Guaranteed Performance Regions in Markovian Systems with Competing Decision Makers

Nahum Shimkin, *Member, IEEE,* and Adam Shwartz, *Senior Member, IEEE*

Abstract—**The paper addresses the problem of (long-term) multiobjective control under dynamic uncertainty, using a game theoretic framework. A decision maker faces a dynamic system, which is also affected by other decision makers (these may stand for other controllers, system users, or dynamic disturbances). He / she considers a vector of time-averaged performance measures. Acceptable performance is defined through a set in the space of performance vectors. Can this decision maker guarantee a performance vector which asymptotically approaches this desired set? We consider the worst-case scenario, where other decision makers may try to exclude his / her vector from the desired set. For a controlled Markov model of the system, we give a sufficient condition for approachability, and construct appropriate control strategies. Under certain recurrence conditions, a complete characterization of approachability is then provided for convex sets. The mathematical formulation leads to a theory of approachability for "stochastic games with vector payoffs." A simple queueing example is analyzed to illustrate this approach.**

I. Introduction

CONSIDER a dynamic system which is influenced by several independent decision makers, for example a multiuser computer system. We take the view of a single decision maker, say DM1. This may be a system user, a central system supervisor, etc. His/her objective is to guarantee acceptable performance according to some individual performance measures, 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 modeled by defining a set in the performance space—in this example $\mathbb{R}^3$—which DM1 wishes to approach.

We model the dynamics of the system as a controlled Markov chain, where each decision maker exerts some control. We make no assumptions on the behavior or objectives of the other decision makers in the system. The question is: For a given set in the performance space, can DM1 guarantee that his/her time-averaged performance vector will converge to this set, even if the other decision makers are doing their best to obstruct him/her (worst case)? Or, can a group of malicious decision makers prevent (exclude) his/her performance vector from approaching this set?

Since we are considering a worst-case scenario, we may as well assume that DM1 is facing a single "opponent." This framework can also be used to model a worst-case analysis (in terms of a performance vector) of a single-controller system, where any dynamic uncertainties or time variations are modeled as control variables chosen by nature.

Similar questions were considered by Blackwell [4], in the context of the basic repeated matrix games model, where an "approachability–excludability" theory has been introduced for these games. A matrix game involves two players, where a payoff $m_{i,j}$ is generated whenever player 1 chooses action i while player 2 chooses action j. Thus, in a repeated matrix game the players face exactly the same situation at each decision epoch. Blackwell's model is therefore a special case of the controlled Markov model, where the state space is reduced to a single state. Let us briefly review the main ideas of Blackwell's results. Consider a two-person matrix game, where the elements of the payoff matrix $M = (m_{i,j})$ are *vectors* in $\mathbb{R}^q$, $q \geq 2$. The following problem was addressed: If the game is repeated infinitely in time, with both players observing and recalling the evolution of the game, can player 1 guarantee that the time-averaged payoff will asymptotically approach a given set (in $\mathbb{R}^q$), no matter what the other player may do? Conversely, can player 2 exclude the average payoff from this set?

For an arbitrary (closed) set B, a sufficient condition for approachability was given, based on the following idea. Player 1 monitors at each stage n the current average payoff. For each possible value $\overline{m}$ of the average payoff which is outside B, consider the hyperplane which passes through C, a closest point in B to $\overline{m}$, and which is perpendicular to the line segment $C - \overline{m}$. Suppose that player 1 has a strategy (i.e., a randomized action) in the matrix game such that, for every possible strategy of player two, the expected one-stage payoff is separated from $\overline{m}$ by this hyperplane. Then, by using such a strategy whenever the average payoff is outside B, the average

Manuscript received August 21, 1991; revised May 25, 1992. Paper recommended by Associate Editor, K. W. Ross. This work was supported in part by the National Science Foundation under Grant ECS-83-51836.

N. Shimkin was with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel, He is now with the Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN 55455.

A. Shwartz is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel.

IEEE Log Number 9204915.

payoff is constantly driven in the direction of B, and finally converges to it.

For convex sets, a complete solution was given. A set is obviously excludable by player 2 in the infinitely repeated game if it is excludable by him/her in the one-shot matrix game. In the convex case, this condition turns out to be both necessary and sufficient for excludability, and its negation is necessary and sufficient for approachability. Further results on approachability in repeated matrix games may be found in [14]–[16], [24], [29]. For some applications, mostly game theoretical, see [3], [5], [6], [12], [13], [17], and [27].

In this paper, the basic ideas of [4] are applied to obtain approachability results for a controlled Markov process with two decision makers and vector payoffs. In game-theoretical terms, this model may be referred to as a two-person stochastic game with vector payoffs. We consider the case of a countable state space, finite action spaces, and a (not necessarily bounded) vector payoff function; the formal setup is given in Section II. A basic assumption which underlies the approach of this paper is the existence of a fixed state, say state 0, for which certain uniform recurrence properties hold. It is then possible to obtain results which are similar to those described above for repeated matrix games, except that strategies in the one-shot matrix game are replaced by certain (stationary) substrategies which are employed between subsequent visits to state 0. Thus, a basic idea in the construction of approaching strategies is to use a fixed substrategy between visits to state 0, and modify this substrategy according to the current average payoff whenever state 0 is reached. (See [1], [2] for a similar approach in Markov decisions problems.)

The paper is organized as follows. The model is formally defined in Section II. Section III contains the main theoretical results, followed by a discussion of computational issues. The proof of the basic Theorem 3.1 is presented in Section IV. In Section V a simple queueing example is analyzed to illustrate the proposed approach, followed by some concluding remarks.

Notation: $|\cdot|$, $\langle\cdot,\cdot\rangle$ and $d(\cdot,\cdot)$ denote the Euclidean norm, inner product and metric in $\mathbb{R}^q$. U denotes the set of unit vectors in $\mathbb{R}^q$.

II. The Model

Consider a controlled Markov process with two independent decision makers, DM1 and DM2. The model is specified by the following objects: a countable state-space S, finite action spaces A_1 and A_2, a state transition law p, and an $\mathbb{R}^q$-valued payoff function r (where $q \geq 2$).

At each stage (time instant) $n = 0, 1, 2, \cdots$, the current state s is observed, and then DM1 chooses an action $a^1 \in A_1$, and DM2 chooses simultaneously and independently an action $a^2 \in A_2$. As a result, a payoff vector $r(s, a^1, a^2)$ is collected, and the next state s' is chosen according to the probability distribution $p(\cdot|s, a^1, a^2)$ on S. The state and action pair at stage n will be denoted by s_n and $a_n = (a_n^1, a_n^2)$, respectively. Let $r_n = r(s_n, a_n)$ be the payoff vector at stage n, and let

$$\bar{r}_n = \frac{1}{n} \sum_{m=0}^{n-1} r_m \tag{2.1}$$

denote the time-averaged payoff vector up to stage n.

Note that we have yet to specify the objectives of either decision maker. This will be done in the next section. For now, the payoff should just be considered as some vector which measures system performance.

A (randomized, history-dependent) strategy π_i for DMi ($i = 1, 2$) is a sequence

$$\pi_i = \{\pi_0^i, \pi_1^i, \cdots\}, \qquad \pi_n^i : H_n \to \mathscr{P}(A_i)$$

where $\mathscr{P}(A_i)$ is the set of probability vectors over A_i, and $H_n = S \times (A_1 \times A_2 \times S)^n$ is the set of possible "histories" up to stage n. Thus, given $h_n = (s_0; a_0, s_1; \cdots; a_{n-1}, s_n)$, the action a_n^i is chosen according to the probability vector $\pi_n^i(h_n)$. Let Π_i denote the class of all such strategies for DMi. A *stationary* strategy for DM1 is specified by a single function $f: S \to \mathscr{P}(A_1)$, so that $\pi_n^1(h_n) = f(s_n)$, $n \geq 0$. The class of stationary strategies for DM1 will be denoted by F, and that of DM2 (defined similarly) by G.

Given the strategy pair $\pi = (\pi_1, \pi_2)$ and an initial state $s_0 = s$, the above description induces a unique probability measure P_π^s and expectation operator E_π^s on the product space $S \times (S \times A_1 \times A_2)^\infty$. When s_0 and π are determined by the context, we just write P and E for the corresponding measure and expectation.

Some definitions from game theory will be required in the sequel. For every vector $u \in \mathbb{R}^q$ and initial state s, consider the case where at each stage, DM2 pays DM1 an amount which is specified by the *scalar* payoff function $r^u = \langle r, u\rangle$. If the objective of DM1 (respectively DM2) is to maximize (respectively minimize) the average expected payoff, then the model becomes a zero-sum stochastic game, which we will denote by $\Gamma_s(u)$. Stochastic games have been extensively studied; for a survey the reader is referred to [19]–[21]. The connection between our model (with vector payoff) and this family of zero-sum stochastic game will be clarified below. We say that $\Gamma_s(u)$ has a value $\operatorname{val}\Gamma_s(u)$ if

$$\operatorname{val}\Gamma_s(u) = \sup_{\pi_1} \inf_{\pi_2} \liminf_{n\to\infty} E_{\pi_1,\pi_2}^s(\langle \bar{r}_n, u\rangle) \tag{2.2}$$

$$= \inf_{\pi_2} \sup_{\pi_1} \limsup_{n\to\infty} E_{\pi_1,\pi_2}^s(\langle \bar{r}_n, u\rangle). \tag{2.3}$$

A strategy $\pi_1 \in \Pi_1$ [$\pi_2 \in \Pi_2$] is *optimal* in $\Gamma_s(u)$ if it satisfies the sup in (2.2) [the inf in (2.3), respectively].

The basic assumptions made in this paper will involve recurrence conditions for some fixed state, which we denote as state 0. Let τ denote the first passage time to state 0:

$$\tau = \inf\{n \geq 1 : s_n = 0\}.$$

A strategy $\pi_1 \in \Pi_1$ is said to be *stable* if there exist positive constants M_2 and R_2 such that:

$$E^0_{\pi_1,\pi_2}(\tau^2) \le M_2 \qquad \forall \pi_2 \in \Pi_2, \tag{2.4}$$

$$E^0_{\pi_1,\pi_2}\left(\sum_{n=0}^{\tau-1} |r_n|\right)^2 \le R_2 \qquad \forall \pi_2 \in \Pi_2. \tag{2.5}$$

Note that (2.5) is redundant in case that the payoff function r is bounded. A set $\Pi_1' \subset \Pi_1$ is *uniformly stable* if (2.4) and (2.5) are satisfied for every $\pi_1 \in \Pi_1'$ with the same constants M_2 and R_2. Stability of DM2's strategies is defined symmetrically.

We introduce now some conditions on the model. Reference to these conditions will be made explicitly when required.

C1) For every unit vector $u \in U$, the game $\Gamma_0(u)$ has a value, and DM1 has a stationary optimal strategy $f^*(u)$ in this game. Moreover, the set $\{f^*(u)\colon u \in U\}$ is uniformly stable.

C2) Condition C1) holds. Furthermore, for each $u \in U$ DM2 has an optimal strategy $g^*(u)$ in $\Gamma_0(u)$ which is stationary and stable.

Existence of stationary optimal strategies is stochastic games has been well studied and established under various conditions (cf. [19], [20]). Conditions which imply stability requirements similar to ours may be found, e.g., in [7]. A particular set of conditions which imply both the existence of stationary strategies as well as all the stability requirements encountered in the sequel is specified in the following lemma. (Compare with [23, ch. 6], where similar assumptions were used in the context of Markov decision processes.)

Lemma 2.1: Assume that:

i) The payoff function r is bounded.

ii) There exists a number M such that the mean first-passage time $E^s_{f,g}(\tau) \le M$ for every $s \in S$ and all stationary nonrandomized strategies $f \in F$, $g \in G$.

Then C1) and C2) are satisfied. Moreover, the entire strategy sets Π_1 and Π_2 are uniformly stable.

Proof: For each $u \in U$, i) implies that the payoff function $r^u = \langle r, u\rangle$ is bounded. Existence of optimal stationary strategies in stochastic games with bounded payoff functions under the recurrence condition ii) was established in [28] (and see also [9] for more general recurrence conditions which imply the same). It remains to establish the stability requirements, for which certain results on Markov decision processes will be utilized. Let Π denote the set of strategies which results when DM1 and DM2 are combined into a single controller, i.e., are allowed to correlate randomized choices at each stage. Let F_D denote the set of nonrandomized stationary strategies in F, and similarly for $G_D \subset G$. Note that $\Pi_1 \times \Pi_2$ may be considered a subset of Π, and that the set Π_D of stationary nonrandomized strategies in Π coincides with $F_D \times G_D$. By standard dynamic programming considerations, it follows from ii) that

$$E^s_\pi(\tau) \le M, \qquad \forall \pi \in \Pi,\ s \in S. \tag{2.6}$$

Indeed, for each $\alpha < 1$ let $J^s_{\pi,\alpha} := E^s_\pi(\Sigma_{k=0}^{\tau-1} \alpha^k)$, and note that $E^s_\pi(\tau) = \lim_{\alpha \to 1} J^s_{\pi,\alpha}$ by monotone convergence. Standard results for discounted cost criteria ([23]) imply that $J^s_{\pi,\alpha}$ is maximized by a stationary nonrandomized strategy in Π_D, so that

$$J^s_{\pi,\alpha} \le \max_{\pi \in \Pi_D} J^s_{\pi,\alpha} \le \sup_{\pi \in \Pi_D} E^s_\pi(\tau) \le M$$

where the last equality follows from ii), and (2.6) follows.

Now, it was established in [7, p. 74], that (2.6) implies $\sup_{\pi \in \Pi} E^0_\pi(\tau^2) < \infty$, so that Π_1, Π_2 (and therefore any of their subsets) are uniformly stable. □

Remark: Lemma 2.1 is most useful in the case of a *finite* state space. Conditions i) and ii) then reduce to the simple requirement that state 0 is recurrent under any pair f, g of stationary nonrandomized strategies (which are now finitely numbered). This requirement may often be verified by a simple inspection of the transition structure.

In addition to conditions C1)-C2), some additional stability conditions for specific sets of strategies will be encountered in the following. It should be emphasized that all these conditions are satisfied automatically under the conditions of Lemma 2.1.

III. Approachability: Definitions and Results

Let us define first the concept of *uniform* almost sure (a.s.) convergence which will be used here. Let $\{X_n, n \ge 0\}$ be a sequence of random variables over some measurable space $(\Omega, \mathcal{F})$, and let $\{P_\lambda,\ \lambda \in \Lambda\}$ be a collection of probability measures on $(\Omega, \mathcal{F})$. It is well known ([26]) that, for a fixed $\lambda \in \Lambda$, $X_n \to 0$ P_λ-a.s. is equivalent to

$$\lim_{N \to \infty} P_\lambda\left(\sup_{n \ge N} |X_n| > \epsilon\right) = 0 \qquad \forall \epsilon > 0. \tag{3.1}$$

Now, we say that $X_n \to 0$ P_λ-a.s., *at a uniform rate over* Λ, if convergence in (3.1) is uniform over Λ, that is

$$\lim_{N \to \infty} \sup_{\lambda \in \Lambda} P_\lambda\left(\sup_{n \ge N} |X_n| > \epsilon\right) = 0. \tag{3.2}$$

The basic concepts of this paper, namely approachability and the dual concept of excludability, are introduced in the following definition. Here and below, $d(r, B)$ denoted the Euclidean point-to-set distance, i.e., $d(r, B) = \inf_{\beta \in B} d(r, \beta)$ where $d(r, \beta) = |r - \beta|$.

Definition 3.1: Let the initial state s be fixed. A set $B \subset \mathbb{R}^q$ is approachable (from s by DM1) if there exists a B-approaching strategy $\pi_1^* \in \Pi_1$ such that

$$d(\bar r_n, B) \to 0 \qquad P\text{-a.s. for every } \pi_2 \in \Pi_2$$

at a uniform rate over Π_2. B is excludable (from s by DM2) if there exists a B-excluding strategy $\pi_2^* \in \Pi_2$ such that, for some $\delta > 0$

$$d(\bar r_n, B_\delta^c) \to 0 \qquad P\text{-a.s. for every } \pi_1 \in \Pi_1$$

at a uniform rate over Π_1, where $B_\delta^c = \{\beta \in \mathbb{R}^q\colon d(\beta, B) \ge \delta\}$.

Remarks:

1) The convergence $d(\bar{r}_n, B_\delta^c) \to 0$ in the definition of excludability may be equivalently written as: $\liminf_n d(\bar{r}_n, B) \geq \delta$. Thus, loosely speaking, a set B is approachable if DM1 can guarantee (irrespective of the other's strategy) that the long-term average payoff vector is in B, and B is excludable if DM2 can guarantee (irrespective of DM1's strategy) that the long-term average payoff is at least a distance $\delta > 0$ away from B.

2) It is obvious that approachability and excludability are contradictory, in the sense that a given set cannot be both approachable by DM1 and excludable by DM2. However, these concepts are not exact opposites of each other. Indeed, it was demonstrated in [4] that even in repeated matrix games, some (nonconvex) sets may be neither approachable nor excludable.

3) In the sequel, it will be convenient to assume that the set B is closed. This involves no loss of generality, since approachability (and excludability) of a set and its closure are plainly the same.

4) An important aspect of the definition is the uniform rate of convergence. This requirement is essential if the infinite stage model is considered as an approximation to the model with very long, but finite, time horizon.

We proceed to formulate the key technical result, which presents a sufficient condition for approachability. To this end, let

$$\phi(\pi_1, \pi_2) = \frac{E^0_{\pi_1, \pi_2}(\sum_{n=0}^{\tau-1} r_n)}{E^0_{\pi_1, \pi_2}(\tau)} \tag{3.3}$$

denote the averaged payoff per "cycle" from state 0 and back. Note that $\phi(\pi_1, \pi_2)$ is well defined if either π_1 or π_2 is a stable strategy. Let $\phi(\pi_1, \Pi_2) := \{\phi(\pi_1, \pi_2)\colon \pi_2 \in \Pi_2\}$. We say that a strategy $\pi_1 \in \Pi_1$ is *started at stage* T if at stage $n = T$ (possibly random) DM1 resets an internal clock to 0 and starts using π_1 as if the state s_T is the initial state.

Let B be a closed set in $\mathbb{R}^q$. For any point $\rho \notin B$ let C_ρ denote a closest point in B to ρ. Let H_ρ be the hyperplane through C_ρ which is perpendicular to $(C_\rho - \rho)$, and let u_ρ be a unit vector in the direction of $(C_\rho - \rho)$ (see Fig. 1).

Theorem 3.1: Assume that the following condition is satisfied:

SC1) For every $\rho \notin B$, there exists a stable strategy $\pi_1(\rho) \in \Pi_1$ such that

$$\langle \phi(\pi_1(\rho), \pi_2) - C_\rho, u_\rho \rangle \geq 0 \qquad \forall \pi_2 \in \Pi_2 \tag{3.4}$$

(equivalently, $\phi(\pi_1(\rho), \Pi_2)$ is weakly separated by H_ρ from ρ). Furthermore, the set $\{\pi_1(\rho)\colon \rho \notin B\}$ is uniformly stable.

Then B is approachable from state 0 by DM1, and a B-approaching strategy is given as follows. Let $0 < T(1) < T(2) < \cdots$ be the subsequent arrival instants to state 0. Let π_1' be some fixed stable strategy for DM1. Then:

- at stages $0 \leq n < T(1)$: use π_1'.
- at stages $T(k) \leq n < T(k+1)$, $k \geq 1$:
  if $\bar{r}_{T(k)} \notin B$, then use $\pi_1(\bar{r}_{T(k)})$, started at $T(k)$.
  if $\bar{r}_{T(k)} \in B$, then use π_1', started at $T(k)$.

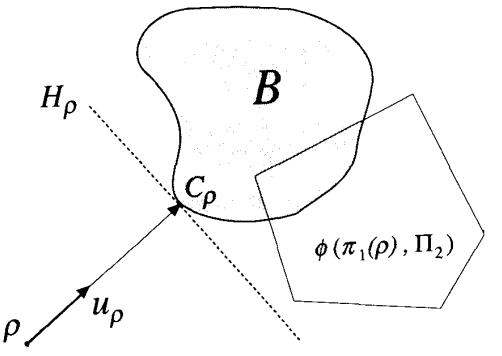


Fig. 1. Geometric interpretation of SC1).

The proof of this theorem is presented in the next section.

Remark: In the approaching strategy presented above, we have specified for simplicity a single strategy π_1' which is employed whenever $\bar{r}_{T(k)} \in B$. More generally, given a *uniformly stable* set Π_1' of DM1's strategies, the results remain valid if an arbitrary $\pi_1' \in \Pi_1'$ is chosen whenever this condition is encountered; indeed, the proof of Theorem 3.1 is given for this more general case. This added freedom may be utilized by DM1 to attend other (secondary) objectives whenever the current average payoff is already in B.

The sufficient condition SC1) and the approaching strategy of Theorem 3.1 admit an intuitively appealing geometric interpretation ([4], [17]). As already noted, (3.4) simply means that H_ρ separates ρ from the set $\phi(\pi_1(\rho), \Pi_2)$ (cf. Fig. 1). Consider then the approaching strategy suggested above: whenever state 0 is reached and the average payoff $\bar{r}$ is outside B, DM1 employs $\pi_1(\bar{r})$ for the next cycle (i.e., up to the next time when 0 is reached). Thus, the averaged payoff in that cycle, as defined in (3.3), will belong to the set $\phi(\pi_1(\bar{r}), \Pi_2)$, and will therefore cause the average payoff $\bar{r}$ to advance towards that set (in some probabilistic sense). Now, the stability conditions imposed on $\{\pi_1(\rho)\}$ imply that, as time progresses, the effect of any one cycle on the average payoff becomes small; therefore, the average payoff actually moves closer to B on each cycle. This suggests that the average payoff will converge to B in the long run.

In Theorem 3.1 it was assumed that the initial state is 0. If not, the conditions of Theorem 3.1 may still be applied provided DM1 can guarantee that state 0 is reached "fast enough." In particular, the proof of Theorem 3.1 applies without modification to the following result:

Corollary 3.2: Assume that condition SC1) of Theorem 3.1 is satisfied. In addition, assume that for some strategy

$\sigma \in \Pi_1$

$$\sup_{\pi_2} E^s_{\sigma,\pi_2}(\tau^2) < \infty \tag{3.5}$$

$$\sup_{\pi_2} E^s_{\sigma,\pi_2}\left(\sum_{n=0}^{\tau-1} |r_n|\right)^2 < \infty. \tag{3.6}$$

Then B is approachable from state s by DM1. An approaching strategy is given as in Theorem 3.1, except that up to time $\tau \equiv T(1)$ the strategy σ is used.

We note, in passing, that when the set B or the payoff function r are bounded, a similar result may easily be established if the equivalents of (3.5) and (3.6) hold only for first (instead of second) moments. From here on, we shall always assume that the initial state is 0, while keeping in mind that the results may be extended to other initial states provided that similar conditions to those of Corollary 3.2 are satisfied.

Under assumption C1) (defined in Section II), the sufficient condition of Theorem 3.1 may be expressed in terms the values of the games $\Gamma_0(u)$. Furthermore, the implied approaching strategy is "piecewise stationary," as specified in the following corollary.

Corollary 3.3: Assume C1). Let B, ρ, C_ρ and u_ρ be as in Theorem 3.1. Then B is approachable from state 0 if

$$\operatorname{val}\Gamma_0(u_\rho) \geq \langle C_\rho, u_\rho\rangle, \qquad \forall \rho \notin B. \tag{3.7}$$

An approaching strategy is then as specified in Theorem 3.1, with $\pi_1(\rho) \triangleq f^*(u_\rho)$, the stationary $\Gamma_0(u_\rho)$-optimal strategy specified in C1).

Proof: It suffices to show that (3.7) implies condition SC1) of Theorem 3.1, with $\pi_1(\rho) = f^*(u_\rho)$. Let $\rho \notin B$ be fixed, and write u, f^*, and C for u_ρ, $f^*(u_\rho)$ and C_ρ. Since f^* is optimal in $\Gamma_0(u)$, then

$$\liminf_{n\to\infty} E^0_{f^*,\pi_2}\langle \bar{r}_n, u\rangle \geq \operatorname{val}\Gamma_0(u) \geq \langle C, u\rangle, \qquad \forall \pi_2 \in \Pi_2. \tag{3.8}$$

For each $\pi_2 \in \Pi_2$, let $\bar{\pi}_2 \in \Pi_2$ be the strategy which starts according to π_2 but regenerates (restarts) whenever state 0 is reached. Since f^* is stable, it follows from the theory of renewal reward processes (cf. [23, sections III-I. and VII-C.]) that

$$\phi(f^*, \pi_2) \equiv \phi(f^*, \bar{\pi}_2) = \lim_{n\to\infty} E^0_{f^*,\bar{\pi}_2}(\bar{r}_n), \qquad \forall \pi_2 \in \Pi_2. \tag{3.9}$$

Thus, from (3.8)

$$\langle \phi(f^*, \pi_2), u\rangle = \lim_{n\to\infty} E_{f^*,\bar{\pi}_2}\langle \bar{r}_n, u\rangle \geq \langle C, u\rangle, \qquad \forall \pi_2 \in \Pi_2 \tag{3.10}$$

which is exactly the required inequality (3.4). □

We consider next the important special case where the set B is convex. It is then possible to obtain [under C2)] a complete characterization of approachability. Some additional notation is introduced first.

For every stable $g \in G$, the long-run average expected payoff:

$$R(f,g) = \lim_{n\to\infty} E^0_{f,g}(\bar{r}_n), \qquad f \in F \tag{3.11}$$

is well defined, and (as noted in the proof of Corollary 3.3) equals $\phi(f,g)$. Further define the following bounded subsets of $\mathbb{R}^q$:

$$R(F,g) = \{R(f,g): f \in F\},$$
$$\bar{R}(F,g) = \operatorname{conv} R(F,g) \tag{3.12}$$

where "conv" denotes the closed convex hull. (In fact, it may be established as in [2] or [8, p. 95] that $R(F,g)$ is convex, so that $\bar{R}(F,g)$ is just its closure. Moreover, if the state space is finite then $R(F,g)$ is just the convex hull of the finite set $\{R(f,g): f \in F_D\}$ where F_D are the *nonrandomized* strategies in F.) Note that boundedness of $R(F,g)$ follows from property (2.5) in the definition of a stable strategy. The sets $R(f,G)$ and $\bar{R}(f,G)$ are similarly defined for any stable $f \in F$, and the same comments apply. Finally, for each convex set B in $\mathbb{R}^q$ define the following set of unit vectors:

$$U(B) = \{u_\rho \in U: \rho \notin B\}.$$

This set represents all directions in which a point outside B might be projected onto B (cf. Fig. 1). Note that $U(B) = U$ if B is bounded.

Theorem 3.4: Assume C2). Let B be a closed convex set in $\mathbb{R}^q$, and let the initial state $s_0 = 0$.

i) B is approachable if and only if either one of the following equivalent conditions are satisfied:

NSC1: There exists a uniformly stable set $\{f(u): u \in U(B)\}$ of stationary strategies for DM1 such that: every $\rho \notin B$ is separated from $\bar{R}(f(u_\rho), G)$ by H_ρ, i.e.,

$$\inf_{g\in G}\left\langle R\big(f(u_\rho), g\big), u_\rho\right\rangle \geq \langle C_\rho, u_\rho\rangle. \tag{3.13}$$

NSC2: The separation condition in NSC1 holds for $f(u) \triangleq f^*(u)$, the optimal strategy of DM1 in $\Gamma_0(u)$, $u \in U(B)$.

NSC3: $\operatorname{val}\Gamma_0(u) \geq \min_{\beta\in B}\langle \beta, u\rangle$ for every $u \in U(B)$.

NSC4: $\bar{R}(F,g)$ intersects B for every stable $g \in G$.

ii) If B is not approachable, then it is excludable by DM2 with a stationary strategy.

Proof: We first establish the equivalence of NSC1–NSC4, and then prove that NSC3 is necessary and sufficient for B to be approachable.

a) Equivalence of NSC1 and NSC2: Since the set $\{f^*(u): u \in U(B)\}$ is uniformly stable by C2), then NSC2 trivially implies NSC1. Conversely, by C2) and (3.11) we have for every $\rho \in B$

$$\inf_{g\in G}\left\langle R\big(f(u_\rho), g\big), u_\rho\right\rangle \leq \operatorname{val}\Gamma_0(u_\rho) = \min_{g\in G}\left\langle R\big(f^*(u_\rho), g\big), u_\rho\right\rangle. \tag{3.14}$$

Thus NSC1 implies NSC2.

b) Equivalence of NSC2 and NSC3: Since B is convex and closed,

$$\langle C_\rho, u_\rho\rangle = \min_{\beta\in B}\langle \beta, u_\rho\rangle, \qquad \rho \notin B. \tag{3.15}$$

The required equivalence follows from this and the last equality in (3.14).

c) Equivalence of NSC3 and NSC4: Assume first that NSC4 holds. By C2) it follows that for every $u \in U(B)$

$$\text{val}\,\Gamma_0(u) = \max_{f \in F} \langle R(f, g^*(u)), u \rangle \geq \min_{\beta \in B} \langle \beta, u \rangle \quad (3.16)$$

where the last inequality follows since $\bar{R}(F, g^*(u))$ intersects B by our assumption. Thus, NSC3 is satisfied.

Assume, conversely, that $\bar{R}(F, g') \cap B = \varnothing$ for some stable $g' \in G$. Since both these sets are closed convex and $\bar{R}(F, g')$ is bounded, then they are strongly separated ([22]), i.e., there exists a vector $u' \in U$ such that

$$\max_{r \in \bar{R}(F, g')} \langle r, u' \rangle < \min_{\beta \in B} \langle \beta, u' \rangle. \quad (3.17)$$

Furthermore u' is necessarily in $U(B)$ since, as may be easily verified, $\inf_{\beta \in B} \langle \beta, u \rangle = -\infty$ for $u \notin U(B)$. Since by C2) there exists an $f^*(u') \in F$ which is optimal (maximin) in $\Gamma_0(u')$, it follows that

$$\text{val}\,\Gamma_0(u') \leq \langle R(f^*(u'), g'), u' \rangle < \min_{\beta \in B} \langle \beta, u' \rangle \quad (3.18)$$

which contradicts NSC3. Thus, equivalence of NSC3 and NSC4 is proved.

d) Sufficiency of NSC3: Follows directly by Corollary 3.3 and (3.15).

e) Necessity of NSC2: Assume that NSC3 is not satisfied, so that (3.18) holds for some $u' \in U(B)$. Let $g' \triangleq g^*(u')$ be the stable optimal strategy of DM2 in $\Gamma_0(u')$, and define

$$\bar{\phi}(\Pi_1, g') \triangleq \text{conv}\{\phi(\pi_1, g') : \pi_1 \in \Pi_1\}. \quad (3.19)$$

It follows now from (3.18) (cf. the proof of Corollary 3.3) that $d(\bar{\phi}(\Pi_1, g'), B) \geq \delta'$. But this implies that B is excludable by DM2 (with his stationary strategy g'). To show this, simply apply Theorem 3.1, with the roles of DM1 and DM2 interchanged, to establish that $B' = \bar{\phi}(\Pi_1, g')$ is approachable by DM2 using g'. Thus, necessity is established.

Finally, note that ii) has already been established in e) above, and the proof is complete. □

Remarks: The following comments point out some consequences of the last theorem.

1) Part ii) of the theorem implies that every convex set is either approachable by DM1 or excludable by DM2. As observed in [4], this dichotomy may be considered a generalization of the minimax theorem for the corresponding game model with scalar payoffs (in the present case, a zero-sum stochastic game).

2) An excludable convex set B may always be excluded by a *stationary* strategy of DM2. Thus, an excludable convex set will remain so even if DM2 is restricted to stationary strategies only (or any superset thereof). Such restriction may be natural in certain cases, particularly if DM2 is not actually a conscious decision maker, but is incorporated in the model to facilitate a worst-case analysis with respect to system uncertainties or state-dependent noise.

3) The excluding strategy may always be chosen as the (stable and stationary) strategy $g^*(u')$, where u' satisfies (3.18). Moreover, any strategy g which violates NSC4, i.e., for which $\bar{R}(F, g)$ and B are disjoint, is obviously a candidate for an excluding strategy. (However, for this to be true in general it is also required that $\bar{\phi}(\Pi_1, g) = \bar{R}(F, g)$. This equality is satisfied in the case of a finite state space, as may be inferred from [8, ch. 7], as well as under the conditions of Lemma 2.1. Other (fairly mild) conditions which imply this equality may be found, e.g., in [2].)

4) The approachability conditions in NSC1–3 are required to hold only for a certain subset of all unit vectors, namely $U(B)$. This may be interpreted as follows. The set $U(B)$ is a *proper* subset of U if and only if B is unbounded, and $u' \notin U(B)$ is equivalent to $\inf_{\beta \in B} \langle \beta, u' \rangle = -\infty$ (i.e., $-u'$ is in the recession cone ([22]) of B). Thus, the average payoff vector may become unboundedly small (negative) in the direction of u' without leaving B, and as far as approaching B is concerned DM1 need not be concerned with "pushing" the payoff in that direction.

In fact, for Theorem 3.4 to hold, condition C2) may be somewhat weakened by requiring that it should hold only over $U(B)$; i.e., existence and stability of stationary optimal strategies in $\Gamma_0(u)$ may be required only for $u \in U(B)$, without affecting the conclusions or proof. This may be significant when existence and stability depend on properties of the payoff functions, such as positiveness or one-side boundedness; cf. [18] and also, e.g., [2], [25] for some relevant results in the single controller case.

Computation and Implementation Issues: We close this section with a brief discussion of certain issues related to the verification of the above approachability conditions and the implementation of approaching strategies. We focus on the case of a finite state space. This case is obviously more tractable computationally than the countable state case; fortunately, it seems that most practical problems with discrete state space are basically of a finite nature, and may be analyzed as such when convenient. This applies, for example, to queueing systems, where the queue length is always bounded in practice.

1) As noted in Section II, conditions C1 and C2, as well as all other stability conditions mentioned in this section, are satisfied under the conditions of Lemma 2.1. For a finite state space, the latter reduce to the recurrence of state 0 under every pair of stationary nonrandomized strategies.

2) One may distinguish two stages in the possible application of the above results for a given model. The first would be the performance evaluation stage, where the approachable sets are identified and feasible control objectives are accordingly determined (as demonstrated in the next section). The second is the computation of an appropriate strategy to implement these objectives, i.e., to approach the desired set in the performance space. The following comments address these two issues in that order.

3) In order to verify approachability of a given set, condition NSC3 of Theorem 3.4 (for convex sets) or Corollary 3.3 (for general sets) require the computation of the values of the zero-sum stochastic games $\Gamma_0(u)$, $u \in U(B)$. In practice, it would be sufficient to compute these values for a "sufficiently dense" set of unit vectors. Moreover, one may concentrate on computing lower bounds on the values. Indeed, if lower bounds are substituted instead of the actual values, the conditions above are still *sufficient* for approachability. The associated loss in accuracy depends of course on the tightness of the bounds.

4) In general, the value of a zero-sum stochastic game is not computable by a finite algorithm. This remains true even under the recurrence conditions of Lemma 2.1. However, under these conditions there exist recursive algorithms which converge to the value, and stopping rules for obtaining ϵ-approximations to the value (and ϵ-optimal strategies) ([10]). Finite algorithms for the exact calculation of the value do exist for various special classes of stochastic games, see [20] for a review. This is the case, for example, in stochastic games with *perfect information*, where at each state (or stage) only one player may choose an action. These games are also distinguished by always having *nonrandomized* optimal strategies, and seem appropriate for engineering applications where the simultaneous choice of actions is not an essential part of the problem.

5) Note that the set of values $\{\Gamma_0(u), u \in U\}$ does not depend on the specific set B. Thus, these values (or their approximation) need to be computed only once, and may then be applied to any set B.

6) The following comments concern the computation of an approaching strategy. Such a strategy would consist of a collection of (stationary) strategies, which are switched at state 0 according to the current average payoff vector.

7) It should be emphasized that this collection of strategies need not contain all the optimal strategies $f^*(u)$. Indeed, concentrating on the convex case, any set of strategies which satisfies NSC1 is appropriate. Moreover, the number of different strategies in this set need not necessarily be large, since a single strategy may satisfy the required inequality in NSC1 for a set of adjacent unit vectors. The required cardinality depends on how ambitious we are in setting the control goals, as reflected by the set B to be approached, and may be reduced by expanding this set.

8) To elucidate the last point, we outline a reasonable procedure for the computation of an approaching strategy. Assume that one starts with a given convex set B, and because of memory limitations the number of different stationary strategies which comprise the approaching strategy is restricted to a certain number N. Assume also that the conditions of Lemma 2.1 hold. One may then choose N unit vectors which are equally spaced in the set $U(B)$, and compute for each such vector u_i an optimal or ϵ-optimal strategy f_i in $\Gamma_0(u_i)$ (see Remark 4 above regarding this computation). Noting (3.15), then NSC1 will be satisfied if the following holds: for every $u \in U(B)$, the inequality

$$J(f_i, u) \triangleq \inf_{g \in G} \langle R(f_i, g), u \rangle \geq \min_{\beta \in B} \langle \beta, u \rangle \quad (3.20)$$

holds for some f_i. Note that the computation of $J(f_i, u)$ is a standard Markov decision problem (with respect to the scalar cost function $\langle r, u \rangle$). Now, a reasonably dense grid of unit vectors has to be chosen in $U(B)$, and the above condition checked over this grid. (A reasonable approach would be to compute $J(f_i, u)$ only for those f_i which correspond to vectors u_i that are close to u.) If it holds, then the set B may be (approximately) approached by an approaching strategy which is comprised of the set $\{f_i\}$. Otherwise, the set B needs to be expanded in the directions where the condition fails, until it is achieved. Obviously, we may or may not be content with the resulting set B, and accordingly may want to reconsider the constraint of N strategies.

IV. Proof of Theorem 3.1

The following martingale-related convergence result will be required.

Proposition 4.1: Let $\{X_n, \mathcal{F}_n, n \geq 0\}$ be a stochastic sequence on some probability space $(\Omega, \mathcal{F}, P)$; that is, $\{\mathcal{F}_n\}$ is an increasing sequence of σ-fields and X_n is $\mathcal{F}_n$-measurable. Assume that $X_0 = 0$, and that there exists a positive constant Q such that

$$E(X_{n+1}^2 \mid \mathcal{F}_n) \leq X_n^2 + W_n^2 \ (P\text{-a.s.}), \qquad n \geq 0 \quad (4.1)$$

where W_n is an $\mathcal{F}_n$-measurable random variable such that $E(W_n^2) \leq Q$. Then $X_n/n \to 0$ a.s., and the rate of convergence depends only on the constant Q. More precisely, for every $\epsilon > 0$ and $\delta > 0$ there exists an integer $N = N(\epsilon, \delta, Q)$ such that:

$$P\left\{ \sup_{n \geq N} \frac{|X_n|}{n} \geq \epsilon \right\} \leq \delta. \quad (4.2)$$

Proof: Note first that (4.1) implies

$$E(X_n^2) \leq nQ, \qquad n \geq 0 \quad (4.3)$$

(which, incidentally, implies that $X_n/n \to 0$ in the mean-square sense).

Let $N \geq 1$ be some fixed integer. Define

$$Z_n = \frac{X_n^2}{n^2} - V_{n-1}, \qquad n \geq N \quad (4.4)$$

where

$$V_n = \sum_{m=N}^{n} \frac{W_m^2}{(m+1)^2}, \qquad n \geq N$$

and $V_{N-1} \triangleq 0$. Note that $\{V_n\}$ is a positive increasing sequence, and

$$E(V_\infty) \leq \sum_{m=N}^{\infty} \frac{Q}{(m+1)^2} \leq \int_N^\infty \frac{Q}{t^2}\, dt = \frac{Q}{N}. \quad (4.5)$$

Now, by (4.4) and (4.1)

$$E(Z_{n+1}|\mathcal{F}_n) = E\left(\frac{X_{n+1}^2}{(n+1)^2}\middle|\mathcal{F}_n\right) - V_n \le \frac{X_n^2}{n^2} + \frac{W_n^2}{(n+1)^2} - V_n = Z_n, \qquad n \ge N$$

so that $\{Z_n, \mathcal{F}_n;\ n \ge N\}$ is a supermartingale. Denoting $Z_n^- = \max\{0, -Z_n\}$, it follows by a standard supermartingale inequality ([26, p. 475]) that, for every $\epsilon' > 0$:

$$\epsilon' P\left\{\sup_{n\ge N} Z_n \ge \epsilon'\right\} \le E(Z_N) + \sup_{n\ge N} E(Z_n^-) \le E\left(\frac{X_N^2}{N^2}\right) + E(V_\infty) \le \frac{2Q}{N} \quad (4.6)$$

where (4.3) and (4.5) were used in the last step. Together with (4.5), this implies that

$$P\left\{\sup_{n\ge N}\frac{|X_n|}{n} \ge \epsilon\right\} = P\left\{\sup_{n\ge N}(Z_n + V_{n-1}) \ge \epsilon^2\right\} \le P\left\{\sup_{n\ge N} Z_n \ge \frac{\epsilon^2}{2}\right\} + P\left\{V_\infty \ge \frac{\epsilon^2}{2}\right\} \le \frac{4Q}{\epsilon^2 N} + \frac{2}{\epsilon^2}E(V_\infty) \le \frac{6Q}{\epsilon^2 N}. \quad (4.7)$$

Thus, (4.2) follows for any $N \ge 6Q/\delta\epsilon^2$. □

Proof of Theorem 3.1: Assume that condition SC1) of the theorem is satisfied. Let DM1 employ the specified approaching strategy (denoted π_1^*), and DM2 any $\pi_2 \in \Pi_2$. In the sequel, all relations between random variables are assumed to hold almost surely with respect to the measure induced by these strategies and the initial state $s_o = 0$.

Recall that

$$T(k) = \inf\{n > T(k-1): s_n = 0\} \quad (4.8)$$

where $T(0) \triangleq 0$. By definition of π_1^* and (2.4) it follows that the $T(k)$'s are all finite (except maybe on a set of measure 0, which we will henceforth ignore). Let $\mathcal{F}_n$ be the σ-algebra generated by the history sequence h_n. Obviously, each $T(k)$ is a stopping time with respect to $\{\mathcal{F}_n\}$.

The following abbreviated notation will be useful:

$$c_n \triangleq C_{\bar r_n}, \quad \text{the closest point in } B \text{ to } \bar r_n$$
$$d_n \triangleq d(\bar r_n, B) = |\bar r_n - c_n|$$
$$\tau_{k+1} = T(k+1) - T(k)$$
$$v_{k+1} \triangleq \sum_{n=T(k)}^{T(k+1)-1} r_n = T(k+1)\bar r_{T(k+1)} - T(k)\bar r_{T(k)}. \quad (4.9)$$

Note that the set of strategies $\{\pi_1(\rho)\} \cup \Pi_1'$ (which was used to define π_1^*; note that we allow for the more general strategy discussed in the remark which follows the theorem) is uniformly stable by assumption. Since on each interval $[T(k), T(k+1) - 1]$ one of these strategies is used, it follows by (2.4), (2.5) that

$$E\left(\tau_{k+1}^2|\mathcal{F}_{T(k)}\right) \le M_2 \quad (4.10)$$

$$E\left(|v_{k+1}|^2|\mathcal{F}_{T(k)}\right) \le E\left(\left(\sum_{n=T(k)}^{T(k+1)-1}|r_n|\right)^2\middle|\mathcal{F}_{T(k)}\right) \le R_2 \quad (4.11)$$

for some constants M_2, R_2, and every $k \ge 0$. Furthermore, since $T_k \ge k$

$$|\bar r_{T(k)}| \le \frac{1}{k}\left|\sum_{m=1}^{k} v_m\right|$$

and it follows easily by (4.11) that

$$E\left(|\bar r_{T(k)}|^2\right) \le R_2, \qquad k \ge 0. \quad (4.12)$$

It is our purpose to prove that $d_n \to 0$. We proceed in three steps: First, the properties of the approaching strategy are used to obtain some bounds on the "sampled" distance sequence $\{d_{T(k)}\}$. Convergence of $d_{T(k)}$ (as $k \to \infty$) is next established, and finally extended to the whole sequence $\{d_n\}$.

i) The Basic Bounds: We set out to establish that

$$E\left(T(k+1)^2 d_{T(k+1)}^2|\mathcal{F}_{T(k)}\right) \le T(k)^2 d_{T(k)}^2 + W_k^2, \quad (4.13)$$

where

$$E(W_k^2) \le Q, \quad (4.14)$$

for some constant Q independent of DM2's strategy.

For every $k \ge 0$, if $d_{T(k)} > 0$ (i.e., $\bar r_{T(k)} \notin B$), then DM1 is using $\pi_1(\bar r_{T(k)})$ on $[T(k), T(k+1) - 1)$. It follows then by (3.3), (3.4) and (4.9) that on $\{d_{T(k)} > 0\}$

$$\left\langle \frac{E\left(v_{k+1}|\mathcal{F}_{T(k)}\right)}{E\left(\tau_{k+1}|\mathcal{F}_{T(k)}\right)} - c_{T(k)}, c_{T(k)} - \bar r_{T(k)}\right\rangle \ge 0. \quad (4.15)$$

Note that (4.15) holds trivially on $\{d_{T(k)} = 0\}$, since then $c_{T(k)} - \bar r_{T(k)} = 0$. Thus, rearranging terms in (4.15) gives

$$E\left(\langle v_{k+1} - \tau_{k+1}c_{T(k)}, c_{T(k)} - \bar r_{T(k)}\rangle|\mathcal{F}_{T(k)}\right) \ge 0. \quad (4.16)$$

Now

$$d_{T(k+1)} = |\bar r_{T(k+1)} - c_{T(k+1)}| \le |\bar r_{T(k+1)} - c_{T(k)}| = \left|\frac{T(k)\bar r_{T(k)} + v_{k+1}}{T(k+1)} - c_{T(k)}\right| \quad (4.17)$$

so that

$$T(k+1)^2 d_{T(k+1)}^2 \le \left|T(k)(\bar r_{T(k)} - c_{T(k)}) + (v_{k+1} - \tau_{k+1}c_{T(k)})\right|^2 = T(k)^2 d_{T(k)}^2 + |v_{k+1} - \tau_{k+1}c_{T(k)}|^2 - 2T(k)\langle v_{k+1} - \tau_{k+1}c_{T(k)}, c_{T(k)} - \bar r_{T(k)}\rangle.$$

Thus, it follows from (4.16) that

$$E\left(T(k+1)^2 d_{T(k+1)}^2|\mathcal{F}_{T(k)}\right) \le T(k)^2 d_{T(k)}^2 + W_k^2 \quad (4.18)$$

where

$$W_k^2 \triangleq E\left(|v_{k+1} - \tau_{k+1}c_{T(k)}|^2 \middle| \mathcal{F}_{T(k)}\right). \quad (4.19)$$

It remains to bound $E(W_k^2)$. By (4.10) and (4.11)

$$W_k^2 \le 2E\left(|v_{k+1}|^2|\mathcal{F}_{T(k)}\right) + 2|c_{T(k)}|^2E\left(\tau_{k+1}^2|\mathcal{F}_{T(k)}\right)$$
$$\le 2R_2 + 2|c_{T(k)}|^2M_2. \quad (4.20)$$

If B is a bounded set, then $c_{T(k)} \in B$ is uniformly bounded and so is W_k^2. However, in the general case we still have to bound $E(|c_{T(k)}|^2)$. Let β be some fixed point in B. Since $c_{T(k)}$ is closest in B to $\bar{r}_{T(k)}$, then $|\bar{r}_{T(k)} - c_{T(k)}| \le |\bar{r}_{T(k)} - \beta|$, and

$$|c_{T(k)}| \le |c_{T(k)} - \bar{r}_{T(k)}| + |\bar{r}_{T(k)}| \le |\bar{r}_{T(k)} - \beta| + |\bar{r}_{T(k)}|$$
$$\le 2|\bar{r}_{T(k)}| + |\beta|.$$

Noting that $(a+b)^2 \le 2(a^2+b^2)$, it follows by (4.12) that $E(|c_{T(k)}|^2) \le 2(4R_2 + |\beta|^2)$, which gives in (4.20)

$$E(W_k^2) \le 2R_2 + 4\left(4R_2 + |\beta|^2\right)M_2 \triangleq Q \quad (4.21)$$

and (4.13) and (4.14) are established.

ii) Convergence of $d_{T(k)}$: By (4.13) and (4.14), and Proposition 4.1 it follows that

$$\lim_{k\to\infty} \frac{1}{k}T(k)d_{T(k)} = 0 \qquad \text{a.s.}$$

and the rate of convergence depends on Q only. Since $T_k \ge k$, the same follows for $d_{T(k)}$, i.e., $d_{T(k)} \to 0$ a.s. at a uniform rate.

iii) Convergence of d_n: Let $\epsilon > 0$ and $\delta > 0$ be fixed. It has just been established in ii) that there exists an integer $k_0 = k_0(\epsilon, \delta, Q)$ such that

$$P\left(\sup_{k\ge k_0} d_{T(k)} \ge \frac{\epsilon}{2}\right) \le \frac{\delta}{2}. \quad (4.22)$$

We proceed to bound $E(D_k^2)$, where

$$D_k \triangleq \sup\left\{|d_n - d_{T(k)}|: T(k) \le n < T(k+1)\right\}.$$

It follows from the definition of d_n and the triangle inequality that, for every n, m:

$$d_n - d_m \le |\bar{r}_n - c_m| - |\bar{r}_m - c_m| \le |\bar{r}_n - \bar{r}_m|.$$

Thus, for $n \ge m$

$$|d_n - d_m| \le |\bar{r}_n - \bar{r}_m| = \frac{1}{n}\left|m\bar{r}_m + \sum_{l=m}^{n-1} r_l - n\bar{r}_m\right| \quad (4.23)$$
$$\le \frac{1}{m}\left((n-m)|\bar{r}_m| + \sum_{l=m}^{n-1}|r_l|\right).$$

Noting that $T(k) \ge k$, this implies that

$$D_k \le \frac{1}{k}\left(\tau_{k+1}|\bar{r}_{T(k)}| + \sum_{l=T(k)}^{T(k+1)-1}|r_l|\right)$$

so that, by (4.10)–(4.12)

$$E(D_k^2) = E\left\{E\left(D_k^2|\mathcal{F}_{T(k)}\right)\right\}$$
$$\le \frac{2}{k^2}E\left\{M_2|\bar{r}_{T(k)}|^2 + R_2\right\} \quad (4.24)$$
$$\le \frac{2(M_2R_2 + R_2)}{k^2} := \frac{Q_0}{k^2}.$$

Therefore

$$P\left\{\sup_{k\ge k_1} D_k > \frac{\epsilon}{2}\right\} \le \sum_{k=k_1}^{\infty} P\left\{D_k > \frac{\epsilon}{2}\right\}$$
$$\le \left(\frac{2}{\epsilon}\right)^2 \sum_{k=k_1}^{\infty} E(D_k^2) \quad (4.25)$$
$$\le \frac{4Q_0}{\epsilon^2}\sum_{k=k_1}^{\infty}\frac{1}{k^2} \le \frac{\delta}{4}$$

where the last inequality holds for some $k_1 = k_1(\epsilon, \delta)$ large enough, which may be chosen to satisfy $k_1 \ge k_0$ [cf. (4.22)]. Finally, for some $N = N(k_1, \delta)$ large enough, it follows by (4.10) that

$$P(T(k_1) > N) \le \frac{1}{N}E(T(k_1))$$
$$\le \frac{k_1M_2}{N} \le \frac{\delta}{4}. \quad (4.26)$$

Now, (4.22), (4.25), and (4.26) imply that

$$P\left(\sup_{n\ge N} d_n > \epsilon\right) \le \frac{\delta}{2} + \frac{\delta}{4} + \frac{\delta}{4} = \delta.$$

Noting that ϵ and δ are arbitrary and $N = N'(\epsilon, \delta)$ does not depend on DM2's strategy, the proof is complete. □

V. A Queueing Example

In this section, we apply the previous results to a simple discrete-time queueing system. Dynamic control of admission, routing and service in queueing systems has been extensively studied in the past decade; see, e.g., [30] and its references. Here we consider the case of service-rate control (by DM1), while the arrival process is not completely specified (or, alternatively, controlled by 'DM2'). The basic problem is maintaining adequate service quality to attract customers, while keeping service costs down. For illustrative purposes, the model was simplified as much as possible so that, while not being trivial, it lends itself to a simple analytic treatment.

Consider the queueing system illustrated in Fig. 2. The time axis is divided into "slots" $n = 0, 1, 2, \cdots$. Only one customer may arrive during each time slot, and the arrival probability is λ. If the queue is empty, he joins the queue and then enters service at the beginning of the next time slot. Otherwise, he may choose either to join the queue, or to leave the system and never return.

Service is applied to one customer at each time slot (provided the queue is not empty at the beginning of the

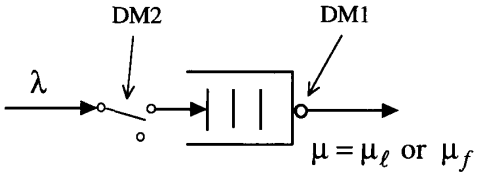


Fig. 2. The queueing system.

slot). The server (DM1) may choose between a slow service mode, where the probability of successful service on each time slot is μ_l, and a fast service mode with success probability μ_f, where $1 \geq \mu_f > \mu_l > \lambda$. If the service is successful the customer leaves the system, otherwise he remains for (at least) another try. We assume that the server may switch service mode *only when the queue is empty*. A fixed cost (which is assumed for convenience to be of μ_f units) is incurred for each service attempt in the fast mode, while slow service is costless.

To fit the model of the previous sections, we may regard all arrival decisions as being made by DM2. The system may be formally described as follows. Let the state $s = (x, M)$, where $x \in \{0, 1, \cdots\}$ is the number of customers in the queue at the beginning of a time slot and $M \in$ {slow, fast, empty} is the service mode in that slot. Thus,

$$x_{n+1} = \begin{cases} A_n & : \quad \text{if } x_n = 0 \\ x_n + A_n a_n^2 - U_n & : \quad \text{if } x_n > 0 \end{cases}$$

$$M_{n+1} = \begin{cases} \text{empty} & : \quad \text{if } x_{n+1} = 0 \\ M_n & : \quad \text{if } x_{n+1} > 0,\ x_n > 0 \\ a_n^1 & : \quad \text{if } x_{n+1} > 0,\ x_n = 0 \end{cases}$$

where $a_n^1 \in$ {slow, fast} and $a_n^2 \in \{0, 1\}$ are DM1's and DM2's choices, $A_n \sim \text{Bern}(\lambda)$ (i.e., $A_n = 1$ w.p. λ and $A_n = 0$ otherwise), $U_n | (M_n = \text{slow}) \sim \text{Bern}(\mu_l)$ and $U_n | (M_n = \text{fast}) \sim \text{Bern}(\mu_f)$.

Note that the transition structure described above corresponds to a stochastic game with *perfect information* ([11]), which means that in each state one of the players (decision makers) is restricted to a single action. It is well known that in zero sum, *finite-state* stochastic games with perfect information the players have optimal *nonrandomized* stationary strategies ([11], [9]); it will be argued below that the same holds in the present (countable state) case. For now, we note that DM1 has only two such strategies, which we denote by $\mu_l^{(\infty)}$ and $\mu_f^{(\infty)}$: in the former, slow service mode is always chosen, and the latter chooses fast service always.

The performance measures which will be considered are the throughput $\bar{\Lambda}$ (rate of successfully served customers) and the average service cost $\bar{C}$, i.e.,

$$\bar{\Lambda}_n = \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}\{U_m = 1,\ x_m > 0\}$$

$$\bar{C}_n = \frac{1}{n} \sum_{m=0}^{n-1} \mu_f \mathbf{1}\{M_m = \text{fast}\}$$

$$\bar{r}_n = \left(\bar{\Lambda}_n, \bar{C}_n\right).$$

We first identify explicitly (in Proposition 5.1 below) the set of approachable convex sets in the resulting performance space $\mathbb{R}^2$. Examples of such sets which are of particular interest will then be discussed.

We note first that conditions C1) and C2) of Section II are satisfied: since $\mu_f > \lambda$, it is easily seen that the entire strategy set of either decision maker is uniformly stable (with $s = (0, \text{empty})$ defined as the '0' state). Next, note that DM1's effective decisions are made only in state 0, and the model is of perfect information; it then follows by elementary considerations that either $\mu_l^{(\infty)}$ or $\mu_f^{(\infty)}$ is optimal in each game $\Gamma_0(u)$, $u \in \mathbb{R}^2$. Now, given that DM1 employs $\mu_l^{(\infty)}$ or $\mu_f^{(\infty)}$ in $\Gamma_0(u)$, DM2 is facing a Markov decision process, and by standard results (see, e.g., [2], [25]) there exist optimal stationary strategies (which minimize the average expected cost), say g_l^* and g_f^*, in either case. It follows that an optimal strategy for DM2 in $\Gamma_0(u)$ is to employ g_l^* if $M_n =$ slow, and g_f^* if $M_n =$ fast.

Let us now calculate the sets $R(\mu_l^{(\infty)}, G)$ and $R(\mu_f^{(\infty)}, G)$ [defined below (3.12)], i.e., the range of the expected average payoff which is induced by $\mu_l^{(\infty)}$ or $\mu_f^{(\infty)}$.

Assume that $\mu_l^{(\infty)}$ is employed. Obviously, the service cost is identically zero, i.e., $\bar{C}_n \equiv 0$. Now, the maximal throughput is clearly λ, while the minimal is achieved when the customers never choose to join the system (unless they have to, when the queue is empty). By considering the induced Markov chain (with the state space reduced to the two states $x = 0$ and $x = 1$), this minimal throughput is easily seen to be

$$\Lambda_l \triangleq \frac{\mu_l \lambda}{\mu_l + \lambda}$$

and thus

$$R(l) \triangleq R(\mu_l^{(\infty)}, G) = \left\{(\bar{\Lambda}, 0):\ \Lambda_l \leq \bar{\Lambda} \leq \lambda\right\}.$$

Assume now that $\mu_f^{(\infty)}$ is used. It follows similarly that the throughput is in $[\Lambda_f, \lambda]$, where

$$\Lambda_f \triangleq \frac{\mu_f \lambda}{\mu_f + \lambda}.$$

Moreover, noting that a cost of μ_f units is incurred for each service attempt, it follows that the average expected cost equals the throughput; thus

$$R(f) \triangleq R\left(\mu_f^{(\infty)}, G\right) = \left\{(\bar{\Lambda}, \bar{C}):\ \Lambda_f \leq \bar{\Lambda} \leq \lambda,\ \bar{C} = \bar{\Lambda}\right\}.$$

Let JK be the line segment between the points $J = (\Lambda_f, \Lambda_f)$ and $K = (\Lambda_l, 0)$, and let LM be the line segment between $L = (\lambda, \lambda)$ and $M = (\lambda, 0)$, (see Fig. 3). We then have the following result.

Proposition 5.1: A convex set $B \subset \mathbb{R}^2$ is approachable by DM1 if and only if it intersects both line segments JK and LM; otherwise it is excludable by DM2.

Proof: Recalling that $f^*(u) \in \{\mu_l^{(\infty)}, \mu_f^{(\infty)}\}$ for every $u \in U$, the proof follows from Theorem 3.4, conditions NSC1 and NSC2, by simple geometric considerations as outlined below.

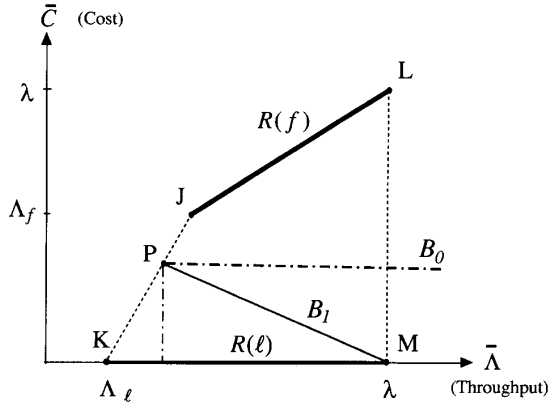


Fig. 3. The performance space, and two approachable sets.

Assume that B satisfies the requirements. It is easily seen that for any $\rho \notin B$, either $R(l)$ or $R(f)$ is separated from ρ by H_ρ. Thus, by NSC1, B is approachable.

Conversely, if B does not intersect either JK or LM, it may be easily verified that there exists a point $\rho \notin B$ such that the required separation does not occur neither for $R(l)$ nor for $R(f)$. Thus, by NSC2 B is not approachable, and by Theorem 3.4-ii) is therefore excludable. □

We now discuss two specific examples of approachable sets.

Example 1: A reasonable objective for the server is to guarantee a throughput above a certain threshold while keeping the average cost below a certain threshold. Thus, we consider the set

$$B_0 = \left\{(\bar{\Lambda}, \bar{C}) \in \mathbb{R}^2 : \bar{\Lambda} \geq \Lambda_0, \bar{C} \leq C_0\right\}$$

where (Λ_0, C_0) are threshold levels to be determined. It follows immediately by Proposition 5.1 that B_0 is approachable if and only if it intersects the line segment JK. Thus, the line segment JK represents the set of Pareto-optimal (undominated) performance vectors (Λ_0, C_0) which may be secured by the server.

Example 2: Let P be a point on JK, and let $M = (\lambda, 0)$. Then, by Proposition 5.1, the set $B_1 \triangleq$ {the line segment PM} is approachable by DM1. We note that this set is minimal, in the sense that no proper convex subset thereof is approachable. This may seem somewhat peculiar at first glance, since the cost on PM is decreasing with increasing throughput. However, this dependence actually reveals an "adaptive" property inherent in the associated approaching strategy. To clarify this point, consider the extreme case where the customers always decide to join the queue. The throughput will obviously be λ, independently of the service mode. Thus, the approaching strategy adapts to this situation (without prior knowledge of the arrival policy) by adhering to the slow service mode, thereby reducing the service cost to 0 (the point M).

In both these examples, the approaching strategy suggested by Theorem 3.1 is obviously nonstationary, since it switches between slow and fast service modes (when in state 0) according to the current average payoff vector. It is important to note that there does *not* exist a stationary approaching strategy in either case, so that dependence on the history is crucial.

VI. Concluding Remarks

The purpose of this paper is to provide an analytic tool for the evaluation of guaranteed performance, from a single controller's viewpoint, in systems where dynamic uncertainty (caused by the actions of other decision makers or disturbances of similar nature) is significant. Our approach is characterized by a worst case viewpoint, and by the explicit consideration of several performance measures, rather than the single "figure of merit" approach which is usually employed in dynamic optimization problems.

The worst-case approach should be contrasted with the "statistical" approach, where a certain (statistical, and usually simplified) model is imposed on the behavior of other decision makers, thus incorporating their actions into the system dynamics. While each approach has its advantages, it is important to realize that the two may be combined by imposing only partial statistical assumptions to yield a more realistic model.

The main results of this paper are Theorem 3.1 and its Corollary 3.3, which give a sufficient condition for any given set to be approachable, and Theorem 3.4 which gives necessary and sufficient conditions for approachability of a *convex* set; in either case, the approaching strategy is specified. These results depend in an essential way on certain recurrence properties of a single fixed state. The verification of the approachability conditions and computation of approaching strategies pose some nontrivial computational problems, tractability of which has been discussed at the end of Section III.

The proposed approaching strategies essentially require our decision maker to monitor system performance (embodied by the current average payoff), in addition to the basic system state, and modify his strategies accordingly. Heuristically, this seems quite reasonable in uncertain dynamic decision problems. It should be stressed that strategy adjustment here is not directly related to learning (as in adaptive control), but rather is required as a response to dynamically changing uncertainties. However, some relations with adaptive behavior were noted at the end of Section V (and see also Remark 2 following Theorem 3.4), and should perhaps be further studied.

We conclude with a few comments regarding possible extensions of these results.

The approaching strategies which were considered here are adapted to the history of the process (i.e., to the relative position of the average payoff compared to the set to be approached) only when the fixed recurrent state is hit. While this is convenient for analysis and easy to implement, it may have the undesirable effect of increasing the "variance" of the payoff if these recurrence times are far apart. It should therefore be of interest to con-

struct approaching strategies which adapt to the current payoff more frequently.

It is quite obvious that some recurrence conditions are required to preserve the basic approach and results of this paper. However, the ones assumed here (namely recurrence of a single fixed state for all relevant strategies) are not the only possibility. Specifically, it is *conjectured* that the basic results hold under the various recurrence conditions considered in [9] (e.g,. when the recurrent state is allowed to depend on the strategies, within a finite set of states).

An interesting problem which has not been directly touched upon is that of partial state information. However, it should be noted that as long as the average reward vector is available, the basic sufficient condition in Theorem 3.1 remains valid even if the strategy set Π_1 of DM1 is limited, either by partial state information or otherwise.

Finally, we note that the main definitions and results of this paper may be straightforwardly generalized to semi-Markov models.

REFERENCES

[1] E. Altman and A. Shwartz, "Nonstationary policies for controlled Markov chains," EE Pub. 633, Technion, June 1987.

[2] ——, "Markov decision problems and state-action frequencies," *SIAM J. Contr. Optimiz.*, vol. 29, no. 4, July 1991.

[3] R. J. Aumann and M. Maschler, "Game theoretic aspects of gradual disarmament," Mathematica, Inc., Princeton, NJ, Rep. to U.S. Arms Control and Disarmament Agency, Contract S.T.80, 1966, ch. V.

[4] D. Blackwell, "An analogue for the minimax theorem for vector payoffs," *Pacific J. Math.*, vol. 6, pp. 1–8, 1956.

[5] ——, "Controlled random walks," in *Proc. Int. Congress Math.*, vol. 3, 1954, pp. 336–338.

[6] ——, "On multicomponent attrition games," *Naval Res. Log. Quart.*, pp. 210–216, 1954.

[7] V. S. Borkar, "Control of Markov chains with long-run average cost criterion," *Stochastic Differential Systems, Stochastic Control and Application*, W. Fleming and P. L. Lions, Eds. New York: Springer-Verlag, IMA vol. 10, pp. 57–77, 1988.

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

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

[10] ——, "Successive approximation methods in undiscounted stochastic games," *Operations Research*, vol. 28, pp. 794–810, 1980.

[11] D. Gillette, "Stochastic games with zero stop probabilities," in *Contributions to the Theory of Games, III* (Annals of Math. Studies 39), M. Dresher *et al.*, Eds. Princeton, NJ: Princeton Univ. Press, 1957, pp. 179–188.

[12] S. Hart, "Nonzero-sum two-person repeated games with incomplete information," *Math. of Oper. Res.*, vol. 10, pp. 117–153, 1985.

[13] J. F. Hannan, "Approximation to Bayes risk in repeated play," in *Contributions to the Theory of Games, III* (Annals of Math. Studies 39), M. Dresher *et al.*, Eds. Princeton, NJ: Princeton Univ. Press, 1957, pp. 97–139.

[14] T. F. Hou, "Weak approachability in a two-person game," *Ann. Math. Stat.*, vol. 40, pp. 789–813, 1969.

[15] ——, "Approachability in a two-person game," *Ann. Math. Stat.*, vol. 42, pp. 735–744, 1971.

[16] M. Katz, "Infinitely repeatable games," *Pac. J. Math.*, vol. 10, pp. 879–885, 1960.

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

[18] A. Maitra and T. Parthasarathy, "On stochastic games, II," *J. Optimiz. Theory Appl.*, vol. 8, pp. 155–160, 1971.

[19] T. Parthasarathy and M. Stern, "Markov games: A survey," *Differential Games and Control Theory*, P. L. E. Roxin and R. Sternberg, Eds. New York: Marcel-Dekker, 1977.

[20] T. E. S. Raghavan and J. A. Filar, "Algorithms for stochastic games—A survey," *Zeit Oper. Res.*, vol. 35, pp. 437–472, 1991.

[21] T. E. S. Ragahavan, T. S. Federgruen, T. Parthasarathy, and O. J. Vrieze, Eds., *Stochastic Games and Related Topics*. The Netherlands: Kluwer, 1991.

[22] R. T. Rockafellar, *Convex Analysis*. NJ: Princeton Univ. Press, 1970.

[23] S. M. Ross, *Applied Probability Models with Optimization Applications*. San Francisco: Holden-Day, 1970.

[24] H. Sackrowitz, "A note on approachability in a two-person game," *Ann. Math. Stat.*, vol. 43, pp. 1017–1019, 1972.

[25] L. I. Sennott, "Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs," *Operations Research*, vol. 37, pp. 626–633, 1989.

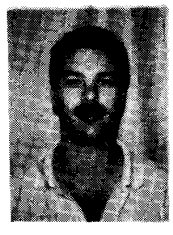
[26] A. N. Shiryayev, *Probability*. New York: Springer-Verlag, 1984.

[27] J. Sorin, "An introduction to two-person-zero-sum repeated games with incomplete information, IMSSS-Economics Tech. Rep.-312, Stanford Univ., memo.

[28] M. A. Stern, "On stochastic games with limiting average pay off," Ph.D. dissertation, Univ. of Illinois, Circle Campus, Chicago, submitted 1975.

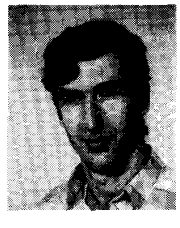
[29] N. Vieille, "Weak approachability," preprint, 1991.

[30] J. Walrand, *An Introduction to Queueing Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1988.



Nahum Shimkin (S'88–M'91) was born on June 6, 1960. He received the B.Sc. degree in 1980, and the M.Sc. and Ph.D. degrees in 1988 and 1991, respectively, from the Technion—Israel Institute of Technology, Israel, all in electrical engineering.

From 1980 to 1985 he served as a control engineer in the Israeli Defense Forces. During 1991–1992 he was on the faculty of the Department of Electrical Engineering at Technion. He now holds a postdoctoral position at the Institute for Mathematics and its Applications, University of Minnesota, Minneapolis. His research interests include stochastic and adaptive control, queueing system, and stochastic games.



Adam Shwartz (SM'89) received the B.Sc. degree from Ben-Gurion University, Israel, in 1979, and the Ph.D. degree from Brown University, Providence, RI, in 1983.

Since 1984, he has been with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Israel. He held a visiting position at Brown University in applied mathematics from 1982–1983 and in electrical engineering from both Brown and the Systems Research Center, University of Maryland, College Park, from 1983–1984 and 1990–1991, respectively. He has been a Consultant with AT & T Bell Laboratories, Murray Hill, since 1987. His research interests include the theory of stochastic processes with applications to computer and communications networks. His most recent contributions concern optimization of controlled Markov models and performance analysis through large deviations.