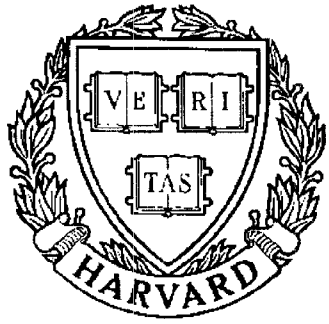
TECHNICAL RESEARCH REPORT

SYSTEMS RESEARCH CENTER



Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), the University of Maryland, Harvard University, and Industry

Research support for this report has been provided by NSFECS-83-51836

Guaranteed Performance Regions for Markov Models

by N. Shimkin and A. Shwartz

TR 91-8

Guaranteed performance regions for Markov models

by

Nahum Shimkin
Electrical Engineering
Technion—IIT
Haifa 32000, Israel

Adam Shwartz
Systems Research Center
University of Maryland
College Park, MD 20742

30th CDC

Keywords: Markov models, stochastic games, approachability, performance

Guaranteed performance regions for Markov models

Nahum Shimkin* and Adam Shwartz*†

Abstract

A user facing a multi-user resource-sharing system considers a vector of performance measures (e.g. response times to various tasks). Acceptable performance is defined through a set in the space of performance vectors. Can the user obtain a (time-average) performance vector which approaches this desired set? We consider the worst-case scenario, where other users may, for selfish reasons, try to exclude his vector from the desired set. For a Markovian model of the system, we give a sufficient condition for approachability (which is also necessary for convex sets), and construct appropriate policies. The mathematical formulation leads to an approachability theory for stochastic games.

I. Introduction.

Consider entering a multi-user resource-sharing system, for example a compter system. The objective is to guarantee acceptable service level for yourself, for example a fast response time of the terminal, adequate computation speed and reasonable delay at the printer queue. Naturally, a somewhat larger delay at the printer would be acceptable if we could gain in the response time. This tradeoff is modelled by defining a set in the performance space—in this example R^3—which we wish to approach.

We model the dynamics of the system as a controlled Markov chain, where each user exerts some control. We make no assumptions on the behavior of the other users. The question is: for a given set in the performance space, can we guarantee that the observed performance will (in the long run) fall into this set, even if the other users are doing their best to obstruct us (worst-case)? Or, can a group of malicious users exclude my performance from approaching this set?

Since we are considering a worst-case scenario, we may as well assume that we are facing a single "opponent" where a gain for us is a loss for our opponent. The performance is captured via a time-average vector (see §2 (iv) and (3.1)), so that we have the setting of a zero-sum game.

This framework can also be used to model a "worst case" analysis (in terms of a performance vector) of a system, where any uncertainties or time variations are modeled as control variables chosen by "nature". The resulting model is again a zero-sum stochastic game with vector payoff. We shall henceforth adopt the terminology of game theory, in order to formulate a precise question and exhibit the answer.

---

* Electrical Engineering, Technion—IIT, Haifa 32000, Israel.

† On leave; currently at the Systems Research Center, University of Maryland, College Park, MD 20742. Research supported in part by NSF Grant ECS-83-51836.

In a fundamental paper [BL3], Blackwell introduced the so called approachability–excludability theory for infinitely–repeated games with vector payoffs. Let us briefly review this theory. Consider a two–person, zero–sum, finite matrix game G_1, where the elements of the payoff matrix $A = (a_{ij})$ are *vectors* in the Euclidean m–space $\mathbb{R}^m$. Blackwell addressed the following question: If this game is repeated infinitely in time, with both players having perfect recall, can player 1 force the average payoff to asymptotically approach a preassigned subset Q of $\mathbb{R}^m$, no matter what the other player may do? Conversely, can player 2 exclude the average payoff from this set?

For an arbitrary set Q, a sufficient condition for approachability was given, based on the following idea. Player 1 monitors his current average payoff $\bar{\alpha}_t$. Suppose that for each value of $\bar{\alpha}$ outside Q, he has a strategy in G_1 which will push this payoff in the direction of Q irrespective of player 2's choice. Then by using such a strategy whenever $\bar{\alpha}_t \notin Q$, the average payoff vector will converge to Q.

A complete characterization of approachability was given for convex sets: a (closed) convex set is approachable unless player 2 has a strategy (in G_1) such that the single–stage payoff does not belong to Q for any strategy of player 1. It follows then that each convex set is either approachable by player 1, or excludable by player 2.

Further results on approachability for repeated games can be found in [HO11, HO12, SA17].

It should be noted here that approachability theory does not deal directly with a competitive–game situation, since no assumptions of rationality on the side of the opponent are made by either player. Rather, each player tries to secure his objective against any possible strategy of the other, so that actually we are dealing with a worst case, or min–max, analysis. However, the theory has important implications for pure game–theoretic models –such as repeated games of incomplete information (see e.g. [AV1, HA9, SO19]) and multicomponent attrition games ([BL5]). Other applications of the theory include compound (statistical) decision problems – see [BL4] and, e.g., [LU13] for a discussion and further references.

The objective of this paper is to extend Blackwell's results to average payoff stochastic games. In stochastic games, the game matrix which is played at each stage depends on some "state" variable, which may change stochastically from stage to stage (over a finite state–space). The players' choices at each stage control not only the payoffs, but also the game to be played at the next stage. These games, with the limit average (scalar) payoff, were introduced in [GI]. The question of existence of a value in the general class of stochastic game with limit average payoff remained open for many years, until answered in the affirmative in [ME14]. Generally, however, the players have only ε–optimal strategies, which are moreover non–stationary*.

For certain sub–classes of stochastic games, stronger results are available (see [PA15, RA16] for a survey). Here we shall be interested in the class of *irreducible* stochastic games, in which the state process forms an irreducible Markov chain for any pair of stationary strategies. For these games, it was established in [GI8, HK10] that both players have optimal strategies within the class of stationary strategies. The recurrence properties implied by the irreducibility assumption are central for the results of this paper.

The paper is organized as follows. In Section 2 the game model is formally defined. In Section 3 approachability and excludability are defined, and the main results are stated; these are analogous to the results of [BL3] for repeated games, as discussed above. The proofs are given in Sections 4 and 5. Finally some concluding remarks are contained in Section 6.

2. The Model

The stochastic game model we shall study is determined by the following objects:

(i) Two players, P1 and P2.

(ii) S – a finite state space, and $z \in S$, a given initial state.

(iii) I, J – finite sets of choices for P1, P2 respectively.

---

* See Section 2 for a definition of stationary policies.

(iv) $A : S \times I \times J \to \mathbf{R}^m$, a vector–valued payoff function.

(v) $\mathcal{P} = \{ P(\cdot \mid \cdot, i, j) : i \in I, j \in J \}$, a collection of transition matrices over S. Thus each $P(\cdot \mid \cdot, i, j) = (P(s' \mid s, i, j))$ is an $|S| \times |S|$ stochastic matrix.

The game starts at stage 0 in the initial state $s_o = z$. Then at each stage $t = 0, 1, 2, \ldots$ the following happens:

(vi) The players observe the current state $s_t \in S$.

(vii) P1 chooses an element $i_t \in I$, and P2 chooses an element $j_t \in J$.

(viii) Each player is told the choices i_t and j_t.

(ix) P1 gets from P2 the payoff vector $A(s_t, i_t, j_t)$.

(x) The game moves to a new state $s_{t+1} = s'$ with probability $p(s' \mid s_t, i_t, j_t)$.

We assume that both players have perfect recall, i.e., they do not forget what they knew at previous stages.

Let us describe the sets of strategies in this game. Let $H_t = (S \times I \times J)^t \times S$ be the set of all possible histories available to the players just before playing at stage t; that is, each $h_t \in H_t$ is a sequence of the form $h_t = (s_o, i_o, j_o, \ldots, s_{t-1}, i_{t-1}, j_{t-1}, s_t)$. A *behavior strategy* σ of P1 is a collection

$$\sigma = (\sigma_t)_{t=0}^{\infty}; \ \sigma_t : H_t \to \mathbf{P}(I) \ ,$$

$\mathbf{P}(I)$ being the set of all probability vectors over I. Thus at state t, P1's choice $i_t \in I$ is determined according to the probability vector $\sigma_t(h_t)$. A behavior strategy of P2 is similarly defined by:

$$\gamma = (\gamma_t)_{t=0}^{\infty}; \ \gamma_t : H_t \to \mathbf{P}(J) \ .$$

The set of all behavior strategies of P1 and P2 will be denoted by Σ and Γ, respectively. Note that since our game is of perfect recall, it follows from the Kuhn–Aumann theorem ([AU1]) that one

needs to consider only behavior strategies. That is, randomizations at each stage can be made independently.

A *stationary strategy* for P1 is a strategy $\sigma \in \Sigma$ in which all the σ_t's are determined by the same function $f : S \to \mathbb{P}(I)$ of the current state, i.e.,

$$\sigma_t(h_t) = f(s_t) \quad , \quad h_t \in H_t \quad .$$

The class of stationary strategies of P1 will be denoted by $\Sigma(st)$, and a typical element of $\Sigma(st)$ by f. The class of stationary strategies g for P2 is defined similarly and will be denoted by $\Gamma(st)$.

Given the strategies σ, γ and the initial state z, the above model induces a probability measure $P^z_{\sigma,\gamma}$ on the product space $\Omega = H_\infty = (S \times I \times J)^\infty$, endowed the product σ–algebra $\mathcal{F}_\infty$. The corresponding expectation operator will be denoted by $E^z_{\sigma,\gamma}(\cdot)$.

In this paper we shall make the following assumption regarding the transition structure $\mathcal{P}$:

Assumption A: The stochastic game determined by the above model is *irreducible*. That is, for each pair (f, g) of stationary strategies, the resulting transition matrix given by:

$$P_{f,g}(s'|s) = \sum_{i,j} p(s'|s,i,j) \cdot f(s)_i \, g(s)_j$$

determines an irreducible (hence positive recurrent) Markov chain.

Note that for the Assumption to hold it is sufficient that $P_{f,g}$ is irreducible for all *pure* stationary strategies.

3. Approachability: Definitions and Main Results

For the definition of approachability and excludability, we shall require a notion of a (uniform) rate of almost sure (a.s.) convergence. Let $(x_n, \ n \geq 0)$ be a sequence of real valued random variables over some probability space $(\Omega, \mathcal{F}, P)$. It is well–known that: $x_n \to 0$ P–a.s. is equivalent to either one of the following conditions:

(i) $\lim_{n \to \infty} P(\sup_{k \geq n} x_k > \varepsilon) = 0$ for every $\varepsilon > 0$,

or:

(i)' for every $\varepsilon > 0$ there exists an n_o such that

$$P(\sup_{k \geq n_o} x_k > \varepsilon) < \varepsilon$$

Let now $(P_\lambda, \ \lambda \in \Lambda)$ be a collection of probability measures on $(\Omega, \mathcal{F})$. We say that $x_n \to 0$ P_λ-*a.s., uniformly in* $\lambda \in \Lambda$, if:

(ii) $\lim_{n \to \infty} \sup_{\lambda \in \Lambda} P_\lambda(\sup_{k \geq n} x_k > \varepsilon) = 0$ for every $\varepsilon > 0$,

or equivalently:

(ii)' for every $\varepsilon > 0$, there exists an n_o such that

$$P_\lambda(\sup_{k \geq n_o} x_k > \varepsilon) < \varepsilon, \text{ for every } \lambda \in \Lambda.$$

Turning back to our model, define:

$\bar{\alpha}_t$ – the average payoff vector up to stage t:

$$\bar{\alpha}_t = \frac{1}{t} \sum_{k=0}^{t-1} \alpha(s_k, i_k, j_k), \quad t \geq 1 \tag{3.1}$$

$d(Q, \alpha)$ – the Euclidean distance of the point $\alpha \in \mathbf{R}^m$ from the set $Q \subset \mathbf{R}^m$.

Definition:

A set $Q \subset \mathbf{R}^m$ is *approachable* by P1, with $\sigma^* \in \Sigma$, if $d(Q, \bar{\alpha}_t) \to 0$ $P^z_{\sigma^*, \gamma}$ – a.s., uniformly in $\gamma \in \Gamma$. That is, for every $\varepsilon > 0$ there exists a t_o such that, for every $\gamma \in \Gamma$:

$$P^z_{\sigma^*, \gamma} \{ \sup_{t \geq t_o} d(Q, \bar{\alpha}_t) > \varepsilon \} < \varepsilon \tag{3.2}$$

A set Q is *excludable* by P2, with $\gamma^* \in \Gamma$, if for some $\delta > 0$:

$$d(Q_\delta^c, \bar{\alpha}_t) \to 0 \quad P_{\sigma,\gamma^*} \quad - \quad \text{a.s., uniformly in } \sigma \in \Sigma \ ,$$

where Q_δ^c is the complement of a δ–neighbourhood of Q.

□

It is evident that any set which is approachable by P1 cannot be excludable by P2, and vice–versa. However, as demonstrated in [BL3], generally there exist (non–convex) sets which are neither approachable nor excludable. Note that approachability and excludability are the same for a set Q and its closure, so we may assume that Q is closed.

The definition given above is the original one used by Blackwell in [BL3]. In some applications (e.g. [HA9], [SO19]), approachability and excludability are defined using (uniform) L^1–convergence in place of (uniform) a.s. – convergence. Since the variables $d(\bar{\alpha}_t, Q)$ are uniformly bounded, L^1 convergence follows from a.s. convergence so that the L^1 definition is weaker. It is then easy to see that all the results of this section remain valid under the L^1 definition.

An important aspect of the definition is the uniform rate of convergence over Γ. This requirement is essential if the infinite stage model is to be considered as an "idealization" of a very long, but finite, stochastic game model.

Let us now proceed to the formulation of the main results. For any pair of stationary policies $f \in \Sigma(\text{st})$ and $g \in \Gamma(\text{st})$, the following limit expected average payoff vector is well defined (see Proposition 4.1):

$$R(f,g) = \lim_{t \to \infty} E^z_{f,g}[\bar{\alpha}_t] \tag{3.3}$$

One can then define the following sets:

$$R(f, *) = \{ R(f,g) : \ g \in \Gamma(\text{st}) \} \ , \tag{3.4}$$

$$R(*, g) = \{ R(f,g) : \ f \in \Sigma(\text{st}) \} \ . \tag{3.5}$$

Namely, $R(f, *)$ is the set of payoffs which P2 can achieve by playing stationary strategies

against P1's stationary strategy f. We note that $R(f, *)$ is a convex polytope in $\mathbb{R}^m$, and in fact equals the convex hull of the finite set $\{ R(f,g):\ g$ is a *pure* stationary strategy of P2 $\}$, (see [DE7, p. 95]). An analogous relation holds for $R(*, g)$.

The following theorem gives a sufficient condition for approachability, as well as the form of the "approaching" strategy:

Theorem 1

Let $Q \subset \mathbb{R}^m$ be a closed set. Assume that for every $\alpha \notin Q$, there exists $f^*(\equiv f^*(\alpha)) \in \Sigma$ (st), a stationary policy of P1, and $y \in Q$ a closest point to α in Q, such that the hyperplane through y perpendicular to $(\alpha - y)$ separates α from $R(f^*, *)$. Then Q is approachable by P1, with the strategy $\sigma^* \in \Sigma$ given below:

Let $0 = T_0 < T_1 < T_2 < \ldots$ be the consecutive arrival times[1] to the initial state z. Then:

– at times $0 \le t < T_1$: play anything.

– at times $T_K \le t < T_{K+1}, K \ge 1$: if $\bar{\alpha}_{T_K} \notin Q$ play according to $f^*(\bar{\alpha}_{T_K})$.

Else if $\bar{\alpha}_{T_K} \in Q$, play anything.

□

The class of P1's strategies suggested by Theorem 1 may seem somewhat restricted, in that each of these strategies is adapted to the payoff history only at the recurrence times (T_n), while in–between these times a fixed stationary strategy is played at each stage. However, the next theorem shows that at least for convex sets this class of strategies is rich enough, in the sense that any convex set which is approachable can in fact be approached by the strategy σ^* of Theorem 1.

In defining the strategy σ^*, we chose the times (T_n) as the arrival times to the *initial state* z. This state was chosen for convenience only, and in fact (T_n) can be defined as the arrival times to any other state $s \in S$ without affecting the results of this section.

---

[1] If z is visited only a finite number N of times, let $T_N = \infty$.

Let us turn now to the case of convex sets. Theorem 1 can then be strengthened to give a complete characterization of approachability, as follows:

Theorem 2

Let $Q \subset \mathbb{R}^m$ be a closed convex set. Then:

(a) Q is approachable by P1 if and only if for every $g \in \Gamma$ (st), $Q \cap R(*, g) \neq \varnothing$.

(b) The condition in (a) is equivalent to the condition of Theorem 1.

(c) If Q is not approachable by P1, it is excludable by P2, with any stationary policy g^* which satisfies $Q \cap R(*, g^*) = \varnothing$.

Note that Theorem 2(c) implies that every convex set is either approachable by P1 or excludable by P2. As noted above, this is *not* the case for arbitrary sets, even in repeated games ([BL3]).

We now proceed to the proof of Theorems 1 and 2.

4. Preparatory Results

In this section some facts and results are collected, most of which are well–known. These will be useful in the proof of the main theorems. The first few are taken from the theory of Markov decision processes.

Proposition 4.1

Let $f \in \Sigma$(st) and $g \in \Gamma$(st) be stationary strategies. Define the stopping time $\tau = \min \{ t > 0 : s_t = z \}$, the first time of return to the initial state z. Then:

$$R(f, g) = \lim_{t \to \infty} E^z_{f,g}(\bar{\alpha}_t)$$

exists, is independent of the state z, and is equal to:

$$R(f,g) = \frac{E^z_{f,g}(\sum_{t=o}^{\tau-1} A(s_t, i_t j_t))}{E^z_{f,g}(\tau)} \tag{4.1}$$

Proof: Recall Assumption A. The claims are then standard results for irreducible, finite Markov chains, see e.g. [DE7, Appendix B].

□

Proposition 4.2

Let z, τ and f be as above. Then:

$$\left\{ \frac{E^z_{f,\gamma}(\sum_{t=o}^{\tau-1} A(s_t, i_t j_t))}{E^z_{f,\gamma}(\tau)} : \gamma \in \Gamma \right\} = R(f, *) \quad . \tag{4.2}$$

That is (recall (4.1) and the definition (3.4) of $R(f, *)$), for a fixed stationary strategy of P1, the set of τ–averaged payoffs which P2 can attain by playing *stationary* strategies does not change even if he is allowed to play *any* strategy $\gamma \in \Gamma$.

Proof: For a fixed $f \in \Sigma$ (st), the game model reduces to a Markov decision process with P2 the only decision maker. The proof then follows from [DE7, pp. 89–90].

□

Proposition 4.3

Let z, τ be as above. Then:

(a) $\bar{\tau} := \sup \{ E^z_{\sigma,\gamma}(\tau) : \sigma \in \Sigma, \gamma \in \Gamma \} < \infty$. (4.3)

(b) $\bar{\tau^2} := \sup \{ E^z_{\sigma,\gamma}(\tau^2) : \sigma \in \Sigma, \gamma \in \Gamma \} < \infty$. (4.4)

Proof: Consider the Markov decision process which results if both players join heads to form a single controller. Then $\Sigma \times \Gamma$ can be regarded as a subset of Π, the set of strategies in this decision process, and it suffices to show that the suprema in (a) and (b) are finite over Π. Note that for every pure stationary policy in Π, the resulting Markov chain is irreducible by Assumption A. (a) is then a standard result (e.g. [DE7, p. 50]), and (b) follows from [BO6, p. 74] using (a).

□

The following theorem is the basic result for irreducible stochastic games:

Theorem 4.4 ([GI8, HK10])

Consider the zero–sum stochastic game model described above, with scalar payoffs (m=1) and the limit expected average payoff. Then the game has a value, and the two players have stationary optimal strategies. In particular:

$$\min_{g \in \Gamma(st)} \max_{f \in \Sigma(st)} \lim_{t \to \infty} E^z_{f,g}(\bar{\alpha}_t) = \max_{f \in \Sigma(st)} \min_{g \in \Gamma(st)} \lim_{t \to \infty} E^z_{f,g}(\bar{\alpha}_t) \tag{4.5}$$

□

Finally, we shall need the following version of the SLLN for martingales:

Theorem 4.5

Let $M = (M_n, \mathcal{F}_n, n \geq 0)$ be a martingale over some probability space, with $M_o = 0$. Let $\Delta M_n = M_n - M_{n-1}$, $n \geq 1$, and assume that $E(\Delta M_n)^2 \leq c$, $n \geq 1$, for some $c < \infty$.

Then:

(a) $\frac{1}{n} M_n \to 0$ a.s.

(b) Moreover, the convergence rate depends only on c; that is, for every $\varepsilon > 0$ there is an $n_1 = n_1(\varepsilon, c)$, such that

$$P\left\{\sup_{k \geq n_1} \frac{M_k}{k} > \varepsilon\right\} < \varepsilon \quad .$$

Proof: M_n is square integrable, since $M_n^2 \le n^2 \sum_{k=1}^{n} (\Delta M_k)^2$. Also,

$$\sum_{k=1}^{\infty} \frac{E(\Delta M_k)^2}{k^2} \le \sum_{k=1}^{\infty} \frac{c}{n^2} < \infty \ , \tag{4.6}$$

and (a) then follows by [SH18, p. 471].

To prove (b), define:

$$m_n = \sum_{k=1}^{n} \frac{\Delta M_k}{k} \quad , \quad n \ge 1$$

$$\Delta m_n = m_n - m_{n-1} = \frac{\Delta M_n}{n} \tag{4.7}$$

where $m_o := 0$. Then:

$$\frac{M_n}{n} = \frac{1}{n} \sum_{k=1}^{n} \Delta M_k = \frac{1}{n} \sum_{k=1}^{n} k \cdot \Delta m_k \tag{4.8}$$

We will now show that (m_n) converges at a uniform rate, and then deduce (b) by (a uniform strengthening of) Kronecker's Lemma, (cf. [SH18]).

Fix $\varepsilon > 0$. Note that for every $n \ge 0$, $\{ (m_{n+k} - m_n)^2, \mathcal{F}_{n+k}, k \ge 0 \}$ is a non–negative submartingale. By Doob's inequality:

$$P\left\{ \sup_{k \ge 0} |m_{n+k} - m_n| \ge \frac{\varepsilon}{4} \right\} \le (\frac{\varepsilon}{4})^{-2} \lim_{k \to \infty} E\,(m_{n+k} - m_n)^2$$

$$= (\frac{\varepsilon}{4})^{-2} \sum_{k=n+1}^{\infty} \frac{E(\Delta M_k)^2}{k^2}$$

$$\le (\frac{\varepsilon}{4})^{-2} \sum_{k=n}^{\infty} \frac{c}{k^2} \quad . \tag{4.9}$$

Hence we can choose $n_o = n_o(\varepsilon, c)$ such that:

$$P\left\{\sup_{k \geq n_o} |m_k - m_{n_o}| > \frac{\varepsilon}{4}\right\} < \frac{\varepsilon}{2} \quad . \tag{4.10}$$

For any $n > n_o$:

$$\left|\frac{M_n}{n}\right| = \left|\frac{1}{n}\sum_{k=1}^{n} k\Delta m_k\right| = \left|\frac{1}{n}\sum_{k=0}^{n-1}(m_n - m_k)\right|$$

$$= \frac{1}{n}\left|M_{n_o} + \sum_{k=0}^{n-1}(m_n - m_{n_o}) + \sum_{k=n_o}^{n-1}(m_{n_o} - m_k)\right|$$

$$\leq \frac{1}{n}|M_{n_o}| + |m_n - m_{n_o}| + \frac{1}{n}\sum_{k=n_o}^{n-1}|m_k - m_{n_o}| \quad . \tag{4.11}$$

Also,

$$P\left[\frac{1}{n}|M_{n_o}| \geq \frac{\varepsilon}{2}\right] \leq \frac{2}{\varepsilon n} E\,|M_{n_o}|$$

$$\leq \frac{2}{\varepsilon n}\sum_{k=1}^{n_o} E\,|\Delta M_k| \leq \frac{2n_o}{\varepsilon n}(1+c) \quad . \tag{4.12}$$

Choose now $n_1 = n_1(\varepsilon, c) > n_o$ such that $\frac{2n_o}{\varepsilon n_1}(1+c) < \frac{\varepsilon}{2}$.

Let

$$\Omega_1 = \{\,\omega : \sup_{k \geq n_o} |m_k - m_{n_o}| < \frac{\varepsilon}{4} \text{ and } \frac{1}{n_1} M_{n_o} < \frac{\varepsilon}{2}\,\} \quad , \tag{4.13}$$

and note that $P(\Omega_1) > 1 - \varepsilon$ by (4.10) and (4.12). On that set, from (4.11) one gets:

$$\sup_{n \geq n_1}\left|\frac{M_n}{n}\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon \tag{4.14}$$

so that (b) is proved.

□

5. Proofs of the Main Results

Let us introduce first the required notation. Recall the definition of (T_n): $T_o = 0$, and

$$T_n = \inf \{ t > T_{n-1} : \ s_t = z \} , \quad n \geq 1 \ .$$

From Proposition 4.3 it follows that all the T_n's are a.s.–finite for any pair of strategies. We may therefore concentrate in the following on the subset of Ω for which the T_n's are finite. Define then:

$$\tau_{n+1} = T_{n+1} - T_n , \quad n \geq 0 \ .$$

Let $\mathcal{F}_t = \sigma(h_t) \subset \mathcal{F}_\infty$ be the finite algebra generated by the history up to time t. Obviously each T_n is a stopping time with respect to $(\mathcal{F}_t)$. Define $\mathcal{F}_{T_n}$ as usual: $\mathcal{F}_{T_n} = \{ A \subset \mathcal{F}_\infty : A \cap (T_n \leq t) \subset \mathcal{F}_n \}$. Note that Proposition 4.3 implies, for any pair of strategies:

$$E^z_{\sigma,\gamma} (\tau_{n+1} | \mathcal{F}_{T_n}) \leq \bar{\tau} \ , \tag{5.1}$$

$$E^z_{\sigma,\gamma} (\tau^2_{n+1} | \mathcal{F}_{T_n}) \leq \bar{\tau^2} \ , \quad P^z_{\sigma,\gamma} - \text{a.s.} \tag{5.2}$$

Since $\bar{\alpha}_t$ is in A_o, the convex hull of the points $\{ A(s,i,j) \}$, there exists a finite constant c_1 such that:

$$|\bar{\alpha}_t| \leq \sup_{\alpha \in A_o} |\alpha| \leq c_1 \tag{5.3}$$

and, furthermore, for any $Q \subset A_o$:

$$d(Q, \bar{\alpha}_t) \leq c_1, \quad t \geq 1 \ . \tag{5.4}$$

Finally, the following shorthand notation will be useful (for $Q \subset \mathbf{R}^m$ is a given set, $t \geq 1, n \geq 1$):

$$
\begin{aligned}
&y_t := \text{ a closest point in } Q \text{ to } \bar{\alpha}_t \\
&\delta_t := d(Q, \bar{\alpha}_t) = |\bar{\alpha}_t - y_t| \\
&\tilde{\delta}_n := \delta_{T_n} \\
&\tilde{\alpha}_n := \bar{\alpha}_{T_n} = \frac{1}{T_n} \sum_{t=0}^{T_n - 1} A(s_t, i_t, j_t) \\
&\tilde{y}_n := y_{T_n}, \text{ a closest point in } Q \text{ to } \tilde{\alpha}_n \\
&\tilde{\mathcal{F}}_n = \mathcal{F}_{T_n}
\end{aligned}
$$

$$
\tilde{\Delta}_{n+1} := \sum_{t=T_n}^{T_{n+1}-1} A(s_t, i_t, j_t) = T_{n+1}\tilde{\alpha}_{n+1} - T_n \tilde{\alpha}_n \tag{5.5}
$$

Proof of Theorem 1:

Suppose the hypotheses of the theorem are satisfied. Since approachability is the same for Q and $Q \cap A_o$, we may assume $Q \subset A_o$, so that (5.4) holds. Let P1 use the specified strategy σ^*, and P2 use any fixed strategy $\gamma \in \Gamma$. Let P and $E(\cdot)$ stand for $P^z_{\sigma^*,\gamma}$ and $E^z_{\sigma^*,\gamma}(\cdot)$, respectively.

The proof will be divided into three parts. In the first two, we establish the convergence of the sub–sequence $(\tilde{\delta}_n)$, which is a "sampling" of the distance sequence (δ_t) at times (T_n). The convergence of (δ_t) will then be deduced in the last part of the proof.

(i) *Bounds on* $(\tilde{\delta}_n)$: For $\tilde{\delta}_n > 0$, one has:

$$
\begin{aligned}
\tilde{\delta}^2_{n+1} &\leq |\tilde{\alpha}_{n+1} - \tilde{y}_n|^2 \\
&= |\tilde{\alpha}_n - \tilde{y}_n|^2 + 2 < \tilde{\alpha}_n - \tilde{y}_n, \tilde{\alpha}_{n+1} - \tilde{\alpha}_n > + |\tilde{\alpha}_{n+1} - \tilde{\alpha}_n|^2 ,
\end{aligned} \tag{5.6}
$$

where $|\cdot|$ and $< \cdot, \cdot >$ are the Euclidean norm and inner product, respectively.

We will now get some bounds on each term in the last expression. In fact, since it is easier to deal with the sequence $(T_n \tilde{\delta}^2_n)$ rather than $(\tilde{\delta}^2_n)$, these bounds will be constructed accordingly.

From (5.5) and (5.3):

$$|\tilde{\alpha}_{n+1} - \tilde{\alpha}_n| = \left| \frac{T_n \tilde{\alpha}_n + \tilde{\Delta}_{n+1}}{T_{n+1}} - \tilde{\alpha}_n \right| = \left| \frac{\tilde{\Delta}_{n+1} - \tau_{n+1}\tilde{\alpha}_n}{T_{n+1}} \right|$$

$$\leq 2c_1 \frac{\tau_{n+1}}{T_{n+1}} \tag{5.7}$$

So that, using (5.2) and ($T_n \geq n$):

$$E\,[T_{n+1}|\tilde{\alpha}_{n+1} - \tilde{\alpha}_n|^2 | \tilde{\mathcal{F}}_n] \leq (2c_1)^2 E\,[\frac{\tau_{n+1}^2}{T_{n+1}} | \tilde{\mathcal{F}}_n]$$

$$\leq \frac{(2c_1)^2 \overline{\tau^2}}{n+1} := \frac{c_2}{n+1} \quad . \tag{5.8}$$

Also,

$$< \tilde{\alpha}_n - \tilde{y}_n, \tilde{\alpha}_{n+1} - \tilde{\alpha}_n > = < \tilde{\alpha}_n - \tilde{y}_n, \frac{\tilde{\Delta}_{n+1} - \tau_{n+1}\tilde{\alpha}_n}{T_{n+1}} >$$

$$= \frac{1}{T_{n+1}} < \tilde{\alpha}_n - \tilde{y}_n, \tilde{\Delta}_{n+1} - \tau_{n+1}\tilde{y}_n > - \frac{1}{T_{n+1}} < \tilde{\alpha}_n - \tilde{y}_n, \tau_{n+1}\tilde{\alpha}_n - \tau_{n+1}\tilde{y}_n > \quad . \tag{5.9}$$

However, the first $< \cdot, \cdot >$ on the right hand side is negative 'on the average'. More precisely, from Proposition 4.2 and the definition of σ^* it follows that:

$$\frac{E\,(\tilde{\Delta}_{n+1} | \tilde{\mathcal{F}}_n)}{E\,(\tau_{n+1} | \tilde{\mathcal{F}}_n)} \in R\,(f^*(\tilde{\alpha}_n), *) \quad \text{a.e. on } \{\, \tilde{\delta}_n > 0 \,\} \quad , \tag{5.10}$$

while the definition of $f^*(\alpha)$ implies that:

$$< \tilde{\alpha}_n - \tilde{y}_n, r - \tilde{y}_n > < 0 \qquad \forall\, r \in R\,(f^*(\tilde{\alpha}_n), *) \quad . \tag{5.11}$$

Therefore:

$$E(<\tilde{\alpha}_n - \tilde{y}_n, \tilde{\Delta}_{n+1} - \tau_{n+1}\tilde{y}_n> | \tilde{\mathcal{F}}_n) = <\tilde{\alpha}_n - \tilde{y}_n, E(\tilde{\Delta}_{n+1}|\tilde{\mathcal{F}}_n) - \tilde{y}_n E(\tau_{n+1}|\tilde{\mathcal{F}}_n)>$$

$$= E(\tau_{n+1}|\tilde{\mathcal{F}}_n) \cdot <\tilde{\alpha}_n - \tilde{y}_n, \frac{E(\tilde{\Delta}_{n+1}|\tilde{\mathcal{F}}_n)}{E(\tau_{n+1}|\mathcal{F}_n)} - \tilde{y}_n> \; < 0 \;, \tag{5.12}$$

$$\text{a.e. on } \{\tilde{\delta}_n > 0\} \quad .$$

Using (5.8), (5.9) and (5.12) in (5.6) then gives:

$$E(T_{n+1}\tilde{\delta}_{n+1}^2|\tilde{\mathcal{F}}_n) \le E(T_{n+1}|\tilde{\mathcal{F}}_n)\tilde{\delta}_n^2 - E(\tau_{n+1}|\tilde{\mathcal{F}}_n)\tilde{\delta}_n^2 + \frac{c_2}{n+1}$$

$$= T_n\tilde{\delta}_n^2 + \frac{c_2}{n+1} \quad , \qquad \text{a.e. on } \{\tilde{\delta}_n > 0\} \quad . \tag{5.13}$$

For $\tilde{\delta}_n = 0$, we similarly have (take $\tilde{y}_n = \tilde{\alpha}_n$ in (5.6)):

$$E(T_{n+1}\tilde{\delta}_{n+1}^2|\tilde{\mathcal{F}}_n) \le E(T_{n+1}|\tilde{\alpha}_{n+1} - \tilde{\alpha}_n|^2|\tilde{\mathcal{F}}_n)$$

$$\le \frac{c_2}{n+1} \quad , \qquad \text{a.e. on } \{\delta_n = 0\} \quad . \tag{5.14}$$

Combining (5.13) and (5.14) then gives, for $n \ge 1$

$$E(T_{n+1}\tilde{\delta}_{n+1}^2|\tilde{\mathcal{F}}_n) \le T_n\tilde{\delta}_n^2 + \frac{c_2}{n+1} \quad \text{a.s.} \tag{5.15}$$

Another bound on $(T_n\tilde{\delta}_n^2)$ can be obtained as follows. Since:

$$\tilde{\delta}_{n+1} - \tilde{\delta}_n \le |\tilde{\alpha}_{n+1} - \tilde{y}_n| - |\tilde{\alpha}_n - \tilde{y}_n| \le |\tilde{\alpha}_{n+1} - \tilde{\alpha}_n| \;,$$

$$\tilde{\delta}_n - \tilde{\delta}_{n+1} \le |\tilde{\alpha}_n - \tilde{y}_{n+1}| - |\tilde{\alpha}_{n+1} - \tilde{y}_{n+1}| \le |\tilde{\alpha}_{n+1} - \tilde{\alpha}_n| \;,$$

it follows that:

$$|\tilde{\delta}_{n+1} - \tilde{\delta}_n| \le |\tilde{\alpha}_{n+1} - \tilde{\alpha}_n| \quad , \quad n \ge 1 \quad . \tag{5.16}$$

Now from (5.4), (5.7) and (5.2):

$$\begin{aligned} |T_{n+1}\tilde{\delta}_{n+1}^2 - T_n\tilde{\delta}_n^2| &\le T_{n+1}|\tilde{\delta}_{n+1}^2 - \tilde{\delta}_n^2| + \tau_{n+1}\tilde{\delta}_n^2 \\ &\le T_{n+1}|\tilde{\delta}_{n+1} + \tilde{\delta}_n| \cdot |\tilde{\delta}_{n+1} - \tilde{\delta}_n| + \tau_{n+1}\tilde{\delta}_n^2 \\ &\le T_{n+1}2c_1|\tilde{\alpha}_{n+1} - \tilde{\alpha}_n| + \tau_{n+1}c_1^2 \\ &\le 3c_1^2\tau_{n+1} \quad , \end{aligned} \tag{5.17}$$

$$E(|T_{n+1}\tilde{\delta}_{n+1}^2 - T_n\tilde{\delta}_n^2|^2 \,|\, \tilde{\mathcal{F}}_n) \le 9c_1^4\overline{\tau^2} := c_3 \quad \text{a.s.} \ , \quad n \ge 1 \tag{5.18a}$$

and similarly,

$$E(T_1\tilde{\delta}_1^2)^2 \le c_3 \quad . \tag{5.18b}$$

(ii) *Convergence of* $(\tilde{\delta}_n)$: It remains now to prove the following assertion: Let $(\tilde{\delta}_n, T_n)_{n \ge 1}$ be any sequence of $(\tilde{\mathcal{F}}_n$–adapted random variables which satisfy (5.15), (5.18) and $(T_n \ge n)$. Then $\tilde{\delta}_n \to 0$ a.s., at a rate depending on c_2 and c_3 only. Indeed, let $x_o := 0$, $T_o := 0$ and

$$x_n := T_n\tilde{\delta}_n^2 - \sum_{k=1}^{n} \frac{c_2}{k} \quad , \qquad n \ge 1 \tag{5.19}$$

Then obviously: $E(x_{n+1}|\tilde{\mathcal{F}}_n) \le x_n$ a.s. From the Doob decomposition for supermartingales ([SH18]), it follows that (x_n) is bounded above by the martingale:

$$\begin{aligned} M_n &:= x_n + \sum_{k=1}^{n} [x_{k-1} - E(x_k \,|\, \tilde{\mathcal{F}}_{k-1})] \\ &\ge x_n \qquad \text{a.s.} \ , \qquad n \ge 0 \quad . \end{aligned} \tag{5.20}$$

Now, for $n \geq 1$:

$$\begin{aligned} \Delta M_n &= x_n - E(x_n \mid \tilde{\mathcal{F}}_{n-1}) \\ &= T_n \tilde{\delta}_n^2 - T_{n-1}\tilde{\delta}_{n-1}^2 + E(T_n \tilde{\delta}_n^2 - T_{n-1}\tilde{\delta}_{n-1}^2 \mid \tilde{\mathcal{F}}_{n-1}) \end{aligned} \tag{5.21}$$

so that from (5.18): $E(\Delta M_n^2) \leq 4c_3$. We can then apply Theorem 4.5 to deduce that $\frac{1}{n} M_n \to 0$ a.s., at a rate which depends on c_3 only.

However,

$$0 \leq \tilde{\delta}_n^2 = \frac{1}{T_n}\left[x_n + \sum_{k=1}^{n} \frac{c_2}{k}\right] \leq \frac{1}{n} M_n + \frac{1}{n} \sum_{k=1}^{n} \frac{c_2}{k} \tag{5.22}$$

and the assertion follows upon noting that the (uniform) convergence of $(\tilde{\delta}_n)$ and $(\tilde{\delta}_n^2)$ are equivalent. (iii) *Convergence of* (δ_n): Let $\varepsilon > 0$ be given. We have just established that: there exists an $n_o = n_o(\varepsilon, c_2, c_3)$ such that:

$$P\left[\sup_{n \geq n_o} \tilde{\delta}_n > \frac{\varepsilon}{2}\right] < \frac{\varepsilon}{2} \quad . \tag{5.23}$$

Consider now the whole sequence (δ_t). For each $n > 0$ and $T_n \leq t < T_{n+1}$, we have similarly to (5.16) and (5.7):

$$|\delta_t - \delta_{T_n}| \leq |\bar{\alpha}_t - \bar{\alpha}_{T_n}| \leq 2c_1 \frac{t - T_n}{t} \leq 2c_1 \frac{\tau_{n+1}}{n} \quad , \tag{5.24}$$

so that:

$$\begin{aligned} P\left\{\sup_{k > n} \sup_{T_k \leq t < T_{k+1}} |\delta_t - \delta_{T_n}| > \frac{\varepsilon}{2}\right\} &\leq \sum_{k=n}^{\infty} P\left\{\sup_{T_k \leq t < T_{k+1}} |\delta_t - \delta_{T_n}| > \frac{\varepsilon}{2}\right\} \\ &\leq \left(\frac{2}{\varepsilon}\right)^2 (2c_1)^2 \overline{\tau^2} \sum_{k=n}^{\infty} \frac{1}{k^2} \leq \frac{\varepsilon}{4} \end{aligned} \tag{5.25}$$

where the last inequality holds for some $n = n_1(\varepsilon, c_1, \bar{\tau}^2)$ large enough, which will be chosen to satisfy $n_1 \geq n_o$.

Finally, we have:

$$P(T_{n_1} > t) \leq \frac{E(T_{n_1})}{t} \leq \frac{n_1 \bar{\tau}}{t} \quad , \tag{5.26}$$

so that $P(T_{n_1} > t_o) < \frac{\varepsilon}{4}$ for some $t_o = t_o(n_1, \bar{\tau})$. It then follows, in conjunction with (5.23) and (5.24) that:

$$P(\sup_{t \geq t_o} \delta_t > \varepsilon) < \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon \quad . \tag{5.27}$$

Noting that c_2, c_3, and therefore t_o depend only on the constants $(c_1, \bar{\tau}, \bar{\tau}^2)$, the theorem follows.

□

An upperbound on the rate of convergence of (δ_t) can in fact be computed by going over the details of the above proof. That is,

$$P(\sup_{t \geq t_o} \delta_t > \varepsilon_1) < \varepsilon_2$$

can be satisfied with $t_o = 0(\varepsilon_1^{-6}\varepsilon_2^{-3})$.

Proof of Theorem 2:

The proof follows from Theorem 1 and Theorem 4.4 along the lines of Theorem 2 of [BL3]. The argument will be repeated here for sake of completeness.

(a) Assume first that $Q \cap R(*, g^*) = \varnothing$ for some $g^* \in \Gamma$ (st). Since $R(*, g^*)$ is compact, $d(Q, R(*, g^*)) > \delta$ for some $\delta > 0$. Noting that $R(*, g^*)$ is convex, it follows easily by Theorem 1 that $R(*, g^*)$ is approachable by P2 with g^*. Hence Q is excludable by P2 with g^*.

Assume next that $Q \cap R(*, g^*) \neq \varnothing$ for all $g^* \in \Gamma$ (st). Approachability will follow by establishing the hypotheses of Theorem 1 for the set Q. Let $\alpha \notin Q$, and y be the closest point in Q to α. We have to show that the hyperplane $\{ q \in \mathbf{R}^m : < q-y, y-\alpha > = 0 \}$ separates α from $R(f^*, *)$ for some $f^* \in \Sigma$(st), i.e.

$$< q-y, y-\alpha > \geq 0, \quad q \in R(f^*, *) \quad . \tag{5.28}$$

Consider the game described by our model, with the scalar payoff function:

$$\tilde{a}(s,i,j) = < a(s,i,j)-y, y-\alpha > \quad . \tag{5.29}$$

Obviously, the average payoff in this game, when played with stationary strategies, is:

$$\tilde{R}(f,g) = < R(f,g)-y, y-\alpha > \quad . \tag{5.30}$$

From Theorem 4.4, P1 has an optimal stationary strategy f^* in this game. So for every $g \in \Gamma$(st):

$$\begin{aligned} \tilde{R}(f^*, g) &\geq \min_{g \in \Gamma(st)} \tilde{R}(f^*, g) = \\ &= \min_{g \in \Gamma(st)} \max_{f \in \Sigma(st)} < R(f,g)-y, y-\alpha > \geq 0 \end{aligned} \tag{5.31}$$

where the last inequality follows from our assumption upon noting that $< q-y, y-\alpha > \geq 0$ for all $q \in Q$. But (5.31) reads exactly (5.28) and the proof of (a) is complete.

(b) now follows since we have just proved that the condition in (a) implies the sufficient condition of Theorem 1. (c) follows from (a) and its proof.

□

6. Concluding Remarks

Approachability results were presented here for the class of *irreducible* stochastic games. The recurrence properties of the state process which are implied by the irreducibility assumption played a central role in the definition of the optimal strategy and in the analysis.

A slight generalization of the class of games considered is possible. Note that we have used directly only the recurrence properties of a single, fixed state. Thus the methods and results of this paper can be extended to the class of games considered in [ST20], where it is only assumed that there is a state z which will eventually be reached from any other state with positive probability, no matter what the strategies of the players might be.

For stochastic games which do not possess such recurrence properties, it seems that different methods are required to derive approachability results, if any. Further research in this direction is required.

References

[AU1] Aumann, R.J., Mixed and behaviour strategies in infinite extensive games. In: Advances in Game Theory, Ann. Math. Studies 52, M. Dresher et al. (eds.), Princeton, N.J., 1964, 627–650.

[AU2] Aumann, R.J. and M. Maschler, Game theoretic aspects of gradual disarmament. Chapter V, Report to the U.S. Arms Control and Disarmament Agency, Contract S.T.80, prepared by Mathematica, Inc., Princeton, N.J., 1966.

[BL3] Blackwell, D.; An analogue for the minimax theorem for vector payoffs. Pacific J. Math., 6, 1956, 1–8.

[BL4] Blackwell, D.; Controlled random walks. Proc. Internat. Congress Math. 3, 1954, 336–338.

[BL5] Blackwell, D.; On multi–component attrition games. Naval Res. Log. Quart. 1, 1954, 210–216.

[BO6] Borkar, V.S., Control of Markov chains with long–run average cost criterion, in Stochastic Differential Systems, Stochastic Control and Application, W. Fleming and P.L. Lions, eds., IMA Vol. 10, Springer–Verlag, 1988, 57–77.

[DE7] Derman, C.: Finite State Markovian Decision Processes, Academic Press, New York, 1970.

[GI8] Gillette, D.: Stochastic games with zero stop probabilities. in: Contributions to the Theory of Games, Vol. III, (Ann. Math. Studies, No. 39). Princeton, N.J., 1957, 179–187.

[HA9] Hart, S.: Nonzero–sum two–person repeated games with incomplete information. Math. of Oper. Res., 10, 1985, 117–153.

[HK10] Hoffman, A.D., and R.M. Karp: On non–terminating stochastic games, Management Science 12, 1966, 359–370.

[HO11] Hou, T.F.: Weak approachability in a two–person game, Ann. Math. Statist. 40, 1969, 789–813.

[HO12] Hou, T.F.: Approachability in a two–person game, Ann. Math. Stat. 42, 1971, 735–744.

[LU13] Luce, R.D. and H. Raiffa, Games and Decisions, Wiley, New York, 1957.

[ME14] Mertens, J.F. and A. Neymann, Stochastic games, International Journal of Game Theory 10, 1981, 53–56.

[PA15] Parthasarathy, T. and M. Stern, Markov games: a survey, in Differential Games and Control Theory, P.L.E. Roxin and R. Sternberg, eds., Marcel Dekker, 1977.

[RA16] Raghavan, T.E.S. and J.A. Filar, Algorithms for stochastic games – a survey. Preprint, June 1989.

[SA17] Sackrowitz, H., A note on approachability in a two–person game, Ann. Math. Stat., 43, 1972, 1017–1019.

[SH18] Shiryayev, A.N., Probability, Springer–Verlag, 1984.

[SO19] Sorin, J., An Introduction to Two–Person–Zero–Sum Repeated Games with Incomplete Information. IMSSS – Economics TR–312, Stanford University, memo.

[ST20] Stern, M.A., On Stochastic Games with Limiting Average Pay–off. Ph.D. dissertation, submitted to the University of Illinois, Circle Campus, Chicago, 1975.