Stochastic Games with Average Cost Constraints

Nahum Shimkin*
Institute for Mathematics and its Applications
University of Minnesota
Minneapolis, MN 55455, USA

Revised, Jan. 1993†

Abstract

The paper introduces the model of undiscounted stochastic games augmented by side constraints on the levels of a player's average costs. Several variants of this problem are suggested, and the special case of zero-sum games with constraints on one side is analyzed; under certain recurrence conditions, existence of the value and (non-stationary) optimal strategies are established.

1 Introduction

In undiscounted stochastic games the purpose of each player is to maximize his average reward, which is the long-term average of a scalar reward sequence. In this paper we consider the situation where, in addition, a player must obey some side constraints on average costs; that is, the player incurs certain costs on each stage, and must verify that the long-term average costs will not exceed pre-specified thresholds.

Dynamic optimization problems with average cost constraints have attracted considerable attention in the context of Markov decision processes (e.g. [7, 9, 2, 10, 14, 1, 5]). Indeed, such constraints follow naturally in various control problems, ranging from economics problems where resources are limited and supplied at a given rate, to communication networks with average delay constraints. The incorporation of such constraints in stochastic games is a natural extension of this line of work, and is similarly motivated.

From a strategic viewpoint, the average cost constraints effectively limit each player to a certain sub-class of his strategies, which generally depends on the strategies employed by other

*Part of this work was performed when the author was with the department of Electrical Engineering at the Technion - Israel Institute of Technology

†Appeared in: *Annals of the International Society of Dynamic Games, vol. 1: Advances in Dynamic Games and Applications*, T. Basar and A. Haurie (Eds.), Birkhauser, 1993

players. Constrained games of this nature have been first considered in [6]; and see also [12, 11]. In the context of dynamic games, a related model is considered in [13], where a two-person repeated matrix game with the average payoff criterion is augmented by the possibility of ruin which occurs if a player's initial wealth decreases to 0.

The general model of undiscounted stochastic games with average cost constraints is introduced formally in Section 2. Several variants are suggested and briefly discussed. For simplicity, only the case of two players and finite state and action spaces are considered. Section 3 presents the analysis of a particular case, namely the zero-sum game with constraints on one side only. Under certain recurrence conditions, existence of a value and optimal strategies for both players are established. The optimal strategy for the constrained player turns out to be non-stationary in an essential way. The analysis relies on the ideas of Blackwell's approachability theory [4], extended to the stochastic game context. The paper is concluded with some remarks and problems for additional research.

2 The General Model

A standard two-person stochastic game is defined by the following elements. The player set $I = \{1,2\}$; a finite state-space S; a finite action space A^i for each player $i \in I$; transition law p, where $p(s'|s,a^1,a^2)$ is the probability of s' being the next state given the present state s and actions $a^1 \in A^1$, $a^2 \in A^2$; and one-step reward functions $r^i : S \times A^1 \times A^2 \to \mathbf{R}, \quad i \in I$. The game is played in stages $n = 1,2,\cdots$. Let s_n and (a_n^1, a_n^2) denote the state and actions at stage n, and let $r_n^i \triangleq r^i(s_n, a_n^1, a_n^2)$ denote the reward of player i at that stage. $\bar{r}_n^i \triangleq n^{-1}\sum_{k=1}^n \bar{r}_k^i$ denotes the average reward up to stage n. Let Π^i be the set of behavioral strategies for player i; thus in each strategy $\pi^i \in \Pi^i$ the choice of action a_n^i at stage n may depend on the entire history $h_n = (s_1, a_1^1, a_2^1, \ldots\ s_n)$, and may include randomizations which are performed independently at each stage. Let F and G denote the set of stationary strategies for players 1 and 2, respectively. A strategy $f \in F$ may be identified with a function $f : S \to \mathcal{P}(A^1)$, where $\mathcal{P}(A^1)$ is the set of probability vectors over A^1, and similarly for $g \in G$. Let $P_{\boldsymbol{\pi}}^s$ and $E_{\boldsymbol{\pi}}^s$ be the probability measure and expectation associated with the strategy pair $\boldsymbol{\pi} = (\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and initial state $s_1 = s \in S$. The initial state will henceforth be assumed to be fixed, and will be omitted from the notation. The undiscounted (limit expected average) reward functional for each player $i \in I$ is now defined as

$$R^i(\boldsymbol{\pi}) = \liminf_{n\to\infty} E_{\boldsymbol{\pi}}(\bar{r}_n^i) \ . \tag{2.1}$$

We note that several other definitions for the undiscounted reward are possible and have been considered in the literature (see, e.g., [3]). For concreteness, only the limit expected average reward criterion will be used here.

We proceed to introduce the cost constraints into the model. Let $c^i : S \times A^1 \times A^2 \to \mathbf{R}^{k^i}$, $k^i \geq 1$, be a (vector valued) cost function for player $i \in I$. Let $\gamma^i \in \mathbf{R}^{k^i}$ be the allowed *cost level* vector for player i. Generally speaking, player i must keep his average costs below the maximal cost levels. The exact statement of this constraint may lead to several different definitions of the payoff, as discussed below.

a. Expected-Cost Constraints:

Let $\bar{c}_n^i = n^{-1} \sum_{k=1}^{n} c^i(s_k, a_k^1, a_k^2)$ denote the average cost for player i up to stage n, and let

$$C^i(\boldsymbol{\pi}) \triangleq \limsup_{n\to\infty} E_{\boldsymbol{\pi}}(\bar{c}_n^i) \ , \quad \boldsymbol{\pi} \in \Pi^1 \times \Pi^2 \ , \tag{2.2}$$

(all limits of vector-valued sequences encountered in this paper are taken over each component separately). Define the following payoff function for player i:

$$\phi^i(\boldsymbol{\pi}) = \begin{cases} R^i(\boldsymbol{\pi}) & \text{if} \quad C^i(\boldsymbol{\pi}) \le \gamma^i \\ -\infty & \text{if} \quad C^i(\boldsymbol{\pi}) \not\le \gamma^i \quad . \end{cases} \tag{2.3}$$

The payoff functions ϕ^1 and ϕ^2 define a strategic-form game over the strategy sets Π^1, Π^2. The Nash equilibrium point for this game may now be defined in a standard manner. This definition of the payoff functions has the following interpretation: player i is primarily concerned with keeping his average expected costs (2.2) below the allowed level γ^i. A violation of these constraints is considered the worst possible outcome for him, as manifested by a payoff of $-\infty$. Otherwise, his payoff is simply given by his average expected reward.

b. Sample Path Constraints:

The expected-cost formulation has the appealing property and both the average reward and the average cost have a similar form. However, such constraints do not exclude the possibility that the actual (sample-path) average costs will exceed the allowed cost levels with positive probability. If this is not acceptable, one is led to the following sample-path formulation. Let

$$\bar{c}_\infty^i \triangleq \limsup_{n\to\infty} \bar{c}_n^i \tag{2.4}$$

denote the (upper) limit average cost vector. Define the payoff function for player i by

$$\phi^i(\boldsymbol{\pi}) = \begin{cases} R^i(\boldsymbol{\pi}) & \text{if} \quad P_{\boldsymbol{\pi}}\{\bar{c}_\infty^i \le \gamma^i\} = 1 \ . \\ -\infty & \text{if} \quad P_{\boldsymbol{\pi}}\{\bar{c}_\infty^i \le \gamma^i\} < 1 \ . \end{cases} \tag{2.5}$$

The payoff now reflects the fact that the allowed cost levels must be obeyed with probability 1.

We note that definitions (2.3) or (2.5) lead the standard definition of Nash equilibrium for constrained games [6, 12]. These definitions may be extended in a straightforward manner to n-person games.

c. Constraints with Ruin:

It has been assumed so far that the violation of a player's constraints affects only his own payoff, while the payoff for the other player is not directly affected. However, it is possible to interpret the violation of a player's constraints as leading to his ruin, or elimination from the game, thereby leading to an increased payoff for the other player. In the extreme case, the ruin of the other player may be a most desired result (cf. [13]). This obviously leads to the following payoff functional for player 1 (where expected-cost constraints are used):

$$\phi^1(\boldsymbol{\pi}) = \begin{cases} R^1(\boldsymbol{\pi}) & \text{if } \; C^1(\boldsymbol{\pi}) \le \gamma^1,\, C^2(\boldsymbol{\pi}) \le \gamma^2 \\ -\infty & \text{if } \; C^1(\boldsymbol{\pi}) \not\le \gamma^1 \\ +\infty & \text{if } \; C^1(\boldsymbol{\pi}) \le \gamma^1,\, C^2(\boldsymbol{\pi}) \not\le \gamma^2 \end{cases} \tag{2.6}$$

The payoff for player 2 may be defined symmetrically. We note that the last definition of payoffs is not directly extendible to games of more than two players. However, it may capture the essence of some two-person conflicts, especially in cases of strict competition. We further note that this definition may be modified in an obvious manner to incorporate sample-path constraints (see the next section).

Our discussion has been limited to the case where the constraints are strict, in the sense that their violation is the worst possible outcome for the player irrespective of his reward. Obviously, one may think of different situations; for example, a player may pay a fixed penalty for violating his constraints, but may consider this worthwhile if appropriately compensated by an increased reward. More generally, the payoff (utility) for a player may be defined through some (partial) order on the multi-dimensional payoff/cost space. Such generalizations fall outside the scope of constrained games, and will not be considered in detail in this paper.

3 The Zero-Sum Game with Constraints on One Side

3.1 Model and Recurrence Conditions

This section focuses on the analysis of a special class of constrained stochastic games. Namely, we consider the case of zero-sum rewards ($r^1 = -r^2$), and constraints with ruin (in the sense of (2.6)) on player 1 only. For concreteness, we restrict attention to sample-path constraints (essentially the same analysis applies for expected cost constraints). Thus, the payoff ϕ^1 for player 1 is given by (2.5), with $i = 1$, and for player 2 by $\phi^2 = -\phi^1$. The game is said to have a value v if

$$\sup_{\pi^1} \inf_{\pi^2} \phi^1(\pi^1, \pi^2) = \inf_{\pi^2} \sup_{\pi^1} \phi^1(\pi^1, \pi^2) \stackrel{\triangle}{=} v\,.$$

A strategy of player 1 is the *optimal* if it guarantees $\phi^1 \ge v$ (i.e., it achieves the supremum on the left). A strategy of player 2 is optimal if it guarantees $\phi^1 \le v$, and is ϵ-*optimal* if it guarantees $\phi^1 \le v + \epsilon$.

The avoid the case where $v = -\infty$ trivially, we assume the following.

Assumption A1: For every strategy of player 2, player 1 has a strategy which satisfies his constraints (w.p. 1).

The conditions for v and π^{1*}, π^{2*} to be the value and optimal strategies of player 1 and 2, respectively, are now equivalent to the following two requirements:

(i) $P_{(\pi^{1*},\pi^2)}\{\bar{c}^1_\infty \leq \gamma^1\} = 1$ for every $\pi^2 \in \Pi^2$.

(ii) $R^1(\pi^1, \pi^{2*}) \leq v \leq R^1(\pi^{1*}, \pi^2)$ for every $\pi^2 \in \Pi^2$, and for every $\pi^1 \in \Pi^1$ which satisfies $P_{(\pi^1,\pi^{2*})}\{\bar{c}^1_\infty \leq \gamma^1\} = 1$.

Existence of a value and optimal strategies will be established under the following recurrence assumption:

Assumption A2: There exists a fixed state, say state 0, which is accessible from any other state in the Markov chain induced by any pair of stationary non-randomized strategies. (In other words, each such Markov chain includes exactly one recurrent class, and state 0 is included in that class.)

3.2 Approachability for Stochastic Games

An approachability theory for repeated matrix games with vector payoffs was introduced in [4]. The basic results of this paper have been generalized to the stochastic games context in [15]. We review below some of these results, which will be applied in the next sub-section to solve the constrained game problem.

Consider the stochastic game model of Section 2, with a fixed initial state s, and the recurrence assumption **A2**. Introduce a vector-valued payoff function $m : S \times A^1 \times A^2 \to \mathbf{R}^b$, $b \geq 2$. Denote $m_n = m(s_n, a^1_n, a^2_n)$, $\bar{m}_n = n^{-1}\sum_1^n m_k$, and let $d(\cdot,\cdot)$ denote the Euclidean point-to-set distance in $\mathbf{R}^b$.

Definition: A set $B \in \mathbf{R}^b$ is *approachable* (by player 1 from initial state s) if there exists a *B-approaching strategy* $\pi^1 \in \Pi^1$ for which

$$\lim_{n\to\infty} d(\bar{m}_n, B) = 0 \ \text{ a.s.}, \quad \forall \pi^2 \in \Pi^2\,,$$

i.e., all limit points of $\{\bar{m}_n\}$ are in B.
A set B is *excludable* (by player 2 from s) if there exists a *B-excluding strategy* $\pi^2 \in \Pi^2$ and a constant $\delta > 0$ for which

$$\liminf_{n\to\infty} d(\bar{m}_n, B) \geq \delta \ \text{ a.s.}, \quad \forall \pi^1 \in \Pi^1\,,$$

i.e., all limit points of $\{\bar{m}_n\}$ are strictly outside B.

We now define an auxiliary family of stochastic games. Let U denote the set of unit vectors in $\mathbf{R}^b$, and let $\langle\cdot,\cdot\rangle$ denote the Euclidean inner product. For each $u \in U$, define $\Gamma(u)$ as the undiscounted (limit expected average) zero-sun stochastic game with the same dynamics as above, a *scalar* payoff function $m^u \triangleq \langle m, u \rangle$ (the projection of m in the direction of u), and player 1 the maximizer. Under assumption **A2**, each game $\Gamma(u)$ admits a value and stationary optimal strategies ([16, 8]); let $\mathrm{val}\,\Gamma(u)$ denote its value, and let $f^*(u)$ and $g^*(u)$ denote the stationary optimal strategy of player 1 and 2, respectively, in $\Gamma(u)$.

For every pair (f, g) of stationary strategies let

$$M(f,g) = \lim_{n\to\infty} E_{f,g}(\bar{m}_n)$$

(existence of the limit follows from **A2**), and define

$$M(F,g) = \{M(f,g) : f \in F\},$$

which is the set of all average payoff vectors generated by stationary strategies of player 1 when player 1 employs g. It is well known ([7, p. 95]) that $M(F, g)$ is a closed convex set under **A2**.

The next theorem, which follows from Theorem 3.4 and Lemma 2.1 in [15], provides a complete characterization of approachability and excludability for convex sets.

Theorem 3.1 *Assume* **A2**. *Let B be a closed* convex *set in $\mathbf{R}^b$. Then*

(i) B is approachable if and only if

$$\mathrm{val}\,\Gamma(u) \geq \inf_{\beta\in B} \langle \beta, u \rangle\,, \quad \forall u \in U\,.$$

(ii) B is excludable if and only if $M(F, g) \cap B = \emptyset$ for some $g \in G$; in that case, the stationary strategy g is B-excluding.

(iii) B is either approachable or excludable (but obviously not both).

(iv) If approachable, then a B-approaching strategy for player 1 is given by the following 'piece-wise stationary' strategy. At each time τ when state 0 is reached, a stationary strategy to be played until the next such time is determined as follows: compute the current average payoff $\bar{m}_\tau$; if $\bar{m}_\tau \notin B$, then play $f^*(u_\tau)$, where u_τ is a unit vector in the direction from $\bar{m}_\tau$ to the closest point in B; otherwise, play any strategy.

3.3 Main Results

Our main result for the game of Section 3.1 is the following.

Theorem 3.2 *Assume* **A1-A2**. *Define the vector-valued payoff function $m \triangleq (r^1, c^1)$, and let ρ^* be the maximal number ρ such that the convex set*

$$B(\rho) = \{(\bar{r}^1, \bar{c}^1) \in \mathbf{R}^{1+k^i} : \bar{r}^1 \geq \rho\,,\, \bar{c}^1 \leq \gamma^1\}$$

is approachable. Then

(i) The game has a finite value, which equals ρ^*.

(ii) An optimal strategy for player 1 is given by any $B(\rho^*)$-approaching strategy.

(iii) For player 2 (the unconstrained player), an ϵ-optimal stationary *strategy exists for any $\epsilon > 0$.*

Proof: We first establish that ρ^* is well defined and finite. Note first that the set $B(-\infty) = \{(\bar{r}^1, \bar{c}^1) : \bar{c}^1 \leq \gamma^1\}$ is approachable; this follows by assumption **A1**, which implies that $B(-\infty)$ is not excludable, and by Theorem 3.1(ii). Since the reward function r^1 is bounded, it follows immediately that $B(\rho)$ is approachable for ρ small enough, and not approachable for ρ large enough. Thus, the set $K = \{\rho : B(\rho) \text{ approachable}\}$ has a finite supremum ρ^*. Finally, it follows readily from Theorem 3.1(i) that K is closed (see also the proof of Corollary 3.1 below). Thus ρ^* is the maximal element and $B(\rho^*)$ is approachable.

To prove the theorem, it is now sufficient to establish that: (i) any $B(\rho^*)$-approaching strategy of player 1 guarantees a payoff $\phi^1 \geq \rho^*$, and (ii) for any $\epsilon > 0$ there exists a stationary strategy of player 2 which guarantees a payoff $\phi^1 \leq \rho^* + \epsilon$. Assume first that player 1 is using π^{1*}, a $B(\rho^*)$-approaching strategy. By definition of approachability and of the set $B(\rho^*)$, it follow that for every strategy of player 2:

$$\limsup_{n\to\infty} \bar{c}_n^1 \leq \gamma^1 \ \text{ a.s.}, \quad \liminf_{n\to\infty} \bar{r}_n^1 \geq \rho^* \ \text{ a.s.},$$

and by Fatou's lemma the last inequality implies that $R^1 \geq \rho^*$. Thus π^{1*} guarantees for player 1 a payoff of ρ^* at least.

Next we identify an ϵ-optimal strategy for player 2. Fix $\epsilon > 0$. By our definition of ρ^*, the convex set $B(\rho^* + \epsilon)$ is not approachable, hence by Theorem 3.1 it is excludable, and there exists an excluding stationary strategy g^ϵ. (Note that this is not true for $\epsilon = 0$, since $B(\rho^*)$ is approachable.) Assume that player 2 employs the strategy g^ϵ, and player 1 any strategy π^1. Then by definition of excludability, all limit points of $\bar{m}_n = (\bar{r}_n^1, \bar{c}_n^1)$ are (almost surely) outside the set $B(\rho^* + \epsilon)$, hence contained in the set $\{\bar{r}^1 \leq \rho^* + \epsilon \text{ or } \bar{c}^1 \not\leq \gamma^1\}$. This implies that either (i) with positive probability, some limit point of $\bar{c}_n^1$ is in $\{\bar{c}^1 \not\leq \gamma^1\}$, or ($ii$) almost surely, all limit points of $\bar{r}_n^1$ are in $(-\infty, \rho^* + \epsilon]$. In case ($i$), we have ($\limsup_n \bar{c}_n^1 \not\leq \gamma^1$) with positive probability, hence the payoff ϕ^1 equals $-\infty$ by its definition. In case (ii) we have ($\limsup_n \bar{r}_n^1 \leq \rho^* + \epsilon$ a.s.), which implies $R^1 \leq \limsup_n E(\bar{r}_n^1) \leq \rho^* + \epsilon$, and the payoff is again not larger than $\rho^* + \epsilon$. Thus, the stationary strategy g^ϵ is ϵ-optimal for player 2. □

Using the above characterization, a more explicit expression may be obtained for the value of the game.

Corollary 3.1 *The value of the game is given by*

$$\rho^* = \min_{x \leq 0} \{\mathrm{val}\, \Gamma((1, x)) - \langle \gamma^1, x \rangle\}.$$

Here x ranges over all vectors with non-positive entries in $\mathbf{R}^{k^1}$, and $(1, x)$ is the vector in $\mathbf{R} \times \mathbf{R}^{k^1}$ obtained from x by adding 1 as the first entry.

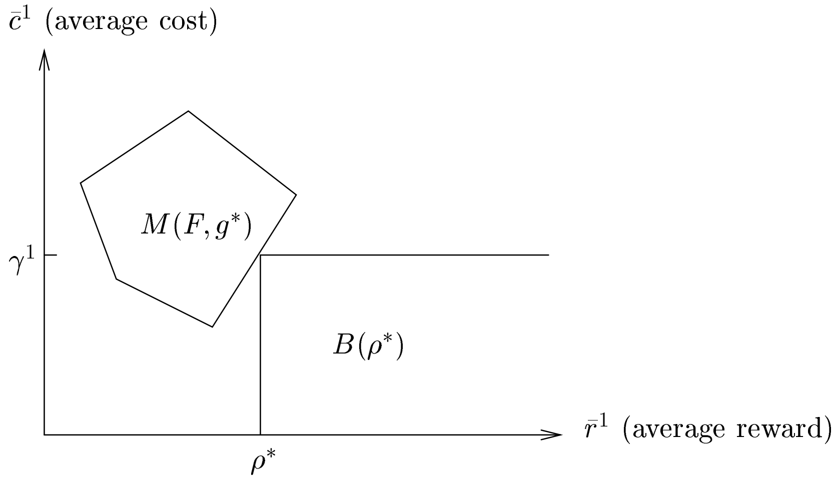

Figure 1: The geometric picture.

Proof: Partition each unit vector as $u = (u_1, u_2) \in \mathbf{R} \times \mathbf{R}^{k^i}$. For every real ρ, the set $B(\rho)$ is approachable if and only if

$$\operatorname{val}\Gamma(u) \geq \inf_{\beta \in B(\rho)} \langle \beta, u \rangle \tag{3.7}$$

is satisfied for every unit vector $u = (u_1, u_2)$ (Theorem 3.1(i)). Now, it may be verified that the right hand side of (3.7) is larger than $-\infty$ if and only if $u_1 \geq 0$ and $u_2 \leq 0$ (component-wise), and is then equal to $\langle (u_1, u_2), (\rho, \gamma^1) \rangle = u_1 \rho + \langle u_2, \gamma^1 \rangle$. Thus $B(\rho)$ is approachable if and only if

$$\operatorname{val}\Gamma((u_1, u_2)) \geq u_1 \rho + \langle u_2, \gamma^1 \rangle$$

for every such $u = (u_1, u_2)$. The case of $u_1 = 0$ need not be considered since ρ is canceled (the inequality is then satisfied as a consequence of **A1**). Thus, dividing by $u_1 > 0$ and noting that $x = u_2/u_1$ ranges over the entire set of vectors with non-positive entries, we get that $B(\rho)$ is approachable if and only if

$$\operatorname{val}\Gamma((1, x)) \geq \rho + \langle x, \gamma^1 \rangle \quad \forall\, 0 \geq x \in \mathbf{R}^{k^i} .$$

Since ρ^* is the largest ρ which satisfies this condition, the required expression is obtained. $\square$

Remarks:

1. An optimal stationary strategy of player 2 can be given a nice geometric interpretation. Consider for simplicity the case of a single (scalar) cost constraint. An optimal stationary strategy g^* of player 2 is such that $M(F, g^*)$ just touches the boundary of $B(\rho^*)$. This is illustrated in Fig. 1. The set $M(F, g^*)$ represents all possible limit points of $(\bar{r}^1_n, \bar{c}^1_n)$ which player 1 can attain against g^* (indeed, interchanging the roles of the players in the definition of approachability, it easily follows from Theorem 3.1 that the set $M(F, g^*)$ is approachable by player 2). Thus, the figure clearly illustrates the interplay between reward and cost: although player 1 may be able to obtain a reward larger than ρ^*, then against g^* this can only be done at the expense of violating his cost constraint γ^1.

2. In this geometric picture, we may interpret the value ρ^* as follows. Considering the set $B(\rho)$, we translate this set to the right (by increasing ρ continuously), and stop just before it becomes disjoint with a set $M(F, g')$ for some g'. Then this point corresponds to ρ^*, and g' is an optimal strategy for player 2.

3. In general, an optimal or even ϵ-optimal strategy for player 1 does not exist within the class of stationary strategies. In fact, with two constraints or more player 1 may have to use a non-stationary (approachability) strategy even just to guarantee that his constraints are met.

4. Although player 2 does have ϵ-optimal stationary strategy, an optimal one may fail to exist (an example follows). We suspect however that generically it does. An obvious candidate for an optimal stationary strategy is the limit of any converging subsequence of the ϵ-optimal set $\{g^\epsilon\}$, as $\epsilon \to 0$. (Note that such a limit point exists since the strategy set G is compact. Also, it may be easily verified as in [7, Theorem 7.3] that under Assumption A2 the payoff vector $M(f, g)$ is continuous in g for each f.)

5. Here is a simple example where an optimal stationary strategy for player 2 does not exist. We assume only one state (a repeated matrix game), with two actions for each player which are also the pure stationary strategies, denoted f_0, f_1 (for player 1) and g_0, g_1. The corresponding payoff vectors in the combined reward-cost space are $M(f_0, g_0) = (-1, 0)$, $M(f_1, g_0) = (1, 0)$, $M(f_0, g_1) = (0, -1)$, $M(f_1, g_1) = (1, 1)$, and the (scalar) constraint level is $\gamma^1 = 0$, It may be verified geometrically that the value is $v = 0$, and the strategy $g_\epsilon = (1-\epsilon)g_0 + \epsilon g_1$ is $(\epsilon/2)$-optimal (it guarantees $\phi^1 \leq \epsilon/2$). Obviously $g_\epsilon \to g_0$, but the latter guarantees only $\phi^1 \leq 1$.

6. The value and optimal strategies for both players as specified above do not depend on the initial state s.

7. The results above are not limited to the specific undiscounted reward functional (2.1), namely the limit expected average reward. Indeed, it was established in the proof that each player can guarantee an average reward of ρ^* in an almost-sure sense, namely the optimal strategy of player 1 guarantees $\liminf_n \bar{r}_n^1 \geq \rho^*$ (a.s.), while the optimal strategy of player 2 guarantees that $\limsup_n \bar{r}_n^1 \leq \rho^*$ (a.s.) for every strategy of player 1 which satisfies the constraints. Thus the main results remain valid under most common undiscounted reward criteria, such as the limit average criterion or the limit discount criterion (in the terminology of [3]).

8. It may also be shown that the main results hold if the sample-path constraints are replaced by (or combined with) expected cost constraints, as defined in Section 2.

9. The definition of approachability which was given here did not refer to the rate of convergence. In fact, it may be shown that the approaching strategy of Theorem 3.1 ensures the

a.s. convergence of $\bar{m}_n$ to B *at a uniform rate over all strategies of the other player* [15]. That is, for any $\epsilon > 0$ there exists N such that

$$P\{d(\bar{m}_n, B) \leq \epsilon \;\forall n \geq N\} \geq 1 - \epsilon\ ,$$

for every strategy of player 2. This may be used in the present context to strengthen the results obtained above. Thus, if player 1 is using the strategy of Theorem 3.1 to approach $B(\rho^*)$, then for every $\epsilon > 0$ there exists N such that

$$P\{\bar{r}_n^1 \geq \rho^* - \epsilon,\, \bar{c}_n^1 \leq \gamma^1 + \epsilon e,\, \forall n \geq N\} \geq 1 - \epsilon$$

(where e is a vector with 1 at each entry) for every strategy of player 2. Similar comments apply with respect to the convergence of $(\bar{r}_n^1, \bar{c}_n^1)$ to $M(F, g^*)$ under g^*.

In conclusion, for the zero-sum constrained game of this section we have established the existence of the value and optimal strategies for both players. While the unconstrained player has an optimal stationary strategy, the constrained player benefits by using non-stationary strategies, and his optimal strategy is in fact an approaching strategy with respect to some set in the reward-cost space.

Obviously, in this paper we have just touched upon the analysis of the general classes of constrained stochastic games as defined in Section 2. We mention some open problems for research. In the zero-sum game of Section 3, it should be interesting to establish existence of the value after removing, or weakening, the recurrence assumption **A2**. We note that the game of Section 3 (constrains with ruin, on one player only) is the only case which has a zero-sum version; all others are essentially non zero-sum (even if the reward functions are zero-sum) due to the constraints. Thus, one may seek to characterize the set of Nash equilibria for these games (folk theorem). If constraints with ruin exist on both sides, then the above analysis implies that equilibrium strategies will be essentially non-stationary for both players, even under recurrence assumptions similar to **A2**. However, for constraints without ruin the existence of equilibrium points in stationary strategies is open for investigation.

References

[1] E. Altman and A. Shwartz, "Markov decision problems and state-action frequencies," *SIAM J. Control and Optimization* **29** No. 4, July 1991.

[2] F. J. Beutler and K. W. Ross, "Optimal policies for controlled Markov chain with a constraint," *J. Math. Anal. Appl.* **12**, pp. 236-256, 1985.

[3] T. Bewly and E. Kohlberg, "On stochastic games with stationary optimal strategies," *Mathematics of Operations Research* **3**, pp. 104-125, 1978.

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

[5] V. S. Borkar, "Controlled Markov chains with constraints," *Proceedings of the Workshop on Recent Advances in Modeling and Control of Stochastic Systems*, Bangalore, Jan. 1991.

[6] G. Debreu, "A social equilibrium existence theorem," *Proceedings of the National Academy of Science* **38**, pp. 886-893, 1952.

[7] C. Derman, *Finite State Markovian Decision Processes*, Academic Press, New-York, 1970.

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

[9] A. Hordijk and L. C. M. Kallenberg, "Constrained undiscounted stochastic dynamic programming," *Mathematics of Operations Research* **9**, pp. 276-289, 1984.

[10] D.-J. Ma, A. M. Makowski and A. Shwartz, "Estimation and optimal control for constrained Markov chains," *Proc. 25th IEEE Conf. on Decision and Control,* pp. 994-999, 1986.

[11] H. Meister, *The Purification Problem for Constrained Games with Incomplete Information*, (Lecture notes in economics and mathematical systems; 295) Springer, Berlin, 1987.

[12] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica* **33**, pp. 520-534, 1965.

[13] R. W. Rosenthal and A. Rubinstein, "Repeated two-player games with ruin," *International Journal of Game Theory* **13**, pp. 155-177, 1984.

[14] K. W. Ross and R. Varadarajan, "Markov decision processes with sample path constraints: the communicating case," *Operations Research* **37**, pp. 780-790, 1989.

[15] N. Shimkin and A. Shwartz, "Guaranteed performance regions in Markovian systems with competing decision makers ," to appear in *IEEE Trans. Automat. Contr.,* Feb. 1993.

[16] M. A. Stern, *On Stochastic Games with Limiting Average Pay-Off.* Ph.D. dissertation, University of Illinois, Circle Campus, Chicago, 1975.