

# NONCOOPERATIVE STOCHASTIC GAMES UNDER A $N$ -STAGE LOCAL CONTRACTION ASSUMPTION\*

MICHÈLE BRETON

*Ecole des Hautes Études Commerciales, 5255 Decelles, Montréal, Canada, H3T 1V6*

PIERRE L'ECUYER

*Département d'informatique, Université Laval, Ste-Foy, Québec, Canada, G1K 7P4*

*(Received 25 January 1988; in final form 20 September 1988)*

We consider a class of noncooperative stochastic games with general state and action spaces and with a state dependent discount factor. The expected time duration between any two stages of the game is not bounded away from zero, so that the usual  $N$ -stage contraction assumption, uniform over all admissible strategies, does not hold. We propose milder sufficient regularity conditions, allowing strategies that give rise with probability one to any number of simultaneous stages.

We give sufficient conditions for the existence of equilibrium and  $\varepsilon$ -equilibrium stationary strategies in the sense of Nash. In the two-player zero-sum case, when an equilibrium strategy exists, the value of the game is the unique fixed point of a specific functional operator and can be computed by dynamic programming.

KEY WORDS: Stochastic games, contraction mappings, dynamic programming, equilibria.

## 1. INTRODUCTION

Noncooperative stochastic games are mathematical models which can represent the behavior of a group of competing decision makers, acting on a dynamic system at discrete points in time, under uncertainty. Discounted stochastic games were introduced by Shapley [13]. His original model was a zero-sum, two player game with finite state and action spaces. More general models of discounted stochastic games were later introduced [3, 9, 11, 12, 14-16], some with an arbitrary number of players, nonzero-sum, state dependent discount factors, and general state and action spaces. These models have been analyzed under contraction assumptions that are uniform over all admissible policies.

The aim of this paper is to analyze a class of stochastic games that violate such uniform contraction assumptions. In these games, any number of successive stages may occur simultaneously, so that the  $N$ -stage expected discount factor is not bounded away from one. We focus on noncooperative Nash equilibria or  $\varepsilon$ -

---

\*Research supported by NSREC-Canada grants A4088 and A5463, and FCAR-Québec grant EQ2831.

equilibria [10], under the total expected discounted reward criterion. Consider for example a system with state space  $S$  and finite player set  $I$ . The system evolves in continuous time but is observed only at discrete points  $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \dots$ . At each observation time  $\tau_i$  (i.e. at each stage of the game), each player selects a decision, and among these decisions, one of the players must decide (or select a probability law according to which will be generated) the time  $\tau_{i+1} - \tau_i$  until the next observation time. Also at each stage, returns are given to each player depending on the state of the system and the decisions selected. All returns are discounted at rate  $\rho > 0$ . The point to note is that for such a system, the expected duration between any two successive stages of the game is not bounded away from zero: it can be smaller than  $\varepsilon$  for any  $\varepsilon > 0$ . Therefore, in principle, there is no upper bound on the number of stages in a given time interval. This substantially complicates the analysis.

Models of this kind can be useful for instance to study the sharing of resources among different agents. A resource may deteriorate with time and break down in some cases. An agent may hold the resource for a given amount of time (which can be zero), before passing it to another agent. When receiving the resource, he may also perform some maintenance on it to improve its utility or lower its failure rate. An observation time occurs every time a resource changes hands.

In this paper, we introduce *local contraction assumptions*, similar to those introduced by L'Ecuyer and Haurie [8] for Markov renewal programs, and show that under these weaker assumptions, it is still possible to analyze the model via a contraction mapping approach [4]. This model generalizes the locally contracting discrete event dynamic programming model developed in [7, 8]. The latter was motivated by the modeling of continuous-time maintenance/replacement systems [6, 7], for which the  $N$ -stage contraction assumption does not hold. The generalization to stochastic games is not straightforward, and the proofs given here differ substantially from those given in [7, 8]. We note that in [7, 8], the development was made under a *one-stage* local contraction assumption, but the properties derived there can be generalized easily to a  $N$ -stage locally contracting model by a reasoning similar to the one used here.

In Section 2, we state our  $N$ -stage locally contracting sequential game model, and examine its relationship with the models studied previously. Basic properties of the associated total expected reward functions are examined in Section 3. In Sections 4 and 5, we derive existence results similar to those obtained by Nowak [11] and Whitt [16] respectively, but under the  $N$ -stage local contraction assumptions. An example is given in Section 6. For the special case of a two-player zero-sum game (also called a *duel*) with Borel state and action spaces, we obtain sufficient conditions for the existence of a Nash equilibrium in stationary strategies (a saddle point). When a Nash equilibrium exists in a duel, the value of the game is the unique fixed point of a specific functional operator and can be computed by dynamic programming. For the general case (many players, non zero-sum), we obtain sufficient existence conditions for an equilibrium if the state space is countable, and for an  $\varepsilon$ -equilibrium if it is uncountable. In this paper, we assume bounded rewards, but the model could be easily modified as in Whitt [16] to

allow for unbounded rewards. For the definitions concerning the measurability concepts used in this paper, we refer to [1].

## 2. A LOCALLY CONTRACTING SEQUENTIAL GAME MODEL

We consider a stochastic sequential game model with Borel *state space*  $S$  and finite *player set*  $I$ . For each player  $i$  in  $I$ , let  $A_i$  be a separable metric space of actions, and for each state  $s$  in  $S$ , let  $A_i(s) \subseteq A_i$  be the non-empty compact set of *admissible actions* to player  $i$  when the system is in state  $s$ . The mappings  $s \rightarrow A_i(s)$  are assumed to be Borel-measurable. To allow for randomized strategies, we assume that each action in  $A_i$  is in fact a *mixed action*, i.e. a probability measure over an underlying set of *pure actions*  $B_i$ . For each  $s$  in  $S$ , let  $B_i(s)$  be a non-empty compact subset of a separable metric space  $B_i$ , and  $A_i(s)$  the set of all probability measures on  $B_i(s)$ , endowed with the weak topology. At each of an infinite sequence of stages (decision times), every player  $i$  observes the state  $s$  of the system and selects an action  $a_i$  in  $A_i(s)$ . Let  $a = (a_i, i \in I) \in A(s) = \prod_{i \in I} A_i(s)$  be the corresponding action vector. The expected return to player  $i$  for the current stage is  $r(s, i, a)$ , and the system moves to a new state  $s'$  according to a probability measure  $q(\cdot | s, a)$  over  $S$ . A new action vector  $a'$  is selected from  $A(s')$ , and so on. The (expected) *one-stage return function*  $r(s, i, a)$  is a bounded Borel-measurable real-valued function of  $s \in S$ ,  $i \in I$  and  $a \in A(s)$ , and the *law of motion* is given by the family of probability measures  $\{q(\cdot | s, a) | s \in S, a \in A(s)\}$ , which form a Borel-measurable stochastic kernel on  $S$  given  $s \in S$  and  $a \in A(s)$  (for a given  $\bar{S}$ , Borel subset of  $S$ ,  $q(\bar{S} | s, a)$  is a Borel-measurable function of  $(s, a)$ ).

Associated with each state  $s$  is a *discount factor*  $\beta(s)$ , where  $\beta: S \rightarrow (0, 1]$  is Borel measurable. Each  $q(\cdot | s, a)$  is assumed to be concentrated on the set of states  $s'$  for which  $\beta(s') \leq \beta(s)$ , so that

$$\alpha(s, a) = \frac{1}{\beta(s)} \int_S \beta(s') q(ds' | s, a) \quad (1)$$

which represents the one-stage expected discount factor, takes its value in  $(0, 1]$ . Thus, with probability one, the discount factor never increases.

A *policy*  $\delta_i$  for player  $i$  is a Borel-measurable function from  $s \in S$  into  $A_i(s)$ , under which player  $i$  takes the mixed action  $\delta_i(s)$  whenever the system is in state  $s$ . Let  $\Delta_i$  denote the set of policies for player  $i$ , and  $\Delta = \prod_{i \in I} \Delta_i$  be the space of policy vectors. A Markov behaviour *strategy* for player  $i$  is a sequence  $\pi_i = (\delta_i^0, \delta_i^1, \dots, \delta_i^t, \dots)$  such that  $\delta_i^t \in \Delta_i$  for each  $t$ . Let  $\Pi_i$  represent the set of all such strategies for player  $i$  and  $\Pi = \prod_{i \in I} \Pi_i$  the set of all strategy vectors, of the form  $\pi = (\delta^0, \delta^1, \dots, \delta^t, \dots)$ , where  $\delta^t \in \Delta$  for each  $t$ . A strategy for player  $i$  is called *stationary* if all  $\delta_i^t$  are identical ( $\equiv \delta_i$ ), i.e. if the same policy is used at each stage. In this case, we also use the symbol  $\delta_i$  to denote that stationary strategy, and the symbol  $\delta$  to denote the stationary strategy vector  $\pi = (\delta, \delta, \dots)$ . In this paper, we consider only Markov strategies. It is well known (see Proposition 8.1 in [1]) that

if all other players use a fixed stationary Markov strategy, no single player acting alone can improve his expected return by using a more general strategy.

From the Ionescu–Tulcea theorem, we know that associated with any initial state  $s \in S$  and strategy vector  $\pi \in \Pi$ , there is a uniquely defined probability measure  $P_{\pi, s}$  and a corresponding mathematical expectation  $E_{\pi, s}$  over the set of all infinite sequences  $(s^0, a^0, s^1, a^1, \dots, s^t, a^t, \dots)$ , where  $s^0 = s$ ,  $s^t \in S$  and  $a^t \in A(s^t)$  for all  $t \geq 0$ . The pair  $(s^t, a^t)$  represents the state of the system and the action vector chosen at stage  $t$ . We also denote by  $\beta^t$  the value of  $\beta(s^t)/\beta(s^0)$ . These upper indexes are not to be confused with exponents.

Nowak [11] has obtained sufficient conditions for the existence of a Nash equilibrium (i.e. a saddle point) for two-player zero-sum stochastic games with general state spaces, and showed that in this particular case, the value of the game can be computed by dynamic programming. His model is in discrete time and assumes a constant discount factor  $\alpha_0$  between the successive stages. Whitt [16], using the monotone contraction operator framework of Denardo [4], has investigated noncooperative sequential games with a countable number of players. He gave sufficient conditions for the existence of a Nash equilibrium [ $\varepsilon$ -equilibrium] when the state set is countable [uncountable]. The models of Nowak and Whitt (and to our knowledge, all other infinite horizon discounted games models) are based on contraction assumptions that are uniform over all admissible policies. For the discounted case, Nowak assumes that

$$\lim_{N \rightarrow \infty} \left[ \sup_{K > N} \sup_{\delta \in \Delta} E_{\delta, s} \left[ \sum_{t=N}^K (\alpha_0)^{t-1} |r(s^t, i, a^t)| \right] \right] = 0 \quad (2)$$

for all  $s \in S$  and  $i = 1, 2$  (he also considers positive games, for which  $\alpha_0 = 1$  and all rewards are non negative). Whitt's  $N$ -stage contraction assumption is that for some integer  $N > 0$ , there exists two constants  $m \geq 0$  and  $0 \leq \alpha_0 < 1$  such that

$$E_{\delta, s}[\beta^1] \leq m \quad \text{and} \quad E_{\delta, s}[\beta^N] \leq \alpha_0 \quad (3)$$

for all  $\delta \in \Delta$  and  $s \in S$ .

In this paper, we investigate a model for which these uniform contraction assumptions are replaced by milder conditions that generalize those introduced by L'Ecuyer and Haurie [8] for Markov renewal programs. It allows for strategies under which the expected time duration between any two stages is not bounded away from zero. Some admissible strategies could even give rise, with probability one, to an unlimited number of successive stages without any movement of the clock (i.e. with the same value of the discount factor  $\beta^t$  at all these stages). The basic idea of our assumptions is to make sure that at least *one* of the players (without loss of generality, say player 1) has the ability to move the clock forward, and that the occurrence of too many stages of the game in a short period of time would be costly to this player.

For any  $i \in I$ ,  $\pi \in \Pi$  and  $\theta_i \in \Pi_i$ , we denote by  $[\pi^{-i}, \theta_i]$  the strategy vector  $\pi' \in \Pi$  such that  $\pi'_i = \theta_i$  and  $\pi'_j = \pi_j$  for  $j \neq i$ . Likewise, for  $\delta \in \Delta$  and  $\gamma_i \in \Delta_i$ ,  $[\delta^{-i}, \gamma_i]$  represents the policy (or stationary strategy) vector  $\delta' \in \Delta$  such that  $\delta'_i = \gamma_i$  and

$\delta'_j = \delta_j$  for  $j \neq i$ . We say that a strategy  $\theta_i = (\gamma_i^0, \gamma_i^1, \gamma_i^2, \dots) \in \Pi_i$  for player  $i$  is  $N$ -stage distinguished, for an integer  $N \geq 1$ , if there exists a constant  $\alpha_0 < 1$  such that

$$E_{[\pi^{-i}, \theta_i], s}[\beta^N] \leq \alpha_0 \quad \text{for all } s \in S, \pi \in \Pi. \quad (4)$$

The stochastic game model is called  $N$ -stage locally contracting if the following condition is satisfied:

CONDITION LC. There exists a  $N$ -stage distinguished strategy  $\tilde{\pi}_1$  for player 1, and two constants  $K_1$  and  $K_2$ , such that

$$K_1 + K_2 < 0 \quad (5)$$

and

$$E_{\pi, s} \left[ \sum_{t=0}^{n-1} \beta^t r(s^t, 1, a^t) \right] \leq K_1 + K_2 E_{\pi, s}[\beta^n] \quad (6)$$

for any  $s \in S, \pi \in \Pi$  and integer  $n$  such that  $N \leq n < 2N$ . ■

If  $K_1 \leq 0$ , then the return to player 1 is always negative. The return function can take positive and negative values if  $K_1 > 0$  and  $K_2 < -K_1 < 0$ . Condition LC ensures that it is possible for player 1 to follow a policy under which, whatever the other players do, the expected discount factor for the following  $N$  stages will be smaller than  $\alpha_0 < 1$ . Hence, player 1 can force the sequence of discount factors  $\{\beta^t, t \geq 0\}$  to converge to 0 with probability one. On the other hand, if the sequence of discount factors does not converge to 0, then the return to player 1 is  $-\infty$ . In terms of Markov renewal games, player 1 always has the possibility to move the clock forward, and has no interest in using a strategy under which the clock will never move. He may, however, use a strategy (possibly optimal for him) under which the  $N$ -stage expected discount factor is equal to one for some of the states.

Obviously, condition LC does not imply (2) or (3). On the other hand, Whitt's assumption (3) implies LC under the model studied here, since one always has  $E_{\pi, s}[\beta^1] \leq 1$ , and the cost function is bounded. More specifically, if  $N$  and  $\alpha_0$  satisfy (3) and  $K_3$  is an upper bound on  $|r(\cdot, 1, \cdot)|$ , then condition LC is satisfied for instance by taking  $K_1 = 2NK_3(1 + \alpha_0)/(1 - \alpha_0)$  and  $K_2 = -4NK_3/(1 - \alpha_0)$ . Also, since Nowak's model is in discrete time with a constant discount factor, all strategies are distinguished under his model when  $\alpha_0 < 1$ , and in that case, LC is satisfied. In general, however, (2) does not imply LC (for instance, if  $\alpha_0 = 1$  and  $r \equiv 0$ , then (2) is satisfied while LC is not).

### 3. VALUE FUNCTIONS AND SOME OF THEIR BASIC PROPERTIES

For every strategy vector  $\pi$  and integer  $n \geq 0$ , we define:  $v_\pi^n: S \times I \rightarrow \mathbb{R}$  as

$$v_\pi^n(s, i) = E_{\pi, s} \left[ \sum_{t=0}^{n-1} \beta^t r(s^t, i, a^t) \right] \quad (7)$$

which represents the  $n$ -stage expected discounted return to player  $i$  under strategy  $\pi$  from initial state  $s_0 = s$ , and (when the limit exists):

$$v_\pi(s, i) = \lim_{n \rightarrow \infty} v_\pi^n(s, i) \quad (8)$$

which represents the total expected discounted return to player  $i$  over the infinite horizon, from initial state  $s_0 = s$ .

An  $\varepsilon$ -equilibrium point ( $\varepsilon$ -EP) in stationary strategies in the sense of Nash [9], for  $\varepsilon \geq 0$ , is a stationary strategy  $\delta$  in  $\Delta$  such that no single player can improve his expected return by more than  $\varepsilon$  by changing unilaterally his strategy. When  $\varepsilon = 0$ , it is called an equilibrium point (EP). More specifically, let (when it exists):

$$f_\delta(s, i) = \sup_{\gamma_i \in \Delta_i} v_{[\delta^{-i}, \gamma_i]}(s, i) \quad (9)$$

for each  $s \in S$  and  $i \in I$ . It represents the optimal return function for player  $i$  when all other players use the stationary strategy  $\delta$ . Then,  $\delta$  is an  $\varepsilon$ -EP if and only if  $f_\delta(s, i) \leq v_\delta(s, i) + \varepsilon$  for all  $s \in S$  and  $i \in I$ .

In the remainder of this section, we show that under condition LC, the expression (8) is well defined for  $i=1$ , and we give an upper bound on this expression. Notice however that for some strategies, it is possible that  $v_\pi(s, 1) = -\infty$ .

LEMMA 1 *Under condition LC, for every strategy  $\pi \in \Pi$  and initial state  $s \in S$ ,  $v_\pi(s, 1)$  is well defined by (8). One also has*

$$v_\pi(s, 1) \leq \max(0, K_1), \quad (10)$$

and for all integers  $k \geq 1$  and  $n \geq kN$ ,

$$v_\pi^n(s, 1) \leq \max(0, K_1) + k(K_1 + K_2)E_{\pi, s}(\beta^n) \leq \max(0, K_1). \quad (11)$$

*Proof* Let  $\pi \in \Pi$ , and for any  $s \in S$  and  $n \geq 0$ , define the following conditional expectations, given the state  $s^n$  at stage  $n$ :

$$v_\pi^{n, N}(s^n, 1) = E_{\pi, s} \left[ \sum_{t=0}^{N-1} (\beta^{n+t}/\beta^n) r(s^{n+t}, 1, a^{n+t}) \mid s^n \right]$$

$$\alpha_\pi^{n, N}(s^n) = E_{\pi, s} [\beta^{n+N}/\beta^n \mid s^n]$$

Notice that these expectations do not depend on  $s$ . For any realization of the stochastic process and integer  $m \geq 0$ , we have from (6):

$$(v_\pi^{n, N}(s^{mN}, 1))^+ = \max(0, v_\pi^{n, N}(s^{mN}, 1)) \leq \max(0, K_1 + K_2 \alpha_\pi^{n, N}(s^{mN})).$$

Let  $k \geq 1$  and  $n = kN$ . Define the set of integers  $\Phi = \{m \mid 0 \leq m \leq k-1 \text{ and } K_1 +$

$K_2\alpha_\pi^{n,N}(s^{mM}) > 0$ ). Let  $v$  be the cardinality of  $\Phi$ , let  $\xi(1), \dots, \xi(v)$  be the elements of  $\Phi$  ranked by increasing order, and define  $\xi(v+1) = k$ . If  $\Phi$  is not empty, then by (5-6), we have  $K_1 > 0$  and  $K_2 < 0$ , and

$$\begin{aligned}
E_{\pi,s} \left[ \sum_{m=0}^{k-1} \beta^{mN} (v_\pi^{n,N}(s^{mN}, 1))^+ \right] &\leq E_{\pi,s} \left[ \sum_{j=1}^v \beta^{\xi(j)N} (K_1 + K_2 \alpha_\pi^{n,N}(s^{\xi(j)N})) \right] \\
&\leq E_{\pi,s} \left[ \sum_{j=1}^v (K_1 \beta^{\xi(j)N} + K_2 \beta^{(\xi(j)+1)N}) \right] \\
&\leq E_{\pi,s} \left[ \sum_{j=1}^v (K_1 \beta^{\xi(j)N} + K_2 \beta^{\xi(j+1)N}) \right] \\
&\leq E_{\pi,s} \left[ K_1 + K_2 \beta^n + (K_1 + K_2) \sum_{j=2}^v \beta^{\xi(j)N} \right] \\
&\leq K_1.
\end{aligned} \tag{12}$$

If  $K_1 \leq 0$ , then  $\Phi$  is empty and each  $(v_\pi^{n,N}(s^{mN}, 1))^+$  is equal to zero. In either case,

$$\sum_{m=0}^{\infty} \beta^{mN} (v_\pi^{n,N}(s^{mN}, 1))^+$$

converges  $P_{\pi,s}$ -almost surely, and from the monotone convergence theorem,

$$\lim_{k \rightarrow \infty} E_{\pi,s} \left[ \sum_{m=0}^{k-1} \beta^{mN} v_\pi^{n,N}(s^{mN}, 1) \right]$$

is well defined.

Now, let  $k \geq 1$  and  $n = (k-1)N + n'$ , where  $N \leq n' < 2N$ . As usual, we define  $\sum_{m=0}^{-1}$  to mean the empty sum, which is always equal to zero. We have

$$\begin{aligned}
v_\pi^n(s, 1) &= E_{\pi,s} \left[ \sum_{t=0}^{n-1} \beta^t r(s^t, 1, a^t) \right] \\
&= E_{\pi,s} \left[ \sum_{m=0}^{k-2} \sum_{t=mN}^{(m+1)N-1} \beta^t r(s^t, 1, a^t) + \sum_{t=(k-1)N}^{n-1} \beta^t r(s^t, 1, a^t) \right] \\
&\leq E_{\pi,s} \left[ \sum_{m=0}^{k-2} \beta^{mN} (v_\pi^{n,N}(s^{mN}, 1)) + \sum_{t=(k-1)N}^{n-1} \beta^t r(s^t, 1, a^t) \right]
\end{aligned}$$


---

$$\begin{aligned} &\leq E_{\pi,s} \left[ K_1(\beta^0 - \beta^n) + (K_1 + K_2)\beta^n + (K_1 + K_2) \sum_{m=1}^{k-1} \beta^{mN} \right] \\ &\leq \max(0, K_1) + k(K_1 + K_2)E_{\pi,s}(\beta^n) \\ &\leq \max(0, K_1), \end{aligned} \tag{13}$$

where the second inequality follows from (6) and the last two inequalities follow from (5). This yields (11).

If  $\lim_{t \rightarrow \infty} E[\beta^t] \neq 0$ , then from (11) and since  $K_1 + K_2 < 0$ ,  $\lim_{n \rightarrow \infty} v_\pi^n(s, 1) = -\infty$ . If  $\lim_{t \rightarrow \infty} E[\beta^t] = 0$ , then the second sum in (13) goes to zero as  $n \rightarrow \infty$ , and since the limit of the expectation of the first sum is well defined, (8) is also well defined. ■

From now on, we assume that condition LC holds.

#### 4. SADDLE POINTS IN TWO-PLAYER ZERO-SUM GAMES

Two-player zero-sum stochastic games, also called *duels*, are a special case of the general model, for which  $|I|=2$  and  $r(\cdot, 1, \cdot) = -r(\cdot, 2, \cdot)$ . At each stage of the game, the returns to player 1 are paid by player 2. Hence, player 1 wishes to *maximize* his total expected returns, while player 2 wishes to *minimize* the total expected returns to player 1. We define the real-valued functions  $\underline{u}$  and  $\bar{u}$ ,  $\underline{u} \leq \bar{u}$ , by

$$u(s) = \sup_{\pi_1 \in \Pi_1} \left( \inf_{\pi_2 \in \Pi_2} v_\pi(s, 1) \right) \quad \text{and} \quad \bar{u}(s) = \inf_{\pi_2 \in \Pi_2} \left( \sup_{\pi_1 \in \Pi_1} v_\pi(s, 1) \right). \tag{14}$$

The game is said to have a *value* if  $\underline{u}(\cdot) = \bar{u}(\cdot)$ , and we define in that case its *value function*  $u_*$  by

$$u_*(s) = u(s) = \bar{u}(s).$$

In duels, an equilibrium point (EP) is also called a *saddle point*. If there exists a saddle point  $\pi \in \Pi$ , then for all  $s \in S$ ,  $v_\pi(s, 1) = u_*(s)$ , the value of the game. When  $\pi$  is also stationary, we say that the game has a saddle point in stationary strategies. In this section, we prove the existence of a saddle point in stationary strategies for a duel, under condition LC and some additional continuity assumptions (condition C below). We show that the value function  $u_*$  is the unique fixed point of a contracting operator and that the saddle point can be computed using dynamic programming.

Let  $U$  be the space of all *return functions*  $u: S \rightarrow [-\infty, \infty]$ , endowed with the supremum norm:

$$\|u\| = \sup_{s \in S} |u(s)|. \tag{15}$$



Let  $U_M$  be the subspace of all Borel-measurable functions in the Banach space of all bounded functions in  $U$ . For the main results of this section, we will need the following continuity condition, which corresponds to assumption  $(M_3)$  in [11].

CONDITION C. For each  $s \in S$ ,  $a_1 \in A_1(s)$ ,  $a_2 \in A_2(s)$  and  $\bar{S}$  Borel subset of  $S$ ,  $r(s, 1, (\cdot, a_2))$  is upper semi-continuous on  $A_1(s)$ ,  $r(s, 1, (a_1, \cdot))$  is lower semi-continuous on  $A_2(s)$ ,  $q(\bar{S} | s, (\cdot, a_2))$  is continuous on  $A_1(s)$  and  $q(\bar{S} | s, (a_1, \cdot))$  is continuous on  $A_2(s)$ . ■

We define the *local return function*  $h$  by

$$h(s, a, u) = r(s, 1, a) + \frac{1}{\beta(s)} \int_S \beta(s') u(s') q(ds' | s, a) \quad (16)$$

for  $s \in S$ ,  $a \in A(s)$  and  $u \in U_M$ . It represents the expected return to player 1 for a fictive auxiliary one-stage game starting in state  $s$ , if the players use the action vector  $a$  and if the expected returns to player 1 from the next stage on are described by the function  $u$ . For every policy  $\delta \in \Delta$ , the associated *return operator*  $H_\delta: U_M \rightarrow U_M$  is defined by:

$$H_\delta u(s) = h(s, \delta(s), u). \quad (17)$$

We define the operator  $F: U_M \rightarrow U_M$  by

$$Fu(s) = \sup_{a_1 \in A_1(s)} \left( \inf_{a_2 \in A_2(s)} h(s, (a_1, a_2), u) \right). \quad (18)$$

Recall that the *modulus* of an operator  $\phi$  mapping a subset of a Banach space into itself is the smallest number  $\alpha$  such that  $\|\phi(u_2) - \phi(u_1)\| \leq \alpha \|u_2 - u_1\|$  for all  $u_1$  and  $u_2$  in that subset. If  $\alpha < 1$ ,  $\phi$  is said to be *contracting* with modulus  $\alpha$ . The operators  $H_\delta$  and  $F$  are monotone, i.e. if  $u_1 \leq u_2$ , then  $H_\delta u_1 \leq H_\delta u_2$  and  $Fu_1 \leq Fu_2$ . However these operators are not necessarily contracting on  $U_M$  or on one of its proper subset. If  $\bar{\delta} = (\bar{\delta}_1, \bar{\delta}_2)$  is such that  $\bar{\delta}_1$  is a distinguished policy for player 1, then  $H_{\bar{\delta}}$  is contracting. In what follows, we show that there exists a closed subset of  $U_M$  and an integer  $n_0 > 0$  such that for all  $n \geq n_0$ ,  $F^n$ , the  $n$ -fold composition of  $F$ , is contracting on that subset.

Let  $K_3$  be such that  $|r(\cdot, 1, \cdot)| \leq K_3$  and define

$$U_C = \left\{ u \in U_M \mid \frac{-(2N-1)K_3}{1-\alpha_0} \leq u \leq \max(0, K_1) \right\} \quad (19)$$

which is a closed subset of  $U_M$ .

LEMMA 2 For each  $n \geq N$ ,  $U_C$  is closed under  $F^n$ .

*Proof* Each  $n \geq N$  can be written as  $n = kN + n'$  where  $k \geq 0$  and  $N \leq n' < 2N$  are integers, and  $F^n$  can be viewed as the composition of  $F^{n'}$  with  $k$  times  $F^N$ . Hence, it suffices to prove the result for  $N \leq n < 2N$ . Let  $u \in U_C$ ,  $s \in S$  and  $N \leq n < 2N$ . From

Theorem 4.1 in [11] (for the  $n$ -stage model, the discount factor need not be bounded away from one), there exists a strategy  $\pi = (\delta^0, \delta^1, \dots, \delta^n, \dots)$  such that

$$F^n u(s) = (H_{\delta^0} H_{\delta^1}, \dots, H_{\delta^{n-1}}) u(s) = E_{\pi, s} \left[ \sum_{t=0}^{n-1} \beta^t r(s^t, 1, a^t) + \beta^n u(s^n) \right].$$

Therefore, from (6),

$$F^n u(s) \leq K_1 + K_2 E_{\pi, s}[\beta^n] + \max(0, K_1) E_{\pi, s}[\beta^n] \leq \max(0, K_1).$$

Conversely, let  $\tilde{\pi}_1$  be an  $N$ -stage distinguished strategy for player 1. Then,

$$\begin{aligned} F^n u(s) &\geq E_{(\tilde{\pi}_1, \pi_2), s} \left[ \sum_{t=0}^{n-1} \beta^t r(s^t, 1, a^t) + \beta^n u(s^n) \right] \\ &\geq -(2N-1)K_3 - \frac{(2N-1)K_3}{1-\alpha_0} E_{(\tilde{\pi}_1, \pi_2), s}[\beta_n] \geq \frac{-(2N-1)K_3}{1-\alpha_0} \end{aligned}$$

because  $E_{(\tilde{\pi}_1, \pi_2), s}[\beta^n] \leq E_{(\tilde{\pi}_1, \pi_2), s}[\beta^N] \leq \alpha_0$ . ■

LEMMA 3 Let  $\alpha_1 \in (0, 1)$  and

$$\eta_1 = \frac{(2N-1)K_3/(1-\alpha_0) + 2\max(0, K_1)}{-(K_1 + K_2)\alpha_1}. \quad (20)$$

Then, under condition C, for any integer  $k \geq \eta_1$  and  $n \geq kN$ ,  $F^n$  is contracting on  $U_C$  with modulus  $\leq \alpha_1$ . Furthermore,  $F$  has modulus  $\leq 1$ .

*Proof* For any positive integer  $n$ ,  $u_1 \in U_C$  and  $u_2 \in U_C$ , from Theorem 4.1 in [11], there exists strategies  $\pi = (\delta^0, \delta^1, \dots, \delta^n, \dots)$  and  $\theta = (\gamma^0, \gamma^1, \dots, \gamma^n, \dots)$  in  $\Pi$  such that

$$F^n u_1(s) = (H_{\delta^0} H_{\delta^1}, \dots, H_{\delta^{n-1}}) u_1(s) = v_{\pi}^n(s, 1) + E_{\pi, s}(\beta^n u_1(s^n))$$

and

$$F^n u_2(s) = (H_{\gamma^0} H_{\gamma^1}, \dots, H_{\gamma^{n-1}}) u_2(s) = v_{\theta}^n(s, 1) + E_{\theta, s}(\beta^n u_2(s^n)).$$

Since the strategies  $\pi$  and  $\theta$  are  $n$ -stage equilibria when the terminal return functions are respectively  $u_1$  and  $u_2$ , we obtain

$$\begin{aligned} &F^n u_2(s) - F^n u_1(s) \\ &\leq v_{(\theta_1, \pi_2)}^n(s, 1) + E_{(\theta_1, \pi_2), s}(\beta^n u_2(s^n)) - v_{(\theta_1, \pi_2)}^n(s, 1) - E_{(\theta_1, \pi_2), s}(\beta^n u_1(s^n)) \\ &\leq \|u_2 - u_1\| E_{(\theta_1, \pi_2), s}(\beta^n). \end{aligned} \quad (21)$$

In particular, for  $n=1$ , we obtain

$$Fu_2(s) - Fu_1(s) \leq \|u_2 - u_1\|.$$

Let  $k \geq \eta_1$  and  $n \geq kN$ . From (11) in Lemma 1, and since  $K_1 + K_2 < 0$ , we have:

$$E_{(\theta_1, \pi_2), s}(\beta^n) \leq \frac{\max(0, K_1) - v_{(\theta_1, \pi_2)}^n(s, 1)}{-k(K_1 + K_2)}. \quad (22)$$

On the other hand, using Lemma 2,

$$\frac{-(2N-1)K_3}{1-\alpha_0} \leq F^n u_2(s) \leq v_{(\theta_1, \pi_2)}^n(s, 1) + E_{(\theta_1, \pi_2), s}(\beta^n u_2(s^n))$$

so that, since  $u_2 \in U_C$ ,

$$-v_{(\theta_1, \pi_2)}^n(s, 1) \leq \frac{(2N-1)K_3}{1-\alpha_0} + \max(0, K_1). \quad (23)$$

From (21–23), one obtains:

$$\begin{aligned} F^n u_2(s) - F^n u_1(s) &\leq \|u_2 - u_1\| \frac{(2N-1)K_3/(1-\alpha_0) + 2\max(0, K_1)}{-k(K_1 + K_2)} \\ &\leq \alpha_1 \|u_2 - u_1\|. \end{aligned}$$

Since  $u_1$  and  $u_2$  can be interchanged, the result follows.  $\blacksquare$

We have just shown that  $F$  has modulus not larger than one, and that for  $n$  large enough,  $F^n$  is a contraction mapping on  $U_C$ , which is a closed subset of  $U_M$ . From the fixed point theorem on  $n$ -stage contraction mappings [4], there is a unique fixed point  $\tilde{u}_*$  in  $U_C$  such that  $F\tilde{u}_* = \tilde{u}_*$  and, for all  $u \in U_C$ ,

$$\lim_{n \rightarrow \infty} \|F^n u - \tilde{u}_*\| = 0.$$

In the next proposition, we show that the game has a value (over the infinite horizon), and that its value function  $u_*$  is equal to the fixed point  $\tilde{u}_*$  of the operator  $F$ .

**PROPOSITION 4** *Under condition C:*

- The infinite horizon game has a value  $u_* \in U_C$ ;*
- For  $u \in U_C$ ,  $Fu = u$  if and only if  $u = u_*$ ;*
- For each  $u \in U_C$ ,  $\lim_{n \rightarrow \infty} \|F^n u - u_*\| = 0$ .*

*Proof* It suffices to prove that  $\tilde{u}_* = \underline{u} = \bar{u} = u_*$ . Let  $\alpha_1, \eta_1, k$  and  $n$  satisfy the

conditions of Lemma 3. Let  $\theta = (\gamma^0, \gamma^1, \dots, \gamma^n, \dots) \in \Pi$  be an  $n$ -stage equilibrium when the terminal return function is zero:

$$(H_{\gamma^0} H_{\gamma^1}, \dots, H_{\gamma^{n-1}})0(s) = u_*^n(s) \stackrel{\text{def}}{=} F^n 0(s), \quad (24)$$

and such that for all  $t \geq n$ ,  $\gamma^t = \tilde{\gamma}_1^t$  where  $\tilde{\theta}_1 = (\tilde{\gamma}_1^0, \tilde{\gamma}_1^1, \tilde{\gamma}_1^2, \dots)$  is an  $N$ -stage distinguished strategy for player 1. For any  $\pi_2 \in \Pi_2$ , from Lemma 2 and since  $0 \in U_C$ , one has

$$\frac{-(2N-1)K_3}{1-\alpha_0} \leq F^n 0(s) \leq v_{(\theta_1, \pi_2)}^n(s, 1)$$

and therefore, using Lemmas 1 and 2 as in (22–23),

$$E_{(\theta_1, \pi_2), s}(\beta^n) \leq \frac{\max(0, K_1) - v_{(\theta_1, \pi_2)}^n(s, 1)}{-k(K_1 + K_2)} \leq \frac{(2N-1)K_3/(1-\alpha_0) + \max(0, K_1)}{-k(K_1 + K_2)} \leq \alpha_1.$$

For  $t > n$ ,  $E_{(\theta_1, \pi_2), s}(\beta^t) \leq \alpha_1 \alpha_0^{(t-n)/N!}$ . Hence,

$$\begin{aligned} u(s) &\geq \inf_{\pi_2 \in \Pi_2} \left[ v_{(\theta_1, \pi_2)}^n(s, 1) + E_{(\theta_1, \pi_2), s} \left( \sum_{t=n}^{\infty} \beta^t r(s^t, 1, a^t) \right) \right] \\ &\geq \inf_{\pi_2 \in \Pi_2} [v_{(\theta_1, \pi_2)}^n(s, 1)] + \inf_{\pi_2 \in \Pi_2} \left[ E_{(\theta_1, \pi_2), s} \left( \sum_{t=n}^{\infty} \beta^t r(s^t, 1, a^t) \right) \right] \\ &\geq u_*^n(s) - \frac{(2N-1)K_3 \alpha_1}{1-\alpha_0}. \end{aligned} \quad (25)$$

On the other hand, let

$$\tilde{\Pi}_1 = \left\{ \pi_1 \in \Pi_1 \mid v_{(\pi_1, \theta_2)}(s, 1) \geq \frac{-4NK_3}{1-\alpha_0} \text{ for all } s \in S \right\}.$$

The set  $\tilde{\Pi}_1$  is not empty (it contains  $\theta_1$ ) and for any  $\pi_1 \in \tilde{\Pi}_1$ ,

$$v_{(\pi_1, \theta_2)}^n(s, 1) \geq -\frac{4NK_3}{1-\alpha_0} - \max(0, K_1).$$

From Lemma 1, this implies that for all  $s \in S$ ,

$$E_{(\pi_1, \theta_2), s}(\beta^n) \leq \frac{\max(0, K_1) - v_{(\pi_1, \theta_2)}^n(s, 1)}{-k(K_1 + K_2)} \leq 2\alpha_1.$$

We then obtain

$$\begin{aligned}
\bar{u}(s) &\leq \sup_{\pi_1 \in \bar{\Pi}_1} v_{(\pi_1, \theta_2)}(s, 1) \\
&\leq u_*^n(s) + \max(0, K_1) \sup_{\pi_1 \in \bar{\pi}_1} E_{(\pi_1, \theta_2), s}(\beta^n) \\
&\leq u_*^n(s) + 2\alpha_1 \max(0, K_1).
\end{aligned} \tag{26}$$

If we let  $\alpha_1 \rightarrow 0$ , then  $n \rightarrow \infty$ , and from (25–26),

$$\underline{\tilde{u}}_*(s) = \lim_{n \rightarrow \infty} u_*^n(s) \leq u(s) \leq \bar{u}(s) \leq \lim_{n \rightarrow \infty} u_*^n(s) = \underline{\tilde{u}}_*(s).$$

This completes the proof.  $\blacksquare$

**PROPOSITION 5** *Under condition C, there exists  $\delta_* \in \Delta$  such that  $Fu_*(s) = H_{\delta_*}u_*(s)$  for all  $s \in S$ , and each such  $\delta_*$  is a saddle point for the infinite horizon game.*

*Proof* From Theorem 4.1 in [11], there exists a policy vector  $\delta_* \in \Delta$  such that for all  $s$  in  $S$ ,

$$H_{\delta_*}u_*(s) = Fu_*(s) = u_*(s).$$

Thus, for any  $s$  in  $S$ ,

$$\begin{aligned}
u_*(s) &= \lim_{n \rightarrow \infty} H_{\delta_*}^n u_*(s) \\
&= \lim_{n \rightarrow \infty} (v_{\delta_*}^n(s, 1) + E_{\delta_*, s}(\beta^n u_*(s^n))) \\
&= v_{\delta_*}(s, 1) + \lim_{n \rightarrow \infty} E_{\delta_*, s}(\beta^n u_*(s^n)).
\end{aligned} \tag{27}$$

Since  $u_*$  is bounded,  $v_{\delta_*}$  is also bounded, and Eq. (11) in Lemma 1 implies that the limit in the r.h.s. of (27) vanishes as  $n \rightarrow \infty$ . We then obtain:

$$v_{\delta_*}(s, 1) = u_*(s)$$

so that  $\delta_*$  defines a saddle point in stationary strategies for the infinite horizon game.  $\blacksquare$

## 5. NON ZERO-SUM GAMES WITH MANY PLAYERS

In this section, we prove the existence of a Nash equilibrium in stationary strategies for a non-zero sum stochastic game with a finite number of players, under condition LC and some additional assumptions (condition D below). Let  $U$ ,  $U_M$ ,  $U_C$ ,  $h$ ,  $H_\delta$  and  $K_3$  be defined as in Section 4. The key to our analysis will be to fix the policy for all the players, except for one of them, say player 1. Player 1

---

thus faces a locally contracting discrete event dynamic programming problem similar to the one analyzed by L'Ecuyer and Haurie [8].

For a fixed policy  $\delta \in \Delta$ , we define the operator  $G_\delta: U_M \rightarrow U_M$  by:

$$G_\delta u(s) = \sup_{a_1 \in A_1(s)} h(s, [\delta(s)^{-1}, a_1], u) \quad (28)$$

where  $[\delta(s)^{-1}, a_1]$  denotes the action vector  $a'$  such that  $a'_1 = a_1$  and  $a'_j = \delta(s)_j$  for  $j \neq 1$ .  $G_\delta u(s)$  represents the expected return to player 1 for a fictive auxiliary one-stage game with initial state  $s$ , if player 1 optimizes his action choice and all players use policy  $\delta$ . We will show that on some closed subset of  $U$ , the  $n$ -fold composition  $G_\delta^n$  is contracting for  $n$  large enough, and that player 1 can restrict himself, without any loss of generality, to a class of policies  $\gamma_1$  under which  $H_{[\delta^{-1}, \gamma_1]}^n$  is contracting for  $n$  large enough. We then show that there exists a Nash equilibrium in stationary strategies for the infinite horizon game. For the rest of this section we will need the following condition D, which corresponds to the assumptions of Theorem 6.2 in [16]. We will make use of this theorem, whose proof is itself based on the Kakutani–Glicksberg–Fan fixed point theorem for point to set mappings.

**CONDITION D.** The state space  $S$  is countable, and for each  $s, s'$  in  $S$  and  $i$  in  $I$ ,  $r(s, i, \cdot)$  and  $q(s' | s, \cdot)$  are continuous on  $A(s)$ .

**LEMMA 6** For every policy  $\delta \in \Delta$  and  $n \geq N$ ,  $U_C$  is closed under the operator  $G_\delta^n$ .

*Proof* Similar to the proof of Lemma 2.  $\blacksquare$

**LEMMA 7** Under condition D, let  $\alpha_1$  and  $\eta_1$  be as in Lemma 3:  $0 < \alpha_1 < 1$  and

$$\eta_1 = \frac{(2N-1)K_3/(1-\alpha_0) + 2 \max(0, K_1)}{-(K_1 + K_2)\alpha_1} \quad (29)$$

then, for any integer  $k \geq \eta_1$ ,  $n \geq kN$  and  $\delta \in \Delta$ ,  $G_\delta^n$  is contracting on  $U_C$  with modulus  $\leq \alpha_1$ .

*Proof* Let  $k \geq \eta_1$ ,  $n \geq kN$ ,  $u_1 \in U_C$ ,  $u_2 \in U_C$  and  $\delta \in \Delta$ . Under condition D, there exists for player 1 a strategy  $\theta_1 = (\gamma_1^0, \gamma_1^1, \dots, \gamma_1^n, \dots) \in \Pi_1$  such that

$$H_{[\delta^{-1}, \gamma_1^0]} H_{[\delta^{-1}, \gamma_1^1]} \dots H_{[\delta^{-1}, \gamma_1^{n-1}]} u_1(s) = G_\delta^n u_1(s). \quad (30)$$

Let  $\pi = [\delta^{-1}, \theta_1]$ . Then, for all  $s \in S$ ,

$$G_\delta^n u_1(s) = v_\pi^n(s, 1) + E_{\pi, s}(\beta^n u_1(s^n))$$

$$G_\delta^n u_2(s) \geq v_\pi^n(s, 1) + E_{\pi, s}(\beta^n u_2(s^n)),$$

so that

$$G_\delta^n u_1(s) - G_\delta^n u_2(s) \leq \|u_1 - u_2\| E_{\pi, s}(\beta^n).$$

It remains to prove that  $E_{\pi,s}(\beta^n) \leq \alpha_1$  and, since  $u_1$  and  $u_2$  can be interchanged, the result will follow.

For all  $s$  in  $S$ , from Lemma 6 and since  $u_1 \in U_C$ , we have

$$\begin{aligned} \frac{-(2N-1)K_3}{1-\alpha_0} &\leq G_\delta^n u_1(s) = H_{[\delta^{-1}, \gamma_1^1]} H_{[\delta^{-1}, \gamma_1^1]} \cdots H_{[\delta^{-1}, \gamma_1^{n-1}]} u_1(s) \\ &= E_{\pi,s} \left( \sum_{t=0}^{n-1} \beta^t r(s^t, 1, a^t) + \beta^n u_1(s^n) \right) = v_\pi^n(s, 1) + E_{\pi,s}(\beta^n u_1(s^n)) \\ &\leq 2 \max(0, K_1) + k(K_1 + K_2) E_{\pi,s}(\beta^n) \end{aligned}$$

where the last inequality follows from (11). Using (29) and since  $k \geq \eta_1$ , we obtain that  $E_{\pi,s}(\beta^n) \leq \alpha_1$ . ■

We now select  $\alpha_1 < 1$  and an integer  $K \geq \eta_1$ , where  $\eta_1$  satisfies (29). Let  $M = KN$  and define the set of policies

$$\tilde{\Delta} = \{ \delta \in \Delta \mid E_{\delta,s}(\beta^M) \leq \alpha_1 \text{ for all } s \in S \}.$$

We will show that  $\tilde{\Delta}$  is a nonempty subset of  $\Delta$  and that when searching for an EP, it is possible to restrain our attention to the set  $\tilde{\Delta}$ . We will then be able to exploit the  $M$ -stage contraction property of the operator  $G_\delta$  on  $U_C$  when  $\delta \in \tilde{\Delta}$ .

LEMMA 8 Under condition D,

- $\tilde{\Delta}$  is nonempty;
- If there exists a stationary equilibrium strategy vector  $\delta \in \Delta$ , then there is one such vector in  $\tilde{\Delta}$ .

*Proof* For any  $\delta \in \Delta$ , since  $G_\delta^n$  is a contraction mapping on  $U_C$  for  $n \geq M$ , there is a unique fixed point  $u_\delta$  in  $U_C$  such that  $G_\delta u_\delta = u_\delta$ , and for all  $u \in U_C$ ,

$$\lim_{n \rightarrow \infty} \|G_\delta^n u - u_\delta\| = 0.$$

Under condition D, there exists a policy  $\tilde{\gamma}_1 \in \Delta_1$  for player 1 such that for all  $s \in S$ ,

$$H_{[\delta^{-1}, \tilde{\gamma}_1]} u_\delta(s) = G_\delta u_\delta(s) = u_\delta(s).$$

Let  $\tilde{\delta} = [\delta^{-1}, \tilde{\gamma}_1] \in \Delta$ . As in the proof of Lemma 7, we have

$$\begin{aligned} \frac{-(2N-1)K_3}{1-\alpha_0} &\leq G_\delta^M u_\delta(s) = H_{\tilde{\delta}}^M u_\delta(s) = E_{\tilde{\delta},s} \left( \sum_{t=0}^{M-1} \beta^t r(s^t, 1, a^t) + \beta^M u_\delta(s^M) \right) \\ &\leq 2 \max(0, K_1) + K(K_1 + K_2) E_{\tilde{\delta},s}(\beta^M). \end{aligned}$$

From (29) and since  $K \geq \eta_1$ ,  $E_{\delta, s}(\beta^M) \leq \alpha_1$ .

Now it remains to prove (b). If  $\delta \in \Delta$  is an EP, then

$$u_\delta(s) = v_\delta(s, 1) = \sup_{\gamma_1 \in \Delta_1} v_{[\delta^{-1}, \gamma_1]}(s, 1) = f_\delta(s, 1).$$

The policy  $\tilde{\delta}$  defined in the proof of (a) above satisfies

$$\begin{aligned} v_{\tilde{\delta}}(s, 1) &= \lim_{n \rightarrow \infty} \left( H_{\tilde{\delta}}^n u_\delta(s) + E_{\tilde{\delta}, s} \left( \sum_{t=n}^{\infty} \beta^t r(s^t, 1, a^t) \right) \right) \\ &= \lim_{n \rightarrow \infty} (G_{\tilde{\delta}}^n u_\delta(s)) = G_{\tilde{\delta}} u_\delta(s) = u_\delta(s) = v_\delta(s, 1). \end{aligned}$$

Hence,  $\tilde{\delta}$  is also an EP.  $\blacksquare$

**PROPOSITION 9** *Under condition D, there exists  $\delta \in \tilde{\Delta}$  such that for all  $s \in S$  and  $i \in I$ ,*

$$v_\delta(s, i) = \sup_{\{\gamma_i \mid [\delta^{-1}, \gamma_i] \in \Delta\}} (v_{[\delta^{-1}, \gamma_i]}(s, i)),$$

i.e. there exists a Nash equilibrium in stationary strategies for the infinite horizon game.

*Proof* For each  $\delta \in \Delta$  and  $i \in I$ , define the local return operator for player  $i$ ,  $H_{\delta, i}: U_M \rightarrow U_M$ , by

$$H_{\delta, i} u(s) = r(s, i, \delta(s)) + \frac{1}{\beta(s)} \int \beta(s') u(s') q(ds' \mid s, \delta(s))$$

and define the set

$$U_D = \left\{ u \in U_M \mid \|u\| \leq \frac{2MK_4}{1-\alpha_1} \right\}$$

where  $K_4$  is a bound on the absolute value of  $r$ :  $|r(s, i, a)| \leq K_4$  for all  $s \in S$ ,  $i \in I$  and  $a \in A(s)$ . Using similar arguments as in the proofs of Lemmas 2 and 3, it is straightforward to show that for any  $\delta \in \tilde{\Delta}$  and  $n \geq M$ ,  $U_D$  is closed under the operator  $H_{\delta, i}^n$ , that  $H_{\delta, i}^n$  is a contraction mapping on  $U_D$  of modulus  $\leq \alpha_1$ , and that for each  $i \in I$ , the function  $v_\delta(s, i)$  is well defined by (8), with

$$v_\delta(s, i) = \lim_{n \rightarrow \infty} H_{\delta, i}^n u(s)$$

for any  $u \in U_D$ .

By restricting the set of admissible strategies to  $\tilde{\Delta}$ , we obtain the  $M$ -contracting game of Whitt [16]. To prove the existence of a Nash equilibrium using a



Kakutani–Glicksberg–Fan fixed point argument, it suffices to show that  $\tilde{\Delta}$  is a convex and compact subset of a Hausdorff topological space, and the result will follow from Theorem 6.2 in [16]. For each  $s \in S$ , define the function  $e_s: \Delta \rightarrow (0, 1]$  by

$$e_s(\delta) = E_{\delta, s}(\beta(s^M)).$$

That function is convex on  $\Delta$ . Therefore, the set

$$\tilde{\Delta}_s = \{\delta \in \Delta \mid e_s(\delta) \leq \alpha_1\}$$

is a closed and convex subset of  $\Delta$ . Since  $S$  is countable,

$$\tilde{\Delta} = \bigcap_{s \in S} \tilde{\Delta}_s$$

is also a closed and convex subset of  $\Delta$ . Under the assumptions of Section 2,  $\Delta$  is a convex and compact subset of a Hausdorff topological vector space, so  $\tilde{\Delta}$  is also compact. ■

For the case of uncountable state spaces, the existence of  $\varepsilon$ -equilibria can be obtained in the same way as in Whitt [16], by constructing a sequence of approximating models, each of which satisfying condition D.

## 6. AN EXAMPLE

We now give a simple illustration of the locally contracting model. More complicated examples, with larger state spaces, can also be considered.

*Example* Let  $I = X = \{1, 2\}$  and  $S = X \times [0, \infty)$ . When the system is in state  $(x, \tau) \in S$ , player  $x$  is said to be in *control* of the game. He incurs a cost  $C > 0$ , and selects a value  $\xi \in [0, 1]$  which corresponds to the time until the next transition. He also incurs a cumulative cost between the transitions, at rate  $c > 0$ . All costs are discounted at rate  $\rho > 0$ , and the players try to maximize their total expected discounted returns. Hence,  $\beta(x, \tau) = e^{-\rho\tau}$  for  $\tau \geq 0$ . The one-stage return function is given by

$$r(x, i, \xi) = \begin{cases} -C - \int_0^\xi c e^{-\rho\zeta} d\zeta & \text{if } i = x; \\ 0 & \text{otherwise.} \end{cases}$$

Given  $(x, \tau)$ , the next state is  $s' = (x', \tau + \xi)$ , where  $x'$  is determined randomly according to the conditional probability  $p(x' | x) = 1/2$  for all  $x, x' \in X$ . The sets of (pure) admissible actions can be defined by:

$$B_i(x, \tau) = \begin{cases} [0, 1] & \text{if } i = x; \\ \{0\} & \text{otherwise,} \end{cases}$$

where when  $i \neq x$ , 0 is a dummy action. The player in control can always select the action zero, in which case the stage has zero duration. If both players always do so, all stages will have zero duration, and the total discounted costs will be infinite to both players.

Let  $\tilde{\gamma}_1$  be the following policy for player one:

$$\tilde{\gamma}_1(x, \tau) = \begin{cases} 1 & \text{if } x = 1; \\ 0 & \text{otherwise.} \end{cases}$$

Under that policy, a stage can still have zero duration, e.g. when player 2 is in control and selects  $\xi = 0$ . However, there is always a probability 1/2 that the next state will be  $s' = (1, \tau)$ , in which case player 1 will then select  $\xi = 1$ . Hence, the expected time for the next two stages is at least 1/2, the expected two-stage discount factor is at most  $(1 + e^{-\rho})/2$ , and the stationary strategy  $\tilde{\gamma}_1$  is 2-stage distinguished for player one. Also, for any strategy and initial state, the expected two-stage reward to player one is bounded above as follows:

$$E_{\pi, s}[r(s^0, 1, a^0) + \beta^1 r(s^1, 1, a^1)] \leq 0 + e^{-\rho}((1/2)(-C) + (1/2)(0)) \leq -C e^{-\rho}/2.$$

Therefore, Eqs (5–6) are satisfied with  $K_1 = -C e^{-\rho}/2$  and  $K_2 = 0$ , and condition LC is satisfied with  $N = 2$ .

We note here that in this example, only the discounting function depends on  $\tau$ , and that the ratio  $\beta(s')/\beta(s) = e^{-\rho\xi}$  does not depend on  $\tau$ . Therefore, it is not really necessary to keep  $\tau$  in the state description, and for practical purposes, we can view  $X$  as the state space.

Let  $\tilde{\gamma}_2(x, \tau) = 0$  for all  $x \in X$  and  $\tau \geq 0$ . If player 2 follows  $\tilde{\gamma}_2$ , then player 1 will always get back control at the same instant as he leaves it, because time can then advance only when he is in control. Player 1 will incur a cost  $C$  every time he gets control, and will always absorb the cumulative cost at rate  $c$ . Therefore, to minimize his costs, he should always keep control as long as possible, i.e. adopt strategy  $\tilde{\gamma}_1$ . On the other hand, if player 1 follows  $\tilde{\gamma}_1$ , suppose player 2 always selects  $\xi > 0$  when he is in control, and let  $v$  be this total expected discounted reward under such a strategy. Then  $v$  obeys:

$$v = -C - \int_0^{\xi} c e^{-\rho\xi} d\xi + e^{-\rho\xi}(v + v e^{-\rho})/2 = -C - \frac{c}{\rho}(1 - e^{-\rho\xi}) + v e^{-\rho\xi}(1 + e^{-\rho})/2.$$

We can isolate  $v$  and derive with respect to  $\xi$  to obtain:

$$\frac{dv}{d\xi} = \frac{e^{-\rho\xi}(-c + (\rho C + c)(1 + e^{-\rho})/2)}{(1 - e^{-\rho\xi}(1 + e^{-\rho})/2)^2}.$$

If  $c > (\rho C + c)(1 + e^{-\rho})/2$ , then  $dv/d\xi < 0$  and  $v$  is maximized over  $\xi \in [0, 1]$  by taking  $\xi = 0$ . Therefore, in that case,  $\delta = (\bar{\gamma}_1, \bar{\gamma}_2)$  is an EP. This is true if we take for instance  $\rho = 1$ ,  $C = 1$  and  $c = 10$ . Notice also that in that case, players 1 and 2 may switch strategies and by symmetry, this yields another EP.

### Acknowledgement

We are grateful to Dr. Alain Haurie and to an anonymous referee for their helpful comments and suggestions, and to Dr. Jerzy A. Filar for pointing out Nowak's paper.

### References

- [1] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, 1978.
  - [2] M. Breton, Equilibres pour des jeux séquentiels, Ph.D. Thesis (in French), Département d'informatique et de recherche opérationnelle, Univ. de Montréal, mai 1986.
  - [3] H. A. M. Couwenbergh, Stochastic games with metric state spaces, *Internat. J. of Game Theory* **9** (1980), 25–36.
  - [4] E. V. Denardo, Contraction mappings in the theory underlying dynamic programming, *SIAM Review* **9** (1967), 165–177.
  - [5] A. Federgruen, On  $N$ -person stochastic games with denumerable state space, *Advances in Applied Probability* **10** (1978), 452–471.
  - [6] A. Haurie and P. L'Ecuyer, Approximation and bounds in discrete event dynamic programming, *IEEE Trans. on automatic control* **AC-31** (1986), 227–235.
  - [7] P. L'Ecuyer, Processus de décision markoviens à étapes discrètes: application à des problèmes de remplacement d'équipement, Ph.D. thesis, published in *Les cahiers du GERAD*, report no. G-83-06, Ecole des H. E. C., Montréal, 1983.
  - [8] P. L'Ecuyer and A. Haurie, Discrete event dynamic programming with simultaneous events, *Mathematics of Operations Research* **13**, 1 (1988), 152–163.
  - [9] A. Maitra and T. Parthasarathy, On stochastic games I, *J.O.T.A.* **8** (1970), 289–300.
  - [10] J. Nash, Noncooperative games, *Annals of Math.* **54** (1951), 286–295.
  - [11] A. S. Nowak, On zero-sum stochastic games with general state space I, *Probability and Math. Statistics* **4** (1984), 13–32.
  - [12] T. Parthasarathy, Discounted, positive and noncooperative stochastic games, *Int. J. of Game Theory* **2** (1973), 25–37.
  - [13] L. S. Shapley, Stochastic games, *Proc. of the National Academy of Sciences of U.S.A.* **39** (1953), 1095–1100.
  - [14] M. J. Sobel, Noncooperative stochastic games, *Annals of Math. Stat.* **42** (1971), 1930–1935.
  - [15] M. J. Sobel, Continuous stochastic games, *J. of Applied Probability* **10** (1973), 597–604.
  - [16] W. Whitt, Representation and approximation of noncooperative sequential games, *SIAM J. on Control and Optimization* **18** (1980), 33–48.
-