Approximations in dynamic zero-sum games, I

Mabel M. Tidball, Eitan Altman

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Approximations in Dynamic Zero-Sum Games, I

Mabel M. TIDBALL, Eitan ALTMAN

N° 2166
Janvier 1994

PROGRAMME 5

Traitement du signal,
automatique et
productique

Rapport
de recherche

1994

Approximations dans les jeux dynamiques à somme nulle, I

Mabel M. TIDBALL et Eitan ALTMAN
INRIA, Centre Sophia-Antipolis
2004 Route des Lucioles, B.P.93
06902 Sophia-Antipolis Cedex
France

Eté 1993

Résumé

Nous développons une méthode unifiée pour approcher un jeu de somme nulle par une suite de jeux approximants. Nous discutons de la convergence des fonctions valeurs, ainsi que de la convergence des stratégies optimales (ou quasi-optimales). De plus, basées sur des politiques optimales pour le jeu limite, nous construisons des politiques qui sont quasi-optimales pour les jeux approximants. Puis, nous appliquons la théorie générale à l'approximation de l'état dans les jeux stochastiques, à la convergence des jeux à horizon fini vers le jeu à horizon infini, à la stabilité en taux d'actualisation et en gain instentané.

APPROXIMATIONS IN DYNAMIC ZERO-SUM GAMES, I

Mabel M. TIDBALL and Eitan ALTMAN
INRIA, Centre Sophia-Antipolis
2004 Route des Lucioles, B.P.93
06902 Sophia-Antipolis Cedex
France

Summer 1993

Abstract

We develop a unifying approach for approximating a "limit" zero-sum game by a sequence of approximating games. We discuss both the convergence of the values and the convergence of optimal (or "almost" optimal) strategies. Moreover, based on optimal policies for the limit game, we construct policies which are almost optimal for the approximating games. We then apply the general framework to state approximations of stochastic games, to convergence of finite horizon problems to infinite horizon problems, to convergence in the discount factor and in the immediate reward.

AMS Subject Classification: 90D05, 93E05
Key words: Zero-sum games, Approximations, Stochastic games.

1 Introduction

In many cases, one encounters dynamic games for which time and space are continuous, and possibly unbounded. In general, numerical solution of such games involve discretization both in time and in space. In pursuit evasion games in particular (see Bardi et al. [6] and Pourtallier and Tidball [21]), and in differential games in general (see Pourtallier and Tolwinsky, [22], Tidball and González [24]), the time and space discretization often lead to dynamic programming that has a stochastic game interpretation. The numerical solution then typically requires a finite state approximation.

Approximations in dynamic games has therefore been an active area of research for several decades. Several schemes for discretization of time and space and for approximations have been developed for differential games [6, 7, 21, 22, 24]. In stochastic games, much attention was devoted to approximations of infinite horizon problems by (long) finite horizon ones, e.g. [14, 20, 18, 26]; discretization of the action and state space have further been considered by Whitt [30, 31]. Except for [30], all the above references consider approximation of the value function of the games.

The aim of this paper is to study in a systematic way approximations in games, not only of the values but also for the policies. We begin by developing a general framework for establishing the convergence of the upper and lower values of a sequence of games G_n, $n = 1, 2, ...$, to a value R (which we assume that exists) of a limit game G_∞. We are further interested in the following questions: (i) do (almost) optimal policies converge (in some sense)? (ii) Assume that u and v are (almost) optimal for some approximating game G_n (where n is large enough in some sense). Can we construct from these almost optimal policies for the limit game? (iii) Assume that u and v are (almost) optimal for the limit game. Can we use them to construct almost optimal policies for the approximating game G_n for n large enough?

Problem (ii) above arizes in the following situation. Suppose that two players use some approximating numerical schemes to obtain "good" policies, e.g. time discretization. Each player might be using a different discretization scheme. Yet, each player would like to ensure that regardless of the discretization scheme used by the other player, he or she can guarantee some value, which would be "almost" the value of the non-discretized game. In fact, since the real game that is played is the non-discretized one, the desired discretization should perform well even if the adversary uses an optimal policy for the non-discretized game.

Problem (iii), on the other hand, arizes in the opposite situation, and this serves as an additional motivation for studying approximations in games. There are many example of dynamic games where one can solve easier an infinite limiting game, where problems related to the boundaries are avoided. Indeed, examples are given in Section 8 of stochastic games with (large) finite state space for which the natural approach for constructing almost optimal policies is to solve a limit game with a countable state space.

After establishing the general theory for approximations in Section 2, we apply it to several approximation problems in discrete time stochastic games with discounted reward and denumerable state space. Applications to other dynamic games are the subject of future research. The basic model of the stochastic games is presented in Section 3. We then present three schemes for state approximation in Section 4 for the case of infinite horizon. This generalizes many results on the convergence of the optimal value in Markov Decision Processes (i.e. stochastic games with a single player), e.g. [11, 15, 16, 28, 29]. Other related work on finite state approximations in Markov Decision Processes are [1, 2, 25]. In Section 5 we extend the results on state approximations for infinite horizon to the case of finite horizon, by a transformation of the state space. In Section 6 we study the convergence of the finite horizon problem to the infinite horizon one, and we combine state approximations with approximation of the horizon. In Section 7 we study the stability of stochastic games in the discount factor and in the immediate reward. Applications of approximation methods developed in this paper are presented in Section 8, and some generalizations are finally discussed in Section 9.

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 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:

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

$G = (S, U, V) \stackrel{\text{def}}{=} (S_\infty, U_\infty, V_\infty)$ will be called the limit game. It will be assumed that it has a value $R \stackrel{\text{def}}{=} R_\infty$.

An example where G_n does not have a value but G does have it will be given in Subsection 6.3 for computing almost optimal stationary policies for stochastic games with long (but finite) horizon.

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

$$\inf_{v \in V_n} S_n(u^*, v) \geq \inf_{v \in V_n} S_n(u, v) - \epsilon \qquad \forall u \in U_n \tag{1}$$

which is equivalent to $\inf_{v \in V_n} S_n(u^*, v) \geq \underline{R}_n - \epsilon$. It is said to be strongly 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

$$\sup_{u \in U_n} S_n(u, v^*) \leq \sup_{u \in U_n} S_n(u, v) + \epsilon \qquad \forall v \in V_n \tag{2}$$

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

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

Note that strongly ϵ-optimality implies ϵ-optimality. If a game has a value $\underline{R}_n = \overline{R}_n$ then strongly 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 for the limit game, can we derive of it an "almost" (strongly) optimal policy for the nth approximating game, for all n large enough?

In most applications that we discuss in this paper $U_n = U$, $V_n = V$ do not depend on n. However, in several applications this is not the case, e.g. approximations in pursuit evasion games, see Bernhard and Shinar [7]. Another example is given for a state approximation scheme for solving stochastic games, see Subsection 4.3.

Theorem 2.1 *Assume that 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*
*(**A1**)* $\overline{\lim}_{n \to \infty}[S_n(u, \sigma_n^2(v)) - S(\pi_n^1(u), v)] \leq 0$ *uniformly in* $u \in U_n$ *for each* $v \in V$.
*(**A2**)* $\underline{\lim}_{n \to \infty}[S_n(\sigma_n^1(u), v) - S(u, \pi_n^2(v))] \geq 0$ *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 $\epsilon' > \epsilon$, *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* $\epsilon' > \epsilon$, *there exists* $N(\epsilon')$ *such that* $\sigma_n^1(u^*)$ *(resp.* $\sigma_n^2(v^*)$*) is strongly* ϵ'*-optimal for the nth approximating game, for all* $n \geq N(\epsilon')$.
(4) Suppose
*(**A3**)* $S(u, v)$ *is a lower semicontinuous function in* u,

(A4) $S(u,v)$ is an upper semicontinuous function in v.

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 ϵ-optimal for the limit game.

Remark 2.1 *Part (1) of Theorem 2.1 is a generalization of Lemma 1 in [14]*

Proof. (1) Choose $\epsilon > 0$. Let $u^* \in U$ and $v^* \in V$ be ϵ-optimal for the limit game S, and choose some sequence $\epsilon(n)$ such that $\lim_{n\to\infty} \epsilon(n) = 0$. Let $u_n \in U_n$ be an $\epsilon(n)$-best response to the policy $\sigma_n^2(v^*)$ in game G_n (i.e., $S_n(u_n, \sigma_n^2(v^*)) \geq S_n(u, \sigma_n^2(v^*)) - \epsilon(n)$ for all $u \in U_n$). Similarly, let $v_n \in V_n$ be an $\epsilon(n)$-best response to the policy $\sigma_n^1(u^*)$ in game G_n.

Choose some $\delta > 0$. By (A1), there exists N such that for all $n \geq N$ and $u \in U_n$, $S_n(u, \sigma_n^2(v^*)) - S(\pi_n^1(u), v^*) < \delta$. Then for all $n \geq N$

$$
\begin{aligned}
\overline{R}_n - R &= \inf_{v\in V_n} \sup_{u\in U_n} S_n(u_n, v_n) - \inf_{v\in V} \sup_{u\in U} S(u,v) \\
&\leq \sup_{u\in U_n} S_n(u, \sigma_n^2(v^*)) - \sup_{u\in U} S(u, v^*) + \epsilon \\
&\leq S_n(u_n, \sigma_n^2(v^*)) - S(\pi_n^1(u_n), v^*) + \epsilon + \epsilon(n) \\
&\leq \delta + \epsilon + \epsilon(n)
\end{aligned}
$$

hence $\overline{\lim}_{n\to\infty} \overline{R}_n \leq R + \delta + \epsilon$.

Similarly, by (A2), one shows that $R \leq \underline{\lim}_{n\to\infty} \underline{R}_n + \delta + \epsilon$. Since $\underline{R}_n \leq \overline{R}_n$, this implies that $\overline{\lim}_{n\to\infty} |R - \underline{R}_n| \leq \delta + \epsilon$ and $\overline{\lim}_{n\to\infty} |R - \overline{R}_n| \leq \delta + \epsilon$. The result follows since ϵ and δ can be chosen arbitrarily small.

(2) Fix some $\delta > 0$. By (A1), as u_n^* is $\epsilon(n)$-optimal for G_n, and by part (1), there exists $N(\epsilon, \delta)$ such that for $n > N(\epsilon, \delta)$ we have

$$
\forall v \in V: \quad S_n(u_n^*, \sigma_n^2(v)) - S(u_n^*, \pi_n^1(v)) < \delta, \qquad \epsilon(n) < \epsilon + \delta, \qquad |\underline{R}_n - R| < \delta \tag{3}
$$

and thus

$$
\inf_{v\in V} S(\pi_n^1(u_n^*), v) \geq \inf_{v\in V} S_n(u_n^*, \sigma_n^2(v)) - \delta \geq \inf_{v\in V_n} S_n(u_n^*, v) - \delta \geq \underline{R}_n - \delta - \epsilon(n) \geq R - 3\delta - \epsilon.
$$

So $\pi_n^1(u_n^*)$ is ϵ'-optimal for S with $\epsilon' = 3\delta + \epsilon$.

In the same way, by assumption (A2) and considering an $\epsilon(n)$-optimal policy v_n^* for G_n, we obtain that $\pi_n^2(v_n^*)$ is ϵ'-optimal for S for all large enough n. The proof follows from the fact that δ was chosen arbitrarily.

(3) Fix some $\delta > 0$. As u^* is an ϵ-optimal strategy in the limit game G and by (A1), for all n large enough, we have

$$\inf_{v\in V_n} S_n(\sigma_n^1(u^*), v) \geq \inf_{v\in V_n} S(u^*, \pi_n^2(v)) - \delta \geq \inf_{v\in V} S(u^*, v) - \delta \geq R - \delta - \epsilon \geq R_n - 2\delta - \epsilon$$

The proof for v^* is obtained in the same way.

(4) Let $\hat{v} \in V$ be such that $\inf_{v\in V} S(\bar{u}, v) \geq S(\bar{u}, \hat{v}) - \delta$. By (A1), (A3) and part (1) of the Theorem, for all $\delta > 0$, there exists $N(\delta)$ such that $n > N(\delta)$ implies $\epsilon(n) < \epsilon + \delta$, and

$$\begin{aligned}\inf_{v\in V} S(\bar{u}, v) \geq S(\bar{u}, \hat{v}) - \delta &\geq S(\pi_n^1(u_n^*), \hat{v}) - 2\delta \\ &\geq S_n(u_n^*, \sigma_n^2(\hat{v})) - 3\delta \\ &\geq \underline{R}_n - 3\delta - \epsilon(n) \\ &\geq R - 4\delta - \epsilon(n)\end{aligned}$$

and hence, $\bar{u}$ is ϵ'-optimal with $\epsilon' = \epsilon + 5\delta$. In the same way, by (A2) and (A4) we prove that $\bar{v}$ is ϵ'-optimal for S. The proof follows from the fact that δ was chosen arbitrarily. ∎

Remark 2.2 *(i) In the rest of the paper, whenever $U_n = U$ and $V_n = V$ do not depend on n, π_n and σ_n will be chosen as the indentity maps.*

(ii) It follows from the proof of part (1) in the above Theorem that if for all 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

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

3 Stochastic games: the model

We are going to use the results from the previous section to study approximations of zero sum stochastic games.

- Let **I** be a denumerable set of states
- $\mathbf{A}_i$ $(\mathbf{B}_i)$ a compact set of actions for players I (resp. II) at state i. Let $K = \{i, \mathbf{A}_i, \mathbf{B}_i\}_{i\in I}$.
- $r : K \to \mathbb{R}$ a bounded immediate reward function (the boundedness condition can be relaxed, see Section 9). Let $M \stackrel{\text{def}}{=} \sup_{i,a,b} |r(i, a, b)|$.

- $P(a,b) = [p(i,a,b,E)]_{i,E}$, $a \in \mathbf{A}_i$, $b \in \mathbf{B}_i$ is a (sub) probability transition (from state i to a set $E \subset \mathbf{I}$) when the players use actions a and b.
- β the discount factor satisfying $0 \leq \beta < 1$

We shall use the following standard assumption (see e.g. Nowak [19]):
$(\mathbf{M}_1)$: $r(i, \bullet, \bullet)$ and $p(i, \bullet, \bullet, E)$ are continuous in both actions for any $E \subset \mathbf{I}$.

The game is played in stages $t = 0, 1, 2, \ldots$. 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 $u = (u_0, u_1, \ldots)$ 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 $q = \{q_0, q_1, \ldots\}$ 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 $p(\bullet \mid j)^g$ over $\mathbf{A}_j$, so that $p^g_{\mathbf{A}_j|j} = 1$, 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 $g(\bullet \mid j) = p(\bullet \mid j)^g$ for stationary g. Let S^A be the set of stationary policies for player 1, and define similarly the stationary policies S^B 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

$$p(j,u,v,k) = \int_{\mathbf{A}_j} \int_{\mathbf{B}_j} p(j,a,b,k) u(da|j) v(db|j). \tag{4}$$

We are concerned in the following section with the infinite horizon discounted problem. It is known that under (M_1), optimal stationary policies exist for both players; i.e., if u and v are optimal policies for both players when both of them restrict to stationary policies, then each one of these policies is also optimal against an arbitrary policy of his/her opponent. We shall therefore restrict to stationary mixed policies, without loss of generality (see [12, 17]).

Next, we introduce a topology on the sets of stationary policies. For any set Γ, let $M(\Gamma)$ denote the set of probability measures on Γ endowed with the weak topology $\xi(\Gamma)$ (see [19]). The class of stationary policies for player 1 (and similarly for player 2) can be identified with the set

$\prod_{i\in\mathbf{I}} M(\mathbf{A}_i) \times M(\mathbf{B}_i)$; moreover it is compact with respect to the product topology $\prod_{i\in\mathbf{I}} \xi(\mathbf{A}_i) \times \xi(\mathbf{B}_i)$.

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. Define the β-**discounted game payoff**

$$S(i, u, v) = E_i^{u,v} \sum_{t=0}^{\infty} \beta^t r(I_t, A_t, B_t) \tag{5}$$

Let $R(i)$ denote the value of the stochastic game for initial state i. For stationary policies u and v, let the expected current payoff be defined by:

$$r(i, u, v) = \int_{\mathbf{A}_j} \int_{\mathbf{B}_j} r(j, a, b) u(da|i) v(db|i) \tag{6}$$

Consider the following (contracting) map:

$$(T_{u,v} f)(i) \overset{\text{def}}{=} r(i, u, v) + \beta \sum_{j\in\mathbf{I}} p(i, u, v, j) f(j) \tag{7}$$

Then $S(i, u, v)$ is known to be the unique solution of (7). The value $R(i)$ is the unique solution of

$$R(i) = val \left[r(i, a, b) + \beta \sum_{j\in\mathbf{I}} p(i, a, b, j) R(j) \right] \tag{8}$$

Moreover, any stationary policies u^* and v^* that chooses at any state j the mixed strategies that are optimal for the matrix game $\left[r(i, a, b) + \beta \sum_{j\in\mathbf{I}} p(i, a, b, j) R(j)\right]_{a,b}$, are known to be optimal for the stochastic game S (see [19] for these statements).

Remark 3.1 *$S(i, \bullet, \bullet) : S^A \times S^B \to \mathbb{R}$ are continuous for all states i. This follows from Corollary 2.2 in Borkar [10]. It will thus follow below that assumptions (A3) and (A4) hold.*

4 State approximations: infinite horizon case

We introduce below several approximating schemes. All of them involve some sequence $\mathbf{I}_n \subset \mathbf{I}$ of sets of states, which are naturally chosen to be increasing. We shall assume

(B1) $\quad \mathbf{I}_n \subset \mathbf{I}_{n+1}, \qquad \bigcup_n \mathbf{I}_n = \mathbf{I}$

The following property will imply conditions (A1)-(A2) in the various schemes that we consider below:

(B2) $\epsilon(r,n) = \sup_{i\in \mathbf{I}_r,a,b}\left\{\sum_{j\notin \mathbf{I}_n} p(i,a,b,j)\right\} \to 0$ as $n\to\infty$ $\forall r$.

Remark 4.1 *Condition (M_1) as well as the compactness of* $\mathbf{A}_i$ *and* $\mathbf{B}_i$ *imply (B2). Indeed, assume that (B2) does not hold. Then, there exists some $\alpha > 0$ such that for some i,*

$$\overline{\lim_{n\to\infty}} \max_{a,b}\left(\sum_{j\in\mathbf{I}} p(i,a,b,j)1\{j\notin \mathbf{I}_n\}\right) = \alpha \tag{9}$$

Let a_n and b_n be some actions achieving the max (the fact that the max is achieved follows from the compactness and continuity assumption (M_1)). Choose a subsequence $n(\ell), \ell = 1,2,\ldots$ along which the limsup is obtained and along which a_n and b_n converge to some actions a^* and b^*. Then $p(i,a_{n(\ell)},b_{n(\ell)},\bullet)$ converges (pointwize) to the probability $p(i,a^*,b^*,\bullet)$ as $\ell\to\infty$ (by (M_1)). But then it follows from a dominant convergence Theorem ([23] Ch. 11 Sec. 4) and from (B1) that

$$\lim_{\ell\to\infty}\sum_{j\in\mathbf{I}} p(i,a_{n(\ell)},b_{n(\ell)},j)1\{j\notin \mathbf{I}_{n(\ell)}\} = \sum_{j\in\mathbf{I}} p(i,a^*,b^*,j)\cdot 0 = 0$$

which contradicts (9). Hence (B2) is established.

For the case of a single player, (B2) was introduced as an assumption for several approximating schemes by Cavazos-Cadena [11]. Note, however, that in [11], (M_1) as well as the compactness of the action spaces are not assumed. In order to obtain conditions (A1) and (A2) for the approximating schemes below, (and hence obtain statements (1), (2) and (3) in Theorem 2.1) one could relax the compactness assumption as well as (M_1); in that case one would indeed need to impose (B2) as an assumption. The compactness and (M_1) (or other similar assumptions, such as (M_2) or (M_3) from [19]) are required however for establishing the continuity conditions (A3) and (A4) required for establishing statement (4) in Theorem 2.1).

Other typical assumptions that imply (B2) have often been used in the literature, see White [28] and Hernández-Lerma [15], as well as $(B3)$ introduced in Altman [2] which will be used occasionally below:

(B3) From any state k, only a finite set of states X_k can be reached.

In all approximations in this section, the approximating games G_n have a value, i.e. $R_n(i) = \underline{R}_n(i) = \overline{R}_n(i)$. Moreover, they will have a saddle point among the stationary policies.

4.1 Approximation Scheme I

We define:

$$\left(H^1_{u,v}f\right)(i) \stackrel{\text{def}}{=} \begin{cases} r(i,u,v)+\beta\sum_{j\in\mathbf{I}_n} p(i,u,v,j)f(j) & \text{if } i\in\mathbf{I}_n \\ 0 & \text{if } i\notin \mathbf{I}_n \end{cases} \tag{10}$$

For this approximating problem we define $S^1_n(i,u,v)$ to be the solution of

$$\left(H^1_{u,v}f\right)(i) = f(i), \qquad \forall i\in\mathbf{I} \tag{11}$$

S^1_n is thus the total discounted payoff (defined in (5)) for the stochastic game whose transition probabilities are $\bar{p}$ instead of p, where $\bar{p}(i,u,v,j) = p(i,u,v,j)$ if $\{i,j\in\mathbf{I}_n\}$, otherwise 0. The value of the game G^1_n is the unique solution of

$$R_n(i) = \begin{cases} val\left[r(i,a,b)+\beta\sum_{j\in\mathbf{I}_n} p(i,a,b,j)R_n(j)\right] & \text{if } i\in\mathbf{I}_n \\ 0 & \text{if } i\notin \mathbf{I}_n \end{cases} \tag{12}$$

Moreover, optimal stationary policies u^*_n and v^*_n for the game G^1_n are obtained by chosing at any state $j\in\mathbf{I}_n$ the mixed strategies that are optimal for the matrix game

$$\left[r(i,a,b)+\beta\sum_{j\in\mathbf{I}_n} p(i,a,b,j)R(j)\right]_{a,b}.$$

The proof of the following Theorem will enable us to evaluate the precision of the approximation. More precisely, it will enable us to get a bound on $|R_n(\bullet)-R(\bullet)|$ which will be uniform in $i\in J$ where J is an arbitrary fixed subset of $\mathbf{I}$.

Theorem 4.1 *All statements of Theorem 2.1 hold for approximating scheme I, where the reward S for the limit game G is defined in (5) and for approximating game G^1_n it is S^1_n.*

Proof. The proof uses an idea by Cavazos-Cadena [11]. We first need to introduce some definitions. Let $\epsilon>0$ we define:

$$g^0(\epsilon,r)=r, \qquad g^k(\epsilon,r)=g(\epsilon,g^{k-1}(\epsilon,r)), \qquad k=1,2,...$$

where

$$g(\epsilon,r)=\min\{m:\epsilon(r,m)\le\epsilon\}$$

and $\epsilon(r,m)$ is defined in property (B2). To understand the meaning of $g^s(\epsilon,r)$, we consider first $\epsilon=0$. Then $\mathbf{I}_{g^{s+1}(0,r)}$ is the set of neighbors of $\mathbf{I}_{g^s(0,r)}$ in the sense that states that are not contained in $\mathbf{I}_{g^{s+1}(0,r)}$ are not reachable from any state in $\mathbf{I}_{g^{s}(0,r)}$. For $\epsilon > 0$, $\mathbf{I}_{g^{s+1}(\epsilon,r)}$ is the set of "ϵ-neighbors" of $\mathbf{I}_{g^{s}(\epsilon,r)}$ in the sense that for any state i in $\mathbf{I}_{g^{s}(\epsilon,r)}$, the states that are not contained in $\mathbf{I}_{g^{s+1}(\epsilon,r)}$ are reachable from i with probability smaller than or equal to ϵ. Note that for any $r \geq 0$,

$$\sum_{j \notin \mathbf{I}_{g^{l+1}(\epsilon,r)}} p(i,u,v,j) \leq \epsilon, \qquad \forall i \in \mathbf{I}_{g^{l}(\epsilon,r)}, \, \forall l \geq 0, \, \forall u,v. \tag{13}$$

Note also that $g^{l}(\epsilon, r)$ need not be increasing in l.

Let $J \subset \mathbf{I}$, $\epsilon(J) = min\,\{m : J \subset \mathbf{I}_m\}$, and suppose $\epsilon(J) < +\infty$ (this is the case if J is chosen to be finite). We define

$$m_k(\epsilon, \epsilon(J)) = \max\left\{\epsilon(J), g(\epsilon,\epsilon(J)), \ldots, g^k(\epsilon,\epsilon(J))\right\} \;\; k = 0,1,2,\ldots \tag{14}$$

We show that assumptions (A1)-(A4) hold for $S_n^1(i,u,v)$ and $S(i,u,v)$ defined above. Below, ϵ and J are held fixed, so that for simplicity of notation we shall write g^l instead of $g^l(\epsilon,\epsilon(J))$. Let $n \geq m_k(\epsilon,\epsilon(J))$, then for all $i \in J$,

$$\begin{aligned} |S_n^1(i,u,v) - S(i,u,v)| \;\; &\leq \;\; \beta \sum_{j \in \mathbf{I}_{g^1}} p(i,u,v,j)\,|S_n^1(j,u,v) - S(j,u,v)| \\ &+ \;\; \beta \sum_{j \notin \mathbf{I}_{g^1}} p(i,u,v,j)\,|S_n^1(j,u,v) - S(j,u,v)| \end{aligned} \tag{15}$$

Note that for any state j, $|S_n^1(j,u,v)| \leq M/(1-\beta)$, and $|S(j,u,v)| \leq M/(1-\beta)$. Hence by (13), (15) and (B2) we obtain:

$$\begin{aligned} &|S_n^1(i,u,v) - S(i,u,v)| \\ &\leq \;\; \beta \sum_{j \in \mathbf{I}_{g^1}} p(i,u,v,j)\,|S_n^1(j,u,v) - S(j,u,v)| + \epsilon\frac{2M\beta}{1-\beta} \\ &\leq \;\; \beta \max_{j \in \mathbf{I}_{g^1}} |S_n^1(j,u,v) - S(j,u,v)| + \epsilon\frac{2M\beta}{1-\beta} \\ &\leq \;\; \beta \max_{j \in \mathbf{I}_{g^1}} \left\{ \beta \sum_{\ell \in \mathbf{I}_{g^2}} p(j,u,v,\ell)\,|S_n^1(\ell,u,v) - S(\ell,u,v)| \right. \\ &\qquad \left. + \beta \sum_{\ell \notin \mathbf{I}_{g^2}} p(j,u,v,\ell)\,|S_n^1(\ell,u,v) - S(\ell,u,v)| \right\} + \epsilon\frac{2M\beta}{1-\beta} \end{aligned}$$

$$\leq \beta^2 \max_{\ell \in \mathbf{I}_{g^2}} |S_n^1(\ell,u,v) - S(\ell,u,v)| + \epsilon\frac{2M\beta^2}{1-\beta} + \epsilon\frac{2M\beta}{1-\beta} \tag{16}$$

$$\leq \beta^k \max_{\ell \in \mathbf{I}_{g^k}} |S_n^1(\ell,u,v) - S(\ell,u,v)| + 2\epsilon\frac{M\beta}{1-\beta}\sum_{\ell=0}^{k-1}\beta^\ell$$

The first inequality follows by (13) since $n \geq m_k \geq g^1$ and since $i \in J \subset \mathbf{I}_{g^0}$. Similarly, (16) follows by (13) since $n \geq m_k \geq g^2$ and since $\ell \in \mathbf{I}_{g^1}$. So we have:

$$|S_n^1(j,u,v) - S(j,u,v)| \leq 2M\frac{\beta(1-\beta^k)\epsilon/(1-\beta) + \beta^k}{1-\beta} \tag{17}$$

Hence, (A1) and (A2) hold true, (where as (A3)-(A4) are established in Remark 3.1). ∎

Combining (17) with Remark 2.2 yields:

Corollary 4.1 *For any $i \in J$, if n is chosen such that $n \geq m_k(\epsilon, \epsilon(J))$, then*

$$|R_n(i) - R(i)| \leq 2M\frac{\beta(1-\beta^k)\epsilon/(1-\beta) + \beta^k}{1-\beta}. \tag{18}$$

In the previous Theorem, the sets $\{\mathbf{I}_n\}$ where given a priory. Next we consider a special choice of $\{\mathbf{I}_n\}$, that will be especially useful under assumption (B3), for a finite set J. This construction will enable us to express in a simple way the set $\mathbf{I}_n$ needed in (10) in order to approximate R by R_n with a given error. This is especially desirable when it is not easy to compute $m_k(\epsilon, \epsilon(J))$ (and thus Corollary 4.1 can not be used).

Let J be a given set (for which we would like to get a computable uniform bound on the error of the approximation), and set $Y(i) = \{j : p(i,u,v,j) > 0 \text{ for some } u, v\}$. Then we define $\mathbf{I}_n$ in the following way:

$$\mathbf{I}_0 = J, \qquad \mathbf{I}_{n+1} = \bigcup_{i \in \mathbf{I}_n} Y(i) \bigcup \mathbf{I}_n \tag{19}$$

Remark 4.2 *Note that if J is finite and if (B3) holds, then all sets $\mathbf{I}_n$ are finite. This construction might be especially useful if the number of states reachable from any given state is small. In that case $\mathbf{I}_n$ do not grow too quickly.*

We now consider $S_n(i,u,v)$ as the solution of (11) with $\mathbf{I}_n$ defined in (19). We have that the following theorem (the analogous of (4.1)) holds:

Theorem 4.2 *(i) Fix a state $i \in J$. Then all statements of Theorem 2.1 hold for approximating scheme I, where the reward S of limit game G is defined in (5) and the reward S_n for the approximating game G_n is the solution of (11) with $\mathbf{I}_n$ defined in (19).*
(ii) For any $i \in J$ and $n = 0, 1, \ldots, |R_n(i) - R(i)| \leq 2M\beta^n/(1 - \beta)$.

Proof. It suffices to prove that (A1) and (A2) are satisfied. Let $i \in J$, then:

$$\begin{aligned}
|S_n(i,u,v) - S(i,u,v)| &\leq \beta \sum_{j \in Y(i)} p(i,u,v,j)|S_n(j,u,v) - S(j,u,v)| \\
&\leq \beta \max_{j \in \mathbf{I}_1} |S_n(j,u,v) - S(j,u,v)| \\
&\leq \beta^2 \max_{j \in \mathbf{I}_1} \sum_{k \in Y(j)} p(j,u,v,k)|S_n(k,u,v) - S(k,u,v)| \\
&\cdot \\
&\cdot \\
&\cdot \\
&\leq \beta^n \max_{j \in \mathbf{I}_n} |S_n(j,u,v) - S(j,u,v)| \leq \frac{2M\beta^n}{1-\beta}
\end{aligned}$$

(ii) then follows from Remark 2.2. ▮

As suggested in Remark 4.2, the above method is useful especially when the approximating games have finite states (i.e. $\mathbf{I}_n$ are finite) and the typical number of states (neighbors) reachable from a state is not too high. If, however, the typical number of neighbors is high, then the sets $\mathbf{I}_n$ become large very rapidly, which suggests that obtaining good estimates of optimal value and policies might require an unexceptably high complexity of computations. We thus present an alternative more general way of constructing finite sets $\mathbf{I}_n$ (even when (B3) does not hold), which will result in a simple expression for $m_k(\epsilon, \epsilon(J))$, and will thus enable to make use of Corollary 4.1 to obtain uniform computable error bound for the approximation for any $i \in J$.

We define a parametrized family $\{\mathbf{I}_n(\epsilon)\}$, where ϵ is a positive real number. Define $\mathbf{I}_0(\epsilon) = J$. $\{\mathbf{I}_n(\epsilon)\}$ are then chosen to be an arbitrary sequence increasing to $\mathbf{I}$ that satisfies the following. If for some $l > 0$, say $l = \hat{l}$,

$$\sup_{a,b,i \in \mathbf{I}_l(\epsilon)} \sum_{j \notin \mathbf{I}_l(\epsilon)} p(i,a,b,j) \leq \epsilon$$

then $\mathbf{I}_n(\epsilon) = \mathbf{I}$ for all $n > \hat{l}$. Otherwize, $\mathbf{I}_{l+1}$ is chosen such that

$$\sup_{a,b,i \in \mathbf{I}_l(\epsilon)} \sum_{j \notin \mathbf{I}_{l+1}(\epsilon)} p(i,a,b,j) \leq \epsilon$$

It follows that $\epsilon(J) = 0$, $g^0 = 0$, and hence $g^k = k$ and $m_k(\epsilon, \epsilon(J)) = g^k = k$ for $k \leq \hat{l}$. (The above quantities were defined in the proof of Theorem 4.1). If J is finite then it follows from the same arguments as in Remark 4.1 that $\mathbf{I}_n(\epsilon)$ can be chosen to be finite for $n \leq \hat{l}$ and hence in particular $\mathbf{I}_{\hat{l}}(\epsilon)$, which is the truncated state space that should be used to perform Approximation Scheme I in order to obtain a precision as in Corollary 4.1.

In this setting, Theorem 4.2 (ii) becomes a special case of Corollary 4.1 with $\epsilon = 0$.

4.2 Approximation Scheme II

In the previous approximation scheme, the dynamics are seen to be a result of transition probabilities that need not sum to one, even if in the limit game they do sum to one. Indeed, (10) can be considered as a stochastic game where we set $p(i, u, v, j) = 0$ for $j \notin \mathbf{I}_n$. In many applications this may be undesirable, and one would like $p(i, u, v, \bullet)$ to remain a probability measure. This is especially the case when we want to learn about the optimal value and (almost) optimal policies for specific given stochastic games with large finite state space, by approximating them through an infinite state game. Indeed, there are cases where one can solve easier an infinite game, since some boundary problems are avoided. Examples are given in Section 8.

We assume that $\sum_{j \in \mathbf{I}} p(i, a, b, j) = 1$ for all $a \in \mathbf{A}_i, b \in \mathbf{B}_i$. We define the following sequence of games. We let $\mathbf{I}_n \subset \mathbf{I}$ be an increasing sequence of sets, converging to $\mathbf{I}$, as in the previous Section. Define

$$\left(H^2_{u,v} f\right)(i) = \begin{cases} r(i, u, v) + \beta \sum_{j \in \mathbf{I}_n} p^*(i, u, v, j) f(j) & \text{if } i \in \mathbf{I}_n \\ 0 & \text{if } i \notin \mathbf{I}_n \end{cases} \tag{20}$$

where:

$$p^*(i, u, v, j) = \begin{cases} p(i, u, v, j) + q_n(i, u, v, j) & \text{if } i \in \mathbf{I}_n,\ j \in \mathbf{I}_n \\ 0 & \text{if } i \notin \mathbf{I}_n \text{ or } j \notin \mathbf{I}_n \end{cases} \tag{21}$$

$q_n(i, u, v, \bullet)$ is some non negative measure satisfying $\sum_{j \in \mathbf{I}_n} p(i, u, v, j) + q_n(i, u, v, j) = 1$. Hence,

$$\sum_{j \in \mathbf{I}_n} q_n(i, u, v, j) = \sum_{j \notin \mathbf{I}_n} p(i, u, v, j) \tag{22}$$

We define S^2_n to be the solution of

$$\left(H^2_{u,v} f\right)(i) = f(i) \tag{23}$$

S^2_n is thus the total discounted payoff (defined in (5)) for the stochastic game whose transition probabilities are p^* instead of p.

All the results of Subsection 4.1 still hold. We demonstrate this with the proof of the analogue of Theorem 4.1. It suffices to show that (A1)-(A2) hold. With the same notation as in Subsection 4.1 we obtain:

$$\begin{aligned}
&|S_n^2(i,u,v) - S(i,u,v)| \\
&\quad \le \beta \sum_{j\in \mathbf{I}_{g1}} p(i,u,v,j)\,|S_n^2(j,u,v) - S(j,u,v)| + \beta \sum_{j\in \mathbf{I}_{g1}} q(i,u,v,j)\,|S_n^2(j,u,v) - S(j,u,v)| \\
&\qquad\qquad + \beta \sum_{j\notin \mathbf{I}_{g1}} p(i,u,v,j)\,|S_n^2(j,u,v) - S(j,u,v)|
\end{aligned}$$

and by (22)

$$\begin{aligned}
&|S_n^2(i,u,v) - S(i,u,v)| \\
&\quad \le \beta \sum_{j\in \mathbf{I}_{g1}} p(i,u,v,j)\,|S_n^2(j,u,v) - S(j,u,v)| + 2\beta \sum_{j\notin \mathbf{I}_{g1}} p(i,u,v,j)\,|S_n^2(j,u,v) - S(j,u,v)| \\
&\quad \le \beta \sum_{j\in \mathbf{I}_{g1}} p(i,u,v,j)\,|S_n^2(j,u,v) - S(j,u,v)| + 2\epsilon\frac{2M\beta}{1-\beta}
\end{aligned}$$

So continuing as in the proof of theorem 4.1 we have:

$$|S_n^2(j,u,v) - S(j,u,v)| \le 2M\frac{2\beta(1-\beta^k)\epsilon(N)/(1-\beta) + \beta^k}{1-\beta} \tag{24}$$

for $n \ge m_k(\epsilon, \epsilon(J))$.

4.3 Approximation Scheme III

The basic idea of the approximation scheme is to fix some stationary policies for both players, and use them in all states except for a subset $\mathbf{I}_n$. The problem is then of determining the optimal mixed strategies for both players in the remaining set of states $\mathbf{I}_n$. We are interested in studying the asymptotic behavior of this approach as $\mathbf{I}_n \to \mathbf{I}$. Similar approaches were used in a framework of Markov decision processes (e.g. [1]), where $\mathbf{I}_n$ were assumed finite. We first fix some arbitrary policies $\hat{u} \in U$, $\hat{v} \in V$. We shall now use the framework of Theorem 2.1. Define

$$U_n = \{u \in U : u(i) = \hat{u}(i),\ \forall i \notin \mathbf{I}_n\}, \qquad V_n = \{v \in V : v(i) = \hat{v}(i),\ \forall i \notin \mathbf{I}_n\}.$$

Fix some $i \in \mathbf{I}$. The limit game is defined as $S(u,v) = S(i,u,v)$, where $S(i,u,v)$ is given in (5). For any $u \in U_n, v \in V_n$, define $S_n(u,v) = S(u,v)$ We set π_n^1 and π_n^2 to be the identity mappings and

$$\sigma_n^1(u)(i) = \begin{cases} u(i) & \text{if } i \in \mathbf{I}_n, \\ \hat{u}(i) & \text{if } i \notin \mathbf{I}_n; \end{cases} \qquad \sigma_n^2(v)(i) = \begin{cases} v(i) & \text{if } i \in \mathbf{I}_n, \\ \hat{v}(i) & \text{if } i \notin \mathbf{I}_n. \end{cases}$$

Theorem 4.3 *Fix a state i. Then*

(i) All statements of Theorem 2.1 hold for approximating scheme III.

(ii) R_n is the unique fixed point of the equation

$$R_n(k) = \begin{cases} val\left[r(k,a,b) + \beta \sum_{j\in\mathbf{I}} p(k,a,b,j) R_n(j)\right] & k \in \mathbf{I}_n \\ r(k,\hat{u},\hat{v}) + \beta \sum_{j\in\mathbf{I}} p(k,\hat{u},\hat{v},j) R_n(j) & k \notin \mathbf{I}_n \end{cases} \tag{25}$$

(iii) Optimal stationary policies u_n and v_n for both players are obtained by using at any state $k \in \mathbf{I}_n$ mixed strategies that achieve the value in (25).

The proof of this theorem is similar to the proof of theorem 4.1.

5 State approximations for the case of finite horizon

Consider the model in Section 3 with, however, a finite horizon reward criterion instead of (5):

$$S^{[m]} = E_i^{u,v}[\sum_{t=0}^{m} \beta^t r(I_t, A_t, B_t)] \tag{26}$$

It is well known that there exist optimal policies for both players within the class of Markov policies. The value R of $S^{[m]}$ is obtained by the recursion:

$$\begin{aligned} R^{m+1} &\overset{\text{def}}{=} 0 & (27) \\ R^k(i) &\overset{\text{def}}{=} val\left[r(i,u,v) + \beta \sum_{j\in\mathbf{I}} p(i,u,v,j) R^{k+1}(j)\right], \qquad k = 0,...,m \\ R &\overset{\text{def}}{=} R^0 \end{aligned}$$

Define $U[m] = (U^0[m], U^1[m], ..., U^m[m])$, $V[m] = (V^0[m], V^1[m], ..., V^m[m])$ where $U^k[m], V^k[m]$ are the set of mixed strategies which are optimal for the matrix game

$$\left[r(i,a,b) + \beta \sum_{j\in\mathbf{I}} p(i,a,b,j) R^{k+1}(j)\right]_{a,b}, \quad k = 0,...,m. \tag{28}$$

for all $i \in \mathbf{I}$. Then, any Markov policies (u, v) such that $u_t \in U^t$, $v_t \in V^t$, $t = 0, ..., m$, are optimal for the stochastic game $S^{[m]}$.

In order to apply the results from Section 4 to the finite horizon case we make the following observation. The finite horizon model is equivalent to the following infinite horizon model with enlarged state space:

- $\hat{\mathbf{I}} = \mathbf{I} \times \{0, ..., m\}$;
- $\hat{\mathbf{A}}_{(i,k)} = \mathbf{A}_i$, $\hat{\mathbf{B}}_{(i,k)} = \mathbf{B}_i$;
- $\hat{r}((i,k), a, b) = r(i, a, b)$;
- $\hat{p}((i,k), a, b, (j,l)) = \begin{cases} p(i, a, b, j) & \text{if } k + 1 = l \leq m \\ 0 & \text{otherwize} \end{cases}$
- $\hat{\beta} = \beta$;

Define

$$\hat{S}(\hat{i}, \hat{u}, \hat{v}) = E_{\hat{i}}^{\hat{u}, \hat{v}} \sum_{t=0}^{\infty} \beta^t r(\hat{I}_t, \hat{A}_t, \hat{B}_t)$$

There is a one to one correspondence between stationary policies in the new model and Markov policies in the original one; if $\hat{u}, \hat{v}$ are stationary in the new model, then the corresponding Markov policies in the original model are given by

$$u_t(\bullet|x) = \hat{u}(\bullet|(x, t)), \qquad v_t(\bullet|x) = \hat{v}(\bullet|(x, t)). \tag{29}$$

and vice versa. Moreover, we have

$$S^{[m]}(u, v) = \hat{S}(\hat{u}, \hat{v})$$

Consequently, the state approximation schemes from the previous section also hold for the case of finite horizon model. The computation of the (approximating) values and (almost) optimal policies can be done by using the above infinite horizon model with enlarged state space, and then applying (29). $\hat{\mathbf{I}}_n$ may be chosen, for example, as $\hat{\mathbf{I}}_n = \mathbf{I}_n \times \{0, ..., m\}$.

6 Successive approximations

We study in this section several new aspects of successive approximations. The convergence of the value of successive approximation is already well known, [20, 18, 26]. By applying Theorem 2.1, we establish in the following subsection, the convergence of (almost) optimal policies. We then study the application of both state approximation and finite horizon approximation. Finally we discuss the restriction of games with finite horizon to stationary policies.

6.1 Convergence of policies for successive approximations

An interesting application of the results in the previous subsections, is the observation that successive approximations (or value iteration) can be viewed as a special case of state approximations. One can define game G_n such that $S_n = S^{[n]}$, where $S^{[n]}$ is given in (26), and consider S as defined in (5). Let u_n^* and v_n^* be a pair of optimal (or ϵ_n-optimal, where $\lim_{n\to\infty} \epsilon_n = 0$) Markov policies for G_n. Let u^* and v^* be any ϵ-optimal stationary (or Markov) policies for the infinite horizon game S. If we use for both G_n and G the equivalent infinite horizon model with the enlarged state space defined in Section 5, then u_n^*, v_n^*, u^* and v^* all have an equivalent representation as stationary policies. The problem becomes one of approximating the state space $\mathbf{I}' = (\mathbf{I} \times \mathbb{N})$ by the subsets $\mathbf{I}'_n = (\mathbf{I} \times \{0, 1, ..., n\})$. Using then Theorems 2.1 and 4.1, we conclude:

Theorem 6.1 *(i)* $\lim_{n\to\infty} R_n = R$.
(2) For any $\epsilon' > \epsilon$, there exists N such that u_n^* (resp. v_n^*) is ϵ'-optimal for the infinite horizon game, for all $n \geq N$.
(3) Let $\bar{u} \in U$ (resp. $\bar{v} \in V$) be a limit point of u_n^* (resp. v_n^*). Then $\bar{u}$ (resp. $\bar{v}$) is ϵ-optimal for the limit game.
(4) For all $\epsilon' > \epsilon$, there exists $N(\epsilon')$ such that u^* is ϵ'-optimal for the nth approximating game, for all $n \geq N(\epsilon')$.

6.2 Successive approximation and finite state approximation

We use the above approach to combine state approximations with finite horizon reward criterion. Such a combination may be especially useful for computational purposes, where $\mathbf{I}_n$ can be chosen to be finite. We can now compute R_n and (Markov) policies which are optimal for G_n, using approximating schemes introduced in the previous Sections, in order to approximate the optimal value R and an almost optimal strategy for the original limit game G. Let again $\mathbf{I}_n \subset \mathbf{I}$ be an increasing sequence of sets of states, converging to $\mathbf{I}$. One can repeat the construction of a model with enlarged state space (that includes, both the original state space and the time), so that, the state space for the nth game G_n is $\hat{\mathbf{I}}_n = (\mathbf{I}_n \times \{0, 1, ..., n\})$, and for the limit game G, it is $\hat{\mathbf{I}} = (\mathbf{I} \times \mathbb{N})$. This would establish the correctness of approximations based on value iteration for a problem with truncated state space. For example, if we adapt the first approach in Section 4, we get the approximating values R_n and Markov policies by performing the following iterations:

$$\begin{aligned} R_n^{n+1}(j) &\overset{\text{def}}{=} 0 \\ R_n^k(j) &\overset{\text{def}}{=} \begin{cases} val\left[r(i,a,b)+\beta\sum_{j\in \mathbf{I}_n} p(i,a,b,j)R^{k+1}(j)\right] & \text{if } i\in \mathbf{I}_n \\ 0 & \text{if } i\notin \mathbf{I}_n \end{cases}, \qquad k=0,...,n \\ R_n &\overset{\text{def}}{=} R_n^0 \end{aligned}$$

Define $U_n = (U_n^0, U_n^1, ..., U_n^n)$, $V_n[m] = (V_n^0, V_n^1, ..., V_n^n)$ where U_n^k, V_n^k are the set of mixed strategies which are optimal for the matrix game

$$val\left[r(i,a,b)+\beta\sum_{j\in \mathbf{I}_n} p(i,a,b,j)R_n^{k+1}(j)\right]_{a,b} \qquad k=0,...,m \tag{30}$$

for any $i \in \mathbf{I}_n$. Then, any Markov policies (u,v) such that $u_t \in U_n^t$, $v_t \in V_n^t$, $t=0,...,m$, are optimal for the stochastic game G_n.

6.3 Finite horizon and stationary policies

For simplicity of implementation, one may be interested to restrict to the class of stationary policies in a stochastic game, rather than use Markovian policies (or others). It is well known, however, that finite horizon games do not have a value within the class of stationary policies. However, it is immediate to see that the conditions (A1)-(A4) hold when restricting to stationary policies, and thus we conclude from Theorem 2.1 that the optimal stationary policies for both players converge (in the sense of Theorem 2.1 (3)) to the strongly-optimal policy of the infinite horizon game, as the horizon goes to infinity. Moreover, the lower and upper values converge to the value of the infinite horizon game.

7 Convergence of the discount factor and immediate reward

We establish in this section the robustness of values and optimal policies with respect to the discount factor and immediate reward. This may be of importance in case that these parameters are not known precisely. One can similarly establish robustness for random time-varying discount factor and immediate reward. We consider a horizon m which may be either infinite or finite.

We consider a sequence of stochastic games G_n, $n = 0, 1, 2, ...$ where the quantities defining each one of them are as in Section 3, except for the immediate reward and discount factor which are replaced by $\beta_n = \beta + \delta_n$, $r_n = r + \rho_n$, where δ_n and ρ_n converge to zero as $n \to \infty$, uniformly in the states and actions. Denote by $S_n(i,u,v)$ the reward for game G_n (as defined either in (5) or in (26)). Then

$$\begin{aligned} &|S(i,u,v) - S_n(i,u,v)| \\ &\quad \le \; E_i^{u,v} \sum_{t=0}^{m} \Big(\beta^t |\rho_n(I_t, A_t, B_t, t)| + |[\beta^t - (\beta+\delta_n)^t]| M + |(\beta+\delta_n)^t \rho_n(I_t, A_t, B_t)| \Big) \end{aligned}$$

and we obtain convergence to zero uniformly over all Markovian policies u and v (to which we may restrict, without loss of generality, as in Section 5). This implies conditions (A1) and (A2) and hence, by Remark 3.1, we see that all statements of Theorem 2.1 hold. This establishes the continuity of the value of the stochastic game as a function of the discount factor β in the open interval $\beta \in (0,1)$, and as a function of the immediate reward. Moreover, it establishes the convergence of (almost) optimal policies (in the sense of Theorem 2.1).

A specially interesting case is the asymptotics of stochastic games as $\beta \to 1$. We restrict for simplicity to the case of finite state and action spaces. The asymptotic behavior of the value of the game was studied by Bewley and Kohlberg [8]. In fact they establish the convergence of $(1-\beta)R_\beta(i)$ to the value $R_{average}$ of the expected long run time-average game (where $R_\beta(i)$ is the value of the game with discount factor β and initial state i).

When trying to apply the approximating theorem 2.1 to the limit as $\beta \to 1$, we are faced with the following problems:
(i) The limit game does not have a value among the stationary (nor even the Markov) policies (see the "big match" by Blackwell and Ferguson [9]).
(ii) The value of the limit game (with the expected average reward) is in general not continuous in the policies (and thus assumptions (A3) and (A4) do not hold in general). This is the case even for a single controller, for which it is known that the value may exhibit discontinuity in the parameters (see Gaitsgory and Pervozvanskii [13] p. 407).

However, both problems are avoided in case that we restrict to either games with perfect information or to irreducible games (see Gillette [14]). Games with perfect information (resp. irreducible games) have a saddle point within the class of stationary deterministic policies (stationary mixed policies, respectively), and (A1) and (A2) hold, see [14]. Within these classes of policies, (A3) and (A4) also hold; indeed, for the perfect information case this follows from the fact that there is only a finite number of stationary deterministic policies. For the irreducible case, this follows e.g. from [2].

8 Applications

We present in this Section a few problems that motivated our research on approximations in stochastic games. As mentioned in the introduction, many discretization schemes of differential games yield dynamic programming that can be interpreted as representing some stochastic game. In some pursuit evasion games, such as the game of the two cars [22], an additional finite-state approximation is then required. The calculations in [22] was done following scheme I (introduced in Subsection 4.1). The state transition in the discretized model satisfied property (B3), and, in fact, each state had at most four neighbors (i.e. four states reachable in one transition). This feature motivates the use of Theorem 4.2 for such applications, which not only establishes the convergence, but also gives the rate of convergence (or, more precisely, enables to compute n for obtaining any required precision).

Another application of the theory we developed in previous sections are stochastic games appearing in queueing systems. Such problems may serve as model for situations of conflicts between users in telecommunication systems, or for worst-case control situation in presence of some unknown disturbance (in production systems, or again in telecommunications applications). An interesting feature in the control of queueing networks is that often, infinite queues are easier to handle than finite queues, as some boundary problems are avoided. Moreover, the optimal policies for infinite horizon problems, being stationary, are easier to implement than those for finite horizon problems. In real applications, however, queues are always finite; moreover, one is often interested in finite horizon problems (e.g., controlling manufacturing during working hours, etc). Our results may thus be applied to obtain almost optimal policies for these cases. Here are some examples:
(i) Altman considered in [3] a stochastic game with an infinite state space in order to solve a flow control problem with an infinite buffer. The solution of the problem with a finite buffer [4] seems more involved, and was only obtained under an important restriction on the actions of the flow controller (namely, it had to contain an action that corresponds to rejection of arriving customers).
(ii) Altman and Koole [5] solved a game where one or more servers have to be assigned to customers of different classes. In a telecommunication context, the different classes may represent different traffic types, such as voice, video and data, and the servers may represent a channel, through which the traffic has to be transmitted. A controller has to decide a customer of which class will be served next (which traffic will have access to the channel). The input traffic was assumed to be controlled as well, e.g. it may have been the output of some dynamic routing mechanism or dynamic flow control. The problem was posed as a zero-sum stochastic game between the service controller and "nature" which represented the unknown input control mechanism. Simple structural results were obtained for the case of infinite queues. In the case of finite queues, the structure of optimal policies

is unknown, even in the case of uncontrolled input. By applying our second state approximation scheme, it follows that the policies obtained for the problem of infinite queues are almost optimal for the case of finite queues which are large enough.

9 Further Generalizations

Although we considered in this paper bounded reward, it is well known that different sets of conditions exist for which problems with unbounded reward can be transformed into ones with bounded reward. Such transformations have been used in the past for finite state approximations of Markov Decision Processes, see White [29]. The generalization of such conditions to games are straightforward (see e.g. Wessels [27]).

There are many other useful directions where the general approximation Theorems are applicable, on which we continue our investigation. Among these are
(i) Differential games, in which a standard problem is to discretize both space and time. Several works have been done in this direction, see [6, 7, 21, 22, 24], where the convergence (and rate of convergence, see [21]) of the values of the approximating games have been established. However, little is known about the convergence of policies. Theorem 2.1 seems to be a suitable tool for approaching these issues.
(ii) Discretization of stochastic games with general state and action spaces. Some results were obtained in the case of a single controller, see Sec. 6 of Hernandez-Lerma and references therein. Further results for stochastic games on the convergence of the value and some results on convergence of policies were obtained by Whitt [30], and then generalized to N person games in [31].

Acknowledgements The authors wish to thank Odile Pourtallier and Pierre Bernhard for fruitful and illuminating discussions.

References

[1] E. Altman, "Denumerable Constrained Markov Decision Problems and Finite Approximations", to appear in *Math. of Operations Research.*

[2] E. Altman, "Asymptotic Properties of Constrained Markov Decision Processes", *Zeitschrift für Operations Research*, Vol. 37, Issue 2, pp. 151-170, 1993.

[3] E. Altman, "Flow control using the theory of zero-sum Markov games," *Proceedings of the 31st IEEE Conference on Decision and Control,* Tucson, Arizona, pp. 1632-1637, December 1992, to appear in *IEEE-AC.*

[4] E. Altman, "Monotonicity of optimal policies in a zero sum game: a flow control model", to appear in *Annals of dynamic games*, Vol. 1, 1993.

[5] 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.

[6] M. Bardi, M. Falcone and P. Soravia, "Fully Discrete Schemes for the Value Function of Pursuit- evation Games", *Annals of dynamic games*, Vol. 1, 1993.

[7] 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.

[8] T. Bewley and E. Kohlberg, "The asymptotic Theory of Stochastic Games" *Math of Oper. Res.*, Vo; 1 No. 3, 1976.

[9] D. Blackwell and T. Ferguson, "The Big match", *Annals of Mathematical Statistics*, **39**, pp. 159-163, 1968.

[10] V. S. Borkar, "A convex analytic approach to Markov decision processes", *Probab. Th. Rel. Fields*, Vol. 78, pp. 583-602, 1988.

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

[12] A. Federgruen, "On N-Person Stochastic Games with denumerable state space", *Adv. Appl. Prob.* **10**, pp. 452-471, 1978.

[13] V. A. Gaitsgory and A. A. Pervozvanskii, "Perturbation Theory for Mathematical Programming Problems", *JOTA*, 389-410, 1986.

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

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

[16] O. Hernandez-Lerma, *Adaptive Control of Markov Processes*, Springer Verlag, 1989.

[17] A. Hordijk, O. Vrieze and G. Wandrooij, "Semi-Markov strategies in stochastic games", *Mathematical Centre Report BW 68/76*, 1976.

[18] L. S. Shapley, "Stochastic Games", *proccedings Nat. Acad. of Science USA*, **39**, pp. 1095-1100, 1953.

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

[20] A. S. Nowak, "Approximation Theorems for zero-sum nonstationary stochastic games" *Proc. of the American Math. Soc.*, vol 92, No 3, pp. 418-424, 1984.

[21] O. Pourtallier and M. Tidball, "A Discrete Scheme of Pursuit-Evasion Games". Work in preparation.

[22] O. Pourtallier and B. Tolwinsky, "Discretization of Isaacs Equation: A Convergence Result", Work in preparation.

[23] H. Royden, *Real Analysis*, New York: Macmillan, 1963.

[24] M. Tidball and R. L. V. González, "Zero Sum Differential Games with Stopping Times. Some Results about its Numerical Resolution", *Annals of dynamic games*, Vol. 1, 1993.

[25] 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.

[26] J. Van der Wal, "Successive Approximations for Average Reward Markov Games", *Int. J. of Game Theory*, Vol. 9, issue 1, pp. 13-24.

[27] J. Wessels, "Markov Games with unbounded rewards", *Dynamische Optimierung*, M. Schäl (editor) Bonner Mathematische Schriften, Nr. 98, Bonn (1977).

[28] 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.

[29] D. J. White, "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.

[30] W. Whitt, "Approximations of Dynamic Programs, I", *Mathematics of Operations Research*, Vol. 3 No. 3, pp. 231-243, 1978.

[31] W. Whitt, "Representation and Approximation of Noncooperative Sequential Games", *SIAM J. Control and Opt.*, Vol 18 No 1, pp. 33-43, 1980.

Unité de Recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Rocquencourt Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399

RR-2166