* of x if $\text{supp}(x^i_s) \subseteq \text{supp}(x'^i_s)$ for every player $i \in N$ and every state $s \in S$.

A set $C \subset S$ is *stable* under x if $q(C|s,x) = 1$ for every $s \in C$. The set is *communicating* under x if for every state $s' \in C$ there exists a perturbation x' of x such that C is stable under x' and

$$\mathbf{P}_{s,x'}(\exists n \in \mathbf{N}. s_n = s') = 1 \qquad \forall s \in C.$$

This is a property of the support of x'. In particular, x' can be chosen arbitrarily close to x. Let $y_{C,x,\epsilon,s}$ be such a perturbation that satisfies $\| y_{C,x,\epsilon,s} - x \| < \epsilon$.

This definition captures the idea that the players can reach any state in C from any other state in C by slightly perturbing the stationary profile x.

We denote by $\mathcal{C}(x)$ the collection of all the sets that communicate under x.

3.2 On Exits

Let $x \in X$ and $C \in \mathcal{C}(x)$.

DEFINITION 3.1 *An* exit *from C (w.r.t. x) is a tuple $e = (s, x^{-L}, a^L)$ where $s \in C$, $\emptyset \neq L \subseteq N$, $a^L \in A^L$, $q(C|s, x^{-L}, a^L) > 0$ while $q(C|s, x^{-L'}, a^{L'}) = 0$ for every strict subset L' of L.*

For simplicity, we sometimes write $e = (s, a^L)$ when no confusion may arise. The set of all exits from C w.r.t. x is denoted by $E(x, C)$. For $e = (s, x^{-L}, a^L) \in E(x, C)$, $L(e) = L$ is the subset of players that need to perturb. If $L(e) = \{i\}$, we say that e is a *unilateral exit* of player i. Otherwise, e is a *joint exit*.

For any subset $C \subset S$ we denote

$$e_S = \inf\{n, s_n \notin C\};$$

that is, the first exit stage from C.

It is well known (see Vieille (1997c)) that for every communicating set $C \in \mathcal{C}(x)$, every probability distribution $\mu \in \Delta(E(x, C))$ and every $\epsilon > 0$ there exists a profile τ in G such that $\| \tau(h) - x \| < \epsilon$ for every history h, and the probability that the play leaves C through any exit in $E(x, C)$ is $\mu[e]$, provided that the initial state is in C. Formally, there exists a profile τ such that, provided that the initial state s is in C, e_C is finite $\mathbf{P}_{s,\tau}$-a.s., and $a_{e_C - 1}$ is distributed as if an exit $e = (s', x^{-L}, a^L) \in E(x, C)$ is first drawn according to μ, and then an action combination is drawn according to (x^{-L}, a^L). We denote such a profile by $\tau_{C,\mu,\epsilon}$.

3.3 On Correlation Devices

For the rest of the section, we fix $M^i = A^i$ for every $i \in N$. Let $\mathcal{D}$ be any autonomous correlation device defined over M. Define a profile σ_0 in $G(\mathcal{D})$ by:

$$\sigma_0^i(s_1, m_1, a_1, \ldots, s_n, m_n) = m_n^i.$$

In words, the players follow the recommendations of the device.

Let $\tau \in \Sigma$ be arbitrary.

Define an autonomous correlation device $\mathcal{D}$ as follows:

$$\mathcal{D}_n(s_0, m_1, s_1, \ldots, m_n, s_n) = \tau(s_0, m_1, s_1, \ldots, m_n, s_n)$$

for every finite history $(s_0, m_1, s_1, \ldots, m_n, s_n) \in H$ (recall that $M = A$). Clearly the probability measure $\mathbf{P}_{s,\tau}$ over $H_\infty = S \times (A \times S)^{\mathbf{N}}$ coincides with the marginal probability measure $\mathbf{P}_{\mathcal{D},s,\sigma_0}$ over H_∞. We say that the device $\mathcal{D}$ *mimics* the profile τ.

4 Sufficient Conditions for Existence of Correlated Equilibrium

For every profile τ, every history $h \in H_m$ and every $n \in \mathbf{N}$ define

$$\gamma_n^i(h, \tau) = \frac{1}{n+m} \mathbf{E}_{h,\tau} \left(r^i(s_1, a_1) + \cdots + r^i(s_{n+m}, a_{n+m}) \right);$$

that is, the expected average payoff in the first n stages, conditional that the history h has occurred.

Let ϵ be fixed. A payoff vector $\gamma \in \mathbf{R}^{N \times S}$ is an *ϵ-payoff* of a profile τ if there exists $n_0 \in \mathbf{N}$ such that for every $n > n_0$ and every finite history h

$$\| \gamma_n(h, \tau) - \gamma \| < \epsilon.$$

In words, whatever be the history, if the current state is s then the expected payoff for the players if they wait long enough is approximately γ_s.

In the following theorem, v^i stands for the punishment level of player i. In the sequel it will either stand for the max-min value or the min-max value, according to whether the players correlate their actions or not in the punishment phase.

THEOREM 4.1 *Let $\gamma \in \mathbf{R}^{N\times S}$ be a payoff vector. If for every ϵ there exists a profile τ such that γ is an ϵ-payoff of τ, and for every state s, every history h whose last state is s, every player $i \in N$, every action $a^i \in \mathrm{supp}(\tau^i(h))$ and every action $b^i \in A^i$*

$$q_{s,\tau\ \ i(h).a^i}\gamma^i \geq q_{s,\tau\ \ i(h).b^i}v^i - \epsilon \tag{1}$$

then γ is an autonomous correlated equilibrium payoff.

Proof: This theorem is a weaker version of Proposition 3.7 in Solan (1998). We provide here a sketch of the proof.

We construct an extensive-form correlation device in the following manner. At every stage the device recommends an action to each player, and reveals to each player the actions that it recommended to *all* the players at the previous stage. This way, a deviation of any player is detected immediately by all other players, and can be punished with the punishment level.

Assuming that the players follow the recommendations of the device, the device knows what is the realized history that occurred, since it knows the previous states that the game has visited. The probability distribution according to which the device chooses actions for the players is $\tau(h)$, where h is the history that the device assumes that occurred.

The strategy profile of the players in the extended game is to follow the recommendations of the device as long as no deviation is detected. Once a deviation is detected, the deviator is punished with his punishment level.

Since γ is an ϵ-payoff of τ and by (1), no player can profit more than 4ϵ by deviating.

The way to construct an autonomous correlation device would be to enlarge the message space. Instead of sending at every stage to each player a single signal, the device sends a vector of signals, one for each possible history. The players, who observe the realized history, know which is the signal that they should follow, and discard all other signals. ■

For the rest of the section we restrict ourselves to recursive games. Hence, we may assume that the payoff function is independent of the action combination, and we denote the payoff for the players in state s by $r(s)$. In the sequel, v^i stands for the min-max value of player i.

PROPOSITION 4.2 *Assume the game is recursive. Let $\gamma \in \mathbf{R}^{S\times N}$ be a payoff vector, $x \in X$ be a mixed action combination, and $(C_1, C_2, \ldots, C_K, T)$ be a partition of $S \setminus S^\star$ such that $C_k \in \mathcal{C}(x)$ for every $k = 1, \ldots, K$, and every $s \in T$ is transient w.r.t x. Assume that for every $k = 1, \ldots, K$ there exists a probability distribution μ_k over $E(x, C_k)$ such that the following hold:*

1. *The Markov chain over S induced by μ_k for every $s \in C_k$ and by x for every $s \in T$ is absorbing.*

2. *For every state $s \in T$ and every player $i \in N$, $q_{s,x}\gamma^i = \gamma^i_s$, and for every action $a^i \in \mathrm{supp}(x^i_s)$ and every action $b^i \in A^i$*

$$q_{s,x_{-i},a^i}\gamma^i \geq q_{s,x_{-i},b^i}v^i.$$

3. *$\gamma^i_s = r^i(s)$ for every player i and every absorbing state $s \in S^\star$.*

Moreover, for every $k = 1, \ldots, K$ we have:

4. *For every unilateral exit $e \in \mathrm{supp}(\mu_k)$ of player i and every action $a^i \in A^i$*

$$q_e\gamma^i \geq q_{e_{-i},a^i}v^i.$$

5. *$\mu_k\gamma^i = \gamma^i_s \geq v^i_s$ for every player i and state $s \in C_k$.*

Then γ is an autonomous correlated equilibrium payoff.

Note that condition 2 implies that $\gamma^i_s \geq v^i_s$ for every state $s \in T$.

Proof: Fix $\epsilon > 0$. We shall construct a profile τ that, together with γ, satisfy the conditions of Theorem 4.1.

The profile τ indicates the players to play x_s in transient states $s \in T$, and to leave each communicating set C_k through the exits in $E(x, C_k)$ according to the probability distribution μ. That is, once a communicating set C_k is reached, then regardless of the past history, the players follow the profile $\tau_{C_k,\mu_k,x,\epsilon}$ until the play leaves C_k.

By condition 1 the game is bounded to be eventually absorbed, and by conditions 2, 3 and 5 the expected payoff for the players if they follow σ is γ_s, where s is the initial state. By condition 2 equation (1) holds for every state $s \in T$, and conditions 4 and 5 and since $\| \tau_{C_k,\mu_k,x,\epsilon}(h) - x \| < \epsilon$ for every history h, it holds for $s \in C_k$ as well. ■

The third condition that we derive is for positive recursive games.

PROPOSITION 4.3 *Assume the game is positive recursive. Let* $\gamma \in \mathbf{R}^{N\times S}$ *be a payoff vector, and* $x \in X$ *a stationary profile. Let* $(C_1,\ldots,C_K,T)$ *be a partition of* $S \setminus S^\star$ *such that* $C_k \in \mathcal{C}(x)$ *for every* $k = 1,\ldots,K$, *and every* $s \in T$ *is transient w.r.t.* x. *Assume that for every* $k = 1,\ldots,K$ *there exists a probability distribution* μ_k *over* $E(x,C_k)$ *such that the following holds:*

1. *The Markov chain over* S *induced by* μ_k *for every* $s \in C_k$ *and by* x *for every* $s \in T$ *is absorbing.*

2. *For every state* s *and every player* $i \in N$, $q_{s,x}\gamma^i = \gamma^i_s$, *and for every action* $a^i \in A^i$
$$q_{s,x[i,a^i]}\gamma^i \leq \gamma^i_s.$$

3. $\gamma^i_s = r^i(s)$ *for every state* $s \in S^\star$ *and every player* $i \in N$.

4. $\gamma^i_s \geq v^i_s$ *for every player* i *and every state* $s \in S$.

Moreover, for every $k = 1,\ldots,K$ *we have:*

5. $\mu_k\gamma^i = \gamma^i_s$ *for every player* i *and every state* $s \in C_k$.

6. *At least one of the following holds:*

   (a) $\mu_k[e] > 0$ *implies that* e *is a unilateral exit of some player.*

   (b) *If* e *is a unilateral exit of player* i *with* $\mu_k[e] > 0$, *then* $q_e\gamma^i = \gamma^i_s$ *for every* $s \in C_k$.

7. *If* $\mu_k[e_1] > 0$ *is a unilateral exits of player* i, *then* $q_{e_2}\gamma^i \leq q_{e_1}\gamma^i \leq \gamma^i_s$ *for every unilateral exit* e_2 *of player* i *from* C_k *and every* $s \in C_k$.

Then γ *is a stationary correlated equilibrium payoff.*

Proof: It is easy to verify that the conditions of the proposition imply the conditions of Proposition 4.2. Indeed, by condition 2 $q_{s,x[i,a^i]}\gamma^i = \gamma^i_s$ for every state $s \in T$, every player i and every action $a^i \in \mathrm{supp}(x^i_s)$. Thus by conditions 2 and 4 it follows that condition 2 of Proposition 4.2 holds. Condition 4 of Proposition 4.2 holds by condition 4, 7 and since the game is positive.

Thus γ is an autonomous correlated equilibrium payoff. In order to transform the device into a stationary one, we note the following:

- By condition 2 each player i is indifferent between actions in $\text{supp}(x_s^i)$, and by conditions 2 and 5 he prefers those actions over playing an action outside $\text{supp}(x_s^i)$, provided he will be punished afterwards by his punishment level.

  Therefore, in states $s \in T$ the players do not need the device: each player can privately choose an action according to x_s^i, and play it. If a player plays an action outside $\text{supp}(x_s^i)$ his deviation is noted, and can be punished by his punishment level.

- If the play enters a communicating set C_k that satisfies condition 6(b) then players are indifferent between their unilateral exits. Since the game is positive, all players prefer absorption to indefinite continuation of the game. It is well known (see Vieille (1997c)) that the profile $\tau_{C_k,\mu_k,x,\epsilon}$ can be constructed in such a way that any deviation of a player that changes the exit distribution from C_k by more than ϵ is noticed by the other players, and can therefore be punished.

  Thus, in this case the players do not need the correlation device as well.

- If the play enters a communicating set C_k that satisfies condition 6(a) then all exits are unilateral exits, and by condition 7 each player is indifferent between his unilateral exits. Moreover, by conditions 5 and 7 the expected payoff for each player if the game leaves C_k, conditional that the exit is *not* a unilateral exit of player i, is at least γ_s^i, for any $s \in C_k$.

  These observations leads us to the desired device. Denote $\alpha_i = \alpha_i(C_k) = \sum \mu[e]$, where the sum is over all unilateral exits e of player i. At the stage in which the game enters the set C_k the device chooses a player according to the probability distribution (α_i). The device then tells the chosen player that he was chosen, and the other players that they were not. The players then play a profile that visits any state in C_k infinitely often without leaving C_k, and the chosen player tries to use one of his unilateral exits, according to the probability distribution induced by μ.

  Since the device does not know when the play enters such a communicating set C_k, it should choose at every stage and for every such set C_k one player, and send to each player the sets for which he is the chosen

one. In the case that the play entered a communication set, each player checks whether he was chosen by the device, and plays accordingly.

■

5 Proofs of the Theorems

As we will see, the technique that is used for the general case is different from the technique used in the special case of recursive games. The first uses the method of Mertens and Neyman (1981), whereas the latter uses methods similar to that of Vieille (1997c).

5.1 Proof of Theorem 2.3

In this subsection we prove Theorem 2.3. The punishment level v^i stands for the min-max value.

Using the method of Mertens and Neyman (1981) for existence of the value in two-player zero-sum stochastic games, we construct for every $\epsilon > 0$ a profile τ that satisfies the assumptions of Theorem 4.1. Mertens and Neyman's work deals with two-player games. However, in an unpublished work Neyman (1988) showed that the results are valid for n-player games as well.

Fix $\epsilon > 0$. For every state $s \in S$ and every vector of discount factors $\lambda = (\lambda^i)_{i \in N}$, let $v^i_\lambda(s)$ be the discounted min-max value of player i when the initial state is s, and each player $j \in N$ uses λ^j as his discount factor. Note that v^i_λ is independent of $(\lambda^j)_{j \neq i}$. Let $x(s, \lambda)$ be an equilibrium strategy profile in the one shot game $G(s, \lambda)$ where the payoffs of each player i are given by

$$\lambda^i r^i(s, a) + (1 - \lambda^i) \sum_{t \in S} p(t|s, a) v^i_\lambda(t).$$

For every state $s \in S$, the set $\{(\lambda, x) \mid x \text{ is an equilibrium in } G(s, \lambda)\}$ is semi-algebraic. Hence there exists a continuous function $x(s, \cdot)$ that is defined in an open set $U = (0, \delta)^N$, and assigns for every vector of discount factors $\lambda \in U$ an equilibrium in $G(s, \lambda)$. Moreover, $x_s = \lim_{\lambda \to \mathbf{0}} x(s, \lambda)$ exists. Denote $x = (x_s)_{s \in S}$.

Let $\lambda_1 = (\lambda_1^i)_{i \in N} \in (0,1]^N$ be sufficiently close to $\mathbf{0}$ (in the supremum topology). Define a strategy profile $\tau(\lambda_1)$ as follows. At stage n the players play the mixed action $x(s_n, \lambda_n)$, where λ_n is calculated inductively from λ_{n-1} as appears in Mertens and Neyman (1981).

By construction we have

$$v^i_{\lambda_n}(s_n) \leq \mathbf{E}_{x(s_n,\lambda_n)} \left(\lambda_n^i r^i(s_n, a_n) + (1 - \lambda_n^i) v^i_{\lambda_n}(s_{n+1}) \right),$$

which is the basic equation needed by Mertens and Neyman.

Mertens and Neyman prove the following, provided that λ_1 is sufficiently close to $\mathbf{0}$.

1. For every $i \in N$, λ_n^i converges to 0 with probability 1.

2. For every $i \in N$, $v^i(s_n)$ converges with probability 1.

3. Given any finite history h, the expected average payoff of each player i in every sufficiently long game, if the players follow $\tau(\lambda_1)$, is, up to an ϵ, at least $v^i(s)$, where s is the last state of h.

It follows that if the players follow $\tau(\lambda_1)$ then with probability 1 the game enters some communicating set $C \in \mathcal{C}(x)$, and stays in it forever. Assume C is a minimal communicating set that satisfies this property. By 2 it follows that v_s is constant over C. Denote this common value by v_C. Denote by $\mathcal{C}^\star$ the set of all these communicating sets.

Fix a set $C \in \mathcal{C}^\star$. Let $\mathcal{E}_C = \{E\}$ be the collection of all ergodic sets w.r.t. $x = (x_s)_{s \in S}$ that are subsets of C, and let r_E be the average undiscounted payoff in the ergodic set E if the players follow x.

By 3 it follows (for a detailed proof see Vieille (1997a)) that there exists a convex combination of $(r_E)_{E \in \mathcal{E}}$ such that for every state $s \in C$, every player i and every action $a^i \in A^i$

$$\sum_{E \in \mathcal{E}_C} \alpha_E r_E \geq q_{s,x^{-i},a^i} v^i.$$

It is easy to construct a cyclic profile τ_C that never leaves C, such that $\| \tau_C(h) - x \| < \epsilon$ for every history h, and if the players follow τ_C then their expected average payoff in every sufficiently long game is approximately $\sum_{E \in \mathcal{E}_C} \alpha_E r_E$.

The desired strategy profile is, then, to follow $\tau(\lambda_1)$ whenever the game is not in any of the communicating sets in $\mathcal{C}^\star$, and once the game enters such communicating set C, to follow τ_C forever. ■

5.2 Recursive Games - Preliminaries

For the rest of the section we restrict ourselves to recursive games. This subsection is essentially a reminder of Vieille (1997b). It contains a number of useful tools. All proofs are omitted. They may be found in Vieille (1997b).

For every stationary profile x, $\gamma(x) = (\gamma_s(x))_{s\in S}$ is the expected undiscounted payoff forthe players if they follow x. A stationary profile $x \in X$ is *absorbing* if for every initial state, the probability to reach an absorbing state is 1, provided the players follow x.

For any stationary profile $x \in X$ and pure action combination $a \in A^S$ we define $x(a) = \prod_{s\in S, i\in N} x_s^i[a_s^i]$.

For every $a, b \in \mathbf{R}^n$, $a \geq b$ if and only if $a^i \geq b^i$ for every $i = 1, \ldots, n$.

It is of special interest to know how the distribution of exit from a set depends upon the (stationary) strategies used by the players.

For $B \subset S \setminus S^\star$, define a B-graph to be a set g of arrows $[s, a \to s']$, where $s \in B, a \in A, s' \in S$, such that :

1. for each $s \in B$, there is a unique pair a, s', such that $[s, a \to s'] \in g$; moreover, $q(s'|s, a) > 0$;

2. for each $s \in B$, there is a path $(s_0, a_0) \to (s_1, a_1) \to \ldots, \to s_N$, such that $s = s_0$, $s_N \notin B$, $[s_n, a_n \to s_{n+1}] \in g$.

The path in condition 2 is unique. We call it the g-path starting from s.

G_B is the set of B-graphs and, for $s \in B$, $s' \notin B$, $G_B(s \rightsquigarrow s')$ is the set of $g \in G_B$, such that the g-path starting from s ends up in s'.

For $x \in X$ and $g \in G_B$, we set

$$p_x(g) = \prod_{[s,a\to s'\in g]} x_s(a) q(s'|s, a).$$

$p_x(g)$ should be interpreted as the probability of g under x.

recall that $e_B = \inf\{n, s_n \notin B\}$ is the first exit stage from B. If B is transient under x, $e_B < +\infty$, $\mathbf{P}_{s,x}$-a.s. for every initial state $s \in B$.

It follows from a Lemma of Freidlin and Wentzell (1984) that

$$\mathbf{P}_{s,x}(s_{e_B} = s') = \frac{\sum_{G_B(s \rightsquigarrow s')} p_x(g)}{\sum_{G_B} p_x(g)}. \tag{2}$$

For simplicity, set $Q_{s,x}(s'|B) = \mathbf{P}_{s,x}(s_{e_B} = s')$. We will have to analyze the limit behavior of the play under a sequence (x_ϵ) of absorbing stationary profiles. From (2), one sees that relevant quantities are the ratios $\frac{x_\epsilon(a_1)}{x_\epsilon(a_2)}$, for $a_1, a_2 \in A^S$. Assume that $(x_\epsilon)_{\epsilon>0}$ is a family of stationary profiles such that

$$\theta_{a_1 a_2} = \lim_{\epsilon \to 0} \frac{x_\epsilon(a_1)}{x_\epsilon(a_2)}$$

exists, for every $a_1 \in A^S, a_2 \in \mathrm{supp}(x_\epsilon)$.

This implies that $\lim_{\epsilon \to 0} x_\epsilon$ exists in X. Moreover, this limit depends only on $(\theta_{a_1 a_2})_{a_1 a_2}$, and not on the exact sequence (x_ϵ). The limit is denoted by $x(\theta)$. One derives from (2) that $\lim_{\epsilon \to 0} Q_{s,x_\epsilon}(\cdot|B)$ exists in $\Delta(S \setminus B)$. It is denoted by $Q_{s,\theta}(\cdot|B)$.

The limit behavior of the play under x_ϵ is best described through a hierarchical decomposition of S into "transient" states and "ergodic" sets. Say that a set $B \subset S \setminus S^\star$ *communicates* for θ if $Q_{s,\theta}(s'|B \setminus \{s'\}) = 1$, for every $s \in B, s' \in B \setminus \{s\}$. This captures the property that, starting anywhere in B, the play will visit "infinitely" many times every state in B before leaving this set (as $\epsilon \to 0$). Denote by $\mathcal{C}(\theta)$ the collection of sets which communicate for θ. The following properties hold:

1. If $C \in \mathcal{C}(\theta)$, $Q_{s,\theta}(\cdot|C)$ is independent of $s \in C$;
2. If $C \in \mathcal{C}(\theta)$, then $C \in \mathcal{C}(x(\theta))$;
3. Given two sets in $\mathcal{C}(\theta)$, they are either disjoint or one is a subset of the other.

Therefore, $S \setminus S^\star$ can be partitioned as $\Pi(\theta) = (C_1, \ldots, C_K, T)$, where $C_1, \ldots, C_K$ are the maximal elements of $\mathcal{C}(\theta)$, and T is the set of those states which belong to no set in $\mathcal{C}(\theta)$ (they are in particular transient under $x(\theta)$).

We now explicit how $Q_\theta(\cdot|C)$ is related to q and $x(\theta)$, for $C \in \mathcal{C}(\theta)$. By possibly duplicating sets, one may assume that, for every $s \in C$, $a \in A$,

$$q(C|s,a) < 1 \Rightarrow q(C|s,a) = 0.$$

For any exit $e = (s, a^L) \in E(x(\theta), C)$, set $q_e = q(\cdot|s, x^{-L}(\theta), a^L)$; it should be thought of as the exit distribution induced by e. Then $Q_\theta(\cdot|C)$ is in the convex hull of $\{q_e, e \in E(x(\theta), C)\}$:

$$Q_\theta(\cdot|C) = \sum_e \mu_{\theta,C}[e] q_e$$

where $\mu_{\theta,C} \in \Delta(E(x(\theta), C)$ is defined as follows. $\mu_{\theta,C}[e]$ is the (limit, as $\epsilon \to 0$) probability that exit from C occurs through e. The following is true.

Let $e = (s, a^L) \in E(x(\theta), C)$. Pick $a^{-L} \in \mathrm{supp}(x^{-L}(\theta))$, $s' \in S \setminus C$ such that $q(s'|s, a^{-L}, a^L) > 0$, and set $a = (a^{-L}, a^L)$. One has $\mu_{\theta,C}(e) > 0$ if and only if there exists a graph $g_1 \in G_{C \setminus \{s\}}$, such that the graph g obtained by taking the union of g_1, and $[s, a \to s']$ satisfies:

$$\forall g' \in G_C, \theta_{gg'} > 0.$$

Although involved, this condition is highly intuitive: if some $g' \in G_C$ did satisfy $\theta_{gg'} = 0$, for every choice of g_1, then exit from C along g' would be infinitely more probable than exit through (s, a^L). This would contradict $\mu_{\theta,C}(e) > 0$.

5.3 Positive Recursive Games

We now prove Theorem 2.5. The proof goes as follows. For $\epsilon > 0$ small enough, we construct a stationary equilibrium x_ϵ of some ϵ-constrained game. We define $\theta = (\theta_{a_1 a_2})$ as the limit (up to a subsequence) of $(\frac{x_\epsilon(a_1)}{x_\epsilon(a_2)})_{a_1,a_2}$, and we prove that $x(\theta), \gamma(\theta)$ and $\Pi(\theta)$ fulfill the requirements of Proposition 4.3.

We assume below, *w.l.o.g.*, that, if $x \in X$ is fully mixed, the only ergodic sets of the corresponding Markov chain are the absorbing states. (If this were not true, then, by turning all ergodic sets w.r.t. such x in $S \setminus S^\star$ into absorbing states with payoff 0, one would get a game with the same set of correlated equilibrium payoffs, and with the desired property).

5.3.1 Constrained Games

For $\epsilon > 0$, define

$$X_\epsilon = \{x \in X \mid x_s^i(a^i) \geq \epsilon^2, \forall i \in N, s \in S, a^i \in A^i\}.$$

We define a continuous map from X_ϵ into itself.

Let $x \in X_\epsilon$, $i \in N$. Define the *continuation cost* of playing a^i in state s against x as

$$c_s(a^i; x) = \max_{\tilde{a}^i \in A^i} q_{s,x^{-i},\tilde{a}^i}\gamma^i(x) - q_{s,x^{-i},a^i}\gamma^i(x), \tag{3}$$

given the continuation payoff of player i is $\gamma^i(x)$, this is the amount that player i gives up by playing a^i, rather than his best reply.

Notice that $0 \leq c_s(a^i; x) \leq 1$, $c_s(a^i; x) = 0$ for each a^i which attains the maximum in (3), and $x \mapsto c_s(a^i; x)$ is continuous over X_ϵ.

For $a^i \in A^i$, $s \in S$, set

$$f_s^i(x)[a^i] = \frac{\epsilon^{c_s(a^i;x)}}{\sum_{\tilde{a}^i \in A^i} \epsilon^{c_s(\tilde{a}^i;x)}}.$$

and $f(x) = (f_s^i(x))_{i \in N, s \in S}$. Observe that $\frac{\epsilon}{|A^i|} \leq f_s^i(x)[a^i] \leq 1$, and $\sum_{a^i \in A^i} f_s^i(x)[a^i] = 1$. Therefore for ϵ sufficiently small $f(x) \in X_\epsilon$. The continuity of f follows from the continuity of continuation costs. By Brouwer's Theorem, f has a fixed point. We denote it by x_ϵ.

Intuitively, in this fixed point each player i plays the action a^i with probability that depends on its cost – the higher the cost, the smaller probability it receives. As ϵ tends to 0, the ratio between the probabilities in which two actions whose cost differ by a constant tends to infinity. Thus, if, as ϵ tends to 0, two actions are played with "comparable" probabilities, then their cost is the same.

5.3.2 Asymptotic Analysis

We study here the asymptotic properties of x_ϵ, as ϵ tends to 0. Up to a subsequence, we may assume that $x = \lim_{\epsilon \to 0} x_\epsilon$ and $g = \lim_{\epsilon \to 0} \gamma(x_\epsilon)$ exist (of course, g needs not be equal to $\gamma(x)$). Thus, $c_s(a^i; x_\epsilon)$ has a limit $c_s(a^i; \theta)$,

for every player $i \in N$ and every state $s \in S$. Note that $c_s(a^i;\theta) = 0$ does *not* imply that $a^i_s \in \mathrm{supp}(x^i_s)$. For every $a_1, a_2 \in A^S$

$$\theta_{a_1 a_2} = \lim_{\epsilon \to 0} \frac{x_\epsilon(a_1)}{x_\epsilon(a_2)}$$

exists, (it is possibly infinite). Moreover, $x(\theta) = x$ and

$$c_s(a^i;\theta) = \max_{\tilde{a}^i \in A^i} q_{s,x^{-i},\tilde{a}^i} g^i - q_{s,x^{-i},a^i} g^i.$$

Write $\Pi(\theta) = (C_1, \ldots, C_K, T)$ the decomposition of $S \setminus S^\star$ into maximal communicating sets and transients states w.r.t. $x_\cdot$. For each k, denote $\mu_k = \mu_{\theta,C_k}$. We shall prove that conditions 1 through 7 of Proposition 4.3 are satisfied. We start with the simplest.

- Condition 1 follows from the definition of $\Pi(\theta)$;
- For every $\epsilon > 0$, $s \in S^\star$, $\gamma_s(x_\epsilon) = r(s)$. Condition 3 follows;
- For every $\epsilon > 0$ and $s \in S$,

$$q_{s,x_\epsilon} \gamma(x_\epsilon) = \gamma_s(x_\epsilon). \tag{4}$$

  The first claim of condition 2 follows by taking the limit $\epsilon \to 0$. On the other hand, property (4) means that, for all s, the sequence $(\gamma_{s_n}(x_\epsilon))$ is a martingale under $\mathbf{P}_{s,x_\epsilon}$ (for the filtration $(\mathcal{H}_n)$, where $\mathcal{H}_n$ is the σ-algebra over the space of infinite histories induced by H_n). By the Optional Sampling Theorem, for every k and every $s \in C_k$,

$$\mathbf{E}_{s,x_\epsilon}[\gamma_{s_{\sigma_B}}(x_\epsilon)] = \gamma_s(x_\epsilon).$$

  Condition 5 follows by letting $\epsilon \to 0$.

- By construction, $x^i_s(a^i) > 0$ implies $\lim_\epsilon c_s(a^i; x_\epsilon) = 0$, therefore $q_{s,x^{-i},a^i} g^i = \max_{A^i} q_{s,x^{-i},\cdot} g^i$. By summation over $a^i \in \mathrm{supp}(x^i_s)$, one gets

$$q_{s,x} g^i = \max_{A^i} q_{s,x^{-i},\cdot} g^i.$$

  Since $q_{s,x} g^i = g^i(s)$, the second part of condition 2 is established.

Condition 7 is implied by the next lemma.

LEMMA 5.1 *Let* $e_1 = (s_1, a_1^{L_1}), e_2 = (s_2, a_2^{L_2}) \in E(x; C_k)$. *If* $\sum_{i \in L_1} c_{s_1}(a_1^i; \theta) < \sum_{i \in L_2} c_{s_2}(a_2^i; \theta)$, *then* $\mu_k(e_2) = 0$.

Proof: For every subset $C \subseteq S$ and every graph $g \in G_C$, define the *overall cost* of g by

$$c(g; \theta) = \sum_{i \in N} \sum_{[s, a \to s'] \in g} c_s(a'; \theta).$$

Under $y_{C_k, x, \epsilon, s_1}$, C_k is stable and, starting from C_k, the play reaches s_1 before c_{C_k}. Therefore, one can construct a graph $g = \{[s, a_s \to s']\} \in G_{C_k \setminus s_1}$ such that a_s^i belongs to the support of $y^i_{C_k, x, \epsilon, s_1}$, for every $s \in C_k \setminus s_1$ and every player $i \in N$.

This implies that $q(C_k | s, x^{-i}, a_s^i) = 1$, thus $c_s(a_s^i; \theta) = 0$.

Let $g' \in G_{C_k \setminus s_2}$ be arbitrary. We now define two graphs $g_0, g_0' \in G_{C_k}$ by adding exits in the support of $a_1^{L_1}$ and $a_2^{L_2}$ respectively. Formally, pick $a_1^{-L_1} \in \text{supp}(x_{s_1})^{-L_1}$ and $s_1' \notin C_k$, with $q(s_1' | s_1, a_1^{-L_1}, a_1^{L_1}) > 0$, and define g_0 as g with the additional arrow $[s_1, (a_1^{-L_1}, a_1^{L_1}) \to s_1']$. Define g_0' in an analogous way.

By the assumption and the above discussion, $c(g_0) < c(g_0')$, which implies that

$$\lim_{\epsilon \to 0} \frac{p_{x_\epsilon}(g_0')}{p_{x_\epsilon}(g_0)} = 0.$$

Since g' is arbitrary, the result follows. ∎

COROLLARY 5.2 *Condition 6 holds.*

Proof: Assume that (a) does not hold: there exist $e_1 = (s_1, a_1^{L_1}) \in \text{supp}(\mu_k)$, with $|L_1| > 1$. For every $j \in L_1$, $q(C_k | s_1, x^{-j}, a_1^j) = 1$, therefore $c_{s_1}(a_1^j; \theta) = 0$.

Let $e_2 = (s_2, a^i)$ be a unilateral exit of player i, with $\mu_k(e_2) > 0$. From Lemma 5.1, one deduces

$$0 \leq c_{s_2}(a_2^i; \theta) \leq \sum_{L_2} c_{s_2}(a_2^j; \theta) \leq \sum_{L_1} c_{s_1}(a_1^j; \theta) = 0$$

Thus $c_{s_2}(a^i; \theta) = 0$, which is (b) of condition 6. ∎

LEMMA 5.3 *Condition 4 holds, that is* $g \geq v$.

Proof: Assume that the inequality $g^i \geq v^i$ does not hold for player i. Let $S_0 \subseteq S$ contain the states where $v^i - g^i > 0$ is maximal. Since $v^i(s) = g^i(s)$ for $s \in S^\star$, $S_0 \cap S^\star = \emptyset$. Since $g^i(s) \geq 0$ for every s, $v^i(s) > 0$ for $s \in S_0$. Let $S_1 \subseteq S_0$ contain the states of S_0 where v^i is maximized. There exist $s \in S_1, a^i \in A^i$, with

$$q(S_1|s, x^{-i}, a^i) < 1 \tag{5}$$

and

$$q_{s,x^{-i},a^i} v^i \geq v^i(s) = \max_{S_0} v^i \tag{6}$$

(otherwise, players $N \setminus \{i\}$ could bring player i's payoff below v^i by playing x^{-i} on S_1, and punishing him if the play leaves S_1). By construction of S_1, (5) and (6) imply that $q(S_0|s, x^{-i}, a^i) < 1$.

On the other hand, $q_{s,x^{-i},a^i} g^i \leq g^i(s)$ by condition 2. Thus

$$q_{s,x^{-i},a^i} v^i - q_{s,x^{-i},a^i} g^i \geq v^i(s) - g^i(s) = \max_S (v^i(\cdot) - g^i(\cdot))$$

This implies $q(S_0|s, x^{-i}, a^i) = 1$ a contradiction. ∎

5.4 Recursive Games

In this section, we modify the foregoing proof to handle the case of recursive games. In doing so, we lose some properties, and will only be able to fulfill the requirements of Proposition 4.2.

It is convenient to divide the construction into two parts. Define $S_1 = \{s \in S \setminus S^\star, v(s) \leq 0\}$, and $S_2 = \{s \in S \setminus S^\star, v^i(s) > 0 \text{ for some } i\}$. Thus $S \setminus S^\star = S_1 \cup S_2$. We first construct a stationary profile on S_1, then adapt to S_2 the previous sequence of constrained games in order to construct the ingredients of Proposition 4.2.

5.4.1 A Profile on S_1

Let $s \in S_1$. For $x = (x^i)_{i \in N} \in \prod_i \Delta(A^i)$, set $\varphi_s^i(x) = \text{argmax}_{\Delta(A^i)} q_{s,x^{-i},\cdot} v^i$: $\varphi_s^i(x)$ is the set of mixed actions of player i which maximize against x^{-i} the expected continuation min max value. Set $\varphi_s(x) = \times_{i \in N} \varphi_s^i(x)$. It is

immediate to check that $\phi_s(x)$ is a non-empty and convex set, and that ϕ_s is upperhemicontinuous on $\prod_i \Delta(A^i)$. Therefore, by Kakutani's Theorem, ϕ_s has a fixed point x_s.

Given the collection $(x_s)_{s\in S_1}$, we simplify the game by assuming that the players follow x_s in each state $s \in S_1$. This amounts to replacing each state $s \in S_1$ by a *dumb* state with transitions $q(\cdot|s, x_s)$. There may exist ergodic sets for $(x_s)_{s\in S_1}$ within S_1. Let C be such a set. Then v is constant over C. Denoting by $v(C)$ its value over C, $v(C) \leq 0$, and every unilateral exit of player i from C lowers the expected level of v^i:

$$\forall i \in N, s \in C, a^i \in A^i, \ q_{s,x_s^{-i},a^i} v^i \leq v^i(s) \leq 0.$$

Therefore, starting from any state in C, the following profile is an ϵ-equilibrium profile: play (x_s), and punish player i (with an ϵ-min max profile) if the play leaves C as a consequence of a unilateral deviation of player i.

In the sequel, C will be identified to an absorbing state with payoff 0 (except when punishment is taking place).

5.4.2 Constrained Games on S_2

The definition of X_ϵ is amended as follows: player i is constrained in a state s only if $v^i(s) > 0$: define

$$\tilde{X}_\epsilon = \{x \in X \mid x_s^i(a^i) \geq \epsilon^2, \text{ if } a^i \in A^i, v^i(s) > 0\}.$$

The definition of continuation cost is unchanged:

$$c_s(a^i; x) = \max_{\tilde{a}^i \in A^i} q_{s,x^{-i},\tilde{a}^i} \gamma^i(x) - q_{s,x^{-i},a^i} \gamma^i(x)$$

To get continuity of the continuation cost, we have to prove that γ is continuous over $\tilde{X}_\epsilon$. It is sufficient to prove that the ergodic decomposition of S into ergodic sets and transient states is independent of $x \in \tilde{X}_\epsilon$.

LEMMA 5.4 *For every $x \in \tilde{X}_\epsilon$, the Markov chain induced by x is absorbing.*

Proof: Assume not. Let C be an ergodic set under $x \in X_\epsilon$. Let i be a player such that $v^i(s) > 0$, for some $s \in C$, and let C_0 be those states in C for which v^i is maximal. Thus, in C_0, player i plays every action with

positive probability. Then, starting from $s \in C_0$, players $N \setminus \{i\}$ can bring player i's payoff below $v^i(s)$ by playing x^{-i} until the first exit from C_0, and from then on an ϵ-min max profile. ■

We now define a "best-reply" map from $\tilde{X}_\epsilon$ into itself. Let $i \in N, s \in S$.

- if $v^i(s) > 0$, define $(\varphi_\epsilon)^i_s(x) \in \Delta(A^i)$ by

$$(\varphi_\epsilon)^i_s(x)[a^i] = \frac{\epsilon^{c_s(a^i;x)}}{\sum_{\tilde{a}^i \in A^i} \epsilon^{c_s(\tilde{a}^i;x)}}$$

- if $v^i(s) \leq 0$, set $(\varphi_\epsilon)^i_s(x) = \text{argmax}_{y^i \Delta(A^i)} q_{s,x^{-i},y^i} v^i$.

Set $\varphi_\epsilon(x) = (\times_{i \in N}(\varphi_\epsilon)^i_s(x))_{s \in S}$. It is obvious that $\phi_\epsilon(x)$ is a non-empty and convex subset of $\tilde{X}_\epsilon$, and that ϕ_ϵ is upperhemicontinuous on $\tilde{X}_\epsilon$. Therefore, by Kakutani's Theorem, φ_ϵ has a fixed point x_ϵ.

5.4.3 Asymptotic Analysis

We study here the asymptotic properties of x_ϵ, as ϵ tends to 0. Up to a subsequence, we may assume that the support of x_ϵ is independent of ϵ and that

$$\theta_{a_1 a_2} = \lim_{\epsilon \to 0} \frac{x_\epsilon(a_1)}{x_\epsilon(a_2)} \tag{7}$$

exists, for every $a_1 \in A^{S_2}, a_2 \in \text{supp}(x_\epsilon)$. Therefore, the ergodic decomposition of S into transient states and ergodic sets is independent of ϵ.

Property (7) implies that $g = \lim_{\epsilon \to 0} \gamma(x_\epsilon)$ exists. Write $\Pi(\theta) = (C_1, \ldots, C_K, T)$: it is a partition of $S \setminus S^\star$. For each k, denote $\mu_k = \mu_{\theta, C_k}$.

We shall prove that conditions 1 through 5 of Proposition 4.2are satisfied.

Conditions 1, 3 and the first part of condition 2 hold for the very same reason as in the case of positive recursive games. The other requirements follow from Lemmas 5.5 through 5.7 below.

We start with a simple observation: for every state $s \in S$, every player $i \in N$ and every mixed action combination $x \in X$ there exists $a^i \in A^i$ such that $q_{s,x^{-i},a^i} v^i \geq v^i(s)$. From this, one deduces that

$$v^i(s) \leq 0 \Rightarrow q_{s,x_\epsilon} v^i \geq v^i(s) \tag{8}$$

LEMMA 5.5 $g \geq v$.

Proof: Assume that $g_s^i < v^i(s)$, for some player $i \in N$ and state $s \in S$. We mimic the corresponding proof for positive recursive games (Lemma 5.3). Denote by $S_0 \subset S$ those states for which $v^i - g^i$ is maximal, and set $S_1 = \text{argmax}_{s \in S_0} v^i(s)$. There are two cases. If $v^i(S_1) > 0$, the proof is identical to that of Lemma 5.3. Otherwise, $v^i(S_1) \leq 0$. By construction, $q_{s,x_\epsilon} v^i \geq v^i(s)$, for every $s \in S_1$, and $\epsilon > 0$. This implies

$$Q_{s,\theta}(\cdot|S_1)v^i \geq v^i(S_1). \tag{9}$$

for every $s \in S_1$. By definition of S_1, this implies in turn $Q_{s,\theta}(S_0|S_1) < 1$. On the other hand, $Q_{s,\theta}(\cdot|S_1)g^i = g^i(s)$. Therefore, using (9),

$$Q_{s,\theta}(\cdot|S_1)v^i - Q_{s,\theta}(\cdot|S_1)g^i \geq v^i(s) - g^i(s).$$

By definition of S_0, this yields $Q_{s,\theta}(S_0|S_1) = 1$. A contradiction. ∎

This readily implies that condition 5 is satisfied.

LEMMA 5.6 *The second assertion in condition 2 is satisfied:*

$$q_{s,x^{-i},a^i} g^i \geq q_{s,x^{-i},b^i} v^i \qquad \forall i \in N, s \in S, a^i \in \text{supp}(x_s^i), b^i \in A^i.$$

Proof: Let s, i, a^i, b^i be as in the statement. There are two cases.

If $v^i(s) \leq 0$, x^i maximizes $q_{s,x^{-i},\cdot} v^i$ over $\Delta(A^i)$. In particular,

$$q_{s,x^{-i},a^i} v^i \geq q_{s,x^{-i},b^i} v^i.$$

The result follows in that case, since $g^i \geq v^i$.

If $v^i(s) > 0$, x^i maximizes $q_{s,x^{-i},\cdot} g^i$ over $\Delta(A^i)$. In particular,

$$q_{s,x^{-i},a^i} g^i \geq q_{s,x^{-i},b^i} g^i.$$

The result follows again in that case. ∎

We finally prove condition 4. It is based on the following technical lemma.

LEMMA 5.7 *Let $s_0 \in C \subseteq S \setminus S^\star$, $x \in X$ and $i \in N$. Assume $v^i(s_0) > 0$. Then there exists a sequence $s_1, \ldots, s_R \in C$ of distinct states, and actions $a_r^i \in A^i$, $r = 0, \ldots R$ such that:*

- a_r *maximizes* $q_{s_r,\cdot,x^{-i}}v^i$ *for each* $r = 0,\ldots,R$.
- $q(C|s_r,a_r^i,x^{-i}) = 1$, *for* $r = 0,\ldots,R-1$;
- s_r *maximizes* $v^i(\cdot)$ *in* $\mathrm{supp}q(\cdot|s_{r-1},a_{r-1}^i,x^{-i})$, *for each* $r = 1,\ldots,R$;
- $q(C|s_R,a_R^i,x^{-i}) < 1$.

Proof: Otherwise, there exists a subset C' of C such that $s_0 \in C'$ and any unilateral exit of player i from C' would lower his expected min max value:

$$\forall s \in C', a^i \in A^i, q(C'|s,a^i,x^{-i}) < 1 \Rightarrow q_{s,a^i,x^{-i}}v^i < v^i(C').$$

Therefore, given the initial state is s_0, players $N \setminus \{i\}$ can bring player i's payoff below $v^i(s_0)$ by playing x^{-i} and punish player i as soon as the play leaves C'. ■

LEMMA 5.8 *Let* $k \in \{1,\ldots,K\}$, $s \in C_k$, $i \in N$ *and* $a^i, b^i \in A^i$. *If* $e = (s,b^i)$ *is a unilateral exit of player* i *such that* $\mu_k(e) > 0$, *then* $q_{s,b^i,x^{-i}}g^i \geq q_{s,a^i,x^{-i}}v^i$.

Proof: There are two cases. Assume first that $v^i(s) \leq 0$. The assumption entails in particular that $x_{s,e}^i(b^i) > 0$. Therefore b^i maximizes $q_{s,\cdot,x^{-i}}v^i$. Hence by Lemma 5.5

$$q_{s,a^i,x^{-i}}v^i \leq q_{s,b^i,x^{-i}}v^i \leq q_{s,b^i,x^{-i}}g^i.$$

Assume now that $v^i(s) > 0$. Set $s_0 = s$, and construct a sequence $s_1,\ldots,s_R$ as in Lemma 5.7. Since $\mu_k(e) > 0$ it follows that

$$c_{s_0}(b^i;\theta) \leq \sum_{r=0}^{R} c_{s_r}(a_r^i;\theta).$$

For $r < R$, $q(C_k|s_r,a_r^i,x^{-i}) = 1$, hence $c_{s_r}(a_r^i;\theta) = 0$. Therefore, $c_{s_0}(b^i;\theta) \leq c_{s_R}(a_R^i;\theta)$, which yields

$$q_{s,b^i,x^{-i}}g^i \geq q_{s_R,a_R^i,x^{-i}}g^i. \tag{10}$$

On the other hand, by construction,

$$q_{s_R,a_R^i,x^{-i}}v^i \geq q_{s_0,a_0^i,x^{-i}}v^i = \max_{A^i} q_{s_0,\cdot,x^{-i}}v^i \geq q_{s,a^i,x^{-i}}v^i. \tag{11}$$

From (10) and (11), and using $g^i \geq v^i$, one deduces $q_{s,b^i,x^{-i}}g^i \geq q_{s,a^i,x^{-i}}v^i$. ■

References

[1] Aumann R.J. (1974) Subjectivity and Correlation in Randomized Strategies, *J. Math. Econ.*, **1**, 67-96

[2] Aumann R.J. (1987) Correlated Equilibrium as an Expression of Bayesian Rationality, *Econometrica*, **55**, 1-18

[3] Blackwell D. and Ferguson T. S. (1968) The Big Match, *Ann. of Math. Stat.*, **39**, 159-163

[4] Everett H. (1957) Recursive Games, *Contributions to the theory of Games*, **3**, 47-78, Princeton N.J., Annals of Mathematical Studies, **39**, Princeton University Press.

[5] Fink A.M. (1964) Equilibrium in a Stochastic n-Person Game, *J. Sci. Hiroshima Univ.*,**28**, 89-93

[6] Forges F. (1986) An Approach to Communication Equilibria, *Econometrica*, **54**, 1375-1385

[7] Forges F. (1988) Communication Equilibria in Repeated Games with Incomplete Information, *Math. Oper. Res.*, **13**, 2, 77-117

[8] Fudenberg D. and Tirole J. (1991) *Game Theory*, The MIT Press

[9] Mertens J.F. (1994) Correlated and Communication Equilibria, In *Game Theoretic Methods in General Equilibrium Analysis*, Mertens J.F. and Sorin S. (eds.), Kluwer Academic Press

[10] Mertens J.F. and Neyman A. (1981) Stochastic Games, *Int. J. Game Th.*, **10**, 53-66

[11] Myerson R.B. (1986) Multistage Games with Communication, *Econometrica*, **54**, 323-358

[12] Neyman A. (1988) Stochastic Games, *preprint*

[13] Nowak A.S. (1991) Existence of Correlated Weak Equilibria in Discounted Stochastic Games with General State Space, In *Stochastic Games an Related Topics*, Raghavan et al. (eds), Kluwer Academic Press

[14] Rosenberg D. and Vieille N. (1998) The MaxMin of Recursive Games with Incomplete Information, *Cahiers du Laboratoire d'Econométrie de l'Ecole Polytechnique*, #476

[15] Shapley L.S. (1953) Stochastic Games. *Proc. Nat. Acad. Sci. U.S.A.*, **39**. 1095-1100

[16] Solan E. (1997) 3-Person Repeated Games with Absorbing States, D.P. #128, Center for Rationality and Interactive Decision Theory, the Hebrew University, Jerusalem

[17] Solan E. (1998) Extensive-Form Correlated Equilibria, , D.P. #175, Center for Rationality and Interactive Decision Theory, the Hebrew University, Jerusalem

[18] Solan E. and Vieille N. (1998) Quitting Games, *preprint*

[19] Thuijsman, F. and Vrieze, K. (1992) A note on Recursive Games, *Game Theory and Economic Applications*, B. Dutta and al. (eds), Lecture Notes in Economics and Mathematical Systems 389, Springer Verlag, Berlin, 133-145

[20] Vieille N. (1997a) Equilibrium in 2-Person Stochastic Games I: A Reduction, CEREMADE D.P. #9745

[21] Vieille N. (1997b) Large Deviations and Stochastic Games, CEREMADE D.P. #9746

[22] Vieille N. (1997c) Equilibrium in 2-Person Stochastic Games II: The Case of Recursive Games, CEREMADE D.P. #9747