APPROXIMATIONS IN DYNAMIC ZERO-SUM GAMES, II

MABEL M. TIDBALL*, ODILE POURTALLIER†, AND EITAN ALTMAN†

Abstract. We pursue in this paper our study of approximations of values and ϵ-saddle-point policies in dynamic zero-sum games. After extending the general theorem for approximation, we study zero-sum stochastic games with countable state space, and unbounded immediate reward. We focus on the expected average payoff criterion. We use some tools developed in the first paper, to obtain the convergence of the values as well as the convergence of the ϵ saddle-point policies in various approximation problems. We consider several schemes of truncation of the state space (e.g. finite state approximation) and approximations of games with discount factor close to one for the game with expected average cost. We use the extension of the general Theorem for approximation to study approximations in stochastic games with complete information. We finally consider the problem of approximating the sets of policies. We obtain some general results that we apply to a pursuit evasion differential game.

Key words. Zero-sum games, Approximations, Stochastic games, pursuit evasion games.

AMS Subject Classification. 90D05, 93E05

1. Introduction. We pursue in this paper our study of approximations of values and saddle-point policies in dynamic zero-sum games. In a previous paper [34], we developed some tools for approximating zero-sum games, and applied them to stochastic games with discounted payoff criterion. In this paper we extend the general theory for approximation to handle cases where a value does not exist for the limit game, and we apply the general theorems for approximation to some dynamic zero-sum games.

We first consider approximation problems arising in stochastic games with expected average cost: finite state approximation of stochastic games with a countable state space, and convergence of stochastic games with discounted cost to the stochastic game with average cost. We then consider approximations in stochastic games with complete information, and problems in dynamic games related to discretization of the strategy sets.

There is a rich literature on finite state approximation in the context of a single controller. The discounted reward was extensively studied, see [2, 13, 21, 22, 35, 36], and [29, 37, 38] for related discretization results. For the expected average cost, there exist only few works on state approximations in the context of control, and none in the context of stochastic games. Even if existing schemes could be extended to the setting of a stochastic game, they are still quite restricted since their convergence (in the setting of control) was established under conditions that seem very strong, and quite often non-applicable. Thomas and Stengos obtained several schemes for finite state approximations. They impose, some scrambling conditions which should hold uniformly in the states. They do not seem to hold for queueing applications, such as the models in [3, 4, 7]. Altman introduced several finite state approximation schemes [1, 2] for constrained control. They do not require the scrambling conditions, but have other restrictive conditions: the scheme in [1] requires some monotone structure on the immediate cost, and holds for immediate costs that are only functions of the state, and not of the actions. The scheme in [2] has the "finite neighbor" restriction,

*Facultad de Ciencias Exactas, Ingeniera y Agrimensura, Universidad Nacional de Rosario, Pellegrini 250, 2000 Rosario, Argentina.

†INRIA, Centre Sophia-Antipolis, 2004 Route des Lucioles, B.P.93, 06902 Sophia-Antipolis Cedex, France.

i.e., from each state, only finitely many states are accessible within one step.

The two approximation schemes that we introduce in the current paper relax the above restrictions, and are thus also useful and new in the case of a single controller. In addition to the convergence of the value, which is the question studied in most of the papers on state approximations, we obtain (i) the convergence of the policies, (ii) the robustness of policies, i.e. an equilibrium point for the limiting (infinite state) stochastic game $G = G_\infty$ is shown to be an ϵ-equilibrium for the approximating games G_n for n large enough. On the other hand, for any ϵ, the equilibrium policies for G_n are almost optimal for the limiting game, for all n large enough.

In the previous paper [34] we focused on approximations of stochastic games with discounted cost and bounded reward, and mentioned that standard techniques can be used to transform problems with unbounded reward to problems with bounded ones. This is, however, not the case for the expected average payoff criterion. The question of existence of value and of equilibrium stationary policies (under some recurrence conditions) for the case of unbounded reward was solved recently in [7, 8, 12, 30]. The growing interest in stochastic games with unbounded cost in recent years was partly driven by applications of stochastic games in telecommunications systems in general, and in queueing systems in particular. Although queues are always finite in practice (which results in a finite state space description), models of infinite queues are frequently more useful since they are usually easier to solve. Indeed, several dynamic games arising in such applications were explicitly solved [7, 9], or at least reduced to the search for equilibrium policies among small classes of policies [4, 5, 7]. The scheduling problem described in [7, 9], the routing problem into two queues [4, 7], the flow and service control in [5] have not been solved for the case of finite state space, since there is an effect of the boundaries due to the finiteness of the queues that destroys the nice structure of the problem with infinite state space. In all the above problems, it is unnatural to consider bounded costs. Since costs represent queue lengths or waiting times, these typically grow to infinity as the number of "customers" in the queues grows to infinity. The theory developed in this paper allows the use of the equilibrium policies obtained for the infinite queues to construct ϵ-equilibrium policies for the corresponding problems with finite queues, provided that they are sufficiently large.

A second issue in this paper is the convergence of stochastic games in the discount factor. The convergence of the value and equilibrium policies for discounted cost stochastic games to those of the average cost game are well known, see e.g. [18]. These were extended recently to unbounded cost (see [8, 30]). We not only obtain an alternative proof for the above convergence of the values and policies, but also obtain new robustness results.

When the players are restricted to use pure strategies in a stochastic game, the game in general does not have a value anymore. Using an extension of the general approximation theorems, we study approximations under that restriction. This yields approximation theorems for stochastic games with complete information (where player 2 knows at time t the action taken by player 1 at time t).

Finally, we consider the problem of approximating the set of policies by other sets. We obtain a general approximating theorem for the case that the strategy sets are endowed with the Hausdorff metric. We apply the theorem to a zero-sum pursuit evasion differential game introduced in [11, 31].

The structure of the paper is as follows. We begin in Section 2 by citing and extending the general theory for approximations, developed in [34]. We then introduce in Section 3 the model, notation and assumptions for the stochastic game. We present two schemes for state approximation in Section 4. The convergence in the discount factor is established in Section 5. In Section 6 we discuss approximations for stochastic games with complete information. The approximation of the strategy sets is finally presented in Section 7 together with the application to the pursuit evasion game.

2. Key theorems for approximations. We consider the following sequence $G_n = (S_n, U_n, V_n)$ $n = 1, 2, ..., \infty$ of generic zero-sum games where U_n is the set of strategies (or policies) of player one and V_n is the set of strategies of player two for the nth game. We assume that both U_n and V_n are endowed with some topology. $S_n : U_n \times V_n \to \mathbb{R}$ is a measurable function for all n. We define the upper (lower) value of the game:

$$\overline{R_n} = \inf_{v \in V_n} \sup_{u \in U_n} S_n(u, v) \quad \left(\underline{R_n} = \sup_{u \in U_n} \inf_{v \in V_n} S_n(u, v) \right). \tag{2.1}$$

$G = (S, U, V) \stackrel{\text{def}}{=} (S_\infty, U_\infty, V_\infty)$ will be called the limit game. It will first be assumed (Theorem 2.1) that it has a value $R \stackrel{\text{def}}{=} R_\infty = \mathbf{Val}\left\{S(u, v)\right\}_{u,v}$. This assumption will be relaxed in Theorem 2.5.

A strategy $u^* \in U_n$ is said to be ϵ-optimal for player one in game n if

$$\inf_{v \in V_n} S_n(u^*, v) \geq \inf_{v \in V_n} S_n(u, v) - \epsilon \qquad \forall u \in U_n, \tag{2.2}$$

which is equivalent to $\inf_{v \in V_n} S_n(u^*, v) \geq \underline{R}_n - \epsilon$. It is said to be strong ϵ-optimal for player one in game n if it satisfies

$$\inf_{v \in V_n} S_n(u^*, v) \geq \overline{R}_n - \epsilon.$$

A strategy $v^* \in V_n$ is said to be ϵ-optimal for player two in game n if

$$\sup_{u \in U_n} S_n(u, v^*) \leq \sup_{u \in U_n} S_n(u, v) + \epsilon \qquad \forall v \in V_n, \tag{2.3}$$

which is equivalent to $\sup_{u \in U_n} S_n(u, v^*) \leq \overline{R}_n + \epsilon$. It is said to be strong ϵ-optimal if

$$\sup_{u \in U_n} S_n(u, v^*) \leq \underline{R}_n + \epsilon.$$

Note that strong ϵ-optimality implies ϵ-optimality. If a game has a value $\underline{R}_n = \overline{R}_n$ then strong ϵ-optimality is equivalent to ϵ-optimality.

Assume that (S_n, U_n, V_n) converge (in some sense) to (S, U, V). We are interested in the following questions:

(Q1) Convergence of the values: does $\underline{R}_n$ (or $\overline{R}_n$) converge to R?

(Q2) Convergence of policies: Fix some $\epsilon \geq 0$. Let ϵ_n be a sequence of positive real numbers such that $\overline{\lim}_{n\to\infty} \epsilon_n \leq \epsilon$. Assume that u_n^* and v_n^* are ϵ_n-optimal policies for the nth game. Are u_n^* and v_n^* "almost" optimal for the limit game, for all n large enough?

(Q3) Let $\overline{u} \in U$ (resp. $\overline{v} \in V$) be some limit point of u_n^* (resp. v_n^*), defined above. Is $\overline{u}$ (resp. $\overline{v}$) ϵ-optimal for the limit game?

(Q4) Robustness of the optimal policy: If u^* (resp. v^*) is an ϵ-optimal policy for the limit game, can we derive of it an "almost" (strong) optimal policy for the nth approximating game, for all n large enough?

A straightforward generalization of Theorem 2.1 in [34] yields :

THEOREM 2.1. *Assume that for any $\epsilon_1 > 0$ there exists a sequence of functions, $\pi_n^1 : U_n \to U$, $\pi_n^2 : V_n \to V$, $\sigma_n^1 : U \to U_n$, $\sigma_n^2 : V \to V_n$, $n = 1, 2, \ldots$ such that*
($\mathcal{A}1$): $\overline{\lim}_{n\to\infty}[S_n(u, \sigma_n^2(v)) - S(\pi_n^1(u), v)] \le \epsilon_1$ uniformly in $u \in U_n$ for each $v \in V$.
($\mathcal{A}2$): $\underline{\lim}_{n\to\infty}[S_n(\sigma_n^1(u), v) - S(u, \pi_n^2(v))] \ge -\epsilon_1$ uniformly in $v \in V_n$ for each $u \in U$.
Then
(1) $\lim_{n\to\infty} \underline{R}_n = \lim_{n\to\infty} \overline{R}_n = R$.
(2) For any $\epsilon' > \epsilon + 3\epsilon_1$, there exists N such that $\pi_n^1(u_n^*)$ (resp. $\pi_n^2(v_n^*)$, see definitions in (Q2)) is ϵ'-optimal for the limit game, for all $n \ge N$.
(3) Let u^* (resp. v^*) be ϵ-optimal for the limit game. Then for all $\epsilon' > \epsilon + 3\epsilon_1$, there exists $N(\epsilon')$ such that $\sigma_n^1(u^*)$ (resp. $\sigma_n^2(v^*)$) is strong ϵ'-optimal for the nth approximating game, for all $n \ge N(\epsilon')$.
(4) Suppose
($\mathcal{A}3$): $S(u, v)$ is a lower semicontinuous function in u for each v,
($\mathcal{A}4$): $S(u, v)$ is an upper semicontinuous function in v for each u.
Suppose $\bar{u} \in U$ (resp. $\bar{v} \in V$) is a limit point of $\pi_n^1(u_n^*)$ (resp. $\pi_n^2(v_n^*)$). Then $\bar{u}$ (resp. $\bar{v}$) is $(\epsilon + 5\epsilon_1)$-optimal for the limit game.

REMARK 2.2. *(i) Whenever $U_n = U$ and $V_n = V$ do not depend on n, π_n and σ_n will be chosen as the identity maps.*
(ii) It follows from the proof of part (1) in the above Theorem that if for every G_n, $n = 1, 2, \ldots, \infty$ there exist optimal policies for both players and if $U_n = U$ and $V_n = V$ do not depend on n, then

$$|\overline{R}_n - R| \le \sup_{u,v} |S_n(u, v) - S(u, v)|, \qquad |\underline{R}_n - R| \le \sup_{u,v} |S_n(u, v) - S(u, v)|$$

REMARK 2.3. *Some results related to those in Theorem 2.1 can be found in [10, 17, 20, 26] and in [27] who also consider additional constraints on the policies is studied there.*

We now relax the assumption that the limit game has a value : $\underline{R}_\infty \neq \overline{R}_\infty$. We show that Theorem 2.1 still holds, by appropriately enlarging the policy spaces and redefining the cost, so that the upper (or lower) value becomes a real value of a new game.

We consider the convergence of the upper values (and corresponding optimal or almost optimal policies) of the approximating games to those of the limit game. The corresponding convergence for the lower values are obtained in a similar way. Define $\mathcal{U}_n = \{$ the class of functions $V_n \to U_n\}$. Define the cost $\hat{\mathcal{S}}_n : \mathcal{U}_n \times V_n \to \mathbb{R}$ by $\hat{\mathcal{S}}_n(\psi, v) = S_n(\psi(v), v)$.

LEMMA 2.4. *(i) For all n, the new game $\mathcal{G}_n = (\hat{\mathcal{S}}_n, \mathcal{U}_n, V_n)$ has a value $\mathcal{R}_n$, and $\mathcal{R}_n = \overline{R}_n$.*
(ii) v^* is $\epsilon-$optimal for player 2 in game $\mathcal{G}_n$ if and only if it is $\epsilon-$optimal in game G_n.

Proof.

$$\inf_{v\in V_n} \sup_{\psi\in\mathcal{U}_n} \hat{\mathcal{S}}_n(\psi, v) = \inf_{v\in V_n} \sup_{\psi\in\mathcal{U}_n} S_n(\psi(v), v) = \inf_{v\in V_n} \sup_{u\in U_n} S_n(u, v) = \overline{R}_n. \tag{2.4}$$

On the other hand,

$$\inf_{v\in V_n} \sup_{\psi\in\mathcal{U}_n} S_n(\psi(v), v) = \sup_{\psi\in\mathcal{U}_n} \inf_{v\in V_n} S_n(\psi(v), v) = \sup_{\psi\in\mathcal{U}_n} \inf_{v\in V_n} \hat{\mathcal{S}}_n(\psi, v). \tag{2.5}$$

The first equality in (2.5) is due to the following. Clearly,

$$\inf_{v\in V_n}\sup_{\psi\in\mathcal{U}_n} S_n(\psi(v),v) \geq \sup_{\psi\in\mathcal{U}_n}\inf_{v\in V_n} S_n(\psi(v),v),$$

and we have to show that the reverse inequality also holds. Fix some $\epsilon' > 0$ and let ψ^* be such that for all v and all $\psi \in \mathcal{U}_n$,

$$S_n(\psi^*(v),v) \geq S_n(\psi(v),v) - \epsilon'.$$

Hence,

$$\inf_{v\in V_n} S_n(\psi^*(v),v) \geq \inf_{v\in V_n}\sup_{\psi\in\mathcal{U}_n} S_n(\psi(v),v) - \epsilon'.$$

We conclude that

$$\sup_{\psi\in\mathcal{U}_n}\inf_{v\in V_n} S_n(\psi(v),v) \geq \inf_{v\in V_n}\sup_{\psi\in\mathcal{U}_n} S_n(\psi(v),v) - \epsilon',$$

which establishes the first equality in (2.5). (i) is obtained by combining (2.4) and (2.5). (ii) follows since for any $v \in V_n$,

$$\sup_{u\in U_n} S_n(u,v) = \sup_{\psi\in\mathcal{U}_n} \hat{\mathcal{S}}_n(\psi,v).$$

□

By using the new games for which the values exist, and applying Theorem 2.1, we may conclude the following convergence properties of the original games.

THEOREM 2.5. *Assume that the functions π_n and σ_n exist as in Theorem 2.1, and that conditions ($\mathcal{A}1$) and ($\mathcal{A}2$) hold. Then*
(1) $\lim_{n\to\infty}\overline{R}_n = \overline{R}$, $\lim_{n\to\infty}\underline{R}_n = \underline{R}$.
(2) For any $\epsilon' > \epsilon + 3\epsilon_1$, there exists N such that $\pi_n^1(u_n^*)$ (resp. $\pi_n^2(v_n^*)$, see definitions in (Q2)) is ϵ'-optimal for the limit game, for all $n \geq N$.
(3) Let u^* (resp. v^*) be ϵ-optimal for the limit game. Then for all $\epsilon' > \epsilon + 3\epsilon_1$, there exists $N(\epsilon')$ such that $\sigma_n^1(u^*)$ (resp. $\sigma_n^2(v^*)$) is ϵ'-optimal for the n approximating game, for all $n \geq N(\epsilon')$.

Proof. Consider the new games $\mathcal{G}_n$ defined above. We show that the assumptions of Theorem 2.1 hold also for $\mathcal{G}_n$. The mapping $\tilde{\pi}_n^2, \tilde{\sigma}_n^2$ for the new games are unchanged;

$$\tilde{\pi}_n^2 = \pi_n^2, \qquad \tilde{\sigma}_n^2 = \sigma_n^2.$$

The mappings $\tilde{\pi}_n^1 : \mathcal{U}_n \to \mathcal{U}$ and $\tilde{\sigma}_n^1 : \mathcal{U} \to \mathcal{U}_n$ for the new games are defined as

$$[\tilde{\pi}_n^1(\psi)](v) = \pi_n^1(\psi(v)), \qquad \forall v \in V, [\tilde{\sigma}_n^1(\psi)](v) = \sigma_n^1(\psi(v)), \qquad \forall v \in V_n.$$

With these definitions as well as the definition of the costs $\hat{\mathcal{S}}_n$, it follows that ($\mathcal{A}1$) and ($\mathcal{A}2$) hold for $\mathcal{G}_n$. The proof now follows by Lemma 2.4. □

We may further obtain convergence results for the optimal (or ϵ-optimal) *responses* (in case that the value of the limit game does not exist). To simplify the formulation, this is done below in terms of the new games $\mathcal{G}_n$:

THEOREM 2.6. *Consider the new games $\mathcal{G}_n$, and let ψ_n^*, v_n^* be defined as u_n^* in (Q2) (above Theorem 2.1). Under the conditions of Theorem 2.5,*

(1) $\lim_{n\to\infty} \mathcal{R}_n = \mathcal{R} = \overline{R}$,
(2) For any $\epsilon' > \epsilon + 3\epsilon_1$, *there exists* N *such that* $\tilde{\pi}_n^1(\psi_n^*)$ *(resp.* $\tilde{\pi}_n^2(v_n^*)$*) is* $\epsilon'-$*optimal for* $\mathcal{G}_\infty$, *for all* $n \geq N$.
(3) Let ψ^* *be* ϵ*-optimal for player 1 in the limit game* $\mathcal{G}_\infty$. *Then for all* $\epsilon' > \epsilon + 3\epsilon_1$, *there exists* $N(\epsilon')$ *such that* $\tilde{\sigma}_n^1(\psi^*)$ *(resp.* $\tilde{\sigma}_n^2(v^*)$*) is* ϵ'*- optimal for the* n *approximating game* $\mathcal{G}_n$ *for all* $n \geq N(\epsilon')$.

Next, we consider the result corresponding to statement (4) in Theorem 2.1.

THEOREM 2.7. *Assume that the conditions of Theorem 2.6 hold, that the set of response-strategies for player 2 in games* $\mathcal{G}_n$ *is endowed with some topology, and that* $(\mathcal{A}3)$ *and* $(\mathcal{A}4)$ *hold for game* $\mathcal{G}_\infty$. *Then statement (4) of Theorem 2.1 holds for games* $\mathcal{G}_n$.

3. Stochastic games with expected average payoff. We consider the two person, zero-sum stochastic game defined by the objects $\{\mathbf{I}, \mathbf{A}, \mathbf{B}, P, r\}$, where

- $\mathbf{I}$ is a countable state space;
- $\mathbf{A}$ and $\mathbf{B}$ are sets of actions for player I and player II, respectively; at each state $j \in \mathbf{I}$, the available actions for the players are $\mathbf{A}_j$ and $\mathbf{B}_j$ respectively. These sets are assumed to be compact metric sets.
- $P(a, b) = [p(i, a, b, j)]_{i,j}$, $a \in \mathbf{A}$, $b \in \mathbf{B}$, $i, j \in \mathbf{I}$ are the transition probabilities, so that $p(i, a, b, j)$ is the probability to move from i to j if the players use actions a and b.
- $r : \mathbf{I} \times \mathbf{A} \times \mathbf{B} \to \mathbb{R}$ is an immediate reward function.

The game is played in stages $t = 0, 1, 2, ...$ If at some stage t the state is i, then the players independently choose actions $a \in \mathbf{A}_i$, $b \in \mathbf{B}_i$. Player II then pays player I the amount $r(i, a, b)$ and at stage $t + 1$ the new state is chosen according to the transition probabilities $p(i, a, b, \bullet)$. The game continues at this new state.
Let U and V be the set of behavioral strategies for both players. A strategy $u \in U$ is a sequence $u = (u_0, u_1, ...)$ where u_t is a probability measure over the available actions, given the whole history of previous states and of previous actions of both players as well as the current state.

A Markov policy $q = \{q_0, q_1, ...\}$ is a policy (for either player one or two) where q_t is allowed to depend only on t and on the state at time t.

A *stationary (mixed) policy* g for player one is characterized by a conditional distribution
$p^g(\bullet \mid j)$ over $\mathbf{A}_j$, so that $p^g(\mathbf{A}_j \mid j) = 1$, which is interpreted as the distribution over the actions available at state j which player I uses when it is in state j. With some abuse of notation, we shall set $g(\bullet \mid j) = p^g(\bullet \mid j)$ for stationary g. Let U_S be the set of stationary policies for player 1, and define similarly the stationary policies V_S for player 2. If both players use stationary policies, say u and v, then $\{X_t\}$ becomes a Markov chain with stationary transition probabilities, given by

$$p(j, u, v, k) := \int_{\mathbf{A}_j} \int_{\mathbf{B}_j} p(j, a, b, k) u(da|j) v(db|j). \tag{3.1}$$

The expected immediate reward at state j becomes

$$r(j, u, v) := \int_{\mathbf{A}_j} \int_{\mathbf{B}_j} r(j, a, b) u(da|j) v(db|j).$$

Denote by $P(u, v)$ the (infinite dimensional) matrix whose (j, k)th component equals $p(j, u, v, k)$. Similarly, denote by $r(u, v)$ the column vector whose jth component equals $r(j, u, v)$.

Next, we introduce a topology on the sets of stationary policies. For any compact metric set Γ, let $M_1(\Gamma)$ denote the set of probability measures on the Borel subsets of Γ endowed with the weak topology $\xi(\Gamma)$ (see [28]). The class of stationary policies for player 1 (and similarly for player 2) can be identified with the set $\prod_{i\in\mathbf{I}} M_1(\mathbf{A}_i) \times M_1(\mathbf{B}_i)$; moreover it is compact with respect to the product topology $\prod_{i\in\mathbf{I}} \xi(\mathbf{A}_i) \times \xi(\mathbf{B}_i)$, and it is metrizable (by virtue of Theorem 4.14 in Kelley [23]).

Let (u, v) be a pair of strategies and let $i \in \mathbf{I}$ be a fixed initial state. Let $I_t, A_t, B_t, t = 0, \ldots$ be the resulting stochastic process of the states and actions of the players. Let $E_i^{u,v}$ denote the expectation with respect to the measure defined by u, v, i.

Let $\mu : \mathbf{I} \to \mathbb{R}$ be some positive function. Following Dekker and Hordijk [16] and Spieksma [32], define the μ-norm of any vector $x \in \mathbb{R}^{\mathbf{I}}$ as

$$\| x \|_\mu = \sup_{i\in\mathbf{I}} \frac{|x_i|}{\mu_i}.$$

In a similar way we will use the μ-J-norm, for any finite subset J of the state space I, defined by

$$\| x \|_\mu^J = \sup_{i\in J} \frac{|x_i|}{\mu_i}.$$

Define the μ-norm of matrices $Q \in \mathbb{R}^{\mathbf{I}\times\mathbf{I}}$ as

$$\|Q\|_\mu = \sup_{i\in\mathbf{I}} \mu_i^{-1} \sum_{j\in\mathbf{I}} |Q_{ij}| \mu_j.$$

We denote $\bigvee_\mu = \{x : \| x \|_\mu < \infty\}$ the space of all vectors that are μ-bounded.

Introduce the following assumptions :

- $(\mathcal{B}1)$
  - i) The instantaneous reward $r(i, a, b)$ is continuous and μ-bounded, i.e.;

    $$\sup_{i\in\mathbf{I}} \sup_{a,b} \frac{|r(i,a,b)|}{\mu_i} \le M < +\infty.$$

    This condition can be rewritten as $\|r(\cdot, u, v)\|_\mu \le M < +\infty$ for all pure stationary policies u and v.
  - ii) The transition probabilities are μ-continuous, i.e. if $a(n) \to a$, $b(n) \to b$ when $n \to +\infty$ then :

    $$\lim_{n\to\infty} \sum_{j\in\mathbf{I}} |p(i, a(n), b(n), j) - p(i, a, b, j)| \, \mu_j = 0 \qquad \forall i \in I.$$

- $(\mathcal{B}2)$
  - i) Under any pure stationary policies for the players, the state space does not contain more than one ergodic class.
  - ii) There exists a finite set $\mathcal{M} \subset E$ and a constant $\beta < 1$ such that :

    $$\sum_{j\in I} {}_{\mathcal{M}}p(i, a, b, j) \, \mu_j \le \beta \, \mu_i \qquad \forall a, b, \forall i \in \mathbf{I}, \tag{3.2}$$

    where ${}_{\mathcal{M}}p(i, a, b, j) = p(i, a, b, j)$ if j does not belong to the set $\mathcal{M}$, and is null otherwise. (3.2) can be rewritten as $\sum_{j\notin\mathcal{M}} p(i, a, b, j) \, \mu_j \le \beta \, \mu_i \;\; \forall a, b, \forall i \in \mathbf{I}$, or as $\|{}_{\mathcal{M}}P(u, v)\|_\mu \le \beta$ for all pure stationary policies u and v.

REMARK 3.1. *If assumptions ($\mathcal{B}1$) and ($\mathcal{B}2$) hold for some $\mathcal{M}$ and μ, then one can choose a state $0 \in \mathcal{M}$ and another μ' such that these assumptions hold with the set $\mathcal{M}' = \{0\}$ replacing $\mathcal{M}$, and μ' replacing μ, (see [32]). Therefore, we assume in the sequel, without loss of generality, that $\mathcal{M} = \{0\}$, for some state* 0.

Define

$$S(i,u,v) = \overline{\lim_{t\to\infty}} \frac{1}{t} E_i^{u,v} \sum_{s=0}^{t-1} r(I_s, A_s, B_s). \tag{3.3}$$

THEOREM 3.2. *(see [8]) Suppose that assumptions ($\mathcal{B}1$) and ($\mathcal{B}2$) hold. Then*
(i) The stochastic game with expected average payoff criterion has a value,
(ii) There exists a unique solution pair (g, v), $g \in \mathbb{R}$, $v \in R^{\mathbf{I}}$ to the functional equation :

$$v(i) = \mathbf{Val} \left\{ r(i,a,b) - g + \sum_{j\in\mathbf{I}} p(i,a,b,j) v(j) \right\}_{a,b} \qquad i \in \mathbf{I}. \tag{3.4}$$

such that $\|v\|_\mu$ is finite and $v(0) = 0$.
(iii) $g = R(i) = \mathbf{Val}\{S(i,u,v)\}_{u,v}$ is the unique value of the stochastic game.
(iv) Let (u,v) be stationary policies such that $u(i), v(i)$ are optimal for the dummy game in the curly brackets in (3.4). Then, they are optimal for the stochastic game.

4. State truncation and approximation. In the following approximating schemes, we modify the "limit" stochastic game in the following way. We consider an increasing sequence of sets of states $\mathbf{I}_1, \mathbf{I}_2, \ldots$ converging to $\mathbf{I}$, such that $0 \in \mathbf{I}_1$. The nth stochastic game is restricted to the set $\mathbf{I}_n$. In the game G_n, we modify the transition probabilities so as to eliminate all transitions outside the set $\mathbf{I}_n$. The two schemes will differ by the way that such transitions will be replaced. Introduce the following assumption

- ($\mathcal{B}3$) $\quad \delta(r,n) = \sup_{\substack{i\in\mathbf{I}_r \\ a\in\mathbf{A}, b\in\mathbf{B}}} \sum_{j\notin\mathbf{I}_n} p(i,a,b,j)\mu(j) \to 0$ as $n \to +\infty, \quad \forall r.$

Under the assumptions of our model (i.e. ($\mathcal{B}1$)-($\mathcal{B}2$)), ($\mathcal{B}3$) holds if $\mathbf{I}_n$ are finite sets $\forall n$ (the proof is similar to the one in [34]).

4.1. Scheme I. In the game G_n, we modify the transition probabilities so as to eliminate all transitions outside the set $\mathbf{I}_n$; we replace transitions outside of $\mathbf{I}_n$ by transitions to state 0. Hence, $p^n(i,a,b,j)$ is defined by :

$$p^n(i,a,b,j) = \begin{cases} p(i,a,b,0) + \sum_{l\notin\mathbf{I}_n} p(i,a,b,l) & j = 0, i \in \mathbf{I}_n \\ p(i,a,b,j) & j \neq 0, i,j \in \mathbf{I}_n \\ 0 & j \notin \mathbf{I}_n \\ 1\{j=0\} & i \notin \mathbf{I}_n \end{cases} \tag{4.1}$$

For $i \in \mathbf{I}_n$, $S_n(i,u,v)$ is defined as (3.3), where the expectation is taken with respect to the signed measure generated by the new transition probabilities (4.1). $S_n(i,u,v)$ is defined to be 0 for $i \notin \mathbf{I}_n$, for all u and v.

For $n \in \mathbb{N}$, let the pair $(g_n, v_n \in \bigvee_\mu)$ be the solutions of the dynamic programming equation :

$$
(4.2) \quad \begin{array}{ll} v_n(i) = \mathbf{Val}\left\{ r(i,a,b) - g_n + \sum_{j\in\mathbf{I}} p^n(i,a,b,j) v_n(j) \right\}_{a,b} & i \in \mathbf{I}_n, i \neq 0 \\ v_n(i) = 0 & i \notin \mathbf{I}_n \text{ or } i = 0 \end{array}
$$

One can show using Theorem 3.2 that for all n, g_n and v_n indeed exist and are unique. Moreover, we have that :

$$
R_n(i) = \mathbf{Val}\left\{ S_n(i,u,v) \right\}_{u,v} = g_n \qquad \forall i \in \mathbf{I}.
$$

In order to prove the convergence of the state approximation scheme we introduce the following quantities :

- $\tau := \inf\{t \geq 1, I_t = 0\}$ is the time to reach state zero (with the convention that $\inf\{t : t \in \emptyset\} = \infty$).
- $w(i,u,v)$:= The total cost to reach zero from state i when policies u and v are used

  $$
  w(i,u,v) = E_i^{u,v} \sum_{s=0}^{\tau} r(I_s, A_s, B_s)
  $$

  which can be rewritten in a vector form as :

  $$
  w(u,v) = \sum_{s=0}^{\infty} [{}_0P(u,v)]^s r(u,v).
  $$

- $w^n(i,u,v)$:= The total cost to reach zero from state i when policies u and v are used, when the transition probabilities are replaced by (4.1).
- $\overline{\tau}(i,u,v), \overline{\tau}^n(i,u,v)$:= The expectations of τ when using the original transition probabilities, and when using those in (4.1), respectively. For $i \notin \mathbf{I}_n$ we have $\overline{\tau}^n(i,u,v) = 1$ for all policies u and v.

We note that $w(\cdot,u,v)$, $w^n(\cdot,u,v)$, $\overline{\tau}(\cdot,u,v)$ and $\overline{\tau}^n(\cdot,u,v)$ are uniformly μ bounded. Indeed,

$$
(4.3) \qquad \|w(u,v)\|_\mu \leq \sum_{s=0}^{\infty} \left[\|{}_0P(u,v)\|_\mu \right]^s \|r(u,v)\|_\mu \leq \frac{M}{1-\beta}
$$

with the same bound for $w^n(u,v)$. Similarly, $\|\overline{\tau}(u,v)\|_\mu$ and $\|\overline{\tau}^n(u,v)\|_\mu$ are bounded by $(1-\beta)^{-1}$. It is easily seen that $w(\cdot,u,v)$, $w^n(\cdot,u,v)$, $\overline{\tau}(\cdot,u,v)$ and $\overline{\tau}^n(\cdot,u,v)$ are the unique solution in $\bigvee_\mu$ of the fixed point equations

$$
(4.4) \qquad w(i,u,v) = r(i,u,v) + \sum_{j\neq 0} p(i,u,v,j) w(j,u,v)
$$

$$
(4.5) \qquad w^n(i,u,v) = \begin{cases} r(i,u,v) + \sum_{j\neq 0} p^n(i,u,v,j) w^n(j,u,v), \text{ for } i \in I_n \\ \\ 0 \text{ for } i \notin I_n \end{cases}
$$

$$
(4.6) \qquad \overline{\tau}(i,u,v) = 1 + \sum_{j\neq 0} p(i,u,v,j) \overline{\tau}(j,u,v)
$$

$$
(4.7) \qquad \overline{\tau}^n(i,u,v) = \begin{cases} 1 + \sum_{j\neq 0} p^n(i,u,v,j) \overline{\tau}^n(j,u,v), \text{ for } i \in I_n \\ \\ 1 \text{ for } i \notin I_n \end{cases}
$$

The uniqueness follows from the fact that the above equations are contracting due to ($\mathcal{B}2$). Note that functions $w(.,u,v)$ and $w^n(.,u,v)$ are μ-bounded for all pair (u,v) on every subset J of $\mathbf{I}$. Since both $\overline{\tau}^n(0,u,v)$ and $\overline{\tau}(0,u,v)$ are both non-zero and finite, it follows (see Chung [14] p. 91-92) that the expected average cost is given by the following ratio between the total cost and the expected hitting time of state zero

$$S(i,u,v) = \frac{w(0,u,v)}{E\tau(0,u,v)} \quad \text{and } S_n(i,u,v) = \frac{w^n(0,u,v)}{E\tau^n(0,u,v)}. \tag{4.8}$$

THEOREM 4.1. *Assume ($\mathcal{B}1$)-($\mathcal{B}3$). All statements of Theorem* 2.1 *hold, where S_n and S are the expected average payoffs defined in (3.3), with the transition probabilities p and p^n (defined in (4.1)) respectively.*

Proof. Fix some initial state i. We use Theorem 2.1. We begin by establishing conditions ($\mathcal{A}1$) and ($\mathcal{A}2$). Since $U = U_n$ and $V_n = V$ for each n, it suffices to show that $S_n(u,v) := S_n(i,u,v)$ converges to $S(u,v) := S(i,u,v)$ uniformly on $\mathbf{I}$. Hence, we set π_n^1, π_n^2, σ_n^1 and σ_n^2 to be identity.
Let J be a given subset of $\mathbf{I}$, and (u,v) a pair of strategies. To avoid cumbersome notations we will write $w(.)$ (resp $w^n(.)$) instead of $w(.,u,v)$ (resp $w^n(.,u,v)$). We first want to prove that

$$\lim_{n\to+\infty} \| w^n - w \|_\mu^J = 0.$$

Once we show that, one obtains in the same way that $\lim_{n\to+\infty} \| E\tau - E\tau^n \|_\mu^J = 0$, and the uniform convergence of $S_n(u,v)$ to $S(u,v)$ now follows from (4.8).

We use an idea introduced by Cavazos-Cadena [13] and used in [34] for a similar problem. Fix ϵ arbitrarily small, and define the sequence g_k in the following way. $g_0 = \min\{m : J \subset \mathbf{I}_m\}$ and recursively,

$$g_k = g(\epsilon, g_{k-1}), \quad g(\epsilon, r) = \min\{m : \delta(r,m) \le \epsilon\},$$

where δ is defined in ($\mathcal{B}3$). Due to assumption ($\mathcal{B}3$) this sequence is well defined, and for all k, g_k is finite. Let ν be a given integer, define also

$$m^\nu(\epsilon) = \max\{g_m, m = 0, 1..., \nu\}.$$

Let $n \ge m^\nu(\epsilon)$, $i \in J$.
Let us now compute $\| w^n - w \|_\mu^J$. We obviously have that $\| w^n - w \|_\mu^J \le \| w^n - w \|_\mu^{I_{g_0}}$ since $J \subset I_{g_0}$, and for $i \in \mathbf{I}_{g_0}$,

$$\begin{aligned}
\frac{1}{\mu_i}|w^n(i) - w(i)| &= \frac{1}{\mu_i}\Big|\sum_{j\neq 0} p^n(i,u,v,j)w^n(j) - p(i,u,v,j)w(j)\Big| \\
&\le \frac{1}{\mu_i}\sum_{j\in\mathbf{I}_{g_1}\setminus\{0\}} |p^n(i,u,v,j)w^n(j) - p(i,u,v,j)w(j)| \\
&\quad + \frac{1}{\mu_i}\sum_{j\in\mathbf{I}_n\setminus\mathbf{I}_{g_1}} |p^n(i,u,v,j)w^n(j) - p(i,u,v,j)w(j)| \\
&\le \sum_{j\in\mathbf{I}_{g_1}\setminus\{0\}} \frac{p(i,u,v,j)\mu_j}{\mu_i}\frac{|w^n(j) - w(j)|}{\mu_j} \\
&\quad + \frac{1}{\mu_i}\sum_{j\in\mathbf{I}_n\setminus\mathbf{I}_{g_1}} p(i,u,v,j)|w^n(j) - w(j)|.
\end{aligned}$$

In the last inequality the first term can be bounded by $\beta \parallel w - w^n \parallel_\mu^{\mathbf{I}_{g_1}}$ because of assumption ($\mathcal{B}$2,ii), and the second by $2\frac{M}{1-\beta}\epsilon$, since

$$\sum_{j \in \mathbf{I}_n \setminus \mathbf{I}_{g_1}} p(i,u,v,j)|w^n(j) - w(j)| \leq \sum_{j \in \mathbf{I}_n \setminus \mathbf{I}_{g_1}} \mu_j p(i,u,v,j) \left[\frac{|w^n(j)|}{\mu_j} + \frac{|w(j)|}{\mu_j} \right],$$

due to the definition of the sequence I_{g_k}, since i belongs to I_{g_0}, due to ($\mathcal{B}$3), and since w^n and w are bounded by $\frac{M}{1-\beta}$, see (4.3). We obtain

$$\parallel w^n - w \parallel_\mu^{\mathbf{I}_{g_0}} \;\leq\; \beta \parallel w^n - w \parallel_\mu^{\mathbf{I}_{g_1}} + 2\frac{M}{1-\beta}\epsilon.$$

Exactly in the same way we get for $k \leq m^\nu(\epsilon)$

$$\parallel w^n - w \parallel_\mu^{\mathbf{I}_{g_k}} \;\leq\; \beta \parallel w^n - w \parallel_\mu^{\mathbf{I}_{g_{k+1}}} + 2\frac{M}{1-\beta}\epsilon,$$

and finally

$$\parallel w^n - w \parallel_\mu^{J} \;\leq\; \beta^\nu \frac{2M}{1-\beta} + \frac{2M\epsilon}{1-\beta}\left(\frac{1-\beta^\nu}{1-\beta}\right). \tag{4.9}$$

Since ν can be chosen arbitrarily large when n tends to infinity, and β is strictly lower than 1, this bound can be as small as needed for n large enough. This establishes ($\mathcal{A}$1)-($\mathcal{A}$2) in Theorem 2.1. It follows from [8] that S is a continuous function of u and v, which implies ($\mathcal{A}$3)-($\mathcal{A}$4) of Theorem 2.1. This completes the proof. □

4.2. Scheme II. In the previous scheme, we replaced transitions outside of $\mathbf{I}_n$ by transitions to state 0. In some applications this may be undesirable; this is the case when the games with truncated space describe real problems that we wish to approximate by some game with an infinite state space. To illustrate this, consider a queue with a finite length L, and assume that the state is the number of customers in the queue. Then, typically, if a transition from state L to state $L+1$ were possible in the case of infinite queue, then in the problem with truncated state space, which corresponds to a finite queue, it is replaced by a transition from L to L. In the previous scheme, it would be replaced by a transition to state 0. This would be especially undesirable, since in queueing problems, we usually have the property of transitions to closest neighbors : from each state, only finitely many neighboring states can be reached in one step. So, having a transition from state L to 0 does not describe a realistic model of a finite queue.

Let $\{q^n(i,a,b,j), i,j \in \mathbf{I}, a \in \mathbf{A}, b \in \mathbf{B}\}$ be sequences of measures such that for all n, $i \in \mathbf{I}$, $a \in \mathbf{A}, b \in \mathbf{B}$,

$$q^n(i,a,b,j) \leq 0 \text{ for } j \in \mathbf{I}_n, \qquad q^n(i,a,b,j) = 0 \text{ for } j \notin \mathbf{I}_n,$$

$$\sum_{j \in \mathbf{I}_n} (p(i,a,b,j) + q^n(i,a,b,j)) = 1.$$

The transitions for the approximating problems are then given by

$$p^n(i,a,b,j) = \begin{cases} p(i,a,b,j) + q^n(i,a,b,j) & i,j \in \mathbf{I}_n \\ 0 & \text{otherwise} \end{cases} \tag{4.10}$$

It follows that

$$\sum_{j \in \mathbf{I}_n} q^n(i,a,b,j) = \sum_{j \notin \mathbf{I}_n} p(i,a,b,j). \tag{4.11}$$

We make the following assumption on μ and on $\mathbf{I}_n$.

$$\text{For any } n > m \text{ and } i \in \mathbf{I}_n \setminus \mathbf{I}_m, \mu(i) \geq \sup_{j \in \mathbf{I}_m} \mu(j).$$

For $i \in \mathbf{I}_n$, $S_n(i,u,v)$ is defined as (3.3), where the expectation is taken with respect to the measure generated by the new transition probabilities (4.10). We set $S_n(i,u,v) = 0$ for all $i \notin \mathbf{I}_n$.

THEOREM 4.2. *Assume ($\mathcal{B}1$)-($\mathcal{B}3$), and consider the above finite approximation scheme. Then all statements of Theorem 2.1 hold.*

Proof. We consider as in the previous section the total expected cost and total expected time till between consecutive epochs that state 0 is reached. By similar arguments to those in the previous Scheme one then establishes that $S_n(u,v)$ converge to $S(u,v)$ uniformly in all stationary policies (the exact proof can be found in [6] Chapter 8). This implies assumptions ($\mathcal{A}1$) and ($\mathcal{A}2$). Assumptions ($\mathcal{A}3$) and ($\mathcal{A}4$) relate only to the limit game, and therefore the proof is the same as in the previous Section. The theorem now follows from Theorem 2.1. □

5. Convergence of the discounted cost to the average cost. Conditions for the convergence of the value and equilibrium policies for discounted cost stochastic games to those of the average cost game are well known, see e.g. [18]. These were extended recently to unbounded cost (see [8, 30]). Theorem 2.1 enables us not only to obtain an alternative proof for the above convergence of the values and policies, but also to obtain new robustness results as in Theorem 5.2 below.

Define the **β-discounted game payoff**

$$S_\beta(i,u,v) = (1-\beta) E_i^{u,v} \sum_{t=0}^{\infty} \beta^t r(I_t, A_t, B_t) \tag{5.1}$$

The following was proved in [8],Theorem 3.4 (see also [15]) :

THEOREM 5.1. *Assume ($\mathcal{B}1$) and ($\mathcal{B}2$). Then*
(i) A value $R_\beta(i)$ exists for the discounted cost.
(ii) Optimal stationary policies exist for both players for any discount factor $0 < \beta < 1$ (they are said to be β-optimal).
(iii) Any limit-point (as β tends to one) of β-optimal stationary policies is expected average optimal; moreover, the value of the discounted games converge to the value of the expected average game.

THEOREM 5.2. *(i) Let (u^*, v^*) be any stationary policy pair which is expected average optimal. Then for any $\epsilon > 0$, (u^*, v^*) is ϵ-optimal for the β-discounted cost, for all β sufficiently close to 1, and for all $u \in U$, $v \in V$,*

$$\overline{\lim_{\beta \to 1}} [S_\beta(i,u,v^*) - S_\beta(i,u^*,v^*)] \leq 0, \qquad \underline{\lim_{\beta \to 1}} [S_\beta(i,u^*,v) - S_\beta(i,u^*,v^*)] \geq 0.$$

(ii) For any $\epsilon > 0$ there exists some $\beta_0 < 1$ such that for any $\beta_0 \leq \beta < 1$ and any stationary pair u^β, v^β which are β-optimal, (u^β, v^β) is ϵ-optimal for the expected average game.

Proof. It is sufficient to prove that $(\mathcal{A}1)$ and $(\mathcal{A}2)$ in Theorem 2.1 hold. This follows indeed from the fact that

$$\lim_{\beta\to 1} \|S_\beta(\cdot,u,v) - S(\cdot,u,v)\|_\mu = 0$$

uniformly over all stationary policies u and v, see [2] p. 166. □

6. Stochastic games with complete information. In Sections 4, 5 and in [34] we described several approximation problems in stochastic games, where we had a value in the limit game. In all those cases, we considered (without loss of optimality) the (randomized) stationary or the (randomized) Markov policies. Consider now the game $G^0 = (S, U_D, V_D)$ where S is given in (3.3) for some fixed initial state i, and where U_D and V_D are the classes of the pure stationary policies. Then the game will not have in general a value : it will have typically a lower and an upper value.

Since we established conditions $(\mathcal{A}1)$ and $(\mathcal{A}2)$ for all the problems considered in Sections 4, 5 (and also for the discounted cost, in [34]), they hold in particular if we restrict to pure stationary strategies. Therefore, the convergence of $\underline{R}_n^D$ and $\overline{R}_n^D$ to the upper and lower values of the limit game ($\underline{R}^D$ and $\overline{R}^D$) as well as the convergence of the policies in Theorems 2.5 and 2.6 hold for all these problems. (G_n^0 is the nth approximating game, and $\overline{R}_n^D$ and $\underline{R}_n^D$ - its upper and lower values.)

Applying Theorem 2.7 is more delicate, since only in special cases, can we define a topology over the space of responses of player 2 such that assumptions $(\mathcal{A}3)$ and $(\mathcal{A}4)$ hold (as opposed to standard stochastic games, where $(\mathcal{A}3)$ and $(\mathcal{A}4)$ need to hold for policies, and not for responses). In case that the action space available to player 1 is finite, one may identify the class of pure stationary response strategies of player 2 (corresponding to pure stationary policies of player 1) with the set of functions $\mathbf{I} \times \mathbf{A} \to \mathbf{B}$, endowed with topology of weak convergence of measures. The continuity assumptions $(\mathcal{A}3)$ and $(\mathcal{A}4)$ can now be established using arguments as in Remark 3.1 in [34] and [8].

We shall now show the usefulness of the above results for stochastic games with complete information, in the sense used e.g. by Küenle [24, 25]. There, one of the players, say player 2, has at each time t the additional information of the action chosen by player 1 at time t (i.e., in addition to the information of all past states and actions of both players, plus the current state). The information structure for the other player is unchanged. We thus define the class of policies V^* to be the set of policies of the form $v = (v_0, v_1, ...)$ where v_t is a probability measure over $\mathbf{B}$ conditioned on the history $(x_0, a_0, b_0, x_1, a_1, b_1, ..., x_{t-1}, a_{t-1}, b_{t-1}, x_t, a_t)$.

Consider the following game : $G^1 = (S, U, V^*)$, where S is given by (3.3) for some fixed initial state i. (The results below will also hold for the expected discounted cost with infinite horizon, considered in [34].) We shall be interested in approximating

$$\underline{R}^1 = \sup_{u\in U} \inf_{v\in V^*} S(u,v).$$

Under quite general conditions [18, 24, 25] there exists optimal policies in U_D for player 1 in the sense that there exists some $u^1 \in U_D$ such that

$$\underline{R}^1 = \sup_{u\in U} \inf_{v\in V^*} S(u,v) = \sup_{u\in U_D} \inf_{v\in V^*} S(u,v) = \inf_{v\in V^*} S(u^1,v).$$

(This property does not hold for non-zero sum games, for which there are counter examples, see [18].)

On the other hand, under the conditions in Section 3, we have for any $u \in U_D$

$$\inf_{v \in V^*} S(u,v) = \inf_{v \in V} S(u,v) = \inf_{v \in V_D} S(u,v).$$

The first equality follows from the fact that $u \in U_D$, so that for any policy $v^* \in V^*$, we have $S(u,v) = S(u,v^*)$ where $v = (v_0, v_1, ...)$ is given by

$$v_t(x_0, a_0, b_0, ..., x_{t-1}, a_{t-1}, b_{t-1}, x_t) = v_t^*(x_0, a_0, b_0, ..., x_{t-1}, a_{t-1}, b_{t-1}, x_t, u(x_t)).$$

The second equality follows since for any $u \in U_D$, player 2 is now faced with a standard Markov decision process (see e.g. [32]). We thus conclude that $\underline{R}^1 = \underline{R}^D$. Thus, the convergence of $\underline{R}_n^D$ to $\underline{R}^D$ imply convergence of lower values for the game with complete information. Similarly, the convergence of the policies in Theorems 2.5 and 2.6 for games G_n^0 imply the convergence of policies for player 1 in the games with complete information that approximate G^1.

Remark 6.1. *An alternative approach to obtain convergence of values and policies for games with complete information is to transform these into standard stochastic games [25]. Note, however, that if the action space is uncountably infinite (e.g. an interval) then the transformation results in an uncountable state space.*

7. A finite approximation of strategy sets. Another type of approximation that arises in dynamic games is the countable or finite approximation of strategy sets. This step is necessary when we want to perform numerical computations, and when the strategy sets are infinite or continuous, or both.

Let $\mathcal{U}$ and $\mathcal{V}$ be metric sets of policies for players one and two, and let $S(u,v)$ correspond to the cost associated to the pair of strategies $u \in \mathcal{U}, v \in \mathcal{V}$. Introduce the following sets of strategies : $U \subset \mathcal{U}$ and $V \subset \mathcal{V}$, and the sequences $\{U_n\}_{n \in \mathbb{N}} \subset \mathcal{U}$ and $\{V_n\}_{n \in \mathbb{N}} \subset \mathcal{V}$. U_n and V_n are assumed to be countable or finite sets of policies.

Theorem 7.1. *Suppose that*
($\mathcal{A}$'1): $\lim_{n \to +\infty} U_n = U$ and $\lim_{n \to +\infty} V_n = V$, in the Hausdorff topology sense,
($\mathcal{A}$'3): $S(u,v)$ is a lower semicontinuous function in $u \in U$ uniformly in $v \in V$,
($\mathcal{A}$'4): $S(u,v)$ is a upper semicontinuous function in $v \in V$ uniformly in $u \in U$,

Then the conclusion of theorem 2.1 hold (where $\overline{R}_n$ and $\underline{R}_n$ are defined in (2.1), and R is defined with respect to the policy sets U and V, and for $\epsilon_1 = 0$).

Proof. We need to prove that under the set of assumptions ($\mathcal{A}$'1), ($\mathcal{A}$'3) and ($\mathcal{A}$'4) the hypothesis of theorem 2.1 are satisfied. ($\mathcal{A}$'3) and ($\mathcal{A}$'4) directly imply ($\mathcal{A}$3) and ($\mathcal{A}$4). We shall only prove that ($\mathcal{A}$1) holds as the proof of ($\mathcal{A}$2) is identical. Choose any ϵ_0 and introduce some sequence of functions :

$$\pi_n^1 : U_n \longrightarrow U, \text{ such that } d(u_n, \pi_n^1(u_n)) < \inf_{u \in U} d(u_n, u) + \epsilon_0$$

and

$$\sigma_n^2 : V \longrightarrow V_n, \text{ such that } d(\sigma_n^2(v), v) < \inf_{v_n \in V_n} d(v_n, v) + \epsilon_0.$$

By ($\mathcal{A}$' 1), for all ϵ_2 there exists $N_1 = N_1(\epsilon_2)$ such that for all $n > N_1$,

$$\max\left(\sup_{u_n \in U_n} \inf_{u \in U} d(u_n, u) \, ; \, \sup_{u \in U} \inf_{u_n \in U_n} d(u_n, u) \right) \leq \epsilon_2,$$

that is for all $u_n \in U_n$,

$$\inf_{u \in U} d(u_n, u) \leq \epsilon_2,$$

so for all $u_n \in U_n$,

$$d(u_n, \pi_n^1(u_n)) \leq \epsilon_0 + \epsilon_2. \tag{7.1}$$

Similarly, for all ϵ_3 there exists $N_3 = N_3(\epsilon_3)$ such that for all $n > N_3$ and for all $v \in V$,

$$d(v, \sigma_n^2(v)) \leq \epsilon_0 + \epsilon_3. \tag{7.2}$$

To prove ($\mathcal{A}1$) we shall show that for all ϵ, there exists $N = N(\epsilon)$ such that for all $n > N$, and for all $u_n \in U_n$, $v \in V$ we have :

$$S(u_n, \sigma_n^2(v)) - S(\pi_n^1(u_n), v) = S(u_n, \sigma_n^2(v)) - S(u_n, v) + S(u_n, v) - S(\pi_n^1(u_n), v) < \epsilon. \tag{7.3}$$

Since S is upper semicontinuous in v uniformly in u, there exists η_1 such that if $d(\sigma_n^2(v), v) \leq \eta_1$, $S(u_n, \sigma_n^2(v)) - S(u_n, v) \leq \epsilon/2$. Choose ϵ_0 and ϵ_3 such that $\epsilon_0 + \epsilon_3 = \eta_1$ in (7.2), there exists N_1 such that for all $n > N_1$,

$$S(u_n, \sigma_n^2(v)) - S(u_n, v) \leq \frac{\epsilon}{2}. \tag{7.4}$$

Similarly, it follows from (7.1) and since S is lower semicontinuous in u uniformly in v that there exists N_2 such that for all $n > N_2$

$$S(u_n, v) - S(\pi_n^1(u_n), v) \leq \frac{\epsilon}{2}. \tag{7.5}$$

(7.4) and (7.5) imply (7.3) by choosing $N = \sup(N_1, N_2)$, which concludes the proof. □

Remark 7.2. *In many applications (e.g. [11, 31]), the strategy sets are compact. Hence it suffices to require in ($\mathcal{A}$' 3) and ($\mathcal{A}$' 4) the semi-continuity properties; the uniform semi-continuity is then a consequence of the compactness of the strategy sets.*

As an application of the Theorem 7.1, we present the following continuous time differential pursuit evasion game by Bernhard and Shinar [11, 31]. We shall use the same notation as in [11, 31]. The game is governed by a differential equation

$$\frac{dx}{dt} = f(x, a, b),\ x \in \mathbf{I}, a \in \mathbf{A}, b \in \mathbf{B}$$

where $\mathbf{A}$ and $\mathbf{B}$ are compact subsets of $\mathrm{I\!R}^{m_1}$ and $\mathrm{I\!R}^{m_2}$ respectively, $\mathbf{I}$ is a domain of $\mathrm{I\!R}^{n}$. Some regularity and growth conditions on f ensure the existence of the solution of the differential equation over $(0, \infty)$ for every pair of measurable functions $a(\cdot)$ (and $b(\cdot)$) from $(0, \infty)$ to $\mathbf{A}$ (respectively, $\mathbf{B}$). The players have access to noisy partial information

$$y_a = h_a(x, w), \qquad y_b = h_b(x, w)$$

where w is a noise and h_a, h_b are globally Lipschitz over $\mathbf{I}$. They are restricted to using feedback strategies $(a(t) = \delta_1(y_a(t))$, $(b(t) = \delta_2(y_b(t)))$ Lipschitz continuous, then the set of strategies is compact in the topology of the uniform convergence. It is assumed that the noise model and the solution concept of the differential equations are such that the payoff P (the expected value of a continuous function of closest approach) is a continuous function of the strategies for the topology of uniform convergence.

If Ω_1 and Ω_2 are compact metric strategies spaces and Δ_1 and Δ_2 are closed subsets of Ω_1 and Ω_2 respectively and $U = M_1(\Delta_1)$ and $V = M_1(\Delta_2)$ are the sets of probability measures over Δ_1 and Δ_2, we know that there exist optimal mixed strategies that achieve the value

$$V(\Delta_1, \Delta_2) = \min_{u \in U} \max_{v \in V} J(u, v) = \int_{\Delta_1} \int_{\Delta_2} P(\delta_1, \delta_2) du(\delta_1) dv(\delta_2)$$

Bernhard and Shinar establish the convergence of the values of some approximating problems to the value of the original one.

We show that, in fact, all the statements concerning the convergence of policies in Theorem 2.1 also hold (with $\epsilon_1 = 0$). In [11], the continuity of $V(.,.)$ is proved, i.e. ($\mathcal{A}$' 3) and ($\mathcal{A}$' 4) are established. They present a finite approximation of this problem by considering finite subsets of Δ_i $i = 1, 2$ that converge to Δ_i in the Hausdorff topology (and thus, ($\mathcal{A}$' 1) holds).

REFERENCES

[1] E. Altman, "Denumerable constrained Markov decision problems and finite approximations", *Math. of Operations Research*, **19**, No. 1, pp. 169-191, 1994.

[2] ———, "Asymptotic properties of constrained Markov decision processes", *ZOR - Methods and Models in Operations Research*, **37**, Issue 2, pp. 151-170, 1993.

[3] ———, "Flow control using the theory of zero-sum Markov games," *IEEE Trans. Automatic Control*, pp. 814-818, 1994.

[4] ———, "A Markov game approach for optimal routing into a queueing network", INRIA report No. 2178, to appear in *Advances of dynamic games and applications.*

[5] ———, "Non zero-sum stochastic games in admission, service and routing control in queueing systems", to appear in *QUESTA*.

[6] ———, *Constrained Markov Decision Processes*, INRIA Report No. 2574, May 1995.

[7] E. Altman and A. Hordijk, "Zero-sum Markov games and worst-case optimal control of queueing systems", invited paper, *QUESTA* **21**, special issue on optimization of queueing systems, Ed. S. Stidham, 415-447, 1995.

[8] E. Altman, A. Hordijk and F. M. Spieksma, "Contraction conditions for average and α-discounted optimality in countable state Markov games with unbounded rewards", to appear in *MOR*, 1994.

[9] E. Altman and G. Koole, "Stochastic scheduling games with Markov decision arrival processes", *Journal Computers and Mathematics with Appl.*, third special issue of "Differential Games", pp. 141-148, 1993.

[10] H. Attouch and J. B. Wets, "A convergence theory for saddle functions", *Transaction of the American Mathematical Society*, **280**, No. 1, 1983.

[11] P. Bernhard and J. Shinar, " On finite approximation of a game solution with mixed strategies", *Appl. Math. Lett.* **3** No. 1, pp 1-4, 1990.

[12] V. Borkar and M. K. Ghosh, "Denumerable state stochastic games with limiting average payoff", *JOTA*, pp. 539-560, 1993.

[13] R. Cavazos-Cadena, "Finite-state approximations for denumerable state discounted Markov decision processes", *J. Applied Mathematics and Optimization* **14** pp. 27-47, 1986.

[14] K. L. Chung, *Markov chains with stationary transition probabilities*, 2nd edition, Springer Verlag, New York, 1967.

[15] H. A. M. Couwenbergh, "Stochastic Games with metric state space", *Int. J. of Game Theory*, **9**, issue 1, pp. 25-36, 1980.

[16] R. Dekker and A. Hordijk, "Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards", *Mathematics of Operations Research*, **13**, pp. 395-421, 1988.

[17] M. C. Delfour and J. Morgan, "Derivative of minmax and saddle points with respect to a parameter", Preprint no. 29, Departimento di Matematica e Applicazioni, Università degli Studi di Napoli "Federico II", 1992.

[18] A. Federgruen, "On N-person stochastic games with denumerable state space", *Adv. Appl. Prob.* **10**, pp. 452-471, 1978.

[19] D. Gillette, "Stochastic games with zero stop probabilities", M. Dresher, A. W. Tucker, P. Wolfe, eds., Princeton University Press, Princeton, 1957, pp. 179-187.

[20] J. Guillerme, "Convergence of approximate saddle point", *J. Mathematics Analysis Applications*, **137**, pp. 297-311, 1989.

[21] O. Hernandez-Lerma, "Finite state approximations for denumerable multidimensional - state discounted Markov decision processes", *J. Mathematical Analysis and Applications* **113** pp. 382-389, 1986.

[22] ———*Adaptive Control of Markov Processes*, Springer Verlag, 1989.

[23] J. L. Kelley, *General topology, Springer Verlag*, 1955.

[24] H. W. Küenle, *Stochastiche Spiele und Entscheidungesmodelle*, Tebuner-Texte, Band 89, 1986.

[25] ———, "On Nash equilibrium solutions in nonzero-sum stochastic games with complete information", *Int. J. of Game Theory* **23**, pp. 303-324, 1994.

[26] M. B. Lignola and J. Morgan, "Existence and approximation for min-sup problems", *Proceedings of the International Conference of Operation Research 90*, Vienna ed. Feichtinger et al., Springer Verlag, pp. 157-164, 1992.

[27] J. Morgan and R. Raucci, "Continuity properties of ϵ-solutions for generalized parametric saddle point problems", Preprint no. 35, Departimento di Matematica e Applicazioni, Università degli Studi di Napoli "Federico II", 1994.

[28] A. S. Nowak, "On zero-sum stochastic games with general state space, I", *Probability and Mathematical Statistics* IV, No. 1, pp. 13-32, 1984.

[29] ———, "Existence of equilibrium stationary strategies in discounted noncooperative stochastic games with uncountable state space", *JOTA* **45**, No. 4, pp. 592-602, 1985.

[30] L. I. Sennott, "Zero-sum stochastic games with unbounded costs : discounted and average cost cases", *ZOR - Methods and Models in Operations Research*, **40**, pp. 145-162, 1994.

[31] J. Shinar and I. Forte, "On the optimal pure strategy sets for a missile guidance law synthesis", *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens, Greece, 1986.

[32] F. M. Spieksma, *Geometrically Ergodic Markov Chains and the Optimal Control of Queues*, Ph.D. thesis, 1990, Leiden University (available on request from the author).

[33] L. C. Thomas and D. Stengos, "Finite state approximation algorithms for average cost denumerable state Markov decision processes", *OR Spectrum*, **7**, pp. 27-37, 1985.

[34] M. Tidball and E. Altman, "Approximations in dynamic zero-sum games, I", *SIAM J. Control and Optimization*, **34**, No. 1, pp. 311-328, 1996.

[35] D. J. White, "Finite state approximations for denumerable state infinite horizon discounted Markov decision processes", *J. Mathematical Analysis and Applications* **74**, pp. 292-295, 1980.

[36] ———, "Finite state approximations for denumerable state infinite horizon discounted Markov decision processes with unbounded rewards", *J. Mathematical Analysis and Applications* **86**, pp. 292-306, 1982.

[37] W. Whitt, "Approximations of dynamic programs, I", *Mathematics of Operations Research*, Vol. 3 No. 3, pp. 231-243, 1978.

[38] ———, "Representation and approximation of noncooperative sequential games", *SIAM J. Control and Opt.*, Vol 18 No 1, pp. 33-43, 1980.